Selection and Optimization of Gene Targets for the Metabolic Engineering of \textit{E. coli}

by

Curt R. Fischer

B.S.E., Case Western Reserve University (2001)
M.E., Tokyo Institute of Technology (2003)

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 2009

© Massachusetts Institute of Technology 2009. All rights reserved.

Certified by ................... Gregory N. Stephanopoulos
Willard Henry Dow Professor of Biotechnology and Chemical Engineering
Thesis Supervisor

Accepted by .......................... William M. Deen
Professor of Chemical Engineering
Chairman, Committee for Graduate Students
Selection and Optimization of Gene Targets for the Metabolic Engineering of *E. coli*

by

Curt R. Fischer

Submitted to the Department of Chemical Engineering on May 15, 2009, in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Chemical Engineering

Abstract

This thesis is about identifying genetic interventions that improve the performance of targeted pathways in the metabolism of the bacterium *Escherichia coli*. Three case studies illustrate three disparate approaches to identifying genetic interventions: (i) combining metabolomic measurements with thermodynamic calculations to identify rate-limiting reaction steps in a target pathway; (ii) use of stoichiometric, optimization-based models of metabolism to predict target genetic interventions in silico; and (iii) the mutagenesis of promoter sequences to fine-tune the expression level of rate-limiting genes. These techniques can be classified by both the number of strain modifications created, and the number of variables measured in each. Taken together, the cases suggest that the best methods for identifying genetic interventions balance the number of strain modifications with the number of measured variables.

The first case is butyrate production in recombinant *E. coli*. A strain of *E. coli* deleted for the production of lactate, ethanol, and acetate was designed to minimize competing pathways for carbon, and was unexpectedly found to exhibit oxygen auxotrophy. Expression of genes from *Clostridium acetobutylicum* resulted in production of 3-hydroxybutyric acid, but not butyric acid. The clostridial genes *ptb* and *buk* were capable of producing S-3-hydroxybutyric acid from the butyrate pathway intermediate metabolite S-3-hydroxybutyryl-CoA. In parallel, the intracellular concentrations of pathway metabolites was measured for a set of strains expressing the clostridial butanol biosynthesis pathway in various configurations. Comparison of measured pool sizes and pool sizes for thermodynamic equilibrium pinpointed the butyryl-CoA dehydrogenase step, encoded by *bcd*, as a bottleneck enzyme. Thus, points for genetic intervention are *ptb*, *buk*, and *bcd*.

The second case is tyrosine overproduction in *E. coli*. Constraints-based models of *E. coli* metabolism proved incapable of predicting gene knockout targets. Therefore, to
understand factors underlying tyrosine overproduction, the intracellular concentrations of amino acids were measured. In tyrosine overproducers, the intracellular concentrations of most proteinogenic amino acids were vastly perturbed relative to non-producing strains. This fact and thermodynamic considerations suggested that the transamination of $p$-hydroxyphenylpyruvate to tyrosine was near equilibrium, and thus that nitrogen supply may be limiting tyrosine production. Culture media amended with glutamate or glutamine, but not with $\alpha$-ketoglutarate or other organic acids, increased tyrosine production in these strains more than 8-fold, showing that interventions which affect nitrogen supply are attractive targets for engineering tyrosine overproduction in *E. coli*.

The last case addresses the question of what types of intervention are best. A series of 22 promoters with well-characterized, variable strengths was created by mutagenesis. This library was used to replace promoters for key genes in the biosynthesis of lycopene or biomass from glucose. These metabolic phenotypes exhibited strain-dependent optima with respect to the expression levels of the key rate-controlling genes. Promoter engineering thus shows that subtle genetic interventions can have profound effects on pathway function.

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

My work on this thesis would not have been possible without the steadfast support and assistance of many colleagues in the Department of Chemical Engineering and elsewhere at MIT. Hal Alper spent countless hours training me when I joined the lab, and was an indefatigable and eager collaborator in the creation and use of promoter libraries (Chapter 6). Hsien-Chung Tseng, Effendi Leonard, and Professor Kristala Prather shared an interest in clostridial butanol biosynthesis pathways and also shared many of the plasmids I used in this work. Through many helpful discussions, Daniel Klein-Marcuschamer helped me understand the area of microbial strain tolerance, and co-authored an early draft of what became Chapter 2 of this thesis. Daniel, along with Andrew Peterson and Tracy Mathews, were always willing to spend long hours with me discussing and analyzing the prospects for a biomass-based chemicals industry. Joanna Mackay, Mitchell Tai, and Alexander Papa-george helped streamline cloning procedures and develop analytical methods. My colleagues in the Stephanopoulos lab, especially Keith Tyo, Christine Santos, Tina Lütke-Eversloh, Chris Loose, Jason Walther, Elke Nevoigt, Jamey Young, and Hang Zhou, were continually generous with advice and support.

My entire thesis committee also provided essential ideas and critical feedback. Thanks especially to Professor Wittrup for wise counsel and pointed questions on topics from research to career advice, and to Professor Prather for willingness to share both her expertise and her laboratory's materials relating to clostridial butanol biosynthesis pathways. Professor Tester, although not technically on my committee, has been a personal role model for me as both a teacher and a researcher. Thanks to my thesis advisor, Professor Stephanopoulos, for being a font of research enthusiasm, a never-ending source of fascinating ideas, and the provider of all the laboratory resources that I could ever hope for.

My gratitude goes to my friends at MIT and elsewhere, too many to list here, for all their support through the years. But an extra hearty thanks goes to my parents John and Barbara Fischer, for their tireless patience as they endured my quarter-centure "career" as a student, and to my fiancee Elizabeth Yin, whose love and devotion never ceases to inspire me.
1 Introduction to the thesis: methods for the selection and optimization of pathway genes for the metabolic engineering of microorganisms
   1.1 Introduction and Motivation .................................. 12
     1.1.1 Two questions for metabolic engineering .................. 12
     1.1.2 Types of bottlenecks in metabolic pathways ................. 14
   1.2 Complexity in metabolic engineering: strain construction vs. strain analysis 16
   1.3 Creating strains and libraries: a survey of relevant methods for metabolic engineering .................................................. 19
     1.3.1 Transposon mutagenesis and gDNA libraries ................. 19
     1.3.2 Transcriptional engineering .................................. 20
     1.3.3 Protein engineering ......................................... 22
     1.3.4 Feeding studies ............................................. 23
   1.4 Common methods relevant for analyzing microbial strains or strain libraries 24
     1.4.1 “omics” ..................................................... 24
     1.4.2 Enzyme activity assays ..................................... 27
     1.4.3 Measurement of single metabolites .......................... 28
   1.5 Outline to the thesis ........................................... 28
   1.6 General conclusions: best experimental practices for identifying genetic interventions for metabolic engineering of E. coli ............................................ 30
2 Metabolic engineering for biofuels production: constraints and opportunities
   2.1 Introduction ..................................................... 37
     2.1.1 Biomass and biofuels ......................................... 38
   2.2 Biofuels feedstocks and products ................................ 41
     2.2.1 Feedstocks .................................................. 42
     2.2.2 Products .................................................... 49
   2.3 Requirements for yield, titer, and productivity ................. 53
   2.4 Biofuels hosts ................................................... 54
     2.4.1 Desirable properties of biofuels hosts ........................ 54
     2.4.2 Strain improvement .......................................... 56
     2.4.3 Engineering environmental tolerance to common process stresses .. 57
| 2.4.4 Biofuels-specific stresses | 59 |
| 2.5 Conclusions | 62 |
| 2.6 Materials and methods | 63 |
| 3 Design and construction of *E. coli* strains and pathways for the production of biofuels | 76 |
| 3.1 Introduction | 76 |
| 3.1.1 Butyric acid fermentation: a case study in synthetic biology for biofuels production | 77 |
| 3.1.2 Metabolic engineering of *E. coli* for the evolution of increased butyrate yield | 79 |
| 3.2 Materials and methods | 82 |
| 3.2.1 Stoichiometric envelopes for gene deletions in *E. coli* | 82 |
| 3.2.2 Strains and plasmids | 82 |
| 3.2.3 Culturing conditions | 84 |
| 3.2.4 Characterization of butyrate pathway function | 86 |
| 3.3 Results | 87 |
| 3.3.1 Butyrate-growth and succinate-growth stoichiometric envelopes | 87 |
| 3.3.2 Expected and unexpected phenotypes of the *pal* strain | 88 |
| 3.3.3 Continuous culture and the evolution of the *pal* strain | 90 |
| 3.3.4 Genetic tests of pathway | 94 |
| 3.4 Discussion and Conclusions | 99 |
| 3.4.1 Unexpected phenotypes of the *pal* strain | 99 |
| 3.4.2 The butyrate pathway: assessing in vivo functionality | 102 |
| 3.4.3 Conclusions | 106 |
| 4 Measurement of acyl-CoA pool sizes in *E. coli* strains expressing clostridial biosynthesis genes | 111 |
| 4.1 Introduction | 111 |
| 4.2 Methods | 114 |
| 4.2.1 Strains, plasmids, and culturing conditions | 114 |
| 4.2.2 Intracellular metabolite measurements | 115 |
| 4.2.3 Equilibrium calculations | 118 |
| 4.3 Results | 125 |
| 4.3.1 Validation of cell lysis my centrifugation into polar extraction fluid | 125 |
| 4.3.2 Physiological and thermodynamic data available from the literature | 125 |
| 4.3.3 Intracellular CoA pools in strain set #1 | 126 |
| 4.3.4 Uncertainty in the pathway thermodynamic model | 129 |
| 4.3.5 A second set of strains for CoA analysis | 129 |
| 4.3.6 Thermodynamic interpretation of pathway metabolite measurements | 133 |
| 4.4 Discussion and conclusions | 135 |
| 4.4.1 Thermodynamic analysis and uncertainty | 135 |
4.4.2 Utility of thermodynamic analysis for understanding the clostridial production of 3HB-CoA ........................................ 136
4.4.3 Conclusions ........................................ 138

5 Tyrosine production in E. coli ........................................ 143
5.1 Introduction ........................................ 143
5.2 Methods ........................................ 144
5.2.1 Construction of E. coli strains ....................... 144
5.2.2 Computational simulations of E. coli metabolism ........ 146
5.2.3 Objective functions for mutant strains ................... 149
5.2.4 Two approaches to predict gene targets using stoichiometric models ........................................ 152
5.2.5 Estimation of intracellular amino acid pool sizes ........... 153
5.2.6 Culturing techniques ........................................ 154
5.3 Results ........................................ 155
5.3.1 MoMA and ROOM converge on a knockout target ........... 155
5.3.2 aceE has an effect on tyrosine opposite to the models' prediction ........................................ 159
5.3.3 Measurement of amino acid pool sizes and a nitrogen limitation of hydrogen production ....................... 163
5.4 Discussion and conclusions ........................................ 165
5.4.1 The pitfalls of stoichiometric models ....................... 165
5.4.2 Other evidence for nitrogen control of tyrosine metabolism ........................................ 166
5.4.3 Conclusions ........................................ 167

6 Molecular tools for optimizing the behavior of targeted metabolic pathways ........................................ 173
6.1 Introduction ........................................ 173
6.2 Materials and methods ........................................ 175
6.2.1 Strains and media ........................................ 175
6.2.2 Library Construction ........................................ 175
6.2.3 Library Characterization ........................................ 176
6.2.4 Promoter Delivery Construction ....................... 179
6.3 Results ........................................ 180
6.4 Discussion and Conclusions ........................................ 191
6.4.1 Discussion ........................................ 191
6.4.2 Conclusions ........................................ 194

A The renewable production of propylene from glucose using a hybrid fermentative-hydrothermal process ........................................ 199

B Toward measuring butanol pathway intermediates in a cell-free system ........................................ 202
B.1 Introduction ........................................ 202
B.2 Materials and Methods ........................................ 202
B.3 Results and discussion ........................................ 203
C Correction of \( K' \) values for use at arbitrary pH and temperature  

D MATLAB code for optimization-based models of \textit{E. coli} metabolism  
D.1 MATLAB .m file for solving the FBA problem  
D.2 MATLAB .m file for solving the MoMA problem  
D.3 MATLAB .m file for solving the ROOM problem  
D.4 MATLAB .m file to vary the forced output level of tyrosine and calculate ROOM fluxes as a result  
D.5 MATLAB .m file for computing min/max flux values around a given flux state  
D.6 MATLAB .m file for making a 2-D projection of the phenotypic phase space  
D.7 MATLAB .m file for plotting a multiple 2-D phenotypic phase spaces
List of Figures

1-1 Common experimental approaches in metabolic engineering categorized along two independent dimensions. ........................................ 18

2-1 Feed and product pathways for the conversion of biomass into biofuels. ........................ 42
2-2 The reversible work of separation of for ethanol, butanol, alpha-pinene, or methyl oleate (biodiesel) from aqueous solution as a function of titer. ......................... 50
2-3 The tolerance of common microbes to butanol or corn-stover hydrolyaste. ... 57

3-1 Pathway of butyrate biosynthesis in Clostridium acetobutylicum ..................... 80
3-2 Stoichiometric product-growth envelopes for succinate and butyrate production in E. coli .................................................. 89
3-3 Oxygen auxotrophy in the pal strain ........................................ 91
3-4 Steady-state metabolite profile of E. coli pal grown in continuous culture under aerobic and anaerobic conditions .......................... 93
3-5 Validation of functional activity of clostridial butyrate biosynthesis genes expressed recombinantly in E. coli. .................................................. 95
3-6 Unexpected activity of the clostridial ptb-buk operon in recombinant E. coli. 97

4-1 Schematic of extraction and derivatization method for the high-sensitivity detection in E. coli of acyl-CoA and related intermediates from the clostridial butanol and 3-hydroxybutyryate biosynthesis pathways .......................... 116
4-2 Comparison of acyl-CoA pool sizes in E. coli strains expressing clostridial butyrate biosynthesis genes .................................................. 128
4-3 Monte-carlo calculation of uncertainty in calculated equilibrium concentrations of the butyrate synthesis pathway .................................................. 130
4-4 The effect of IPTG induction on observed ΔG' for component reactions of the clostridial butyrate or butanol biosynthesis pathway .......................... 135

5-1 Schematic illustration of mathematical optimization algorithms required for various flux prediction algorithms .......................... 153
5-2 Single-gene knockout predictions for increased tyrosine flux .......................... 156
5-3 The effect of aceE deletion on tyrosine production in various E. coli strains . 160
5-4 Inverse variation between tyrosine and alanine pools in tyrosine overproducing E. coli .................................................. 164
5-5 Nitrogen control of tyrosine production in tyrosine-overproducing E. coli ................................................................. 165

6-1 Representative false-color fluorescence photomicrographs of E. coli clones with varying degrees of expression level heterogeneity ........................................... 181
6-2 Work flow of promoter mutagenesis and screening ............................................................. 183
6-3 Multidimensional characterization of promoter strength ...................................................... 184
6-4 Effect of ppc promoter replacement on E. coli biomass yield on glucose .......................... 191
6-5 Effect of dxs promoter replacement on lycopene production in recombinant E. coli ............... 192

A-1 A renewable route for the production of high-purity propylene from carbohydrates ...................................................... 200

B-1 An attempt at analysis of clostridial butanol biosynthesis pathway function in a cell-free extract of E. coli .......................................................... 204
List of Tables

2.1 Energy densities, maximum yields on glucose, and infrastructure comparability of biomass and biofuels .................................................. 40
2.2 ATCC numbers for strains used in tolerance testing .................................. 64
3.1 Strains and plasmids used in this chapter. ........................................... 85
3.2 Substrate range of phosphotransbutyrylase (ptb) and butyrate kinase (buk) from Clostridium acetobutylicum ATCC824 ........................................ 104
4.1 New strains and plasmids introduced in this chapter ............................. 122
4.2 Thermodynamic and biological parameters estimated from the literature .................................................. 122
4.3 Probability distributions used for the Monte-Carlo estimation of uncertainty in the equilibrium pool sizes of CoA intermediates in butyrate biosynthesis. 124
4.4 Intracellular pool sizes of coenzyme A and related metabolites in E. coli strains expressing clotridial butanol biosynthesis genes ........................................ 132
5.1 The enzymatic reactions of tyrosine biosynthesis and their apparent equilibrium constants .................................................. 162
6.1 Sequences and properties of the 22-member E. coli promoter library ........ 185
A.1 Gas-phase concentration of propylene and byproducts produced from hydrothermal treatment of aqueous 3-hydroxybutyrate .................. 201
1.1 Introduction and Motivation

This thesis is about finding ways to identify genetic interventions that improve the performance of targeted pathways in microbial metabolism. The problem of identifying genetic interventions comprises two parts. The first is to decide, where to intervene, and the second, related, question is how to intervene?.

1.1.1 Two questions for metabolic engineering

Why is it important for metabolic engineers to consider and answer these questions? In a sense, all metabolic engineering projects aimed at developing new microbial strains have tried to answer these questions. However, in the past, metabolic engineers had limited experimental tools at their disposal. Approaches to the two questions in this thesis were frequently dictated by what was possible.
Yet, in a new era of “synthetic biology” and “systems biology”, both strain analysis and strain construction are experimentally easier than ever before. For example, cheap, large-scale DNA synthesis has facilitated the synthesis of large-scale, multi-gene DNA constructs. These constructs can easily be codon-optimized for expression in a desired host, and at the same time, clever schemes for the digestion, ligation, and rearrangement of these constructs hint that in the near future, facile, one-pot synthesis of a hugely diverse libraries of completely synthetic DNA constructs will also become routine.

Similarly, cheaper and more data-rich analytical technologies are making it easier to build huge data sets on the performance of desired metabolic pathways, microbial strains, and strain libraries. An example is the precipitous fall in the cost of DNA sequencing. This is commoditizing genome sequencing and is also behind the rise of newer methods for profiling microbial transcriptomes.

But what variations in native or synthetic DNA constructs will be the most informative to a metabolic engineer? What analytical strategies will best reveal routes to strain improvement? This thesis focuses on a specific type of question, specifically, on how to identify and target genes in a given metabolic pathway as the focus of further engineering effort. Certain genes may encode reaction steps which need no further engineering or improvement. A (usually much smaller) subset of genes will relate to the performance of limiting pathway steps. These pathway steps are bottlenecks. The task of identifying genetic interventions which improve the performance of a targeted metabolic pathway, then, is one of identifying and alleviating bottlenecks.
1.1.2 Types of bottlenecks in metabolic pathways

In this work, a metabolic pathway is taken to mean an intracellular series of enzymes, and the metabolites that react with or are produced by these enzymes. A “bottleneck” is a particular step or set of steps in the pathway whose rate constrains the rate of the overall system. Bottlenecks may be broadly classified into three categories:

- **thermodynamic:** If a particular metabolic step is at or close to equilibrium, it can not be accelerated by adding more enzymes or by adding or removing enzymatic “cofactors”. The only means to alleviate the bottleneck is to shift the equilibrium, following Le Chatelier’s principle: the concentration of substrates must be increased, or that of products must be decreased. An example of a metabolic reaction which lies close to equilibrium in *E. coli* is the phosphoglucone isomerase-mediated interconversion of glucose-6-phosphate and fructose-6-phosphate [31].

- **kinetic:** A strong thermodynamic driving force may be present, and enzymes capable of catalyzing the metabolic reaction of interest are expressed in the cell, but an intrinsically inefficient enzyme mechanism limits the rate of the relevant metabolic reactions. Adding more enzyme may help remove kinetic bottlenecks, but only to a point. Complete removal of the bottleneck could require protein engineering to reconfigure the enzyme’s active site, substrate specificity, stability, or even its mechanism. An example of an *E. coli* enzymatic reaction in this category might be the aerobic ribonucleotide reductase, responsible for reducing ribonucleotides to deoxyribonucleotides, necessary for the biosynthesis of DNA [7].
• regulatory: This class of bottleneck is distinguished from a kinetic bottleneck only by degree. Although the cell possesses genes capable of carrying out a given activity, these genes are not expressed or are otherwise downregulated during exposure to certain physiological conditions. Alternatively, small-molecule regulators downmodulate the desired enzymatic activity by binding to allosteric sites, blocking substrate access to the active site, or be stabilizing non-active states of conformations of the enzyme. Removing the allosteric inhibitor, or deactivating the transcriptional repressor, will alleviate the bottleneck. An example of a regulatory bottleneck in *E. coli* metabolism is *tyrA*-encoded chorismate mutase / prephenate dehydrogenase [16]. It is allosterically inhibited by tyrosine and transcriptionally repressed by the tyrosine-responsive transcriptional regulatory *tyrR*.

The task of the metabolic engineer is to identify the types and locations of bottlenecks which constrain the pathway of interest (*where to intervene*), as well as to make the appropriate genetic changes at these loci (*how to intervene*) to boost or restore pathway activity to higher levels. The boundary between kinetic and regulatory classifications for bottlenecks is often arbitrary, but the distinction is useful, as it often can help resolve which types of genetic interventions are most likely to be efficacious for the improvement of target pathway activity. For example, the *tyrA* bottleneck to tyrosine overproduction in *E. coli* was successfully eliminated by mutating the putative allosteric regulatory site and deleting the *tyrR* transcriptional regulator [17], but removal of the ribonucleotide reductase bottleneck in DNA synthesis would likely require an speed-up of a specific mechanistic step (the 1,2-hydroxyl shift) in the enzyme’s kinetic mechanism.
1.2 Complexity in metabolic engineering: strain construction vs. strain analysis

The central metabolism of *E. coli* consists of about 600–1000 enzymes and transporters, which operate on a set of a few hundred small-molecule metabolites [6]. In theory, any of these reactions could be a bottleneck for the overproduction of a target metabolite. To face this potential complexity, methods for strain/library design and strain/library analysis must be widely scoped. However, in practice, usually preexisting knowledge or heuristics can usually focus a metabolic engineering investigation on a much smaller subset of enzymes and transporters.

To illustrate this, all possible metabolic engineering experiments can be represented in a 2-D diagram as shown in Figure 1-1, where one dimension measures the complexity of strain modification, and the other measures the complexity of strain analysis. For example, strains can be modified by completely random mutagenesis with broad-spectrum untargeted DNA damaging reagents such as nitrosoguanidine (NTG). The effects of these mutations can be very wide-ranging, and it is difficult to limit the number of mutations in a given strain. An upper bound for the possible complexity of this method in *E. coli* might be $4^{4,600,000}$: in theory, each of the 4,600,000 nucleotides in the *E. coli* genome can independently assume one of four possible values (A, G, C, or T). These methods of mutagenesis were in common use until the advent of recombinant DNA technology in the 1980s [29].

In contrast, transposon mutagenesis and selection [3] uses selectable markers to control the number of genetic modifications per clone to about 1, and because large scale insertion
events, rather than single-base deletions, transitions, or transversions, are the only type of mutation which occurs, most of the mutations are gene disruption events. In *E. coli*, an approximate upper bound on the complexity of this technique might be $2^{4,300}$, arising from the independent possibilities for the presence or absence (i.e. binary outcomes) of each of *E. coli*’s ~ 4300 genes.

Cloning a single gene is perhaps the simplest type of strain or “library” construction, but environmental perturbations, such as changes in nutrient supply or the introduction of toxic agents, can be viewed as “modifications” which complexify a strain or strain library.

The other dimension of complexity is what is measured. For example, measuring microbial growth is perhaps the simplest metabolic engineering experiment. Measuring the production of a single, secreted metabolite is another very common, if still very simple, mode of measurement. “Omics” techniques, in constrast, measure up to hundreds or thousands of unique proteins, transcripts, or metabolites.

The remainder of this chapter outlines the three case studies detailed in this thesis’s remaining chapters, and places each case in the 2-D complexity landscape shown in Figure 1-1. It is already apparent from Figure 1-1 that the majority of the techniques used in this thesis fall in the middle of the *x* and *y*-axes in Figure 1-1. The goal of the remaining discussion in this chapter is to answer why this is so.
Figure 1-1. Common experimental approaches in metabolic engineering categorized along two independent dimensions: the complexity of strain construction (y-axis) and complexity in strain analysis (x-axis). The points along each axis are meant as reference points only; many possible experimental measurements or strain construction techniques may be similar in complexity to promoter replacement, for example, despite having nothing in common with promoters. The chapters in this thesis are shown in this categorization scheme.
1.3 Creating strains and libraries: a survey of relevant methods for metabolic engineering

This section introduces some useful approaches to the creation of strains and strain libraries. Rather than serve as an exhaustive list, this discussion is meant to sample a few key techniques in order to illustrate the range in possible strain/library construction complexity.

1.3.1 Transposon mutagenesis and gDNA libraries

Unlike the methods discussed below, what is varied in these libraries is not a small number of targeted nucleotides which comprise part of a well-understood biologically functional module. Instead, the variable is the identity of a long piece of DNA. There are several popular formats for these studies. Perhaps the oldest is the expression of genomic DNA libraries derived from a wild-type strain in a mutant host (or vice-versa). If the mutant has been made deficient in the biosynthesis of a growth-essential metabolite, for example, a genomic fragment from its wild-type parent strain, or even a completely unrelated strain of bacteria, could restore growth if it carries the gene for the missing enzyme needed for biosynthesis of the growth-essential metabolite. Many, many genes were first cloned in this way. This method is of undisputed utility for the isolation and functional characterization of unstudied genes and genomic regions.

This format of library creation can be applicable to metabolic engineering efforts only if a library screening method can be directly linked to pathway function. For example, Yong-Su Jin showed [13] through this type of analysis that \textit{idi} and \textit{dxs} were rate-controlling genes in...
the biosynthesis lycopene by recombinant E. coli.

Untargeted insertions in the E. coli chromosome will randomly disrupt open reading frames (ORFs) and thus inactivate or "knock out" the activity associated with the ORF. Knockout libraries are useful when limits on pathway flux are not endogenous to the pathway but instead are caused by redistribution of carbon or energy to competing, possibly unknown pathways. An example is the discovery by Alper et al. that the deletion of a regulatory gene, hnr, led to increased lycopene accumulation [2].

Both the gDNA overexpression and knockout studies depended on colony screening based on the color of lycopene: the redder a colony, the more lycopene it produced. Thus, the complexity of the measurement in these studies was low: the lycopene production level of transposon-insertion or gDNA-overexpression libraries forms a 1-dimensional data set.

1.3.2 Transcriptional engineering

This broad category includes promoter engineering, global transcription machinery engineering, engineering RNA secondary structure, and, broadly construed, even methods for controlling the DNA copy number of metabolic genes.

The past years have seen an explosion of engineering approaches to modulating the transcriptional programs for the expression and translation of target enzyme mRNAs. The diversity in methods stems from an explosion in fundamental knowledge on the regulation of mRNA synthesis, as well as the translation of mRNA into protein.

The control of RNA polymerase binding and transcription initiation by "promoter" DNA sequences has long been recognized, and since the 1980s systems have been available for the very high-level overexpression of recombinant proteins. These systems were used mainly as
tools for the preparation of large amounts of highly purified enzymes and other proteins of interest. The metabolic engineering utility of many of these systems was limited by their strength: in metabolic engineering, a metabolite derived from an enzyme is the target product, and the enzyme(s) overexpressed are rarely the goal in and of themselves.

Many tools have since been developed to experimentally access the full continuum of gene expression space. The binding of RNA polymerase to "promoter" regions of DNA has long been recognized to be an important control point in transcription. Chapter 6 of this thesis explores how random mutagenesis to promoter regions can be exploited to develop new promoters of interest. Recent developments have also included the development of DNA cassettes which are inactive as promoters until a specific sub-region is inverted by a specific DNA recombinase system, allowing inducible control of the reconfiguration of a DNA sequence into a functional promoter system [9].

Since mRNA synthesis is a template-driven reaction, at full induction gene expression can be controlled by the copy number of the genes, promoters, and transcriptional terminators present in a given cellular system. Metabolic engineers have also developed tools for controlling the copy number of desired expression cassettes [14, 27].

However, it has become increasingly clear that mRNA secondary and tertiary structures can play major roles in controlling the translation of mRNAs into protein, and in the control of transcription itself. These processes, too, are amenable to exploitation by engineers for the introduction of desired regulatory programs. Perhaps the most prominent recent example is the development of modular elements for the control of mRNA expression through modulation of secondary structure by Christina Smolke and co-workers [4, 20, 24, 25, 30]. These methods are best for optimizing or eliminating known bottlenecks, or introducing de-
sired programs for the control of gene expression. More direct methods may be of utility for identifying which genes or enzymes are bottlenecks in the first place.

An example of using transcriptional engineering, in this by varying gene expression by duplication (or not) of the chromosomal gene on a plasmid-based overexpression system, as a means to identifying pathway bottlenecks, is Lütke-Eversloh’s work on combinatorial overexpression of tyrosine biosynthesis pathway genes in order to identify which pathway genes lead to improved tyrosine production [17].

1.3.3 Protein engineering

If a given metabolic pathway is limited by a single metabolic reaction, protein engineering to alter the chemical architecture of the enzyme’s active site may potentially alleviate the rate limitation. However, enzymes have evolved over billions of years to optimize the rate of biochemical reactions, so radical improvements in catalytic efficiency can happen only rarely, and usually when the change is an expansion or alteration of substrate specificity, rather than a boost in the catalytic activity with the enzyme’s preferred substrate.

Perhaps the most frequent modulation of enzymatic behavior by protein engineering is the addition or removal of allosteric regulation. Allosterically regulated enzymes have catalytic efficiencies that are controlled by the binding of third-party ligands to sites distal to the active site. Usually these regulators effect conformational changes in the enzyme structure, which block the active site or reposition catalytically essential residues or elements in to inert configurations. One recent example of protein engineering for the improvement of allosteric enzyme regulation include the modulation or replacement of *E. coli* citrate synthase with variants which were not allosterically inhibited by NADH [28]. These variants improved
the production of ethanol from five-carbon sugars in *E. coli*, by speeding the removal of acetyl-CoA. Another is Lütke-Eversloh's isolation, by screening against a non-metabolizable inhibitor analog, of a variant of the TyrA prephenate dehydrogenase which is not sensitive to feedback inhibition by tyrosine [16].

1.3.4 Feeding studies

Feeding pathway pre-cursors is a very common method for identifying or circumventing pathway bottlenecks. One example is Kristala Prather's use of butyrate and 3-hydroxybutyrate as butanol precursors [19]. In these examples, provision of a metabolite which enters a pathway downstream of all or part of rate-limiting enzymes bypass this enzyme. In these cases high flux can be observed. By systematic variation in the substrate, it can be possible to resolve exactly which enzymatic steps are most limiting. A good example is a study on the production of benzoisoquinoline alkaloids in yeast [10], from Christina Smolke's group. Benzoisoquinoline alkaloids are synthesized from tyrosine, but the pathway steps from tyrosine to a common intermediate called norcoclaurine are poorly characterized. Feeding the norcoclaurine analogue norlaudanosine circumvented the need for these poorly characterized steps and allowed for validation of the functional activity of all downstream pathway steps.

Another variant of this technique is to exploit the reversibility of relevant pathway reactions, and to test all or part of the pathway in reverse. An example of this method is illustrated in Chapter 3, where it is applied to ~8 microbial strains.

A limitation of feeding studies is that many pathway intermediates, especially if phosphorylated, are impermeable to the cell wall or cell membrane. In these cases, pathway precursors can be fed in permeable or transportable forms, but doing so introduces complex-
ity: the import of the “precursor intermediate” must not be rate limiting in order for this type of experiment to be informative. A description of a preliminary attempt at such an experiment is described in Appendix B of this work.

1.4 Common methods relevant for analyzing microbial strains or strain libraries

This section introduces some recent approaches analyzing strain libraries. The assumed goal for such analyses is the extraction of information about which genes or environmental conditions are most relevant to the overproduction of a desired target metabolite. As above, rather than serve as an exhaustive list, this discussion is meant to sample a few key techniques in order to illustrate the range in possible strain/library analysis complexity, and to show that the techniques used in this thesis fall at intermediate levels of complexity, as shown in Figure 1-1.

1.4.1 “omics”.

Transcriptomics and genomics

These two “omics” technologies are briefly discussed as example high-complexity methods for analysis of strains and strain libraries in metabolic engineering.

Genome analysis is beginning to emerge as a technique for metabolic engineering [12]. I classify this technique as very complex, because it entails measuring the identity of every nucleotide in a strain’s (or set of strains’) genome. This is equivalent to making 4.3 million
measurements, each with four possible outcomes.

Transcriptomics has a long history in metabolic engineering. Transcriptomics can provide relative assessments of the concentrations of (for \textit{E. coli}) \(\sim 4300\) gene transcripts. Although in principle the concentration of an mRNA transcript is a continuous variable, in practice, experimental errors and uncertainties will make it impossible to resolve arbitrarily small differences in concentration. Today's microarrays provide perhaps 20\~{}50 distinguishable, concentration levels for mRNA transcripts, across a wide dynamic range. The complexity of this technique might thus be \(\sim 2^{4.300}\), a number considerably smaller than for completely random single-nucleotide mutagenesis.

Combinations of transcriptomics and genomics have recently been applied to metabolic engineering problems in powerful ways [11, 12, 5, 8]. The study by Herring et al. is perhaps the most compelling. Using a microarray-based comparative genome sequencing method, Herring and his co-workers tracked a population of \textit{E. coli} as they evolved to improve their growth rate on glycerol as a sole carbon source, over hundreds of generations [12]. By effectively sequencing the entire genome of many clones isolated from various points of the evolutionary process, they were able to exhaustively identify mutations in only eight key genes, and show that artificially introducing these mutations, both singly and in combination, was able to recover rapid-growth phenotypes on glycerol. However, even though costs for these types of analyses are falling, they are still much more costly and time consuming than simpler single-species analytical techniques. Additionally, it is unclear whether all metabolic phenotypes of interest are likely to be controlled by such a small number of genes.
Metabolomics

Because of the chemical diversity of metabolites, as compared to other classes of biomolecules, methods for the rapid, large scale analysis of many intracellular metabolites were comparatively more difficult to develop and have become commonplace only recently. Many microbial metabolomic analyses have been reported [15, 21, 22, 26].

How can metabolomic information be exploited in the identification of pathway bottlenecks? Perhaps the best study to date is reported from Kummel et al. [15]. The key point is that thermodynamics apply to enzymatic and metabolic reactions just as well as it does to traditional chemical reactions. The driving force for any reaction is a minimization of free energy. When the changing the ratio between "reactants" and "products" in a metabolic reaction leads to the release of free energy, the reaction is thermodynamically favorable.

The key point is that the driving force for metabolic reactions depends on the intracellular concentrations of substrates and products. Metabolomics offers the possibility of measuring many of these concentrations of a large set of these metabolites. When coupled to basic thermochemical data for metabolites [1], metabolomic data can point out which metabolic reactions enjoy large thermodynamic driving forces, and which are closer to equilibrium. Such knowledge is important, because although enzymatic reactions are highly specific and vastly accelerated to the uncatalyzed reaction, enzymes can do no more than exploit a pre-existing thermodynamic driving force for a reaction to its fullest extent. The rate of reactions which lie close to equilibrium cannot usually be boosted by simply increasing enzyme activity. Important exceptions may apply be when (i) enzyme concentrations are in excess of substrate and/or product concentrations, and (ii) the enzyme-product complex is differen-
tially stabilized relative to the enzyme-substrate complex. Under these conditions, the major fraction of product or substrate molecules may exist as complexes with the relevant enzyme, and not necessarily exist as freely diffusing solutes in the cellular cytoplasm. In effect, these two factors amount to enzymatic manipulation of the thermodynamic equilibrium constant. Small molecule metabolites usually have intracellular concentrations in the range of tens of micromolar to tens to hundreds of millimolar, but enzyme concentrations are usually tens of micromolar or lower in concentration [18, 23].

The maximal number of small molecules reported as measured or detected in single metabolomics studies is usually in the hundreds or low thousands. The dynamic range and robustness of quantitative results for various reported metabolomic methods varies considerably [26, 32]. If we assume that perhaps 20~50 concentration levels are distinguishable, metabolomics studies may have complexities comparable or slightly less than microarray-based transcriptomics techniques.

In Chapter 4 and Chapter 5 of this thesis, I use focused “metabolomic” analysis techniques that quantitatively assay for pathway-focused subsets of all intracellular metabolites. Chapter 4 describes a technique for the measurement of ~9 intracellular CoA esters and ATP, ADP, and AMP. Chapter 5 describes a technique for the estimation of the intracellular concentrations of ~15 amino acids. These techniques are thus of intermediate analytical complexity in the axis shown in Figure 1-1.

1.4.2 Enzyme activity assays

Enzyme activity assays for metabolic engineering are perhaps most informative when conducted on crude cell extracts. Usually large excesses of all relevant substrates are added, and
detection of substrate depletion or product formation is done by the facile means available. The spectrophotometric monitoring of NADH or NADPH depletion is a common choice.

