Model Guided Engineering of *Escherichia coli* for Chemical Production

By

Xiaolin Zhang

A dissertation submitted in partial fulfillment of

the requirements for the degree of

Doctor of Philosophy

(Cellular and Molecular Biology)

at the

UNIVERSITY OF WISCONSIN-MADISON

2014

Date of final oral examination: 8/12/2014

This dissertation is approved by the following members of the Final Oral Committee:

Jennifer L. Reed, Associate Professor, Chemical and Biological Engineering

Timothy J. Donohue, Professor, Bacteriology

Brian F. Pfleger, Associate Professor, Chemical and Biological Engineering

Edward G. Ruby, Professor, Medical Microbiology and Immunology

John Yin, Professor, Chemical and Biological Engineering

Acknowledgment

First and foremost, I would like to thank my advisor, Professor Jennifer Reed, for having me in this research group, for her valuable guidance over these years, for the weekly meeting to keep me on track, for being patient with me, and for teaching me everything about systems metabolic engineering. I like to have freedom to do experiments and try my own ideas. I thank her for giving me such freedom, and when things do not work, giving me encouragements. I have been somewhere between biology and engineering over these years and take different classes, research-related or unrelated. Thanks for being supportive and encourage me to learn different things. I am also grateful for helping me apply for fellowships and supporting me to attend conferences almost every year. I really enjoy the experience of going outside to talk with researchers from academia and industry.

Secondly, I would like to thank the graduate students and former and current postdocs in the group for providing constructive suggestions about research and advices at my presentations. I would particularly like to thank Joonhoon Kim who has the patience to answer my questions for six years, teaches me optimization and debugs my code. Thank David Baumler for helping me with experiments and inviting me to the lake house. I will never forget the experience of water tubing. I also appreciate the works that have been done by Chris Tervo and Josh Hamilton. Without your previous works, I could not finish my projects. I also want to thank Camo Cotton, Wai Kit Ong, and Matt Long for their feedbacks about my presentations and Paul Adamczyk for teaching me new experimental techniques and taking care of the lab. I have also worked with undergraduates: Mink Arunrattanamook, Sanjan Tp, Yifu Chen. Thank them for the hard work.

I would also like to express my thanks toward my thesis committee members: Timothy Donohue, Brian Pfleger, Edward Ruby and John Yin. Thank you for finding time to meet me every year and providing critical comments and insights which are extremely valuable to my work.

My research projects are funded by BACTER fellowship and KECK foundation. I greatly appreciate the financial support. I also like the fun brain storm discussion in the KECK group.

Over these years, I obtained many experimental materials provided by Rebecca Lennen, Mark Politz, Matthew Begemann, Tyler Youngquist, Matthew Copeland and Wenjing Zhao. I also got tremendous help from GLBRC HPLC facility, particularly Mick Mcgee and Brendan Thomson. Thank them for providing help for my research.

Finally, thank my parents for supporting me to study abroad. Thank Qi for taking care of me and always being there for me. Thank my whole family for their endless love and support. Without you, I could not live happily.

Table of Contents

List of Tables

List of Figures

Chapter 1 Metabolic engineering for chemical productions at a system level

1.1 Systems metabolic engineering

Metabolic engineering emerged as an independent discipline over a decade ago. In the first book of this field, metabolic engineering was defined as "directed improvement of product formation or cellular properties through the modification of specific biochemical reactions or the introduction of new genes with the use of recombinant DNA technology"[1]. While recombinant DNA technology is still widely used to introduce genetic changes, many advanced techniques have been developed which have expedited our understanding of bacterial metabolism and strain development. Engineering via evolution has often been used to improve product yield or select desired cellular phenotype (i.e. stress-tolerant strains) [2]. The data collected from high-throughput measurement technologies (transcriptomics, proteomics, metabolomics) have been used for strain engineering by deciphering cellular control mechanisms, identifying metabolic bottlenecks and identifying gene targets to improve strain performance [3,4]. Novel genetic tools facilitated by synthetic biology have allowed researchers to design and precisely control metabolic pathways and cellular phenotypes [5,6]. The quantity and scale of omics data sets has led to the development of computational analysis tools. One of the most influential computational tools in metabolic engineering is genome-scale metabolic models. These models can efficiently predict cellular phenotypes and suggest beneficial gene targets to improve chemical production [7,8]. Genome-scale metabolic models have guided most of the experiments described in this dissertation. Metabolic engineering has traditionally

focused on studying the properties of individual metabolic pathways, enzyme and metabolite interactions and genetic regulatory circuits. Computational models allow one to predict how local pathway changes may affect global network properties, and vice versa. The development of these new computational and experimental techniques provides unprecedented opportunities in metabolic engineering to improve product yield and cellular performance at a systems-level.

1.2 Branched-chain amino acids and biofuels

L-valine, L-leucine and L-isoleucine are three branched-chain amino acids (BAA). They have been used for various applications including ingredients for cosmetics and pharmaceuticals, additives in diet supplements and animal feeds [3], and building blocks for the synthesis of herbicides [9]. Increasing attention to branched-chain amino acids is due to the discovery of how the BAA biosynthetic pathways can be used to produce α-keto acid intermediates which can be converted into branched-chain alcohols (e.g. 1-propanol, isobutanol, 1-butanol, 2-methyl-1-butanol) [10]. Considered as substitutes for gasoline, branched-chain alcohols have significant advantages over ethanol, with their higher energy content, lower volatility and decreased hygroscopicity. BAA producing strains with high yields could possibly meet demands not only for industrial applications, but also for the increasing global energy consumption. The branched-chain amino acid producing strains have been generated using *Brevibacterium lactofermentum* and *Corynebacterium glutamicum* mutants [9,11-13]. An efficient L-valine production strain in *E. coli* was also developed [3]. In this study, we evaluated two different approaches to develop *E. coli* strains for BAA production by

combining directed evolution and gene knockout simulations using genome-scale metabolic models.

1.3 Summary for this work

The first approach we used was to construct and evolve a synthetic mutualistic community consisting of leucine and lysine auxotrophs of *E. coli*. The two mutants can only grow in glucose minimal medium when they exchange leucine (or its precursors) and lysine. We hypothesized that the lysine auxotroph would be forced to produce extra leucine and secrete it into medium if the co-culture improved its growth rate. Although the evolved lysine auxotroph turned out not to secrete leucine when grown in isolation, the establishment and evolution of the co-culture system and the population dynamics were studied in details described in Chapter 2. We found a viable co-culture using these two auxotrophs could be established and adaptively evolved to increase growth rates (by ~3 fold) and optical densities (by ~2.5 fold). While evolved isolates had increased fitness in co-culture, they exhibited significantly decreased fitness in mono-culture (when supplemented with leucine or lysine). When the estimated leucine and lysine secretion rates were used, the model simulation results agreed well with experimental data and suggested a better exchange of leucine (or its precursors) and/or lysine improved the growth phenotype of the co-cultures. This study will contribute more knowledge for future utilization of co-culture in metabolic engineering.

The second approach was to first build a pyruvate (the BAA precursor) producing strain with high yields, and then produce BAA in the pyruvate strains by increasing the BAA synthesis fluxes. In order to maximize the pyruvate production yield, we designed gene deletion strains

using a genome-scale metabolic model of *E. coli*. The computationally designed mutant strains were constructed, characterized and one of them achieved yields of more than 0.92 g pyruvate per g of glucose (95% of theoretical yield) under aerobic conditions, which was more than the yield of the strain previously reported: 0.76 g of pyruvate per g of glucose (78% of theoretical yield) [14]. This project was described in Chapter 3. The BAA production strains based on the pyruvate producing strains will be developed in near future.

Pyruvate is a precursor to BAA and many other fine chemicals, commodity chemicals and high-value pharmaceutical ingredients. The pyruvate overproducing strains could be used as platform strains to produce other important chemicals. In Chapter 3, we demonstrated the usage of pyruvate strains by genetically modifying them to produce ethanol. The re-engineered strains achieved yields up to 0.43 g ethanol per g of glucose (~85% of theoretical yield) in batch fermentation. In Chapter 4, we computationally explored what additional chemicals could be derived from pyruvate. A comprehensive investigation of the production capabilities of *E. coli* as a background host to produce non-native products was performed. We found 1,793 non-native products could be produced in *E. coli*, confirmed that 284 of them had commercial value and identified 64 valuable non-native products within 5 reaction steps of pyruvate. The 64 valuable non-native products would be good targets for re-engineering the pyruvate strains (described in Chapter 3) to produce other important chemicals.

Chapter 2: Adaptive evolution of synthetic mutualistic communities improves growth performance

2.1 Introduction

Microbes are affected by their physical and chemical environment, and they naturally encounter other species that can also influence their behaviors. Symbiotic interactions between microbes and higher organisms can lead to stable interactions and microbial communities. Mutualism is one type of symbiotic interaction, where both species benefit from the interaction. The existence of cooperation between members of a community appears to violate evolutionary theory that natural selection favors selfish behaviors, and therefore different theories have been proposed to explain how cooperation arises and evolves [15-19]. While symbiotic interactions are important, most of our knowledge of bacterial metabolism has been gathered from studies of individual strains in pure cultures. However, more than 99 percent of microbes cannot be cultured nor studied in mono-culture, since their growth depends on the presence of other species [20]. Additionally, the phenotypes of cultivatable strains may drastically change when grown in a mixed community as compared to mono-culture [21,22]. Therefore, studies are needed on how bacterial metabolism is influenced by interactions with other organisms.

In the last decade, experimental efforts have been made to build and study controlled multispecies systems [23-26]. Hosada et al. used amino acid *Escherichia coli* auxotrophs to investigate requirements for nascent mutualism, including how initial cell concentrations affect

co-culture dynamics [25]. Wintermute and Silver screened 1,035 combinations of *E. coli* auxotrophs to identify pairs of strains that could grow in co-culture and estimated cooperation levels and costs associated with cooperation between strains grown in co-culture [24]. Kerner et al. created a tunable system using tyrosine and tryptophan *E. coli* auxotrophs containing inducible genetic circuits that control production of tyrosine and tryptophan, and thus growth rates and strain ratios [23]. Recently, Pande et al. studied co-cultures of cross-feeding *E. coli* mutants which consumed (due to an amino acid auxotrophy) and produced amino acids. Surprisingly, they showed that most co-cultures with cross-feeders had faster growth rates than the wild-type strains and were stable in the presence of non-cooperators [26].

While these studies investigated initial stages of mutualism in co-culture, other studies have also investigated how adaptive evolution alters community behaviors. Harcombe used co-cultures of a methionine *E. coli* auxotroph and a methionine secreting *Salmonella typhimurium* to select for improved methionine secretion [27]. Harcombe showed that adaptive evolution of co-cultures, made up of three strains, selected for cooperators (methionine secreting *S. typhimurium*) over non-cooperators (wild-type *S. typhimurium*) and that loss of spatial structure (by using flasks rather than agar plates) led to a loss of cooperators over time. Hillesland et al. adaptively evolved co-cultures of a sulfate reducing bacterium and a methanogenic archae and found growth rates and biomass yields improved significantly (by 80% and 30%, respectively). When evolved populations were co-cultured with their ancestral partner, antagonistic interactions were found between the two evolved populations [28].

Mathematical models have also been used to explore natural and synthetic co-cultures of microbes. Using parameters measured in co-cultures of two auxotrophic yeast strains, Shou

et al. delineated requirements for initial cell densities and cell numbers needed to achieve an initial viable co-culture [29]. Bull and Harcombe used a model of two cross-feeding species to show how population dynamics affected the fitness of the microbial community [30]. Constraint-based metabolic models have also been used to study natural and synthetic microbial communities. These models have be used to identify strains capable of cooperating [24] [26], predict intra- and extra-cellular flux distributions in co-cultures [31,32], and evaluate which co-culture objective (e.g. individual or community growth) best matches experimental data[31,33].

The idea of using microbial consortia to solve multiple tasks in complex environments has also drawn tremendous attention [34-36], and successful examples have illustrated the use of consortia for industrial applications [37-39]. In addition to these studies, many new tools have been developed to create novel microbial cross-feeding interactions, structured consortia, as well as, quorum-sensing communication. Creating novel interacting systems allows hypotheses to be tested and reveals ecological principles [40].

Despite these promising findings, the study of microbial consortia has just recently begun and many questions remain. How do species first establish a mutualistic community? Does cooperation persist during evolution? When does community or strain fitness increase and what mechanisms drive such improvements? How does the population structure change over time? How do phenotypes of individual strains change during evolution?

To answer these questions, we constructed a synthetic mutualistic community of two auxotrophic *E. coli* mutant strains to study how adaptive evolution influences community phenotypes and structure, as well as, individual strain behaviors. In our synthetic community,

strain L (which is unable to catalyze an intermediate step in leucine biosynthesis) and strain K (which is unable to catalyze the last step in lysine biosynthesis) can only grow in glucose minimal medium if they exchange leucine (or its precursors) and lysine. The community was adaptively evolved and its growth rate improved by almost three-fold. Monitoring the population dynamics during evolution showed a decrease in the ratio of lysine to leucine auxotrophs over time. Isolates from evolved co-cultures showed improved growth when co-cultured with their un-evolved partner strain compared to the un-evolved K and L co-culture. We additionally used a genome-scale metabolic model of the co-culture to investigate how uptake and release of essential amino acids would influence co-culture fitness and composition, and suggest mechanisms for observed adaptive evolutionary changes. This study provides insights into the evolution of mutualistic communities and how microbial phenotypes are altered during adaptive evolution in a co-culture environment. In addition, this study for the first time investigates how individual isolates in the evolved community influence community fitness and composition.

2.2 Materials and methods

2.2.1 Strains and plasmids

E. coli BW25113 and the plasmids pKD46, pCP20, and pKD13 were obtained from *E. coli* genetic stock center. The pKD3 plasmid was provided by Dr. Brian Pfleger (UW, Madison). The *E. coli* knockout strains Δ*leuA::kan*, Δ*lysA::kan*, and Δ*recA::kan* mutants were obtained from the Keio collection (Open Biosystems). An *E. coli* BW25113 Δ*recA::cat* strain was constructed using a PCR-based method [41]. A PCR product was generated that contains the chloramphenicol

resistance cassette (*cat*) from pKD3 and has homology to the upstream and downstream sequences of *recA*. The following primers were used in the PCR reaction with pKD3 as a template, 5'-ATGCGACCCTTGTGTATCAAACAAGACGATTAAAAATCTTCGTTAGTTTCGTGTAGGCTG GAGCTGCTTC-3' and 5'-CAGAACATATTGACTATCCGGTATTACCCGGCATGACAGGAGTAAAAATGC ATATGAATATCCTCCTTAG-3'. *E. coli* BW25113 containing pKD46 was transformed with the PCR product using electroporation. Cells were added into 1mL SOC medium (Fisher Scientific) with addition of 5mM L-arabinose, incubated at 37°C for 2 hours, and plated on a LB agar plate containing 34 μg/L chloramphenicol. To generate the double *E. coli* mutants used in the co-cultures, Δ*leuA recA::cat* (referred to as strain L since it is a leucine auxotroph) and Δ*lysA recA::kan* (referred to as strain K since it is a lysine auxotroph), the temperature-sensitive plasmid pCP20 was used to remove the *kan* gene from the BW25113 Δ*leuA::kan* and Δ*lysA::kan* mutants. The Δ*recA::kan* and Δ*recA::cat* mutations were then moved into these two kanamycin sensitive strains by P1 transduction [42] and selected on LB agar plates with kanamycin (50 μg/L) or chloramphenicol (34 μg/L).

2.2.2 Media and culture conditions

Most liquid co-cultures were grown at 37°C in M9 minimal medium (pH 7.0, 100 μM $CaCl_2$, 2 mM $MgSO_4$, 6.4 g/L $Na_2HPO_4 \bullet 7\ H_2O$, 1.5g/L KH_2PO_4, 0.25g/L NaCl, 0.5g/L NH_4Cl) supplemented with 2g/L glucose. For some mono-culture experiments, L-lysine or L-leucine was added into the medium at different concentrations. A concentration of 10 mg/L leucine or lysine was used for the un-evolved strains, since this allowed for significant growth while still ensuring that the amino acid was the limiting nutrient. Higher (16 mg/L) and lower (1.6 mg/L)

lysine and leucine concentrations were used to evaluate the evolved isolates in mono-culture, so that growth rates could be measured (for concentrations below 1 mg/L, the change in optical density (OD) during growth was too small to estimate growth rates accurately). For mono-culture experiments, cells were inoculated on LB agar plates with kanamycin (50 μg/L) or chloramphenicol (34 μg/L) for 24 hours and resuspended in glucose minimal media. The starting OD600 was 0.01 and 0.05 for un-evolved and evolved strains, respectively. For co-culture experiments, cells from frozen stock were first grown separately in glucose M9 minimal media with 10% (v/v) luria broth (LB) for 24 hours at 37°C, and then pelleted and washed twice using minimal medium without glucose, to remove any residual nutrients from the preculture. Strains were then combined into a co-culture in glucose minimal media.

2.2.3 Adaptive evolution

Multiple parallel co-cultures of K (lysine auxotroph) and L (leucine auxotroph) strains were each started with a 1:1 ratio based on OD600 values. Co-cultures were started with an initial OD600 of 0.0065 and were grown in 250 mL flasks containing 100 mL glucose minimal medium. Co-cultures were grown aerobically in a shaking incubator at 37°C. The OD600 of the co-culture was monitored and when it reached ~0.2 the co-culture was transferred to fresh media (resulting in an OD600 between 0.001 and 0.01) and 3 mL of culture was stored at -80°C. The growth rate of adaptively evolved co-cultures at each passage was approximated using the duration and the change in OD600 value of the passage. The percent of dead cells for the first 5 passages was determined using SYTOX Green nucleic Acid Stain (Molecular Probes, Invitrogen, cat. no. S7020). Frozen co-cultures were later recovered by growing them in 2 mL glucose

minimal medium and transferring them into 200 μL of fresh medium (such that the starting OD600 was 0.01) in a 96 well plate and grown at 37°C for 4 days. OD600 values were measured in a Tecan microplate reader and the changes in OD600 values and growth rates for the co-culture were calculated.

2.2.4 Mono-culture and hybrid co-culture of evolved strains

Evolved isolates from the frozen co-culture samples were obtained by selecting colonies from LB+kanamycin (50 μg/L) and LB+chloramphenicol (34 μg/L) agar plates. For mono-culture and hybrid co-culture experiments (consisting of evolved isolates [L^{ev} or K^{ev}] and their un-evolved partner strain [K or L]), evolved isolates were grown on LB+kanamycin (50 μg/L) or LB+chloramphenicol (34 μg/L) plates and a single colony was used to inoculate cells into glucose M9 minimal medium with (for mono-culture experiments) or without (for hybrid co-culture experiments) leucine or lysine. Mono-cultures and hybrid co-cultures were started with an initial OD600 of 0.05 and 0.01, respectively. Each evolved isolate mono-culture was repeated in triplicate in 384 well plates and grown for 48 hours at 37°C, where OD600 values were measured every 15 minutes. Growth rates were determined by searching for the maximum growth rate in a 3 hour window during exponential growth. A 3 hour window was used because this was less than the exponential growth period for the different cultures and it had enough data points (>10) to get a good estimate for the growth rate. Hybrid co-cultures containing a 1:1 mixture of evolved isolates and un-evolved K or L strains were carried out in 96 well plates. Hybrid co-cultures were grown in glucose medium at 37°C for 72 hours. A total of four

replicates were done, two each on different plates. The OD600 values were monitored every 4 to 6 hours and used to estimate the growth rates.

2.2.5 Concentration measurements

A bioassay was used to measure the concentration of amino acids. A standard curve for converting a change in OD600 values of strain K to lysine concentrations was generated by growing the K strain (*ΔlysA recA*::*kan*) to stationary phase in glucose minimal medium with various concentrations of lysine for 48 hours. The change in OD600 was proportional to the concentration of lysine, with a proportionality constant of 25.91 mg/L lysine per OD (**Figure 2.1B**). To measure the concentration of lysine in the culture medium, we passed the culture medium through a 0.2 μm nylon membrane to remove cells. The filtrate was then mixed with an equal volume of glucose minimal medium, inoculated with the K mutant and grown at 37°C for 48 hours. The concentration of lysine present in the filtrate was then estimated from the change in OD600 and the proportionality constant.

To estimate the levels of leucine, *Lactobacillus casei* 12A (provided by James L. Steele, UW Madison) was used as a leucine biosensor, since it is incapable of synthesizing leucine. A standard curve was generated using the same method described above, but the growth medium was comprised of equal parts by volume, 2 g/L glucose M9 minimal media with various concentrations of leucine and CDM medium without leucine [43]. The proportionality constant was 20.45 mg/L leucine per OD (**Figure 2.1A**). To quantify the amount of leucine in the culture medium the same procedure described above was used, except *L. casei* was used instead of strain K and the filtrate was mixed with an equal volume of CDM medium without leucine.

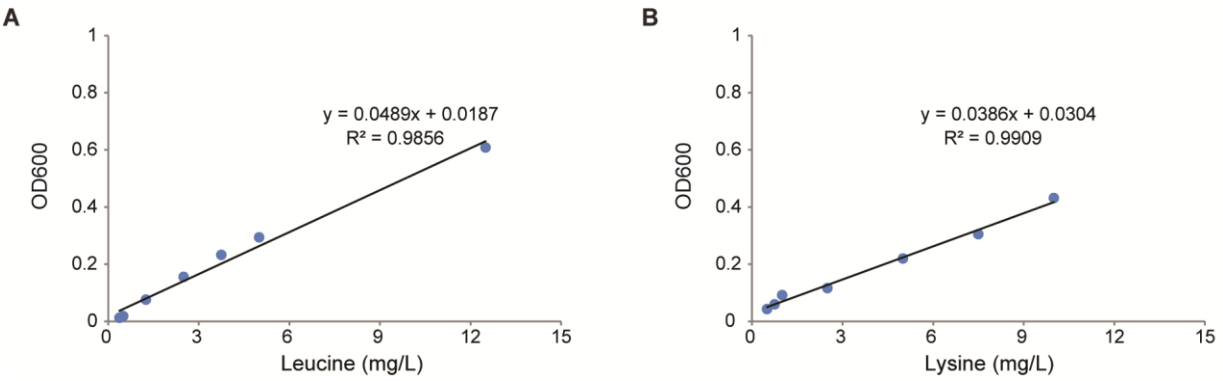


Figure 2.1. Standard curves for amino acid bioassays.
(A) The standard curve for leucine concentration versus OD600 values for *L. casei*. (B) The standard curve for lysine concentrations versus OD600 values for strain K. Tests were performed in 96 well plates and OD600 values were monitored every 15 minutes in microplate reader. The OD values were then converted to OD600 values in a standard spectrophotometer with 1cm pathlength using a standard curve (not shown). Each dot shows the amino acid concentration versus the converted OD600 value minus the OD600 value when the strains were grown in medium without amino acids.

The lower limit of detection for leucine and lysine that can be measured accurately using the bioassays was ~3.5 μM. One limitation of the bioassay is that the filtrate could contain chemicals that inhibit or enhance cell growth causing the bioassay to underestimate or overestimate the amino acid concentrations. To minimize the effects of other chemicals the filtrate was diluted two fold.

Glucose concentrations were measured using a glucose assay from Sigma (GAGO20) after cells were removed using a 0.2 μm nylon filter.

2.2.6 Estimation of growth and uptake rates

The growth rate and biomass requirements in mono-cultures were estimated from concentration measurements. First, the growth rate (μ) during exponential growth was calculated from the slope of a linear fit between ln(OD) and time (given by lnOD = μ·t+constant).

To estimate the biomass requirements (mmol substrate/gDW) for glucose, lysine, or leucine (Y_{Glc}, Y_{Lys} and Y_{Leu}), the OD600 values were converted to biomass concentration (g dry weight/L; gDW/L) using a conversion factor of 0.415 gDW/(L·OD) [44]. A linear regression between the substrate concentrations (glucose, leucine or lysine) and biomass concentrations (X_K or X_L) was performed, and the resulting slopes corresponded to the biomass requirements (e.g., [lysine] = $-Y_{Lys} \cdot X_K$+constant). Substrate uptake rates (mmol/gDW/hour) for glucose, lysine and leucine (U_{Glc}, U_{Lys} and U_{Leu}) in mono-cultures and co-cultures were then estimated by multiplying the measured biomass requirements by the growth rate (Equation 2.1). Release rates (mmol/gDW/hour) for lysine (R_{Lys}) and leucine (R_{Leu}) in co-cultures were estimated by equating the amount of amino acid produced by one strain to the amount consumed by the other strain (Equations 2.2 and 2.3).

$$U_i = Y_i \cdot \mu \qquad \forall\, i = \{Glc, Lys, Leu\} \tag{2.1}$$

$$R_{Lys} = U_{Lys} \cdot (X_K / X_L) \tag{2.2}$$

$$R_{Leu} = U_{Leu} / (X_K / X_L) \tag{2.3}$$

2.2.7 Quantifying relative populations in the co-culture

Standard plating methods measuring the colony forming units (CFUs) on LB, LB+kanamycin (50 μg/L) and LB+chloramphenicol (34 μg/L) agar plates were initially used to quantify the relative abundance of strains K and L in the co-culture. However, the adaptively evolved strains grew poorer on LB plates and the CFUs/(mL·OD) decreased (data not shown). Thus, we decided to use quantitative PCR (qPCR) to determine the relative abundance of the two populations in co-culture based on genomic DNA abundance rather than CFUs. The

genomic DNA of 500 μl of the frozen co-cultures was extracted using the Qiagen DNeasy Blood and Tissue Kit (cat. no. 69504). Fragments of the *kan* and *cat* genes were amplified from genomic DNA using qPCR with primers, qkan-L (5'-CTCGTCCTGCAGTTCATTCA-3'), qkan-R (5'-AGACAATCGGCTGCTCTGAT-3'), qcat-L (5'-CGTAATTCCGGATGAGCATT-3'), and qcat-R (5'-TCCGGCCTTTATTCACATTC-3'). Each 20 μL PCR reaction contained 10 μL SsoAdvanced SyBR Green supermix (Bio-Rad), 500 nM forward primers, 500 nM reverse primers and 20 ng genomic DNA. Each assay included triplicates for each co-culture, duplicate no DNA control, and positive controls of 0.1 ng, 1 ng, 10 ng, 100 ng of a 1:1 mixture containing genomic DNA from the parental K and L strains. The positive controls were used to generate a standard curve. The uncertainty for the estimated DNA concentration using the standard curve was calculated based on the error propagation method as following [45]:

Quantitative PCR requires using a standard curve (Equation 2.4) generated from samples with a known DNA concentration. For the standard curve, the log10 transformed DNA concentration is the independent variable (denoted as x), while the measured quantification cycle (C_q) is the dependent variable (y). Parameters m and b are determined from a linear least squares estimate. The standard deviations for y and m are denoted as s_y and s_m, respectively.

For a sample of unknown DNA concentration, its log 10 transformed concentration (x) can be estimated using the standard curve and the unknown sample's C_q number. The uncertainty in x (s_x) is calculated using Equations 2.5-2.7, where n is the number of data points used to generate the standard curve and k is the number of measurements for the unknown sample. The standard deviation for the unknown non-transformed DNA concentration (s_z) is calculated from s_x using Equation 2.8.

$$y = m \cdot x + b \tag{2.4}$$

$$s_m = \sqrt{\frac{\sum(y - b - m \cdot x)^2}{n - 2}} \cdot \frac{1}{\sqrt{\sum(x_i - x)^2}} \tag{2.5}$$

$$D = \frac{s_y^2}{s_m^2} \cdot n \tag{2.6}$$

$$s_x = \frac{s_y}{|m|} \sqrt{\frac{1}{k} + \frac{x^2 n}{D} + \frac{\sum(x_i^2)}{D} - \frac{2x\sqrt{x_i}}{D}} \tag{2.7}$$

$$s_z = s_x \cdot 10^x \cdot \ln 10 \tag{2.8}$$

2.2.8 Dynamic metabolic model of co-culture

A dynamic co-culture model was constructed which uses a stoichiometric matrix for each strain (based on the previously published stoichiometric matrix for the iAF1260 model [46]). The fluxes through reactions associated with the deleted genes in the K and L strain were constrained to be zero in the corresponding network. The amount of leucine or lysine consumed by one cell type in the co-culture was constrained to be equal to the amount released by the other cell type. The concentration of glucose, amino acids and biomass were calculated at 0.1 hour intervals using dynamic flux balance analysis (dFBA) [47]. At each time step in dFBA model, the metabolic model(s) were assumed to be at steady state and flux balance analysis (FBA) was used to predict the fluxes in that time step with the objective function of the combined growth rate of the two strains. Here, a FBA problem was formulated and solved to find the flux distributions in strain K ($\boldsymbol{v}^K$), strain L ($\boldsymbol{v}^L$), and media concentration rate of change ($\boldsymbol{v}^{Media}$). The units for the fluxes ($\boldsymbol{v}^K$ and $\boldsymbol{v}^L$) are mmol/gDW/hour (except for

the biomass flux which is 1/hour) and the units for the concentration rate changes ($\boldsymbol{v}^{Media}$) are mmol/L/hour. The FBA problem is shown below in Equations 2.9-2.17.

max $$\boldsymbol{c}^K \cdot \boldsymbol{v}^K + \boldsymbol{c}^L \cdot \boldsymbol{v}^L \quad (2.9)$$

such that $$\boldsymbol{S} \cdot \boldsymbol{v}^K = 0 \quad (2.10)$$

$$\boldsymbol{S} \cdot \boldsymbol{v}^L = 0 \quad (2.11)$$

$$LB^K \leq \boldsymbol{v}^K \leq UB^K \quad (2.12)$$

$$LB^L \leq \boldsymbol{v}^L \leq UB^L \quad (2.13)$$

$$v_i^{Media} = -\sum_j S_{i,j} \cdot \left(v_j^K \cdot x_t^K + v_j^L \cdot x_t^L\right) \quad \text{for } i \in I_{Ex} \text{ and } j \in J_{EX} \quad (2.14)$$

$$LB^{Media} \leq \boldsymbol{v}^{Media} \leq UB^{Media} \quad (2.15)$$

$$v_{DAPDC}^K, v_{IPPS}^L = 0 \quad (2.16)$$

$$v_{leu(e)}^{Media}, v_{lys(e)}^{Media} = 0 \quad (2.17)$$

Here $\boldsymbol{S}$ is the stoichiometric matrix reported in the iAF1260 model and contains all metabolic, transport and exchange reactions. Steady-state mass balance constraints are imposed for each strain (Equations 2.10 and 2.11). The lower bounds (LB) and upper bounds (UB) are used to constrain the flux distributions (Equations 2.12 and 2.13) and concentration rate changes (Equation 2.15). For metabolic and transport reactions the upper limits on fluxes were set to 1000 mmol/gDW/hour, and the lower limits were set to 0 or -1000 mmol/gDW/hour for irreversible and reversible reactions, respectively. The upper and lower limits for the ATPM (ATP maintenance) reaction were set to 8.39 mmol/gDW/hour. The upper and lower bounds for the set of exchange fluxes (set J_{EX}) and concentration rate changes used in the simulations are shown in **Table 2.1** in page 19.