Enzyme activity assays are valuable because they offer kinetic information about individual pathway steps, but especially when large excesses of substrates are added to the extract, the rates of the enzyme reaction may not reflect rates under \textit{in vivo} conditions, with possibly much lower substrate concentrations.

1.4.3 Measurement of single metabolites

This method is perhaps the most common in metabolic engineering studies of strain libraries. Because massive numbers of microbial cells (> $10^{12}$) can easily be cultured in the laboratory, this technique is preferred for analyzing very large strain libraries created through transposon mutagenesis, gDNA cloning, or random mutagenesis.

Single-metabolite analyses focused on non-growth compounds can also be conducted in a very high throughput fashion, either in 96-well plates or in colonies on agar plates, if colorimetric or fluorometric screens are available.

In this thesis, a spectrophotometric technique to analyze lycopene production is used in Chapter 6 to analyze three collection of ~15 promoter replacement strains of \textit{E. coli}.

1.5 Outline to the thesis

This thesis compares approaches to answering the two central questions, \textit{where} to intervene, and \textit{how} to intervene, by a series of case studies. Chapter 2 motivates the industrial relevance of these central metabolic engineering questions.
Chapters 3, 4, and 5 all seek to answer the question, *where to intervene*, in the context of butyric acid and butanol production by recombinant *E. coli*. These chapters illustrate two main approaches to answering this question. These chapters are concerned mainly with techniques for *identifying* the location of pathway bottlenecks. In these chapters, the goal was identify which steps in the pathway were rate-limiting. In Chapter 3, the core approach relies on the use of genetic complementation and feeding studies to test the functional activity of each of the pathway enzymes. Chapter 4 combines metabolomic measurements with a thermodynamic model of the butyrate biosynthesis pathway to identify which pathway steps are farthest from equilibrium, and thus most likely to be kinetic or regulatory bottlenecks.

Chapter 5 describes a case study for the production of a non-biofuel small-molecule metabolite, tyrosine, for possible application in renewable materials industry. It also deals primarily with the problem of *identifying* pathway bottlenecks. Unlike the work described in Chapter 3 and Chapter 4, it deals with a pathway which is entirely endogenous to the host organism. The work in this chapter began from a pathway that was already observably active at tyrosine production; however, the pathway performance was sub-optimal.

Chapter 6 describes the development of a molecular method useful for answering the question, *how to intervene*. Specifically, this chapter describes the creation of a promoter library in *E. coli* and the application of this library to the optimization of metabolic pathways whose bottleneck steps had already been identified.

This chapter, the first in the thesis, describes in general terms the types of approaches that are applied in the later chapters, and places these techniques in a larger context of possible and frequently used approaches in metabolic engineering.

In the remainder of this chapter, I touch on a number of the more common techniques for
introducing both modifying strains and strain libraries, and on the common techniques for measuring strains and strain libraries. References to later chapters and to the literature serve to place the approaches described in later chapters into the broader context of metabolic engineering as it is practice today.

Figure 1-1 also shows where the case studies comprising this thesis fit into this 2-D complexity landscape.

1.6 General conclusions: best experimental practices for identifying genetic interventions for metabolic engineering of *E. coli*

The evidence presented in the other chapters of this thesis illustrated a variety of approaches for strain and library analysis. These case studies are not an exhaustive sampling of the possible combinations of strain/library creation complexity and analytical complexity, but as shown in Figure 1-1, most of the techniques in this thesis combined techniques of intermediate strain/library complexity and intermediate analytical complexity.

Why is this so? This central question can be broken into two parts. The first: why not use higher-complexity techniques for both strain modification and strain analysis? The answer here is straightforward: an obvious tradeoff exists between the complexity of a resulting data set and the ease of acquisition of the data. To use an extreme example, imagine collecting full metabolomic, proteomic, and transcriptomic analyses on a strain collection in which contained also possible genomes of 4,600,000 nucleotides in length. The strain collection would
consist of $4^{4,600,000}$ strains, and full "multi-omic" analysis would provide measurements of approximately 600 metabolites + 4600 transcripts + 4600 proteins = 9800 independent variables for each strain. Such a data set is obviously inaccessible. No current technology is available to create and analyze so many unique strain variants, and even if it were, the time necessary to acquire the data would be centuries, and petabytes of memory would not be enough to store the resultant data set. Although this example is extreme, it serves to illustrate, the fundamental tradeoff between the richness and quality of a data set and the ease of its acquisition.

When the pathway for a desired phenotype, usually the overproduction of a desired metabolite, is well-characterized, it is best to focus strain and library design to this pathway. With the complexity of the strains and libraries to be analyzed thus limited, the question becomes, why focus on techniques that are intermediate in complexity in both strain modification and in strain analysis? That is, why not use a high-complexity strain analysis technique on a very simple set of strains? Or conversely, why not use very high-complexity strain libraries and simplify the analytical technique, perhaps measuring only the metabolite of interest?

This is a much deeper question. The answer lies in the structure and complexity of metabolism. The function of a metabolic network depends on the interaction of its components. For example, if in one experiment boosting the level of enzyme 1 was found to improve the desired phenotype, and in another experiment boosting the level of enzyme 2 was found to have the same effect, it does not follow that the improvement, if any, of simultaneously boosting both enzyme 1 and 2 will be the sum of each enzyme's individual effects.

How do these interactions affect how informative a given metabolic engineering approach
will be? Consider two strain variants, X and Y, and suppose that Y has an improved phenotype relative to X. Suppose that we analyze these two strains by multiple high-complexity analysis techniques, say, metabolomics and proteomics. Such an approach would fall in the lower right-hand corner of Figure 1-1. In general, the levels of many metabolites and proteins to differ between the two strains. But which differences are important for the improved phenotype? Do all of the differences stem from a single key mutation in strain Y? Such an analysis cannot easily address these questions, because the data does not allow inferences on which interactions in the metabolic networks of strains X and Y are important for the desired phenotype.

Alternatively, consider a library of single-gene knockout mutants analyzed for the production of a single metabolite, from which several single-gene knockouts A, B, and C are found to improve production. Such an experimental approach would lie on the upper left corner of Figure 1-1. Do these knockouts share a common mechanism of action? How do they act to improve production? Without a richer data set from which we might infer commonalities between the single-gene knockouts, and the ways in which knockouts A, B, and C interact with the metabolic pathway for the production of the target, it is impossible to say.

To understand how beneficial strain modifications interact with the metabolic network, a metabolic engineering experimentation should rely on approaches as close as possible to the upper-right corner of Figure 1-1. But as discussed previously, the extreme upper-right corner of Figure 1-1 is constrained by the practical exigencies of the laboratory. Thus, approaches which combine moderately complex strain modification techniques and moderately complex strain analysis techniques are most likely optimize the data on important interactions in the metabolic engineering network, subject to practical constraints. Such approaches should
be the first place metabolic engineers turn to identify pathway bottlenecks, and identify beneficial genetic interventions which improve target metabolite overproduction.


This Chapter:

- Describes the challenges and opportunities for the metabolic engineering of biofuels-producing microbes.
- Reviews the microbially relevant characteristics of prevalent biomass feedstocks and reviews the metabolic pathways which have been proposed as future biofuels.
- Illustrates with two examples that host organisms respond very differently and in non-generalizable ways to biofuels-related stresses.
- Discusses approaches for selecting and optimizing host organisms as a starting point in the engineering of a biofuels production pathway.

2.1 Introduction

Recent years have seen a surge in public concern about both global warming and the security of the United States’ energy supply. One proposed solution to decrease reliance on foreign oil and to potentially mitigate greenhouse gas emissions from the burning of transportation

†Note: This chapter is derived largely from material previously published by the author [28].
fuels is to rely on domestic and renewable feedstocks for fuels production. The feedstock which to date has received the most attention is plant-derived organic materials, or biomass.

In this chapter, I survey the feedstocks that may be useful for biofuels production, and similarly survey the molecules that have been proposed as second-generation liquid biofuels, and consider how various feedstocks and products influence the choice and optimization of suitable host organisms for biofuels production. Although substantial opportunities for metabolic engineering of higher plants exist [75], the instant focus is limited to examining microbial conversions of biomass.

The interest in fuel production from biomass has also spurred interest in the production of other commodity chemicals for the materials and pharmaceutical industries. Fuels usually (i) cost less than other commodity chemicals on a per-weight or per-energy (heat of combustion) basis, and (ii) enjoy far larger market sizes than commodity chemicals. Biomass might thus first become attractive as a feedstock for higher-value, lower-demand commodity chemicals as a natural path to reaching the economies of scale needed for profitable fuels production. However, this chapter focuses on fuels because (i) this market is where turning to biomass can best improve energy security and reduce fossil fuel-associated CO$_2$ emissions, and (ii) the metabolic engineering challenges associated with biomass use are most starkly illustrated in the context of fuels production, because of fuels’ lower price.

2.1.1 Biomass and biofuels

The creation of liquid transportation fuels from renewable biomass has been a long-standing research goal [17]. Biomass is renewable, and is abundant in places that other liquid fuels, chiefly petroleum and its byproducts, are not readily available [70].
Biomass resources are widely distributed, but often with densities of 0.4 kg / L or lower. The challenge of fuels production from biomass is to liquefy and increase the bulk density of the resource, all while preserving its energy content. The final fuel product should (i) have a high energy density on a mass as well as a volume basis, (ii) be produced at yields near the stoichiometric maximum from a given biomass feed, and (iii) enjoy compatibility with existing fuel distribution infrastructure. No single fuel currently satisfies all of these criteria Table 2.1.
Table 2.1. Energy densities, maximum yields, and compatibility with existing fuel distribution infrastructure for various types of biomass and liquid fuels.

<table>
<thead>
<tr>
<th>Petrofuels</th>
<th>Molar Energy Density (kJ / mol)</th>
<th>Formula Mass</th>
<th>Volumetric Energy Density (MJ / L)</th>
<th>Mass Energy Density (MJ / kg)</th>
<th>Bulk Density (kg / L)</th>
<th>Maximum stoichiometric yield from glucose (kg/kg)</th>
<th>Infrastructure compatibility?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gasoline</td>
<td>100 - 105</td>
<td>34.7</td>
<td>42.3</td>
<td>0.72 - 0.78</td>
<td>-</td>
<td>-</td>
<td>Yes</td>
</tr>
<tr>
<td>No. 2 Diesel</td>
<td>~200</td>
<td>38.3</td>
<td>45.3</td>
<td>0.80 - 0.89</td>
<td>-</td>
<td>-</td>
<td>Yes</td>
</tr>
</tbody>
</table>

Example Biofuels Feedstocks

| Hybrid poplar wood  | -                               | 6 - 7.1      | 19.38                             | 0.310 - 0.370               | -                    | -                                               | No                             |
| Glucose             | 2803                            | 180.155      | 24.3                              | 15.6                        | 1.562                | -                                               | No                             |
| Syngas (CO + H₂)    | 569                             | 30.026       | 18.95                             | -                           | -                    | -                                               | No                             |

Candidate Biofuels

| Ethanol             | 1367                            | 46.07        | 23.6                              | 29.9                        | 0.792                | 0.511                                           | No                             |
| n-butanol           | 2676                            | 74.121       | 29.2                              | 36.1                        | 0.81                 | 0.411                                           | Probable                       |
| Vegetable oil (canola) | 35390                         | ~887         | 36.2 - 36.7                       | 39.8-40.0                   | 0.910-0.917          | ~0.353                                          | No                             |
| Biodiesel           |                                  |              |                                   |                             |                      |                                                 |                                |
| (methyl-esterified virgin canola) | 12080                    | ~298         | 35.7                              | 40.6                        | ~0.88                |                                                 | Partial and increasing       |
| α-pinene (representative terpenoid) | 6205                     | 136.234      | 38.9                              | 45.5                        | 0.8539               | 0.324                                           | Unknown                        |

\(^a\) Data compiled from [10, 13, 20, 35, 51, 84, 98].
Ethanol as made from corn or sugar cane has an acceptable energy density and can be produced from a variety of biomasses at excellent yields. However, ethanol largely fails the requirement for compatibility with existing fuel infrastructure. Gasoline, the predominant liquid fuel in the U.S., enjoys over 95,000 miles of dedicated pipeline for its distribution, over 160,000 filling stations, usually with underground gasoline storage tanks, and over 240,000,000 heavy- and light-duty passenger vehicles already in use. Together, this infrastructure is worth hundreds of billions of dollars. Ethanol can cause materials corrosion and draw water into the fuel mixture, properties which render it unsuitable for use in the existing fuel distribution infrastructure. Increased use of ethanol in large quantities may thus necessitate large infrastructure investments.

Other proposed fuels, in particular, biodiesel, butanol, and various terpenoid compounds may be compatible with existing fuels infrastructure, but efficient, high-yielding processes for their production are not yet commercially feasible. The principal barrier to the production of these advanced biofuels is the development of robust, high-yielding microbes and processes for their production.

2.2 Biofuels feedstocks and products

Metabolic pathways for the synthesis of most proposed biofuels proceed through common metabolic intermediates such as acetyl-CoA or pyruvate. The pathways can thus be effectively divided into feed pathways, as shown in Figure 2-1(a), which convert biomass to the common metabolic intermediates, and production pathways, which convert the intermediate to the chosen fuel, as shown in Figure 2-1(b). Generally speaking, the feed pathways
create reducing equivalents which are required by production pathways for the synthesis of the biofuel. Consumption of reducing equivalents by other pathways, for example aerobic respiration, is therefore generally undesirable.

*Figure 2-1. Pathways for the metabolic conversion of biomass into biofuels can be loosely divided into (a) feed pathways, which convert carbohydrate biomass into the central metabolic intermediates pyruvate and acetyl-CoA; and (b) product pathways, which convert these central intermediates into fuels. Reducing equivalents, shown with yellow arrows, are generated in (a) and consumed in (b). For reasons of clarity and familiarity, molecular structures are not drawn in (a).*

### 2.2.1 Feedstocks

**Starches and simple sugars**

Ethanol, today's predominant biofuel, is presently manufactured using feedstocks such as cane-derived sucrose and corn-derived starch. In Brazil, sugar cane juice and sugar cane molasses are used as sources of sucrose. Using *Saccharomyces cerevisiae* and closely related yeast strains as the host, industrial ethanol yields on sucrose are up to 93% of the stoichiometric maximum. Both continuous and batch production is used, with residence times in
the fermenters being 6-10 hrs [18]. Production is more than 15 billion gallons (57 billion liters) per year.

In the US, the predominant feedstock is maize. Studies have reported industrial yields in the range of 2.65 to 2.71 gallons per bushel (0.388 to 0.397 L/kg) of maize, which amounts to approximately 93.6 to 95.8% of the stoichiometric maximum [45, 59]. Presently, industrially attained ethanol titers post-fermentation are around 9% by weight [59], although this figure may increase with improvements in strain ethanol tolerance (see Section 2.4.2, below).

The process for production of ethanol from these sources is highly efficient, with feedstock cost generally amounting to 60-80% of the cost of production, although this figure is subject to the usual vagaries of the agricultural commodities markets. The metabolic pathway for starch or sucrose conversion into fuel begins with enzymatic hydrolysis, which yields glucose. The glucose is converted to pyruvate via the familiar pathway of glycolysis Figure 2-1(a). Because processes for starch- or sugar-derived ethanol are so efficient, they are a convenient benchmark by which to evaluate other proposed biofuels processes.

**Lignocellulosics**

Cellulosic biomass enjoys a much more massive resource base than what is available from maize or sugarcane [70]. It is expected to play a dominant role in biofuels production in the near future.

However, cellulosic biomass is more difficult to convert into fermentable sugars than is corn or sugar cane, because (i) five-carbon sugars, mainly xylose, account for 10-25% of the total carbohydrates; (ii) of the presence of lignin, a highly recalcitrant network polymer of aromatic alcohols that accounts for 17-25% of common cellulosic biomasses [100]; and (iii)
cellulose is much more resistant to hydrolysis than starches and simple oligosaccharides. The first obstacle can be overcome through the selection and/or engineering of microbes capable of the anaerobic fermentation of xylose and other five-carbon sugars to ethanol (see below). These five carbon sugars are metabolized via the pentose phosphate pathway Figure 2-1(a). Enzymatic degradation of lignin is too slow to be practical industrially, but lignin can be productively burned for power production, or gasified for thermochemical conversion to fuels or chemicals.

The third obstacle is the most important. Cellulosic biomass is recalcitrant to biochemical degradation and conversion because cellulose’s highly crystalline structure protects inner layers from exposure to degradative conditions, and because lignin is a highly branched polymer held together by relatively inert ether- or carbon-to-carbon-bonds. This recalcitrance necessitates chemical pre-treatment of cellulosic biomass to a partially hydrolyzed product that cellulase enzymes can more easily digest. A number of pretreatment variations have been proposed, but many generate fermentation inhibitors (see Section 2.4.2, below). Costs for the conversion of cellulose to fermentable sugars include the capital and operating costs of chemical pretreatment, as well as the cost of cellulase enzymes which catalyze the final stages of polysaccharide hydrolysis to fermentable sugars.

Two main metabolic engineering approaches have long been studied for provision of cellulase and sugar fermentation to ethanol. The first approach separates the production of cellulase from the fermentative conversion of biomass sugars to ethanol. That is, in one process step cellulase is produced by a specialist (set of) microbe(s) engineered for high-level cellulase overproduction; in another process step, it is added to a “pre-treated” slurry of partially degraded lignocellulosic biomass to convert the cellulose to fermentable sugars;
in a third step, the fermentable sugars are converted by a separate (set of) microbe(s)
into desired fuel products. Work to improve this approach has been aimed at genetically
modifying \textit{S. cerevisiae}, \textit{E. coli}, \textit{Zymomonas mobilis} or other organisms so that pathways
for utilization of five carbon sugars and for ethanol synthesis exist in the same organism
[38, 42, 62]. Recent, promising results with \textit{S. cerevisiae} have been reviewed [100]. The
advantages of \textit{Saccharomyces} are high ethanol tolerance, and the ability to ferment at low
pH and forgo medium sterilization. The disadvantage is that \textit{Saccharomyces} (like \textit{E. coli}
and \textit{Z. mobilis}) cannot utilize cellulose or derived oligosaccharides directly. This necessitates that
a biorefinery install its own dedicated on-site cellulase production system, usually involving
the aerobic cultivation of organisms such as \textit{Trichoderma reseei} [112].

The diversion of feedstock to cellulase production, and the expenses incurred by aerobic
cell growth (e.g. electricity and capital for air compression and culture agitation) add to
the cost of fuel production. A 2000 study estimated that cellulase production accounted
for about $0.073/\text{kg} (\$0.22/\text{gal}) of ethanol, or about 14\%, of the production cost [59]
in a \textit{Zymomonas}-based process, in rough agreement with past estimates [56]. The assumed
biomass cost in these studies was $35 - $42 per dry ton.

The second metabolic engineering approach, called consolidated bioprocessing or (CBP)
and formerly referred to as direct microbial conversion (DMC), uses highly cellulolytic or-
ganisms like the thermophilic \textit{Clostridium thermocellum} either exclusively or in co-culture
with other thermophilic, higher-producing sugar fermenters [57, 65]. These organisms per-
mit cellulase production, cellulose hydrolysis, and fermentation to occur anaerobically in the
same process vessel. Advantages are (i) that the thermophilic nature of these organisms may
also obviate requirements for medium sterilization by allowing operation at 60 °C, (ii) most
clostridia can efficiently use xylose without recombinant modification, and (iii) the cellulase production and cellulose digestion can occur the same vessel. Disadvantages of CBP are (i) the lower solvent resistance of clostridia [21], (ii) comparative difficulty in genetic modification of clostridia, even despite recent developments [99], and (iii) increased energetic demands for cellulase production in anaerobic environments [57].

Thus, the effect of consolidated processing on effective cellulose conversion costs is unclear. Anaerobic, single-vessel cellulase production may appear to eliminate the tradeoff plant designers and engineers must make: should a given unit of feed be diverted to dedicated cellulase production, or should it be sent to the main fermentation tank so that it can become biofuel? However, an analogous tradeoff exists for single organisms that both produce cellulase and ferment the resulting hydrolysis products [56]. CBP may have the potential to lower cellulose conversion costs, but to be realized, the development of strains and processes which dramatically increase rates of anaerobic cellulose conversion must be developed [57].

**Syngas**

Syngas, sometimes called producer gas, is a mixture of carbon monoxide, hydrogen, and carbon dioxide. It can be prepared easily from biomass (or fossil fuels) by treatments at high temperature in the absence of oxygen [94]. Syngas can serve as the sole carbon and energy source for a variety of microorganisms, including ethanol and butanol producers. Carbon monoxide is oxidized to carbon dioxide by water via carbon monoxide dehydrogenase (CODHs). This reaction provides reducing power for anabolic CO metabolism via the Wood-Ljungdahl pathway [76], where CO is converted to acetyl-CoA via acetyl-CoA synthetase
(ACS) enzyme Figure 2-1(a).

Hydrogen can be metabolized by many of the same organisms which metabolize carbon monoxide. In some cases, however, hydrogen metabolism may be inhibited by carbon monoxide itself, or by other trace syngas components such as nitric oxide [1, 76]. Increasing the tolerance of microbes to syngas-associated stresses is an intriguing possibility, but approaches are just beginning to be studied. Nonetheless, a variety of clostridia such as Clostridium ljungdahlii, Clostridium autoethanogenum, and Clostridium carboxidivorans can simultaneously utilize both of these substrates and overproduce ethanol, butanol, and/or acetate [77, 89]. For example, Lewis and his colleagues [19, 77] have reported ethanol yields on CO of 45% the theoretical, with productivities in the neighborhood of 1 g/L/day and titers of up to 3 g/L.

Relative to synthesis of fuels from syngas via traditional chemical processes such as heterogeneous catalysis [72], fermentation-based processes may (i) be operated at lower pressures, (ii) enjoy superior specificity for the desired products, and (iii) are likely to be more tolerant of sulfur impurities in syngas than traditional heterogeneous catalysts.

**Light and carbon dioxide**

Photosynthetic organisms tap the energy in light to generate reducing equivalents from water, thus allowing fixation and use of carbon dioxide as a growth substrate. The pathway for incorporation of CO₂ is through the reductive pentose phosphate pathway, also known as the Calvin cycle. The enzyme Rubisco catalyzes the conversion of CO₂ and ribulose-5-phosphate to two trioses, which can subsequently be converted to pyruvate Figure 2-1(a).

Biomass from microbial photosynthesis is being explored for biofuels production. Com-
pared to higher plants, microbial photosynthesizers can be genetically modified more easily, and also enjoy faster life cycles and rapid growth.

Maximum attainable cell densities and productivities are important determinants of the cost of algal biomass. Generally, cell densities in outdoor algae ponds reach no more than 0.5–1.0 g/L of dry cell weight, even when supplying highly enriched sources of carbon dioxide (such as the flue gas of a coal or natural gas-fired power plant). This necessitates mobilization of considerable land and water resources for algal culture. Cost estimates made in the 1980s by a Department of Energy program for algal biomass grown in large-scale ponds was $0.19 to $0.41 per kg of biomass (unadjusted for inflation) [85].

An alternative is to use photobioreactors for growth. Algal photobioreactors can obtain cell densities near 20 g/L [50], but estimates of the cost of algal biomass grown in closed bioreactor systems have been up to $12 - $32 / kg dry weight [33, 50], indicating that these systems are likely far more costly than open ponds. Various elaborate modulations of the intensity and the spatial and temporal distribution of the incident light flux have been proposed [30, 68] as methods for increasing productivities.

A U.S. Department of Energy review of its extensive research program on algal biomass production in the 1970s and 1980s details the isolation of several promising algal strains, as well as methods for genetic transformation of algae, and improvement of lipid yields [85]. Representative productivities in outdoor pond systems were 15–20 g/m²/d. Lipid contents of ~40% w/w and specific growth rates of 2 d⁻¹ were representative. More recently, robust genetic methods have been reported for *Chlamydomonas reinhardtii*, *Chlorella* subspecies, and other algae [16, 92].

Algal biomass may become an attractive option for fuels production if metabolic engi-
neering can boost productivities, photosynthetic efficiencies, or lipid content dramatically over historical levels, or if carbon taxes or emissions caps necessitate CO₂ capture from flue gas emissions from fossil fuel power plants.

2.2.2 Products

Ethanol

Ethanol is produced from pyruvate in two steps: pyruvate decarboxylase converts pyruvate to acetaldehyde, and alcohol dehydrogenase reduces the acetaldehyde to ethanol Figure 2-1. This pathway is commonly exploited in Saccharomyces yeast hosts, or in the gram-negative gamma-proteobacteria Z. mobilis or recombinant E. coli.

The maximum attainable ethanol concentration depends on the host and medium composition, but is usually between 4 and 16 wt %, depending on the host. Processes for the distillation of ethanol from dilute fermentation broth were known even in ancient Egypt [93], but distillative separation of ethanol from the fermentation broth requires considerable energy input. To provide a basis for comparison with other fuels without specifying separation process specifics, we estimated the reversible work of separation for ethanol and several other fuels from an aqueous mixture as a function of concentration Figure 2-2†. For reference, energy input required to separate 9 wt% ethanol from water is about 5 MJ / kg for corn ethanol in the US [45]. (This implies that distillation operates at about 5% of the thermodynamically maximum efficiency).

Distillation represents a higher percentage of the cost of production for corn ethanol than for cellulosic ethanol, because feedstock pretreatment and cellulase production still dominate

†the Materials and Methods section occurs at the end of this chapter
projected costs for cellulosic ethanol (see Section 2.2.1 above and references therein, above).

**Butanol**

1-Butanol is a fermentation product of *Clostridium acetobutylicum* and *Clostridium bjerinkci* [25]. It is produced from acetyl-CoA through the dimerization of two acetyl-CoAs into acetoacetyl-CoA. Subsequent enzymes catalyze the four-electron reduction and dehydration of acetoacetyl-CoA to butyric acid. The butyrate can then be re-metabolized by some clostridial species via a further four-electron reduction to 1-butanol. In clostridia, acetone (not shown in Figure 2-1(b) is often produced concomitantly with butanol by decarboxylation of acetoacetyl-CoA.

The total pathway from glucose to butanol is, in principle, redox balanced. However, in some organisms, including *Clostridium acetobutylicum* and *E. coli*, hydrogen is produced via formate-hydrogen lyase [52] and thus hydrogen production represents a competing outlet for reducing equivalents. In principle, hydrogenase activity or alternate formate dehydrogenases [90] could recover these reducing equivalents.
The clostridial pathway has been functionally expressed in recombinant *E. coli* hosts, although the observed productivity fell short of what was observed for clostridia [9].

An alternative approach to butanol production that starts from pyruvate instead of acetyl-CoA has been proposed in the patent literature [22]. Decarboxylative dimerization of pyruvate to acetolactate is followed by reduction to α-ketovalerate, decarboxylation, and further reduction to butanol. The enzymes composing this pathway are not usually found in the same organism. Notably, this pathway results in production of the 2-butanol isomer. Relative to 1-butanol, the 2-butanol isomer may be preferred as a fuel.

Butanol has a considerably lower reversible work of separation from water than ethanol, which implies that its separation from fermentation broth will be considerably more energy efficient (Figure 2-2). However, it is far more toxic to most microorganisms. In clostridia, 13 g/L has been reported as a usual 1-butanol titer [24]. Butanol yields on glucose of 0.76 mol / mol and productivities of 0.76 g/L/hr have been reported for clostridial cultures sparged with hydrogen [24].

**Higher lipids**

Acetyl-CoA is the ultimate source of carbon for fatty acid biosynthesis. In many organisms, long-chain fatty acids biosynthesized through the ATP-requiring carboxylation to malonyl-CoA, followed by cycles of decarboxylative addition of malonyl-CoA to acyl units and β-reduction Figure 2-1(b). Long-chain fatty acids, as either methyl esters [111] or decarboxylated hydrocarbon derivatives [58] are attractive as renewable substitutes for diesel fuel.

Oils from higher plants are also synthesized in this way. Easily extracted oils from palm trees and soybeans can be converted to biodiesel and are widely used for biofuels production.
The obstacle that has stopped wider adoption of this technology is the insufficient availability of low-cost feedstocks [36]. Nonetheless, about 0.225 billion gallons (0.805 billion liters) were produced in the United States in 2006. Like corn ethanol, government subsidies for biodiesel production are currently required to sustain commercial interest in these technologies.

Lipid production by algae has the potential to get around current biodiesel feedstock limitations [15, 34]. Improving lipid production by algae has been an important metabolic engineering goal for many years, and remains so today (see Section 2.2.1 above). More recently, interest has developed in the generation of biofuels via lipogenic (or oleaginous) yeasts and other microbes [102]. This technology would allow microbial conversion of cellulosic materials to lipids.

Lipids do not mix with water, and as such, the reversible work of separation from aqueous solution is near zero. One drawback may be that lipids are often accumulated intracellularly, which could require extraction of the lipids from crude cell pastes [33]. Separation of the lipids from undesirable biomass components, perhaps by solvent extraction or other means, will require its own energy inputs. An alternative metabolic engineering strategy is to engineer secretion of the lipid products.

Many details of fatty acid secretion even in model organisms like *E. coli* and *S. cerevisiae* are poorly understood. These areas may be attractive targets for metabolic engineering.

**Terpenoids**

Isoprenoids are a broad class of metabolites synthesized from isoprenyl pyrophosphate (IPPP), the pyrophosphate ester of 3-methylbut-3-en-1-ol, or its isomer dimethylallyl pyrophosphate (DMAP), the pyrophosphate ester of 3-methylbut-2-en-1-ol. These molecules are synthe-
sized either from glyceraldehyde-3-phosphate and pyruvate via the methylerthyritol pathway (omitted from Figure 2-1(b)), or from acetyl-CoA via the mevalonate pathway, as shown in Figure 2-1(b).

IPPP and DMAP can be dimerized or polymerized to an astonishing array of olefinic hydrocarbons or their alcohol derivatives. For fuel applications, water-insoluble liquid products such as the hemiterpenoid alcohol 3-methylbut-2-en-1-ol (prenol) [107], or perhaps monoterpenes and sesquiterpene hydrocarbons such as limonene, pinene, or cadinene, may be preferred targets.

Strains and processes capable of converting sugars to terpenoids at yields similar to the ethanol process have not yet been reported in the scientific literature. For example, a representative titer for the sesquiterpene hydrocarbon amorphadiene has been reported as about 0.5 g/L [80].

Terpenoid production has usually been examined only under aerobic culture conditions, but for large-scale, high-yield production, anaerobic processes are desired. Obviation of the oxygen requirement for terpenoid overproduction is one key metabolic engineering objective for biofuels applications of this pathway.

2.3 Requirements for yield, titer, and productivity

Ultimately, any host organism for biofuels production must perform at high yield, because of the significant cost for feedstock. Years of research has revealed the enzymological and genetic details of metabolic pathways for ethanol and butanol fermentation, and allowed their functional pathway expression in convenient host organisms [9, 42]. Carbon monoxide
dehydrogenase has been functionally expressed in recombinant *E. coli* [55], suggesting that research to develop recombinant syngas fermenting organisms may also be attractive.

Equally important, comprehensive stoichiometric models of metabolism permit the design of efficient genetic backgrounds for fuel production [2, 29, 41]. These models can also be used to design genetic backgrounds in which biomass yield is positively correlated with the desired product yield, allowing simple evolutionary strategies to be used to optimize the yield of the desired product. In this thesis, the application of these stoichiometric models to the metabolic engineering of *E. coli* for butyrate production is discussed in Chapter 3, and application to the metabolic engineering of *E. coli* for tyrosine production is discussed in Chapter 5.

However, a similarly diverse and effective array of techniques is not available to optimize productivities, or, in particular, titers. Mechanisms for dealing with the stresses associated with biofuels production, such as osmotic stress, membrane disruption, and the presence of various toxins appear to be host specific and often poorly understood [14, 95]. The selection of a fitting host organism is thus essential for the development of biofuels production processes.

### 2.4 Biofuels hosts

#### 2.4.1 Desirable properties of biofuels hosts

An optimal host for the production of possess all of the following functions: degradation of lignocellulosic biomass or at least a component thereof, metabolic conversion of the biomass to the desired product, and tolerance of high end-product and by-product titers. Given nature’s incredible biodiversity, such a host may exist in nature, but if so, no one has yet
reported it. Instead, researchers have sought to combine desirable characteristics through the introduction of recombinant DNA encoding for one of the key functions into hosts which can endogenously provide the other functions.

The characteristics that have made laboratory strains like *E. coli*, *S. cerevisiae*, and *Bacillus subtilis* so popular are relevant for the development of industrially useful strains: genetic competence, the availability of well-characterized metabolic engineering modules for gene expression or deletion (e.g. promoter libraries [3], termination sequences, repressor-inducer systems, well-characterized plasmids, chromosomal integration cassettes, mathematical network models of metabolism, etc.) that allow manipulation of the genotype in different ways, and perhaps most importantly, the vast literature on organismal physiology stemming from decades of fundamental research.

Industrially useful organisms must be genetically stable without antibiotic-based selection, be active without the use of expensive inducers, and, as much as possible, not require exacting sterilization procedures in order to prevent contamination. Thermophilic [81] or acidophilic organisms are especially well suited to the latter requirement. They must also exhibit well-understood production kinetics. For example, solventogenesis in *C. acetobutylicum* is a stationary-phase phenotype controlled by global transcriptional regulators that respond to different cues [78, 106]. The behavior of such systems must be understood in order to implement butanol fermentations. Many of these traits are phenotypes conferred by expression of many, often unknown genes, and so these traits are more difficult to transfer into a desired host. For this reason, strains that are easy to manipulate and are well characterized are good starting points for the development of production platforms.
2.4.2 Strain improvement

Improving environmental tolerance has been an active area of study because maximum productivity and titer are key determinants for the profitable fermentative production of commodity chemicals. In general, genetic changes for strain improvement may be of two types: rational (or directed) and random (or combinatorial). The first is straightforward when a genetic sequence that is likely to impact a trait of interest is known. The second is preferred when this knowledge is lacking, and relies in our ability to construct genetically diverse populations and screen them.

Random approaches have gained special attention, because detailed physiological mechanisms of tolerance are largely unknown. Traditionally, these efforts have been focused on long term adaptation or serial rounds of mutagenesis and selection, usually referred to as classical strain improvement (CSI). The mutagenesis step is accomplished with chemical or physical mutagens (e.g. nitrosoguanidine or ultraviolet light), producing libraries of variants that can be then screened for improved tolerance to the stress of interest. A clear limitation of this approach is the non-transferability and non-traceability of the resulting genotypic changes. More recently, a variety of random-search (i.e. library-based) methods for strain improvement have emerged that are based on phenotypic alteration using tractable elements. These include artificial transcription factors [69], global transcription machinery engineering [4, 5], libraries of siRNAs [11], randomized ribozymes [61], knockout and overexpression libraries [44], and others.

This fact is underscored by the data in Figure 2-3, which shows the tolerance of commonly-used mesophilic, facultatively anaerobic biofuels production hosts to two stresses. *B. subtilis*
outperforms the other six hosts by a large margin in the case of tolerance to 1-butanol, but is very sensitive to the fermentation inhibitors in dilute-acid pretreated [83], quicklime-neutralized corn stover hydrolysate (CSH). Conversely, Corynebacterium glutamicum is comparatively much more tolerant to the CSH than to butanol. The yeasts S. cerevisiae and Yarrowia lypolytica are reasonably tolerant to both of these stresses. Data like that in Figure 2-3 is useful for qualitatively comparing the different hosts. But, without tailoring temperatures, growth media, and, in the case of CSH, pretreatment and neutralization conditions to each particular host, it is difficult to extrapolate such data to process conditions.

2.4.3 Engineering environmental tolerance to common process stresses

Environmental stresses come in many disparate forms, only some of which have been reasonably well-studied. Heat tolerance has received significant attention, as running fermentations
at higher temperatures can reduce cooling costs and the risk of contamination. Heat increases the fluidity of the cell membrane (Balogh et al., 2005; Mansilla et al., 2004; Shigapova et al., 2005), which in turn interferes with energy transduction and pH maintenance [88]. In addition, it causes protein denaturation and aggregation, directly affecting most cellular functions [79]. Thermotolerance has been enhanced both by rational and random schemes. Overexpression of heat shock protein hsp22.4 (from *Chaetomium globosum*) in *S. cerevisiae* increased cell viability after a 4-hr shift to 51°C [53]. This approach has been extended to other organisms, from bacteria [27] to very complex multicellular systems [26, 105]. Thermotolerance has also been improved by pre-conditioning [86], overproduction of metabolites such as trehalose [73], and supplementation with protectants such as betaine and choline [39]. These three and other compounds act similarly: they change the osmolarity of the cytoplasm and stabilize water-protein interactions, thereby preventing protein aggregation (Cayley and Record, 2003; Conlin and Nelson, 2007; Ignatova and Gierasch, 2006).

Combinatorial approaches have also been successful. For example, subjecting a mutagenized *Saccharomyces* population to serial freeze-thaw cycles followed by selection has delivered mutants with increased tolerance to multiple stresses, including heat [14]. Screening of artificial transcription factor libraries has been used to engineer thermotolerance in *S. cerevisiae* [69] and to determine a gene responsible for this phenotypic alteration in *E. coli* [69].

Osmotolerance has also been studied as a target for strain improvement, as high substrate, product or salt concentrations that increase osmotic pressure are commonly encountered in industrial fermentations [97, 101]. Some of the same strategies outlined for enhancing thermotolerance can be or have been applied. For example, trehalose and betaine have been
reported to increase osmotolerance in *E. coli* [60], and the protective role of trehalose in yeast has been reported [40]. Transcriptional engineering has delivered yeast strains improved for growth in high salt [69] and glucose [4]. Similarly, this method has improved the tolerance of *E. coli* to a variety of stresses, as well as improved metabolite overproduction [5].

### 2.4.4 Biofuels-specific stresses

The problem of environmental tolerance is not only a challenge for fermentation technology in general, but is particularly relevant to the production of biofuels for two reasons. First, when the production strain cannot degrade lignocellulosic material directly, the substrate for fermentation may contain toxic compounds. This has been widely discussed for feedstocks derived from lignocellulosic hydrolysates, as pre-treatment by acid hydrolysis produces a mixture of oligosaccharides, organic acids, phenolic derivatives, and furans [82], all of which can be inhibitors of growth for many microorganisms. Second, the product of fermentation may itself be toxic, in which case the titer could be inherently limited. A classic example of this is ethanol production in *Saccharomyces* [4].

Lignocellulosic derivatives that have received most attention are furfural, hydroxymethyl-furfural (HMF), acetic acid, and phenolic compounds. The amount and identity of the inhibitors after detoxification of hydrolysates depends on the method used (overliming, laccase treatment, charcoal, etc.) [47]. Although toxicity and detoxification issues have been mostly explored for ethanol, some studies for other fermentations such as butanol exist [23]. In general, there seems to be no “magic bullet”: because different compounds exert toxicity through different mechanisms, and because the inhibitors may interact in complex ways [23, 47, 95], endowing tolerance to inhibitors in the feedstock through rational approaches
and/or simple process modifications seems unlikely. However, partial successes on this front have been reported. Overexpression of ADH6 in S. cerevisiae, an 5-HMF reducing enzyme, has enhanced detoxification of HMF, but no increase in ethanol productivity was reported [71]. A similar effort with ZWF1, encoding a glucose-6-phosphate dehydrogenase, resulted in higher furfural tolerance [31]. Overexpression of the enzyme phenylacrylic acid decarboxylase (from PAD1) resulted in Saccharomyces strains improved in ethanol productivity in the presence of ferulic and cinnamic acids [48].

Manipulation of the fermentation pH has been used for alleviating toxicity to acetic acid, as toxicity is mainly effected by protonated carboxylic acid species rather than carboxylate anions [49, 96]. However, pH control is undesirable because of the additional cost associated with it, and because low pH reduces the risk of contamination as discussed above. Transferring a stress-response gene from an acid-resistant Oenococcus oeni strain resulted in an E. coli strain with improved low pH tolerance [63]. These or similarly constructed hosts may be better suited for fermentations at low pH.

Improvements from random approaches have also been reported. Genome shuffling of ethanologenic Candida krusei has delivered acetic acid-resistant mutants that perform better than the parent in ethanol fermentations in the presence of the inhibitor [104]. The usefulness of classical strain improvement methods for improving tolerance of yeasts to lignocellulosic hydrolysate components has also been reported [54, 91]. Similarly, studies describe that Pichia stipitis long-term adapted to increasing concentrations of hardwood hydrolysate partially neutralized or alkalinized with lime had higher ethanol productivity and titer [67, 66].

Product inhibition has also been addressed, mainly for yeast ethanol fermentations. Most
fuels have solvent-like properties, and are thought to have broadly similar impacts on cell physiology. In general, solvents and other lipophilic hydrocarbons partition into the membrane and disrupt its fluidity, which in turn results in ion leakage [32, 43]. As with heat, solvent-induced increases in membrane fluidity cause dissipation of transmembrane pH gradients and thus leads to cellular energy deprivation and eventually growth ceases.

Many ethanol-tolerant yeast strains have been independently isolated in the context of alcoholic beverage production, through hundreds of years of artificial trait selection. These are usually poorly characterized, but application of modern techniques has helped in the elucidation of possible ethanol-tolerance mechanisms in these strains [87, 103]. Following a comparison of sake-brewing and laboratory strains using microarrays, overexpression of tryptophan-biosynthesis genes (TRP1-5) resulted in enhanced ethanol tolerance [37]. The mechanism behind this phenotype seems to be similar to that of strains with higher levels of trehalose [46], and for the same reasons outlined for thermotolerance.

The complexity of fuel/solvent tolerant phenotypes has invited the use of random approaches. A mutant of the TATA-binding protein, coded by SPT15, increased yeast tolerance to high-glucose, high-ethanol conditions through the alteration in expression of hundreds of genes [4]. The improved phenotype could not be recovered from localized changes in gene expression, which is in tune with the complex nature of the stresses and underlines the need for global cellular engineering for strain improvement. A similar effort in *E. coli* delivered a mutant sigma factor with improved ethanol tolerance [5]. Overexpression of two genes in *C. acetobutylicum* that were found by screening a genomic DNA library resulted in increased butanol tolerance [12]. Other localized genetic manipulations have been explored, like null mutantions ura7 and gal6 that were selected by challenging a knockout library of
Saccharomyces in ethanol [108]. The mutant strains grew and consumed glucose faster in the presence of 8% ethanol compared to the wild-type, through changes in membrane lipid composition and other mechanisms that were not fully understood. CSI programs have also delivered ethanol tolerant E. coli [110] and butanol-tolerant C. beijerinckii mutants [8, 74].

Even with the substantial research interest that this problem has attracted, environmental tolerance for biofuel fermentations continues to be a challenge. A main reason for Saccharomyces fermentations of ethanol being traditionally preferred is that this microorganism has been bred to precisely this end for millions of (microbial) generations. Artificial evolution has helped address many of the limitations, but it has not yet had time to solve others (like tolerance to lignocellulosic hydrolysates) that are of practical significance today. Rational genetic manipulations based on omics data will likely be part of the answer, but random methods will also play a central role.

2.5 Conclusions

Today's predominant microbially produced biofuel is starch-derived ethanol. However, further expansion of production capacity will require use of lignocellulosic feedstocks. The two best-developed technologies for conversion of lignocellulosics to fuel are (i) pretreatment and enzymatic hydrolysis to fermentable sugars, or (ii) gasification to syngas. Microbial photosynthetic processes may also render carbon dioxide as an attractive biomass feedstock.

An additional and orthogonal problem with starch-derived ethanol is the chemical properties of ethanol itself. Other fuel molecules, such as butanol, or liquid hydrocarbons of lipid or terpenoid origin, have higher energy densities, mix less readily with water, can be easier
to separate from fermentation broth, and are compatible with the existing fuel distribution infrastructure. A primary barrier to the introduction of these technologies is insufficient microbe performance. That is, metabolic engineering research to improve the yield, titer, and productivity of microbial strains will be essential to the development of later-generation biofuels. We conjecture that yield may be the easiest of the trifecta to engineer, because the experimental tools for pathway manipulation, and the metabolic models required for evolutionary strain optimization are already in place. However, a similar toolset for improvement in productivity and titer is not available; in many cases regulatory pathways governing are incompletely understood. Further elucidation of these pathways is a much-needed avenue for future research.

2.6 Materials and methods

The reversible work of separation was calculated using ASPEN Plus by simulating the mixing of a pure fuel and pure water stream. The Gibbs energy change on mixing is equivalent to the reversible work of separation. The property model was UNIFAC-LL, which is suited to cases of liquid-liquid phase equilibria. For butanol, alpha-pinene and methyl oleate, phase separation was predicted. The shown data is the sum of the Gibbs energy of each phase, normalized by the fuel content of each phase.

A table of the strains used is shown in Table 2.2. All cultures were grown at 30 °C in RM medium [7]. For butanol tolerance, overnight cultures were inoculated 1:10 into fresh RM medium, and after 2 hr, a 1:10 dilution of the fresh culture was made into RM medium containing various amounts of 1-butanol (Sigma Aldrich, St. Louis MO). To prevent butanol
Bibliography


66


[55] H. K. Loke, G. N. Bennett, and P. A. Lindahl, “Active acetyl-coa synthase from
clostridium thermoaceticum obtained by cloning and heterologous expression of acsab
2000.

[56] L. R. Lynd, “Overview and evaluation of fuel ethanol from cellulosic biomass:
Technology, economics, the environment, and policy,” Annual Review of Energy and


deoxygenation of fatty acids and their derivatives,” Energy & Fuels, vol. 21, no. 1,

[59] A. McAloon, F. Taylor, W. Yee, K. Ibsen, and R. Wooley, “Determining the cost of
producing ethanol from corn starch and lignocellulosic feedstocks,” National Renewable
Energy Laboratory, National Renewable Energy Laboratory, US Department of

[60] E. N. Miller and L. O. Ingram, “Combined effect of betaine and trehalose on osmotic
tolerance of escherichia coli in mineral salts medium,” Biotechnol Lett, vol. 29, no. 2,

M. Sano, Y. Kato, Y. Takagi, Y. Tanaka, M. Warashina, T. Kuwabara, S. Y. Sawata,
technologies: demonstration of usefulness of libraries of ribozymes and short hairpin
rnas (shrnas),” Nucleic Acids Symp Ser (Oxf), no. 49, pp. 91–2, 2005.

xylose, and arabinose by mixed cultures of two genetically engineered zymomonas
1998.

of a recombinant strain of escherichia coli expressing genes from the acidophilic


This Chapter:

- Introduces the clostridial pathways for the biosynthesis from glucose of butyric acid and butanol.
- Describes the construction and characterization of an *E. coli* host strain for the expression of clostridial pathway genes.
- Illustrates two classical approaches for identifying rate-limiting genes in a metabolic pathway: genetic complementation and substrate feeding.

### 3.1 Introduction

Butyric acid is a saturated C4 fatty acid that can be reduced to butanol, and as such, an efficient means to produce it from glucose or other biomass-derived starches would enable efficient biofuels production from agricultural products or other biomass sources, as discussed in Chapter 2.

However, as is well-known, typical fermentation products for *E. coli* do not include butyrate, and include instead a mixture of ethanol, lactic acid, and acetic acid. This chapter describes the two related lines of research: (i) the engineering and characterization of an
E. coli strain in which pathways for ethanol, lactate, and acetate formation pathways have been deleted, rendering it unable to form its usual fermentation products, and (ii) the insertion into this host of clostridial butyrate biosynthesis genes with the goal of characterizing and optimizing this pathway as recombinantly expressed in E. coli.

3.1.1 Butyric acid fermentation: a case study in synthetic biology for biofuels production

In Clostridium tyrobutyricum ATCC 25755, butyrate is the final product of glucose fermentation [29]. No secondary fermentation of butyrate to butanol or acetone is observed. However, this organism is comparatively unstudied compared to the clostridial type strain Clostridium acetobutylicum 168, which also naturally ferments glucose to butyrate. This and many other clostridial strains, however, subsequently ferment butyrate to a mixed spectrum of butanol, acetone, and ethanol following glucose depletion from the medium. This secondary fermentation is believed to be a part of the sporulation process of Clostridium spp. The two-phase nature of the fermentation, the difficulty of genetic manipulation of clostridia, and the lack of reproducibility and process continuity associated with spore formation render Clostridium acetobutylicum 168 an unattractive host for metabolic engineering.

An attractive possibility is to simply move the pathway for clostridial butyrate production into a well-studied, robust host organism such as E. coli. Many of the tools of metabolic engineering and synthetic biology discussed in Chapter 1 operate exclusively in E. coli, or are more advanced in E. coli. For example, understanding the sequence-level factors which govern promoter function and thus control gene expression are far better understood in E. coli. This understanding has been exploited by the creation of a variety of molecular
tools for the control and modulation of gene expression (for example, see Chapter 6.

The pathway for butyrate production in *Clostridium* is shown in Figure 3-1. The genes involved in the pathway are:

1. **acetoacetyl-CoA thiolase.** In clostridia, this gene is known as *thl1*. *E. coli* possesses this activity as part of the beta oxidation pathway for catabolism of fatty acids, but this activity can be induced only by long chain (>C12) fatty acids [8]. Unlike any of the other pathway enzymes, this enzyme is present in *E. coli*, where it is encoded by *atoB*.

2. **beta-hydroxybutyryl-CoA dehydrogenase.** In clostridia, this gene is known as *hbd*, and requires NADH for activity. Overexpression of clostridial *hbd* in the *E. coli* cytoplasm resulted in a functional gene product. *E. coli* possesses this activity as part of the beta oxidation pathway for catabolism of fatty acids, but this activity can be induced only by long chain (>C12) fatty acids [8]. A similar activity has been expressed in *E. coli* strains engineered for the production of poly-hydroxybutyric acids, but the version of the activity used for PHB synthesis requires NADPH, not NADH, and yields the *R* enantiomer of 3-hydroxybutyryl-CoA, unlike the clostridial version.

3. **crotonase.** In clostridia, this gene is known as *crt*. Overexpression of clostridial *crt* in the *E. coli* cytoplasm resulted in a functional gene product [4]. *E. coli* possesses this activity as part of the beta oxidation pathway for catabolism of fatty acids, but this activity can be induced only by long chain (>C12) fatty acids [7].

4. **butyryl-CoA dehydrogenase.** In clostridia, this gene is known and *bcd* and requires a redox-active flavin cofactor for activity. *E. coli* natively has this activity but it
is induced aerobically only by long chain (>C12) fatty acids [7]. Overexpression of clostridial \( bcd \) in the \( E. \ coli \) cytoplasm resulted in a functional gene product in the hands of at least one group. It is believed that interaction between Bcd and an electron transfer flavoprotein (ETF) encoded by clostridial genes \( etfAB \) may be required for redox turnover of the flavin cofactor and thus for activity [4, 8].

5. **phosphotransbutyrylase.** In clostridia, this gene is known as \( ptb \). \( E. \ coli \) does not have this activity natively, but the clostridial \( ptb \) gene has been expressed in \( E. \ coli \) and was functional [9].

6. **butyrate kinase.** In clostridia, this gene is known as \( buk \). \( E. \ coli \) does NOT have this activity natively, but the clostridial \( buk \) gene has been expressed in \( E. \ coli \) and was functional [9].

These clostridial enzymes are the only activities which must be provided to an \( E. \ coli \) host in order to provide a capacity for fermentation of glucose to butyrate. Thus, a set of apparently well-characterized "building blocks" are available. This pathway thus appears well-suited to a synthetic biology approach. The building blocks which require assembly into a metabolically functional systems all come from a comparatively well-studied clostridium, \( C. acetobutylicum \).

### 3.1.2 Metabolic engineering of \( E. \ coli \) for the evolution of increased butyrate yield

The yield of product per unit of nutrient fed is a critical parameter which determines the eventual cost of production for nearly any commodity bioproduct, as discussed in Chapter 2.
E. coli is not a native butyrate producer; its native fermentative pathways produce ethanol, lactate, and acetate. Eliminating pathways that compete for carbon with target pathways has long been an important metabolic engineering strategy for increasing the yield.

Recently, this concept has been extended into engineering genetic backgrounds which genetically favor the evolution of increased production through a desired pathway [37]. These methods are based on linear programming and parameter-free, stoichiometric models of metabolic networks, as discussed in Chapter 5. The key observation is that in strains rendered stoichiometrically restricted by gene deletion or enzyme inactivation, the omnipresent selective pressure for faster cell growth can be coupled to the flux through a particular pathway. Thus, not only can eliminating competing pathways stop metabolic flux of carbon into
unwanted intermediates, it can also couple the flux through a target pathway to a flux for cell growth.

A variety of computational approaches have been reported for predicting “optimal” sets of stoichiometric restrictions for the coupling of biomass formation to flux through a target pathway.

In simple cases, involving changes in a small number of fluxes for native \textit{E. coli} metabolic pathways, adaptive evolution has been shown to occur. Examples include the evolution of wild-type \textit{E. coli} toward optimal growth on glycerol, as discussed by Ibarra et al. [24], or the evolution of acetate and ethanol-deleted \textit{E. coli} toward homofermentative lactate production, as studied by Fong and co-workers [16, 14].

Transcriptional, enzymatic studies, as well as metabolic flux analysis on the resulting strains revealed common patterns to the adaptive metabolic evolution of deletion strains of \textit{E. coli}. As Fong et al. put it [15]:

\begin{quote}
\text{[A]n introduced genetic perturbation should force a redistribution of fluxes through the network as an immediate rescue solution, and evolution would then entail a process of refining this newly established initial state. Generally, our results fully support this hypothesis.}
\end{quote}

The evolutionary refinement in question is believed to involve subtle changes in transcriptional regulatory networks of various metabolic operons. Thus, it is unknown how the process of adaptive evolution will respond to the presence of artificially-introduced, non-native metabolic pathways. Put differently, if adaptive evolution proceeds by regulatory modifications, it is unclear if the evolutionary process could “see” or respond to the presence of the recombinant pathway.
3.2 Materials and methods

3.2.1 Stoichiometric envelopes for gene deletions in *E. coli*

The correlation between the yield on glucose of *E. coli* cell biomass and either succinate or butyrate was made using the iJE660b stoichiometric model of *E. coli* metabolism described in Chapter 5. The MATLAB routines scanProdGrowthEnv1 and makeProdGrowthEnv1, given in Appendix D were applied to the iJE660b model in order to derive the growth-product stoichiometric envelopes.

3.2.2 Strains and plasmids

Strains and plasmids used in this work are summarized in Table 3.1.

The *pal* host strain

*E. coli* K12 Δpta, *E. coli* K12 ΔldhA, and *E. coli* K12 ΔadhE were constructed via the method of Datsenko and Wanner. These genotypes were combined through a series of P1-mediated transductions of the corresponding pta::kan, ldhA::kan, and adhE::kan alleles followed by elimination of the kanamycin marker with FLP-mediated recombination by pCP20. The resulting triple Δpta ΔldhA ΔadhE triple knockout was called the *pal* strain.

For expression of the lambda Red expressing electrocompetent cells, an IPTG-inducible variant of the pKD46 plasmid described by Wanner that we have found to give superior repeatability was used.

The *pal*(DE3) derivative of *pal* was created with the DE3 lysogenization kit (Merck
KGaA, Darmstadt, Germany) as per the manufacturer's directions.

The \textit{pal}-\textit{le} strain was created by evolving \textit{E. coli pal} pZE21-pab pHATet-CBEH (see below) in MOPS-butyrate medium for 130 hours after inoculation from an overnight 4 mL culture in LB broth. After a 110 hr, this strain eventually grew on butyrate as a sole carbon source. Colonies were streaked from the exponentially growing culture in MOPS-butyrate onto LB agar. Two colonies were picked, grown overnight in LB broth, and re-inoculated into MOPS butyrate, where they were found to have a lag phase of less than 1 hr. This strain is likely a \textit{fadR} deletion mutant but has not been exhaustively characterized.

\textbf{Construction of plasmids expressing clostridial butyrate biosynthesis genes}

The \textit{ptb/buk} operon was cloned from \textit{Clostridium acetobutylicum} ATCC 824 genomic DNA using primers \texttt{MT_ptb-buk.Sense} (\texttt{TCCCCCGGGGGAGTTTAGGGCAAAAGTTTTTATAAACATGGGTACTGG}) and \texttt{MT_ptb-buk.Anti} (\texttt{GGGTTACCCCGAGTGTACGACCAGTGATTAAGAGTTTTAA}) endowed with Xmal and KpnI restriction sites (emphasized in primer sequences. This PCR product was digested and ligated to KpnI/XmaI digested pZE21-gfp(ASV) in order to create pZE21-\textit{ptb/buk}. The \textit{atoB} gene was cloned from \textit{E. coli} K12 MG1655 genomic DNA with primers \texttt{CF184_atoB Sense_KpnI} (\texttt{ACTCGGGTACCATGAAAAATTGTGTCATCGTCAGTG}) and \texttt{CF185_atoB antiHindIII} (\texttt{AGAGGAAGCTTTAATTCAACCGTCAATCACCAC}) (restriction sites emphasized), digested with HindIII and KpnI, and ligated to the HindIII/KpnI fragment of pZE21-gfp(ASV) to create pZE21-\textit{atoB}. The promoter-gene \texttt{P}_{LTetO1}\texttt{-atoB} expression cassette from pZE21-\textit{atoB} was amplified with primers \texttt{CF188_P(LTetO1)_atoB_Sense} (\texttt{AATII}) (\texttt{GGCTTCCCAACCTTACCAGG}) and \texttt{CF189_P(LTetO1)_atoB_AntiAatII(Xbal)} (\texttt{CCTCGACGTCTAGATATTACCGCCTTGGTACGCTG}) and the resulting fragment was digested with AatII and ligated into a an AatII-digested frag-
ment of pZE21-ptb/buk in order to create pZE21-pab. pRK415-BCS was digested with SacI and AvrII, and the fragment containing the BCS operon with genes *crt, etfAB, bcd,* and *hbd* was ligated to XbaI-digested pHACm in order to create pHACm-CBEH. This plasmid carries genes for both tetracycline resistance (from pRK415-BCS) and chloramphenicol resistance (from pHACm). The CmR-conferring *cat* gene was disrupted in pHACm-CBEH by removing the 38bp MscI and NcoI linker region and re-ligating the plasmid, in order to form pHATet-CBEH.

### 3.2.3 Culturing conditions

Batch anaerobic cultures were maintained in 15 mL glass tubes (Bellco Glass, Inc.) stoppered with a butyl rubber septum. For oxygen-limiting culture, the septum was pierced with a 26-gauge or 27-gauge syringe needle. Antibiotics were used as required [38].

Anaerobically, *E. coli* requires bicarbonate for growth at low cell densities [23]. In batch anaerobic cultures without bicarbonate or CO₂ supplementation, this requirement results in long lag phases as sufficient bicarbonate must be accumulated by the metabolism of inoculated cells before growth can begin. In continuous cultures, gas sparging can strip away bicarbonate as CO₂, stopping growth. To avoid either possibility, MOPS minimal medium (Teknova Inc., Hollister, CA) was supplemented with 10 mM ammonium bicarbonate.

The chemostat design used here was modified from a previously reported [13, 31] configuration. Medium flow was driven by cylinder pressure (through sterile gas filters) rather than a peristaltic pump, into a three-hole-stoppered 250 mL flask immersed in a 30 °C water-bath incubator. One stopper hole provided for drop-wise continuous medium entry. Bacterial growth in the medium inlet tube was prevented by performing all experiments in the cold
Table 3.1. Strains and plasmids used in this chapter.

<table>
<thead>
<tr>
<th>Source strains</th>
<th>Description</th>
<th>Reference or Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>E. coli K12 MG1655</td>
<td>wild-type</td>
<td>ATCC 47076, [20]</td>
</tr>
<tr>
<td>pal</td>
<td>K12 Δpta ΔldhA ΔadhE</td>
<td>this work</td>
</tr>
<tr>
<td>BL21(DE3)</td>
<td>wild-type E. coli strain with chromosomal T7 TRNAP</td>
<td>Novagen (now Merck KGaA, Darmstadt, Germany)</td>
</tr>
<tr>
<td>pal(DE3)</td>
<td>pal strain infected with defective lambda phage in order to confer chromosomal copy of T7 RNAP</td>
<td>this work</td>
</tr>
<tr>
<td>L32</td>
<td>K12 fadR601 atoC512(Con) atoD32</td>
<td>Yale E. coli Genetic Stock Center</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Plasmids</th>
<th>Description</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>pKD46</td>
<td>temperature-sensitive, low-copy AmpR plasmid for expression of lambda Red recombinase system</td>
<td>[11]</td>
</tr>
<tr>
<td>pCP20</td>
<td>temperature-sensitive, AmpR expression module for the FRT recombinase</td>
<td>[11]</td>
</tr>
<tr>
<td>pJM12</td>
<td>derivative of pKD46 with PBAD promoter replaced by Ptac</td>
<td>this work</td>
</tr>
<tr>
<td>pR415-BCS</td>
<td>source of cloned clostridial CBEH operon</td>
<td>[41]</td>
</tr>
<tr>
<td>pZE21-gfp(ASV)</td>
<td>medium-copy kanR plasmid with constitutive PtetO1 promoter</td>
<td>[12], [30]</td>
</tr>
<tr>
<td>pZE21-ptb/buk</td>
<td>pZE21 with gfp replaced by the ptb/buk polycistron from C. acetobutylicum</td>
<td>this work</td>
</tr>
<tr>
<td>pZE21-pab</td>
<td>pZE21-pab with an insertion of an atoB expression cassette driven by a separate PtetO1 promoter</td>
<td>this work</td>
</tr>
<tr>
<td>pHAc-CBEH</td>
<td>low-copy SC101 derivative for the aTc-inducible expression of the CBEH operon from C. acetobutylicum, contains both Tetr and Cmr</td>
<td>this work</td>
</tr>
<tr>
<td>pHATet-CBEH</td>
<td>variant of pHAc-CBEH with Cmr removed</td>
<td>this work</td>
</tr>
<tr>
<td>pRARE</td>
<td>pl35A-derived, Cmr plasmid expressing tRNAs for rare codons in E. coli for the improvement of the translation of rare codons (now Merck KGaA, Darmstadt, Germany)</td>
<td>[1]</td>
</tr>
<tr>
<td>pHACm-0</td>
<td>low-copy SC101 derivative without any expression cassette</td>
<td>[1]</td>
</tr>
<tr>
<td>pETDuet1-thi1-hbd</td>
<td>T7-RNAP-based expression vector for thi1 and hbd</td>
<td>[41]</td>
</tr>
<tr>
<td>pETDuet1-thi1-phaB</td>
<td>like pETDuet1-thi1-hbd, but with NAPDH/R-3HB-CoA specific phaB in place of hbd</td>
<td>[41]</td>
</tr>
<tr>
<td>pCDFDuet1-thi1-phaB</td>
<td>T7-RNAP-based expression vector for clostridial thi1-phaB operon</td>
<td>[41]</td>
</tr>
<tr>
<td>pCDFDuet1-tesB</td>
<td>like pCDFDuet1-thi1-phaB but with tesB in place of thi1-phaB</td>
<td>[41]</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Study strains</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>pal pZE21-pab pHATet-CBEH</td>
</tr>
<tr>
<td>2</td>
<td>pal pZE21-ptb/buk pHATet-CBEH</td>
</tr>
<tr>
<td>3</td>
<td>pal pZE21-atoB pHATet-CBEH</td>
</tr>
<tr>
<td>4</td>
<td>pal pZE21-pab pHACm-</td>
</tr>
<tr>
<td>5</td>
<td>pal pZE21-gfp pHACm-</td>
</tr>
<tr>
<td>6</td>
<td>pal pZE21-pab pHATet-CBEH pRARE</td>
</tr>
<tr>
<td>7</td>
<td>pal pZE21-pab pHATet-CBEH pRARE</td>
</tr>
<tr>
<td>8</td>
<td>pal pZE21-atoB pHATet-CBEH pRARE</td>
</tr>
<tr>
<td>9</td>
<td>pal pZE21-pab pHACm- pRARE</td>
</tr>
<tr>
<td>10</td>
<td>pal pZE21-gfp pRARE</td>
</tr>
<tr>
<td>11</td>
<td>pal(DE3) pETDuet1-thi1-hbd pCDF-tesB</td>
</tr>
<tr>
<td>12</td>
<td>pal(DE3) pETDuet1-thi1-phaB pCDF-tesB</td>
</tr>
<tr>
<td>13</td>
<td>pal(DE3) pETDuet1-thi1-hbd pCDF-tesB</td>
</tr>
<tr>
<td>14</td>
<td>pal(DE3) pETDuet1-thi1-hbd pCDF-tesB</td>
</tr>
<tr>
<td>15</td>
<td>BL21(DE3) pETDuet1-thi1-hbd pCDF-tesB</td>
</tr>
<tr>
<td>16</td>
<td>BL21(DE3) pETDuet1-thi1-hbd pCDF-tesB</td>
</tr>
<tr>
<td>17</td>
<td>BL21(DE3) pETDuet1-thi1-hbd pCDF-tesB</td>
</tr>
<tr>
<td>18</td>
<td>BL21(DE3) pETDuet1-thi1-hbd pCDF-tesB</td>
</tr>
<tr>
<td>19</td>
<td>pal pZE21-pab pHATet-CBEH evolved on butyrate</td>
</tr>
<tr>
<td>911</td>
<td>strain 11 transformed with pHATet-CBEH</td>
</tr>
<tr>
<td>912</td>
<td>strain 12 transformed with pHATet-CBEH</td>
</tr>
<tr>
<td>914</td>
<td>strain 14 transformed with pHATet-CBEH</td>
</tr>
</tbody>
</table>
room, and by using a 1 mL filter pipette tip as a protective nozzle at the flask end of the medium inlet tube. One other stopper hole served as the gas inlet, and the third stopper hole served as gas and liquid outlet. A U-shaped hook of glass tubing extended down from the stopper into the liquid in the flask. By adjusting the level of the glass hook, the liquid holdup in the flask could be adjusted between 50 and 250 mL. Dilution rates could vary between 0.0 and 0.5 hr⁻¹. Gas sparging was at 1 vvm with either 99:1 N₂:CO₂, or with 74:21:5 N₂:O₂:CO₂. Sampling done by tilting flask to force 2-3 mL of liquid into outlet tube. This minimizes fluid residence in outlet tube to a few seconds.

### 3.2.4 Characterization of butyrate pathway function

All growth experiments with non-glucose carbon sources were in MOPS minimal medium [32] supplemented with 10 mM ammonium bicarbonate and 20 mM carbon source. Stocks of carbon sources were adjusted to neutral pH before addition to the medium. IPTG was used at a concentration of 1 mM as required. *E. coli* LJ32 [26] carries *atoC* constitutive mutation, a *ΔfadR* mutation, which normally allows growth on butyrate as the sole carbon source, but for an *ΔatoD* mutation which abolishes growth. This strain was transformed with pZE21-ptb/buk or pZE21-pab in order to test for *atoB, ptb*, and *buk* functionality by monitoring growth (as A₆₀₀ at 600 nm) in oxygen-limited hungate tubes. Strain *E. coli* pal(DE3) was transformed with pETDuet1-thi1-hbd and pCDFDuet1-tesB, grown in LB-glucose (10 g/L) overnight, inoculated into 100 mL of MOPS-glucose (50 g/L) in a 250 mL flask under oxygen-limiting conditions (the flask was sealed with a butyl rubber stopper which had been pierced with a 23-gauge syringe needle). After 123 hr of culture, the pH had fallen to below 6.0, and 9.0 g of CaCO₃, along with 30 μm of fresh IPTG and antibiotics, was added to the
culture. Crotonic and 3-hydroxybutyric acids were detected by injection of 10 μL of late exponential-phase culture supernatant into a Shimadzu (Shimadzu Corp., Japan) HPLC system equipped with a Biorad HP87X-H column (Bio-rad, Inc., Hercules, CA). Detection was by UV absorption at 210 nm, and quantification was relative to authentic standards of sodium 3-hydroxybutyrate or crotonic acid (Sigma-Aldrich, St. Louis, MO). With a mobile phase of 14 mM sulfuric acid at 50°C, the retention time of 3-hydroxybutyrate was 12 min and crotonate eluted at 22.5 min. Glucose consumption was monitored with a YSI 7100 bioanalyzer equipped with glucose-sensitive enzyme electrodes.

3.3 Results

3.3.1 Butyrate-growth and succinate-growth stoichiometric envelopes

The deletion of the genes *pta*, *adhE*, and *ldhA* from *E. coli* K12 metabolic model iJE660b led to a stoichiometric environment where biomass formation was positively linked to succinate production, as shown in Figure 3-2(a), although a correlation was also apparent in a double-knockout Δ*adhE*Δ*ldhA* background, unable to produce ethanol or lactate, metabolites which provide outlets for the cellular reducing power (NADH and NADPH) generated by glycolysis. In the pal strain, growth without any succinate production was possible only up to biomass yields on glucose of 0.003 gDCW / 1 mmol glucose, or about 0.01 gDCW / g glucose. This is about ten times lower than usual anaerobic biomass yields on glucose for *E. coli*. At the maximal biomass yield in the pal strain of 0.04 gDCW / g glucose, the stoichiometric succinate yield approaches 1 mol / mol glucose. Succinate production in *E. coli* has been studied by several workers [25, 39], who have found that improved succinate production does
result from eliminating pathways for the production of acetate, ethanol, and lactate. These results were consistent with the theoretical prediction. Previous workers have noted that the theoretical yield of succinate on glucose is highest when the *E. coli* metabolic network is given unrestricted ability to uptake carbon dioxide. Figure 3-2(b) shows the stoichiometric envelopes when CO₂ was freely available. Growth yields for the *pal* strain can reach 0.017 gDCW / mmol glucose, or about 0.094 gDCW / g glucose. For this level of growth, succinate must be produced at a rate near 1 mol / mol glucose, as in (a).

Introduction of the clostridial butyrate biosynthesis pathway reactions (Figure 3-1) into the iJE660b stoichiometric model of *E. coli pal* metabolism caused a shift in the growth-optimal outlet for reducing equivalents. Compare the stoichiometric envelopes for succinate production and for butyrate production in Figure 3-2(b) and (c). When the butyrate pathway is available, growth yields exceed 0.094 gDCW / g glucose at a butyrate yield on glucose of only 0.1 mol / mol. The maximum biomass yield of the *pal* triple knockout is the same as the (butyrate-producing) wild-type strain, 0.032 gDCW / mol glucose, and it occurs at a butyrate yield of 7.5 mol butyrate / mol glucose.

Thus, although the *E. coli* metabolic network has alternate routes for clearing glycolysis-generated reducing equivalents under anaerobic conditions (e.g., the production of succinate), introduction of butyrate production provides a more facile route for anaerobic biomass production, at least stoichiometrically.

### 3.3.2 Expected and unexpected phenotypes of the *pal* strain

The relevance of the stoichiometric analysis was unexpectedly brought into question when characterization of the *pal* triple-knockout strain revealed that it is auxotrophic for oxygen,
Figure 3-2. Stoichiometrically feasible region for the anaerobic biosynthesis of (a) succinate and *E. coli* cell biomass from glucose as the sole carbon source; (b) succinate and *E. coli* cell biomass from glucose and carbon dioxide; (c) butyric acid and *E. coli* cell biomass from glucose as the sole carbon source. Panel (d) shows the data in (c) from an alternate visual perspective for clarity. The stoichiometric model for *E. coli* metabolism used was iJE660b, described in Chapter 5.
as shown in Figure 3-3. While many previous reports have characterized the metabolism of similar *E. coli* strains, but no one had reported oxygen auxotrophy [25, 3]. Several types of characterization support our conclusion that *E. coli* K12 MG1655 ΔldhA Δpta ΔadhE depends on oxygen for growth. No growth of the strain was observed under anaerobic conditions in very rich clostridial reinforced medium, which contains glucose, acetate, cysteine, beef and yeast extracts, starch, casein, and peptone. Nor did growth occur in minimal MOPS glucose medium. Addition of alternative electron acceptors nitrate (as 100 mM sodium nitrate) and fumarate (as 30 mM sodium fumarate) did not restore growth, showing that the auxotrophy was specific for oxygen, and not just any respiratory electron acceptor. Neither of these supplements inhibited the aerobic growth of the *pal* strain or the anaerobic growth of wild-type *E. coli* K12. In oxygen-limiting culture, the *pal* strain grew linearly, as expected for batch growth with a constant influx of a limiting nutrient, as shown in Figure 3-3(b).

Low (~4 mL) volumes of medium in 15 mL Hungate tubes allowed the growth of *pal* when the tube headspace was not sparged with nitrogen, but not when the headspace was sparged (data not shown). This experiment proves that the rate limitation in *pal* growth is gas transport of oxygen *in to* the culture, and eliminates the possibility that the linear growth results from rate-limiting transport of some growth-inhibiting waste gas *out of* the culture.