The media concentration rate change ($\boldsymbol{v}^{Media}$) was calculated using Equation 2.14 from the exchange fluxes for each strain and the cell concentration of strain at time step t (x_t^K or x_t^L with units of gDW/L). Here J_{EX} is the set of exchange fluxes and I_{Ex} is the set of extracellular metabolites. Fluxes through reactions associated with LeuA in strain L and LysA in strain K were set to zero to reflect deletions in these two strains (Equation 2.16). The concentration rate changes for lysine and leucine were set to zero to ensure all the lysine produced by strain L is consumed by strain K, and leucine produced by stain K is consumed by strain L (Equation 2.17).

The objective function (Equation 2.9) in the FBA problem was first set to maximize the sum of the fluxes through the two biomass equations (by setting $c_{Biomass}^K$ and $c_{Biomass}^L$ equal to 1 and all other *c* values equal to 0). The biomass fluxes ($v_{Biomass}^K$ and $v_{Biomass}^L$) were then fixed to their optimal values and then the objective function was changed to minimize glucose uptake rates (by maximizing flux through glucose exchange reactions) needed to achieve these growth rates (by setting $c_{EX\_glc\_e}^K$ and $c_{EX\_glc\_e}^L$ equal to 1 and all other *c* values equal to 0).

These two optimizations were done for each time step (t) and the results used to calculate the glucose ($C_{t+1}^{Glucose}$) and cell (x_{t+1}^K and x_{t+1}^L) concentrations in the next time step (Equations 2.18-2.20) of the dFBA problem. This process was repeated until the total cell concentration reached 0.083 gDW/L (corresponding to an OD600 value of 0.2). At this point the K:L ratio and average biomass flux (i.e. growth rate) was calculated and used in **Figures 2.13** and **2.14**.

The ratio of the FBA predicted growth rates ($v_{Biomass}^K$ and $v_{Biomass}^L$) and glucose uptake rates ($-v_{EX\_glc(e)}^K$ and $-v_{EX\_glc(e)}^L$) is the the biomass yield and used in Equation 2.20. The

starting concentrations of glucose ($C_0^{Glucose}$) and each cell type (x_0^K and x_0^L) were set to 11.96 mmol/L and 0.0026975 gDW/L). The time step used (Δt) was 0.1 hour.

$$x_{t+1}^K = x_t^K \cdot e^{(v_{Biomass}^K \cdot \Delta t)} \quad (2.18)$$

$$x_{t+1}^L = x_t^L \cdot e^{(v_{Biomass}^L \cdot \Delta t)} \quad (2.19)$$

$$C_{t+1}^{Glucose} = C_t^{Glucose} - \frac{v_{EX\_glc(e)}^K}{v_{Biomass}^K}(x_t^K - x_{t+1}^K) - \frac{v_{EX\_glc(e)}^L}{v_{Biomass}^L}(x_t^L - x_{t+1}^L) \quad (2.20)$$

Table 2.1: Upper and lower bounds for fluxes.
(LB^K, UB^K, LB^L and UB^L with units of mmol/gDW/hour) and concentration rate changes (LB^{Media}, UB^{Media} with units of mmol/L/hour).

Metabolite	LB^K	UB^K	LB^L	UB^L	LB^{Media}	UB^{Media}
Glucose	-10	1000	-10	1000	-20	1000
Oxygen	-15	1000	-15	1000	-30	1000
Ions (ca2,cl,cobalt2,cu2,fe2,k, mg2,mn2,mobd,na1,tungs,zn2)	-1000	1000	-1000	1000	-2000	1000
fe3	-1000	0	-1000	0	-2000	0
Other (nh4,pi,so4,h,h2o,cbl1)	-1000	1000	-1000	1000	-2000	1000
Leucine	value[a]	value[a]	value[b]	value[b]	0	0
Lysine	value[a]	value[a]	value[b]	value[b]	0	0
All other external metabolites	0	1000	0	1000	0	1000

[a] for uptake rate simulations the LB and UB for lysine were set to negative the specified value and for leucine were set to 0 and 1000, respectively; for release rate simulations the LB and UB for leucine were set to positive the specified value and for lysine were set to -1000 and 1000, respectively.

[b] for uptake rate simulations the LB and UB for leucine were set to negative the specified value and for lysine were set to 0 and 1000, respectively; for release rate simulations the LB and UB for lysine were set to positive the specified value and for leucine were set to -1000 and 1000, respectively.

2.3 Results

2.3.1 Characterization of individual auxotrophs in mono-culture

Lysine (strain K) and leucine (strain L) *E. coli* auxotrophs were used in this work to study microbial interactions in co-culture. To reduce the chance of horizontal gene transfer between the two auxotrophs, we deleted *recA* from Δ*lysA* and Δ*leuA* mutants and replaced it with an antibiotic resistance marker to generate strain K (Δ*lysA recA*::*kan)* and strain L (Δ*leuA recA::cat*). Strain K requires lysine for growth, while strain L requires leucine. The additional deletion of *recA* did not reduce the mutant growth rates compared to the *recA* positive Δ*lysA* and Δ*leuA* mutants (in LB the growth rates were ~1.32 h^{-1} and ~ 1.36 h^{-1} for the *recA* negative and positive strains, respectively).

Both the K and L strains were characterized individually in mono-culture during growth in glucose minimal medium when supplemented with lysine and leucine, respectively. When grown in mono-culture where the essential amino acid (lysine or leucine) is limiting, the strains exhibited constant amino acid consumption rates and growth rates (**Figure 2.2A** and **2.2B**), which were estimated from the concentration data. In mono-culture, strains K and L had similar growth rates; however, strain K had a lower essential amino acid uptake rate than strain L, indicating that *E. coli* needs more leucine than lysine for biomass production. The amino acid requirements were also estimated from the biomass and concentration measurements (**Table 2.2**, see Methods for details), and they represent the amount of amino acid needed to produce

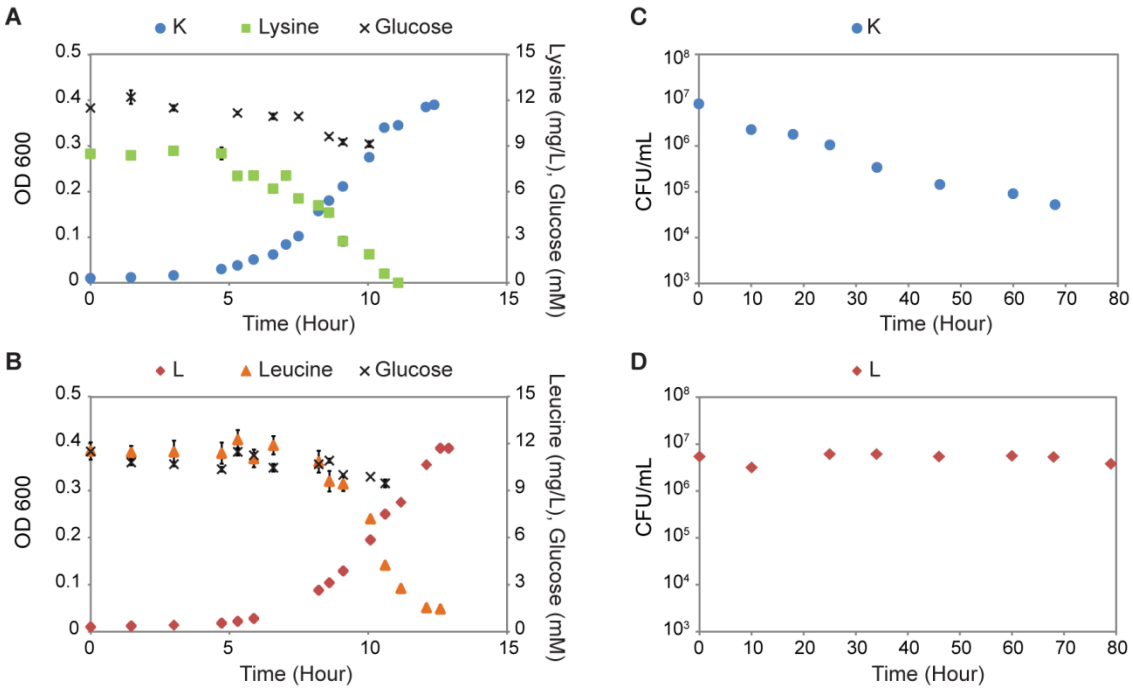


Figure 2.2: Characterization of mutant growth in mono-culture.
(A and B): Un-evolved K and L strains were grown in mono-culture in glucose minimal medium supplemented with 10 mg/L lysine or leucine, respectively. Concentrations of strain K (blue circles), strain L (red diamonds), leucine (orange triangles), lysine (green squares), and glucose (black x) in mono-cultures of strains K and L are shown. (C and D): Survival of strains K (panel C) and L (panel D), in mono-culture in glucose minimal medium without amino acid supplementation. The error bars represent standard deviations across three replicate measurements.

Table 2.2. Mutant phenotypes during growth in mono-culture.

	Δ*lysA* mutant estimated value	Δ*leuA* mutant estimated value
Growth rate (hour^{-1})	0.461	0.465
Amino acid requirement (mmol/gDW)*	0.350	0.473
Amino acid uptake rate (mmol/gDW/hour)ǂ	0.161	0.220

*The amino acid requirements represent the amount of leucine or lysine required for production of 1 gDW of cells.

ǂThe uptake rates are estimated as the product of the growth rate and amino acid requirements.

1 gDW of cells. Specifically, 0.350 mmol of lysine was needed for the formation of 1 gDW of strain K and 0.473 mmol leucine for 1 gDW of strain L. These values are close to the reported biomass composition of *E. coli* B/r, which contains 0.326 mmol lysine and 0.428 mmol leucine per gDW of cells [48]. Accordingly, if these strains have the same growth rate, the leucine uptake rate by L will be higher than the lysine uptake rate by K.

Since co-cultures of K and L would be carried out without supplementation of leucine and lysine, we also evaluated the survival of strains in mono-culture in glucose minimal medium without addition of amino acids. Cell viability was monitored over time by quantifying the number of colony forming units (CFU) per mL (**Figure 2.2C** and **2.2D**) and percent of dead cells using Sytox green nucleic acid stain (**Figure 2.3**). Interestingly, the two strains showed different resistances to starvation, which has been reported for other amino acid auxotrophs [25,29]. For the K strain, the number of CFUs/mL decreased within 10 hours, while the L strain did not show a large drop in CFUs/mL over 80 hours.

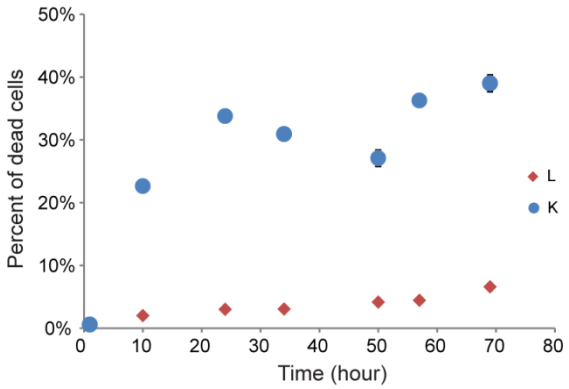


Figure 2.3. Survival of mutants in starvation.
The percent of dead cells in the mono-culture in minimal glucose medium without supply of amino acids was measured using Sytox green nucleic acid stain. The error bars indicate standard deviations.

2.3.2 Characterization of un-evolved co-cultures

We next explored the behavior of a co-culture of K and L when grown in glucose minimal media (without amino acid supplementation). Both mutants were inoculated with the same initial density in glucose minimal medium and growth of the co-culture was monitored over 70 hours. One surprising feature of the co-culture was that there was no lag phase at the beginning of the co-culture, even though the strains were precultured separately. The co-culture had an exponential growth rate equal to 0.056 h^{-1} (**Figure 2.4B**), which was around ~12% of the mono-culture growth rates (**Table 2.2** in page 21). The glucose uptake rate for the co-culture was estimated to be 2.42 mmol/gDW/hour. We also quantified the relative size of the two mutant populations at different time points of the co-culture by extracting genomic DNA and amplifying the *kan* and *cat* genes using qPCR. In the co-culture, the K and L strains proliferated at very similar rates with L growing slightly slower than K (**Figure 2.4C**). An equal mixture of K and L (based on OD600 values) corresponds to a K:L ratio (based on genomic DNA levels) of 1.59 ± 0.18, and the average K:L ratio determined by genomic DNA levels during exponential growth of the co-culture was 1.6 (**Figure 2.4D**). These results indicate that an exchange of leucine (or its precursors) and lysine happened immediately when the two mutants were grown together and was enough to support stable exponential growth.

2.3.3 Evolution of co-culture

To improve growth of the co-culture and establish a more cooperative artificial microbial community, we adaptively evolved the co-culture for short (one week) and exponential growth rate that was 88% lower than the strains grown in mono-culture

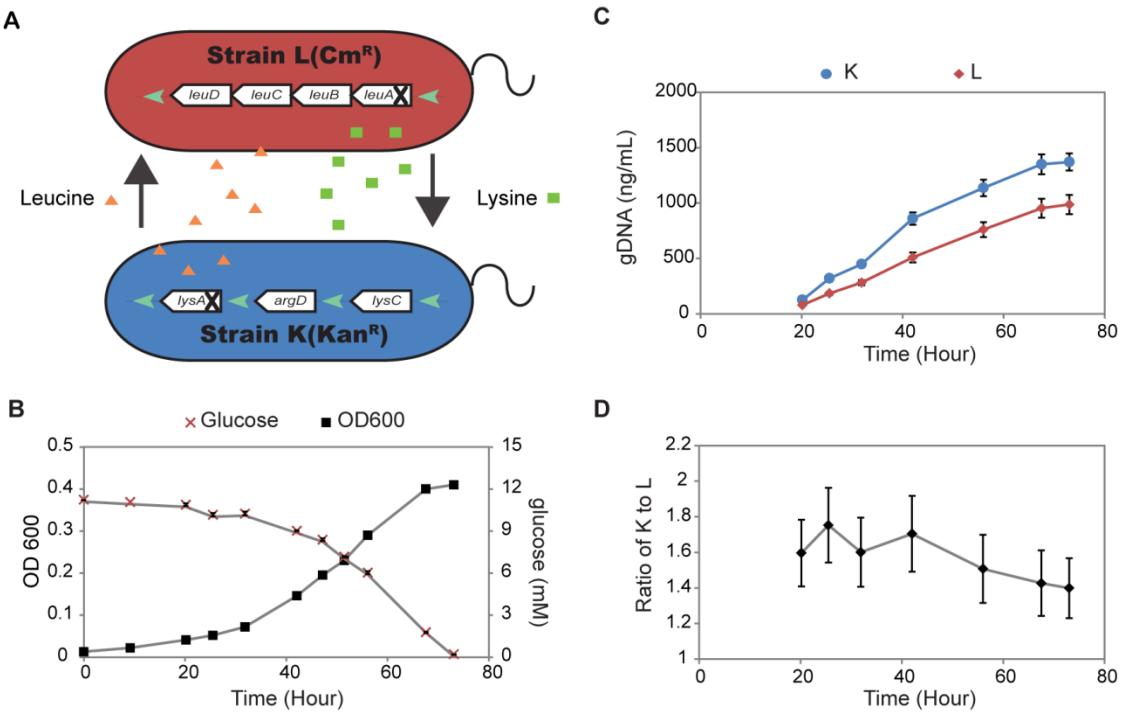


Figure 2.4: The un-evolved co-culture of strains K and L.
(A) Nutrient exchange and dependence in co-culture of two *E. coli* strains L and K. Strain L is incapable of synthesizing leucine, while strain K is unable to synthesize lysine. In the co-culture, if exchange of leucine and lysine occurs then both strains can grow in glucose minimal medium. Panel (B) shows the concentration profiles of glucose (red x) and optical density (black squares) during batch growth of the co-culture. The error bars indicate the standard deviations across replicates. (C) Genomic DNA from the two mutants were extracted from the co-culture at several time points during batch growth of the co-culture and analyzed by qPCR. Blue circles and red diamonds represent the K and L strains, respectively. The error bars were calculated by the error propagation method described in method. (D) The ratio of K to L was calculated from the qPCR results. The K to L ratio measured using qPCR was 1.59±0.18 for a 1:1 mixture of un-evolved cells based on OD600. The error bars indicate standard deviations.

supplemented with amino acids, so there was significant room for improving the fitness of the co-culture. We first adaptively evolved three replicate co-cultures for five passages starting with equal amounts (based on OD) of strains K and L in glucose minimal medium. The co-cultures were maintained in prolonged exponential growth by serially transferring cells into fresh medium, and the OD was monitored over the five passages (**Figure 2.5A**). In all three independent co-cultures, the growth rate was constant over the first two passages ($\mu \sim 0.05\ h^{-1}$)

and improved by 3-fold during the third passage and then stabilized (μ~0.14 h^{-1}**, Figure 2.5B**). Interestingly, just like in the initial co-culture the cells did not appear to have any lag phase during the later passages. The average percent of dead cells across the three co-cultures decreased over the first five passages (Spearman Rank Correlation, R^2 = 1, p=0.016), ranging between ~5% and ~2% at mid-exponential growth (**Figure 2.6**).

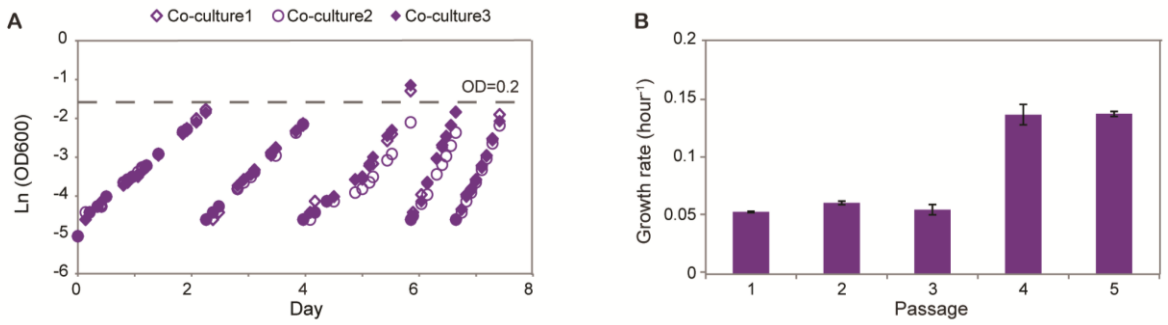


Figure 2.5. Changes in growth behavior over short adaptive evolutionary time frames.
(A) Natural log of the OD over time for three parallel co-cultures (co-cultures 1, 2, and 3) for five passages. Cultures were passed to fresh medium at OD600 ≈ 0.2 (marked as dashed line). The open diamond, open circle and solid diamond denote co-culture1, 2, and 3, respectively. (B) Average growth rates of each passage across the three independently evolved co-cultures. The error bars represent the standard deviations of the growth rate over the three co-cultures.

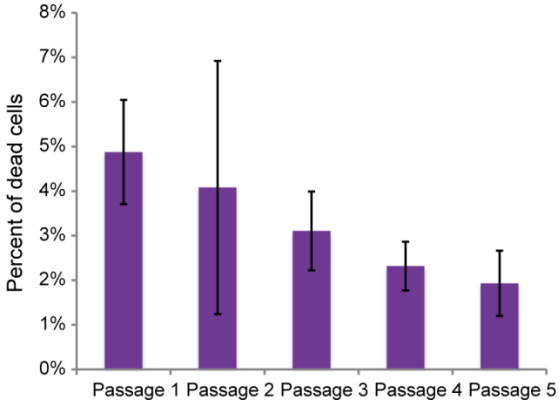


Figure 2.6. Percent of dead cells in co-cultures.
Percent of dead cells in five different passages from co-cultures 1, 2, 3 were measured when OD≈0.2. Each column represents the average percent of dead cells from the three co-cultures. The error bars indicate standard deviations.

After the short-term evolution experiments, we performed three parallel long-term adaptive evolutionary experiments of the co-culture using the same serial transfer process. The adaptive evolution lasted between 30 and 40 days, and included over 20 passages and 160 generations. Periodically, a small amount of co-culture was spread on LB agar plates and subsequently transferred to glucose, LB+kanamycin and LB+chloramphenicol agar plates, to check that one strain did not become independent of the other and take over the culture. For these co-cultures, we did not observe any isolates that were able to grow on glucose plates. The growth rate for each passage was estimated from the change in OD values and duration of each passage (**Figure 2.7A**). Similar to our short-term adaptive evolution results, the growth rate increased around day six in these independent co-culture experiments. After 10 days (5 passages), the growth rates oscillated around the same value. The three parallel co-cultures showed similar endpoint growth rates, which has been observed during evolution of individual

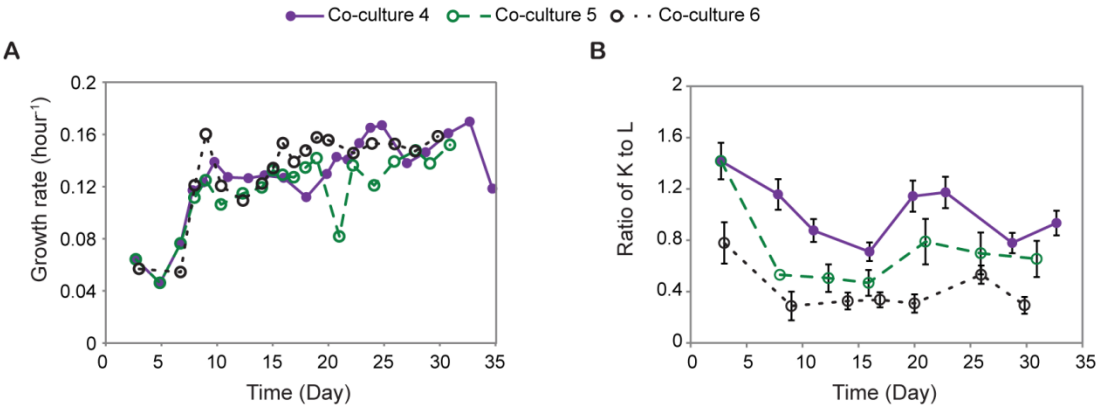


Figure 2.7: Adaptive evolution of the co-culture.
Three parallel co-cultures were performed, represented as a purple solid line (co-culture 4), a green short dashed line (co-culture 5), and a black dotted line (co-culture 6). (A) Growth rates were calculated based on the starting and ending OD values for each passage. (B) Genomic DNA was extracted from frozen samples of the co-culture taken at the end of each passage (OD≈ 0.2). Relative populations of K and L were estimated using qPCR and used to calculate the ratio of K to L. The error bars represent standard deviations calculated using the error propagation method.

strains [49]; however, the evolutionary trajectories of the co-cultures were different, indicated by different growth rates on the same day of evolution.

At the end of each passage, a sample of each co-culture was frozen and stored at -80°C. These frozen co-culture samples were later recovered and further evaluated to study the population dynamics of the co-culture and monitor the evolution of each strain. We monitored the relative abundance of the two strains over adaptive evolution. To estimate the relative abundance of the two strains at the end of each passage, genomic DNA from the frozen co-culture was extracted and qPCR was used to estimate the cell ratios (**Figure 2.7B**). The ratio of K to L decreased in all three evolved co-cultures and the final K:L ratios varied across the different parallel co-cultures between 0.93 and 0.29. The lower K:L ratio indicates that a smaller population of K cells can maintain a larger population of L strains. This could be due to a higher release rate of leucine (or its precursors) via secretion or cell lysis compared to lysine or a higher uptake rate of leucine compared to lysine. Since we did not detect any leucine or lysine in the co-culture medium, we cannot exclude either possibility.

While the growth rates of the co-culture were higher after evolution, it was unclear if the biomass yields of the evolved co-culture increased in the same fashion since the strains were transferred before reaching stationary phase. When frozen co-cultures were transferred directly into glucose minimal media, the frozen co-cultures tended to grow faster than the fresh co-cultures, which could be due to cell lysis caused by the freeze-thaw process. So we first recovered frozen co-cultures in glucose minimal medium and then passed the exponentially growing culture into fresh medium. Cells were then grown to stationary phase in a microplate reader, allowing growth rates and changes in OD600 values to be quantified (**Figure 2.8**).

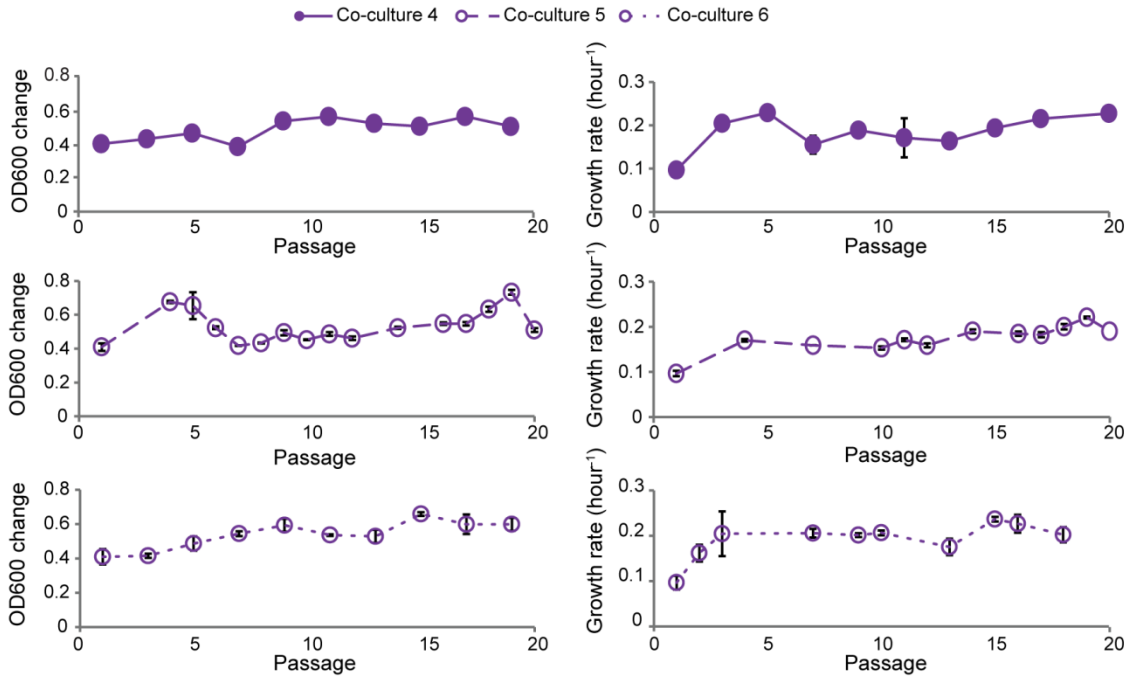


Figure 2.8. The change in OD600 and growth rates of evolved co-cultures 4, 5 and 6.
Frozen co-cultures were recovered and then grown in glucose minimal medium to stationary phase in 96 well plates. The change in OD600 shown in the figure was maximum OD600 value minus the initial value. The unevolved co-culture had a growth rate ~0.05 (hour^{-1}) and OD600 change ~ 0.26 in 96 well plates. Co-cultures 4, 5 and 6 are denoted by solid line, dashed line and dotted lines, respectively. The error bars indicate standard deviations.

In microplates, the change in OD600 of the un-evolved co-culture was 0.26 and all evolved co-cultures showed higher changes in OD600 than the un-evolved co-culture. The growth rate of the un-evolved co-culture was ~0.05 h^{-1} (similar to the value observed in flask experiments), while the growth rates of evolved co-cultures were 2- to 3-fold faster (**Figure 2.8)**.

2.3.4 Characterization of evolved strains in mono-culture

During adaptive evolution, co-cultures of K and L strains achieved higher growth rates and biomass yields. However, these experiments were done with a heterogeneous population and not using individual isolates. To further investigate how adaptive evolution affected individual strain behaviors we isolated strains from different passages of the co-culture and first

evaluated their growth in mono-culture. We randomly selected colonies of evolved K (or L) strains from different passages in co-culture 4 and 6 (since these co-cultures were the most different), and grew individual evolved isolates (K^{ev} and L^{ev}) in mono-culture in glucose minimal medium with different concentrations of lysine or leucine.

We first selected 3 colonies of K^{ev} (or L^{ev}) from late passages of co-cultures 4 and 6 and inoculated them in medium with increasing amounts of lysine or leucine. Surprisingly, the K^{ev} and L^{ev} strains had lower growth rates and changes in OD600 values compared to the un-evolved K and L strains, except for some K^{ev} isolates in co-culture 6 which had higher changes in OD600 values (**Figure 2.9**). We subsequently evaluated 10 isolates from different passages of co-cultures 4 and 6 for growth in mono-culture in the presence of high (16mg/L) and low (1.6 mg/L) concentrations of lysine or leucine. In general, we found that some isolates from earlier passages did show improved growth phenotypes in high and low concentrations of amino acids, but that most isolates from later passages had decreased growth rates and changes in OD600 values than the un-evolved K and L strains (**Figure 2.10**).

In addition to growing individual isolates in mono-culture with exogenous amino acid supplementation, we also evaluated the survival of isolates in mono-culture without exogenous lysine and leucine by measuring the percent of dead cells after 24 hours in glucose minimal medium. Compared to the un-evolved K strain, the evolved K^{ev} isolates from co-culture 4 and 6 had a lower percentage of dead cells (**Figure 2.11**). On the other hand, the evolved L^{ev} isolates from both co-cultures had a higher percentage of dead cells compared to the un-evolved L strain. These data indicate that possible mechanisms for improving growth of the co-culture could be due to a decreased viability of the L strain and/or increased viability of strain K.