### 3.3.3 Continuous culture and the evolution of the *pal* strain

The *E. coli* ΔadhE single knockout does not grow in anaerobic minimal medium [19]. To check that our continuous culture apparatus can select for improved growth under anaerobic conditions, we inoculated 3 mL of an aerobically grown LB culture of *E. coli* K12 ΔadhE
Figure 3-3. Oxygen auxotrophy in the pal strain. (a) The growth of the pal strain in mineral salts medium with unrestricted (open squares) or impeded access to oxygen (filled squares). (b) Overnight growth, or lack thereof, of either the pal strain or its parent, E. coli K12 MG1655, in either mineral salts or rich medium, as indicated.
into the anaerobic continuous culture growing with M9 mineral medium with 2.5 g/L of glucose operating at a dilution rate of 0.042 hr⁻¹. Nitrogen:CO₂ (99:1 v/v%) was sparged through the chemostat vessel to maintain anaerobiosis. Within one day, the turbidity had increased to an A600 of ~1.0. After several more days of culture, we isolated three single colonies from the reactor. All were kanR⁺, and produced very high amounts of lactate, no ethanol, and little to no acetate. The growth rates for the three isolates as measured in batch MOPS/bicarbonate medium were 0.226 ± 0.018 hr⁻¹, as compared to -0.037 ± 0.015 for the original E. coli K12 ΔadhE strain. This result validated that our continuous culture apparatus was capable of evolving E. coli strains under anaerobic conditions.

As shown in Figure 3-3, the triple-knockout E. coli ΔptaΔadhEΔldhA strain could not be grown in purely anaerobic culture. When inoculated into the chemostat vessel at a dilution rate of 0.06 hr⁻¹ to an initial turbidity of A₆₀₀ = 0.18, no growth was observed, and instead the A₆₀₀ fell to less than 0.04 in 15 hr. After the initial decrease, the dilution rate was reduced to 0. Over the next 150 hr, the pal strain grew linearly, despite anaerobic gas sparging. We attribute this observation to a slight oxygen leak into the reactor. During this period, the accumulation and glucose depletion of all metabolic byproducts shown in Figure 3-4 was also linear (R² > 0.9) over 8 time points, and the slopes of these lines were used to calculate the yields. Carbon recovery relative to glucose consumed was 135% on a C-mole basis.

In a second experiment, the sparging gas was changed to O₂:CO₂:N₂ 21:5:74%, and the dilution rate was 0.096 hr⁻¹. At steady state, glucose conversion was 39.7 ± 1.3%, or 2.11 ± 0.08 mM of the 5.83 mM (1 g/L) in the feed. The high residual glucose at this steady state is further evidence for the oxygen requirement of the pal strain. The fate of the converted
glucose is shown in Figure 3-4. The chief metabolic byproduct of the pal strain at these conditions was malate, with lesser quantities of succinate and pyruvate. Carbon recovery relative to glucose consumed was 127 ± 8% on a C-mole basis. In either study, consumption of bicarbonate in the liquid medium and carbon dioxide in the gas feed was not measured, but both may be supplemental sources of carbon for the pal strain. If all of the observed malate and succinate resulted from the incorporation of one C-mole of CO₂ per three C-mole of glucose through the action of phosphoenolpyruvate carboxylase or malic enzyme, carbon recovery relative to glucose consumed could approach approximately ∼120%.

No experiment gave any evidence for a metabolically altered pal strain, although the total number of generations was <16 for the anaerobic experiment, and 40 for the second experiment. Weighting each generation by its average optical density (and thus population size, assuming $2.5 \times 10^8$ cells / (A₆₀₀ · mL)), leads to estimates of total cell division events as $4.4 \times 10^{10}$ or $4.4 \times 10^{11}$, respectively.
3.3.4 Genetic tests of pathway

We hypothesized that introduction of the clostridial butyrate biosynthesis pathway to *E. coli* could restore an anaerobic growth phenotype. To that end, we introduced the clostridial pathway to *E. coli* pal by transformation of two plasmids: pZE21-pab, containing two operons for the expression of the bicistronic *ptb-buk* operon, and the monocistronic *atoB* operon, both from constitutive, strong P_LtetO1 promoters [2]; and pHACm-CBEH, expressing the polycistronic *crt-bcdetfAB-hbd* operon from an IPTG inducible P_{lac} promoter.

No butyric acid was detected in cultures of the transformed strain. Micromolar amounts of crotonic acid were the only observable product of the transformed pal strain (strain #1 in Table3.1). The integrity of all plasmids and genes was confirmed by DNA sequencing of all coding regions for metabolic genes (data not shown). Obtained DNA sequences matched DNA sequences reported in sequence databases.

To validate the functional expression of pathway genes in *E. coli*, we relied on the ability of *ptb-buk* to complement an Δ*atoD* mutation in a Δ*fadR atoC*con background [9] when butyrate is the sole carbon source (the but\(^+\) phenotype). Strain LJ32 (3.1) contains all of these mutations, and as Figure 3-5(a) shows, the LJ32 transformed with pZE21-pab plasmid can grow on butyrate as a sole carbon source. Notably, plasmid pZE21-ptb/buk also confers a but\(^+\) phenotype, but the growth rate of LJ32 pZE21-ptb/buk is slower than observed for LJ32 pZE21-pab, showing that the *atoB* operon on pZE21-pab is also functionally expressed and is assisting in metabolic assimilation of butyrate.

Using plasmids pETDuet1-thl-hbd and pCDFDuet1-tesB, we also validated that *hbd* and the clostridial variant of *atoB* were also functionally expressed in the *E. coli* cytoplasm.
Figure 3-5. Validation of functional activity of clostridial butyrate biosynthesis genes in recombinant E. coli. (a) Plasmids constitutively expressing both the *atoB* and *ptb-buk* operons, but not only *gfp* as a negative control, confer the ability to grow on butyric acid as a sole carbon source in *E. coli* strain LJ32. (b) Expression of pathway clostridial butyrate biosynthesis genes *thi1* and *hbd* from an IPTG-inducible plasmid leads to the formation of 3-3-hydroxybutyric acid from glucose, if a 3HB-CoA hydrolyzing enzyme encoded by *tesB* is co-expressed. Together these results demonstrate functional expression of *ptb-buk*, *atoB* and/or its equivalent *thi1*, and *hbd* in recombinant *E. coli* hosts.
These two plasmids express the indicated genes from T7 promoters, and thus were introduced into a DE3 derivative of \textit{pal} in order to provide the T7 RNA polymerase necessary for transcription from T7 promoters. 3-hydroxybutyric acid accumulated to more than 0.5 g/L (>6 mM) after fermentation of \textit{pal}(DE3) pETDuet1-thi1-hbd pCDFDuet1-tesB, showing that \textit{hbd} and \textit{thi1} were functionally expressed. In further shake flask experiments with CaCO$_3$ emendations for the control of pH, the 3HB titer could be increased to more than 3 g/L (data not shown) and yields relative to total consumed glucose were $\sim$25% mol/mol.

It is worth noting that since the reactions coded by AtoB and Hbd are reversible, the 3HB produced may be subject to reuptake and oxidized to AcCoA, after which it can enter the TCA cycle. Thus data reported here may thus underestimate the capacity of \textit{E. coli} \textit{pal} for producing 3HB.

With all but \textit{bcd}, \textit{etfAB}, and \textit{crt} shown to be functionally expressed in the cytoplasm of \textit{pal} \textit{E. coli} strains, it may seem that one or more of these genes is responsible for the lack of butyrate production. This conclusion is premature. Chapter 4 does detail an method for analysis of this part of the pathway, but observations made in the course of the present chapter suggest that the \textit{ptb-buk} operon is responsible for activities it was not originally believed to possess.

First, experiments with racemic 3-hydroxybutyrate as a sole carbon source in MOPS bicarbonate minimal medium revealed that \textit{E. coli} strains expressing \textit{hbd} and \textit{ptb-buk} operons grew, whereas strains substituting \textit{tesB} for \textit{ptb-buk} or \textit{phaB} for \textit{hbd} did not (Figure 3-6(a)). Unlike the Ptb-Buk system, TesB does not regenerate ATP in the CoA-hydrolysis direction, and thus, is effectively irreversible. Thus, failure of \textit{tesB}-expressing strains to grow on 3HB as sole carbon source should not be surprising; however, given the previously observed activity
Figure 3-6. Unexpected activity of the clostridial ptb-buk operon in recombinant E. coli. (a) Growth of E. coli expressing 3HB pathway genes for the reduction of acetoacetyl-CoA (phaB, hbd) and hydrolysis (tesB, ptb-buk) of 3-hydroxybutyryl-CoA to 3-hydroxybutyric acids. Only strains expressing clostridial butyrate biosynthesis genes hbd and ptb-buk can grow; further stereochemical tests revealed on the S enantiomer of 3HB permitted growth. (b) Production of S-3-hydroxybutyrate by E. coli pal strains expressing clostridial biosynthesis genes hbd and ptb-buk. Production of the S-enantiomer has not been observed using ptb-buk. (c) Crotonic acid is a byproduct of co-expressing both hbd and ptb-buk. The expression of butyrate biosynthesis genes (crotonase, crt, hbd, ptb-buk; or phaB) is shown semi-quantitatively by color. Red indicates omission of a gene, shades of green indicate semiquantitative expression level. Crotonate production does not depend on the presence of crt, correlates with the expression level of hbd, and is much higher when growing on glucose than when growing on butyrate. Strain numbers refer to Table 3.1.
of Ptb and Buk on R-3HB, the failure of the phaB-pto-buk expressing strain to grow on 3HB as a sole carbon source is noteworthy.

Further experiments with E. coli strains expressing hbd and pto-buk revealed that R-3-hydroxybutyrate could not be used as a sole carbon source, but that S-3-hydroxybutyrate could serve as a sole carbon source. Secondly, in the pal(DE3) strain, co-expression of hbd and pto-buk led to the accumulation of 3HB in the culture supernatant. Chiral separation and GC-MS analysis of the culture supernatant showed that the 3HB produced was entirely the S enantiomer. Since the hydroxybutyryl-CoA dehydrogenase encoded by the hbd gene is known to be quite stereospecific, forming only S-3HBCoA from acetoacetyl-CoA, the data in Figure 3-6(a) and (b) shows that the Ptb and Buk enzymes are together capable of forming S-3HBCoA from free S-3-hydroxybutyrate, and that E. coli cells expressing hbd and pto-buk operons were capable of forming the free S-3-hydroxybutyrate from S-3HB-CoA.

Secondly, analysis of the E. coli pal(DE3) expressing hbd and pto-buk revealed that these two operons together were sufficient for crotonate production, as shown in Figure 3-6(c). This panel shows crotonate production for several strains growing under various conditions in minimal medium. Strain 1, expressing all the genes of the clostridial biosynthesis pathway, produced 76 μM crotonic acid during growth on MOPS glucose, but this strain expressed both crotonase (crt) and 3-hydroxybutyryl-CoA dehydrogenase (hbd) at relatively low levels, due to the use of the SC101-derived 1-copy pHATet-CBEH plasmid. In MOPS butyrate medium under oxygen-limiting conditions, Strain 1 did not show signs of growth until more than 110 hours of culture, but eventually cell growth began. Two colonies isolated after growth proved to show no lag phase on inoculation into MOPS butyrate, and gave identical growth behaviors, showing that a genetic change had taken place and resulted in the ability
to grow on butyrate. The change presumably resulted in constitutive expression of E. coli’s native fad operon for the β-oxidation of fatty acids. This evolutionary change resulted in a 10× increase in the production crotonic acid, and was specific to growth on glucose. Surprisingly, a 3× increase was observed between strains 1 and 12. These strains differ in two important ways: (i) hbd is expressed using the higher-copy pETDuet1-thi1-hbd plasmid and the strong T7 promoter, and (ii) crt is not present in strain 12. Thus, the production of crotonate does not depend on the crotonase enzyme. Crotonate was produced at much, much lower levels in strains expressing tesB in place of ptb-buk such as 911, 11, or 13, illustrating that crotonate production depends on expression of both hbd and ptb-buk but not crt. Strains expressing phaB in place of or in addition to hbd also showed nearly no crotonate production. Taken together, the results shown in Figure 3-6 illustrates that ptb-buk is active on the S enantiomer of 3HBCoA, but that the activity of this two-enzyme Ptb-Buk system on S-3HBCoA is non-specific, resulting in the co-production of crotonic acid.

3.4 Discussion and Conclusions

3.4.1 Unexpected phenotypes of the pal strain

Figure 3-2 shows that the pal triple-knockout genotype creates a positive correlation between the anaerobic production of succinate from glucose and biomass. As such it suggests that pal strains will evolve in a direction of increased succinate flux. Adding the clostridial biosynthesis pathway reactions (Figure 3-1 to the iJE660b model used reveals that butyrate formation has become the favored product for biomass formation, suggesting that if endowed
with the capacity to ferment glucose to butyric acid, pal strains will evolve to higher butyrate production fluxes. The pal genotype was predicted not by solving sophisticated and computationally intensive dual-layer mixed-integer linear programs [5, 36], but by simply recognizing that the regeneration of glycolysis-generated NADH and the ability to synthesize ATPs from acyl-CoAs by substrate-level phosphorylation are factors which limit biomass synthesis in E. coli. This suggests that the even simple, intuitive methods are also valid in choosing metabolic genotypes for to apply strain evolution.

In contrast, the actual experimental effort to evolve the pal strain proved unsuccessful, even after a total of 65 generations of growth and \( > 10^{11} \) cell divisions of the pal population. This evidence that the bottleneck in applying combined engineering / evolution approaches to the optimization of recombinant metabolic pathways may be in the experimental evolution step itself, not the computational prediction of necessary engineering modifications. In particular, the iJE660b metabolic model of E. coli does not predict the oxygen auxotrophy of the E. coli pal strain, as depicted in the Figure 3-3. The failure of pal to invoke any of its anaerobic metabolic pathways, including use of alternative electron acceptors for respiration, or fermentative pathways leading to succinate, shows that the auxotrophy is a regulatory bottleneck. An \( O_2 \)-auxotrophic strain of E. coli may be useful in applications requiring continuous culture of E. coli. The \( O_2 \) auxotrophic strain can be grown in continuous culture at arbitrarily high residual glucose concentrations, and the evolutionary pressures which can result from glucose depletion [33, 34, 35] can be avoided. E. coli doesn’t seem to be capable of forming biofilms under anaerobic conditions [6], but the aerobic conditions required by pal may make avoiding long-term biofilm formation difficult.

Selection by serial transfer can favor the selection of mutants with shorter lag phases or
better survival in stationary phase. Additionally, serial transfer requires human intervention at regular and often inconvenient time periods over very long periods. Continuous culture at first glance seems to suffer from neither of these setbacks. However, as others have noted, selections in chemostats of other continuous cultures can select for a variety of undesirable mutations, included increased propensity for biofilm formation, or increased specific glucose uptake rates, or decreased Monod constants for glucose.

Succinate production approaching 1.71 mol per mol of glucose has been observed in *E. coli* K12 ΔldhA Δpta ΔadhE mutants which were additionally modified by (i) deletion of iclR, (ii) expression of a NADH-insensitive isocitrate synthase, and (iii) recombinant expression of pyruvate carboxylase [39]. Since several single-base pair mutations in *E. coli*’s native citrate synthase have been reported to diminish NADH sensitivity [40], the first two mutations should be evolutionarily accessible to *E. coli* K12 ΔldhA Δpta ΔadhE. The failure to obtain improved pal mutants suggests that these mutations do not occur quickly enough for the strain to overtake its parent in <40 generations of continuous culture, or that none of three mutations has a large impact on succinate production by itself.

In the course of performing this research, a report appeared from Ingram and co-workers which created a strain of *E. coli* C deleted for acetate, ethanol, and lactate production by the introduction of a ΔackA ΔldhA ΔadhE genotype [25]. This strain was evolved by serial transfer into a high-yielding succinate producer, after 2,000 generations of growth. Interestingly, the *E. coli* C ΔackA ΔldhA ΔadhE mutant also produced large amounts of malate before it underwent evolution. However, unlike the pal strain, this mutant reportedly grew well anaerobically if provided with a rich medium. *E. coli* C is known to differ from *E. coli* K12 in a number of important characteristics, including the sequence of its primary
sigma factor rpoD [10]. On the other hand, acetyl phosphate, the metabolic intermediate between AcCoA and free acetic acid, is a known global regulatory signal in E. coli K12 [43], and though the degradation of acetyl phosphate to acetate is blocked in ackA deletions, it is the synthesis of acetyl phosphate from AcCoA which is blocked pta. The source of the differences between Ingram’s study [25] and the present one is unclear.

3.4.2 The butyrate pathway: assessing in vivo functionality

Several other workers [27, 17, ?] have shown that the ptb-buk operon could be used to produce R-3HB at high titers and productivities. The finding that these clostridial genes were active in the E. coli cytoplasm (Figure 3-5) was not surprising. But these same groups have reported that ptb-buk seemed incapable of producing the S-3HB enantiomer, presumably because of a lack of reactivity by PtB or Buk on S-3HB-CoA or S-3HB-phosphate, respectively. These results stand in contrast to the data in Figure 3-6, which shows that the PtB-Buk two-enzyme system is capable of converting S-3HB to S-3HB, and that S-3HB is formed in cells expressing ptb-buk.

What happens when PtB and Buk encounter S-3HBCoA? Traditional in vitro biochemical studies with purified enzymes have shown that Buk is likely slow, rate-limiting enzyme enzyme. As shown in Table 3.2, it has been reported to react with 3-hydroxybutyryl phosphate at only 3% of the rate of its reaction with butyryl phosphate, the native substrate. Unfortunately, the stereochemistry of the 3HBCoA used in this in vitro study was not reported. However, the 3HB-CoA isostere isovaleryl-CoA also reacts more quickly with PtB (relative to the native substrate) than does Buk. The very high titers of R-3HB obtained in E. coli strains expressing ptb-buk suggest that activity of the combined enzyme system, and
thus of Buk, is much higher with the \( R \) substrates than with their \( S \) enantiomers. I suggest the data given in Table 3.2 for 3-hydroxybutyryl-CoA more closely corresponds to activities for \( R \) substrates than for \( S \) substrates.

The fact that crotonate is produced as a byproduct from \( hbd \)-expressing strains but not \( phaB \) expressing strains (Figure 3-6) indicates that crotonate is likely a non-specific product of clostridial butanol biosynthesis pathway in \textit{E. coli} cells expressing \textit{ptb} and \textit{buk}. The fact that growth on butyrate, in the “upwards” pathway direction (Figure 3-6), results in vastly lowered crotonate production relative to growth on glucose, is consistent with the idea that Buk is the limiting enzyme. In support of this idea is the activity of Ptb on 3HBCoA and on its isostere valeryl-CoA, as compared to the activity of Buk on 3HB and its isostere, valeric acid, as shown in Table 3.2.

Thus, the data in Figure 3-6 and Figure 3-5 suggest that \( S \)-3-hydroxybutyryl phosphate or a similar intermediate accumulates to very high levels in \textit{ptb-buk} expressing strains. It is converted to free \( S \)-3-hydroxybutyric acid only slowly, be it by Buk or some other enzyme present in the \textit{E. coli} cytoplasm.

Regardless of the precise mechanism of Ptb-Buk activity on \( S \)-3-hydroxybutyryl-CoA, the data in Figure 3-6 and Figure 3-5 also suggest that current understanding of butyrate biosynthesis as it occurs in \textit{Clostridium acetobutylicum} is incomplete. Figure 3-1 shows the currently prevailing model. \( S \)-3-hydroxybutyryl-CoA is clearly an intermediate in clostridial butyrate biosynthesis. At the same time, \textit{ptb} and \textit{buk} are believed to be responsible for conversion of butyryl-CoA to butyrate. \textit{Clostridium acetobutylicum} has not been reported to produce any measurable crotonic acid or \( 3 \)-hydroxybutyric acid during butyrate fermentation. The fact that this work shows Ptb and Buk \textit{are} active on \( S \)-3-hydroxybutyrate is
Table 3.2. Substrate range of phosphotransbutyrylase (ptb) and butyrate kinase (buk) from Clostridium acetobutylicum ATCC824. The ptb enzyme is much more highly active on the 3-hydroxybutyryl-CoA isostere isovaleryl-CoA than the buk enzyme is on the 3HB isostere isovalerate; also, Ptb and has been reported to be more active on 3-hydroxybutyryl-CoA of uncertain stereochemistry than Buk on free 3HB (also of uncertain stereochemistry).

<table>
<thead>
<tr>
<th>Enzyme</th>
<th>Substrate</th>
<th>Activity</th>
<th>Notes</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>phosphotransbutyrlase</td>
<td>3-hydroxybutyryl-CoA</td>
<td>25</td>
<td>substrate was prepared by enzymatic synthesis with ptb-buk from 3HB acid of unspecified stereochemistry</td>
<td>[28]</td>
</tr>
<tr>
<td></td>
<td>n-propionyl-CoA</td>
<td>23%</td>
<td></td>
<td>[42]</td>
</tr>
<tr>
<td></td>
<td>4-hydroxybutyryl-CoA</td>
<td>3%</td>
<td></td>
<td>[28]</td>
</tr>
<tr>
<td></td>
<td>acetyl-CoA</td>
<td>1.6%</td>
<td></td>
<td>[42]</td>
</tr>
<tr>
<td></td>
<td>3-hydroxyvaleryl-CoA</td>
<td>0.3%</td>
<td></td>
<td>[28]</td>
</tr>
<tr>
<td>butyrate kinase</td>
<td>n-butyrate</td>
<td>100% = 237 U/mg</td>
<td></td>
<td>[28, 21]</td>
</tr>
<tr>
<td></td>
<td>n-valerate</td>
<td>89%</td>
<td></td>
<td>[21]</td>
</tr>
<tr>
<td></td>
<td>isobutyrate</td>
<td>54%</td>
<td></td>
<td>[21]</td>
</tr>
<tr>
<td></td>
<td>propionate</td>
<td>43%</td>
<td></td>
<td>[21]</td>
</tr>
<tr>
<td></td>
<td>isovalerate</td>
<td>32%</td>
<td></td>
<td>[21]</td>
</tr>
<tr>
<td></td>
<td>acetate</td>
<td>23%</td>
<td></td>
<td>[21]</td>
</tr>
<tr>
<td></td>
<td>4-hydroxyvalerate</td>
<td>10%</td>
<td></td>
<td>[28]</td>
</tr>
<tr>
<td></td>
<td>4-hydroxybutyrate</td>
<td>7%</td>
<td></td>
<td>[28]</td>
</tr>
<tr>
<td></td>
<td>vinyl acetate</td>
<td>6%</td>
<td></td>
<td>[21]</td>
</tr>
<tr>
<td></td>
<td>crotonate</td>
<td>4%</td>
<td></td>
<td>[21]</td>
</tr>
<tr>
<td></td>
<td>3-hydroxybutyrate</td>
<td>3%</td>
<td>substrate was of unspecified stereochemistry</td>
<td>[28]</td>
</tr>
<tr>
<td></td>
<td>hexanoate</td>
<td>0.5%</td>
<td></td>
<td>[21]</td>
</tr>
</tbody>
</table>
incompatible with the current model of clostridial butyrate biosynthesis (Figure 3-1). The S-3-hydroxybutyryl-CoA pool produced by Hbd does apparently not encounter PtB or Buk in the clostridial cytoplasm. Interestingly, several pieces of historical evidence support this conclusion:

1. Crude extracts of \textit{buk} disruption mutants of \textit{Clostridium acetobutylicum} produced showed significant butyrate kinase activity (around 20\% of around the level of the wild-type) [18].

2. \textit{buk} disruption mutants of \textit{Clostridium acetobutylicum} still produced butyrate in fermentation, albeit at a \(~3\times\) lower level [18].

3. In fermentations of \textit{buk} disruption mutants of \textit{Clostridium acetobutylicum}, butyrate concentrations never decreased in later stages of the fermentation, in contrast to fermentations in wild-type clostridia [18].

4. Enzyme activities for phosphotransbutyrylase and butyrate kinase measured in crude extracts of wild-type \textit{Clostridium acetobutylicum} were anti-correlated. PtB activities were highest at the beginning stages of fermentation and decreased to undetectable levels in the later stage fermentations, when butanol productivity was highest. In contrast, butyrate kinase activities showed the opposite pattern – they sharply increased in the later stage of the fermentation [22].

These evidences all point to a role for Buk, but not PtB, in the uptake of butyrate in the second phase of clostridial two-phase fermentations. One hypothesis is that directionality of flux through Buk, but not PtB, in butyrate fermentation by clostridium has been mis-
assigned. Indeed, Hartmanis and Gatenbeck hypothesized in 1984 that “a direct reduction of butyryl phosphate to butyraldehyde and further to butanol may be possible” [22].

3.4.3 Conclusions

In this chapter, we demonstrated that the “well-understood” components whose existence is an assumed *sine qua non* of synthetic biology were anything but! First, the deletion of three “well-characterized” metabolic genes from the *E. coli* K12 genome resulted in an unexpected oxygen auxotrophy. Second, the Ptb-Buk enzyme system from *Clostridium acetobutylicum* was shown to be active on S-3-hydroxybutyryl-CoA. This observation throws into doubt current understanding of the clostridial biosynthesis pathway. However, the synthetic biology approach revealed bottlenecks in the clostridial pathway and simultaneously illustrated areas of incomplete understanding for the native clostridial pathway. Additionally, the approach here was successful at the creation of a strain for the production of 3HB from glucose at macroscopic yields. We also demonstrated that this 3HB may be specifically converted to propylene gas by post-fermentation exposure of aqueous 3HB solutions to hydrothermal conditions A. This scheme forms a practicable route from renewable biomass to propylene gas.


CHAPTER 4

Measurement of acyl-CoA pool sizes in \textit{E. coli} strains expressing clostridial biosynthesis genes

This Chapter:

- Introduces thermodynamic approaches to characterizing pathway activities.
- Describes an analytical method for the measurement of intracellular CoA concentrations.
- Combines the thermodynamic approach, the analytical CoA method, and literature data to show that the $bcd$ reaction step of the clostridial biosynthesis pathway is a limiting reaction.

4.1 Introduction

Maximal yields of glucose conversion to butyric acid or butanol observed to date for \textit{E. coli} strains expressing clostridial butanol biosynthesis genes has been under 15\% of the theoretical\cite{20, 27, ?}, far short of what has been observed in native clostridial hosts. To identify deficiencies in the recombinant pathway, we developed an analytical procedure for the determination of intracellular pools of short-chain acyl-coenzyme A and applied the technique to the analysis of two sets of \textit{E. coli} strains expressing butyrate biosynthesis genes. In the first set, clostridial butyrate biosynthesis genes were expressed on low-copy vectors from native clostridial operons. In the second set, clostridial genes for butyrate and butanol biosynthesis were expressed...
on high-copy vectors from engineered monocistronic operons, in combination with accessory
genes that allowed for the production of butanol or of 3-hydroxybutyric acid. By comparing
the CoA pool sizes we measured to a thermodynamic model of the butyrate biosynthesis
pathway, we were able to identify the reduction of acetoacetyl-CoA by \textit{hbd} as a rate-limiting
step in the production of butyrate for the first set of strains. At the higher expression levels
of \textit{hbd} found in the second set of strains, the \textit{bcd} reaction step was found to be rate limiting.

The relationships between metabolic fluxes, metabolite pool sizes, \textit{in vivo} enzymatic
activity, and gene expression levels of a given metabolic pathway are difficult to decipher.
Nonetheless, a goal of synthetic biology is the \textit{de novo} construction of metabolic pathways
from disparate sets of genes, enzymes, and host organisms. An example pathway is the
recombinant production of butanol in \textit{E. coli}. Several laboratories have investigated this
pathway \cite{4, 20, 27}, which involves the synthesis of butanol starting from two molecules
of acetyl-CoA (see Figure 3-1). Since native clostridial hosts are comparatively difficult to
engineer, an attractive possibility was incorporation of the butanol production pathway into
a easily manipulable recombinant host like \textit{E. coli}. Despite the investigation of a variety
of strain backgrounds and gene variants, Liao and co-workers obtained maximal yields of
\sim 12\% of the theoretical \cite{4}, with final titers of 5 mM. Using an alcohol dehydrogenase with
enhanced specificity to butanol, Yukawa and co-workers were able to increase butanol titers
to 16 mM, but the maximum reported yield was 7.3\% of the theoretical \cite{20}. These groups
offered various hypotheses to explain the difficulties, including the regulation of CoA pools
\cite{4}, insufficient availability of NADH for \textit{3HBCoA} or \textit{ButCoA} reduction \cite{27}, or insufficient
activity of butyryl-CoA dehydrogenase \cite{20}.

Given the high capacity for acetyl-CoA production inherent in \textit{E. coli} metabolism \cite{8}, and
observations for near-theoretical yields of butanol and butyrate production in Clostridia [25], we were puzzled by the poor performance of the butyrate pathway reported in *E. coli* strains. What limits the flux of carbon through the recombinant butanol biosynthesis pathway? Thermodynamics provides an ultimate bound on any enzymatic reaction occurring in a metabolic network: every individual enzymatic reaction must be energetically favorable at physiologic conditions for it to occur *in vivo*. The free energy change of a given biochemical reaction depends on the equilibrium constant of the reaction, and the concentrations of all reacting species. In principle, reaction equilibrium constants can depend on accurate knowledge of a slew of physiological variables, including the pH, the concentration of all relevant coenzymes and reacting metabolites, and the ionic strength [3]. Measurement of these variables *in vivo* is difficult, especially since many relevant metabolites or species are derivatized with phosphate, CoA, or other groups which can complicate GC-MS-based metabolomic analyses [10].

Nonetheless, prior applications of thermodynamics to metabolic network models [12, 13, 22], including of *E. coli* metabolism, have predicted regulatory control points in metabolic networks through identification of reaction steps farthest from thermodynamic equilibrium. However, these studies have relied in general on constraint-based approaches, and assumed values or ranges for many intracellular concentrations of interest. In addition, they were limited in scope to native metabolic networks unperturbed by the addition of recombinant genes.

Thus, the question remains: do pathway thermodynamics provide insight into the function of engineered pathways? Does thermodynamic assessment of pathway equilibrium constitute a useful tool to metabolic engineers interested in strain analysis? This chapter de-
scribes the development of an analytical technique for measuring intracellular concentrations of butanol pathway intermediates, and the application of thermodynamic principles to this data as a means of identifying specific problematic reaction steps in two sets of recombinant *E. coli* strains.

### 4.2 Methods

#### 4.2.1 Strains, plasmids, and culturing conditions

The first set of *E. coli* strains expresses butyrate or butanol biosynthesis genes from *Clostridium acetobutylicum* on low copy vectors. The bulk of the strains and plasmids described in Chapter 3 are used in this chapter without modification. The only new strains introduced in this chapter are shown in Table 4.1.

For the first set of strains, metabolite extracts were prepared from exponentially growing cultures at 30 °C in MOPS minimal medium amended with 10 mM ammonium bicarbonate, 1 mM IPTG, and antibiotics as appropriate.

For the second set of strains, cultures were grown overnight in LB + 10 g/L glucose at 37 °C, inoculated 1:2 into fresh LB glucose, and allowed to grow for two hours to return to exponential growth phase. They were then harvested by centrifugation and resuspended in MOPS minimal medium with 5 g/L glucose to an *A*₆₀₀ > 0.5. After two hours to condition the cells to the minimal medium, the cultures were induced with either 1 mM IPTG (or with water for negative controls). Metabolites were extracted after 4 hr of induction.
4.2.2 Intracellular metabolite measurements

Our method for the determination of intracellular levels of short-chain acyl-CoA esters is derived from the method reported by Shimazu et al, but is modified in several aspects. This approach is illustrated schematically in Figure 4-1. Briefly, a 1-mL aliquot of bacterial culture was suspended above 400 µL of bromododecane, which itself was layered above 100 µL of metabolite extraction fluid. We originally used 1 M trifluoroacetic acid as the extraction fluid, but a recent report [29] on the degradation of nucleotide triphosphates in *E. coli* metabolite extracts suggested that lower concentrations of water in metabolite extraction fluids better inhibited the autodegradation of of certain metabolites. Because our cell harvesting method requires a dense, hexane- and bromododecane-immiscible extraction fluid, we formulated an extraction fluid of acetonitrile, 2,2,2-trifluoroethanol, water, and trifluoroacetic acid in a 45:45:9:1 ratio (ρ ~ 1.08 g/cm³). For acyl-CoAs, we found no significant difference between the two extraction fluids (data not shown).

After the initial centrifugation, culture medium and bromododecane supernatants were removed by aspiration, the sample was twice washed by the addition of 0.5 mL of hexanes and subsequent vortexing and re-centrifugation. 70 µL of the original 100 µL of extract was transferred to a clean microcentrifuge tube in order to separate metabolite extracts from cell debris, and all samples were evaporated to dryness in a vacuum centrifuge (vaccufuge, Eppendorf AG, Hamburg, Germany) at 30 °C. Dried samples were stored indefinitely at -20 °C. This procedure avoids the use of both exotic extractants like Freon and trioctylamine as well as solid-phase extraction columns. Instead of post-column derivatization with bromoacetaldehyde, and fluorescent detection of the resulting 1,N⁶-ethenoadenine adduct, we used
Figure 4-1. Schematic of extraction and derivatization method for the high-sensitivity detection in *E. coli* of acyl-CoA and related intermediates from the clostridial butanol and 3-hydroxybutyrylate biosynthesis pathways. (a) The clostridial pathway for butyrate biosynthesis, highlighting the adenine moieties (orange) present in all acyl-CoAs as well as in the adenosine phosphates ATP and ADP. (b) Extraction of intracellular metabolites by centrifugation of cells through bromododecane into a dense, polar extraction fluid. (c) Adenine moieties are converted to the fluorescent 1,6-ethenoadenine derivative by treatment with chloroacetaldehyde, which forms the basis of analysis by HPLC with fluorescence detection. (d) Shows sample chromatograms for authentic CoA standards analyzed using method V and standard curves for authentic CoA standards analyzed using method ▲.
pre-column derivatization with 150 μL of 1M chloroacetaldehyde in 150 mM sodium citrate buffer (pH=4.0).

By optimizing derivatization temperatures and times in a gradient-capable thermal cycler, we found that 70 °C derivatization for 60 min gave response factors 8-10% higher than previously reported derivatization conditions [23]. Recently, LC-MS methods for acyl-CoA analysis have also been reported [26, 28], but they necessitate an extra step of solid-phase extraction in order to obtain preparations clean enough for LC-MS analysis. Beta-conjugated acyl CoAs (acetoacetyl-CoA and crotonyl-CoA) were either partially degraded or less efficiently derivatized under derivatization conditions, as slopes of standard curves for these compounds were 10x lower than for acetyl CoA, 3-hydroxybutyryl-CoA, and butyryl-CoA.

We used two chromatographic methods: one which we called Method N was a simple modification of the protocol of Shimazu “Method 1” [31], in which buffer A was substituted with a 50% mixture of (100 mM sodium phosphate, 75 mM sodium acetate, pH 4.5) and 60 mM magnesium chloride, and both buffers were amended by addition of 1% v/v unstabilized tetrahydrofuran. The addition of magnesium chloride was necessary to resolve acetyl-CoA and 3-hydroxybutyryl-CoA, which co-elute in the method originally reported by Shimazu and co-workers. As such, data reported by Shimazu [31] for acetyl-CoA pool sizes in *E. coli* and *Mixococcus xanthus* may be overestimates, due to conflation of 3HB-CoA and AcAcCoA pool sizes.

In a second method we called Method 7, buffer A was 5 mM ammonium phosphate / 1%v/v THF adjusted to pH 6.5 with phosphoric acid, and buffer B was 99:1 methanol:THF. The column temperature was 60 °C and flow rate was 1.2 mL/min and the gradient regime was: 0% B for 2 min; then from 0 % B to 6.25%B by 3 min; from 6.25% B to 14.6% B
by 7.5 min; from 14.6% B to 75% B by 13.3 min; a constant flow of 75% B until 15.3 min; to 0% B by 16 min; and a constant flow of 0% B for 2 min. We found this method is preferred for measurement of adenosine phosphates and free CoASH, but was not suitable for determination of AcAcCoA. For all samples analyzed with this method, hexanoyl-CoA was used as an internal standard by diluting a 10 mM stock of hexanoyl-CoA dissolved in PBS 1000× into metabolite extraction fluid, immediately before beginning extractions. The peak sizes of all analytes were normalized to the HexCoA peak size, and the normalized peak size was calibrated against known standards in order to quantify metabolite pool sizes.

The concentrations of acyl-CoAs were calculated from linear ($R^2 > 0.9$) standard curves obtained from stock solutions (acetyl-CoA), or from spiking known amounts of acyl-CoA into wild-type *E. coli* K12 metabolite extracts (all other acyl-CoAs). The uncertainty in measured pool sizes was the root-mean-square of the standard error between duplicate or triplicate measurements and the standard error associated with the linear regression. This latter error term was dominant for low pool-size measurements (< 100 pmol/(A$_{600}$ mL)).