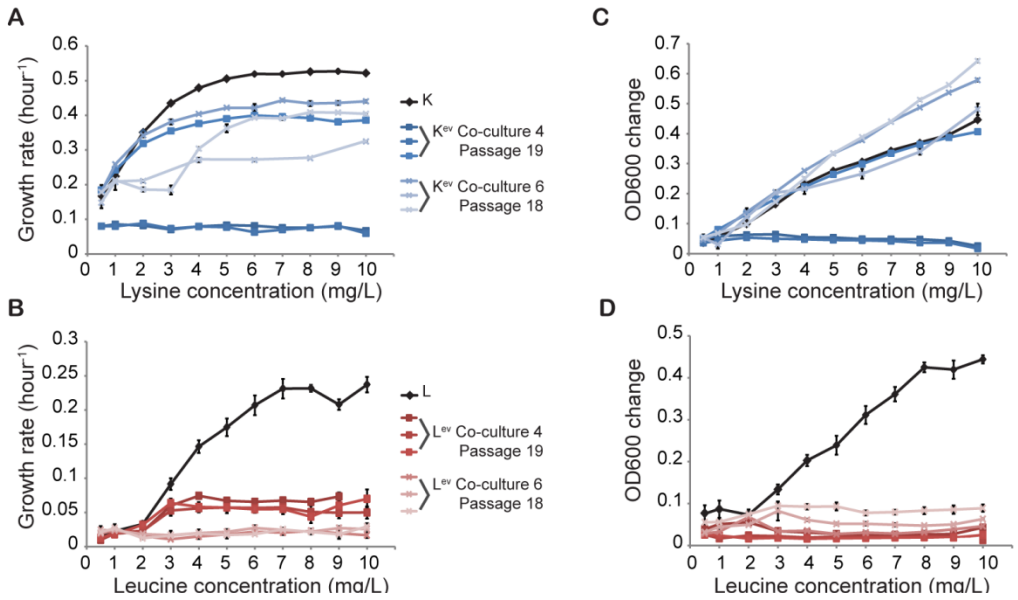


Figure 2.9: Mono-culture of K^{ev} and L^{ev}.
Three randomly selected colonies of K^{ev} (or L^{ev}) from passage 19 of co-culture 4 and passage 18 of co-culture 6 were inoculated in glucose minimal medium with various amounts of lysine (for K^{ev} strains, panel A and C) or leucine (for L^{ev} strains, panel B and D). Each colony was tested in three replicate mono-cultures. The growth rates and change in OD600 were calculated for the evolved and un-evolved parental strains (control). The error bars represent the standard deviations across biological replicates.

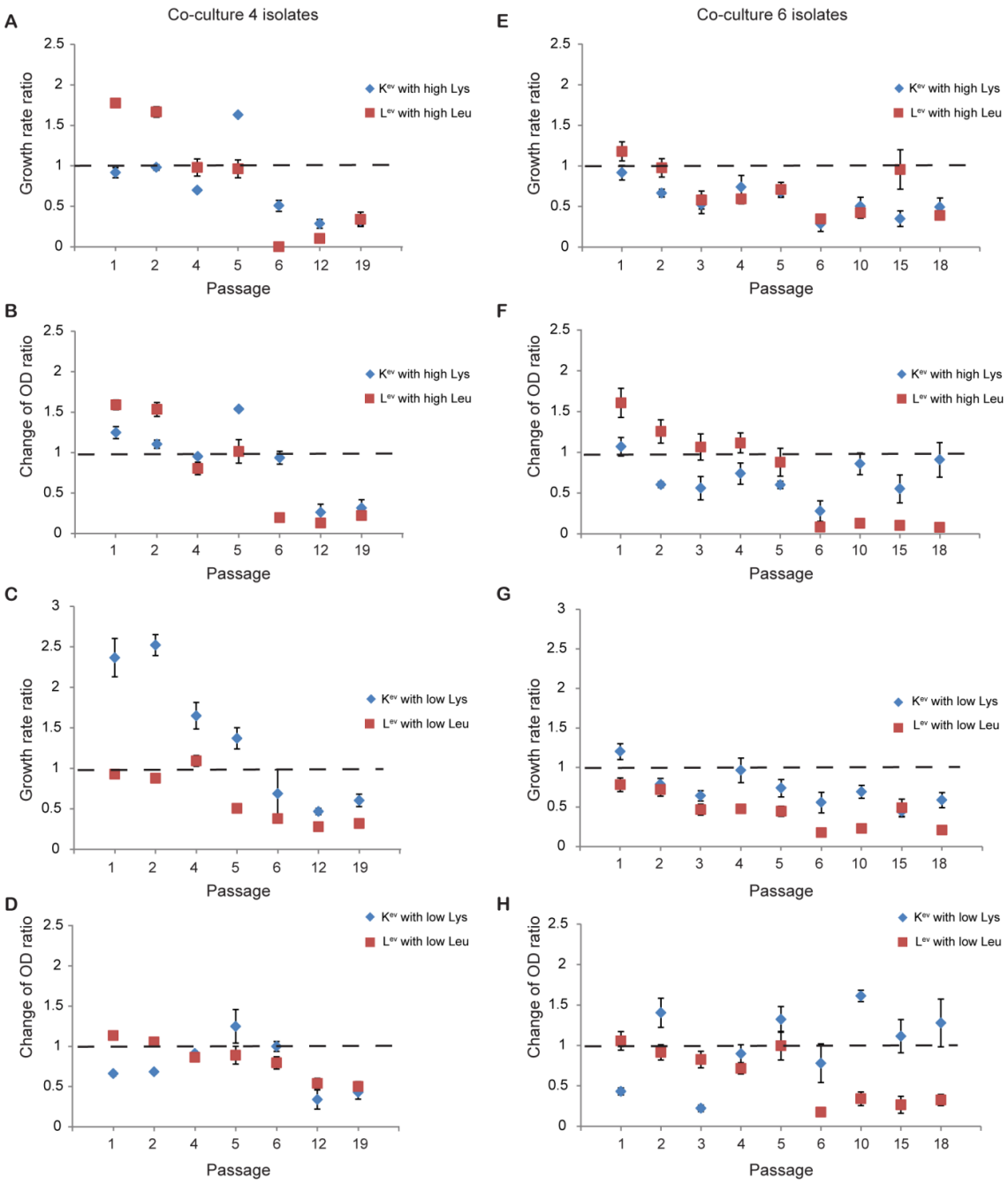


Figure 2.10. Growth phenotypes of K^{ev} and L^{ev} isolates grown in mono-culture with high and low levels of amino acids.

Ten randomly selected K^{ev} (or L^{ev}) isolates from different passages were inoculated in glucose minimal medium supplemented with Lys (or Leu) at high (16 mg/L, panels A, B, E and F) and low (1.6g/L, panels C,D, G and H) concentrations. Each isolate was tested in three replicate mono-cultures in 384 well plates. The growth rates and changes in OD600 were calculated for evolved isolates and normalized to the values for their respective un-evolved parental strains (K or L). The mean of the growth rate ratios and change in OD600 ratios are shown as blue diamond (K^{ev}/K) and red square (L^{ev}/L), respectively, in panels A-D (Co-culture 4) and E- H (Co-culture 6). The error bars indicate standard deviations.

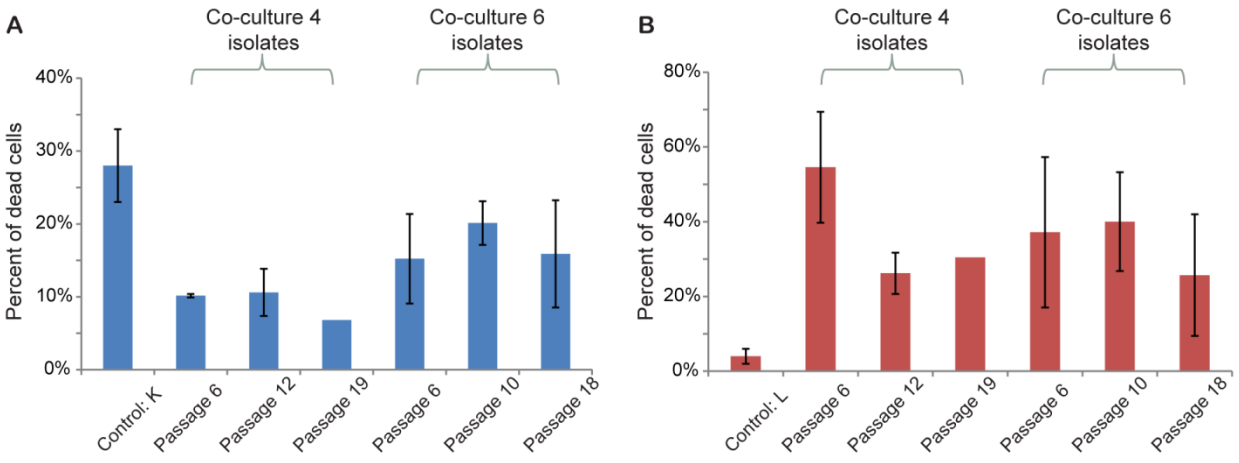


Figure 2.11. The percent of dead cells in mono-culture.
Three evolved K^{ev} (panel A) or L^{ev} (panel B) isolates from different passages were grown in mono-culture in glucose minimal medium. The number of dead cells was measured using Sytox green nucleic acid stain at 24 hour. The control is the percent of dead cells in mono-culture containing un-evolved K or L strains. The bar indicates the standard deviation across replicates.

2.3.5 Properties of evolved isolates in hybrid co-culture

Since the strains were evolved in co-culture and not mono-culture we also sought to evaluate changes in growth phenotypes of individual isolates when grown in co-culture with their un-evolved partner strains (referred to here as a hybrid co-culture). To find out how evolved isolates derived from each strain affect growth of the co-culture, we evaluated hybrid co-cultures containing evolved isolates (K^{ev} or L^{ev}) with their un-evolved partner strains (L or K) in glucose minimal media. The growth rates and biomass yields of L^{ev}+K (or L+K^{ev}) hybrid co-cultures were then compared to those of the initial un-evolved co-culture (L+K).

In co-culture 4, the growth rates of L+K^{ev} and L^{ev}+K hybrid co-cultures containing isolates from the first five passages were similar to the initial co-culture (K+L) (**Figure 2.12A**), while increased growth rates were observed in hybrid L+K^{ev} and L^{ev}+K co-cultures containing isolates from later passages. Growth rate improvements in the hybrid co-cultures were slightly delayed

compared to our earlier analysis of the evolved co-culture (**Figure 2.5B**), where the biggest growth rate improvements happened after three passages. This delayed improvement in growth rate could be due to the fact that only single evolved isolates were evaluated (rather than a mixed population) and that evolved isolates were tested in combination with un-evolved partner strains (rather than evolved partner strains). Compared to co-culture 4, isolates from co-culture 6 **(Figure 2.12D)** had larger variations across isolates from the same passage and earlier increases in growth rates. Interestingly, none of the evolved isolates co-cultured with their un-evolved partner strains led to a three-fold improvement in growth rate as observed in the evolved co-culture, indicating that synergistic effects between evolved isolates may exist in the co-culture. In both co-culture 4 and 6, the growth rate of L^{ev}+K hybrid co-cultures increased faster than the corresponding L+K^{ev} co-cultures, indicating that the L strains adapt more quickly to enhance co-culture growth. Hybrid co-cultures containing evolved isolates from co-culture 4 and 6 also exhibited higher biomass yields (measured by changes in OD600, **Figure 2.12B** and **2.12E**).

Since each hybrid co-culture contains at least one of the un-evolved parental strains, if the L+K^{ev} (or L^{ev}+K) co-culture grows better than L+K co-culture, then the evolved isolates likely have increased uptake and/or release of leucine (or its precursors) or lysine. An improved uptake rate would increase the abundance of the evolved strain in the co-culture while a higher release rate would benefit its partner strain. The ratio of K:L in the hybrid co-cultures during exponential growth was also measured using qPCR of genomic DNA, and compared to the K:L ratio in the un-evolved co-culture. These ratio measurements allowed us to find out which

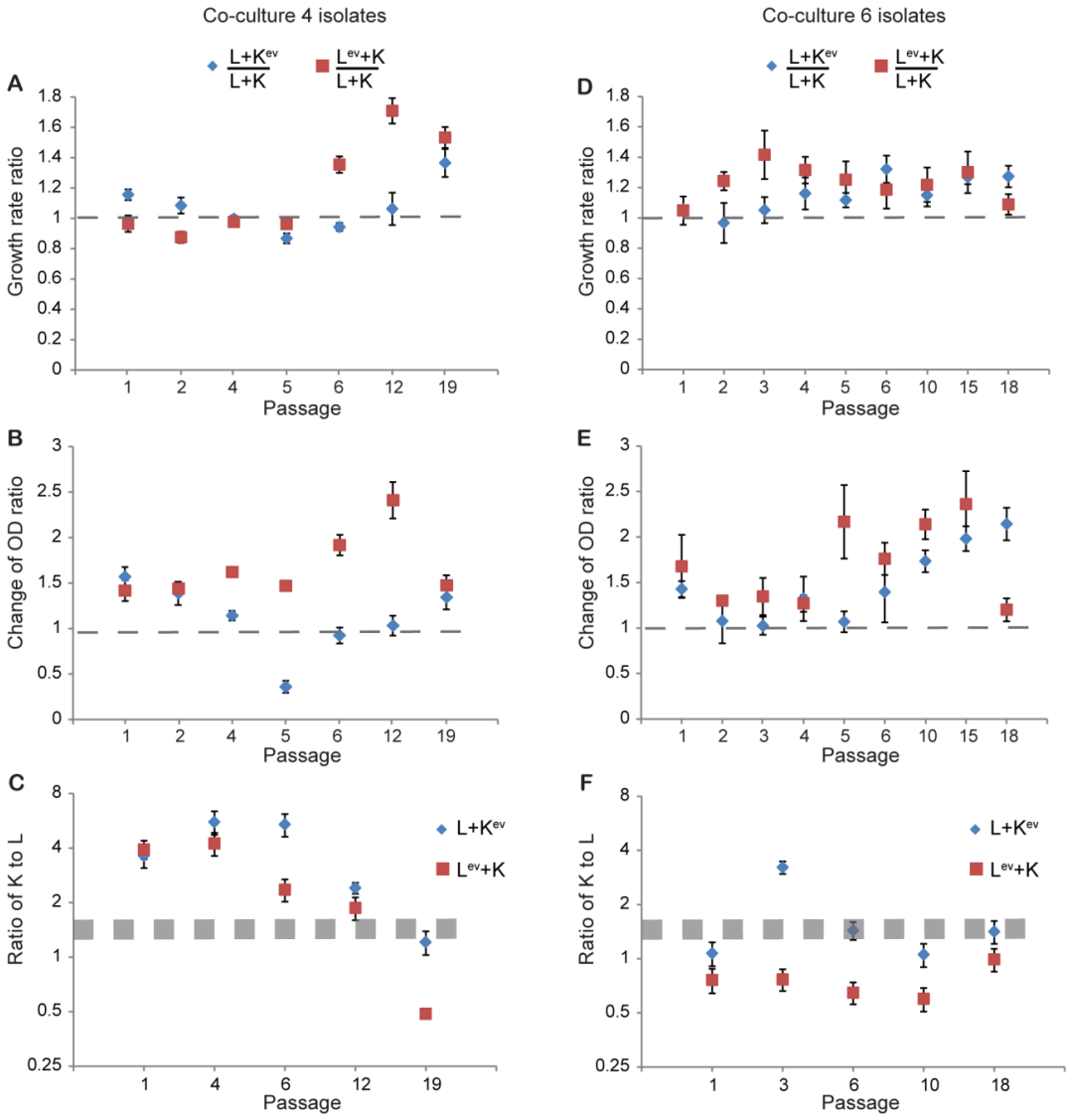


Figure 2.12: Comparisons between un-evolved co-cultures L+K and hybrid co-cultures containing L+K^{ev} or L^{ev}+K.

Cells from 10 colonies of K^{ev} (or L^{ev}) at each selected passage were grown individually in co-culture with the un-evolved partner strain (L or K). The growth rate and change in OD600 for each hybrid co-culture was normalized to the growth rate and change in OD600 of the un-evolved co-culture grown on the same microplate. The resulting growth rate ratios and change in OD600 ratios are shown as blue diamonds (L+K^{ev}) and red squares (L^{ev}+K), respectively, in panels A and B (isolates from co-culture 4) and panels D and E (isolates from co-culture 6). The error bars indicate the standard deviations based on 10 separate hybrid co-cultures each with four replicates (n=40). The dashed lines indicate the behavior of the un-evolved co-culture (L+K). Panels C and F shows the K:L ratio in L+K^{ev} and L^{ev}+K in hybrid co-cultures and the un-evolved co-culture. The hybrid co-cultures contained evolved isolates from co-culture 4 (panel C) or co-culture 6 (panel F). The error bars indicate the standard deviations based on hybrid co-cultures using three different isolates and three measurements for each passage (n=9). The shaded bands in C and F show the mean $\pm$ the standard deviation for the K:L ratio in the un-evolved co-culture at an OD600 of 0.2 when grown in 96 well plates (1.62 ± 0.14).

strain if any dominated the hybrid co-culture. Three hybrid co-cultures at five different passages were selected for this analysis. They represent the slowest, medium and fastest growing hybrid co-cultures within a given passage. For comparison, the co-culture of un-evolved strains (L+K) was measured and had a K:L ratio 1.62±0.14. With isolates from co-culture 6, the K:L ratio of L+K^{ev} hybrid co-cultures at mid-exponential growth were all less than 1.6 (except for passage 3) indicating that the K^{ev} strains improved growth of the L strain more than the original K strain (**Figure 2.12F**). In addition, the K:L ratios in L^{ev}+K hybrid co-cultures also showed a decreasing trend, implying that the L^{ev} strains became dominant in the hybrid co-cultures. These results suggest that the K^{ev} strains may increase release of leucine (or its precursors) and/or the L^{ev} strains increase uptake rates of leucine. The hybrid co-culture with isolates from co-culture 4 showed a very different pattern. The K:L ratio initially increased for both of L+K^{ev} and L^{ev}+K hybrid co-cultures compared to the L+K un-evolved co-culture, suggesting possible better exchange of lysine, while the K:L ratio decreased at later passages, suggesting a better exchange of leucine (or its precursors) (**Figure 2.12C**).

2.3.6 Simulation of batch co-cultures

A number of possible mechanisms associated with amino acid exchange could explain the improvements in growth of the co-culture over adaptive evolution. These include increased uptake or release rates of leucine (or its precursors) or lysine, or combinations of these. Direct measurements of cross-feeding rates could not be made, so metabolic modeling was used to gain additional insights. To further evaluate the co-culture evolution, we developed a computational model of the co-culture using a genome-scale metabolic model of *E. coli* [46].

Dynamic flux balance analysis (dFBA) simulations were performed where the uptake/release of leucine and lysine were varied and the growth rates and K:L ratios were predicted at an OD600 of 0.2. At each time step in dFBA, metabolism was assumed to be at a steady-state and a flux distribution maximizing the combined growth rate was found. Since we did not detect any leucine or lysine in the co-culture media, we additionally constrained the dFBA model to ensure that there was no net accumulation of leucine or lysine in the media.

As expected, changing the uptake and release rates of the essential amino acids affected the community composition and the average growth rate (**Figure 2.13**). The model predicted that higher uptake or release rates of lysine will result in a larger K:L ratio, while larger rates of leucine uptake or release will decrease the ratio (**Figure 2.13A** and **2.13B**). What we did not anticipate is that the strain ratio was predicted to be more sensitive to the uptake rates than release rates. The strain ratio ranged between 0.04 and 17.54 when consumption rates were constrained, compared to 0.26 and 2.5 when release rates were constrained to the same range of values. The growth rate of the co-culture was predicted to improve by increasing uptake and/or release of leucine or lysine (**Figure 2.13C** and **2.13D**).

A major obstacle in studying the co-culture is an inability to directly measure the real-time uptake and release rates of the exchanged amino acids. We estimated the uptake and release rates of leucine (or its precursors) and lysine for the evolved K (or L) strains using the measured growth rates, biomass requirements and K:L ratios (assuming amino acid requirements did not change, see Methods for details). These estimated uptake and release rates (**Table 2.3** in page 40) were used to project the evolutionary trajectories for co-cultures 4 (**Figure 2.13**) and co-cultures 6 (**Figure 2.14**). The estimated uptake and release rates of leucine

and lysine both increased in co-culture 4, while only leucine exchange increased dramatically in co-culture 6. Using the estimated uptake or release rates as inputs, the model was then used to predict the K:L ratio and average co-culture growth rate. The experimentally measured K:L ratios and growth rates in the evolved co-culture were highly correlated to model predictions when release rates were constrained, but not uptake rates (**Figure 2.13E**). Since the uptake rate was estimated by multiplying the growth rate with the lysine (or leucine) requirement per gDW cells (**Table 2.2**), constraining the uptake rates effectively constrains the model growth rates to be close to the measured values, resulting in a K:L ratio always close to 1.

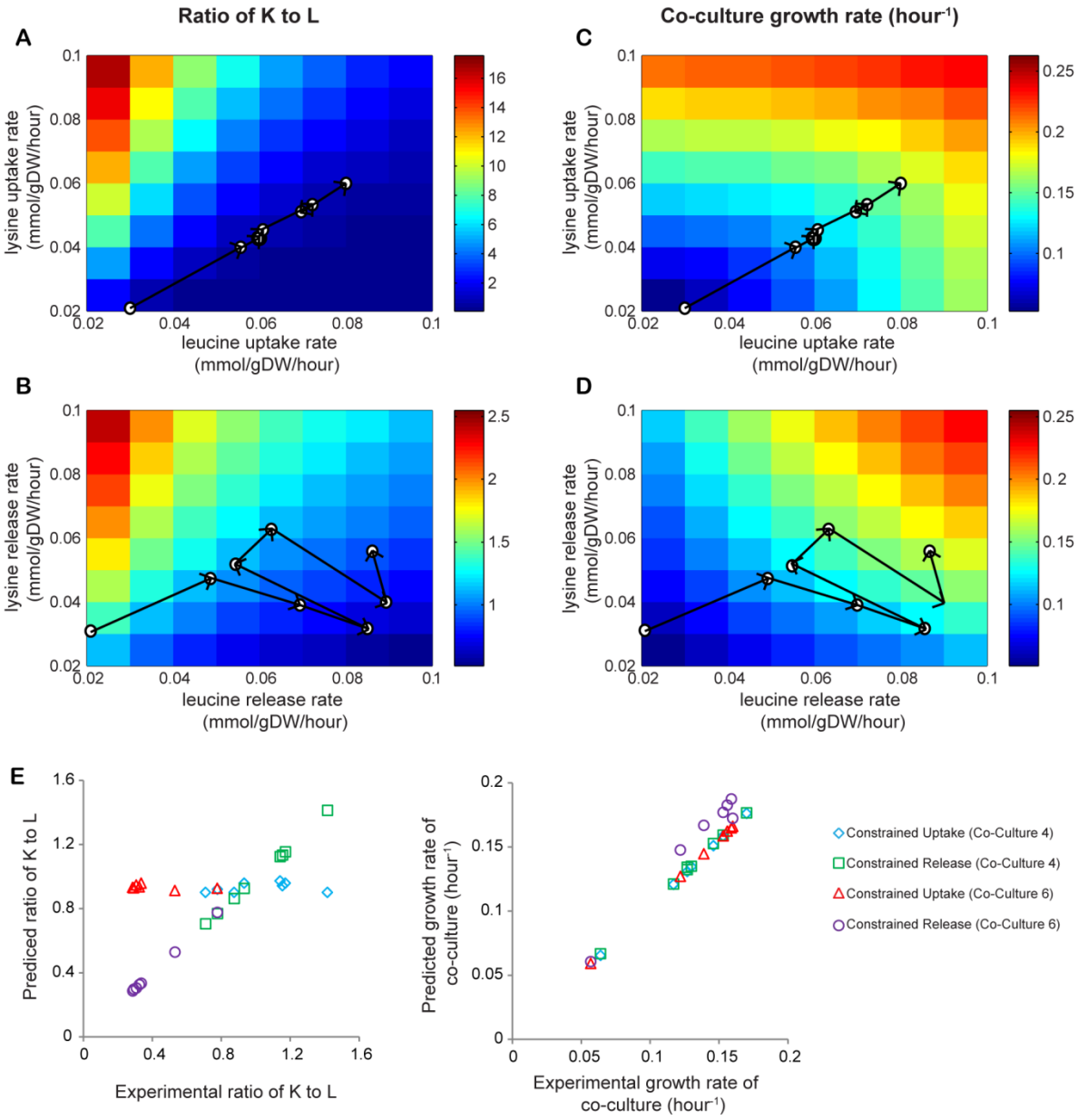


Figure 2.13: Computational model predictions of co-culture composition and growth rates.
The model was constrained using either amino acid uptake (panels A and C) or release rates (panels B and D). Panels A and B display the predicted K:L ratio at a co-culture OD≈0.2. The color map indicates the value of K:L ratio. Panels C and D show the predicted average growth rate of co-culture, indicated by the color map. The evolutionary trajectory of co-culture 4 is shown on panels A through D, where the open circles indicate passages 1,4,7,10,12,15,19 and 21. The estimated uptake or release rates for evolved K^{ev} and L^{ev} strains in each passage were then used to constrain the model. Panel E compares the model predicted K:L ratio and average growth rate of the co-culture near OD600≈0.2 to the estimated experimental values. Blue diamonds and red triangles denote the predictions when the model was constrained by the estimated uptake rates for co-culture 4 and 6, respectively. Green squares and purple circles denote the predictions when the model was constrained by the estimated release rates for the two co-cultures for co-culture 4 and 6, respectively.

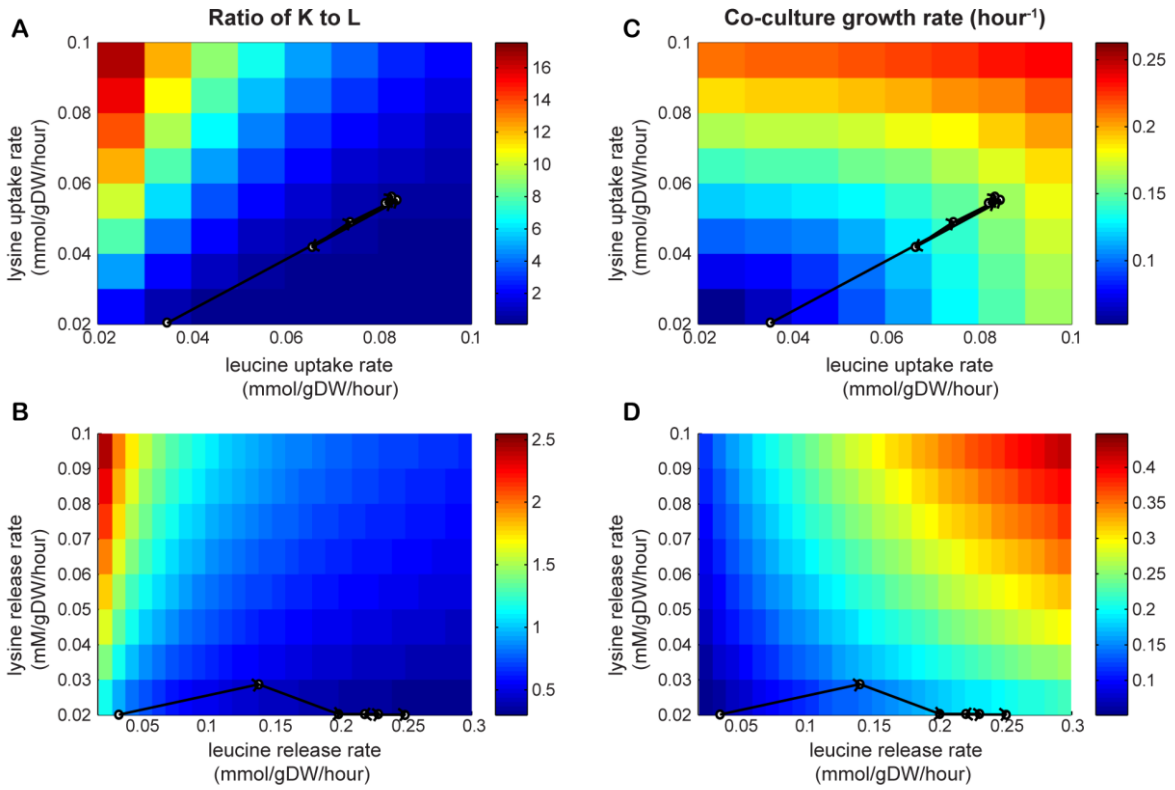


Figure 2.14. Computational model predictions of co-culture composition and growth rates.

The model was constrained using either amino acid uptake (panels A and C) or release rates (panels B and D). Panels A and B display the predicted K:L ratio at a co-culture OD≈0.2. The color map indicates the predicted K:L ratio. Panels C and D show the predicted average growth rate of co-culture, indicated by the color map. The evolutionary trajectory of co-culture 6 is shown on panels A through D, where the open circles indicate passages 1,4,7,10,13, 16 and 18.

Table 2.3. Estimated uptake and release rates of lysine and leucine during adaptive evolution.

Passage	Lysine uptake rate (mmol/gDW/hour)	Leucine uptake rate (mmol/gDW/hour)	Lysine release rate (mmol/gDW/hour)	Leucine release rate (mmol/gDW/hour)
Un-evolved Co-culture				
Un-evolved	0.020	0.027	0.020	0.027
Co-culture 4				
1	0.022	0.030	0.032	0.021
4	0.041	0.055	0.047	0.048
7	0.044	0.060	0.039	0.069
10	0.044	0.060	0.032	0.085
12	0.046	0.061	0.052	0.054
15	0.054	0.072	0.063	0.062
19	0.051	0.069	0.040	0.089
21	0.060	0.080	0.056	0.086
Co-culture 6				
1	0.020	0.027	0.016	0.035
4	0.056	0.076	0.030	0.142
7	0.043	0.058	0.013	0.197
10	0.049	0.066	0.014	0.229
13	0.055	0.074	0.018	0.226
16	0.054	0.072	0.018	0.215
18	0.056	0.075	0.017	0.245

The uptake and release rates of lysine and leucine were estimated using equations 2.1-2.3 in methods section.

2.4 Discussion

We built a synthetic mutualism system with two *E. coli* auxotrophs (Δ*lysA recA::kan* and Δ*leuA recA::cat)* and adaptively evolved the co-culture. In their initial encounter, both strains exchanged leucine (or its precursors) and lysine cooperatively to support the partner's growth.

Replicate co-cultures all maintained persistent cooperation during adaptive evolution and achieved similar growth rates but resulted in different population compositions and evolutionary trajectories. The experimental data and computational model predictions were combined to analyze mechanisms for improving cooperation during evolution and suggested that one evolved co-culture benefited from a better exchange of leucine (or its precursors), while another evolved co-culture experienced better exchange of both amino acids. Interesting, the timing needed to improve co-culture growth rates (~10 days) was similar to a previous study on evolving individual strains [50]. Based on our results, adaptive evolution of the co-culture led to: (1) improved exchange of essential metabolites between strains, (2) altered survival during starvation, and (3) fitness tradeoffs.

2.4.1 Metabolite exchange between strains

E. coli does not normally secrete amino acids, and amino acid synthesis is well controlled by regulatory mechanisms so that the cellular inputs are best used for growth. In previous studies of auxotrophs, starvation led to cell death and release of some metabolites (amino acids and nucleic acids) [25,29]. In lysine-limiting media, a Δ*lysA E. coli* mutant (lacking diaminopimelate decarboxylase) has been shown to secrete various metabolites, including diaminopimelate (DAP), an important cell wall constituent [51,52]. In our study, we observed that in mono-culture without amino acid supplementation, the L and K strains showed different death rates. Given the different death rates of the two strains we expected to see an initial one-way cross-feeding from K to L (not cooperative) and a lag phase prior to exponential growth in co-culture. However, we found reproducible growth of both strains in co-culture and an

absence of a lag phase in replicate co-cultures, indicating consistent two-way cross-feeding of leucine (or its precursors) and/or lysine.

During adaptive evolution the estimated uptake rates of leucine and lysine increased (**Figure 2.13**), making this one possible mechanism for improvements in the co-culture. We did not detect any leucine (or its precursors) or lysine in the co-culture media, indicating that the levels of these essential metabolites are below our bioassay detection limit (~3.5 μM) and are quickly consumed. Based on the co-culture growth rates and amino acid requirements, we estimated the possible uptake rates for lysine and leucine (or its precursors) to be ~0.02 and ~0.03 mmol/gDW/hour, respectively, for un-evolved co-cultures and ~0.06 and ~0.08 for evolved co-cultures. These values are close to the reported transport rates ~0.048 mmol/gDW/hour for leucine with a concentration of ~2 μM and ~0.011 to 0.044 mmol/gDW/hour for lysine with a concentration ranging from 0.2 μM to 10μM [53,54].

We used a genome-scale metabolic model to get a better understanding of how metabolite uptake/release rates affect the fitness landscape (**Figures 2.13** and **2.14**). In general, increases in both release and uptake rates will enhance proliferation of strains and alter community composition. A prior study by Wintermute and Silver developed models to evaluate invested benefits and cooperation levels in *E. coli* co-cultures [24,55]. They found that when one strain overshares (i.e., is highly cooperative), the other strain becomes dominant in the co-culture. The oversharing strain can only improve its growth if its partner cooperates. Our computational results (**Figure 2.13B** and **2.13D**) are consistent with these findings. When leucine (or lysine) release is higher in strain K (or L), its corresponding partner strain L (or K)

dominates. When its partner strain produces more lysine (or leucine), K (or L) will begin to increase its relative population in the community.

2.4.2 Altered viability during starvation

Un-evolved K and L strains exhibited different survival rates during lysine and leucine starvation (**Figure 2.2**). We observed that the un-evolved K strain (Δ*lysA recA::kan*) quickly underwent cell death in the absence of lysine. Cell death and lysis were also observed in a yeast *lysA* mutant [29]. In a previous study of *E. coli* co-cultures, a Δ*lysA* mutant behaved as a universal cooperator, supporting growth of a variety of other auxotrophs in co-culture, while other strains (including Δ*leuB*, Δ*leuC*, and Δ*leuD* mutants) grew with a smaller number of partner strains [24]. Based on our results, cell death could explain how universal cooperators enable co-culture growth through the release of many different metabolites by cell lysis. Another previous study suggests that evolution of cooperative cross-feeding requires an initial one-way cross-feeding by one species [30]. The stability of our K and L cooperative system could be due to strain K's ability to cross-feed metabolites due to cell death.

We additionally observed that the evolved K^{ev} and L^{ev} isolates displayed altered survival during amino acid starvation. The K^{ev} strains adapted to reduce their death rates during lysine starvation, while evolved L^{ev} strains died more quickly during leucine starvation. Increased cell death by L^{ev} strains and decreased cell death by K^{ev} strains could contribute to better metabolite exchange and improvement of the co-culture.

2.4.3 Fitness tradeoffs

In single species evolution, evolved strains often gain fitness in one environment at the expense of reduced fitness in another environment. Populations evolved in glucose media can lose fitness in the presence of other carbon sources [56]. Mutants adapted to low temperatures may have reduced fitness at higher temperatures [57]. The environment in the co-culture is complex, and strains adapted to the co-culture might gain fitness in the co-culture but lose fitness in mono-culture. Our experiments demonstrated that evolved L^{ev} and K^{ev} isolates were able to improve growth of co-cultures (**Figure 2.12**), but had reduced fitness in mono-culture when supplemented with their essential amino acids (**Figures 2.9** and **2.10**). Assimilation of amino acids is important for improving co-culture growth and the reduced growth in mono-culture was unexpected. It may imply that strains in evolved co-cultures become dependent on other strains and/or that additional metabolites are being exchanged. Growth in mono-culture could decrease due to a downregulation or loss of essential genes, whose biological roles are fulfilled by the other strain in co-culture. This has been recently referred to as the black queen hypothesis [17]. Further investigation of these evolved strains using gene expression analysis and genomic sequencing could potentially identify genetic reasons for the observed changes in co-culture and mono-culture phenotypes.

In this study, we performed a series of experiments to investigate the behaviors of un-evolved and evolved co-cultures and how individual evolved isolates contribute towards improving co-culture fitness. Metabolite (lysine and leucine or its precursors) cross-feeding is essential for co-culture growth but unfortunately could not be quantified directly. Estimated uptake and release rates of essential metabolites increased over adaptive evolution, except for

lysine release rates in co-culture 6. In addition to genome and mRNA sequencing, future experimental approaches enabling the direct measurement of nutrient exchange rates in co-cultures would aide in pinpointing the mechanism(s) for the observed growth rate improvements. While this study and others [23-27,29] have focused on mutualistic interactions, the adaptive evolution of communities with other types of symbiotic interactions (e.g., commensalism, amensalism or parasitism) would be of interest as well [38].

Chapter 3: Metabolic engineering of *Escherichia coli* for production of pyruvate

3.1 Introduction

Microbes can produce diverse useful chemicals [58]. However, microbes have not evolved to just produce a specific product of interest. Metabolic engineering of microbes aims to improve production rates, yields, and titers, generating microbial factories for cost-effective production of desired chemicals [1,59]. A broad range of products have been successfully produced by engineered strains, such as transportation fuels (ethanol, butanol and diesel) [10,60-63], pharmaceuticals (alkeloids, polyketides, nonribosomal peptides and isoprenoids) [64-69], and bulk and fine chemicals (amino acids, organic acids, industrial solvents and polymer precursors) [3,70-73]. Some chemicals are starting points (or precursors) for other important materials, and engineered strains which overproduce these precursors can be further modified to produce other important chemicals. This strategy has been used previously, since one of the first steps to produce a desired end-product is to increase the supply of its precursor(s). For example, a strain with elevated malonyl-CoA levels was used to produce an important polyketide, phloroglcinol [74]. Pyruvate overproducing strains have also been altered to produce L-alanine and diacetyl [75]. Similarly, strains with higher levels of oxaloacetate showed increased succinate, threonine and lysine in the production strains [76].

Pyruvate is a key metabolite in central metabolism and is a precursor for acetyl-CoA, acetaldehyde and several amino acids (including alanine, lysine, valine, isoleucine, and leucine).

Commodity chemicals (e.g., ethanol, acetic acid, lactic acid, and acrylic acid), as well as active pharmaceutical ingredients (e.g., polyketides and isoprenoids) can also be derived from pyruvate. Furthermore, pyruvate itself has various applications as a food addictive, nutriceutical, weight loss agent, and anti-ageing skin treatment. Pyruvate is commercially manufactured by a chemical method involving dehydration and decarboxylation of tartaric acid. The process is not cost-effective and requires the use of toxic solvents [77]. Therefore, microbial production of pyruvate is an attractive alternative to a chemical process. A number of strains have been genetically modified to improve pyruvate production in *Escherichia coli*, yeast and *Corynebacterium glutamicum* [14,78-82]; however, high yields have not been achieved.

The metabolic engineering strategies applied previously to *E. coli* strains mainly focused on blocking pyruvate consuming pathways which produce phosphoenolpyruvate (PEP), acetyl-CoA, ethanol, acetate, lactate and formate. Other strategies prevented the conversion of PEP to oxaloacetate by deleting PEP synthase, and increasing glycolytic flux by disrupting oxidative phosphorylation using a F_1-ATPase-defective mutant or reducing the NADH availability [14,78,79]. Pyruvate production was also improved by reducing CO_2 formation by the TCA cycle through deletion of α-ketoglutarate decarboxylase [14].

A primary goal for developing a pyruvate overproducer is of course high yield. Until now, the highest yield (0.75 g/g) reported is 78% of theoretical maximum yield [14]. In order to push the pyruvate production towards higher yields, it is important to understand the impact of metabolic engineering interventions on metabolism at a systems-level. Given the scale and

complexity of the metabolic and gene networks, predicting systems-level impacts on metabolism can be facilitated through computational modeling.

Genome-scale metabolic models and their associated analytical tools are useful for predicting cellular phenotypes in response to genetic perturbations and suggesting gene target modifications to improving chemical production [8]. A bi-level optimization algorithm (OptKnock) has been successfully used to design strains for overproducting of succinate, lactate, 1,3-propanediol, and 1,4-butanediol [70,83,84]. OptKnock identifies reaction deletions that couple cellular growth and chemical production. When growth rate is used as a selection pressure, OptKnock designed strains should increase production of the target chemical through adaptive evolution. OptKnock only considers the reaction network and does not account for gene to reaction associations. Another algorithm (OptORF) was developed based on OptKnock, but searches for gene deletions instead of reaction deletions [85].

In this study, we sought to design and construct pyruvate overproduction strains using OptORF and a genome-scale metabolic model of *E. coli*. Four computationally designed mutant strains were constructed and characterized for growth and pyruvate production under aerobic conditions. Two strains were adaptively evolved, which increased growth rates and pyruvate production. Finally, the pyruvate strains were used as platform strains to develop other chemical production strains. This was demonstrated by modifying the pyruvate strains to produce ethanol.

3.2. Materials and methods

3.2.1 Strains and plasmids

The parental strain, *E. coli* BW25113, and pCP20 plasmid were obtained from the *E. coli* genetic stock center (CGSC, Yale University). The *E. coli* single gene deletion strains Δ*aceE::kan*, Δ*cyoA::kan*, Δ*pta::kan*, Δ*ldhA::kan*, Δ*aceA::kan*, Δ*lpdA::kan*, Δ*gnd::kan*, Δ*sdhA::kan*, Δ*poxB::kan*, Δ*gdhA::kan*, and Δ*pflB::kan* mutants were obtained from the Keio collection (Open Biosystems). To generate mutants (listed in **Table 3.1**) with multiple gene deletions, the kanamycin resistance gene (*kan*) was removed using the pCP20 plasmid [41]. An additional gene was deleted (and *kan* re-inserted) using P1 transduction [86] with selection on LB agar plates with 50 μg/mL kanamycin. This process was repeated for each additional knockout and the gene deletions were verified by PCR after each round.

The pJGG2 plasmid and its corresponding empty vector, pBBR1-MSC5, were obtained from Robert Landick (University of Wisconsin-Madison). The pJGG2 plasmid is a low copy number plasmid with a *lac* promoter that controls the expression of the *Zymomonas mobilis* PET cassette genes (*pdc* and *adhB*), which are responsible for ethanol synthesis. The *E. coli* K-12 ethanologen, GLBRCE1, was obtained from Robert Landick and is missing *ldhA, pflB* and *ackA* [87]. GLBRCE1 also contains pJGG2 and a chromosomal copy of the PET cassette inserted in the *pflB* locus.

3.2.2 Media and culture conditions

M9 minimal medium (pH 7.0, 100 μM $CaCl_2$, 2 mM $MgSO_4$, 6.4 g/L $Na_2HPO_4 \bullet 7H_2O$, 1.5 g/L KH_2PO_4, 0.25 g/L NaCl, 0.5 g/L NH_4Cl) supplemented with glucose and acetate (at varying

concentrations) was used throughout the study. For pyruvate and ethanol production experiments, wild-type and mutant strains were precultured at 37°C overnight in Luria Broth (LB) and then pelleted and washed twice with M9 minimal medium to remove any residual nutrients from the preculture. For aerobic experiments, cultures were started with an initial OD600 of 0.01 and then grown aerobically in 250 mL flasks containing 100 mL glucose minimal media. Ethanol fermentation experiments were performed anaerobically in hungate culture tubes containing 10 mL of medium with an initial OD600 of 0.01. The pJGG2 and pBBR1-MSC5 plasmids confer gentamicin resistance and 15 μg/mL gentamicin was used in the ethanol experiments. IPTG was added to a final concentration of 200 μM to induce the expression of PET cassette in the ethanol experiments. To remove oxygen, the hungate tubes were vacuumed and flushed with argon three times. All experiments were carried out at 37°C in a shaking incubator with a shaking speed of 150 rpm and performed with three replicates. Samples were periodically taken for further analysis and cells were removed using 0.2 μm nylon filter.

3.2.3 Metabolite uptake and secretion rate measurements

The concentration of glucose in the medium was determined using an enzyme assay from Sigma (GAGO20). The concentrations of pyruvate, lactate, acetate, succinate, formate and ethanol in the medium were measured by HPLC using an Aminex HPX-87H with Cation-H guard column (Bio-Rad, cat# 125-0140). The mobile phase contained 0.02N H_2SO_4 and was run at a flow rate of 0.5 mL/min at 50°C. The products were detected and quantified (from standard curves) based on their refractive index. The uptake/secretion rates were determined from the

metabolite and biomass concentration data from the exponential growth phase. The biomass concentration (gram of cell dry weight per liter) was calculated using the measured OD600 values and a conversion factor 1 OD600 = 0.415 gDW/L [44]. The maximal yield per glucose is 0.51 gDW cell/g glucose [88].

3.2.4 Adaptive evolution

Adaptive evolution of two mutant strains was performed for 20 passages. The initial cultures were inoculated at an OD600 of 0.01 and grown at 37°C in 100 mL of M9 minimal medium supplemented with 1.6 g/L glucose and 0.4 g/L acetate. When cells reached an OD600 ~ 0.2 OD600, the cells were transferred to fresh medium (such that starting OD600 = 0.01). During adaptive evolution, the amount of acetate in the minimal medium was gradually reduced, while the glucose concentration was increased so that the total carbon source concentration was maintained at 2 g/L. After 15 passages, the medium contained 1.98 g/L glucose and 0.02 g/L acetate. Cultures from each passage were frozen and stored at -80°C.

3.2.5 *In silico* computations

The optimization algorithm OptORF was used to identify genetic strategies which couple growth and pyruvate production [85]. This method finds solutions that ensure pyruvate is produced when cells are at their highest biomass yield. Thus, selection for increased growth rates will subsequently select for increased pyruvate production. The OptORF algorithm was run using a tilted inner objective function (growth rate – 0.001 •pyruvate production rate) that

finds the minimum pyruvate production at the maximum growth rate [89]. A penalty equal to 1 for gene deletions was used in OptORFs outer objective function. All simulations were done for glucose aerobic conditions using the iJR904 *E. coli* genome-scale metabolic network [90], where the maximum glucose uptake rate was set to 10 mmol/gDW/hour and the maximum oxygen uptake was unlimited.

3.3. Results

3.3.1 *In silico* strain design for pyruvate production

Four different strain design strategies for pyruvate production were selected from those suggested by OptORF: (1) *aceE, cyoA, cydB, pta, eutI, ldhA, dld*, (2) *lpdA, gnd, sdhA, poxB, pflB, pflD, tdcE, purU*, (3) *aceE, gdhA, poxB, ldhA, dld, atpE, pflB, pflD, tdcE*, and (4) *aceE, gnd poxB, ldhA, dld, atpE, pflB, pflD, tdcE*. Given the large number (7-9) of gene deletions suggested, we sought to prioritize the gene targets. Gene targets that might be inactive under glucose aerobic conditions (e.g. due to regulation) were first excluded. PflB is expressed under microaerobic and anaerobic conditions [91] and a PflD (encoding a putative pyruvate formate lyase) deletion mutant had a similar fermentation pattern as its parent strain [91,92]. In addition, *eutI, dld* and *tdcE* encode minor isozymes for Pta, LdhA and PflB, respectively [93-96]. PurU hydrolyzes 10-formyltetrahydrofolate into formate for use in purine biosynthesis by PurT. A PurU deletion mutant grows well aerobically in glucose, since another enzyme (PurN) can be used instead of PurT [97,98]. These genes (*pflB*, *pflD*, *eutI*, *dld*, *tdcE* and *purU*) were not selected for deletion since they were thought to have low (if any) activity and would thus have weaker metabolic

effects. Additionally, the *cydB* and *atpE* deletions were lethal in combination with other gene deletions (data not shown) and could not be included in the constructed strains. The remaining genes identified by OptORF were deleted in the engineered strains (**Figure 3.1**).

The resulting engineered strains each involved deletions that impacted metabolism and pyruvate production differently. Deletion of *aceE, pta, poxB* and *ldhA* will reduce the conversion of pyruvate into acetyl-CoA, acetate and lactate. Deletion of *cyoA* or *sdhA* serves to slow down the citric acid (TCA) cycle which would otherwise divert flux away from pyruvate. The rationale for deleting *gdhA* and *gnd* is less obvious. *E. coli* has two primary pathways for glutamate synthesis from NADPH, ammonia and α-ketoglutarate. The GDH pathway (encoded by *gdhA*) does not require ATP, while the other GS-GOGAT pathway consumes one ATP per glutamate produced. Deleting *gdhA* will force the GS-GOGAT pathway to be used, increasing ATP consumption and decreasing growth rate. Similarly, deleting *gnd* prevents NADPH production by the pentose phosphate pathway and increases NADPH production from NADH via pyridine nucleotide transhydrogenases. The transhydrogenase consumes energy, thereby lowering the maximum growth rate. In both cases, lowering the maximum growth rate (via *gdhA* or *gnd* deletions) increases the pyruvate production rates (since both pyruvate and biomass compete for carbon). The gene deletions suggested by OptORF either prevent pyruvate consumption or reduce growth, and both synergistically enhance pyruvate production. Based on the computational results, four strains (PYR001-PYR004) were constructed and tested experimentally. Each strain contained four gene deletions and is listed in **Table 3.1**.

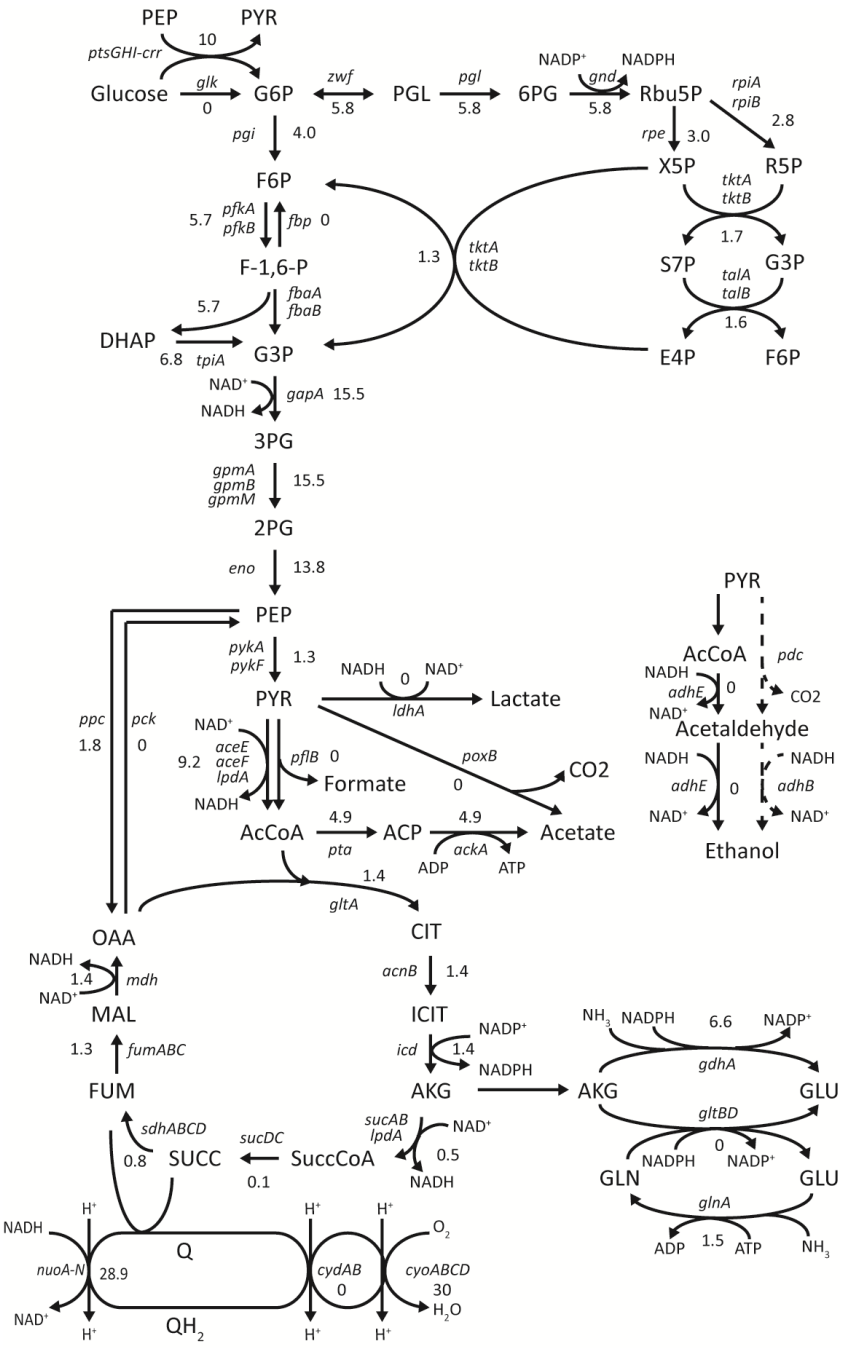


Figure 3.1. Central metabolic pathway of wild-type *E. coli*.

Genes associated with each reaction in the central metabolic network are shown. The metabolic flux distribution for wild-type strain under aerobic condition was predicted by flux balance analysis and flux values were labeled. Glucose uptake rate was set as 10 mmol/gDW/hour. The dash line represents the ethanol synthesis pathway (PET operon) from *Zymomonas mobilis*.

Table 3.1 Strains and plasmids used in this study.

Strains/plasmid	Relevant characteristics	Reference
E. coli strains		
BW25113	*lacI*^{q} *rrnBT*14 Δ*lacZWJ*16 *hsdR*514 Δ*araBADAH*33 Δ*rhaBADLD*78	[41]
PYR001	BW25113 Δ*aceE*::*kan* Δ*cyoA* Δ*pta* Δ*ldhA* Δ*aceA*	This study
PYR002	BW25113 Δ*lpdA*::*kan* Δ*gnd* Δ*poxB* Δ*sdhA*	This study
PYR003	BW25113 Δ*aceE*::*kan* Δ*gdhA* Δ*poxB* Δ*ldhA*	This study
PYR004	BW25113 Δ*aceE*::*kan* Δ*gnd* Δ*poxB* Δ*ldhA*	This study
PYR010	Evolved strain of PYR001	This study
PYR020	Evolved strain of PYR002	This study
GLBRCE1	MG1655 Δ*ackA* Δ*ldhA* Δ*pflB*::PET *crl*(70insIS*1*) *ylbE*(253insG) *gltB*(G3384A) *yodD*(A85T) *glpR*(150delG) *gatC*(916insCC) / pJGG2	[87]
EH010	PYR010 / pJGG2	This study
EH020	PYR020 / pJGG2	This study
EH010-pflB	PYR010 derivative with *kan* removed and addition of ΔpflB::*kan* / pJGG2	This study
EH020-pflB	PYR020 derivative with *kan* removed and addition of Δ*pflB*::*kan* / pJGG2	This study
EH030-pflB	PYR003 derivative with *kan* removed and addition of Δ*pflB*::*kan* / pJGG2	This study
EH040-pflB	PYR004 derivative with *kan* removed and addition of Δ*pflB*::*kan* / pJGG2	This study
WT_empty	BW25113 / pBBR1-MSC5	This study
WT_pJGG2	BW25113 / pJGG2	This study
Plasmids		
pJGG2	pBBR1-MSC5 with *adhB* and *pdc* (PET cassette) from pLOI295; Gent^{R}	[87]
pBBR1-MSC5	pBBR oriT; P_{lac}; Gent^{R}	[87]

Abbreviations: *kan*, kanamycin resistance gene; Gent^{R}, gentamicin resistance.

3.3.2 Pyruvate production

Parental and mutant strains PYR001, PYR002, PYR003 and PYR004 were characterized for pyruvate production in 100 mL glucose minimal medium for 60 hours (**Figure 3.2A**). Due to the *aceE* and *lpdA* deletions (subunits of pyruvate dehydrogenase) the synthesis of acetyl-CoA from pyruvate was prevented and the mutants were unable to grow with glucose as a sole carbon. To allow growth, acetate was added to the media to generate acetyl-CoA by acetyl-CoA synthetase (**Table 3.2**). The four mutants grew significantly slower than the parental strain, but

produced pyruvate as predicted (**Figure 3.2A**). The parental strain did not secrete any pyruvate. Strain PYR001 grew the slowest and consumed only ~40% of the glucose (~4.0 mM) within 60 hours; however, PYR001 converted most of the glucose consumed to pyruvate with a yield of 80% of the theoretical yield (**Table 3.2**). Strains PYR003 and PYR004 completed growth within 20 hours and produced 17.0 and 19.4 mM pyruvate, respectively (79% and 87% of theoretical yield). Among the four mutants, PYR002 had the lowest pyruvate yield (46% of theoretical yield) and also exhibited a low growth rate.

The final pyruvate titers were proportional to the pyruvate yields, except for strain PYR001, which did not consume all the glucose (**Figure 3.2A**). The volumetric production rate of pyruvate depends on the pyruvate titers and growth rates of the mutants. Strain PYR003 and PYR004 had the best volumetric production rates among the four mutants because of their high yields and growth rates. On the contrary, PYR001 only reached 13% of the volumetric production rate of PYR003, as a result of its poor growth and low titer. The production rate per cell dry weight was also calculated. PYR001 and PYR002 both had low production rates, ~30% of that of PYR003.

The secretion of by-products from strains, such as succinate, formate, acetate, lactate and ethanol, was analyzed by HPLC. Acetate was the main byproduct of the parental strain. PYR001 and PYR002 formed 1-2 mM acetate (even though acetate was required for growth), while the other mutants consumed acetate, presumably for acetyl-CoA generation (**Figure 3.3**). Only one strain PYR002 produced lactate with a concentration 9.8 mM, which could explain its low pyruvate yield and low production per gram of cell dry weight. Concentrations of succinate,

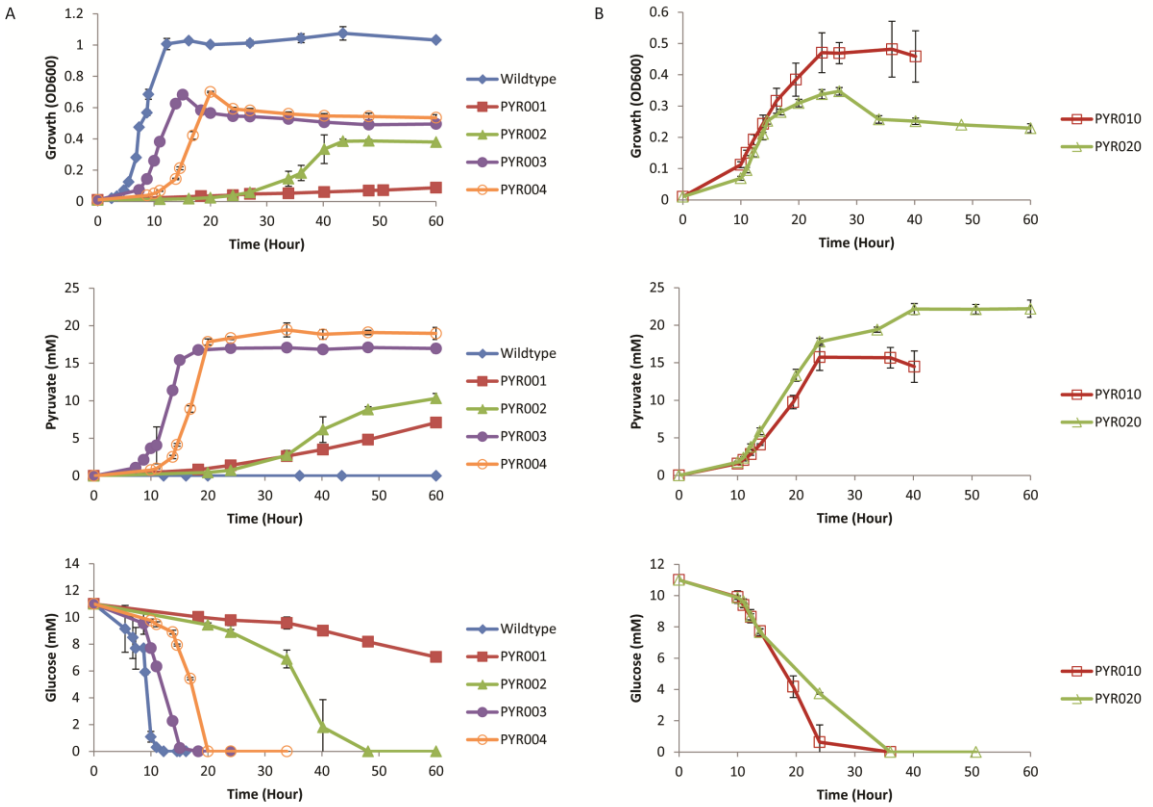


Figure 3.2. Cellular phenotype of wild-type and mutant strains designed *in silico*.
(A) Growth and production of pyruvate by wild-type, PYR001, PYR002, PYR003, and PYR004 strains is represented by blue diamond, red square, green triangle, purple solid circle and orange open circle. The medium is M9 minimal medium containing glucose with/without acetate. (B) The growth and production phenotype by evolved strains PYR010 and PYR020 is represented by open red square and open green triangle.

formate and ethanol secreted by all strains were too low to be detected by HPLC. A carbon balance was calculated for each strain. In the parental strain, 61% of carbon was used for biomass and acetate formation, and the remainder was presumably oxidized to CO_2. On the contrary, the four mutants used 90-111% of the carbon for biomass and pyruvate with less carbon for CO_2 formation. This was consistent with the predictions from flux balance analysis using the genome-scale metabolic model. The reported yields and carbon balances being higher than 100% is likely due to evaporation in the shake flasks (estimated to be ~0.13 mL per hour) which would concentrate the media.