### 4.2.3 Equilibrium calculations

**Calculation of equilibrium states**

In principle, the equilibrium composition of a single chemical reaction of *n* species can be given by specifying *n* – 1 initial or final concentrations and from the familiar relationship for the equilibrium constant, $\frac{\Delta G^\circ}{RT} = \ln K = \sum_{i=1}^{n} \nu_i \ln C_i$, where $C_i$ represents the concentration of species *i*, and $\nu_i$ represents the stoichiometric coefficient of species *i* in the chemical reaction of interest. $\Delta G^\circ$ is the free energy change of the reaction at a standard state. In
chemical thermodynamics, the standard state is usually interpreted as a hypothetical ideal solution (with unit activity coefficients and concentrations for all species) at a temperature of 298.15 K and atmospheric pressure.

For biochemical reactions it is convenient to treat certain intensive system variables, such as the pH or the concentration of a coenzyme, as a fixed, independent variable of the system. However, the standard chemical formalism is to treat species extensive variables, including the total number of hydrogen ions, as fixed and independent. Conversion from \(\{T, P, n_i\}\) coordinates to \(\{T, P, pH, n_{i\neq H^+}\}\) coordinates requires the Legendre transformation of the equilibrium expression. Transformed Gibbs free energies are related to the untransformed values according to (4.1),

\[
\Delta G' = \Delta G - \mu_H \Delta n_H = \Delta G + (RT \ln(10)pH) \Delta n_H \tag{4.1}
\]

where \(\Delta n_H\) is the change in the number of hydrogen ions for the reaction, and \(\mu_H\), the chemical potential for aqueous protons, is related to the pH as shown. The resulting equilibrium equation is shown in (4.2),

\[
\frac{\Delta G'_{\circ}}{-RT} = \ln K' = \sum_{i=1}^{n} \nu_i \ln C_i \tag{4.2}
\]

where the prime mark \(\prime\) is added to refer to transformed Gibbs free energies and equilibrium constant. As discussed thoroughly by Alberty and others [3], the identity of the “species” participating in each reaction and thus the meaning of \(C_i\) is changed by application of the Legendre transform. In the transformed case, each species is now a sum of all chemical species differing only in their number of hydrogen ions. For example, instead of
counting for unionized butyric acid and the corresponding butyrate ion as separate species,
when using transformed equations, the relevant “butyrate” species would be the sum of these
two forms.

This fact is essential for the use of experimental equilibrium data, in which e.g. the total
butyrate concentration, but not the individual amounts of butyric acid and butyrate ion, is
accessible to experimental measurement and manipulation.

Even then, the experimentally observed equilibrium constants $K'$ provided by most
databases (e.g. Goldberg) are valid only at given temperatures and pHs. To correct $K'$
values observed at one pH and temperature for use at another value of pH and temperature,
the corrections described in Appendix C can be used. For a system of $m$ chemical reactions
and $n$ total species, the equilibrium criteria can be written in matrix form, as shown in (4.3),

$$\ln K' = \mathbf{Y} \cdot \ln C$$

(4.3)

where $\ln K'$ represents an $m$-vector of the logarithm of observed equilibrium constants,
$\ln C$ an $n$-vector of the logarithm of observed equilibrium concentrations, and $\mathbf{Y}$
is an $m \times n$ stoichiometric matrix. In many cases of biochemical interest, $m < n$, and thus, the system
is undetermined. To calculate equilibrium concentrations, $n - m$ additional concentration
values must be measured or assumed. If these concentrations are “initial” concentrations
of species before the approach to equilibrium, non-linear equation solvers are required for
the calculation of equilibrium compositions, but if the equilibrium concentration of certain
species can be measured directly or assumed, calculation of $\ln C$ is a straightforward exercise
in linear algebra. We used the latter approach to estimate equilibrium concentrations.
Available thermodynamic data for the butyrate biosynthesis pathway

The six reactions for butyrate synthesis from acetyl-CoA are shown in Figure 3-1. These reactions involve 14 species (after Legendre transformation with respect to pH). To calculate equilibrium concentrations for the butyrate biosynthesis system, this requires the knowledge of 6 pH-and temperature-corrected apparent equilibrium constants, and specification of eight out of the 14 equilibrium concentrations of reactant species.

As shown in Table 4.2, measurements of the six relevant reaction equilibrium constants are available, but these are not at the temperature and pH relevant to the cytoplasm of \( E.\ coli \) growing at 30 C. These values can be corrected to relevant temperatures and pHs as discussed above, if reaction enthalpies and acid-base ionization pKa’s for all involved species are known. Alberty has compiled ionization pKas for a variety of biochemical species [3], including all species of interest here, except for butyryl phosphate and butyrate. We assumed that the pKa’s for these species were identical to those for their homologues acetyl phosphate and acetate.

As shown in Table 4.2, we estimated reaction enthalpies by using data from the literature or NIST. Subscripts refer to chemical reactions numbered as shown in Figure 3-1.

To completely determine the system, we sought to specify the concentrations of phosphate and water, and the ratios of NADH:NAD, ATP:ADP, and acetyl-CoA:CoASH from literature data. See Table 4.2 for references and the values reported. These parameters, along with measured values for the absolute concentration of acetyl-CoA, are enough to estimate the equilibrium composition of the butyrate biosynthesis pathway.
Table 4.1. New strains and plasmids introduced in this chaptera.

<table>
<thead>
<tr>
<th>Plasmids</th>
<th>Description</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>pETDuet1-atoB</td>
<td>AmpR plasmid for T7-based IPTG-inducible expression of <em>E. coli</em> thioesterase atoB</td>
<td>[27]</td>
</tr>
<tr>
<td>pETDuet1-atoB-gapA</td>
<td>AmpR plasmid for simultaneous expression of atoB and gapA</td>
<td>[27]</td>
</tr>
<tr>
<td>pCDFDuet1-hbd-crt</td>
<td>SpecR plasmid for expression of clostridial <em>crt</em> and <em>hbd</em> genes</td>
<td>[27]</td>
</tr>
<tr>
<td>pCOLADuet1-bcd-efAB</td>
<td>KanR plasmid for expression of clostridial <em>bcd</em> and <em>efAB</em> genes</td>
<td>[27]</td>
</tr>
<tr>
<td>pACYCDuet1-adhE2</td>
<td>CmR plasmid for expression of broad-spectrum butanol dehydrogenase <em>adhE2</em></td>
<td>[27]</td>
</tr>
<tr>
<td>pACYCDuet1-adhE2</td>
<td>AmpR plasmid for simultaneous expression acyl-CoA dehydrogenase <em>acd</em> and the <em>Ralstonia eutropha bktB</em> gene</td>
<td>[27]</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Strains</th>
<th>Description</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>21 pal(DE3) pETDuet1-atoB pCDFDuet1-hbd-crt pCOLADuet1-bcd-efAB this work</td>
<td></td>
<td></td>
</tr>
<tr>
<td>22 pal(DE3) pETDuet1-atoB-gapA pCDFDuet1-hbd-crt pCOLADuet1-bcd-efAB this work</td>
<td></td>
<td></td>
</tr>
<tr>
<td>23 pal(DE3) pETDuet1-acd-bktB pCDFDuet1-hbd-crt pACYCDuet1-adhE2 this work</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

a - For other strains and plasmids, see Table 3.1.

Table 4.2. Thermodynamic and biological parameters estimated from the literature

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
<th>Units</th>
<th>Tref</th>
<th>pHref</th>
<th>Notes</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>( K_1 )</td>
<td>1.56E-05</td>
<td></td>
<td>298.15</td>
<td>7</td>
<td></td>
<td>[14]</td>
</tr>
<tr>
<td>( K_2 )</td>
<td>1.48E-05</td>
<td></td>
<td>298.15</td>
<td>6</td>
<td></td>
<td>[14]</td>
</tr>
<tr>
<td>( K_3 )</td>
<td>0.18</td>
<td></td>
<td>298.15</td>
<td>7.5</td>
<td></td>
<td>[15]</td>
</tr>
<tr>
<td>( K_4 )</td>
<td>7.36</td>
<td></td>
<td>311.15</td>
<td>7.09</td>
<td></td>
<td>[16]</td>
</tr>
<tr>
<td>( K_5 )</td>
<td>20.75</td>
<td></td>
<td>311.15</td>
<td>7.1</td>
<td></td>
<td>[16]</td>
</tr>
<tr>
<td>( K_6 )</td>
<td>3023.542</td>
<td></td>
<td>298.15</td>
<td>7</td>
<td></td>
<td>[17]</td>
</tr>
<tr>
<td>( K_7 )</td>
<td>7.50E-02</td>
<td></td>
<td>303.15</td>
<td>7.6</td>
<td></td>
<td>[14]</td>
</tr>
<tr>
<td>( K_8 )</td>
<td>0.00787</td>
<td></td>
<td>298.15</td>
<td>7.4</td>
<td></td>
<td>[14]</td>
</tr>
<tr>
<td>( K_9 )</td>
<td>0.00855</td>
<td></td>
<td>296.15</td>
<td>7.4</td>
<td></td>
<td>[14]</td>
</tr>
<tr>
<td>NADH:NAD+</td>
<td>0.62</td>
<td></td>
<td></td>
<td></td>
<td>anaerobic chemostat, higher overexpression of <em>pncB</em></td>
<td>[5]</td>
</tr>
<tr>
<td>NADH:NAD+</td>
<td>0.91</td>
<td></td>
<td></td>
<td></td>
<td>anaerobic chemostat, mutant <em>pncB</em> overexpression</td>
<td>[5]</td>
</tr>
<tr>
<td>NADH:NAD+</td>
<td>0.71</td>
<td></td>
<td></td>
<td></td>
<td>anaerobic chemostat, <em>pncB</em> overexpression</td>
<td>[5]</td>
</tr>
<tr>
<td>NADH:NAD+</td>
<td>1.06</td>
<td></td>
<td></td>
<td></td>
<td>anaerobic chemostat, control strain</td>
<td>[5]</td>
</tr>
<tr>
<td>NADH:NAD+</td>
<td>0.7</td>
<td></td>
<td></td>
<td></td>
<td>anaerobic</td>
<td>[9]</td>
</tr>
<tr>
<td>NADH:NAD+</td>
<td>0.05</td>
<td></td>
<td></td>
<td></td>
<td>fully aerobic</td>
<td>[9]</td>
</tr>
<tr>
<td>ATP:ADP</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td>value at intermediate O2 tension</td>
<td>[9]</td>
</tr>
<tr>
<td>ATP:ADP</td>
<td>10</td>
<td>maximum observed</td>
<td></td>
<td></td>
<td></td>
<td>[22]</td>
</tr>
<tr>
<td>intracellular phosphate</td>
<td>3150</td>
<td>pmol/(A600nm mL)</td>
<td></td>
<td></td>
<td>most probable value</td>
<td>[7, 35]</td>
</tr>
<tr>
<td>( \Delta H^\circ_{\text{i}} )</td>
<td>126000</td>
<td>J/mol</td>
<td></td>
<td></td>
<td>maximum value</td>
<td>[1]b</td>
</tr>
<tr>
<td>( \Delta H^\circ_{\text{i}} )</td>
<td>41000</td>
<td>J/mol</td>
<td></td>
<td></td>
<td>minimum value</td>
<td>[1]b</td>
</tr>
<tr>
<td>( \Delta H^\circ_{\text{i}} )</td>
<td>-42500</td>
<td>J/mol</td>
<td></td>
<td></td>
<td>most probable value</td>
<td>[6]</td>
</tr>
<tr>
<td>( \Delta H^\circ_{\text{i}} )</td>
<td>-37000</td>
<td>J/mol</td>
<td></td>
<td></td>
<td>minimal value</td>
<td>[34]</td>
</tr>
<tr>
<td>( \Delta H^\circ_{\text{i}} )</td>
<td>-16000</td>
<td>J/mol</td>
<td></td>
<td></td>
<td>maximum value</td>
<td>[24]</td>
</tr>
<tr>
<td>( \Delta H^\circ_{\text{i}} )</td>
<td>-95200</td>
<td>J/mol</td>
<td></td>
<td></td>
<td>maximum value</td>
<td>[30]</td>
</tr>
<tr>
<td>( \Delta H^\circ_{\text{i}} )</td>
<td>12128</td>
<td>J/mol</td>
<td></td>
<td></td>
<td>maximum value</td>
<td>c</td>
</tr>
<tr>
<td>( \Delta H^\circ_{\text{i}} )</td>
<td>10546</td>
<td>J/mol</td>
<td></td>
<td></td>
<td>minimal value</td>
<td>c</td>
</tr>
<tr>
<td>( \Delta H^\circ_{\text{i}} )</td>
<td>-9576</td>
<td>J/mol</td>
<td></td>
<td></td>
<td>maximum value</td>
<td>c</td>
</tr>
<tr>
<td>( \Delta H^\circ_{\text{i}} )</td>
<td>-14647</td>
<td>J/mol</td>
<td></td>
<td></td>
<td>minimum value</td>
<td>c</td>
</tr>
<tr>
<td>intracellular water</td>
<td>60</td>
<td>M</td>
<td></td>
<td></td>
<td>required for consistency with a reference state of unit water activity</td>
<td>[2]</td>
</tr>
<tr>
<td>acetyl-CoA-CoASH ratio</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td>exponential phase</td>
<td>[8]</td>
</tr>
<tr>
<td>acetyl-CoA-CoASH ratio</td>
<td>0.3</td>
<td></td>
<td></td>
<td></td>
<td>stationary phase</td>
<td>[8]</td>
</tr>
<tr>
<td>intracellular pH</td>
<td>6.5</td>
<td></td>
<td></td>
<td></td>
<td>anaerobic</td>
<td>[33]</td>
</tr>
<tr>
<td>intracellular pH</td>
<td>7.6</td>
<td></td>
<td></td>
<td></td>
<td>average intracellular value for aerobic, growing <em>E. coli</em></td>
<td>[36]</td>
</tr>
<tr>
<td>intracellular pH</td>
<td>8.1</td>
<td></td>
<td></td>
<td></td>
<td>peak transient for stationary phase cells hit with glucose bolus</td>
<td>[36]</td>
</tr>
</tbody>
</table>

b Values for this equilibrium constant were calculated from the standard reduction potentials of the butyryl-CoA / crotonyl-CoA couple and the NADH/NAD+ couple by using the Nernst equation.

c Limiting values for this reaction enthalpy were estimated from maximal and minimal Gibbs free energies by assuming that the entropy change of the reaction was zero.
Monte Carlo error propagation

All of the thermodynamic and biological data are subject to considerable uncertainty. For example, many reports for the measurement of the equilibrium constants for enzymatic reactions come from the 1950s and 1960s, when pure preparations of many reactants were not available. In addition, for pathways containing reactants and products of significantly different electrical charge, the solution ionic strength can strongly affect measured equilibria. Extrapolating previously measured NADH:NAD, ATP:ADP, and acetyl-CoA:CoASH ratios for *E. coli* to the strains and culture conditions used in this study is also bound to introduce considerable uncertainty. To represent this uncertainty, each of the reaction enthalpies, measured apparent equilibrium constants, and specified concentrations or ratios substituted by probability distributions, as shown in Table 4.2. Monte-carlo techniques can then be used to calculate distributions on the resulting equilibrium concentrations of acyl-CoAs. In these techniques, random values for the input parameters are drawn from the specified distributions, and resulting output parameter values are calculated. Many repetitions of this process allow for the construction of histograms of the output parameters. Monte Carlo simulation for the estimation of the uncertainties in equilibrium concentrations were done in Mathematica 6.0, with 100,000 points per simulation. The distributions for input parameters are shown in Table 4.3.

Cancelation of error through calculation of $\Delta_{\text{IPTG}} \Delta G''$

The strains used in this study express butyrate or butanol biosynthesis genes from IPTG-inducible promoters. If the induction of gene expression with IPTG changes the pool sizes of pathway metabolites, the free energy change associated with any given pathway step will
Table 4.3. Probability distributions used for the Monte-Carlo estimation of uncertainty in the equilibrium pool sizes of CoA intermediates in butyrate biosynthesis.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Distribution</th>
<th>Meana</th>
<th>√variancea</th>
<th>Max</th>
<th>Min</th>
<th>Mode</th>
<th>Units</th>
</tr>
</thead>
<tbody>
<tr>
<td>ΔH_1^a</td>
<td>uniform</td>
<td>126</td>
<td>41</td>
<td>kJ/mol</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ΔH_2^a</td>
<td>uniform</td>
<td>-16</td>
<td>-37</td>
<td>kJ/mol</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ΔH_3^a</td>
<td>uniform</td>
<td>12.1</td>
<td>10.5</td>
<td>kJ/mol</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ΔH_4^a</td>
<td>uniform</td>
<td>-9.6</td>
<td>-14.5</td>
<td>kJ/mol</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ΔH_5^a</td>
<td>triangular</td>
<td>-38</td>
<td>-47</td>
<td>-42.5</td>
<td>kJ/mol</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ΔH_6^a</td>
<td>triangular</td>
<td>-90</td>
<td>-110</td>
<td>-95.2</td>
<td>kJ/mol</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AH</td>
<td>uniform</td>
<td>12.1</td>
<td>10.5</td>
<td>126</td>
<td>41</td>
<td></td>
<td>kJ/mol</td>
</tr>
<tr>
<td>AHO</td>
<td>uniform</td>
<td>-9.6</td>
<td>-14.5</td>
<td>-16</td>
<td>-37</td>
<td></td>
<td>kJ/mol</td>
</tr>
<tr>
<td>AH2</td>
<td>triangular</td>
<td>-38</td>
<td>-47</td>
<td>-42.5</td>
<td>kJ/mol</td>
<td></td>
<td></td>
</tr>
<tr>
<td>AH4</td>
<td>triangular</td>
<td>-90</td>
<td>-110</td>
<td>-95.2</td>
<td>kJ/mol</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ATP:ADP ratio</td>
<td>triangular</td>
<td>3</td>
<td>10</td>
<td>10</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NADH:NAD ratio</td>
<td>triangular</td>
<td>0.05</td>
<td>1.06</td>
<td>0.7</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>phosphate</td>
<td>triangular</td>
<td>0.05</td>
<td>1.06</td>
<td>0.7</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>K_7^a</td>
<td>lognormal</td>
<td>-1.11E+01</td>
<td>1.91E-01</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>K_8^a</td>
<td>lognormal</td>
<td>2.91E+00</td>
<td>3.19E-01</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>K_9^a</td>
<td>lognormal</td>
<td>-1.81E+00</td>
<td>4.27E-01</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>K_10^a</td>
<td>lognormal</td>
<td>7.92E+00</td>
<td>4.27E-01</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>K_11^a</td>
<td>lognormal</td>
<td>-2.68E+00</td>
<td>4.27E-01</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>K_12^a</td>
<td>lognormal</td>
<td>4.78E+00</td>
<td>2.39E-01</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>pH</td>
<td>triangular</td>
<td>6.5</td>
<td>7.6</td>
<td>8.1</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

^a For lognormally distributed variables, the mean and variance given refer to the underlying normal distribution.

also change. The difference between the induced state of the pathway and uninduced state of the pathway is shown in (4.4),

\[
\Delta_{\text{IPTG}} \Delta G' = (-RT (\ln K' - \ln Q_{\text{IPTG}})) - (-RT (\ln K' - \ln Q_{\text{no IPTG}}))
\]

\[
= RT \left( \ln \frac{Q_{\text{IPTG}}}{Q_{\text{no IPTG}}} \right) \tag{4.4}
\]

where \( Q \) indicates the actual, possibly non-equilibrated reaction quotient \( \prod_{\text{species } i} C_i^{\nu_i} \), where \( C_i \) is the concentration of reactant or product \( i \) and \( \nu_i \) is the stoichiometric coefficient of the reactant or product in the chemical reaction of interest. As shown in (4.4), in the calculation of \( \Delta_{\text{IPTG}} \Delta G' \), the thermodynamic equilibrium constant \( K' \) cancels out, and thus uncertainty in \( \Delta G' \) stemming from uncertainty in \( \ln K' \) cancels out. To the extent that assumed physiological parameters like the NADH:NAD ratio or the energy charge stay constant after transformation, the uncertainty in these values will also cancel out.
4.3 Results

4.3.1 Validation of cell lysis my centrifugation into polar extraction fluid

Tests with gfp-expressing E. coli strains showed that our metabolite extraction procedure resulted in quick and complete cell lysis and protein denaturation. Cell pellets spun into a bromododecane layer without extraction fluid did not lose their gfp-induced fluorescence at an appreciable rate, but spinning cells through a bromododecane layer into our extraction fluid (a mixture of acetonitrile, trifluoroethanol, water, and trifluoroacetic acid) resulted in complete disappearance of the GFP fluorescence, in less than the 30 seconds required for centrifugation and sample examination. This shows that cell disruption happens only on contact between cells and the extraction fluid, and is not caused by centrifugation conditions or exposure to bromododecane.

4.3.2 Physiological and thermodynamic data available from the literature

Equilibrium constants for all six reactions are available in the Thermodynamics of Enzyme-Catalyzed Reactions database curated by Goldberg and his co-workers. The curation of the data is described in several reports [18, 14, 15, 16, 17]. In many cases, the database contains multiple values for the same equilibrium constant, stemming from both one lab's examination of multiple pHs, ionic strengths, and temperature conditions, and from duplication of experimental effort between laboratories. Usually these multiple values were inconsistent, often by
more than twofold. The database curators evaluate the laboratory methods and practices of literature-reported determination of equilibrium constants and assign each database entry a letter grade. The values reported in Table 4.2 are from database entries of the highest grade available, B or better in all cases, and in cases where equilibrium constants were given as a function of ionic strength, we used the value from the lowest ionic strength reported. The uncertainty in the equilibrium constants was estimated as the difference between values for closely matched experimental conditions from the same data set, or, if no duplicate values were available, as +/- 20%.

Multiple studies on *E. coli* have reported measurements of the intracellular pH and the intracellular NADH:NAD+ ratio. These studies have examined a variety of culture conditions and used a variety of strains. The range spanned by the extrema reported in these studies is thus very likely to contain the NADH:NAD+ ratios and intracellular pHs relevant for the *E. coli* strain and culture conditions used here. The ATP:ADP ratio was back-calculated from the energy charge from a single study [21], assuming AMP and ADP levels were similar. Although the accuracy of this value is thus less certain, the ATP:ADP ratio affects only the equilibrium concentrations of butyryl phosphate and butyrate, not of the upstream CoA esters which are the main focus of this work.

4.3.3 Intracellular CoA pools in strain set #1

In the basal butyrate pathway-expressing strain and in all of its variants (#1-#10 in Table 3.1), acetyl-CoA was the sole short-chain acyl-CoA detected from growth-phase cultures (0.72 < $A_{600} < 0.95$) by use of HPLC method R. The levels of acetyl-CoA were similar in all of the strains, ranging from between 100 to 160 pmol/($A_{600}$· mL) in the stationary phase, and
30~50 pmol/(A_{600} \cdot mL) for all of the strains but strains but #5, #9 and #10, control strains not expressing the butyrate pathway. With method R, underivatized co-enzyme A could not be quantified, but by using the level of AcCoA detected in strain #1 (160 pmol/(A_{600} \cdot mL)) and literature data (Table 4.2), we could estimate the value of free co-enzyme A. This estimate and the measured AcCoA level was sufficient to calculated equilibrium levels of CoAs for downstream pathway metabolites. If each pathway reaction was functioning as completely and as efficiently as possible, all pathway intermediates would equilibrate with each other. Comparing observed metabolite levels with calculated equilibrium levels can illustrate which pathway steps have the least efficient function.

Figure 4-2 shows a comparison of measured/assumed pool sizes to equilibrium pool sizes. For strain #1 from Table 3.1, AcAcCoA, 3HBCoA, CrtCoA, and ButCoA could not be detected. The equilibrium model of the pathway suggests that even at equilibrium, AcAcCoA pools will be difficult to detect due to an unfavorable equilibrium constant for its formation from two AcCoA molecules, but that the equilibrium concentration of 3HBCoA should be easily detectable at equilibrium, when it is not. This fact shows that either the AcAcCoA-forming step (thi1/atoB) or the 3HB-forming step (hbd) has insufficient activity in strain #1.

A strain of E. coli with the entire Ralstonia eutropha-derived pathway for polyhydroxybutyrate biosynthesis chromosomally integrated [32] contains an alternate route for the synthesis of 3-hydroxybutyryl-CoA by use of the R. eutropha phaB gene. In a clone of this strain transformed with pHATet-CBEH and pZE21-pab (the same butyrate-producing plasmids as in strain #1), 3HB-CoA was detectable in cell extracts (Figure 4-2). Also, as described in Chapter 3, increasing the expression level of hbd by expressing it from an artifi-
Figure 4-2. Comparison of acyl-CoA pool sizes in *E. coli* strains expressing clostridial butyrate biosynthesis genes. In (a), CoA pool sizes are shown for a basal strain expressing the complete clostridial butyrate biosynthesis pathway from low-copy plasmids. Measured pool sizes are shown in pink, while CoA pool sizes calculated from measurements and literature data are shown in orange, and equilibrium pool sizes are shown in blue. In (b), the same calculation is repeated for pool sizes measured in an *E. coli* strain bearing a chromosomal copy of the PHB biosynthesis pathway. The equilibrium calculations suggest that the hbd reaction step may be rate-limiting in the basal strain, either because of insufficient cofactors or precursors, or because of insufficient enzyme activity. For reference, the light gray range bars indicate approximate limits of detection for each metabolite, and the dark gray range bars indicate an approximate range for the \( K_m \) for the binding of the indicated metabolite to the immediately downstream pathway enzyme.
cial monocistronic operon under the control of a high-strength T7-derived promoter resulted in the accumulation of 3HB in medium supernatants. In contrast, no 3HB production was ever observed in strain #1.

Combined with data from Chapter 3 which indicates the pZE21-pab plasmid was functionally active (Figure 3-5), these facts illustrate that \textit{hbd} was the bottleneck step in the clostridial butyrate pathway.

4.3.4 Uncertainty in the pathway thermodynamic model

No error bars are shown in Figure 4-2. How uncertain are the estimated equilibrium concentrations? Monte Carlo simulation of the uncertainty in system parameters described in Table 4.3 and in Section 4.2.3, revealed that the uncertainty was quite large, and mainly stemmed from uncertain values of the NADH:NAD ratio and the intracellular pH (Figure 4-3). Uncertainty in reaction equilibrium constants or in reaction enthalpies were far less important. Calculated equilibrium concentrations spanned two orders of magnitude when the effect of both thermodynamic and physiological uncertainty was included, and uncertainty was greater for pathway metabolites farther downstream from those with measured (or assumed) concentrations.

4.3.5 A second set of strains for CoA analysis

After confirming that \textit{hbd} was the pathway bottleneck through this analysis, we removed the bottleneck by focusing on a second set of strains engineered to exhibit much higher expression levels of \textit{hbd}. Strains with this higher expression level of \textit{hbd} production could be engineered to produce 3-hydroxybutyric acid, as shown in Chapter 3, showing that the
Figure 4-3. Monte-carlo calculation of uncertainty in calculated equilibrium concentrations of the butyrate synthesis pathway stemming from uncertainty in thermodynamic data (white), uncertainty in physiological data (light gray), or the combined effects of both (dark gray). These calculations correspond to panel (a) of Figure 4-2. Although the uncertainty in the thermodynamic data alone provides an uncertainty of ~1 order of magnitude in equilibrium concentrations, the uncertainty in physiological parameters, primarily in the NADH:NAD\(^+\) ratio and the intracellular pH, is a far greater driver of uncertainty in the estimated equilibrium concentrations.
identification of \textit{hbd} as a bottleneck enzyme was correct.

However, strains with increased \textit{hbd} expression did not produce butyric acid. Empowered by our prediction of \textit{hbd} as the rate limiting enzyme, this fact suggested to us that the elimination of the \textit{hbd} expression-level bottleneck had simply allowed another pathway bottleneck to manifest itself. To test this hypothesis, we measured CoA pool sizes for pathway intermediates under both induced and non-induced conditions. Table 4.4 shows CoA pool sizes in this second set of strains for further insights into the bottleneck.
Table 4.4. Intracellular pool sizes in pmol/(A_{600}•mL of free coenzyme A, acyl CoAs, and adenosine phosphates in E. coli strains expressing clostridial butanol biosynthesis genes. Uncertainties shown are the root-mean-square of (i) the standard error in the mean calculated from either 2 or 3 duplicate measurements and (ii) the standard error in the y estimate stemming from linear regression against HPLC calibration curves.

<table>
<thead>
<tr>
<th>Strain</th>
<th>CoASH</th>
<th>AcCoA</th>
<th>AcCoA:CoASH ratio</th>
<th>3HBCoA</th>
<th>ButCoA</th>
<th>ATP</th>
<th>ADP</th>
<th>AMP</th>
<th>EC = ATP + ADP - AMP</th>
</tr>
</thead>
<tbody>
<tr>
<td>NO IPTG</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>676 ± 50</td>
<td>1170 ± 48</td>
<td>1.7 ± 0.13</td>
<td>18 ± 24</td>
<td>8 ± 74</td>
<td>547 ± 37</td>
<td>214 ± 31</td>
<td>179 ± 58</td>
<td>0.61 ± 0.051</td>
</tr>
<tr>
<td>12</td>
<td>278 ± 31</td>
<td>1059 ± 47</td>
<td>3.8 ± 0.24</td>
<td>27 ± 25</td>
<td>19 ± 74</td>
<td>220 ± 46</td>
<td>150 ± 10</td>
<td>105 ± 29</td>
<td>0.62 ± 0.039</td>
</tr>
<tr>
<td>13</td>
<td>477 ± 36</td>
<td>1116 ± 201</td>
<td>2.3 ± 0.43</td>
<td>96 ± 54</td>
<td>16 ± 74</td>
<td>478 ± 61</td>
<td>184 ± 29</td>
<td>71 ± 15</td>
<td>0.78 ± 0.022</td>
</tr>
<tr>
<td>14</td>
<td>412 ± 82</td>
<td>1216 ± 123</td>
<td>3.0 ± 0.63</td>
<td>14 ± 25</td>
<td>10 ± 74</td>
<td>379 ± 105</td>
<td>195 ± 27</td>
<td>113 ± 29</td>
<td>0.69 ± 0.033</td>
</tr>
<tr>
<td>21</td>
<td>391 ± 222</td>
<td>627 ± 110</td>
<td>1.6 ± 0.94</td>
<td>120 ± 59</td>
<td>16 ± 74</td>
<td>250 ± 68</td>
<td>148 ± 32</td>
<td>78 ± 5</td>
<td>0.68 ± 0.026</td>
</tr>
<tr>
<td>22</td>
<td>635 ± 38</td>
<td>943 ± 197</td>
<td>1.5 ± 0.32</td>
<td>109 ± 26</td>
<td>13 ± 74</td>
<td>583 ± 395</td>
<td>228 ± 63</td>
<td>76 ± 17</td>
<td>0.79 ± 0.032</td>
</tr>
<tr>
<td>23</td>
<td>262 ± 150</td>
<td>503 ± 71</td>
<td>1.9 ± 1.12</td>
<td>25 ± 34</td>
<td>16 ± 74</td>
<td>184 ± 2</td>
<td>114 ± 13</td>
<td>104 ± 37</td>
<td>0.58 ± 0.058</td>
</tr>
<tr>
<td>24</td>
<td>475 ± 149</td>
<td>419 ± 64</td>
<td>0.9 ± 0.44</td>
<td>56 ± 39</td>
<td>11 ± 74</td>
<td>288 ± 102</td>
<td>110 ± 6</td>
<td>128 ± 38</td>
<td>0.65 ± 0.047</td>
</tr>
<tr>
<td>pal</td>
<td>570 ± 201</td>
<td>1031 ± 354</td>
<td>1.8 ± 0.78</td>
<td>62 ± 48</td>
<td>13 ± 74</td>
<td>441 ± 197</td>
<td>181 ± 56</td>
<td>131 ± 61</td>
<td>0.71 ± 0.065</td>
</tr>
</tbody>
</table>

| + IPTG |       |       |                   |        |        |     |     |     |                     |
|--------|-------|-------|-------------------|        |        |     |     |     |                     |
| 11     | 342 ± 36 | 1218 ± 50 | 3.6 ± 0.52      | 6 ± 24 | 20 ± 74 | 5 ± 1 | 4 ± 1 | 31 ± 20 | 0.17 ± 0.028 |
| 12     | 528 ± 38 | 1365 ± 57 | 2.5 ± 0.31      | 210 ± 56 | 0 ± 74 | 450 ± 188 | 159 ± 21 | 144 ± 79 | 0.70 ± 0.008 |
| 13     | 818 ± 49 | 861 ± 155 | 1.1 ± 0.39     | 170 ± 89 | 16 ± 74 | 0 ± 0 | 0 ± 0 | 247 ± 104 | 0.00 ± 0.000 |
| 14     | 1442 ± 271 | 1564 ± 158 | 1.1 ± 0.46     | 763 ± 440 | 25 ± 74 | 610 ± 504 | 289 ± 80 | 331 ± 170 | 0.69 ± 0.066 |
| 21     | 98 ± 553 | 665 ± 116 | 0.7 ± 0.80     | 322 ± 146 | 35 ± 74 | 542 ± 295 | 78 ± 34 | 211 ± 27 | 0.70 ± 0.054 |
| 22     | 364 ± 32 | 476 ± 142 | 1.9 ± 0.79     | 33 ± 24 | 9 ± 74 | 218 ± 229 | 94 ± 52 | 60 ± 27 | 0.71 ± 0.057 |
| 23     | 1244 ± 703 | 788 ± 109 | 0.6 ± 0.74     | 407 ± 407 | 32 ± 74 | 611 ± 15 | 126 ± 29 | 235 ± 167 | 0.69 ± 0.138 |
| 24     | 564 ± 177 | 571 ± 223 | 1.0 ± 1.01     | 45 ± 34 | 17 ± 74 | 342 ± 242 | 125 ± 14 | 200 ± 119 | 0.61 ± 0.088 |
| pal    | 298 ± 108 | 1134 ± 279 | 3.8 ± 3.27   | 87 ± 63 | 13 ± 74 | 337 ± 301 | 177 ± 109 | 187 ± 174 | 0.61 ± 0.113 |
The adenylate energy charge calculated from measurements of ATP, AMP, and ADP pool sizes was generally consistent with previously obtained measurements [31]. The extremely low values for EC obtained in strains #11 and #13 after IPTG induction reflect the activity of the tesB gene instead of the ptb-buk gene system. The TesB enzyme does not regenerate ATP during the hydrolysis of 3HB-CoA to form free 3-hydroxybutyric acid; whereas with Ptb and Buk, ATP is regenerated during the two-step formation of 3-hydroxybutyric acid from 3HB-CoA.

In IPTG-induced strains of 3HB-producing strains, 3HB-CoA was readily detectable in E. coli cell extracts. This metabolite was not reliably detected in non-induced strains of 3HB-producing E. coli or in other E. coli strains engineered to produce butanol.

In E. coli strains engineered to produce < 600 mg/L of butanol, butyryl-CoA was not reliably detected in IPTG-induced cultures. In contrast, slight amounts of butyryl-CoA were detected in non-induced butanol-producing strains, but the uncertainty in the level of butyryl-CoA detected under non-inducing conditions was still quite large.

Butyryl-CoA levels were too low to be reliably detected in these extracts, as illustrated by the uncertainties given in Table 4.4 for ButCoA. This observation suggests that the bcd reaction step or the crt reaction step (Figure 3-1), were the problematic ones.

4.3.6 Thermodynamic interpretation of pathway metabolite measurements

To test this hypothesis, the data in Table 4.4 were used to calculate the effect of induction on the free energy change of reaction for pathway reactions, as shown in Figure 4-4. This figure shows that IPTG increased (made less thermodynamically favorable) the conversion
of AcCoA to 3HB-CoA for two of the 3HB producing strains (strain #12 and #14), as shown by the positive values for $\Delta_{\text{IPTG}}\Delta G'$. The $\Delta_{\text{IPTG}}\Delta G'$ value for the 3HBCoA-from-AcCoA reactions for strain #13 was also likely positive, but a value of 0 cannot be excluded because of the slight overlap of the error bar and the $\Delta_{\text{IPTG}}\Delta G' = 0$ center axis. In strain #11, measurements did not reliably detect intracellular 3HB-CoA, making reliable estimation of $\Delta_{\text{IPTG}}\Delta G'$ impossible.

Calculation of $\Delta_{\text{IPTG}}\Delta G'$ removes much of the thermodynamic uncertainty inherent in calculation of equilibrium concentration (Figure 4-3), but because of the low level of detection and correspondingly high relative uncertainty in the 3HBCoA pool size for non-induced conditions, $Q_{\text{no IPTG}}$ in (4.4) is highly uncertain. The least uncertain $\Delta_{\text{IPTG}}\Delta G'$ in Figure 4-4 still has an uncertainty of 8 kJ / mol. The impact of the uncertainty in $Q_{\text{no IPTG}}$ only is represented in Figure 4-4 with thick gray bars. The total uncertainty in $\Delta_{\text{IPTG}}\Delta G'$, stemming from both uncertainty in $Q_{\text{no IPTG}}$ and $Q_{\text{IPTG}}$ is shown with thin, colored error bars. With few exceptions, the uncertainty in $Q_{\text{no IPTG}}$ accounts for the vast majority of the uncertainty in $\Delta_{\text{IPTG}}\Delta G'$.