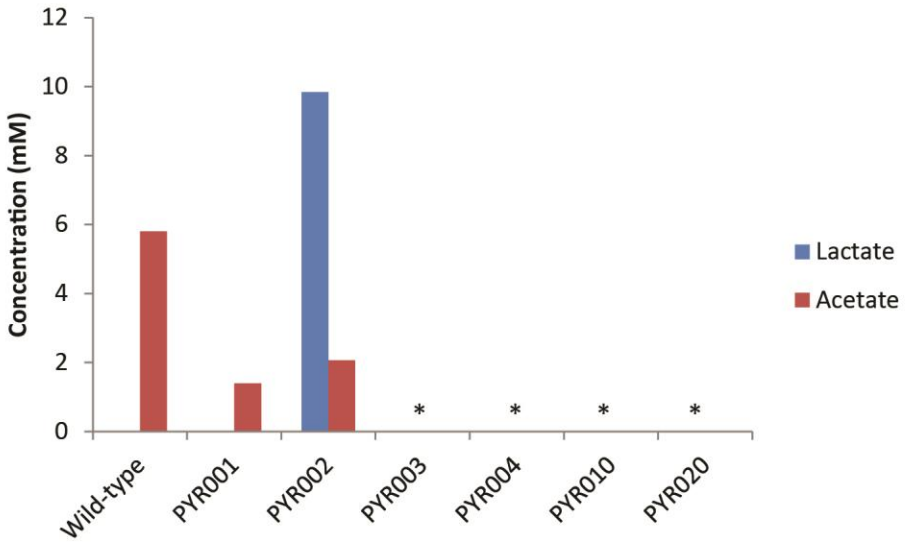


Figure 3.3: Lactate and acetate secretion for wild-type and mutant strains in aerobic condition.
The concentration is the maximum acid produced within 60 hour. Acetate accumulation in the culture of wild-type, PYR001 and PYR002 strains was observed by HPLC analysis. * indicates the concentrations of acetate and lactate were below the detection level of HPLC.

3.3.3 Adaptive evolution

Strain PYR003 and PYR004 produced large amounts of pyruvate. To improve pyruvate production by strain PYR001 and PYR002, adaptive evolution was conducted under aerobic conditions for 20 passages at 37°C in glucose+acetate minimal medium. Acetate was added to the medium to enable the cell growth, but the concentration added was reduced over the different passages (**Table 3.2**). The pyruvate strains were designed such that growth rate and pyruvate production would increase simultaneously. Therefore, adaptive evolution could select for faster growing strains with enhanced pyruvate yields. The evolved strains PYR010 and PYR020, progenies of PYR001 and PYR002 from the last passage, were characterized and their performance is shown in **Table 3.2** and **Figure 3.2B**. The evolved strains had a 10- and 3- fold increase in growth rate and an almost 2-fold increase in pyruvate titers. In terms of pyruvate

Table 3.2 Production of pyruvate from the wild-type and mutant strains.

Strains	Medium		Growth rate ($hour^{-1}$)	Pyruvate yield		Pyruvate concentration		Pyruvate production rate		Carbon balance [f] %
	Glucose (g/L)	Acetate (g/L)		% theoretical [b]	conversion [c] (g/g)	(g/L) [d]	(mmol/gDW) [d]	Volumetric (g/L/hour)	Specific [e] (mmol/gDW/ hour)	
Wild-type(WT)	2	0	0.59±0.01	0	0	0	0	0	0	60.83±0.78
PYR001	1.9	0.1	0.02±0.00	80.24±4.63	0.78±0.05	0.62±0.04	194.0±11.2	0.01±0.00	6.04±0.24	111.0±6.03
PYR002	1.8	0.2 [a]	0.12±0.01	46.24±2.89	0.43±0.03	0.91±0.06	65.56±4.09	0.02±0.00	5.47±0.04	103.7±4.27
PYR003	1.9	0.1	0.45±0.03	79.05±0.63	0.75±0.00	1.50±0.01	75.10±0.60	0.08±0.00	20.36±0.47	99.24±1.00
PYR004	1.9	0.1	0.30±0.00	86.97±4.12	0.82±0.04	1.71±0.08	83.63±3.97	0.07±0.01	19.11±0.25	110.9±4.71
PYR010	1.98	0.02	0.20±0.04	69.34±7.81	0.67±0.08	1.39±0.16	80.71±9.09	0.06±0.00	14.91±1.68	90.51±10.7
PYR020	1.98	0.02	0.34±0.00	95.16±3.12	0.92±0.03	1.95±0.06	212.3±6.96	0.05±0.00	23.73±0.88	110.6±3.72

[a] This mutant requires more acetate than others to start growth within 48 hour.

[b] Percent of theoretical yield is calculated as the pyruvate concentration is divided by the theoretical maximum production of pyruvate (2 mmol of pyruvate per mmol of glucose). Acetate is also taken account for calculating the theoretical maximum production (0.5 mmol of pyruvate per mmol of acetate). The yield is adjusted by the culture volume loss due to the liquid evaporation in aerobic condition.

[c] The conversion is expressed as the gram of pyruvate produced per gram of total carbon source. It is adjusted by the culture volume loss due to the liquid evaporation in aerobic condition.

[d] The concentration is the value reported from HPLC analysis.

[e] The specific production rate is the pyruvate production rate per gram of cell dry weight (gDW) during exponential growth.

[f] The carbon balance is calculated as the percent of carbon used for the biomass formation, pyruvate production and byproduct formations.

yield, PYR010 had a 10% lower yield than its unevolved strain while PYR020 had ~2- fold increase. The combination of improved growth rate and pyruvate yields increased the volumetric pyruvate production rate of PYR010 and PYR020 compared to their unevolved strains. Interestingly, both evolved strains both needed less acetate (5-fold and 10-fold decrease) in the medium to support their growth.

Among the four designed strains and two evolved strains, PYR020 performed best with respect to yield, titer and volumetric productivity, followed by PYR004. To account for the slower growth rates of engineered strains, the pyruvate specific production rate (normalized by gram per cell dry weight instead of volume) was calculated. PYR003 and PYR004 had 3-4 fold higher specific production rates than PYR001 and PYR002 (**Table 3.2**). The difference was caused by the different gene knockout strategies for the four strains. PYR003 and PYR004 had deletions which affected the NADPH and ATP supply to decrease growth rates, while deletions in PYR001 and PYR002 affected the TCA cycle to reduce the biomass yield.

3.3.4 Production of ethanol in the altered pyruvate producing strains

Pyruvate is a precursor for many metabolites. To test the idea of using the engineered pyruvate strains to produce other chemicals, we further altered the strains to enhance ethanol production by expressing the *Z. mobilis* PET cassette. The pJGG2 plasmid included the PET cassette containing pyruvate decarboxylase (*pdc*) and alcohol dehydrogenase (*adhB*) under the control of an IPTG inducible *lac* promoter.

NADH is converted back into NAD using oxidative phosphylation or by secreting fermentation byproducts, such as ethanol and lactate. Since ethanol was the desired product, the strains were fermented in hungate tubes with 1.98 g/L glucose minimal medium supplemented with 0.02 g/L acetate. Ethanol production was analyzed after 48 hours. Three controls were included: the parental strain with empty vector (WT_empty), the parental strain with pJGG2 plasmid (WT_pJGG2) and an ethanol production strain (GLBRCE1, which lacks *ackA*, *pflB*, and *ldhA* and expresses the PET cassette from the chromosome and pJGG2 plasmid) [87]. Pyruvate formate lyase (PflAB) is active under anaerobic conditions but not aerobic conditions, converting pyruvate to acetyl-CoA and formate. Since the OptORF strategies deleted reactions converting pyruvate into acetyl-CoA, *pflB* was deleted from the four pyruvate mutants to make ethanol production strains EH010-pflB, EH020-pflB, EH030-pflB and EH040-pflB. The WT_empty strain (containing the empty vector) had a lower growth rate compared to WT_pJGG2 (containing the PET cassette on pJGG2) (**Table 3.3**). Moreover, WT_pJGG2 strain had around a 2- fold higher ethanol yield, ethanol titer and ethanol production rate compared to WT_empty. The improved growth and ethanol production is likely a result of enhanced NADH recycling. Aerobically, the pyruvate mutants had growth rates between 37 and 77% of the growth rate of parental strain, while anaerobically the mutants (EH010-pflB, EH020-pflB, EH030-pflB and EH040-pflB) derived from the pyruvate mutants exhibited growth rates between 64 and 136% of the growth rate of WT_empty. Three mutants (EH020-pflB, EH030-pflB and EH040-pflB) had 1.3-fold higher ethanol yields compared to WT_pJGG2 strain, and had higher specific production rates than GLBRCE1, even though the yields were similar. The EH010 and EH020 strains (both *pflB*$^{+}$) had lower ethanol yields and higher production of other byproducts—

succinate, acetate and formate (**Figure 3.4** blue color). Surprisingly, EH010-pflB did not show increased ethanol yield compared to WT_pJGG2 but did have a higher specific production rate.

During the 48 hour fermentations about half the culture volume was removed for HPLC analysis, which doubled the headspace in the hungate tube. Since ethanol is highly volatile, ethanol could enter the headspace and escape from the hungate tube during sampling, leading to an underestimation of ethanol concentrations. The fermentation experiments were repeated taking fewer samples (at 16, 20 and 24 hours). All strains consumed glucose completely within 24 hours (data not shown). Four of the nine strains re-tested (EH010, EH020-pflB, EH030-pflB and EH040-pflB) showed significantly higher ethanol yields when fewer samples were taken (**Figure 3.4** red color). Three of the engineered ethanol production strains from this study (EH020-pflB, EH030-pflB and EH040-pflB) had higher ethanol yields and lower succinate production than GLBRCE1.

The ethanol strains were grown with 0.02 g/L acetate in the medium and it was unclear how acetate concentrations would affect ethanol yields. Additional fermentation experiments were performed using medium with 0.1g/L acetate and 1.9 g/L glucose. The ethanol yields and byproduct concentrations did not appear to differ when a higher concentration of acetate was used (**Figure 3.4** green color).

Table 3.3 Production of ethanol from the wild-type and mutant strains.

Strains[a]	Growth rate ($hour^{-1}$)	Ethanol yield		Ethanol concentration		Production rate		Carbon balance [e] %
		% theoretical [b]	Conversion [c] (g/g)	(g/L)	(mmol/gDW)	Volumetric (g/L/hour)	Specific [d] (mmol/gDW/ hour)	
WT_empty	0.28±0.00	38.04±1.70	0.19±0.01	0.39±0.02	39.15±1.75	0.02±0.00	6.26±0.10	91.57+5.21
WT_pJGG2	0.37±0.02	63.06±2.59	0.32±0.01	0.64±0.03	56.71±2.33	0.04±0.00	11.71±1.09	114.3+9.91
GLBRCE1	0.16±0.02	82.21±0.91	0.42±0.01	0.83±0.01	128.8±1.43	0.03±0.00	16.08±0.78	92.42+1.90
EH010	0.25±0.01	62.45±2.54	0.32±0.01	0.63±0.03	80.12±3.25	0.03±0.00	17.91±2.96	87.00+4.50
EH010-pflB	0.18±0.01	61.81±6.77	0.31±0.03	0.62±0.07	88.96±9.75	0.02±0.00	16.61±1.15	80.70+7.75
EH020	0.38±0.00	58.05±5.06	0.29±0.03	0.59±0.05	59.29±5.17	0.04±0.00	19.75±2.03	97.15+8.65
EH020-pflB	0.25±0.02	80.23±4.84	0.41±0.02	0.81±0.05	134.0±8.09	0.04±0.00	23.10±1.48	95.90+4.64
EH030-pflB	0.19±0.05	79.47±7.12	0.40±0.04	0.80±0.07	122.4±11.0	0.02±0.00	19.29±1.12	90.13+4.26
EH040-pflB	0.22±0.03	84.59±7.03	0.43±0.04	0.85±0.07	107.6±8.94	0.04±0.00	22.37±2.28	100.2+2.93

[a] WT_empty, WT_pJGG2 and GLBRCE1 were grown in M9 minimal medium with 2 g/L glucose. The rest strains were grown in minimal medium with 0.02 g/L acetate and 1.98 g/L glucose.

[b] Percent of theoretical yield is calculated as the ethanol concentration divided by the theoretical maximum production of ethanol (2 mmol of ethanol per mmol of glucose). Acetate is also taken account for calculating the theoretical maximum production (0.67 mmol of ethanol per mmol of glucose).

[c] The conversion is expressed as the gram of ethanol produced per gram of carbon.

[d] The specific production rate is the ethanol production rate per gram of cell dry weight during exponential growth.

[e] The carbon balance is calculated as the percent of carbon used for the biomass formation and fermentation byproduct productions including succinate, lactate, formate, acetate and ethanol.

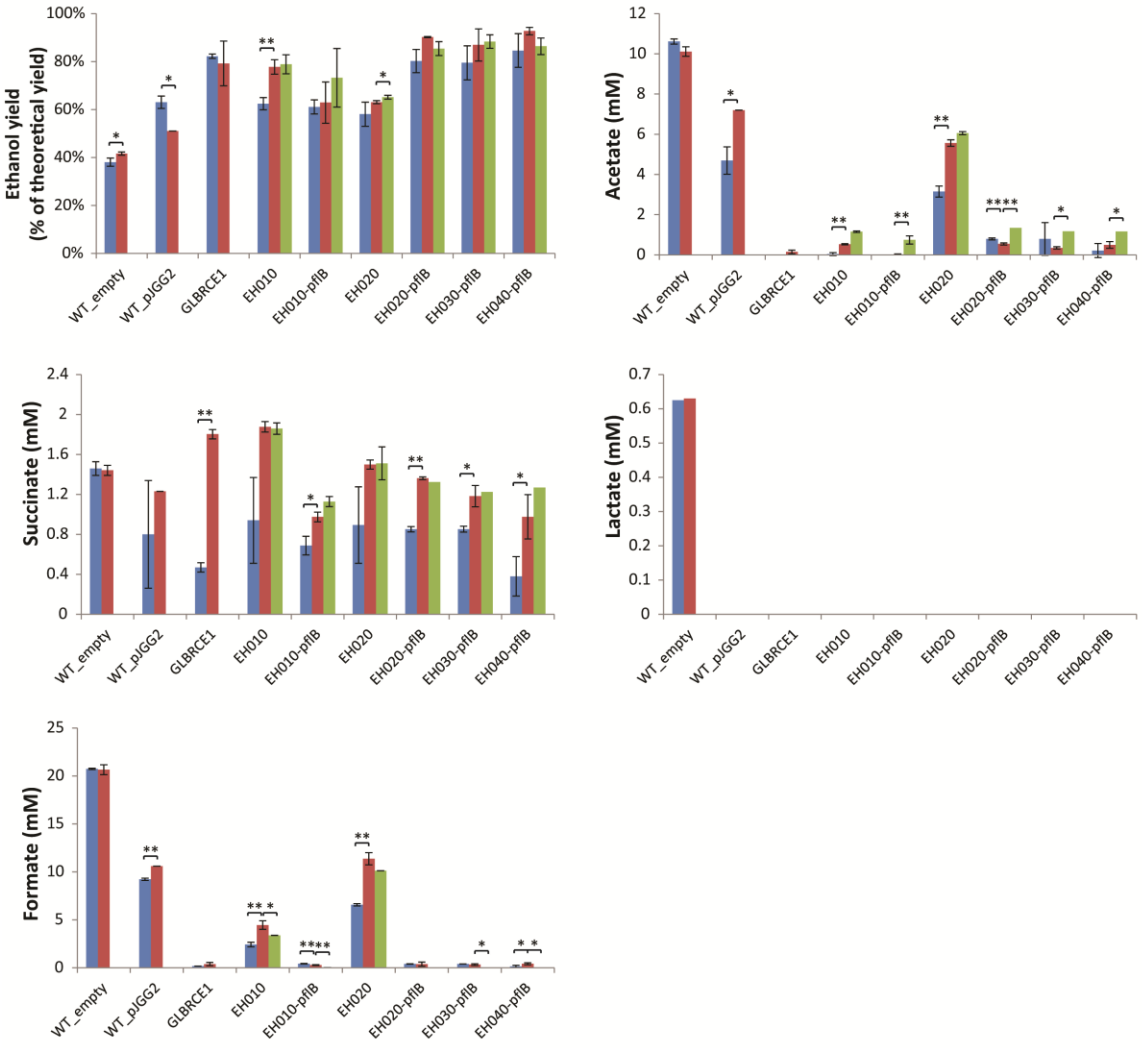


Figure 3.4: The byproducts secretion in anaerobic condition.

The blue columns represent the fermentation conducted in M9 minimal medium containing glucose and 0.02 g/L acetate for 48 hours and multiple samples (about half the culture volume) were removed for HPLC analysis. The red columns represent the fermentation in the same medium for 24 hours and three samples were taken at 16, 20 and 24hours. The green columns denote the fermentation in the minimal medium with more acetate (0.1 g/L) for 24 hours, and three samples were taken as well. Error bars represent standard errors among three replicates. Percent of theoretical yield is calculated as the ethanol concentration divided by the theoretical maximum production of ethanol (2 mmol of ethanol per mmol of glucose). Acetate is also taken account for calculating the theoretical maximum production (0.67 mmol of ethanol per mmol of glucose). The difference of product concentrations in 48 hour and 24 hour fermetnation (blue and red columns) were analyzed by t-test, and the same was done for the 24 hour fermentions containing different acetate concentrations in the medium (red and green columns). * and ** indicate p-value is between 0.01 and 0.05 or less than 0.01, respectively.

3.4 Discussion

Optimizing production of a specific metabolite usually involves increasing synthesis of its precursors in bacteria. Pyruvate is a starting compound for synthesizing a variety of biofuels (e.g., ethanol, 1-butanol and isobutanol) and chemicals. A high-yield pyruvate producing strain has great potential for creating strains to produce valuable chemicals. In this study, a genome-scale metabolic model of *E. coli* and OptORF were used to identify gene deletion targets to improve pyruvate production. Strains constructed based on the computational predictions produced high levels of pyruvate and adaptive evolution of two strains increased pyruvate yields, titers and volumetric and specific production rates. Further engineering of these platform pyruvate strains resulted in strains with high ethanol production.

3.4.1 Similar flux distribution patterns shared by *in silico* designed strains

All computationally designed strains over-produced pyruvate. The gene targets suggested by OptORF prevented pyruvate consumption by removing competing pathways and reduced growth by eliminating more energetically efficient routes for NADPH and glutamate production. The mutations involved shutting down the pentose phosphate pathway, reducing TCA cycle flux, and lowering biomass production (**Figure 3.5**). All of the computationally designed mutants were predicted to have increased glycolytic fluxes and coupling between growth and pyruvate production. Two of the strains immediately exhibited high pyruvate yields, while two other strains were adaptively evolved to improve production rates and/or yields.

3.4.2 The resource for synthesis of acetyl-CoA

All the pyruvate strains have pyruvate dehydrogenase subunits deleted (either *aceE* or *lpdA*). The model predicted that other pathways (besides pyruvate-formate lyase) could be used to produce acetyl-CoA. Acetyl-CoA could be made from acetaldehyde via acetaldehyde dehydrogenase (MhpF), where acetaldehyde is produced by threonine degradation and other reactions. Acetyl-CoA could also be produced by 2-amino-3-ketobutyrate CoA ligase (Kbl) from threonine degradation. However, all of the mutants were unable to grow in the absence of acetate, suggesting that these other pathways are not active at high enough levels. Acetate was consumed by all the pyruvate strains, except PYR001, presumably to generate acetyl-CoA by acetyl-CoA synthetase. The amount of acetate available (0.34 – 3.4 mM) was greater than or close to the amount acetyl-CoA needed for biomass (estimated the product of the biomass concentration and biomass requirement for 3.7 mmol acetyl-CoA per gDW)[99]. In ethanol production study, the mutants with increased fluxes of ethanol synthesis were observed to grow faster, which is also probably caused by the generation of acetaldehyde and then converted to acetyl-CoA, while another possibility is the balancing of NADH.

3.4.3 Reduced ethanol yield in one evolved strain

When the resulting pyruvate strains were re-engineered for ethanol production, three of the resulting strains achieved high ethanol yields (EH020-pflB, EH030-pflB and EH040-pflB) during fermentation. Deleting *pflB* and expressing the PET cassette increased ethanol as expected, except for EH010-pflB. EH010-pflB (derived from PYR010), had the lowest yield of the

mutants with *pflB* deleted and PET added. Among all the strains tested, EH010-pflB is closest genetically to GLBRCE1. Both EH010-pflB and GLBRCE1 have *ldhA*, *pta* and *pflB* deletions. Even though EH010-pflB has two additional deletions, *aceE* and *cyoA*, neither gene would be expected to be expressed anaerobically [100]. Thus, the significantly lower ethanol yield in EH010-pflB compared with GLBRCE1 was unexpected. GLBRCE1 was derived from a closely-related background strain (MG1655, compared to BW25113) and has an extra chromosomal copy of the PET cassette. This additional copy of the PET cassette could lead to higher PET expression levels and ethanol production in GLBRCE1. When compared to EH010, EH010-pflB should have reduced formate production (which it does, see **Figure 3.4**) and increased availability of pyruvate; however, EH010-pflB and EH010 exhibited similar ethanol yields. For the EH010-pflB strain, only 80% of the carbon was recovered in the biomass and measured products (which is lower than the other strains) and so it is possible that some other metabolite (not detected by HPLC) was secreted by EH010-pflB.

3.4.4 Survey of available pyruvate strains

Yeast and bacterial strains have previously been engineered for pyruvate production. Performance metrics for pyruvate producing strains reported over the last decade are compared to metrics for PYR020 in **Table 3.4**. Previous strains had volumetric production rates up to 1.2 g/L/hour with yields between 24% and 78% of the maximum theoretical production. The strains usually require additional nutrients besides glucose (e.g., yeast extract, tryptone, thiamine) which will increase the cost for commercial production. While PYR020 requires acetate for growth, acetate is commonly found in lignocellulosic hydrolysates. Currently, *E. coli*

TC44 has the best performance considering the pyruvate yields, production rate, and titer. Our strain, PYR020, uses only mineral salt medium and reaches significantly higher yield, but had lower titers and volumetric production rate because of the lower glucose concentrations used in the media. The next step is to investigate pyruvate productivity of these engineered strains in minimal medium with higher concentrations of glucose, or in cheaper hydrolysate feedstock under batch or fed batch process.

Figure 3.5 A: Strain PYR001, designed as Δ*aceE*, Δ*cyoA*, Δ*cydB*, Δ*pta*, Δ*eutI*, Δ*ldhA and* Δ*dld*

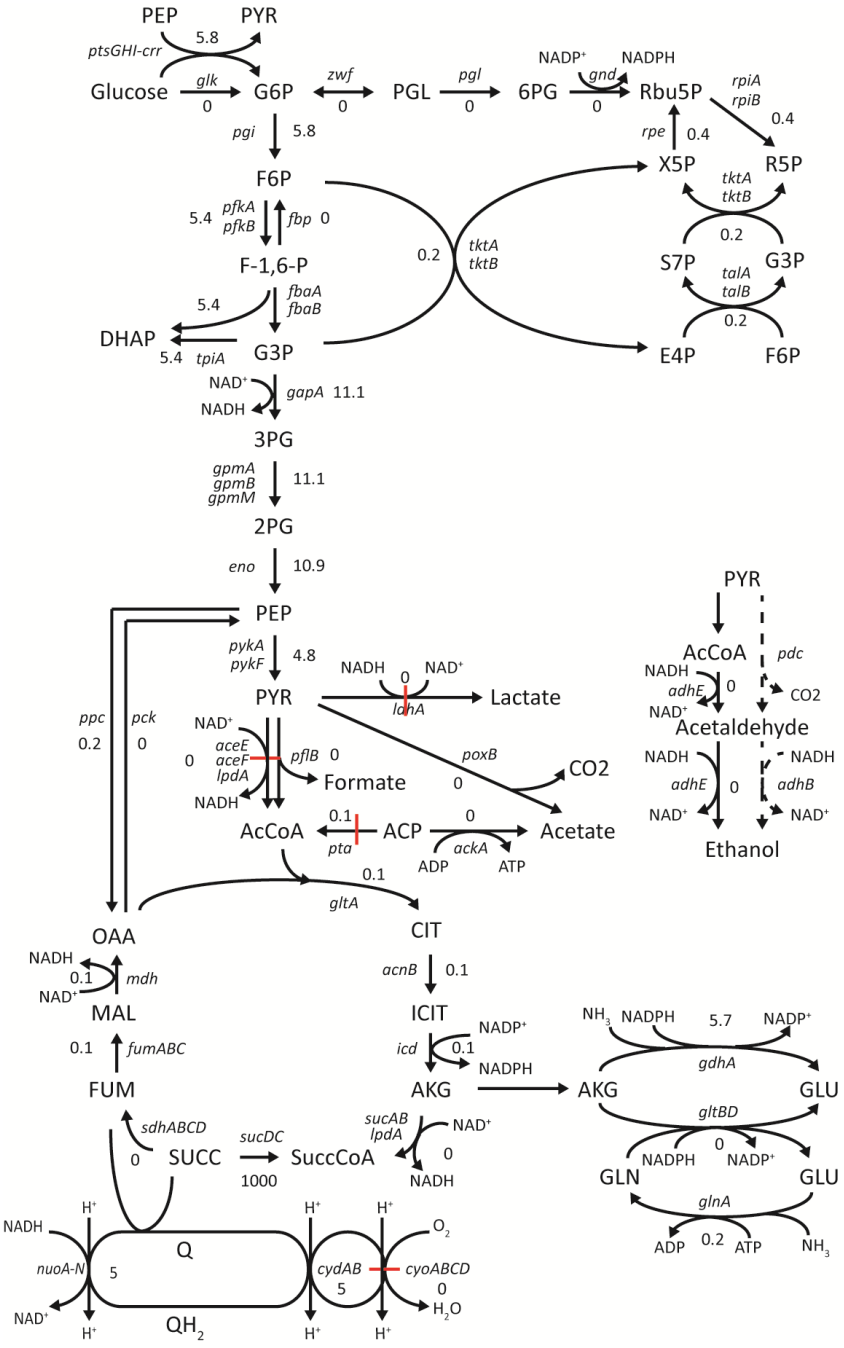


Figure 3.5 B: Strain PYR002, designed as Δ*lpdA*, Δ*gnd*, Δ*sdhA*, Δ*poxB*, Δ*pflB*, Δ*pflD*, Δ*tdcE* and Δ*purU*

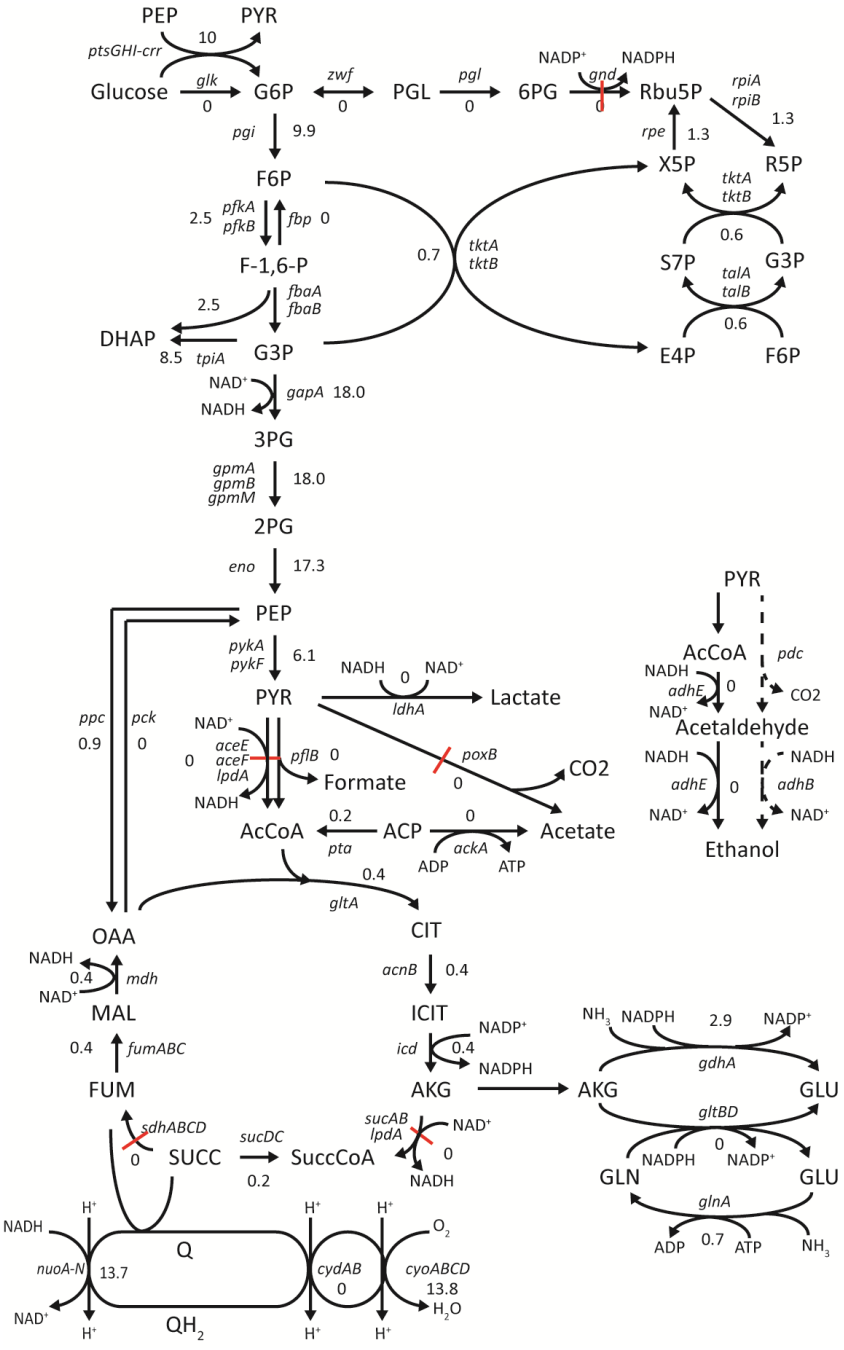


Figure 3.5 C: Strain PYR003, designed as Δ*aceE*, Δ*gdhA*, Δ*poxB*, Δ*ldhA*, Δ*dld*, Δ*atpE*, Δ*pflB*, Δ*pflD*, Δ*tdcE*