The failure to detect appreciable levels of butyryl-CoA in butanol-producing *E. coli* under either inducing or non-inducing conditions made accurate estimation of $\Delta_{\text{IPTG}}\Delta G'$ impossible. However, it is notable that the estimated ranges for $\Delta_{\text{IPTG}}\Delta G'$ shown for the 3HBCoA → ButCoA set of pathway reactions all overlap 0. The wild-type pal strain, with no plasmids or pathway genes, also had estimated $\Delta_{\text{IPTG}}\Delta G'$ values near 0 kJ / mol, as would be expected for a strain with no pathway activity of any type.
Figure 4-4. The change from adding IPTG inducer in the observed $\Delta G'$ for component reactions of the clostridial butyrate or butanol biosynthesis pathway. A positive $\Delta_{IPTG}G'$ indicates that induction of pathway genes has filled pool sizes for reaction products, or depleted pool sizes for reactants, of the given reaction, and as such, represents evidence that pathway induction leads to conversion of thermodynamically available potential into flux through the given reaction. The graph shows $\Delta_{IPTG}G'$ for nine strains numbered according to Tables 3.1. For strains 11-14, only the $\Delta_{IPTG}G'$ for the lumped reaction corresponding to 3HBCoA synthesis from AcCoA is shown, because these strains lack downstream pathway genes. In butyrate producing strains 21-24, data for both 3HBCoA synthesis from AcCoA and ButCoA synthesis from 3HBCoA is shown. For some strains, induction clearly shifts the 3HBCoA synthesis reaction toward equilibrium. In contrast, no such change is shown for the ButCoA synthesis reaction, indicating that this pathway step is likely rate-limiting.

4.4 Discussion and conclusions

4.4.1 Thermodynamic analysis and uncertainty

By measuring intracellular pool sizes of acyl-CoA intermediates of the butyrate biosynthesis pathway, we were able to estimate thermodynamic equilibrium pool sizes. However, error analysis by Monte Carlo simulation showed that equilibrium pool estimates were highly uncertain, in large part due to uncertainty in key physiological parameters, mainly the NADH:NAD ratio and the adenylate energy charge.

Calculating $\Delta_{IPTG}G'$ instead of equilibrium concentrations of intermediate metabolites can mitigate the effect of the uncertainty to some extent. However, the cancelation of
thermodynamic uncertainty that occurs in the calculation of $\Delta_{\text{IPTG}}\Delta G'$, as shown in (4.4), depends on the physiological parameters of the cytoplasm being unaffected by IPTG induction. The data in Table 4.4 for the energy charge shows that this assumption was not true in general (viz. strains #11 and #13), but was adequate for most examined strains. However, techniques to directly measure the NADH:NAD ratios would still be preferred to reliance on literature estimates.

4.4.2 Utility of thermodynamic analysis for understanding the clostridial production of 3HB-CoA

Strains expressing clostridial butyrate biosynthesis genes from low-copy plasmids were shown in this work to have insufficient $hbd$ activity, as indicated by the thermodynamic analysis in Figure 4-2. The insufficient activity was alleviated by expressing $hbd$ from higher-copy plasmids using stronger promoters, as shown in Table 4.4 and by the data on the production of 3HB, as described in Chapter 3. This work shows that thermodynamic methods can be useful in the diagnosis of pathway behavior.

Thermodynamic analysis of CoA pools in 3HB-producing second set of strains revealed that induction drove the 3HB synthesis reactions closer to thermodynamic equilibrium in two (and possibly three) cases, as shown by the $\Delta_{\text{IPTG}}\Delta G'$ values calculated for strains #12, 14 (and 13) in Figure 4-4. The fact that strain #11 had a much smaller 3HB-CoA pool after induction than the other strain suggests that in this strain, downstream pathway steps, such as the conversion of 3HB-CoA to free 3-hydroxybutyric acid, may be limiting.

Moving other clostridial butyrate biosynthesis genes to high-copy vectors did not improve the production of butanol, butyrate, or of the key butyryl-CoA intermediate (Table 4.4). This
fact shows that gene expression level does not explain the deficient pathway behavior.

In the course of this work, several other laboratories have measured the enzyme activities of the bcd-encoded butyryl-CoA dehydrogenase and other clostridial pathway enzymes in crude extracts prepared from recombinant E. coli [20, 27]. Measured activities for the BCD enzyme were far lower than for other pathway genes, suggesting that this enzyme was a bottleneck. These observations are consistent with the observations of in-vivo pathway thermodynamics measured here.

One other study has compared observed butanol production in recombinant E. coli expressing bcd to a similar strain with a Streptomyces colinus crotonyl-CoA reductase gene in the place of bcd [4]. Surprisingly, replacement of bcd by ccr gene halved the observed butanol production from ~500 mg/L to ~250 mg/L. One possibility for this observation is that the activity of the crotonyl-CoA enzyme is not as originally hypothesized: the original report of the cloning of the ccr gene and its expressing in E. coli, the activity was measured by spectrophotometric consumption of NADPH in the presence of the substrate crotonyl-CoA, not by direct detection of the hypothesized reaction product, butyryl-CoA. A more recent study [11] has found that the ccr gene from a closely related Streptomyces species catalyzes the reductive carboxylation of crotonyl-CoA to ethylmalonyl-CoA, with NAPDH as the electron donor:

\[
\text{CrtCoA} + \text{NADPH} + \text{H}^+ + \text{CO}_2 \rightarrow \text{ethylmalonyl-CoA} + \text{NADP}^+.
\]

Additionally, a recent investigation of BCD activity in Clostridium kluyveri [19, 24] has suggested that the reaction attributed to this enzyme is also incorrect. The C. kluyveri BCD enzyme was shown to reduce two equivalents of NADH, one of which was used in the exergonic step of reducing CrtCoA to ButCoA. The free energy change of this step was used
to drive the endergonic step of reducing ferredoxin with a second NADH. It is unknown if the *C. acetobutylicum* BCD enzyme used in the present study behaves similarly, but if it does, it is notable that none of the butanol or butyrate production pathways in this study have included the clostridial ferredoxin gene as part of the pathway.

### 4.4.3 Conclusions

The measurement of the intracellular concentration of metabolic intermediates allows estimation of the thermodynamic driving force through each step of the pathway. This information can be used to pinpoint underperforming or inadequate pathway components, as shown here for the *hbd* gene as expressed from low-copy plasmids, and for the *bcd* gene, regardless of its expression level.
Bibliography


CHAPTER 5

Tyrosine production in *E. coli*

This Chapter:

- Motivates the need for tyrosine overproduction in *E. coli*.
- Describes stoichiometric and optimization-based models.
- Details the application of these models to the prediction of gene targets for the metabolic engineering of tyrosine overproduction in *E. coli*.
- Examines the conflict between model predictions and experimental results, and the role of nitrogen in tyrosine overproduction.
- Gives a method for the intracellular analysis of amino acids in *E. coli*.

5.1 Introduction

Petroleum-derived aromatics such as styrene are key chemical intermediates in the manufacture of a wide range of products. Polystyrene, as an example, is a thermoplastic material whose worldwide production is greater than 3 megatons per year [30], and is used in a applications ranging from building insulation to hard plastic packaging. Styrene is produced commercially from benzene and ethylene [6], both of which derive ultimately from crude oil. Other aromatics are widely used in the pharmaceutical and fragrance industries. Recent oil price shocks and long term concern over petroleum availability have spurred interest in the
production of the production of aromatic chemical intermediates from renewable resources. Although styrene and polystyrene were originally discovered by isolation and distillation from plant resins [33], the yields obtained in plant extracts are insufficient for commercial production.

The most common biologically derived aromatic compounds are the familiar proteinogenic amino acids tyrosine and phenylalanine. Overproduction of these compounds via microbial fermentation would thus be an attractive route for the production of aromatic chemical intermediates from renewable feedstocks such as sugars or other biomass sources. Phenylalanine and tyrosine can be converted to related aromatic compounds such as coumaric acid and cinnamic acid, which are of interest as aromatic feedstocks [32].

As described in Chapter 1, reliable, general methods for the identification of the key genetic modifications required to elicit enhanced metabolite overproduction are unknown. In this chapter, we compare two methods of gene target identification for the enhancement of tyrosine overproduction in *E. coli*. One uses optimization-based stoichiometric models, and the other relies on thermodynamic information about the biochemical pathway and metabolomic information about the levels of pathway intermediates.

### 5.2 Methods

#### 5.2.1 Construction of *E. coli* strains

*E. coli* K12 MG1655 is the parent all all other strains used in this work. Tyrosine-producing strains T1 (K12 Δ*tyrR* pCL1920::P_{tet-tyrAfbr-aroGfbr}) and T2 (K12 Δ*tyrR* pCL1920::P_{tet-tyrAfbr-aroGfbr-tktA-pps}) were constructed by Tina Lütke-Eversloh [25]. These strains ex-
press feedback-resistant mutants of the AroG 2-dehydro-3-deoxyphosphoheptonate aldolase [18] and the TyrA chorismate mutase / prephenate dehydrogenase [24]. The ΔtyrR derivative was prepared by phage transduction [27] from strains constructed by Lütke-Eversloh [25].

The aceE knockout could not be successfully obtained via usual methods for gene deletion [8], and was constructed using the Targetron Gene Knockout System (Sigma-Aldrich, St. Louis, MO, USA) as per the manufacturer’s directions. Briefly, a group II intron system originally programmed to insert itself at precise points in a Lactococcus genome was appended to a Kan^r selection marker and cloned into a p15a-derived plasmid called pACD4K-C (obtained from Sigma-Aldrich). Using primers “CF55 aceE 153:154a-EBS2”, “CF56 aceE 153:154a-IBS”, and “CF57 aceE 153:154a-EBS1d”, with sequences TGAACGCAAGTTTCTAATTCTGGTTGCACATCGATAGAGGAAAGTGTCT, AAAAAGCTTATAATTATCCTTATGTGCCTGCGGCGTGCGCCCAGATAGGGTG, and CAGATTGTACAAATGTGGTGATAACAGATAAGTCTGCGGCTATAACTTACCTTTCTTTGT, respectively, the targeting sequence of the group II intron was changed to match the E. coli K12 aceE sequence at nucleotides 153 and 154. The resulting mutant plasmid was transformed into LB cultures of E. coli K12, E. coli T1 and E. coli T2, and induced with IPTG. Colonies isolated from these cultures were tested for kanamycin resistance and chloramphenicol sensitivity, and the ΔaceE genotype was verified by colony PCR with primers “CF52 aceE::kanR verif Sense”, “CF53 aceE::kanR verif Anti”, “CF54 kt IMPROVED”, and “CF55 k2 IMPROVED”, with sequences of GAATTGCTCTATTCGCGTCGCGAGA, TCTTCTCTTCTGCCTGAGCAGGTGCA, AGCCGGCCACACGTGGA-TGAATCCAGAAA, and CCTGTCCGCTGCCCTGAATGAATCGCA, respectively.
5.2.2 Computational simulations of *E. coli* metabolism

Stoichiometric models for wild-type *E. coli*

Most intracellular metabolites are consumed as rapidly as they are formed; they do not build up as long as the cell is undergoing "balanced growth" [10]. Under the assumption of balanced growth, the steady-state conservation equation for any metabolite *i* is given by (5.1), where *S*<sub>ij</sub> is the stoichiometric coefficient for metabolite *i* in reaction *j*, and the sum is taken over all of the *n* metabolic reactions in the system.

\[
\sum_{j=1}^{n} S_{ij} v_j = 0
\]

(5.1)

How to account for the net import or export of nutrients, metabolic wastes, or other by-products from the cell? These metabolites are not at steady state; nutrients such as glucose or ammonium and metabolic by-products such as CO₂ or lactate are being consumed or accumulated. One way is to add unbalanced "generation" or "consumption" fluxes to the model. For example, for glucose, a "generation" flux $\emptyset \rightarrow GLC_{xt}$, where GLC<sub>xt</sub> symbolizes the extracellular glucose pool, and the symbol $\emptyset$ symbolizes "nothing", so that the reaction allows for the net import of glucose from some source external to the model. In this way, the metabolite mass balance constraints can be written as in (5.2).

\[
S \cdot \mathbf{v} = 0
\]

(5.2)

The stoichiometric model of *E. coli* metabolism considered here is derived from the iJE660 model constructed by Jeremy Edwards and his co-workers [9]. The 660 refers to the
number of non-transport metabolic reactions; the full stoichiometric matrix considers 536 metabolites and 953 fluxes. The model as originally formulated by Edwards et al. included redundant columns (fluxes) in an attempt to model stoichiometrically equivalent reactions. For example, the original model included 5 duplicate columns for the NADH-driven reduction of l-lactate to lactaldehyde (column indices were 109, 110, 111, 113, and 114). Each column in theory represented a distinct enzyme, in recognition of the fact that aldA, aldB, adhE, adhC, and aldH were believed to encode enzymes capable of catalyzing the given reaction to at least some extent.

In some cases, such as when applying the model to predict lethal gene knockouts [11, 35], such complexity may be useful. When searching for targets genes to modify in order to enhance the production of a desired metabolite, however, this extra complexity is not useful. Model-predicted “genes”are best interpreted not as a specific gene, but instead as a specific flux which needs to be down-regulated. Existing literature, database, or bioinformatic information can point to the gene(s) best-linked to a specific metabolic flux.

In light of these considerations, the IJE660 model was simplified by removing 117 duplicate fluxes, resulting in a stoichiometric matrix with dimensions 536 × 836. This simplified model was called ”iJE660b”. Even after model reduction, the number of columns (fluxes) is more than the number of rows (metabolite balances), and therefore the metabolic system is underdetermined, meaning that specific values for the \( v_j \) cannot be found from metabolite mass balances alone. To use stoichiometric models to predict flux states, two additional criteria must be applied.

One such criterion is a set of lower and upper bounds for each of the constituent fluxes. The flux \( v_j \) of a reversible reaction \( j \) is unbounded; that is, \(-\infty \leq v_j \leq \infty\). For an irreversible
reaction \( j \), however, the lower bound of the flux is zero: \( 0 \leq \nu_j \leq \infty \). The uptake rates of specific nutrients may be further constrained. For example, in the model of Edwards et al., the glucose uptake rate is constrained to be less than 10 mmol gDCW\(^{-1}\) h\(^{-1}\) [9, 14], because this is the maximal rate observed in the many literature reports on glucose uptake in \( E.\ coli \).

Thus, in addition to the equality constraints on the metabolic flux state \( \nu \) given in (5.2), the model also imposes upper and lower bounds, as shown in (5.3),

\[
lb_j \leq \nu_j \leq ub_j \quad \forall j \in \{1, 2, 3, \ldots, n\}
\]  

(5.3)

where \( lb_i \) is the lowest possible value, or lower bound, for the \( i \)th flux, and \( ub_i \) is the upper bound for the same flux.

The second criterion is an objective function. The pioneering work of Ibarra et al.[14] showed that the \( E.\ coli \) metabolic network evolves toward optimizing biomass formation as constrained by (5.2) and (5.3). The biomass formation flux \( \nu_{\text{biomass}} \) thus becomes the objective function in a linear optimization problem, as shown in (5.4).

\[
\begin{align*}
\text{max} & \quad \nu_{\text{biomass}} \\
\text{subject to:} & \\
S \cdot \nu & = 0 \\
lb_j \leq \nu_j \leq ub_j & \quad \forall j \in \{1, 2, 3, \ldots, n\}
\end{align*}
\]

(5.4)

Linear optimization of the biomass formation flux is referred to in the literature as “flux
balance analysis” (FBA). A well-defined stoichiometry for biomass is obviously required in order to define the “biomass flux” \( \nu_{\text{biomass}} \). This biomass formation stoichiometry is a key component of optimization-based stoichiometric models. Edwards et al. [9] cite only the two-volume treatise on *Escherichia coli* by Neidhardt et al.[28] as the source for the biomass stoichiometry used in their model. Exactly how they extracted a single stoichiometric equation from this work is unclear. The model handles the biomass formation flux by adding both a balanced reaction for biomass formation reaction, and a biomass export flux \( \text{BMxt} \rightarrow \emptyset \) to the model.

### 5.2.3 Objective functions for mutant strains

Ibarra et al. showed that in order to achieve optimal biomass formation, a strain culture must undergo hundreds or thousands of generations of adaptive evolution [14]. The FBA criterion of maximal biomass formation cannot be used to accurately predict the metabolic flux state \( \nu \) in strains which have not been adaptively evolved.

Targeted genomic modifications are just that – targeted. In general, strains engineered by targeted gene deletion or targeted insertion or overexpression of other genes have no chance to undergo multigenerational adaptive evolution under relevant environmental conditions before they are tested or studied. At least two research groups have made efforts to modify optimization-based models of *E. coli* metabolism in order to predict flux states in engineered derivative strains. The ROOM criterion [35] and the MoMA criterion [34] have both been proposed as computational methods for the prediction of metabolic flux states in genetically modified strains of *E. coli*.

The MoMA criterion specifies that the flux state in the mutant cell is as close as possible
to the flux state in its parental wild-type, where closeness is judged by the Euclidean distance [34]. That is, first a flux state for the wild-type strain growing under specified conditions is found using FBA, as described in (5.2.2). Call this flux state $\nu_{wt}$. MoMA predicts the flux state of a mutant strain by solving the quadratic optimization problem in (5.5).

$$\min ||\nu_{mut} - \nu_{wt}||^2$$

subject to:

$$S \cdot \nu_{mut} = 0$$

$$lb_{mut,j} \leq \nu_{mut,j} \leq ub_{mut,j} \quad \forall \ j \in \{1, 2, 3, \ldots, n\}$$

The Euclidean distance in (5.5) has its usual meaning: $||\nu_{mut} - \nu_{wt}||^2 = \sum_{j=1}^{n} (\nu_{mut,j} - \nu_{wt,j})^2$.

The changes engineered in the mutant strain are modeled through changing the upper and lower bounds $lb_{mut,j}$ and $ub_{mut,j}$. For example, in iJE660 flux #110 refers to the NADH-driven reduction of lactate to lactaldehyde. This reaction is reversible, and so for the wild-type $E. coli$, $-\infty \leq \nu_{110} \leq \infty$. The model for a mutant strain in which this reaction was deleted would have $0 \leq \nu_{110} \leq 0$. All of the other $lb_{j,mut}$ and $ub_{j,mut}$ are equal to $lb_{j,wt}$ and $ub_{j,wt}$. The MoMA objective favors small changes to a large number of fluxes.

The ROOM criterion species that the minimal number of fluxes is significantly perturbed in the mutant strain relative to the wild-type [35]. Mathematically, it is a mixed-integer linear programming (MILP) problem, as shown in (5.6).
\[
\min \sum_{j=1}^{n} y_j \quad (5.6)
\]

subject to:

\[
\begin{align*}
S \cdot \mathbf{v}_{\text{mut}} &= 0 \\

\nu_j - y_j (ub_{\text{mut},j} - w_j^u) &\leq w_j^u \\

\nu_j - y_j (lb_{\text{mut},j} - w_j^l) &\geq w_j^l \\
y_j &\in 0, 1 \\
w_j^u &= \nu_{\text{wt},j} + \delta |\nu_{\text{wt},j}| + \epsilon \\
w_j^l &= \nu_{\text{wt},j} - \delta |\nu_{\text{wt},j}| - \epsilon \\
\forall \ j &\in \{1, 2, 3, \ldots, n\}
\end{align*}
\]

In (5.6), the binary variables \( y_j \) take values of either 0 or 1. If \( y_j = 1 \), it indicates the flux \( \nu_{\text{mut},j} \) is significantly perturbed relative to the wild-type flux \( \nu_{\text{wt},j} \). Significance is specified by the parameters \( \delta \) and \( \epsilon \), which allow for slight amounts of relative and absolute deviation, respectively, of any individual flux \( \nu_{\text{mut},j} \) in the mutant, even if \( y_j = 0 \). In this work, we set \( \delta = 0.14 \) and \( \epsilon = 0.021 \), as these values are similar to those used by the original report introducing the ROOM criterion [35]. ROOM favors as large a change as is needed in the minimal number of fluxes necessary.

The optimization problems in (5.4), (5.5), and (5.6) were solved in MATLAB by using the TOMLAB software package to invoke CPLEX 9.0. MATLAB code for these routines is given in D. For some LP and QP calculations, the free linear and quadratic programming
solver CLP [23] and its free interface to MATLAB Mexclp [22] were used, but these packages were not capable of solving MILP problems (5.6).

5.2.4 Two approaches to predict gene targets using stoichiometric models

In the “genome scan” approach, gene knockouts or constrained overexpressions are simulated in a combinatorial fashion, and the flux state from each knockout or overexpression simulation is recorded. Knockouts or overexpressions whose flux states include an elevated tyrosine export flux are chosen as potential gene targets. For this work, we considered only single-gene overexpressions or single-gene knockouts in an effort to minimize computational complexity. To perform an *in silico* overexpression, the target flux was constrained to be $x$-fold its value in the wild-type strain.

In the “inverse” approach, the tyrosine export flux is artificially constrained at a higher-than-normal level, and the model is simulated. Here, the fluxes which must be perturbed (relative to a wild-type) in order to support the higher tyrosine export flux dictate the choice of gene targets. The genome scan approach assumes a given genotype, and from that genotype (as expressed in the modified stoichiometric constraints) it attempts to predict the resulting phenotype. In contrast, the inverse mode of simulation imposes a phenotype (high tyrosine flux) and asks what genotypes are necessary to best support the desired phenotype. ROOM is uniquely suited to this mode of use of the stoichiometric models, because it gives a minimum number of alterations required to support a given flux state.

These various optimization methods and computational approaches are illustrated schematically in Figure 5-1.
5.2.5 Estimation of intracellular amino acid pool sizes

To analyze the intracellular concentrations of amino acids, between 2 and 10 mL of *E. coli* cell culture was applied to a vacuum filtration unit using 47mm-diameter polysulfone filters with a pore size of 22μm (Millipore Corp., Bedford, MA). The bacteria adhered to the filter, and were washed one time with 2 to 10 mL of phosphate-buffered saline (150 mM sodium chloride, 10 mM sodium phosphate pH 7.2). The filter with adhered bacteria were immediately plunged into 2 mL of chilled chloroform. The polysulfone filters were solved by the chloroform, ensuring release of bacteria from the filter and cell lysis. One mL of methanol and 800 μL of tricine buffer were then added [40]. The extract was centrifuged at 20,000×g for 3 minutes, the upper aqueous layer was transferred to a fresh tube, and 800 μL each of methanol and buffer were added again to the chloroform/filter mixture, which was re-vortexes and centrifuged. The second upper aqueous layer was added to the first.

The resulting extract was analyzed for amino acids by derivatization with *o*-phthalaldehyde/3-mercaptopropionic acid to form the corresponding 1-(3-carboxypropyl)-2-alkylisoindole derivat...
tizes, which were detected by high pressure liquid chromatography on an Agilent 1100 HPLC system with a Zorbax Eclipse AAA column (4.6×75 mm with 3.5μm resin) and 40 mM sodium phosphate, pH 7.8, as the mobile phase. Detection was by absorbance at 338 nm [12], and quantification was by comparison to a commercial standard mix of amino acids (Agilent Technologies No. 5061-3330).

5.2.6 Culturing techniques

Cultures of *E. coli* were done in 250 mL unbaffled flasks growing at 37 °C with orbital shaking at 200~250 rpm. The culture medium was 50 mL of MOPS minimal medium, with 5 g/L of glucose. MOPS is usually formulated with 9.52 mM (0.5 g/L) of ammonium chloride [29], but we routinely supplemented ammonium chloride to 28.6 mM (1.5 g/L) for better stimulation of tyrosine production. The only exception was in the nitrogen supplementation studies, where MOPS with basal levels of ammonium chloride was supplemented with various nitrogen sources as indicated. Spectinomycin and kanamycin were used as required at standard concentrations [31]. Induction was by addition of IPTG to 1 mM at the time of inoculation.

For ΔaceE strains, which are auxotrophic for acetate [1, 21], 1 mM of sodium acetate was added to the culture medium. In aceE+ strains used as controls in experiments with ΔaceE, 1 mM sodium acetate was also added, but this supplementation did not significantly affect tyrosine production.
5.3 Results

5.3.1 MoMA and ROOM converge on a knockout target

Figure 5-2(a) shows the results of a “genome scan” approach to identifying single-gene deletion targets, using both MoMA and ROOM optimization criteria. Each point corresponds to the deletion of a particular gene. The full flux state \( v \) is projected into the two key dimensions of growth and tyrosine formation. Using the MoMA optimization criterion, only two gene deletions are predicted to result in both (i) a strain with increased flux to tyrosine export, and (ii) a viable cell (i.e. with non-zero growth rate). Those genes are \( aceE \) and \( ppc \). The \( aceE \) gene encodes the pyruvate dehydrogenase subunit of the pyruvate dehydrogenase multienzyme complex, and is responsible for the conversion of pyruvate to acetyl-CoA under aerobic culture conditions. The \( ppc \) gene encodes the enzyme phosphoenolpyruvate carboxylase, responsible for the anaplerotic carboxylation of PEP to oxaloacetate [17]. These knockout targets were found in all three of the tyrosine production backgrounds tested.

In contrast, applying ROOM to the genome-scan approach gave results that depended on the imposed level of tyrosine production, as shown in the second column of Figure 5-2(a). With no imposed tyrosine production, the only deletions which resulted in both non-zero growth yield and in better tyrosine production were \( eno \) and \( sdhA \), which encode for enolase, and succinate dehydrogenase, respectively. At an imposed tyrosine production flux of 0.01 mol / mol GLC, the identified knockout targets become four in number. \( sdhA \) is again observed as a deletion target, but \( eno \) is not. Instead, the model predicts that \( sucA \), \( ribE \), and \( adhC \) deletions will lead to greater tyrosine production. \( sucA \) encodes a subunit
Figure 5-2. Single-gene knockout predictions for increased tyrosine flux as predicted by (a) the genome-scan approach or (b) the inverse approach. In (a), the six panels differ by the optimization criterion used (left column, MoMA; right column, ROOM) and by the amount of forced tyrosine production imposed on the model (none, first row; 0.01 mol / mol GLC, second row, 0.05 mol / mol GLC, third row).
of the α-ketoglutarate dehydrogenase multienzyme complex responsible for the conversion of α-ketoglutarate to succinyl-CoA. ribE encodes 6,7-dimethyl-8-ribityllumazine synthase, responsible for the penultimate step in the biosynthesis of riboflavin. adhC is an older name for adhE, the familiar alcohol dehydrogenase responsible for NADH-driven reduction of acetyl-CoA to acetaldehyde and of acetaldehyde to ethanol.

Figure 5-2(b) shows the results of applying ROOM to the inverse approach to the identification of gene targets. The x-axis indicates the imposed tyrosine production flux, and the y-axis indicates the number of significantly perturbed fluxes, that is, the number of genes for which \( y_j = 1 \). As expected, the number of fluxes whose significant perturbation from the wild-type flux state is required to support tyrosine production increases as the imposed level of tyrosine production increases. The strains T1 and T2 produce tyrosine at yields on glucose of less than 0.1 g/g. The lowest levels of imposed tyrosine flux shown in Figure 5-2(b) is most relevant to these strains. Also, the number of significant flux perturbations required to support higher tyrosine production increases sharply, from under 10 significantly perturbed genes at tyrosine production of 0.1 g/g GLC or less, to greater than 50 for tyrosine production of 0.15 g/g GLC or higher. For these reasons, only the perturbations significant at the lowest levels of imposed overproduction were examined for further study.

The six identified perturbations included ppc and aceE, which had previously been identified in the MoMA-based genome-scan approach. Also identified were (i) trxB, encoding thioredoxin reductase, which is responsible for the NADPH-driven reduction of thioredoxin; (ii) the isoleucine/valine/leucine biosynthesis pathways encoded by the ilv and leu operon; and (iii) the pyrimidine biosynthesis pathway encoded by the pyr operon. Unlike the genome-scan simulations, the targets identified via the inverse approach do not have the correspond-
ing flux set to zero in the simulation. Rather, these genes encode for enzymes whose fluxes are necessarily significantly “perturbed” when tyrosine export is constrained to a certain level. The perturbation can be either towards higher flux relative to the wild-type parental strain, or toward lower flux. In the case of aceE, the perturbation was in the direction of lower flux.

Even cursory surveys of the literature revealed that some of the model predictions were highly problematic, especially for simulations using the ROOM algorithm. For example, eno deletions have long been known to render E. coli incapable of growth on minimal glucose medium [13, 5]. The ROOM criterion instead predicts that an eno deletion will be viable with glucose as the sole carbon source. The gene ribE was also identified as a single-gene knockout target. However, this gene is required for riboflavin biosynthesis, which is also required for E. coli viability on minimal medium [17].

Why does ROOM give these erroneous predictions? It is difficult to know exactly. One source of error may be the lack of gene and enzyme regulatory information in the stoichiometric model. In genetically perturbed strains, cellular metabolism may not be able to circumvent what in the wild-type strain is an advantageous regulatory barrier. Stoichiometric models incorporating regulatory information have been developed, but are mathematically even more complex than the MILP-based ROOM model described here.

Viewed another way, ROOM encourages the extreme modification of only a few metabolic fluxes. Take the case of eno. It catalyzes the conversion of 2-phosphoglycerate to PEP. It is the final step in the lower glycolytic pathway responsible for converting glyceraldehyde-3-phosphate (GAP) to PEP. E. coli must find an alternate route to synthesize PEP and pyruvate if enolase is deleted. Stoichiometrically, a route from GAP through a reverse
tartronate semialdehyde pathway (in which 2 glyoxylates are produced per GAP, and are in turn are converted by TCA and glyoxylate shunt enzymes into pyruvate, PEP, oxaloacetate, or acetyl-CoA) is possible and indeed is predicted by ROOM. In reality, this path is impossible, at least in part due to enzyme regulation. The glyoxylate shunt is controlled by the $iclR$ repressor, which suppresses transcription of these genes during growth on glucose, and additionally, the tartronate semialdehyde pathway is repressed by glyoxylate through the action of $allR$ [15]. Additional thermodynamic and regulatory constraints may also prevent pathway flux in this direction; such a pathway has never been observed in any bacterium, despite recent discoveries in glyoxylate assimilation [36].

Some investigators have attempted to extend the stoichiometric models to account for these types of regulatory behaviors, usually by adding Boolean constraints between fluxes. These models are considerably more complex and have not been as widely studied[7].

Despite the obvious limitations of the purely stoichiometric models, we were encouraged by the emergence of $aceE$ as a knockout target under several scenarios. Thus, it was chosen for experimental study.

5.3.2  $aceE$ has an effect on tyrosine opposite to the models’ prediction

The $\Delta aceE$ gene was deleted in $E. coli$ K12, $E. coli$ K12 $\Delta tyrR$, $E. coli$ T1, and $E. coli$ T2 backgrounds. These strains were tested for tyrosine production, and the results are shown in Figure 5-3.

Counter to the prediction of both the genome-scan MoMA and the inverse ROOM approaches to stoichiometric modeling, deletion of $aceE$ did not enhance tyrosine production.
In the Δ\textit{tyr}R background, the addition of Δ\textit{ace}E seemed to have little effect on tyrosine production. However, both T1 and T2 backgrounds Δ\textit{ace}E was less than half of its value in the parental strains.

Stoichiometric modeling had predicted \textit{ace}E deletion to enhance tyrosine production in a variety of contexts, and so the experimental repudiation of this finding suggests that the assumptions of the stoichiometric model are not met for the system of tyrosine overproduction in \textit{E. coli}. In particular, the stoichiometric model cannot account for thermodynamic limitations in pathway flux, and can only account for regulatory effects through the crude, indirect proxy of the optimization criterion. If a particular enzyme reaction step is at near-equilibrium conditions, the flux through this pathway step can be increased no further. As discussed above for \textit{eno}, relevant regulatory information suggests that the ROOM model’s prediction for the \textit{eno} deletion is inaccurate.

What does this discussion have to do with \textit{ace}E? One interpretation of the effect of \textit{ace}E on cellular metabolism is that this deletion should vastly reduce the oxidation of pyruvate to acetyl-CoA. Likely this would lead to an increased supply of the tyrosine precursor PEP
as well as pyruvate. Additionally, decreased acetyl-CoA levels would lead to lower fluxes through the TCA cycle, thus conserving carbon that would normally be lost to CO₂, and suppressing the formation of NADH from NAD⁺. Since tyrosine biosynthesis requires NAD⁺ for the oxidative decarboxylation of prephenate, lower NADH and higher NAD⁺ levels could also be advantageous. This interpretation is "stoichiometric". Since fewer moles of carbon are pushed towards acetate, and fewer moles of oxidation equivalents are wasted in the TCA cycle, more are available for tyrosine.

Of course, this explanation is contraindicated by the data. Stoichiometric considerations thus do not appear to be controlling the formation of tyrosine. What was? To test the hypothesis that thermodynamic limitations at a particular reaction step were related to limitations in tyrosine flux, we examined the equilibrium constants at physiological conditions of the tyrosine biosynthesis pathway. As shown in Table 5.1, nearly all the equilibrium constants were very large, with the exception of the final step in tyrosine biosynthesis, the tyrosine transaminase reaction. The only reactions which are possible reversible under physiological conditions are 3-dehydroquinate dehydratase, shikimate dehydrogenase, shikimate kinase, and tyrosine transaminase. No experimental equilibrium constant is available for shikimate kinase, but for analogous reactions such as glucose kinase or glycerol kinase, equilibrium constants are known to be $\approx 200$ [4]. Workers studying the enzymes in the pathway from 5-enoylpyruvylshikimate-3-phosphate synthase to prephenate dehydrogenase have described all of these enzymes as "irreversible" and have never been able to observe in vitro flux in the backwards direction [2, 19, 16]. In contrast, the transamination step has an apparent equilibrium constant $K'$ of 1.28.

This suggests that any bottlenecks upstream of tyrosine transaminase in the tyrosine
Table 5.1. The enzymatic reactions of tyrosine biosynthesis and their apparent equilibrium constants

<table>
<thead>
<tr>
<th>Enzyme</th>
<th>E. coli</th>
<th>Gene</th>
<th>Reaction</th>
<th>EC Number</th>
<th>Equilibrium Constant</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>2-dehydro-3-deoxy-D-arabino-heptonate-7-phosphate synthase (DAHP synthase)</td>
<td>aroG</td>
<td></td>
<td>PEP(aq) + E4P(aq) + H2O ⇌ DAHP(aq)</td>
<td>4.1.2.15</td>
<td>5 × 10^6</td>
<td>[38]</td>
</tr>
<tr>
<td>3-dehydroquinate synthase</td>
<td>aroB</td>
<td>aroD</td>
<td>DAHP(aq) ⇌ 3-dehydroquinate(aq) + P(aq)</td>
<td>4.6.1.3</td>
<td>3 × 10^13</td>
<td>[37]</td>
</tr>
<tr>
<td>3-dehydroquinate dehydratase</td>
<td>aroE, ydiB</td>
<td></td>
<td>3-dehydroquinate(aq) ⇌ 3-dehydroshikimate(aq)</td>
<td>4.2.1.10</td>
<td>4.8</td>
<td>[37]</td>
</tr>
<tr>
<td>Shikimate dehydratase</td>
<td>aroE, ydiB</td>
<td></td>
<td>3-dehydroshikimate(aq) + H2O ⇌ shikimate + NADP⁺(aq) + H⁺</td>
<td>1.1.1.25</td>
<td>12.6</td>
<td>[41]</td>
</tr>
<tr>
<td>Shikimate kinase</td>
<td>aroK, aroL</td>
<td></td>
<td>shikimate(aq) + ATP(aq) ⇌ 3-phosphoshikimate(aq) + ADP(aq)</td>
<td>2.7.1.71</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5-enolpyruvylshikimate-3-phosphate synthase (EPSP synthase)</td>
<td>aroA</td>
<td></td>
<td>3-phosphoshikimate(aq) + PEP(aq) ⇌ EPSP(aq) + P(aq)</td>
<td>2.5.1.19</td>
<td>180</td>
<td>[2]</td>
</tr>
<tr>
<td>chorismate synthase</td>
<td>aroC</td>
<td></td>
<td>EPSP(aq) ⇌ chorismate(aq) + P(aq)</td>
<td>4.2.3.5</td>
<td>&gt; 2.4 × 10^{-3}</td>
<td>[3]</td>
</tr>
<tr>
<td>chorismate mutase</td>
<td>tyrA</td>
<td></td>
<td>chorismate(aq) ⇌ prephenate(aq)</td>
<td>5.4.99.5</td>
<td>&gt; 9700</td>
<td>[16]</td>
</tr>
<tr>
<td>prephenate dehydrogenase</td>
<td>tyrA</td>
<td></td>
<td>prephenate(aq) + NAD⁺(aq) ⇌ 4-hydroxyphenylpyruvate + NADH(aq) + CO₂(g)</td>
<td>1.3.1.12</td>
<td>&gt; 1 × 10^{12}</td>
<td></td>
</tr>
<tr>
<td>Tyrosine aminotransferase</td>
<td>tyrB</td>
<td></td>
<td>4-hydroxyphenylpyruvate + glutamate(aq) ⇌ tyrosine(aq) + α-ketoglutarate(aq)</td>
<td>2.6.1.5</td>
<td>1.28</td>
<td>[39]</td>
</tr>
</tbody>
</table>

Pathway will stem from insufficient enzymatic activity. The thermodynamic driving force for these reactions is very likely to be positive (at least unless metabolite pool sizes at successive steps of the reaction are not several orders of magnitude different). Enzyme activity can be boosted by increasing enzymatic expression level (see Chapter 6) or by eliminating activity-repressing phenomena such as allosteric inhibition [24]. The transamination step may limit tyrosine production flux regardless of the activity TyrB or other transaminases. If so, tyrosine production flux will be limited not by pathway enzyme activity, but by the pool sizes of glutamate and α-ketoglutarate.

Since glutamate is a nitrogen donor for the transamination of phenylalanine, aspartate, isoleucine, leucine, valine, and alanine, and histidine as well as tyrosine, we hypothesized that tyrosine overproduction may be constrained by equilibrium at the tyrosine aminotransferase reaction, and that this perturbation would affect the intracellular concentration of a variety of amino acids. To investigate this possibility, we measured the intracellular concentration of 16 of the 20 proteinogenic amino acids.
5.3.3 Measurement of amino acid pool sizes and a nitrogen limitation of hydrogen production

Our extraction technique for isolation of intracellular amino acids took less than 3 minutes from perturbation of the *E. coli* culture to cell lysis by immersion in cold methanol/chloroform. The turnover time for amino acids in *E. coli* is in many cases faster than the 2-3 minutes required for our procedure [42]. Our data are thus best interpreted not as precise, absolute measurements of intracellular pool sizes, but as pool size estimates suitable for comparing the intracellular amino acid pools of various strains.

In tyrosine overproducing strains T1 and T2 the intracellular tyrosine pool was between 3.5 and 12 mM. In contrast non-overproducing strains *E. coli* K12, *E. coli* Δ*tyrR* had intracellular tyrosine concentrations of less than 1 mM. The data show that the phenylalanine pool is positively correlated with the tyrosine pool. This is not surprising, as phenylalanine in synthesized from the same precursors as tyrosine, and shares all but two reactions in its biosynthesis. Interestingly, histidine, lysine, and to a lesser extent arginine, were also positively correlated to measured tyrosine pools.

Aspartate and alanine pool sizes were strongly anti-correlated with the tyrosine pool size. As an example, Figure 5-4 shows a scatter plot of tyrosine concentration vs. alanine concentration for several strains.

Glutamate, the direct nitrogen donor to tyrosine, was not appreciably changed in the tyrosine overproduction strains, perhaps because it has a much larger pool size in wild-type *E. coli* than many other amino acids [42].

The data suggested that the supply and distribution of nitrogen in the *E. coli* amino acid
pools was seriously perturbed by tyrosine overproduction. To further test this hypothesis, we tried supplementing the MOPS minimal medium with two other nitrogen sources: glutamate and glutamine. As controls, metabolically related carboxylic acids were used in place of the amino acids. Yields of tyrosine on total carbon (glucose + supplement) in the medium is shown in 5-5.

Figure 5-5 illustrates that the supply of nitrogen is limiting tyrosine production in \textit{E. coli}. Supplementation of nitrogen sources glutamine or glutamate led to increases in tyrosine yield to nearly 10% g/g for strain T2, an increase of nearly 8-fold. In contrast, the by-product of amino-group transfer from glutamate to tyrosine, \( \alpha \)-ketoglutarate, abolished overproduction of tyrosine when supplemented to the media. Glutamine, a major source of glutamate in the ammonium assimilation of wild-type \textit{E. coli}, gave an increase in tyrosine production nearly
the same as did glutamate. Citrate or acetate did not enhance tyrosine production, but also did not show the strongly inhibit tyrosine production to the same degree as α-ketoglutarate.

5.4 Discussion and conclusions

5.4.1 The pitfalls of stoichiometric models

The model system for this chapter was tyrosine overproduction in *E. coli*. Initially, a variety of computational approaches rooted in stoichiometric, optimization based models of *E. coli* metabolism were used to search for gene targets leading to enhanced tyrosine overproduction. This search failed: the predictions of the various flavors of models unanimously agreed that *aceE* was a gene whose deletion would enhance tyrosine overproduction. Experimental prediction failed to validate the model prediction. In fact, *aceE* deletion halved the observed tyrosine yield on glucose.
5.4.2 Other evidence for nitrogen control of tyrosine metabolism

The failure of stoichiometric models to predict successful gene deletion targets, and the identification of other bottlenecks through measurement of intracellular amino acids is in accord with other reports on tyrosine overproduction in *E. coli*. It is interesting to compare Table 5.1 to research performed subsequently to these experiments. Lütke-Eversloh and Stephanopoulos [26] have combinatorially overexpressed genes in the tyrosine biosynthesis pathway to identify enzymatic bottlenecks. They found that overexpression of *ydiB* and *aroK* both led to greater tyrosine production. These enzymes code for shikimate dehydrogenase and shikimate kinase. Tyrosine production was increased by modulating the gene expression level, showing that enzymatic activity, and not equilibrium constraints, are controlling flux through these pathway steps. Interestingly, when overexpressing both *ydiB* and *aroK*, an additional combinatorial overexpression search revealed the tyrosine aminotransferase *tyrB* as an overexpression target. This datum undergirds our conclusion that in the strains T1 and T2, expression of TyrB activity is sufficiently high that the amino transfer reaction approaches equilibrium. In combining T2 with overexpressions of *ydiB* and *aroK*, however, apparently the upstream metabolite pools in the tyrosine biosynthesis pathway could no longer be processed to near-equilibrium levels by TyrB.

Santos and Stephanopoulos have, in several unpublished studies, isolated strains with enhanced tyrosine production identified through high-throughput screening that were believed to carry unknown chromosomal mutations partially responsible for the enhanced tyrosine production. Interestingly, microarray analysis showed that two of these strains, designated rpoA27 and rpoD3, had vastly elevated levels of expression of the *gadABC* genes. GadA
and GadB together comprise glutamate decarboxylase, responsible for decarboxylating glutamate to \(\gamma\)-aminobutyrate. This molecule can then fuel GadC, a glutamate/\(\gamma\)-aminobutyrate antiporter. One function of the operon is the mitigation of acid toxicity: it imports an extracellular glutamate (GadC) and decarboxylates it (GadAB), thereby removing acidity. The less acidic product, \(\gamma\)-aminobutyrate, is exported in place of the original glutamate. In the minimal medium used by both Santos cite(something) and in this work, extracellular concentrations of both \(\gamma\)-aminobutyrate and glutamate are very low, so the relevance of the GadC reaction is unclear. The exact relationship between glutamate’s enhancement of tyrosine production shown in Figure 5-5 and these findings is thus difficult to pinpoint. Perhaps the pathway through \(\gamma\)-aminobutyrate serves only to deplete pools of \(\alpha\)-ketoglutarate, which is aminated by \(\gamma\)-aminobutyrate through the action of GabT to form glutamate and succinate semialdehyde. However, the microarray analysis showed that \(gdhA\), \(gltD\), and \(gltB\) were all down-regulated. Whatever the precise mechanism, the microarray evidence adds support to the conclusion that nitrogen assimilation and metabolism exerts important control over tyrosine production.

\[5.4.3\ \text{Conclusions}\]

Through examination of thermodynamic data for the tyrosine biosynthesis pathway shown in Table 5.1, and by measurement of intracellular concentrations of amino acids for wild-type and tyrosine-overproducing strains, we were led to hypothesize two facts. The first was that for any bottlenecks upstream of tyrosine aminotransferase, increasing enzyme activity was likely to overcome the bottleneck, but that the supply of nitrogen, specifically the equilibration of the tyrosine transaminase reaction, controlled tyrosine overproduction in \(E.\ coli\).
In the absence of other upstream perturbations, thermodynamic considerations suggest it is unlikely that increasing the enzyme activity of the transaminase is likely to alleviate this bottleneck.

This case study in tyrosine accords with Chapter 4, in which similar metabolite-measuring strategies were used to examine steps in a metabolic pathway from the standpoint of equilibrium. Such approaches have been used before to identify points of regulatory and allosteric regulation in wild-type bacterial strains [20]. This work represents an extension of similar approaches to the system of tyrosine biosynthesis in *E. coli*. 
Bibliography


CHAPTER 6

Molecular tools for optimizing the behavior of targeted metabolic pathways

This Chapter:

- Motivates the need for molecular tools useful in perturbing gene expression across a continuum of values.
- Describes an error-prone mutagenesis and characterization program for the creation of E. coli promoter variants.
- Applies the promoter library to several metabolic phenotypes of interest, showing two examples of optimal phenotypes at intermediate values of gene expression.

6.1 Introduction

Protein engineering via directed evolution and gene shuffling [10, 28] has been extensively applied for the systematic improvement of protein properties such as antibody binding affinity [4], enzyme regulation [23], and increased or diverse substrate specificity [8]. A similar approach whereby continuously improved mutants are generated along a selection-defined trajectory in the sequence space can also be applied for the systematic improvement or modification of other types of biological sequences, e.g. ribozymes [9, 30]. We show here that

Note: This chapter is based on material previously published by the author [1].
promoters can also be engineered via directed evolution to achieve precise strengths and regulation, and, by extension, can constitute libraries exhibiting broad ranges of genetic control. Typically, the deletion [33] and the strong over-expression [24] of genes have been the principal strategies for elucidation of gene function. These two methods sample the continuum of gene expression at only a few discrete points, determined by experimental feasibility [12] and not necessarily biological significance. Thus, the full dependency of phenotype on gene expression may not be accessible due to the limitations inherent in these methods. Despite prior attempts [13, 14, 16], no previous work has developed a fully-characterized, homogeneous, broad-range, functional promoter library and demonstrated its applicability to the analysis of such a genetic control. While inducible promoters allow for a continuous control of expression at the macroscopic level, practical applications of these systems are limited by prohibitive inducer costs, hypersensitivity to inducer concentration, and transcriptional heterogeneity at the single-cell level [22, 27]. The latter factor in particular, can limit the effect of inducers in a culture to a simple increase of the number of cells expressing the gene of interest instead of the overexpression of the gene in all cells. Inducible systems are suitable in certain applications (e.g. recombinant protein overproduction) [26]; however, the elucidation of gene function and genetic control on phenotype requires well characterized promoter libraries which are homogeneous at the single cell level.
6.2 Materials and methods

6.2.1 Strains and media

*E. coli* DH5α (Invitrogen) was used for routine transformations as described in the protocol. *E. coli* K12 (MG1655) and *E. coli* K12 PT5-dxs, PT5-idi, PT5-ispFD (provided by DuPont) were used for promoter engineering examples. In specified strains the lycopene expression was performed using the pAC-LYC plasmid [5] and assayed as described previously [1]. Assay strains were grown at 37 °C with 225 RPM orbital shaking in M9-minimal media [25] containing 5 g/L D-glucose. When necessary, the M9 media was supplemented with 0.1% casamino acids. All other strains and propagations were cultured at 37°C in LB media. Media was supplemented with 68 µg/mL chloramphenicol, 20 µg/mL kanamycin, and 100 µg/mL ampicillin as necessary. Glucose monitoring was conducted using the r-Biopharm kit. Cell density was monitored spectrophotometrically at 600 nm. All PCR products and restriction enzymes were purchased from New England Biolabs; PCR was with Taq polymerase. M9 minimal salts were purchased from US Biological and all remaining chemicals were from Sigma-Aldrich.

6.2.2 Library Construction

Nucleotide analogue mutagenesis was carried out in the presence of 20µM 8-oxo-2-deoxyguanosine (8-oxo-dGTP) and 6-(2-deoxy-β-D-ribofuranosyl)-3,4-dihydro-8H-pyrimido-[4,5-c][1,2]oxazin-7-one (dPTP) [31]. Using plasmid pZE-gfp(ASV) kindly provided by M. Elowitz as template [7] along with the primers PL_sense_AatII: PRIMER SEQUENCE and PL_anti_EcoRI:
CCGGAATTCCGGTCAGTGCGTCCTGCTGAT, 10 and 30 amplification cycles with the primers mentioned above were performed. The 151 bp PCR products were purified using the GeneClean Spin Kit (Qbiogene). Following digestion, the product was ligated overnight at 16 °Covernight and transformed into library efficiency *E. coli* DH5α (Invitrogen). About 30,000 colonies were screened by eye from minimal media-casamino acid agar plates and 200 colonies, spanning a wide range in fluorescence intensity, were picked from each plate.

### 6.2.3 Library Characterization

#### Initial Characterization

About 20 µL of overnight cultures of library clones growing LB broth were used to inoculate 5mL M9G medium supplemented with 0.1% w/v casamino acid (M9G/CAA) and the cultures were grown at 37 °C with orbital shaking. After 14 h, a sample of the culture was centrifuged at 18,000 x g for 2 minutes and the cells were resuspended in ice-cold water. Flow cytometry was performed on a Becton-Dickinson FACScan and the geometric mean of the fluorescence distribution of each clonal population was calculated. In order to ensure that bulk, population-averaged measurements could reflect the underlying single-cell behavior, only clones with clean, monovariate distributions of fluorescence were retained for further analysis. Twenty-seven clones were isolated in this way. Promoters were sequenced using primers PL_Left_seq (AGATCTGGCCGCAAGAAA) and PL_Right_seq (GCCATGGAACAGGTAGTT-TTCCAG). Sequencing revealed that these 27 clones represented 22 unique promoter sequences.
Promoter Strength Metric

Shake flasks containing 50 mL of M9G/CAA medium were inoculated with 1% v/v of an overnight LB culture of a library clone. The culture turbidity (A600nm) and fluorescence (Packard Fusion microplate fluorescence reader, Perkin-Elmer, Boston, MA) were monitored as a function of time. Fluorescence readings taken during the exponential growth phase were plotted as a function of turbidity. The best-fit slope to this line represents the exponential-phase steady-state concentration of GFP, \( f_{ss} \). Because \( f_{ss} \) is affected by the cell growth rate, the oxygen-dependent maturation constant of GFP, and the protease-mediated degradation of GFP as in addition to the promoter-driven synthesis of new GFP, it is not a suitable metric for promoter strength. Instead, we used a previously published dynamic model [18] that accounts for all of these factors. Under this model and under the assumption that the rate constant of protease-mediated degradation is the same for mature GFP as its precursor polypeptide, \( P \), the rate of promoter-driven production of GFP, can be expressed as in Equation 1.

\[
P = f_{ss} \mu \left( 1 + \frac{\mu}{m} + m \right) + f_{ss} D \left( \frac{2\mu}{m} + \frac{D}{m} + 1 \right)
\]  

(6.1)

In Equation 1, \( f_{ss} \) is the steady-state fluorescence intensity of a culture, \( \mu \) is the culture growth rate, \( m \) is the maturation constant for oxygen-dependent activation of the GFP fluorophore, \( D \) is the first-order rate constant for protease-mediated degradation. Estimates of \( m \) and \( D \) of 1.5 h\(^{-1}\) and 0.23 h\(^{-1}\), respectively (10, 11), were obtained from the literature. The parameters \( f_{ss} \) and \( \mu \) were measured separately for each member of the promoter library. \( P \), in relative fluorescence units per absorbance unit per hour, was calculated from Equation 6.1
for each clone. We performed duplicate cultures for each clone.

**Transcriptional Analysis**

Cultures inoculated as previously were grown for 3 h and the total RNA was extracted from a 1.5 mL sample with a commercial kit (RNEasy, Qiagen Corp). All samples were diluted to a final concentration of 20 µg/mL and stored at -20 °C. A commercial kit for RT-PCR (iScript One-Step RT-PCR Kit with SYBR Green, Bio-Rad) was used with a CCD-equipped thermal cycler (iCycler, Bio-Rad) for RT-PCR of the gfp transcript. Primers (sense\texttt{ATGGCTAGCAAGGGAGAAGA} and antisense\texttt{ATCCATGCCATGTGTAATCC}) were used at a final concentration of 100 nM and 20 ng of RNA was used as template in each 50 µL reaction. We performed duplicate cultures for each clone and duplicate extractions for each culture. The threshold cycles for each sample were calculated from the fluorescence data with proprietary software (Bio-Rad, Inc).

**Chloramphenicol Resistance**

pZE-promoter-cat plasmids were created by PCR of the CAT gene from pACYC184 using primers CAT\_Sense\_MluI: \texttt{CGACGCGTATTTCTGCCATTCATCCGCTTATTATCA} and CAT\_Anti\_KpnI: \texttt{CGGGGTACCTTTACGGCTTAAGGAAGCTAAATGGA} and ligated into the proper pZE-promoter construct which was previously digested by KpnI and MluI. Exponential-phase cultures grown in LB supplemented with kanamycin were plated onto LB agar supplemented with kanamycin and various concentrations of chloramphenicol ranging from 0 to 500 µg/mL. After overnight incubation at 37 °C, the lowest concentration of chloramphenicol that inhibited the growth of a clone was recorded.
6.2.4 Promoter Delivery Construction

Promoter replacements were conducted using PCR product recombination [6] using the pKD46 plasmid expressing the lambda red recombination system and pKD13 as the template for PCR. Promoter replacements were verified through colony PCR using the k1, k2 and kt primers along with the verification primers listed below. To create the cassette for promoter replacement, two fragments were amplified via PCR. Fragment 1 contained the promoter with primer homology to the upstream region of the endogenous promoter. Fragment 2 contained the kanamycin maker from pKD13 and had homology to an area downstream of the endogenous promoter or gene. These two fragments had an internal homology to each other of 25 base pairs to allow for self-annealing and subsequent amplification of a single cassette which was used (100 ng) for the transformation. For the case of *dxs*, the entire gene was amplified and used as a third fragment which was annealed with the previous two. This provided higher recombination efficiency due to the increased homology region. The following sets of primers were used in the construction of these fragments:

**ppc fragment**

ppc-pze Sense: GTTGTAGCCCTGTATCCTTCAGTCCGACTTGGCGCGAATATGCTCGGCATCTTCTTTCTCCTCTTTAAT-GAATTCGG

pze-pkd13 shunt: GAAGCAGCTCCAGCCTACACTCCGACGCTCTAAGAAAACCATTATTA

pkd13 sense: GTTCCGGCTGAGCTGCTTC

pkd13-ppc anti: CATTTCCATAAGTTACGCTTATTTAAAGCGTCGTGAATTTAATGACGTAATCCGTCGACCTGCAGTTCA

verification: CCGATCCCTGGCTATGAATGC

**dxs fragment**
dxs-pze Sense: TGGGTGGAGTCGACCAGTGCCAGGGTCGGGTATTTGGCAATATCAAAACTCATCACTCCTCTTTAATGAATT-
CGG

pze-pkd13 shunt: GAAGCAGCTCCAGCCTACACTCCGACGTCTAAGAAACCATTATTA

pkdl3 sense: GTGTAGGCTGGAGCTGCTTC

pkd13-dxs anti: ACTCGATACCTCGGCACTGGAAGCGCTAGCGGACTACATCATCCAGCGTAATAAAAATCGTGCACCTGCAGT-
TCGA

dxs sense: ATGAGTTTTTGATATTCGA

dxs anti: TTATGCCAGCCAGGCCCTTG

verification: GTCAGAGCGTCGGAATAGCCAGAC

6.3 Results

A derivative of the constitutive bacteriophage PL-λ promoter [20] was mutated through error-
prone PCR [31], cloned into a reporter plasmid upstream of a degradation-tagged GFP gene
[3], and screened in E. coli based on the fluorescence signal in a glucose minimal medium.
Medium was supplemented with 0.1% casamino acids to attenuate GFP toxicity. Nearly
200 promoter mutants, spanning a wide range of GFP fluorescence, were selected. Many of
these initially screened promoters exhibited large variations in fluorescence between several
trials or did not have an acceptable single-cell level homogeneity. Twenty-two mutants were
finally chosen to form a functional promoter library based on reproducible and homoge-
neous single-cell fluorescence distributions as measured by flow cytometry and fluorescence
microscopy. Figure 6-1 displays representative fluorescence photomicrographs of homo- and
heterogeneously expressing library clones. Figure 6-2 illustrates the process of creating and subsequently selecting these promoters.

![Image](image_url)

Figure 6-1. Representative light-field (right) and false-color dark-field (left) photomicrographs of clones with (a) highly heterogeneous or (b) homogeneous expression of GFP. Only promoters with relatively homogeneous fluorescence distributions (b) were chosen for further analysis. The two images in (b) correspond to promoters with a relative promoter strength metric of 0.124 and 0.417 from top to bottom.

In light of the uncertainty surrounding the concept of promoter strength [11] and the poor reliability of single reporter-gene-based systems, we performed a multi-faceted characterization of each library member. We first determined the promoter strength in the library strains (in units of GFP fluorescence per cell per hour) by measuring culture fluorescence and using a dynamic equation balancing GFP production and degradation [18]. The promoter strength of the library members was found to span a 196-fold range with a mean spacing of
29% between adjacent members (Figure 6-5).

Next, to characterize the promoter library directly at the transcriptional level, we measured the relative mRNA levels of *gfp* transcripts in the above cultures by quantitative RT-PCR. The high correlation between fluorescence and mRNA level (Figure 6-5) confirmed that expression was transcriptionally controlled. The mRNA level spanned a 325-fold range with a mean spacing of 32% between adjacent members. We then formed an average promoter strength metric for each promoter by averaging the scaled mRNA and fluorescence data. This promoter strength metric is shown, along with promoter sequences, in Table 6.1.

Finally, to verify the constitutive nature of all the promoters, each was redeployed into a new construct driving the reporter gene cat. Cultures bearing these constructs were assayed for resistance to chloramphenicol on a rich, solid-phase medium. The MIC spanned a 26-fold range with a mean spacing between MIC values was 17% which is biased due to a discrete levels of chloramphenicol tested. Figure 2 displays the high correlation between these three metrics of promoter performance. These data indicate that the library exhibits a high dynamic range which behaves similarly regardless of the gene being regulated. Moreover, these conditions test the promoter library in contrasting medium and growth environments (liquid minimal medium vs. solid complex medium) further underscoring the constitutive nature of the library promoters.
Figure 5.2. A variant of the constitutive bacteriophage PL-\(\lambda\) promoter was mutated through error-prone PCR, used in a plasmid construct to drive the expression of \(\text{gfp}\), then screened based on fluorescence of colonies. The chosen constructs have a wide range of fluorescence both on a culture-wide level and on a single-cell level as illustrated by representative flow cytometry histograms. All of the selected promoters have a uniform expression level on a single cell level as measured by GFP signal.
Several orthogonal metrics were employed to characterize the promoter library and ensure the consistent behavior of all its members for various genes and culturing conditions. We show here three metrics that were chosen to quantifying transcriptional of the promoters: (1) The dynamics of GFP production based on fluorescence, (2) measurement of the relative mRNA transcript levels in the cultures, and (3) testing of the MIC for chloramphenicol in an additional library of constructs where the promoter drove the expression of CAT. The overall strong correlation between the various metrics suggests a broad-range utility of the promoter library for a variety of genes and conditions.
<table>
<thead>
<tr>
<th>Promoter Name</th>
<th>Promoter Sequence</th>
<th>Promoter Strength $P$</th>
<th>mRNA Cm</th>
<th>MIC</th>
<th>Average Promoter Strength Metric</th>
<th># Mutations</th>
</tr>
</thead>
<tbody>
<tr>
<td>AA</td>
<td>CAATTCCGAGGCTCTAAGAAGGACCATTATTATTATGAC-ATTAACTCTAATAGCCGTATACAGCCGCGCCCTT-CCGCTCTCGCTAGTCGCAAAGGCGTGACCTCAGTGATGTTTTTGTACAATTCATACGTGATGAGAATACGTGACAT-CAGCAGGAGCAGCACTGACC</td>
<td>40942.21</td>
<td>0.236732</td>
<td>225</td>
<td>0.219304</td>
<td>3</td>
</tr>
<tr>
<td>B</td>
<td>CAATTCCGAGGCTCTAAGAAGGACCATTATTATCATTGAC-ATTAAACCTAATAAGGCGGTACAGAGGGCGCCCTT-TCGCTCTACCTAATAGGCGTGACCTCAGTGATGTTTTTGTACAATTCATACGTGATGAGAATACGTGACAT-CAGCAGGAGCAGCACTGACC</td>
<td>143258.01</td>
<td>0.906126</td>
<td>372</td>
<td>0.813623</td>
<td>7</td>
</tr>
<tr>
<td>BB</td>
<td>CAATTCCGAGGCTCTAAGAAGGACCATTATTATCATTGAC-ATTAAACCTAATAGGCGGTACAGAGGGCGCCCTT-TCGCTCTACCTAATAGGCGTGACCTCAGTGATGTTTTTGTACAATTCATACGTGATGAGAATACGTGACAT-CAGCAGGAGCAGCACTGACC</td>
<td>85320.056</td>
<td>0.407102</td>
<td>338</td>
<td>0.418989</td>
<td>7</td>
</tr>
<tr>
<td>E</td>
<td>CAATTCCGAGGCTCTAAGAAGGACCATTATTATCATTGAC-ATTAAACCTAATAGGCGGTACAGAGGGCGCCCTT-TCGCTCTACCTAATAGGCGTGACCTCAGTGATGTTTTTGTACAATTCATACGTGATGAGAATACGTGACAT-CAGCAGGAGCAGCACTGACC</td>
<td>183693.69</td>
<td>0.743061</td>
<td>375</td>
<td>0.834644</td>
<td>4</td>
</tr>
<tr>
<td>F</td>
<td>CAATTCCGAGGCTCTAAGAAGGACCATTATTATCATTGAC-ATTAAACCTAATAGGCGGTACAGAGGGCGCCCTT-TCGCTCTACCTAATAGGCGTGACCTCAGTGATGTTTTTGTACAATTCATACGTGATGAGAATACGTGACAT-CAGCAGGAGCAGCACTGACC</td>
<td>194748.86</td>
<td>0.686715</td>
<td>372</td>
<td>0.826154</td>
<td>11</td>
</tr>
<tr>
<td>Promoter Name</td>
<td>Promoter Sequence</td>
<td>Promoter Strength $P$</td>
<td>mRNA</td>
<td>Cm MIC</td>
<td>Average Promoter Strength Metric</td>
<td># Mutations</td>
</tr>
<tr>
<td>---------------</td>
<td>-------------------------------------------------------------------------------------------------------</td>
<td>-----------------------</td>
<td>------</td>
<td>--------</td>
<td>---------------------------------</td>
<td>-------------</td>
</tr>
<tr>
<td>II</td>
<td>CAATTCCGACGTCTAAGAAACCATTATTATCATGAC-ATTACCTATTAAATAGGGGTATCACGAGGCCCTT-TCGCTCTACCCCATGATGAGAT-GGACATCCCTATCAGTGATAGATCTGACACAT-CAGCAGGACGCACCTGACC</td>
<td>5781.5417</td>
<td>0.063856</td>
<td>34</td>
<td>0.038711</td>
<td>1</td>
</tr>
<tr>
<td>JJ</td>
<td>CAATTCCGACGTCTAAGAAACCATTATTATCATGAC-ATTACCTATTAAATAGGGGTATCACGAGGCCCTT-TCGCTCTACCCCATGATGAGAT-GGACATCCCTATCAGTGATAGATCTGACACAT-CAGCAGGACGCACCTGACC</td>
<td>18896.294</td>
<td>0.159416</td>
<td>68</td>
<td>0.120082</td>
<td>1</td>
</tr>
<tr>
<td>K</td>
<td>CAATTCCGACGTCTAAGAAACCATTATTATCATGAC-ATTACCTATTAAATAGGGGTATCACGAGGCCCTT-TCGCTCTACCCCATGATGAGAT-GGACATCCCTATCAGTGATAGATCTGACACAT-CAGCAGGACGCACCTGACC</td>
<td>50376.463</td>
<td>0.29891</td>
<td>282</td>
<td>0.274361</td>
<td>4</td>
</tr>
<tr>
<td>L</td>
<td>CAATTCCGACGTCTAAGAAACCATTATTATCATGAC-ATTACCTATTAAATAGGGGTATCACGAGGCCCTT-TCGCTCTACCCCATGATGAGAT-GGACATCCCTATCAGTGATAGATCTGACACAT-CAGCAGGACGCACCTGACC</td>
<td>129869.65</td>
<td>0.757858</td>
<td>338</td>
<td>0.707972</td>
<td>3</td>
</tr>
<tr>
<td>N</td>
<td>CAATTCCGACGTCTAAGAAACCATTATTATCATGAC-ATTACCTATTAAATAGGGGTATCACGAGGCCCTT-TCGCTCTACCCCATGATGAGAT-GGACATCCCTATCAGTGATAGATCTGACACAT-CAGCAGGACGCACCTGACC</td>
<td>90494.778</td>
<td>0.366629</td>
<td>372</td>
<td>0.409463</td>
<td>2</td>
</tr>
<tr>
<td>Promoter Name</td>
<td>Promoter Sequence</td>
<td>Promoter Strength $P$</td>
<td>mRNA</td>
<td>Cm MIC</td>
<td>Average Promoter Strength Metric</td>
<td># Mutations</td>
</tr>
<tr>
<td>---------------</td>
<td>----------------------------------------------------------------------------------</td>
<td>-----------------------</td>
<td>------</td>
<td>--------</td>
<td>----------------------------------</td>
<td>-------------</td>
</tr>
<tr>
<td>OO</td>
<td>CAATTCCGACGTCTAAAGAAACCATTATCATGACAT-CAACCTATAAAATAGGGCTATCAGGAGCCCTTCTCTGCTCTCCACTCAAGCTCCCTATCTAGTGATAGGCGAT-TGACATCCCTATCATGACGATTTATTCGAGGAGATTTGAGCACAT-CAGCAGGGAGCAGCTGAGC</td>
<td>996.99754</td>
<td>0.003075</td>
<td>17</td>
<td>0</td>
<td>57</td>
</tr>
<tr>
<td>P</td>
<td>CAATTCCGACGTCTAGGAAACGTTATCATGACGCCCCCTATCAAAGGAGCCCTTCTCTGCTCTCCACTCAAGCTCCCTATCTAGTGATAGGCGAT-TGACATCCCTATCATGACGATTTATTCGAGGAGATTTGAGCACAT-CAGCAGGGAGCAGCTGAGC</td>
<td>160631.4</td>
<td>0.575729</td>
<td>425</td>
<td>0.686331</td>
<td>55</td>
</tr>
<tr>
<td>PLTetO1</td>
<td>CAATTCCGACGTCTAAGAAACCATTATCATGACAT-ATTAACCTATAAAATAGGGCTATCAGGAGCCCTTCTCTGCTCTCCACTCAAGCTCCCTATCTAGTGATAGGCGAT-TGACATCCCTATCATGACGATTTATTCGAGGAGATTTGAGCACAT-CAGCAGGGAGCAGCTGAGC</td>
<td>145806.85</td>
<td>1</td>
<td>338</td>
<td>0.862491</td>
<td>0</td>
</tr>
<tr>
<td>Q</td>
<td>CAATTCCGACGTCTAGGAAACGTTATCATGACAT-ATTAACCTATAAAATAGGGCTATCAGGAGCCCTTCTCTGCTCTCCACTCAAGCTCCCTATCTAGTGATAGGCGAT-TGACATCCCTATCATGACGATTTATTCGAGGAGATTTGAGCACAT-CAGCAGGGAGCAGCTGAGC</td>
<td>195662.5</td>
<td>0.736003</td>
<td>450</td>
<td>0.857432</td>
<td>1</td>
</tr>
<tr>
<td>R</td>
<td>CAATTCCGACGTCTAGGAAACGTTATCATGACAT-ATTAACCTATAAAATAGGGCTATCAGGAGCCCTTCTCTGCTCTCCACTCAAGCTCCCTATCTAGTGATAGGCGAT-TGACATCCCTATCATGACGATTTATTCGAGGAGATTTGAGCACAT-CAGCAGGGAGCAGCTGAGC</td>
<td>32141.611</td>
<td>0.126381</td>
<td>225</td>
<td>0.140672</td>
<td>6</td>
</tr>
<tr>
<td>Promoter Name</td>
<td>Promoter Sequence</td>
<td>Promoter Strength $P$</td>
<td>mRNA</td>
<td>Cm MIC</td>
<td>Average Promoter Strength Metric</td>
<td># Mutations</td>
</tr>
<tr>
<td>---------------</td>
<td>-----------------------------------------------------------------------------------</td>
<td>-----------------------</td>
<td>------</td>
<td>--------</td>
<td>----------------------------------</td>
<td>-------------</td>
</tr>
<tr>
<td>S</td>
<td>CAATTCCGACGTTCTAAGAAACCATTATTATCATGAC-ATTACCTATAAAAATAGGCGATACGCCGCCCC-TCGCTCCCTACCTGAGTACCCTACATGATAGAGATGACATCCCTAACCCTGTGATAGAGATCAGCAGGAGCCACTGACC</td>
<td>41376.671</td>
<td>0.393587</td>
<td>282</td>
<td>0.285051</td>
<td>1</td>
</tr>
<tr>
<td>T</td>
<td>CAATTCCGACGTCTAAGAAACCATTATTATCATGAC-ATTACCTATAAAAATAGGCGATACGCCGCCCC-TCGCTCCCTACCTGAGTACCCTACATGATAGAGATGACATCCCTAACCCTGTGATAGAGATCAGCAGGAGCCACTGACC</td>
<td>63080.517</td>
<td>0.333052</td>
<td>338</td>
<td>0.324903</td>
<td>14</td>
</tr>
<tr>
<td>U</td>
<td>CAATTCCGACGTCTAAGAAACCATTATTATCATGAC-ATTACCTATAAAAATAGGCGATACGCCGCCCC-TCGCTCCCTACCTGAGTACCCTACATGATAGAGATGACATCCCTAACCCTGTGATAGAGATCAGCAGGAGCCACTGACC</td>
<td>43264.021</td>
<td>0.376762</td>
<td>282</td>
<td>0.285285</td>
<td>69</td>
</tr>
<tr>
<td>V</td>
<td>CAATTCCGACGTCTAAGAAACCATTATTATCATGAC-ATTACCTATAAAAATAGGCGATACGCCGCCCC-TCGCTCCCTACCTGAGTACCCTACATGATAGAGATGACATCCCTAACCCTGTGATAGAGATCAGCAGGAGCCACTGACC</td>
<td>161088.47</td>
<td>0.574349</td>
<td>450</td>
<td>0.686484</td>
<td>21</td>
</tr>
<tr>
<td>W</td>
<td>CAATTCCGACGTCTAAGAAACCATTATTATCATGAC-ATTACCTATAAAAATAGGCGATACGCCGCCCC-TCGCTCCCTACCTGAGTACCCTACATGATAGAGATGACATCCCTAACCCTGTGATAGAGATCAGCAGGAGCCACTGACC</td>
<td>109462.28</td>
<td>0.525424</td>
<td>450</td>
<td>0.540319</td>
<td>0</td>
</tr>
<tr>
<td>Promoter Name</td>
<td>Promoter Sequence</td>
<td>Promoter Strength $P$</td>
<td>mRNA</td>
<td>Cm MIC</td>
<td>Average Promoter Strength Metric</td>
<td># Mutations</td>
</tr>
<tr>
<td>---------------</td>
<td>-------------------</td>
<td>----------------------</td>
<td>------</td>
<td>--------</td>
<td>----------------------------------</td>
<td>-------------</td>
</tr>
<tr>
<td><strong>Y</strong></td>
<td>CAATTCCGACGTCTAAGAAACATTATTATCATGAC-ATTAACTTATAAAATAGGGGTATACAGGCACCCCTT-TCGCTTTCCACCTCGAGTCCTATCGTAGGAGAT-TGACATCCTCTATCGTAGATAGAGACACTGAGCAGCAT-CAGCAGGACCACCTGACC</td>
<td>70234.541</td>
<td>0.273757</td>
<td>338</td>
<td>0.31076</td>
<td>2</td>
</tr>
<tr>
<td><strong>Z</strong></td>
<td>CAATTCCGACGTCTAAGAAACATTATTATCATGAC-ATTAACTTATAAAATAGGGGTATACAGGCACCCCTT-TCGCTTTCCACCTCGAGTCCTATCGTAGGAGAT-TGACATCCTCTATCGTAGATAGAGACACTGAGCAGCAT-CAGCAGGACCACCTGACC</td>
<td>146023.14</td>
<td>0.75968</td>
<td>450</td>
<td>0.751938</td>
<td>6</td>
</tr>
</tbody>
</table>

Table 6.1. (Continued)
We applied the functional promoter library to introduce precise transcriptional control in the investigation of specific genetic effects on a cellular phenotype. We performed chromosomal promoter delivery into the region upstream of the targeted gene, replacing the native promoter and its inherent regulation modality. The utility of the promoter library was tested by investigating the effect of two endogenous genes (\textit{ppc} and \textit{dxs}) on two divergent phenotypes—growth yield and lycopene production. First, we investigated the growth yield from glucose as a function of the expression level of the \textit{ppc} gene in \textit{Escherichia coli}. \textit{E. coli}'s native \textit{ppc} promoter was replaced with varying-strength promoter-\textit{ppc} constructs, and these mutants were cultured while biomass and glucose concentrations were periodically monitored. Figure 3a presents the exponential-phase biomass yields as a function of the average promoter strength metric. Increasing \textit{ppc} levels have a positive effect on the biomass yield only to a certain point. This increase reaches a plateau, and further increases in the \textit{ppc} level have a negative effect on the biomass yield. Nevertheless, these results illustrate an optimum in the expression level of \textit{ppc} that is above that found from endogenous expression.