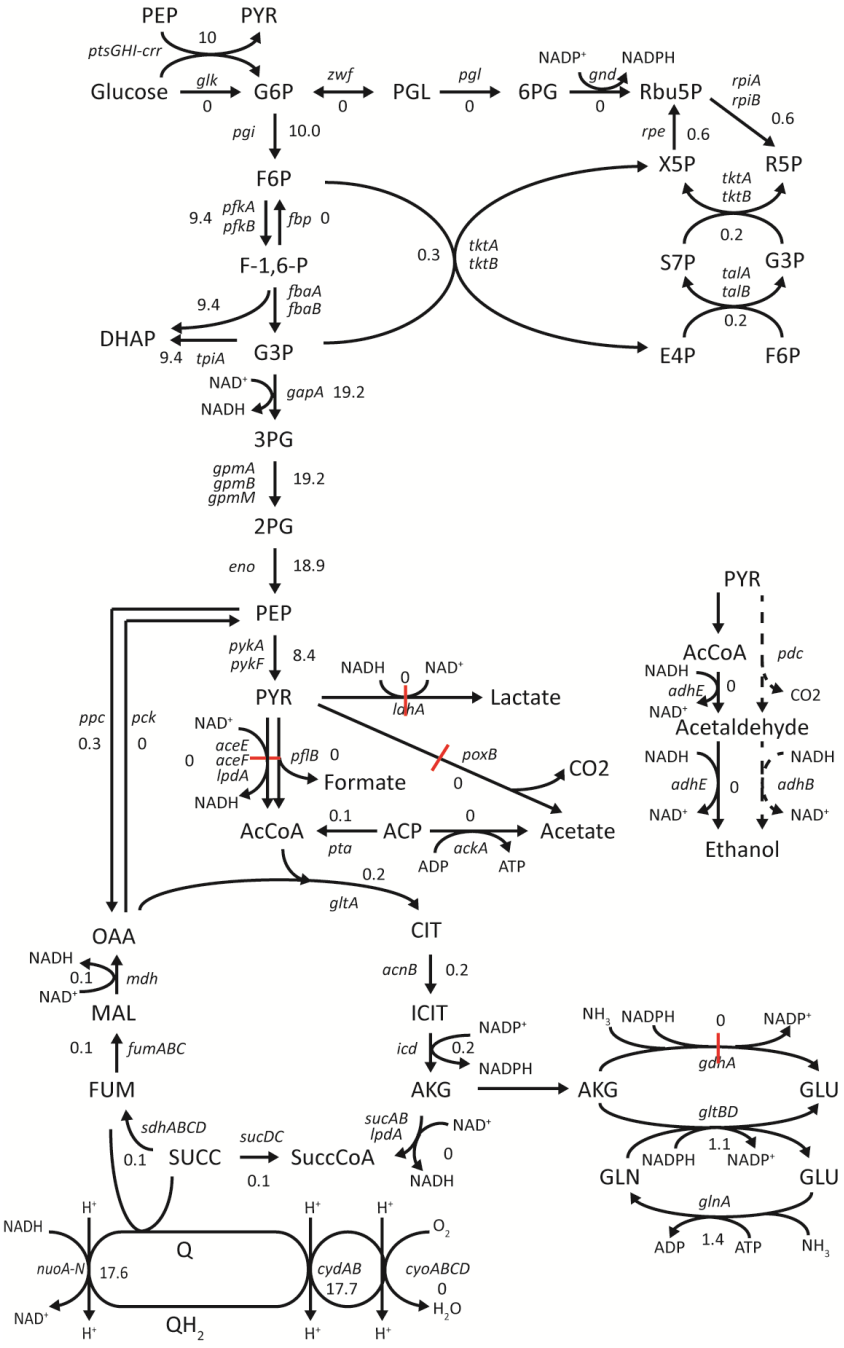


Figure 3.5 D: Strain PYR004, designed as Δ*aceE*, Δ*gnd*, Δ*poxB*, Δ*ldhA*, Δ*dld*, Δ*atpE*, Δ*pflB*, Δ*pflD*, Δ*tdcE*

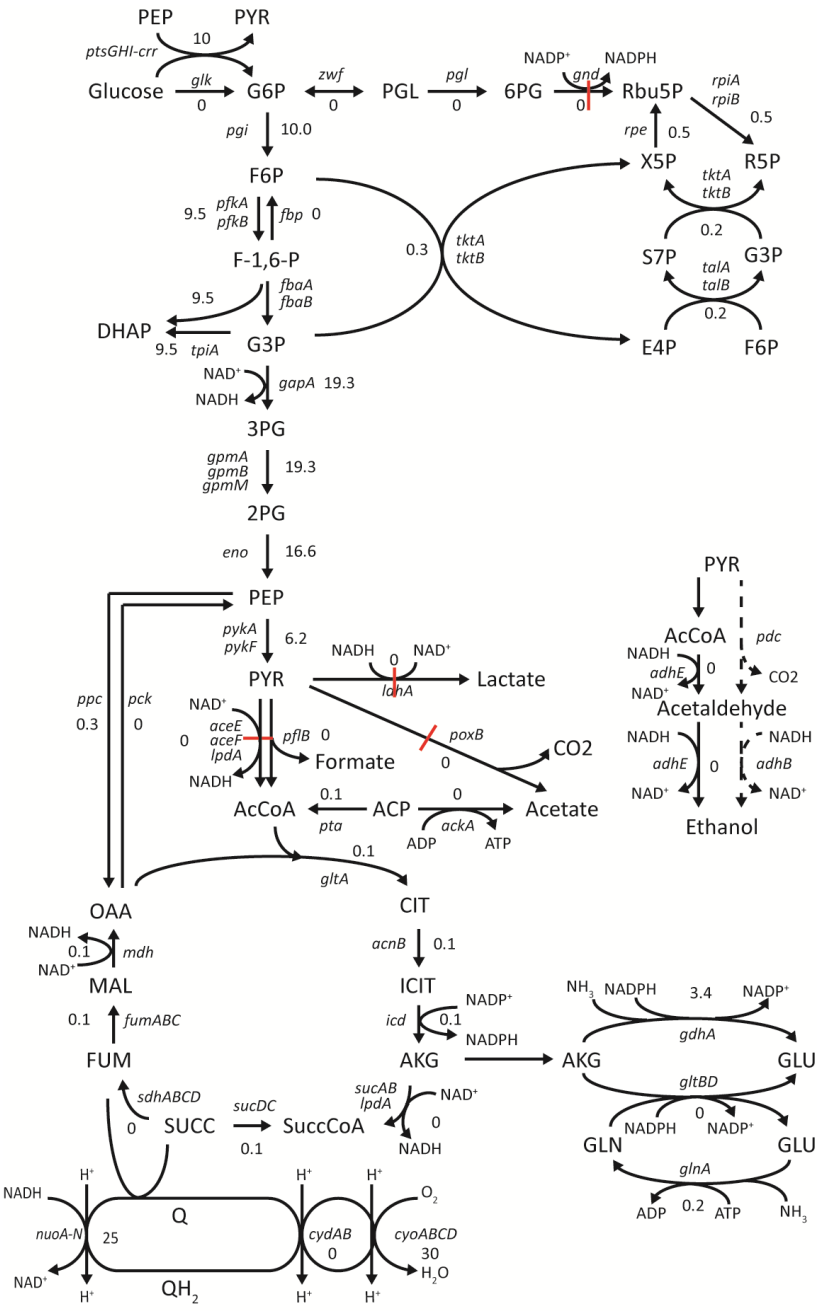


Figure 3.5. Central metabolic pathway of designed mutant strains for pyruvate production.
Genes associated with each reaction in the central metabolic network are shown. The reactions marked by red bars correspond to the deletion targets calculated by computational method. The labeled metabolic flux distribution for each strain was predicted by flux balance analysis. Glucose uptake was limited by 10 mmol/gDW/hour. Oxygen uptake was unlimited for strain PYR002, PYR003 and PYR004, but limited to 3 mmol/gDW/hour for strain PYR001.

Table 3.4 Existing engineered strains for pyruvate production.

Strains	Carbon and nitrogen Source	Genotype	Pyruvate yield (g/g)	Pyruvate titer (g/L)	Volumetric production rate (g/L/hour)	Fermentation time (hour)	Reference
E. coli TC44	60 g/L Glucose $(NH_4)_2HPO4$	*ΔpflB ΔfrdBC ΔldhA ΔadhE ΔsucA ΔackA ΔpoxB ΔatpFH*	0.76	52	1.2	43	[14]
S. cerevisiae FMME-002ΔTHI2	35 g/L Glucose NH_4Cl Thiamine	*MAT**a** leu2-3, 112 ura3-52 his3-Δ1 trp1-289 ΔTHI2*	0.23	8.21	0.09	96	[80]
E. coli CGSC791	40 g/L Glucose 3 g/L Acetate Tryptone $(NH_4)_2HPO4$	*ΔaceF ΔfadR ΔadhE Δppc*	0.65	35	0.97	36	[79]
C. glutamincum	30 g/L Glucose 10 g/L Acetate L-alanine $(NH_4)_2HPO4$	*ΔaceE Δpqo ΔldhA ΔalaT ΔavtA C-T ilvN*	0.59	17.6	0.32	55	[82]
S. cerevisiae G2U1-A_0	150 g/l Glucose Yeast extract Tryptone	*Δpdc udhA+*	0.63	75.1	0.63	120	[81]
E. coli PYR020	1.98 g/L Glucose 0.02 g/L Acetate NH_4Cl	*ΔlpdA ΔsdhA ΔpoxB Δgnd*	0.92	1.95	0.05	27	This study

Chapter 4 Discovery of non-native products produced from pyruvate and their synthesis pathways

4.1 Introduction

The goal of metabolic engineering is to manufacture chemicals in a cost-effective manner using living organisms. Identifying which metabolites are capable of being produced by microbes and what biosynthesis pathways are required to enable metabolite production are fundamental questions. *E. coli* is the most widely used host for metabolic engineering, since it is one of the best-studied microbes and it has a variety of advanced genetic, synthetic biology, and systems biology tools available. Currently, it is unknown how many non-native products *E. coli* could potentially produce by introducing heterologous enzymes and reactions. Metabolic databases, such as KEGG[101], provide a listing of known biochemical enzymes and their associated reactions and metabolites. Previously, heterologous KEGG reactions have been integrated with genome-scale metabolic network models of *E. coli* to identify missing *E. coli* metabolic reactions [102,103] and to identify a small set of heterologous reactions that need to be added to the *E. coli* network to optimize production of a desired native or non-native product [104,105]. While some analysis has been done for individual products, a broader characterization of the non-native metabolites that could be produced, including the pathways and central metabolic precursors needed to produce them, has not been done.

Approaches for finding pathways use either a known set of reactions (e.g., reactions from KEGG or a genome-scale reconstruction) or propose novel reactions. The BNICE

framework generates pathways using a set of enzyme reaction rules based on the enzyme classification system [106]. Since BNICE takes into account the chemical structure of reactants and products, it can suggest novel biosynthesis pathways for known and novel biological compounds. BNICE and related algorithms have been successfully applied to enable production of 1,4-butanediol and 1,2,4-trichlorobenzene [107,108]. Although BNICE is a well developed tool for producing novel biosynthesis pathways, it generates a large number of possible pathways that use novel biochemical reactions for which enzymes must be engineered.

Different approaches can be used to find a synthesis pathway(s) for products using a set of reactions. These approaches use optimization-based or graphical-based searches on a collection of known or hypothesized biochemical reactions. A pathway search specifies a starting and ending metabolite, such as a central metabolic precursor and a desired non-native product. Several metabolites are centrally located in metabolism, have high degrees of connection, and can be converted into a variety of other chemicals. Thirteen central metabolic precursors (including glucose-6-phosphate, glyceraldehyde-3-phosphate, 3-phosphoglycerate, pyruvate, oxaloacetate, and α-ketoglutarate) are used in *E. coli* to make all cellular components (lipids, proteins, RNA and DNA) [99]. These metabolic precursors can be considered as the starting points for synthesis pathways for most metabolites [107,109]. Usually many different pathways can convert a starting metabolite to a desired end-product, and enumerating alternative pathways is useful for comparing them to determine which would be best to produce a desired target molecule.

Graph-based methods have been used to find pathways from a starting metabolite to the target product. These graph-based methods represent metabolites as nodes and reactions

as edges, and the results will be sensitive to which edges are included in the network (e.g. edges between all reactants and products for a given reaction versus edges only between reactants and products that share carbon atoms). Unfortunately, these graph-based methods can identify paths which cannot be mass-balanced by the cell. In this case, the paths proposed produce (or consume) metabolites that cannot be consumed (or produced) by other parts of the network.

An alternative to graphical-based approaches is to use an optimization-based method. Optimization-based methods can incorporate network stoichiometry and rule out pathways that cannot satisfy mass balance constraints under steady state conditions [110]. More recently, another optimization based method, PathTracer, was developed [111]. PathTracer uses reaction stoichiometry to identify feasible pathways, but also eliminates undesired pathways with internal cycles. Optimization-based methods [110,111] can also only use connections between a reactant and a product if they share carbon atoms, so that the pathways that are proposed follow carbon through the network. .

In this study, we used the genomic-scale metabolic model of *E. coli* to identify all non-native products that could be produced in *E. coli*. For the subset of non-native products with commercial applications, we also identified pathways for producing these products from different central metabolic precursors. We were particularly interested in compounds that could be derived from pyruvate, since we have previously developed pyruvate overproducing strains which could serve as platforms to generate other chemical producing strains. The synthesis pathways from pyruvate to those non-native products were searched by PathTracer

algorithm. The pathways solutions provide a valuable resource for selecting pathways and target products for metabolic engineering.

4.2 Materials and methods

4.2.1 Maximum yield calculations

Reactions in the KEGG database that were not in the *E. coli* metabolic network were classified as heterologous reactions [101,112]. All 4,740 heterologous reactions utilized in this study were elementally balanced (by adding H and water as reactants or products), and represented 56% of the 8,452 reactions in the KEGG database. The heterologous reactions involve 4,725 metabolites, and 694 of these metabolites also participate in reactions included in the genome-scale metabolic model of *E. coli* iJO1366 which has a total of 1136 metabolites [112]. The heterologous metabolic reactions were combined with the *E. coli* metabolic reactions, to generate an integrated metabolic network with 5,167 unique metabolites and 6991 reactions. The maximal yield of each metabolite from glucose minimal medium under aerobic conditions by this integrated network was calculated using flux balance analysis (FBA) [47]. To calculate the maximum yield for a target metabolite using FBA, flux through a sink variable (y_i) was maximized, where production by the network reactions resulted in a positive sink variable. A steady-state mass balance constraint was imposed (Equation 4.1), which accounted for flux in the *E. coli* (defined as set Rxn), heterologous (defined as set $KEGG$) and sink reactions.

$$\sum_j (S_{i,j} \cdot v_j) - y_i = 0 \qquad \forall\, i \in Met, j \in Rxn \cup KEGG \qquad (4.1)$$

Here $S_{i,j}$ was the stoichiometric matrix representing the reaction stoichiometry for all metabolic reactions (j) in the Rxn and $KEGG$ sets. The set Met contained all metabolites (i) used in the *E. coli* and heterologous KEGG reactions. The v_j variable was the flux through a reaction (j). In addition to the mass balance constraint (Equation 4.1), enzyme capacity and thermodynamic constraints were also included in FBA:

$$v_{j,min} \leq v_j \leq v_{j,max} \qquad \forall\, j \in Rxn \cup KEGG \qquad (4.2)$$

$$0 < y_i \leq v_{max} \qquad \forall\, i \in Target \qquad (4.3)$$

$$y_i = 0 \qquad \forall\, i \in Met \backslash Target \qquad (4.4)$$

where $v_{j,min}$ and $v_{j,max}$ were the lower and upper bounds imposed on the fluxes. $v_{j,max}$ was set to 1000 mmol/gDW/hour and $v_{j,min}$ was set to 0 or -1000 mmol/gDW/hour for irreversible and reversible reactions, respectively. The lower limit for the ATP maintenance (ATPM) reaction was 3.15 mmol/gDW/hour. The maximum glucose uptake rate was set to 10 mmol/gDW/hour and oxygen uptake was unlimited. The set $Target$ contains the metabolite of interest whose maximum yield was being calculated. Metabolite production ($\sum_i y_i$) was maximized subject to the constraints in Equations 4.1-4.4.

4.2.2 Minimal number of heterologous reactions required

For the metabolites that could be produced aerobically from glucose using the integrated metabolic network, the minimal number of heterologous reactions required to produce at least 50% of the maximal yield was calculated. First, flux through the sink variable was constrained to be at least 50% of its maximal value found by FBA:

$$y_i \geq 0.5 \cdot M_i \qquad \forall\, i \in Target \qquad (4.5)$$

where M_i was a parameter corresponding to the maximal production rate for metabolite i calculated by FBA (Equations. 4.1-4.4). A binary variable z_j was used to indicate the flux status of the heterologous reactions. If the heterologous reaction j was active then z_j was equal to 1; however, if z_j was 0 then the heterologous reaction j was inactivated (Equation 4.6):

$$v_{j,min} \cdot z_j \leq v_j \leq v_{j,max} \cdot z_j \qquad \forall\, j \in KEGG \qquad (4.6)$$

A constraint on the maximum number of heterologous reactions that could be used was imposed:

$$\sum_j z_j \leq N \qquad \forall\, j \in KEGG \qquad (4.7)$$

where N was set to 20 in this study. The minimum number of heterologous reactions required to produce a target metabolite was found using the constraints shown in Equations 4.1, 4.4-4.7, while minimizing the objective (Obj) shown in Equation 4.8.

$$Obj = \sum_j z_j \qquad \forall\, j \in KEGG \qquad (4.8)$$

4.2.3 Finding paths between precursors and targets (PathTracer)

Synthesis pathways (involving heterologous and *E. coli* reactions) for non-native products from central metabolic precursors (e.g. pyruvate, oxaloacetate, α-ketoglutarate, glyceraldehyde-3-phosphate, and glucose-6-phosphate) were determined using the PathTracer algorithm [111]. First, the MapMaker algorithm was used to identify all the elemental transfers

between reactants and products of a reaction based on the metabolites' molecular formula. For example, a carbon transfer indicates how many carbon atoms are transferred between a reactant-product pair. The MapMaker results were used to create carbon transfer mappings between reactants and products only if carbon atoms are transferred between them. MapMaker was run for all *E. coli* and heterologous reactions included in the integrated metabolic network. Carbon maps involving currency metabolites (e.g. CO_2 and ATP) were omitted (see **Appendix 1** for complete listing).

The carbon maps generated by MapMaker were used by the PathTracer algorithm to identify pathways connecting precursor metabolites to non-native products. PathTracer was formulated as a network flow problem where the metabolites were nodes connected through edges corresponding to the carbon maps generated by MapMaker. PathTracer determined the shortest path from a specified precursor to the target metabolite using *E. coli* and/or the heterologous reactions from KEGG database. Alternative paths connecting the two metabolites were also found using integer cut constraints (either reaction or path cuts) [111].

Importantly, the PathTracer algorithm guaranteed a path was feasible by applying mass balance (Equation 4.1), enzyme capacity and directionality constraints (Equation 4.2), and the following constraints (Equations 4.9-4.12):

$$v_j \geq v_{min} \qquad \forall\, j \in Path \cup Forward \qquad (4.9)$$

$$v_j \leq -v_{min} \qquad \forall\, j \in Path \cup Reverse \qquad (4.10)$$

$$v_j = 0 \qquad \forall\, j \in KEGG \backslash Path \qquad (4.11)$$

$$\mu \geq 0.01 \qquad (4.12)$$

where v_{min} was a parameter, set to 0.001 in this study. The set $Path$ contained the reactions (from Rxn and $KEGG$) in a proposed path. The reactions moving in forward and reverse directions in a proposed path were included in sets $Forward$ and $Reverse$, respectively. Equations 4.9-4.11 ensured that all reactions in a selected path were active (i.e., have non-zero flux) and that heterologous reactions not included in the path were inactive (i.e., have zero flux). Equation 4.12 imposed a minimal growth rate (μ).

The net reaction for a path was determined from the stoichiometric coefficients of the reactions in the path (Equation 4.13). The net reaction coefficient for the starting precursor metabolite was constrained to be negative (Equation 14), in order to prevent paths from being proposed where the precursor started the path but was later produced by the path. These two additional constraints were formulated as:

$$S_i^{net} = \sum_{j \in Path} S_{i,j} \tag{4.13}$$

$$S_i^{net} < 0 \qquad \forall\, i \in StartMet \tag{4.14}$$

where S_i^{net} was the coefficients for the metabolites in the net reaction. The set $StartMet$ referred to the starting precursor metabolite. To ensure that a reaction or metabolite was only used once in a path, PathTracer was run using reaction blocking and loop killing constraints (see [111] for details).

A variation of the PathTracer algorithm was implemented by including the constraints described above (Equations 4.1, 4.2, 4.4, 4.5, 4.9-4.14) and reaction blocking and loop killing constraints, into the basic algorithm (see [111] for details). PathTracer found the shortest and alternative paths from a starting precursor metabolite to a desired product. PathTracer was

implemented in GAMS (GAMS Development Corporation, Washington, DC) and solved with the CPLEX solver with a default CPU-time limit of 1000 s. The results were subsequently filtered to ensure that reactants in the heterologous reactions used in a path were available (i.e., reactants were part of iJO1366 or produced by other reactions in the path).

4.3. Results

4.3.1 Non-native products that could be produced by *E. coli*

The maximal yields for 5,167 unique metabolites were calculated using FBA for an integrated metabolic network containing *E. coli* reactions and 4,740 heterologous reactions from KEGG. Under glucose aerobic conditions, 2,510 metabolites (49% of total metabolites) could be produced by the integrated network. The integrated model could not produce all the metabolites (including 265 out of 1,136 *E. coli* metabolites and 2,392 out of 4,031 non-native metabolites), which could be due to network gaps (e.g., unknown routes for precursor synthesis or by-product degradation), reaction directionality, or medium conditions (e.g., only produced from carbon/nitrogen sources).

Among the 2,510 metabolites that could be produced, 871 participated in *E. coli* reactions (77% of iJO1366 metabolites) and the remaining 1,639 participated only in heterologous reactions from the KEGG database. For each producible metabolite, the minimum number of heterologous reactions needed to achieve at least 50% of the maximal yield (for the given producible metabolite) was determined. Metabolites that required heterologous reactions for production were classified as non-native products, while metabolites that did not require any heterologous reactions were classified as native products. Of the 871 producible

E. coli metabolites, 716 were made using only *E. coli* metabolic reactions (i.e. native products), while 155 required heterologous reactions indicating they were non-native products. A total of 1,793 non-native products (including 155 *E. coli* metabolites and 1,638 KEGG metabolites) needed a minimum of 1 to 16 heterologous reactions to allow their production (**Figure 4.1**); one additional non-native product (sinapine) required more than 20 heterologous reactions to be produced. This set of 1,793 metabolites will be referred to as the set of non-native products. Of the non-native products, 29% of them (522 metabolites) needed only one heterologous reaction to enable production, and 35% (631 metabolites) required only two or three heterologous reactions. On the contrary, very few non-native products (<4%) required more than ten heterologous reactions. These results indicate that the *E. coli* metabolic network is suit to make many non-native metabolites with only few additional non-native reactions.

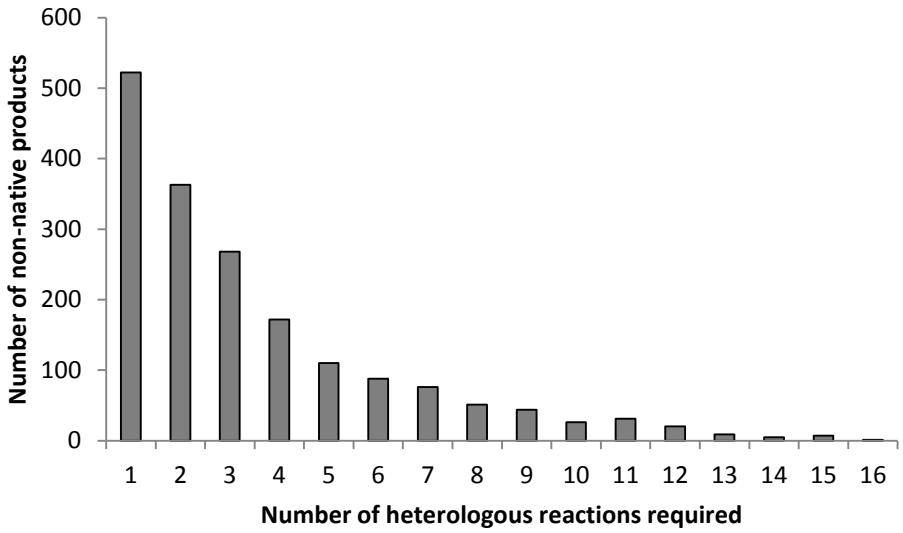
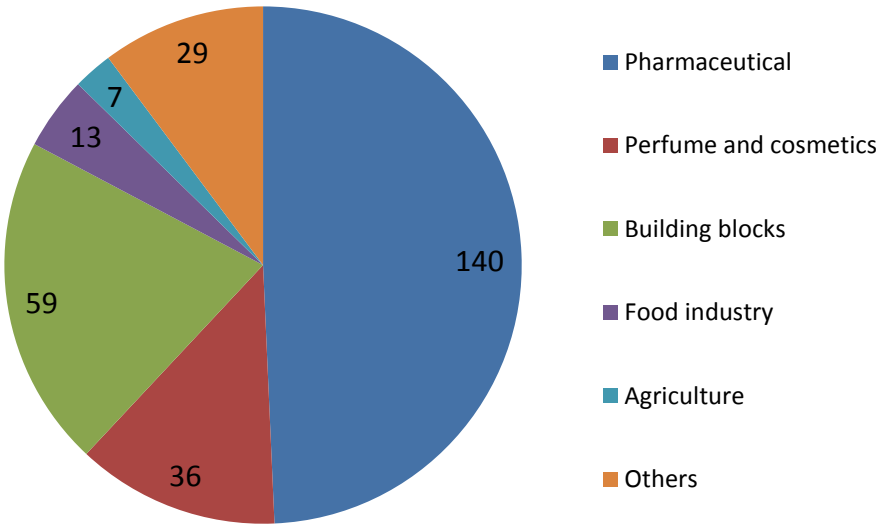


Figure 4.1: Minimal heterologous reaction requirements for the non-native products.
The minimal number of heterologous reactions needed to produce each product at 50% of its maximum theoretical yeild was calcualted. The minimal number of heterologous reactions required for producting 1,793 non-native products ranged between 1 and 16. One additional non-native product required over 20 heterologous reactions.

The 1,793 non-native products include many metabolic intermediates which do not have any commercial value. Before identifying paths between precursors and desired targets, which is a time-consuming process, we first identified how many of the 1,793 non-native products had commercial applications. Five databases were queried, including CAS (Chemical Abstracts Service), WikiPedia, DrugBank, Sigma-Aldrich and KEGG, to gather information of commercial availability, general usage, drug data, retail price and chemical structure. Of the 1,793 non-native products, 284 were manually confirmed to have applications in a variety of industries, including pharmaceuticals, food industry, cosmetic and perfume, agriculture, manufacture and others (**Figure 4.2A**). According to the CAS database, 279 of the 284 metabolites are commercially available. The 284 valuable non-native products could be made using *E. coli* by expressing enzymes to catalyze 1 to 16 heterologous KEGG reactions. About 27%

A



B

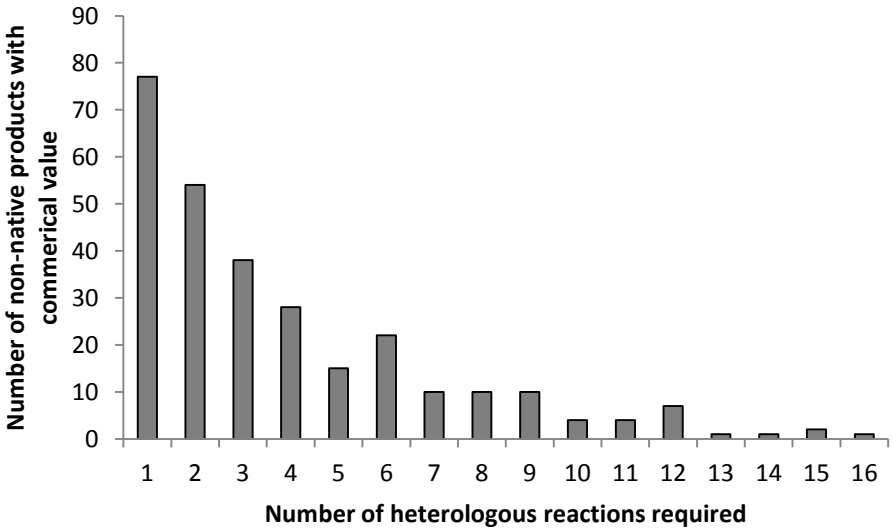


Figure 4.2: Commerical applications for non-native products and the non-native reaction requirements for produciton.

(A) The primary applications for the 284 non-native products. (B) The minimal number of heterologous reactions required for producing the valuable non-native products.

of these valuable non-native products required only one heterologous reaction, while 33% needed two or three reactions (**Figure 4.2B**). Additional details about the 284 valuable non-native products can be found in **Appendix 2.**

4.3.2 Paths to valuable non-native products from pyruvate

After identifying commercial applications for 284 non-native products, paths containing *E. coli* and heterologous reactions from pyruvate to these valuable compounds were found using PathTracer. Pyruvate was chosen as a precursor since it is part of central metabolism and over-producing strains have been developed (Chapter 3). Out of the 284 non-native products, 64 were found to be within 5 reactions of pyruvate. The shortest path and alternative paths (with equal or greater length) were identified, and the number of different PathTracer solutions for the 64 metabolites varied between 1 and 72 (**Figure 4.3**). Fourteen of the 64 metabolites (~22%) had 5 or fewer paths from pyruvate, and ~69% had fewer than 20 different paths from pyruvate. The small number of paths for most metabolites made it easy to evaluate the individual solutions. Some of the 64 metabolites have been produced previously using *E. coli.* The following sections compares the PathTracer identified paths to the strategies used previously for engineering *E. coli* to produce 2,3-butanediol, 1-propanol and acrylic acid.

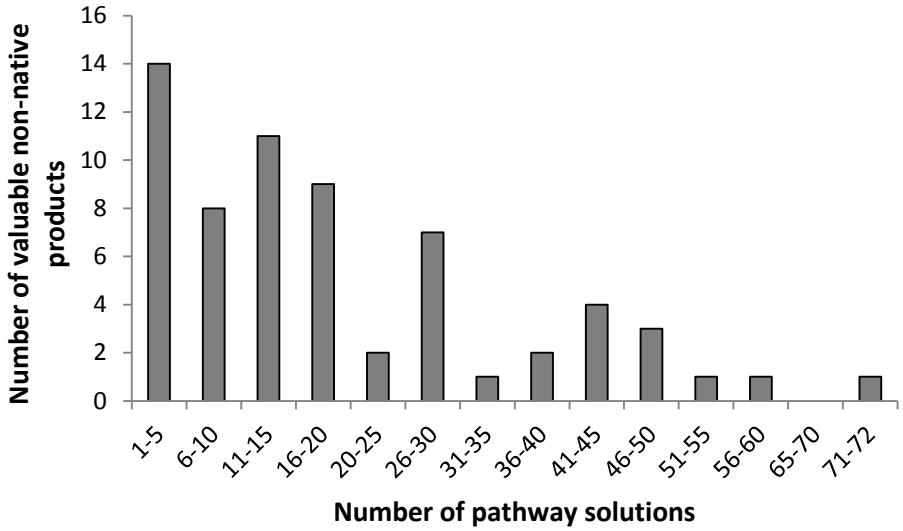


Figure 4.3: The number of alternative paths to valuable non-native products from pyruvate.
PathTracer algorithm identified numerous paths with less than or equal to 5 reactions for producing 64 non-native products from pyruvate. The metabolites were divided into groups based on how many alternative paths were found.

4.3.3 Pathways for 2,3-butanediol synthesis

2,3-butanediol (2,3-BD) is a bulk fuel and industrial solvent, and is also used in manufacturing plasticizers, inks and explosives. Yeast, *E. coli,* cyanobacteria and *Klebsiella pneumoniae* have been metabolically engineered to produce 2,3-BD [113-118] using similar synthesis routes for 2,3-BD (**Figure 4.4A**). One approach converted two molecules of pyruvate into α-acetolactate which was decarboxylated to produce R-acetoin, which was converted into (R,R)-2,3-BD. α-acetolactate can also be spontaneously converted into diacetyl. Another approach used diacetyl reductase to convert diacetyl into R-acetoin or S-acetoin, which were then converted into (R,R)-2,3-BD or (S,S)-2,3-BD, respectively. The necessary enzymes (α-acetolactate decarboxylase, (R,R)-butanediol dehydrogenase, diacetyl reductase and (S,S)-butanediol dehydrogenase) are found in many organisms including yeast,

Enterobacter aerogenes, *Klebsiella pneumoniae* and *Bacillus subtilis*, but not in *E. coli*. PathTracer found all of the enzymatic reactions that have been used previously for 2,3-BD synthesis (**Figure 4.4B**). The spontaneous (non-enzymatic) reaction from α-acetolactate to diacetyl was not included in the model and thus not found as a pathway solution. In addition, PathTracer found two alternative reactions using different co-factors (NAD^+ or $NADP^+$) for converting R-acetoin to diacetyl, and an additional reaction from R-acetoin to S-acetoin, which could participate in additional routes for (S,S)-2,3-BD. All predicted paths had the same maximal yield (1.08 mol/mol glucose). No other pathways involving known enzymes were discovered for synthesizing 2,3-BD.

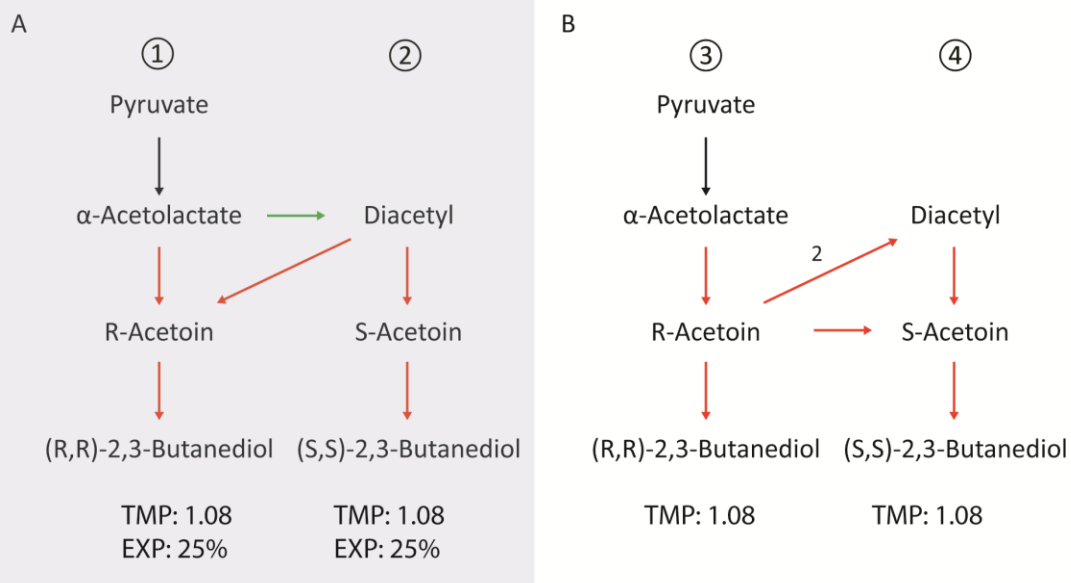


Figure 4.4: Synthesis pathways for 2,3-Butanediol.
The pathways that have been implemented in different species are in grey area, shown in (A). The black and red arrows mean *E. coli* and heterologous reactions, respectively. The green arrow indicates the reaction is spontaneous and not included in the integrated metabolic model. The predicted pathways for the production of 2,3-BD are shown in (B). The number above a reaction indicates the number of reactions that can connect two metabolites. The theoretical maximum yield using each pathway was calculated and labeled at the bottom. TMP: theoretical maximum yield; EXP: reported experimental yield.

4.3.4 Pathways for 1-propanol synthesis

1-propanol is used as a liquid fuel and industrial solvent, and is used for manufacturing drugs and cosmetics. Microbial production of 1-propanol has been engineered in *E. coli*, *Propionibacterium freudenreichii* and *Thermobifida fusca* using three pathways [119-124]. The first route, converts pyruvate into α-ketobutyrate (by the branched-chain amino acid biosynthesis pathway), and the α-ketobutyrate is converted to 1-propanol using an α-keto-acid decarboxylase and an alcohol dehydrogenase (**Figure 4.5A**, path ①) [119,120]. The engineered *E. coli* strain with this first route has the highest reported yield for 1-propanol [120]. Unfortunately, the α-keto-acid decarboxylase reaction is not included in KEGG database and so PathTracer was unable to find this solution.

The second route starts by forming α-ketobutyrate as well (**Figure 4.5A,** path ②) [121]. Propionyl-phosphate is produced from α-ketobutyrate spontaneously and converted to propionate by a propionate kinase (encoded by *tdcD*). Then acetyl-CoA synthetase produces propionyl-CoA, and the propionyl-CoA is converted into propionyl-aldehyde and then 1-propanol. The enzymes needed for this pathway are all present in *E. coli*; however PathTracer only found the connections between propionate to 1-propanol (**Figure 4.5B**, path ④), because the spontaneous reaction from α-ketobutyrate to propionyl-phosphate was not part of the integrated model. PathTracer identified several other sets of reactions to generate propionate and propionyl-CoA from pyruvate.

The third route for producing 1-propanol uses methylglyoxal synthase to convert dihydroxyacetone-phosphate into methylglyoxal (**Figure 4.5A**, path ③). Methylglyoxal is then

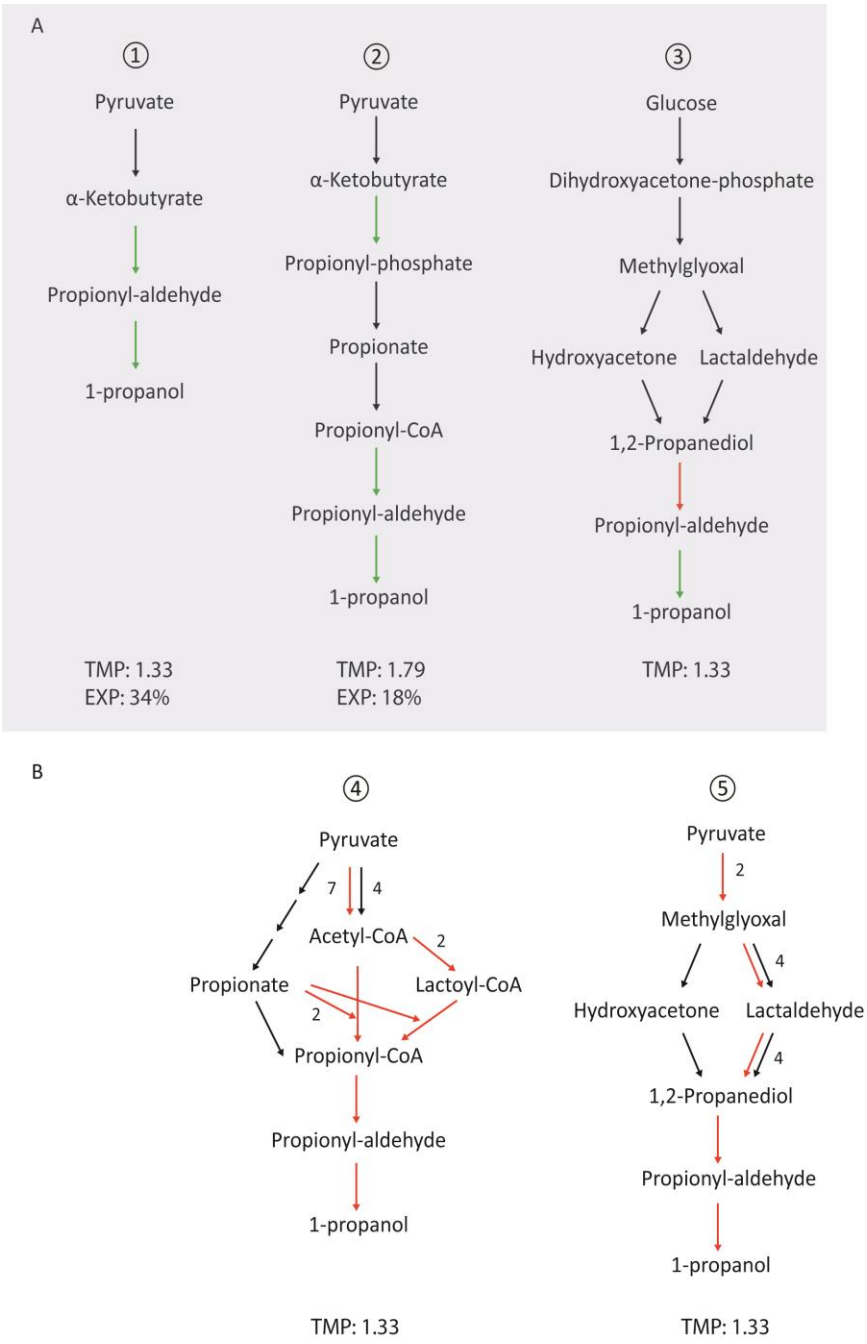


Figure 4.5: Synthesis pathways for 1-propanol.

The pathways that have been implemented in different species are in grey area, shown in (A). The black and red arrows mean *E. coli* and heterologous reactions, respectively. The green arrow indicates the reaction is either spontaneous or recently found occurring *E. coli*, but not updated in the integrated metabolic model. The predicted pathways for the production of 1-propanol are shown in (B). The number above a reaction indicates the number of reactions that can connect two metabolites. The theoretical maximum yield using each pathway was calculated and labeled at the bottom. TMP: theoretical maximum yield; EXP: reported experimental yield.

converted into 1,2-propanediol, using either hydroxyacetone or lactaldehyde as an intermediate, 1,2-propanediol is then be converted into propionyl-aldehyde, and the propionyl-aldehyde is used to make 1-propanol. Similar routes were predicted by the PathTracer algorithm (**Figure 4.5B**, path ⑤), except methylglyoxal was produced from pyruvate using two heterologous reactions and more alternative reactions could convert methylglyoxal to lactaldehyde and then to 1,2-propanediol.

The maximum theoretical yields and experimental yields for 1-propanol using the three experimentally implemented pathways were calculated (**Figure 4.5**). Surprisingly, the second route which uses only *E. coli* reactions to make 1-propanol from α-ketobutyrate showed the highest theoretical yield, but had a lower experimental yield than the first route.

4.3.5 Pathways for acrylic acid synthesis

Acrylic acid is a commodity chemical with a global production of 4.7 million tons in 2012 [125]. Acrylic acid and its esters are widely used for manufacturing plastics, coatings, paints and adhesives. Several *Clostridium propionicum* strains have been engineered to produce acrylic acid, however the yield is very low [126-128]. To date, two pathways have been used to make acrylic acid, one via lactate and the other via 3-hydroxypropanoate (3-HP). Both pathways were found by the PathTracer algorithm (**Figure 4.6**). A variety of *E. coli* and heterologous reactions can be used to produce lactate from pyruvate and lactate can be converted to acrylic acid in three steps using heterologous reactions. The overall reaction consumes one pyruvate and one NADH to generate acrylic acid and one molecule of NAD^+.

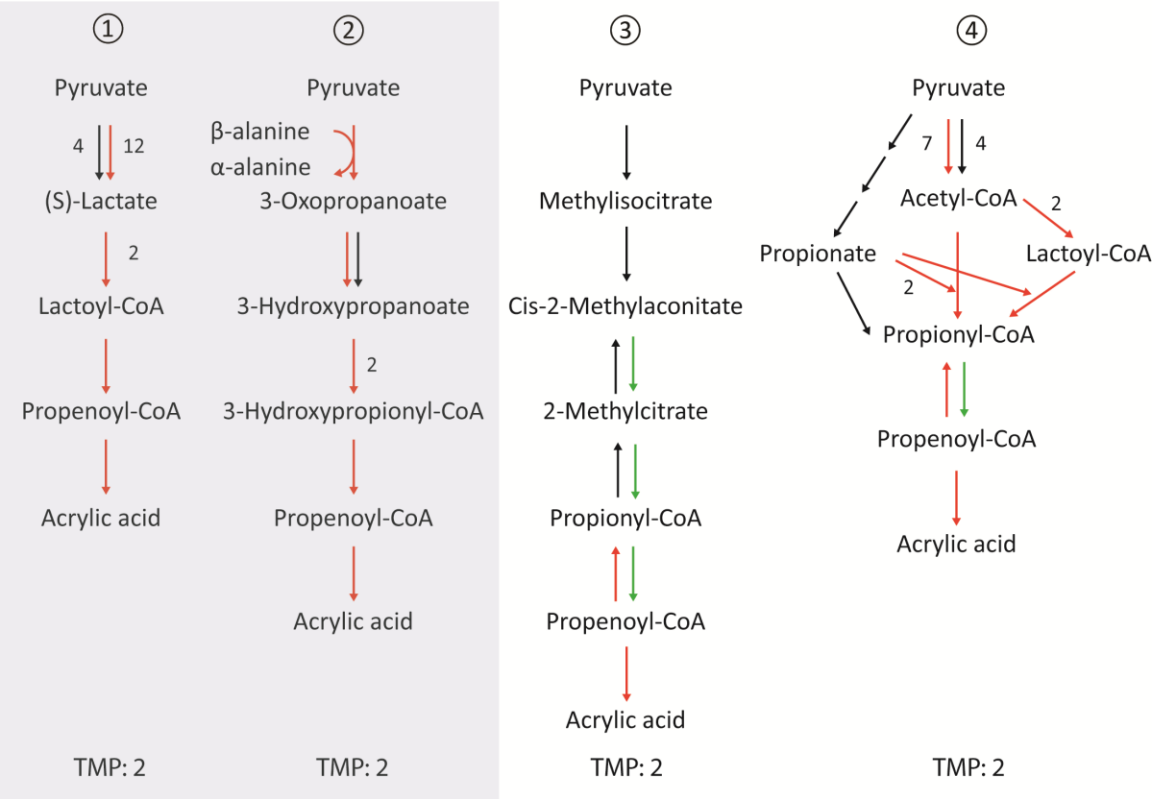


Figure 4.6: Synthesis pathways for acrylic acid.
Four pathways were predicted by model and two of them have been implemented in different species, shown in the grey area. The black and red arrows mean *E. coli* and heterologous reactions, respectively. The green arrows indicate the reactions have to be changed to forward direction, in order to find this pathway. The number above a reaction indicates the number of reactions that can connect two metabolites. The theoretical maximum yield (TMP) using each pathway was calculated and labeled at the bottom.

The second pathway that has been used experimentally starts by converting pyruvate into 3-oxopropanoate using an additional reactant, β-alanine (which can also derived from pyruvate). 3-oxopropanoate is reduced to 3-HP and then converted to acrylic acid in three additional steps. The two pathways have the same maximal yield; however the 3-HP pathway required more reactants in the overall reaction, including pyruvate, acetate, β-alanine, ATP,

NADH and CoA. Thus, it may take more engineering effort to increase flux through the 3-HP pathway as compared to the lactate pathway.

A third pathway using methylcitrate as an intermediate has been proposed previously (**Figure 4.6**) [129]; however, PathTracer did not find it because three reactions would need to go in the thermodynamically opposite direction. When the three reactions were changed to reversible reactions in the integrated model then a path involving 6 reactions was found, where the overall reaction was: pyruvate + succinate + CoA + $NADP^+$+ acetate → oxaloacetate + NADPH + acetyl-CoA + acrylic acid. Allowing the reaction from propionyl-CoA to acryloyl-CoA to be reversible in the model, allowed another pathway to be found. This last pathway resembled one of the 1-propanol synthesis pathways (**Figure 4.5B**), where propionyl-CoA was converted into propenoyl-CoA and then acrylic acid. These last two pathways involve thermodynamically unfavorable reactions, and detailed experiments of studying the thermodynamic parameters need to do before utilizing the pathways for the acrylic acid production.

4.3.6 The properties of pathway solutions for producing valuable non-native products

The paths for generating valuable non-native products from pyruvate were further analyzed to (1) evaluate their dependence on heterologous reactions and (2) assess where in the paths alternative reactions can be used. If strains were developed that could make precursors at high rates and yields, then these could be modified to make specific chemicals of interest. While shorter synthesis pathways might be desirable since there are fewer parts to control, it is also important to consider the number of heterologous genes that would need to be introduced in the host (which can have advantages and disadvantages). To find out which

products might be the easiest to make, the length of the shortest pathway to non-native products from pyruvate was plotted against the number of heterologous reactions required by the pathway (**Figure 4.7**). For this analysis only the 64 non-native products that could be produced from pyruvate using 5 or less reactions were considered. The closest non-native product to pyruvate was 2-butynedioic acid, which was produced in one step and with one heterologous reaction. The furthest non-native product was 1,3-propanediol, which was made from pyruvate in 5 steps using all heterologous reactions. The majorities of non-native products were within 3-4 steps of pyruvate and required 1-3 heterologous reactions to enable their production.

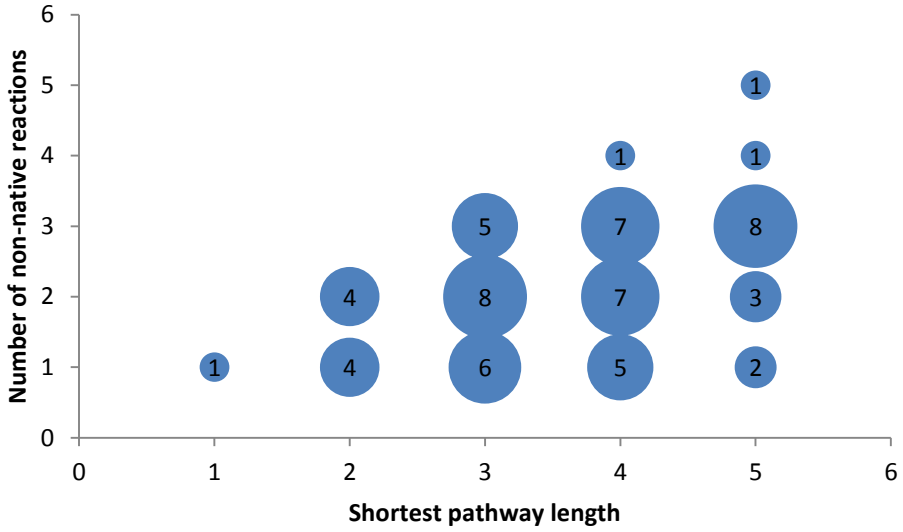


Figure 4.7: The number of heterologous reactions vs the shortest synthesis pathway.
The 64 metabolites able to be produced from pyruvate in 5 steps were grouped by the length of their shortest pathway and the number of heterologous reactions present in the shortest pathway. The size of the bubble reflects the number of metabolites. The number of metabolites in each group was labeled in the center.

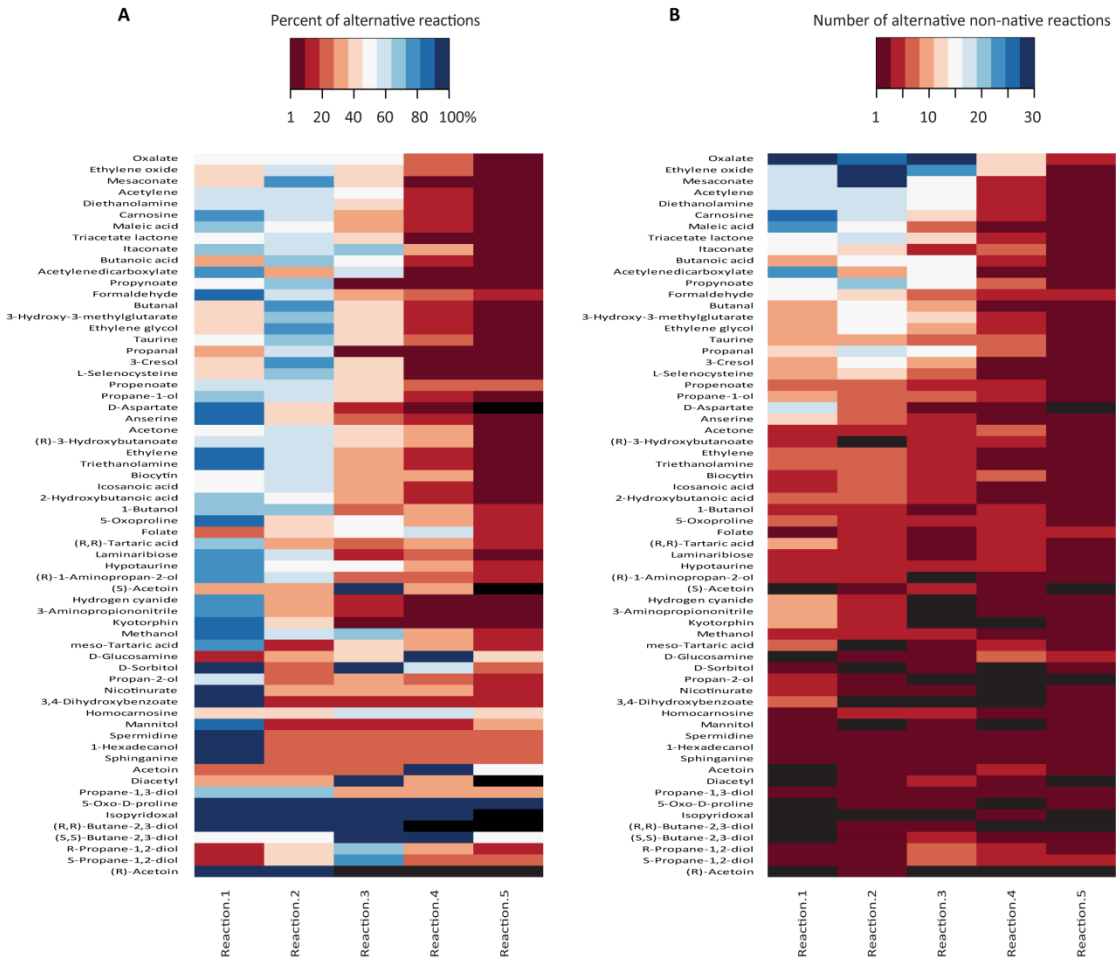


Figure 4.8: The pattern of alternative relations in all pathway solutions within 5 steps.
(A) For a product in the list of 64 valuable non-native products produced from pyruvate, the number of different reactions (*E. coli* or heterologous) appearing in each step of its synthesis pathway was normalized by the number of pathway solutions. (B) The number of alternative heterologous reactions present in each step of its synthesis pathway was calculated. The black color indicates no reaction (heterologous and/or *E. coli*) in that position.

Expressing a heterologous enzyme in a host organism can be problematic. Correct expression and maturation is not always guaranteed and heterologous enzymes could be toxic to the host. Reducing the number of heterologous reactions needed would help avoid some of these problems, as well as using reactions with alternatives that could be used instead if

enzymes for part of the pathway are difficult to express. For each of the 64 valuable non-native products, the number of alternative heterologous reactions that could be used at each reaction step of the synthesis pathways was determined (**Figure 4.8B**). The non-native products shown at the top had the most alternative reactions, which included oxalate, ethylene oxide, and mesaconate.

We also investigated how many alternative reactions exist in each step across all identified pathways for synthesizing a product of interest. For each of the 64 non-native products within 5 steps of pyruvate, the number of different reactions (**Figure 4.8A**) used at each step was normalized by the number of paths found. For all products evaluated, the first several steps in the pathways had multiple equivalent reactions and the number of alternative reactions in the last step was less than that in at least one of other steps. It is not surprising, because the 64 metabolites are not essential metabolites, only existing in several organisms and certainly have fewer production reactions.

4.3.7 Alternative precursors for producing valuable non-native products

Of the 284 valuable non-native products, only 64 could be produced from pyruvate within 5 reactions. The remaining non-native products required longer pathways for their production and might be better produced from other central metabolic precursors. In order to discover pathways for producing other valuable non-native products additional central metabolic precursors were considered, including glucose-6-phosphate, glycerate-3-phosphate, glyceraldehyde-3-phosphate, oxaloacetate, and α-ketoglutarate. These precursors are the hubs in the metabolic networks of many different organisms [109]. Paths for producing a non-native

product from each precursor were found and a total of 81 valuable non-native products were found to be within 5 steps of the 6 different precursors considered (**Figure 4.9**). The 6 precursors could be converted into each, and thus many non-native products could be produced from several precursors. Pyruvate, oxaloacetate and α-ketoglutarate were the most common precursors for which non-native products could be made within 5 steps. They showed a highly overlapping profile of metabolites that could be produced within 5 steps from any of the three (**Figure 4.10**). For a given non-native product, the shortest pathways from each of the 6 starting precursors were compared to identify the closest precursor. The precursor that could produce the non-native product with the fewest reactions was considered as the closest precursor. Pyruvate was shown to be the closest precursor for 40 of the 64 non-native metabolites that could be synthesized from it. While oxaloacetate could produce more non-native products in 5 steps than any of the other precursors tested, it was only the closest precursor for 27 of the 68 non-native products that could be produced from oxaloacetate in 5 steps. Glyceraldehyde 3-phosphate uniquely supported the production of one non-native product and was that product's closest precursor. Additional metabolite hubs should be explored as potential precursors (e.g., succinate, L-glutamate, L-aspartate [109]), as well as top value building blocks (e.g., glycerol [130]).

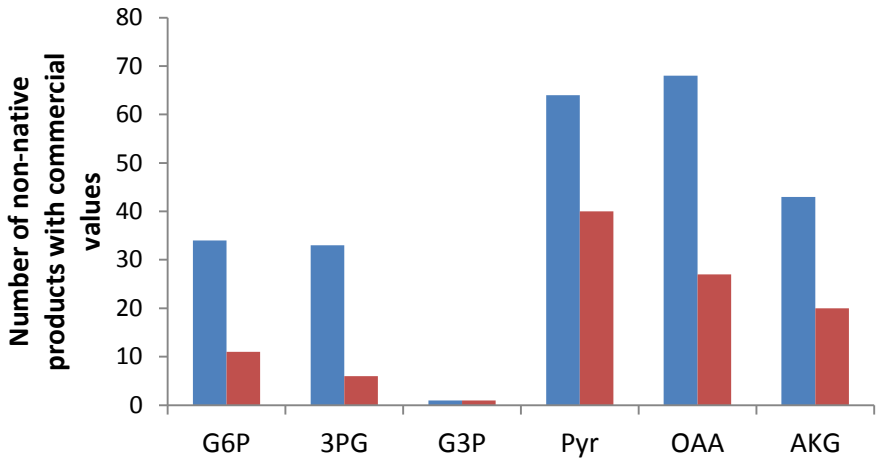


Figure 4.9: Alternative precursors for the production of valuable non-native products.

The number of products able to be produced from the precursors within 5 steps is indicated by blue column. The red columns indicate that synthesis from that precursor is shortest among the six precursors. Seventeen products could be produced from two or more precursors with fewest reactions. Metabolite abbreviations: G6P: glucose-6-phosphate; 3PG: D-glycerate-3-phophate; G3P: D-glyceraldehyde 3-phosphate; Pyr: pyruvate; OAA: oxaloacetate; AKG: α-ketoglutarate.

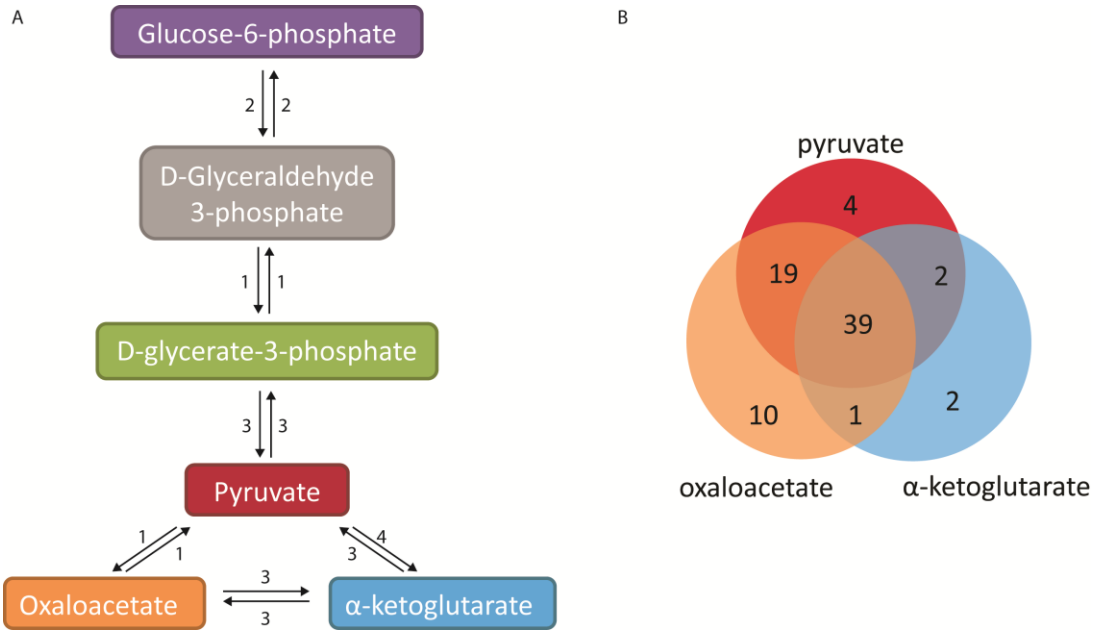


Figure 4.10: Six precursors for the synthesis of metabolites.