In this second case, volumetric productivity of lycopene accumulation in glucose medium was investigated as a function of the expression levels of the \textit{dxs} gene in two different \textit{E. coli} strains: the wild-type K12 strain and a previously engineered strain which already produces lycopene in high titers [2]. Figure 3b shows the lycopene production in these \textit{dxs} constructs in a wild-type (K12) background. Elevating \textit{dxs} expression increases lycopene accumulation only until a certain point. Beyond this optimum, increased \textit{dxs} expression is detrimental for lycopene production. Finally, the strength of the native \textit{dxs} promoter can be inferred from this analysis as is illustrated on the graph. In contrast to the above results, a linear relationship was obtained when similar promoter-\textit{dxs} constructs were placed in an
engineered strain [2] overexpressing downstream genes in the isoprenoid pathway (ispFD and idi). Figure 3c illustrates a nearly linear response of lycopene production to varying levels of dxs expression, suggesting that in the new genetic background, dxs has become rate-limiting.

![Graph showing selected promoters were integrated into the promoter region of ppc and strains were cultured in M9-minimal media with glucose as the sole carbon source.](image)

**Figure 6-4.** Selected promoters were integrated into the promoter region of ppc and strains were cultured in M9-minimal media with glucose as the sole carbon source.

### 6.4 Discussion and Conclusions

#### 6.4.1 Discussion

The 200 random promoter mutants we screened varied widely in their expression strength and clonal expression heterogeneity. Screening for only those promoters which drive stable, monovariate expression in culture by flow cytometry was critical for deployment of our promoter constructs in pathway analysis and expression optimization. Working only with ho-
Figure 6-5. Selected promoters were integrated in front of the dxs gene in an (a) wild-type E. coli strain or (b) a recombinant strain also overexpressing ispFD and idi, and these promoter-replacement strains were later assayed for the production of lycopene. A clear maximum in lycopene production was obtained in (a). From the wild-type production level, the native dxs promoter strength can be inferred to be around 0.26 according to our metric. In (b), the linear response of lycopene yield to the promoter strength illustrates a rate limiting behavior of dxs across all tested promoter strengths.

Homogeneous expressers enabled establishment of a multi-dimensional quantitative assessment of promoter strength. Reliance on bulk averages would obscure the underlying relationship between expression and phenotype. Furthermore, the use of an integrated system allowed us to bypass the instabilities and inherent mutation rates associated with the over-expression of endogenous genes using plasmid-based systems [32]. We tested the promoter engineering concept for the analysis of two different phenotypes in E. coli. In the first, the expression of ppc was modulated to effect biomass yield from glucose. This gene expresses phosphoenol pyruvate (PEP) carboxylase, a key anaplerotic enzyme. A ppc knockout is lethal for E. coli in glucose minimal medium [21]. Furthermore, overexpression of this gene has been shown to improve the growth yield on glucose [19]. These data imply two possibilities: either biomass yield is a monotonically increasing function of ppc expression or there exists a
particular $ppc$ expression level which maximizes yield. Our data show that the latter is the case. Possible reasons why ever-increasing $ppc$ levels lead eventually to a decrease in yield include the metabolic burden of severe overexpression of $ppc$ or the creation of a futile ATP-wasting cycle in metabolism where PEP is converted to oxaloacetate by $ppc$ and back again by pck, the gene for PEP carboxykinase. In addition to the global, pleiotropic phenotype of growth yield, we also employed promoter engineering in the study of a single metabolic pathway, by modulating $dxs$ expression and measuring lycopene biosynthesis. Kinetic control of metabolic pathways is often distributed and dependent on the expression level of several genes within the pathway [29]. The gene $dxs$ represents the first committed step in isoprenoid synthesis in $E. coli$ and has been implicated in control of lycopene production [17]; however, the quantitative nature of this control was unclear, and promoter delivery experiments also allowed us to quantify this control in multiple backgrounds (Figures 3b and 3c). In the case of wild-type $E. coli$, an optimal $dxs$ expression was again apparent. Past the optimum, increasing $dxs$ expression lowers lycopene yield, presumably due to the inadequate activity of downstream enzymes in the isoprenoid pathway and resulting toxic buildup of DXP. In contrast, in a strain already engineered to overexpress $idi$, $ispF$, and $ispD$, downstream genes in lycopene biosynthesis, no maximum is apparent. A linear response to an enzyme concentration is expected for rate-controlling genes exhibiting a high flux control coefficient for a given pathway [15], suggesting that even at the highest expression levels examined in this study, the $dxs$-catalyzed reaction is rate-limiting for lycopene biosynthesis. We also note that cell density in both strains was greatly reduced in the constructs harboring low-strength promoters, which was expected, as $dxs$ is an essential gene (CITE). The creation of a library of promoter mutants in yeast illustrates the applicability of
this approach in both prokaryotic and eukaryotic contexts. As with *E. coli*, flow cytometry allowed isolation of only those promoters with relatively homogeneous reporter gene expression. It is possible to further extend and refine the selection process to create libraries of conditional promoters, active only under specified conditions. We have recently applied this selection methodology to create conditional genetic control elements which are responsive to environmental perturbations (e.g. oxygen concentration) (data not shown).

6.4.2 Conclusions

For the first time, we have created a general framework for the precise, quantitative control of gene expression in vivo. Our strategy allows (1) achievement of any desired expression level for a specific gene, (2) optimization of gene expression for maximal (or minimal) pathway function, and (3) a means for the analysis of the distribution of genetic control on pathway behavior. In two disparate examples we have shown that pathway function can exhibit well-defined extrema with respect to levels of gene expression. The existence of these extrema evinces the need for precise gene-dosage studies for the full understanding of pathway behavior. The creation and detailed characterization of a promoter library as described here is a facile and robust means to such an end.
Bibliography


APPENDIX A

The renewable production of propylene from glucose using a hybrid fermentative-hydrothermal process

The initial goal of the study described in Chapter 3 was to produce butyric acid in recombinant E. coli. However, no butyrate production could be observed in strains expressing clostridial biosynthesis genes. Further analyses pinpointed that the ptb-buk reaction step (Figure 3-6) and the bcd reaction step (Chapter 4 as liming pathway flux towards butyrate. This bottlenecks could be avoided if 3-hydroxybutyric acid was chosen as a target. This molecule is more oxidized than butyrate and less likely to be useful in biofuels applications, but is receiving increasing attention has a stereochemical specialty product.

In collaboration with Andrew Peterson in Prof. Jefferson Tester’s Supercritical Fluids Laboratory at MIT, we have demonstrated the conversion of 3-hydroxybutyric acid at concentrations near what we have obtained by fermentation (Figure 3-5) to propylene at high specificities and yields under hydrothermal conditions, according to the reaction 3HB → H2O + propylene + CO2. This process is shown schematically in Figure A-1.

The propylene is obtained as a component of a highly concentrated gas stream as shown in Table A.1, with very low amounts of other hydrocarbons, especially ethylene. Such a gas stream considerably facilitates the separation and purification of propylene. In petroleum
Fermentatively produced 3-hydroxybutyric acid can be hydrothermally converted to high-purity propylene gas. (a) shows a schematic chemical reaction and characteristic reaction conditions. (b) shows graphs demonstrating each process step, individually. On the left and as shown in Chapter 3, 3HB is produced from glucose. On the right, aqueous solutions of pure 3-hydroxybutyric acid are decarboxylated and dehydrated to form propylene. The shift from sub-critical to supercritical conditions at the time of 20:24 and concomitant decrease in propylene output suggests the reaction may proceed by a non-radical, two-electron mechanism with charged intermediates/transition states.
Table A.1. Fermentation-derived 3-hydroxybutyric acid is converted to propylene under hydrothermal conditions. Pure-water solutions of 3-hydroxybutyric acid (20 g/L) was fed at 0.2 mmol / min to a constantly-stirred hydrothermal reactor vessel operating at 371 °C and 250 bar, where it was allowed to react for a mean residence time was approximately 14 min. Outlet gas properties are given below.

<table>
<thead>
<tr>
<th>Component</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>TOTAL FLOW</td>
<td>0.1 mmol / min</td>
</tr>
<tr>
<td>Propylene</td>
<td>69%</td>
</tr>
<tr>
<td>Carbon dioxide</td>
<td>20%</td>
</tr>
<tr>
<td>Ethane*</td>
<td>8%</td>
</tr>
<tr>
<td>Hydrogen</td>
<td>5%</td>
</tr>
<tr>
<td>Propane</td>
<td>&lt; 3%</td>
</tr>
<tr>
<td>Methane</td>
<td>&lt; 3%</td>
</tr>
</tbody>
</table>

* The identification of the GC-FID peak corresponding to this species as ethane is preliminary and awaits further confirmation.

Refining operations today, propylene is produced by the cracking of naphtha hydrocarbons, which also produce high levels of propane, ethylene, and other small hydrocarbons. These impurities necessitate purification by fractional distillation, a relatively expensive process. In addition, a high-yielding route to propylene from renewable biomass has not yet been reported in the literature.
APPENDIX B

Toward measuring butanol pathway intermediates in a cell-free system

B.1 Introduction

As described in Chapter 1 of this thesis, the cell wall can make it impossible to perform feeding studies on the function of metabolic pathways whose intermediates are not permeable to the cell wall. This brief appendix describes an initial attempt at using cell-free extracts of *E. coli* cytoplasm to circumvent this limitation.

B.2 Materials and Methods

Cell-free extracts were prepared from essentially wild-type *E. coli*, and mixed with appropriate nutrients, co-factors, and buffers as described [1]. At time zero, plasmid DNA isolated from an *E. coli* strain expression butanol biosynthesis gene (Chapter 4) was added to the cell-free reaction. At the same time or at 1 hour later, acetyl CoA and NADH were

---

Note: This appendix describes a single experiment done in collaboration with M. Jewett of Harvard Medical School and Northwestern University. Consult his Ph.D. thesis [1] for details on the types and sources of the cell-free extracts used in this work.
spiked to final concentrations of 5 mM and 1 mM, respectively.

The 15 μL cell free reactions were quenched after four hours of reaction by the addition of 85 μL of 1 M trifluoroacetic acid.

I made standards by spiking aliquots 1 M trifluoroacetic acid with known amounts of various CoAs so that the final volume was 100 μL. I used hexanoyl-CoA as an internal standard, adding 31 nmol to each standard and sample before derivatization. I took 85 μL of these samples and vacuum-fuged them before derivatization with chloroacetaldehyde and HPLC analysis.

B.3 Results and discussion

Acetyl-CoA and free coenzyme A were the only CoA intermediates I reliably detected in cell-free reactions containing plasmids for butyrate biosynthesis Figure B-1. For reference, the amount of acetyl-CoA added to our reactions was 4.2 mM. Amounts shown below range from 2 to 3.6 mM. Error bars are the standard error in the mean. For the PANOx system, AcCoA levels were noticeably lower in the experimental reactions than in the no-plasmid controls, suggesting that perhaps AcCoA was consumed in these reactions. This was not true in the Cytomim system.

I was also able to detect a peak corresponding to free coenzyme A. No distinct trend was apparent in these data either, although perhaps the PANOx reactions had slightly higher CoAs. Somewhat surprisingly, CoASH levels were much higher than AcCoA levels. Whereas in wild-type aerobically grown E. coli cells, the AcCoA:CoASH ratio has been reported to be up to 5:1, in the extracts, the ratio is between 0.047 and 0.081. The level of CoASH observed
Figure B-1. (a) Analysis of clostridial butyrate biosynthesis gene expression by SDS-PAGE in cell-free E. coli extracts spiked with four butanol biosynthesis plasmids. Protein samples were taken six hours after adding plasmid DNA to the reaction. (b) Measurement of CoA pool sizes in the cell-free reactions. Only acetyl-CoA and free coenzyme A were detected in this initial experiment.
is far in excess of both the added AcCoA to CoASH, suggesting that possibly this CoASH derives from the extract itself.

However, if the CoASH analysis above is accurate, another reason may be a thermodynamically unfavorable driving force. The thermodynamic driving force depends on the AcCoA:CoASH ratio, and on the NADH:NAD ratio. If the NADH:NAD+ ratio is 1.0, and the AcCoA:CoASH ratio is 5, pathway intermediates should be easily observable. However, if these ratios are each drastically lower, the maximum, equilibrium, level of downstream CoAs would be much, much lower. Perhaps these ratios could be increased after a time period for protein synthesis, possibly by the addition of respiration inhibitors like cyanide, piericidin, or rotenone.
Assuming that enthalpies of reaction do not vary with temperature, the van't Hoff correction can be used to estimate $K'$ values at arbitrary temperatures, as shown in (C.1).

$$\ln \frac{K'(T)}{K'(T_{ref})} = -\frac{\Delta H'}{R} \left(\frac{1}{T} - \frac{1}{T_{ref}}\right)$$  \hspace{1cm} (C.1)

To correct $K'$ values to an arbitrary pH when they have been measured at a reference pH of pH$_{ref}$, an analogous equation, shown in (C.2), can be used.

$$\ln \frac{K'(pH)}{K'(pH_{ref})} = -\ln(10) \left[pH \Delta n_H (pH) - pH_{ref} \Delta n_H (pH_{ref})\right]$$  \hspace{1cm} (C.2)

The change with reaction of the number of hydrogen ions, $\Delta n_H$, is in general a function of pH, so that the term in brackets in (C.2) indicates that the change between two pHs of the product pH $\times \Delta n_H$ must be evaluated to make the correction.

Calculation of $\Delta n_H$ as a function of pH for given reactions requires the pKa's for all acid-base ionizations involving all the species participating in the reaction, as described by Alberty.

The corrections in (C.2) and (C.1) can be combined, yielding (C.3)
\[ \ln \left( K'(T, \text{pH}) \right) = \]
\[ \ln \left( K'(T_{ref}, \text{pH}_{ref}) \right) - \frac{\Delta_r H^\circ}{R} \left( \frac{1}{T} - \frac{1}{T_{ref}} \right) \]  
\[ - \ln(10) \Delta_{\text{pH}}(\text{pH} \, \Delta_r n_H) - \ln(10) \Delta_{\text{pH}}(\text{pH} \, \Delta_r n_H) \left( 1 - \frac{T}{T_{ref}} \right) \]  

where \( \Delta_{\text{pH}} \) refers to the change in a quantity at one pH relative to another, and \( \Delta_r \) refers to the change of a quantity for a given chemical reaction. The extra term \( - \ln(10) \Delta_{\text{pH}}(\text{pH} \, \Delta_r n_H) \left( 1 - \frac{T}{T_{ref}} \right) \) arises from the pH dependence of the transformed reaction enthalpy \( \Delta_r H^\circ \).
APPENDIX D

MATLAB code for optimization-based models of E. coli metabolism

D.1 MATLAB .m file for solving the FBA problem

```matlab
% solveFBA.m
% FBA on a stoichiometric model of metabolism
% solves the FBA problem: optimizing the "growth flux" of a
% inputs
% objective: a length n vector which "points to" the growth flux
% Smat: an m-by-n matrix with stoichiometric information for the
% model
% bounds.L: lower bounds on the fluxes
% bounds.U: upper bounds on the fluxes
% zeroRHS: an m-vector of zeros which forms the RHS of S.nu = 0
% forceOUTgn: an integer pointing to a flux which should be below a
% given threshold in the perturbed but not wt strain.
% forceOUTfx: a lower bound constraint to be put on the 'forceOUTgn'th
% flux
% forceKO: a vector of integers which force certain fluxes to be
% identically zero in all perturbed strains
% cpxControl: []
% outputs
% a MATLAB struct FBAsoln with fields:
% fluxes: the optimal fluxes (may not be unique)
% lambda: the "shadow prices" of metabolites
% red.cost: the reduced costs of fluxes
% mu: the optimal cost, i.e., the growth rate
% result: a flag which reports termination condition of the cplex
% solver; is 1 if an optimum is found
% basis: an n*m-length vector containing the basis of the found
% optimum
```
% targ target product
% ko.mat an m x n matrix of indices to ko. m different sims run,
% the n indices in each row point to positions to ko in each run
%outputs:
% prodXTo n vector of product fluxes
% mu n vector of growth rates
% workedQ n vector whose nth element is 1 if the optimization at the
% nth pt worked

function [prodXTo,mu,workedQ,yield] =
scanProdGrowthEnvl(n, targ, ko.mat, obj, S, lb, ub, zeroRHS)

hold all;
ub.ko = ub;
lb.ko = lb;

for ind = 1:length(ko.mat');
    [foo, goo, kos] = find(ko.mat(ind,:));
    ub.ko(kos) = 0;
lb.ko(kos) = 0;
    [prodXTo,mu,workedQ,yield] =
        makeProdGrowthEnvl(n, targ, obj, S, lb.ko, ub.ko, zeroRHS);
    ub.ko = ub;
lb.ko = lb;
clear kos;
clear foo;
clear goo;
end

hold off;

return
function FBAsoln = solveFBA(objective, Smat, bounds_L, bounds_U, zeroRHS, forceOUTgn, forceOUTfx, forceKO);

if nargin < 5
    forceOUTgn = 729;  % location of TYRxtO flux in iJE660b is row 729
    forceOUTfx = 0;    % force yield of 1 mol TYR per 10 mol GLC
    forceKO = [];      % location of GLCUP2R in iJE660b is row 520
    PriLev = 0;
    cpxControl = struct();
end

[m,n] = size(Smat);

% set the forced overproduction constraints
bounds_L(forceOUTgn) = forceOUTfx;

% set the forced KO constraints
for ind = 1:length(forceKO)
    bounds_L(forceKO(ind)) = 0;
    bounds_U(forceKO(ind)) = 0;
end

[fluxes, slack, lambda, redcosts, mu, n_inf, s_inf, Inform, basis] =
    cplex(objective, Smat, bounds_L, bounds_U, zeroRHS);

FBAsoln.fluxes = fluxes;
FBAsoln.lambda = lambda;
FBAsoln.redcosts = redcosts;
FBAsoln.growth = -mu;
FBAsoln.result = Inform;
FBAsoln.basis = basis;

return

D.2 MATLAB .m file for solving the MoMA problem

solveMoMA.m
MoMA on a stoichiometric model of metabolism
solves the MoMA problem: minimizing the Euclidean distance between a
"wild-type" flux-state and a perturbed flux state while still satisfying
all stoichiometric constraints.

% inputs
% wt: a length n vector of fluxes to which the "perturbed" MoMA strain
% will be compared.
% Smat: an m-by-n matrix with stoichiometric information for the
% model
% bounds_L: lower bounds on the fluxes
% bounds_U: upper bounds on the fluxes
% zeroRHS: an m-vector of zeros which forms the RHS of S.nu = 0
% forceOUTgn: an integer pointing to a flux which should be below a
given threshold in the perturbed but not wt strain.
% forceOUTfx: a lower bound constraint to be put on the 'forceOUTgn'th
% flux
% forceKO: a vector of integers which force certain fluxes to be
% identically zero in all perturbed strains
% in same units of Smat (most likely mol tyr per 10 mol glc consumed)
% cpxControl: []

% outputs
% a MATLAB struct ROOMsoln with fields:
% fluxes: the optimal fluxes (may not be unique)
% lambda: the "shadow prices" of metabolites
% red_cost: the reduced costs of fluxes
% mu: the optimal cost, i.e., the minimum number of
% significantly perturbed genes
% result: a flag which reports termination condition of the cplex
% solver; is 1 if an optimum is found
% basis: an n*m-length vector containing the basis of the found
% optimum

function MoMAsoln = solveMoMA(wt, Smat, bounds_L, bounds_U, zeroRHS,
forceOUTgn, forceOUTfx, forceKO, cpxControl, PriLev)
% set defaults on inputs not supplied by user

if nargin < 5
    forceOUTgn = 729; % location of TYRxtO flux in iJE660b
    forceOUTfx = 1; % force yield of 1 mol TYR per 10 mol GLC
    forceKO = []; % location of GLCUP2R in iJE660b
    PriLev = 0;
    cpxControl = struct();
elseif nargin < 8
    PriLev = 0;
    cpxControl = struct();
end

[m,n] = size(Smat);

% set the forced overproduction constraints
bounds_L(forceOUTgn) = forceOUTfx;

% set the forced KO constraints
for ind = 1:length(forceKO)
    bounds_L(forceKO(ind)) = 0;
    bounds_U(forceKO(ind)) = 0;
end
bounds.Lmoma = bounds.L - wt;
bounds.Umoma = bounds.U - wt;
F = speye(n,n);
zeroLinObj = zeros(n,1);

[dif_fluxes, slack, lambda, red.costs, mu, n.inf, s.inf, Inform, basis] =
cplex(zeroLinObj, Smat, bounds.Lmoma, bounds.Umoma, zeroRHS,
[],[],[],[],[],[],[],[],[],[F,[],[],[],[],[],[]],
MoMAsoln.fluxes = dif_fluxes + wt;
MoMAsoln.lambda = lambda;
MoMAsoln.red.costs = red.costs;
MoMAsoln.opt = mu;
MoMAsoln.result = Inform;
MoMAsoln.basis = basis;

return

D.3 MATLAB .m file for solving the ROOM problem

solveROOM.m
ROOM on a stoichiometric model of metabolism
solves the ROOM problem: minimizing the number of "significantly
perturbed" fluxes to predict flux states for perturbed strains, given te
flux state of a wild-type strain (found by FBA).
inputs
wt: a length n vector of fluxes to which the "perturbed" ROOM strain
will be compared.
Smat: an m-by-n matrix with stoichiometric information for the
model
bounds.L: lower bounds on the fluxes
bounds.U: upper bounds on the fluxes
zeroRHS: an m-vector of zeros which forms the RHS of S.nu = 0
forceOUTgn: an integer pointing to a flux which should be below a
given threshold in the perturbed but not wt strain.
forceOUTfx: a lower bound constraint to be put on the 'forceOUTgn'th
flux
forceKO: a vector of integers which force certain fluxes to be
identically zero in all perturbed strains
in same units of Smat (most likely mol tyr per 10 mol glc consumed)
cpxControl: []

outputs
a MATLAB struct ROOMsoln with fields:
fluxes: the optimal fluxes (may not be unique)
lambda: the "shadow prices" of metabolites
red_cost: the reduced costs of fluxes
mu: the optimal cost, i.e., the minimum number of
significantly perturbed genes
result: a flag which reports termination condition of the cplex
solver; is 1 if an optimum is found
basis: an n*m-length vector containing the basis of the found
optimum

function ROOMsoln = solveROOM(wt, Smat, bounds_L, bounds_U, zeroRHS,
forceOUTgn, forceOUTfx, forceKO, del, eps, cpxControl, PriLev)

% set defaults on inputs not supplied by user
if nargin < 5
    forceOUTgn = 729;  % location of TYRxtO flux in iJE660b
    forceOUTfx = 1;    % force yield of 1 mol TYR per 10 mol GLC
    forceKO = [];      % location of GLCUP2R in iJE660b
    del = 0.03;
    eps = 0.01;
    PriLev = 0;
    cpxControl = struct();
elseif nargin < 8
    del = 0.03;
    eps = 0.01;
    PriLev = 0;
    cpxControl = struct();
end

[m,n] = size(Smat);

% set the forced overproduction constraints
bounds_L(forceOUTgn) = forceOUTfx;

% set the forced KO constraints
for ind = 1:length(forceKO)
    bounds_L(forceKO(ind)) = 0;
    bounds_U(forceKO(ind)) = 0;
end

%%% this whole setup for the ROOM procedure assumes a state variable which
%%% is a column vector where the 1st n elements are the fluxes, and the
%%% 2nd n elements are the y_i, the integer vars for each rxn which
%%% tells whether it is strongly perturbed or no

%%% it also assumes that constraints involving integer variables will be
%%% added to the BOTTOM of the list of constraints

% construct the w_i_L and w_i_U parameters, which specify how big of a
% change in a flux is allowed before it counts as "big".
% this can be thought of as a window in which fluxes are absolutely free to
% move however they want
window_L = wt - del*abs(wt) - eps;
window_U = wt + del*abs(wt) + eps;

% form the n * 2*n matrix of lower bound constraints
% (on the window) involving integer vars
left.block = speye(n, n);
right.block.coefs = -(bounds_L - window_L);
right.block.temp = speye(n,n);
for ind = 1:n
    right.block(ind,:) = right.block.coefs' .* right.block.temp(ind,:);
end
lb.mat = [left.block, right.block];
lb_rhs = window.L;

% form the n * 2*n matrix for % upper bound constraints
% (on the window) involving integer vars
left.block = speye(n, n);
right_block.coefs = -(bounds_U - window.U);
right.block.temp = speye(n,n);
for ind = 1:n
    right.block(ind,:) = right.block.coefs' .* right.block.temp(ind,:);
end
ub.mat = [left.block, right.block];
ub_rhs = window.U;

% concatenate ub.mat and lb.mat into a single matrix for all of the
% constraints involving integer variables. for the upper bound
% constraints, the inequality sign is mat.state <= rhs, but we multiply by
% negative one so the inequality sign will reverse and be the same as with
% the lower bound constraints. All inequalities are now greater than signs
% >=, so are constraints only on the lower side
intvar.mat = [lb.mat, -ub.mat];
intvar.rhs = [window.L; -window.U];

%% combine these constraints with the stoichiometric constraints
intvar.lb = zeros(n,1);
intvar.ub = ones(n,1);

bounds_L = [bounds_L; intvar.lb];
bounds.U = [bounds_U; intvar.ub];

% the stoichometric Smat is m*n, but now the state var is 2*n+1, so we
% need to pad zeros in the UR quadrant of Smat to get the appropriate no.
% of cols to concatenate with the integer constraints
URquad = spalloc(m,n,1);
Smat = [Smat,URquad];
Smat = [Smat;intvar.mat];
% the state vector construction in this problem has the noninteger
% variables (i.e. the fluxes) in the 1st n elements, and the integer
% variables in the last n elements
intvarList = linspace(n+1,2*n,n); % due to multiplying by -1 above, all constraints are now of the form
% S.nu_i > rhs_i
RHS_L = [zeroRHS; intvar_rhs];
RHS_U = [zeroRHS; Inf*ones(n+2,1)];

% the objective in ROOM is minimization of the sum of integer variables
room_obj = [zeros(n,1); ones(n,1)];

[fluxes, slack, lambda, red_costs, mu, n_inf, s_inf, Inform, basis] =
    cplex(room_obj, Smat, bounds_L, bounds_U, RHS_U, RHS_L,
            [], [], PriLev, [], intvarList,
            [], [], [], [], [], [], [], [], [], []);
ROOMsoln.fluxes = fluxes(1:n);
ROOMsoln.intvars = fluxes((n+1):2*n);
ROOMsoln.lambda = lambda;
ROOMsoln.red_costs = red_costs;
ROOMsoln.opt = mu;
ROOMsoln.result = Inform;
ROOMsoln.basis = basis;

return

D.4 MATLAB .m file to vary the forced output level of tyrosine and calculate ROOM fluxes as a result

% scanTYR_ROOM.m
% a function to scan over forced tyrosine output levels
% and make ROOM predictions of the fluxes at each point.
% inputs
% wt: a length n vector of fluxes to which the "perturbed" ROOM strain
% will be compared.
% Smat: an m-by-n matrix with stoichiometric information for the
% model
% bounds_L: lower bounds on the fluxes
% bounds_U: upper bounds on the fluxes
% zeroRHS: an m-vector of zeros which forms the RHS of S.nu = 0
% forceOUTgn: an integer pointing to a flux which should be below a
% given threshold in the perturbed but not wt strain.
% forceOUTfx: a vector of lower bound constraint to be put on the 'forceOUTgn'th
% flux, with each element being applied in successive iterations
% forceKO: a vector of integers which force certain fluxes to be
identically zero in all perturbed strains
in same units of Smat (most likely mol tyr per 10 mol glc consumed)
del: a parameter in ROOM specifying the relative drift fluxes are
allowed to take before being considered "perturbed"
eps: a parameter in ROOM specifying the absolute drift fluxes are
allowed to take before being considered "perturbed"
PriLev: a integer either 0, 1, or 2 specifying how much output CPLEX
should print
cpxControl: []

outputs
a MATLAB struct scanTYR-ROOM with fields:
fluxes: the optimal fluxes (may not be unique)
lambda: the "shadow prices" of metabolites
red.cost: the reduced costs of fluxes
mu: the optimal cost, i.e., the minimum number of
significanlty perturbed genes
result: a flag which reports termination condition of the cplex
solver; is 1 if an optimum is found
basis: an n*m-length vector containing the basis of the found
optimum

function TYRvarROOM = scanTYRROOM(wt, Smat, boundsL, boundsU, zeroRHS,
forceOUTgn, forceOUTfx, forceKO, del, eps, cpxControl, PriLev);

n = 1;
while n < length(forceOUTfx)
    TYRvarROOM(n) = solveROOM(wt, Smat, boundsL, boundsU, zeroRHS,
forceOUTgn, forceOUTfx(n), forceKO, del, eps, cpxControl, PriLev);
    temp = find([TYRvarROOM.result]==119);
    if length(temp)#0
        forceOUTfx(n)
        del = del + 0.01
eps = eps + 0.001
        n=1;
clear TYRvarROOM;
    else n = n+1;
end
end

D.5 MATLAB .m file for computing min/max flux values around a given flux state

%minmaxLP
%a function to take a feasible (assumed optimal) solution to an
%optimization problem, and return two vectors: one giving the smallest
%value of each solution vector element consistent with the given optimum
%*when each of the other elements are held constant*, and also the maximum
%
% inputs
% soln: an n-length feasible (assumed optimal) solution to an optimization
% problem with equality constraints S*nu = rhs and lower and upper
% bounds lb and ub
% Smat: an m*n matrix of coefficients for equality constraints on the
% optimization variables (nu)
% rhs: the right-hand side (often indicated b) of the equality
% constraints S*nu = rhs
% lb: an n-vector whose elements bound the corresponding optimization
% variables, from below
% ub: an n-vector whose elements bound the corresponding optimization
% variables, from above
% obj: an n-vector comprising the objective function
% (it points to optimization variables that
% are (assumed to be) optimized in the calculation of soln.
% variables that obj points to will not be minmaxed

% outputs
% minsoln an n-vector whose ith element bounds from below the range
% of values of nu(i) consistent with obj*nu = opt
% maxsoln an n-vector whose ith element bounds from above the range
% of values of nu(i) consistent with with obj*nu = opt

function [minsoln, maxsoln, minsolnInform, maxsolnInform, nomin, nomax] = 
    minmaxLP(soln, obj, Smat, rhs, lb, ub)

n = length(soln);
m = n;
min_obj = zeros(n,1);
max_obj = zeros(n,1);
minsoln = zeros(m,1);
maxsoln = zeros(m,1);

minsolnInform = zeros(m,1);
maxsolnInform = zeros(m,1);

opt = -obj'*soln;
tol = 1e-5;

if opt > 0
    opttol.lb = (1-tol)*opt;
    opttol.ub = (1+tol)*opt;
else
    opttol.lb = (1+tol)*opt;
    opttol.ub = (1-tol)*opt;
end
cpxControl.EPRHS = 3e-5  ;

for ind=1:m
    min_obj(ind) = 1;
    max_obj(ind) = -1;

    [nu.min, slack, lambda, red.costs, mu, n.inf, s.inf, minInform, basis] =
        cplex(min_obj, [Smat;-obj'], lb, ub, [rhs;opt], [], cpxControl);
    %[rhs;opttol.lb],[rhs;opttol.ub]
    [nu.max, slack, lambda, red.costs, mu, n.inf, s.inf, maxInform, basis] =
        cplex(max_obj, [Smat;-obj'], lb, ub, [rhs;opt], [], cpxControl);

    minsoln(ind) = nu.min(ind);
    maxsoln(ind) = nu.max(ind);

    minsolnInform(ind) = minInform;
    maxsolnInform(ind) = maxInform;

    min_obj(ind) = 0;
    max_obj(ind) = 0;
end

nomin = find(minInform>1);
nomax = find(maxInform>1);

return

D.6 MATLAB .m file for making a 2-D projection of the phenotypic phase space

% makeProdGrowthEnvl.m
% a function to plot the S.nu convex hull in two dimensions: product output
% vs. growth rate
% inputs:
%   n number of data points
%   targ target product
% outputs:
%   prodXTo n vector of product fluxes
%   mu n vector of growth rates
%   workedQ n vector whose nth element is 1 if the optimization at the
%   nth pt worked
function [prodXTo,mu,workedQ,yield] =
    makeProdGrowthEnvl(n, targ, obj, S, lb, ub, zeroRHS)
\texttt{a = linspace(0,1,n);}

\texttt{\% calc max yield}
\texttt{objPROD=Obj;}
\texttt{objPROD(627)=0;}
\texttt{objPROD(targ)=-1;}

\texttt{yieldsoln = solveFBA(objPROD, S, lb, ub, zeroRHS,[],[],[]);}
\texttt{yield = yieldsoln.growth;}

\texttt{\% calc max growth}
\texttt{growthsoln = solveFBA(obj, S, lb, ub, zeroRHS,[],[],[]);}
\texttt{maxgrowth = growthsoln.growth}

\texttt{ub.loop = ub;}
\texttt{lb.loop= lb;}

\texttt{for n=1:length(a)}
\texttt{lb.loop(627) = a(n) * maxgrowth;}
\texttt{ub.loop(627) = a(n) * maxgrowth;}

\texttt{maxsoln = solveFBA(objPROD, S, lb.loop, ub.loop, zeroRHS,[],[],520);}
\texttt{minsoln = solveFBA(-objPROD, S, lb.loop, ub.loop, zeroRHS,[],[],520);}

\texttt{prodXTo((2*n-1):(2*n)) = [maxsoln.fluxes(targ) minsoln.fluxes(targ)];}
\texttt{mu((2*n-1):(2*n)) = [maxsoln.fluxes(627) minsoln.fluxes(627)];}
\texttt{workedQ(n) = maxsoln.result;}

\texttt{lb.loop = lb;}
\texttt{ub.loop = ub;}
\texttt{end}

\texttt{plot(mu,prodXTo,'.');}

\texttt{return}

\textbf{D.7 MATLAB .m file for plotting a multiple 2-D phenotypic phase spaces}

\texttt{%scanProdGrowthEnv1.m}
\texttt{\%a function to plot the S.nu convex hull in two dimensions: product output}
\texttt{\%vs. growth rate, for a number of koed genes}
\texttt{\%inputs:}
\texttt{\% n number of data points}

\texttt{219}
% targ target product
% ko_mat an m x n matrix of indices to ko. m different sims run,
% the n indices in each row point to positions to ko in each run
%outputs:
% prodXTo n vector of product fluxes
% mu n vector of growth rates
% workedQ n vector whose nth element is 1 if the optimization at the
% nth pt worked

def function [prodXTo, mu, workedQ, yield] = 
  scanProdGrowthEnvl(n, targ, ko_mat, obj, S, lb, ub, zeroRHS)

hold all;
ub.ko = ub;
lb.ko = lb;

for ind = 1:length(ko_mat');

  [foo, goo, kos] = find(ko_mat(ind,:));
  ub.ko(kos) = 0;
  lb.ko(kos) = 0;
  [prodXTo, mu, workedQ, yield] = 
    makeProdGrowthEnvl(n, targ, ko_mat, obj, S, lb.ko, ub.ko, zeroRHS);
  ub.ko = ub;
  lb.ko = lb;
clear kos;
clear foo;
clear goo;
end

hold off;
return