(A) The six precursors are in the central metabolism. The number of reactions converting two precursor metabolites was predicted using the integrated metabolic model of *E. coli* and heterologous reactions. (B) shows the number of valuable non-native products produced from any of the three precursors.

4.4 Discussion

This study was motivated by two practical questions regarding metabolic strain engineering: (1) what valuable non-native biochemicals can be produced from a given metabolic precursor, and (2) what precursors should be produced to synthesize a variety of non-native products. To answer these questions, a comprehensive investigation of the production capabilities of *E. coli* as a background host was performed. Considering all the heterologous reactions in KEGG, *E. coli* could be engineered (by adding up to 16 heterologous reactions) to produce 1,793 non-native products (more than double the number of native products). Of these 1,793 non-native products, 284 were confirmed to have commercial value. Using pyruvate as a starting precursor, we used PathTracer to identify 64 valuable non-native products within 5 reaction steps. Pyruvate production strains have been engineered previously (described in Chapter 3), and these 64 non-native products would be good targets for re-engineering these pyruvate strains to produce other valuable chemicals. By identifying alternative paths for synthesizing non-native products, potential engineering strategies can be compared based on requirements for non-native reactions, length of pathways from central metabolic intermediates, maximal theoretical yields, and pathway requirements with regard to reactants and co-products in the overall reactions.

Comparisons between predicted pathways and experimentally implemented pathways showed that, while substantial overlap exists, more pathway variations are possible and expanded reaction databases are needed. PathTracer found many alternative reactions for connecting pathway intermediates that could be used to improve experimental yields. The PathTracer results depend on the completeness and accuracy of the heterologous reaction

database and host metabolic network being considered. PathTracer missed some experimentally used pathways that involved spontaneous and recently characterized reactions. Recent efforts to build larger and more comprehensive reaction databases [131], as well as tools to identify novel putative reactions [106,132,133], will expand the number of non-native products that could be potentially produced as well as the number of different pathways that can be used to generate these products.

Besides pyruvate, five other central metabolites were considered as starting precursors to synthesize 284 non-native products with commercial value. When only 5 reaction steps were allowed, 81 out of the 284 products could be produced from at least one of the precursors considered, while the rest 203 products could not be produced. One possible reason is that their biosynthesis pathways have more than 5 reaction steps from the selected starting precursors, and thus no solution was found. Another possibility is that the 6 precursors considered are not the closest *E. coli* metabolite, and so more *E. coli* reactions were needed extending the length of the pathway. From looking at the minimum path lengths between products and precursors, it appears that pyruvate, oxaloacetate and α-ketoglutarate are the closest precursors to 69 out of the 81 non-native products considered. Strains capable of producing these precursors at high rates and yields (e.g. pyruvate overproducing strains described in Chapter 3) would be good background strains to produce these 69 products.

The results presented here focused on analyzing connections between a select number of precursors to a wide variety of non-native products. Future work should also evaluate which native metabolites (not just the 6 precursors considered) are closest to non-native products, and should also consider putative reactions. The current results provide guidance as to what

precursors are most important to make and what non-native products should be made using these strains.

Chapter 5 Conclusions and future directions

In this work, we combined genetic tools, adaptive evolution techniques and *in silico* simulations to explore different metabolic engineering strategies for strain development and product optimization. First, we investigated the phenotypes of un-evolved and evolved co-cultures consisting of two *E. coli* auxotrophs cross-feeding leucine (or its precursors) and lysine, found the improved community phenotype was attributed to a better exchange of the two amino acids (or precursors) and also found fitness tradeoffs occurred in individual strains. Though the isolated strains did not show an overproduction of leucine or lysine when grown in isolation, the co-culture studies will contribute more knowledge to utilizing microbial consortia in metabolic engineering. Second, aided by the genome-scale metabolic model of *E. coli*, we constructed gene deletion strains capable of producing pyruvate at high yields. Pyruvate is a starting material for synthesizing many important chemicals. We successfully converted the pyruvate strains to produce ethanol and anticipated that the developed strains would serve as a platform to produce other valuable chemicals. Last, to explore all possible chemicals that could be made from pyruvate, we did a comprehensive study of the production capability of *E .coli* for known biochemical and identified numerous valuable non-native products able to be made in *E. coli* and their synthesis pathways from different precursors, including pyruvate. The data generated from this work will contribute to expedite the productions of non-native products in microbes and the pathway engineering.

5.1 Future directions

5.1.1 Produce other native products in designed pyruvate producing strains

Pyruvate is the precursor for native metabolites such as branched-chain amino acids, L-alanine and ethanol. We have used pyruvate strains to produce ethanol and reached 58 to 85% of theoretical yield. One ethanol molecule formation costs one pyruvate and two NADH molecules. In a similar fashion, L-alanine synthesis by *Bacillus sphaericus* alanine dehydrogenase requires one molecule pyruvate and one molecule NADH. This *Bacillus* gene *alaD* encoding the alanine dehydrogenase has been successfully expressed in *E. coli* and produced L-alanine [134]. We will try to express *alaD* in the pyruvate producing strains and evaluate the L-alanine production, which we expect to be similar to the yield of ethanol (up to ~85% of the theoretical maximum yield) produced in the modified pyruvate producing strains.

The three branched-chain amino acid synthesis pathways are interconnected. Pyruvate is the sole precursor for L-valine synthesis, and L-leucine is made from intermediates of L-valine. Thus we should first develop a L-valine producing strain by overexpressing genes (*ilvBNCDE*) involving the synthesis in the pyruvate overproducing strains and then convert it into a leucine producing strain by overexpressing more genes (*leuABCD*). The synthesis of the third BAA, L-isoleucine, requires L-threonine and pyruvate as precursors, and it could be studied last, since additional pathways leading to threonine production would need to be manipulated.

5.1.2 Produce non-native products based on pyruvate producing strains

Pyruvate can be used to produce 64 non-native products within 5 reactions, and the synthesis pathways were also identified in this study (Chapter 4). Eighteen of non-native products only require one heterologous reaction and would thus be good targets for producing in the *E. coli* pyruvate producing strains. The priority of the target products can be evaluated by

multiple criteria such as: the availability and number of genes required for the heterologous reaction, the catalytic activity of the heterologous enzymes (which could be searched in BRENDA database), toxicity of intermediates and end products, the expression and maturation of heterologous enzymes in *E. coli* and the cost of the target product. The downstream quantification method for the target product should also be available. After the target is decided, the heterologous gene can be cloned or synthesized by company, and then introduced into pyruvate producing strains.

5.2 Concluding remarks

Bacteria, the smallest living organisms, provide many benefits to us. After working on metabolic engineering for the past six years, I am amazed by the metabolic and phenotypic plasticity of bacteria and have seen they can produce useful chemicals that may impact the lives of everyone. I am glad that I learned optimization skills to analyze genomic-scale models to better understand bacterial metabolism and to apply the models for strain development. From my experimental and computational experience, I believe that experiments are the foundation to provide information, to validate hypotheses, and to control the bacteria in the way we want. I also believe that computational methods can dramatically expedite the discovery of complex mechanism, the design of strains at a system level, and the optimization of product yields and the development process. The combination of both will create new products, pathways and functions. I expect more exciting discoveries in this field and more efficient microbial factories developed to manufacture the products used in our life.

Appendix 1: Currency metabolites removed from the carbon transfer map

KEGG ID	Common name	Formula	KEGG ID	Common name	Formula
C00237	CO	CO	C01367	3'-AMP	C10H14N5O7P
C00011	CO2	CO2	C00286	dGTP	C10H16N5O13P3
C01353	Carbonic acid	H2CO3	C00044	GTP	C10H16N5O14P3
C00288	HCO3-	HCO3	C00361	dGDP	C10H15N5O10P2
C00365	dUMP	C9H13N2O8P	C00035	GDP	C10H15N5O11P2
C00105	UMP	C9H13N2O9P	C00362	dGMP	C10H14N5O7P
C01368	3'-UMP	C9H13N2O9P	C00144	GMP	C10H14N5O8P
C01346	dUDP	C9H14N2O11P2	C00459	dTTP	C10H17N2O14P3
C00015	UDP	C9H14N2O12P2	C00363	dTDP	C10H16N2O11P2
C00460	dUTP	C9H15N2O14P3	C00364	dTMP	C10H15N2O8P
C00075	UTP	C9H15N2O15P3	C01345	dITP	C10H15N4O13P3
C00239	dCMP	C9H14N3O7P	C00081	ITP	C10H15N4O14P3
C00055	CMP	C9H14N3O8P	C01344	dIDP	C10H14N4O10P2
C05822	3'-CMP	C9H14N3O8P	C00104	IDP	C10H14N4O11P2
C00458	dCTP	C9H16N3O13P3	C00130	IMP	C10H13N4O8P
C00063	CTP	C9H16N3O14P3	C00229	acyl carrier protein	C11H22N2O7PRS
C00705	dCDP	C9H15N3O10P2	C00390	Ubiquinol	C14H20O4
C00112	CDP	C9H15N3O11P2	C00003	NAD+	C21H28N7O14P2
C00131	dATP	C10H16N5O12P3	C00004	NADH	C21H29N7O14P2
C00002	ATP	C10H16N5O13P3	C00006	NADP+	C21H29N7O17P3
C00008	ADP	C10H15N5O10P2	C00005	NADPH	C21H30N7O17P3
C00206	dADP	C10H15N5O9P2	C00010	CoA	C21H36N7O16P3S
C00360	dAMP	C10H14N5O6P	C00016	FAD	C27H33N9O15P2
C00020	AMP	C10H14N5O7P	C01352	FADH2	C27H35N9O15P2

Appendix 2: The 284 non-native products in *E. coli* that have commerical value

	KEGG ID	Formula	Common name	Maximal yield*	Minimal number of heterologous reaction required	Application‡
1	C00067	CH2O	Formaldehyde	6	1	Pha
2	C00114	C5H14NO	Choline	1.2	6	Pha
3	C00146	C6H6O	Phenol	1	1	Pha
4	C00230	C7H6O4	Protocatechuate	0.85714	1	Pha
5	C00243	C12H22O11	Lactose	0.5	1	Pha
6	C00245	C2H7NO3S	Taurine	3	2	Pha
7	C00272	C9H15N5O3	Tetrahydrobiopterin	0.66667	2	Pha
8	C00300	C4H9N3O2	Creatine	1.5	3	Pha
9	C00315	C7H19N3	Spermidine	0.85714	1	Pha
10	C00329	C6H13NO5	D-Glucosamine	1	1	Pha
11	C00355	C9H11NO4	Levodopa	0.66667	1	Pha
12	C00378	C12H17N4OS	Thiamine	0.5	2	Pha
13	C00386	C9H14N4O3	Carnosine	0.66667	1	Pha
14	C00389	C15H10O7	Quercetin	0.4	7	Pha
15	C00392	C6H14O6	Mannitol	1	1	Pha
16	C00398	C10H12N2	Tryptamine	0.6	1	Pha
17	C00402	C4H7NO4	D-Aspartate	1.5	1	Pha
18	C00483	C8H11NO	Tyramine	0.75	1	Pha
19	C00504	C19H19N7O6	Folate	0.31579	1	Pha
20	C00509	C15H12O5	Naringenin	0.4	4	Pha
21	C00519	C2H7NO2S	Hypotaurine	3	3	Pha
22	C00547	C8H11NO3	L-Noradrenaline	0.75	4	Pha
23	C00552	C4H6O6	meso-Tartaric acid	1.5	2	Pha
24	C00556	C7H8O	Benzyl alcohol	0.85714	6	Pha
25	C00628	C7H6O4	2,5-Dihydroxybenzoate	0.85714	2	Pha
26	C00643	C11H12N2O3	5-Hydroxy-L-tryptophan	0.54545	2	Pha
27	C00745	C10H14N2	Nicotine	0.6	4	Pha
28	C00757	C20H18NO4	Berberine	0.3	13	Pha
29	C00780	C10H12N2O	Serotonin	0.6	3	Pha
30	C00788	C9H13NO3	L-Adrenaline	0.66667	5	Pha
31	C00794	C6H14O6	D-Sorbitol	1	1	Pha
32	C00795	C6H12O6	D-Tagatose	1	2	Pha
33	C00805	C7H6O3	Salicylate	0.85714	1	Pha
34	C00808	C10H16O	(+)-Camphor	0.6	3	Pha
35	C00811	C9H8O3	4-Coumarate	0.66667	2	Pha
36	C00814	C16H12O5	Biochanin A	0.375	8	Pha

Appendix 2, continued

	KEGG ID	Formula	Common name	Maximal yield*	Minimal number of heterologous reaction required	Application‡
37	C00836	C18H39NO2	Sphinganine	0.33333	2	Pha
38	C00880	C6H12O7	D-Galactonate	1	1	Pha
39	C00884	C10H16N4O3	Homocarnosine	0.6	1	Pha
40	C00916	C16H21N3O8S	Cephalosporin C	0.375	7	Pha
41	C00954	C10H9NO2	Indole-3-acetate	0.6	3	Pha
42	C00955	C10H11NO	Indole-3-ethanol	0.6	3	Pha
43	C01026	C4H9NO2	N,N-Dimethylglycine	1.5	3	Pha
44	C01157	C5H9NO3	trans-4-Hydroxy-L-proline	1.2	1	Pha
45	C01197	C9H8O4	Caffeate	0.66667	6	Pha
46	C01262	C10H16N4O3	Anserine	0.6	2	Pha
47	C01424	C7H6O5	Gallate	0.85714	5	Pha
48	C01441	C12H26N4O6	Neamine	0.5	9	Pha
49	C01467	C7H8O	3-Cresol	0.85714	2	Pha
50	C01477	C15H10O5	Apigenin	0.4	6	Pha
51	C01494	C10H10O4	Ferulate	0.6	9	Pha
52	C01514	C15H10O6	Luteolin	0.4	6	Pha
53	C01537	C3H7NO2	Urethane	2	2	Pha
54	C01598	C13H16N2O2	Melatonin	0.46154	5	Pha
55	C01617	C15H12O7	Taxifolin	0.4	7	Pha
56	C01678	C2H7NS	Cysteamine	3	4	Pha
57	C01701	C20H18O5	(-)-Glyceollin I	0.3	12	Pha
58	C01717	C10H7NO3	Kynurenate	0.6	4	Pha
59	C01737	C23H46N6O13	Neomycin B	0.26087	16	Pha
60	C01759	C17H34N4O10	Ribostamycin	0.35294	11	Pha
61	C01765	C10H18O	(+)-Borneol	0.6	2	Pha
62	C01850	C18H16O8	Rosmarinate	0.33333	10	Pha
63	C01983	C8H8O3	(R)-Mandelate	0.75	4	Pha
64	C01984	C8H8O3	(S)-Mandelate	0.75	3	Pha
65	C02105	C19H23NO4	(S)-Reticuline	0.31579	8	Pha
66	C02107	C4H6O6	(S,S)-Tartaric acid	1.5	1	Pha
67	C02378	C6H13NO2	6-Aminohexanoate	1	5	Pha
68	C02442	C9H13NO	N-Methyltyramine	0.66667	2	Pha
69	C02670	C6H8O6	D-Glucuronolactone	1	4	Pha
70	C02890	C21H25NO4	Tetrahydropalmatine	0.28571	12	Pha
71	C02906	C15H12O8	Dihydromyricetin	0.4	6	Pha
72	C02993	C15H23N5O4	Kyotorphin	0.4	1	Pha

Appendix 2, continued

	KEGG ID	Formula	Common name	Maximal yield*	Minimal number of heterologous reaction required	Application‡
73	C03329	C20H21NO4	(S)-Canadine	0.3	12	Pha
74	C03375	C6H17N3	Norspermidine	1	3	Pha
75	C03582	C14H12O3	Resveratrol	0.42857	3	Pha
76	C03758	C8H11NO2	Dopamine	0.75	2	Pha
77	C03761	C6H10O5	Meglutol	1	2	Pha
78	C04444	C18H16O7	3,7,4'-Tri-O-methylquercetin	0.33333	12	Pha
79	C04548	C9H13NO2	Synephrine	0.66667	4	Pha
80	C04858	C26H28O14	Apiin	0.23077	8	Pha
81	C05178	C19H23NO4	(R)-Reticuline	0.31579	11	Pha
82	C05332	C8H11N	Phenethylamine	0.75	1	Pha
83	C05335	C5H11NO2Se	L-Selenomethionine	1.2	4	Pha
84	C05422	C6H6O6	Dehydroascorbate	1	6	Pha
85	C05551	C16H18N2O4S	Penicillin G	0.375	4	Pha
86	C05587	C9H13NO2	3-Methoxytyramine	0.66667	3	Pha
87	C05623	C21H20O12	Quercetin 3-O-glucoside	0.28571	8	Pha
88	C05625	C27H30O16	Rutin	0.22222	11	Pha
89	C05627	C8H8O	Phenylacetaldehyde	0.75	3	Pha
90	C05670	C3H6N2	3-Aminopropiononitrile	2	2	Pha
91	C05829	C6H10N2O5	N-Carbamyl-L-glutamate	1	5	Pha
92	C05853	C8H10O	Phenylethyl alcohol	0.75	2	Pha
93	C05903	C15H10O6	Kaempferol	0.4	6	Pha
94	C05905	C15H11O6	Cyanidin	0.4	9	Pha
95	C05908	C15H11O7	Delphinidin	0.4	8	Pha
96	C05984	C4H8O3	4-Hydroxybutanoic acid	1.5	1	Pha
97	C06044	C8H10O2	4-Hydroxyphenylethanol	0.75	2	Pha
98	C06046	C14H20O7	Salidroside	0.42857	3	Pha
99	C06051	C8H9NO3	Isopyridoxal	0.75	1	Pha
100	C06124	C18H38NO5P	Sphingosine 1-phosphate	0.33333	4	Pha
101	C06142	C4H10O	1-Butanol	1.5	2	Pha
102	C06173	C19H21NO3	Thebaine	0.31579	15	Pha
103	C06186	C12H16O7	Arbutin	0.5	2	Pha
104	C06199	C10H15NO	Hordenine	0.6	4	Pha
105	C06213	C11H14N2	N-Methyltryptamine	0.54545	2	Pha
106	C06308	C10H16	(-)-alpha-Pinene	0.6	1	Pha
107	C06562	C15H14O6	(+)-Catechin	0.4	8	Pha

Appendix 2, continued

	KEGG ID	Formula	Common name	Maximal yield*	Minimal number of heterologous reaction required	Application‡
108	C06563	C15H10O5	Genistein	0.4	6	Pha
109	C07130	C7H8N4O2	Theophylline	0.85714	2	Pha
110	C08299	C12H16N2O	Bufotenine	0.5	6	Pha
111	C08615	C30H50O	alpha-Amyrin	0.2	3	Pha
112	C08616	C30H50O	beta-Amyrin	0.2	3	Pha
113	C08628	C30H50O	Lupeol	0.2	3	Pha
114	C08650	C15H12O4	Isoliquiritigenin	0.4	3	Pha
115	C09094	C20H34O	Geranylgeraniol	0.3	2	Pha
116	C09126	C21H20O10	Genistein	0.28571	7	Pha
117	C09629	C15H24	beta-Caryophyllene	0.4	1	Pha
118	C09684	C15H24	Humulene	0.4	1	Pha
119	C09762	C15H12O4	Liquiritigenin	0.4	4	Pha
120	C09789	C27H32O14	Naringin	0.22222	8	Pha
121	C09826	C15H12O5	Pinobanksin	0.4	6	Pha
122	C09827	C15H12O4	Pinocembrin	0.4	5	Pha
123	C10028	C15H10O4	Chrysin	0.4	6	Pha
124	C10044	C15H10O5	Galangin	0.4	7	Pha
125	C10107	C15H10O8	Myricetin	0.4	7	Pha
126	C10192	C15H10O7	Tricetin	0.4	6	Pha
127	C10208	C15H10O4	Daidzein	0.4	6	Pha
128	C10216	C21H20O9	Daidzin	0.28571	7	Pha
129	C10443	C21H20O6	Curcumin	0.28571	9	Pha
130	C10509	C22H22O9	Ononin	0.27273	8	Pha
131	C10520	C16H12O6	Pratensein	0.375	9	Pha
132	C10521	C16H12O5	Prunetin	0.375	9	Pha
133	C12127	C15H14O7	(+)-Gallocatechin	0.4	8	Pha
134	C12138	C21H21O12	Delphinidin 3-O-glucoside	0.28571	9	Pha
135	C15511	C20H18O5	Glyceollin III	0.3	12	Pha
136	C15652	C23H46N6O13	Neomycin C	0.26087	15	Pha
137	C16829	C15H24	gamma-Humulene	0.4	1	Pha
138	C17742	C20H18O5	Demethoxycurcumin	0.3	9	Pha
139	C17743	C19H16O4	Bisdemethoxycurcumin	0.31579	4	Pha
140	C18023	C20H20O5	Sophoraflavanone B	0.3	5	Pha
141	C00423	C9H8O2	trans-Cinnamate	0.66667	2	Per and Cos
142	C00466	C4H8O2	Acetoin	1.5	3	Per and Cos
143	C00521	C10H16	(-)-Limonene	0.6	1	Per and Cos
144	C00823	C16H34O	1-Hexadecanol	0.375	2	Per and Cos

Appendix 2, continued

	KEGG ID	Formula	Common name	Maximal yield*	Minimal number of heterologous reaction required	Application‡
145	C00964	C10H16O	(-)-trans-Carveol	0.6	2	Per and Cos
146	C01126	C15H26O	trans-Farnesol	0.4	1	Per and Cos
147	C01389	C20H40O	Phytol	0.3	4	Per and Cos
148	C01499	C10H16O	Geranial	0.6	2	Per and Cos
149	C01724	C30H50O	Lanosterol	0.2	3	Per and Cos
150	C01769	C4H8O2	(S)-Acetoin	1.5	2	Per and Cos
151	C01879	C5H7NO3	5-Oxoproline	1.2	1	Per and Cos
152	C02237	C5H7NO3	5-Oxo-D-proline	1.2	1	Per and Cos
153	C02344	C10H18O	(-)-endo-Fenchol	0.6	1	Per and Cos
154	C02348	C4H6N4O3	(R)(-)-Allantoin	1.5	4	Per and Cos
155	C02350	C4H6N4O3	(S)(+)-Allantoin	1.5	3	Per and Cos
156	C02394	C9H10O	Cinnamyl alcohol	0.66667	5	Per and Cos
157	C02576	C10H14O	Perillyl aldehyde	0.6	3	Per and Cos
158	C03220	C15H26O	2-cis,6-trans-Farnesol	0.4	2	Per and Cos
159	C05413	C40H64	Phytoene	0.15	2	Per and Cos
160	C05421	C40H64	15-cis-Phytoene	0.15	3	Per and Cos
161	C06074	C10H16	Myrcene	0.6	1	Per and Cos
162	C06099	C10H16	(+)-Limonene	0.6	1	Per and Cos
163	C06231	C6H10N2O2	Ectoine	1	3	Per and Cos
164	C06358	C10H10O2	Methyl cinnamate	0.6	6	Per and Cos
165	C06359	C11H12O2	Ethyl cinnamate	0.54545	4	Per and Cos
166	C06394	C15H24	(+)-delta-Cadinene	0.4	1	Per and Cos
167	C07086	C8H8O2	Phenylacetic acid	0.75	1	Per and Cos
168	C09183	C20H36O2	Sclareol	0.3	2	Per and Cos
169	C09621	C15H26O	(-)-alpha-Bisabolol	0.4	1	Per and Cos
170	C09704	C15H26O	Nerolidol	0.4	1	Per and Cos
171	C09847	C10H16O	cis-Citral	0.6	3	Per and Cos
172	C09871	C10H18O	Nerol	0.6	4	Per and Cos
173	C09893	C10H16O	Pulegone	0.6	5	Per and Cos
174	C11388	C10H18O	(-)-Linalool	0.6	1	Per and Cos
175	C11389	C10H18O	(+)-Linalool	0.6	1	Per and Cos
176	C11409	C10H16O	(+)-trans-Carveol	0.6	2	Per and Cos
177	C00090	C6H6O2	Catechol	1	1	Bui
178	C00132	CH4O	Methanol	6	2	Bui
179	C00180	C7H6O2	Benzoate	0.85714	6	Bui
180	C00207	C3H6O	Acetone	2	2	Bui
181	C00218	CH5N	Methylamine	6	3	Bui

Appendix 2, continued

	KEGG ID	Formula	Common name	Maximal yield*	Minimal number of heterologous reaction required	Application‡
182	C00246	C4H8O2	Butanoic acid	1.5	1	Bui
183	C00261	C7H6O	Benzaldehyde	0.85714	4	Bui
184	C00292	C6H7N	Aniline	1	1	Bui
185	C00409	CH4S	Methanethiol	6	2	Bui
186	C00472	C6H4O2	p-Benzoquinone	1	2	Bui
187	C00479	C3H6O	Propanal	2	1	Bui
188	C00490	C5H6O4	Itaconate	1.2	1	Bui
189	C00511	C3H4O2	Propenoate	2	3	Bui
190	C00804	C3H2O2	Propynoate	2	1	Bui
191	C00818	C6H10O8	Glucarate	1	5	Bui
192	C00870	C6H5NO3	4-Nitrophenol	1	4	Bui
193	C00898	C4H6O6	(R,R)-Tartaric acid	1.5	1	Bui
194	C00903	C9H8O	Cinnamaldehyde	0.66667	4	Bui
195	C00986	C3H10N2	1,3-Diaminopropane	2	2	Bui
196	C01089	C4H8O3	(R)-3-Hydroxybutanoate	1.5	2	Bui
197	C01147	C6H10O2	2-Hydroxycyclohexan-1-one	1	6	Bui
198	C01263	C15H12O5	(-)-Glycinol	0.4	11	Bui
199	C01326	CHN	Hydrogen cyanide	6	2	Bui
200	C01380	C2H6O2	Ethylene glycol	3	2	Bui
201	C01384	C4H4O4	Maleic acid	1.5	1	Bui
202	C01403	C7H8O	Anisole	0.85714	2	Bui
203	C01408	C14H12O2	Benzoin	0.42857	5	Bui
204	C01502	C7H8O2	o-Methoxyphenol	0.85714	2	Bui
205	C01837	C2H5NO2	Nitroethane	3	3	Bui
206	C01845	C3H8O	Propan-2-ol	2	3	Bui
207	C01902	C30H50O	Cycloartenol	0.2	3	Bui
208	C01987	C6H7NO	2-Aminophenol	1	3	Bui
209	C01998	C3H3N	Acrylonitrile	2	5	Bui
210	C02372	C6H7NO	4-Hydroxyaniline	1	1	Bui
211	C02457	C3H8O2	Propane-1,3-diol	2	2	Bui
212	C02505	C8H9NO	2-Phenylacetamide	0.75	1	Bui
213	C02752	C6H6O3	Triacetate lactone	1	1	Bui
214	C02912	C3H8O2	(R)-Propane-1,2-diol	2	1	Bui
215	C02917	C3H8O2	(S)-Propane-1,2-diol	2	1	Bui
216	C02954	C8H12N2O3S	6-Aminopenicillanate	0.75	4	Bui
217	C03194	C3H9NO	(R)-1-Aminopropan-2-ol	2	1	Bui

Appendix 2, continued

	KEGG ID	Formula	Common name	Maximal yield*	Minimal number of heterologous reaction required	Application‡
218	C03248	C4H2O4	Acetylenedicarboxylate	1.5	1	Bui
219	C05380	C8H8N2O3	Nicotinurate	0.75	1	Bui
220	C05986	C3H4O	2-Propyn-1-ol	2	4	Bui
221	C06104	C6H10O4	Adipate	1	3	Bui
222	C06202	C7H6O2	Salicylaldehyde	0.85714	2	Bui
223	C06425	C20H40O2	Icosanoic acid	0.3	2	Bui
224	C06547	C2H4	Ethylene	3	3	Bui
225	C06548	C2H4O	Ethylene oxide	3	1	Bui
226	C06593	C6H11NO	epsilon-Caprolactam	1	6	Bui
227	C06813	C6H5NO2	Nitrobenzene	1	6	Bui
228	C08063	C6H8O2	1,4-Cyclohexanedione	1	4	Bui
229	C10833	C9H10O5	Syringic acid	0.66667	7	Bui
230	C11506	C3H6O	(R)-1,2-Epoxypropane	2	4	Bui
231	C11507	C3H6O	(S)-1,2-Epoxypropane	2	4	Bui
232	C16028	C15H24	Amorpha-4,11-diene	0.4	1	Bui
233	C16074	C8H7N	Phenylacetonitrile	0.75	2	Bui
234	C16521	C5H8	Isoprene	1.2	1	Bui
235	C17277	C15H24	(+)-Valencene	0.4	1	Bui
236	C00072	C6H8O6	Ascorbate	1	5	Fod
237	C00741	C4H6O2	Diacetyl	1.5	2	Fod
238	C00755	C8H8O3	Vanillin	0.75	9	Fod
239	C00810	C4H8O2	(R)-Acetoin	1.5	1	Fod
240	C00852	C16H18O9	Chlorogenate	0.375	4	Fod
241	C01613	C24H42O21	Stachyose	0.25	3	Fod
242	C01767	C10H14O	(-)-Carvone	0.6	3	Fod
243	C02477	C29H50O2	alpha-Tocopherol	0.2069	7	Fod
244	C02483	C28H48O2	gamma-Tocopherol	0.21429	6	Fod
245	C05629	C9H10O2	Phenylpropanoate	0.66667	3	Fod
246	C06672	C8H8O4	Vanillate	0.75	12	Fod
247	C08604	C21H21O11	Chrysanthemin	0.28571	9	Fod
248	C14151	C27H46O2	delta-Tocopherol	0.22222	5	Fod
249	C01500	C10H18O	Geraniol	0.6	1	Agr
250	C01566	CH2N2	Cyanamide	3	1	Agr
251	C04720	C9H9NO5	DIMBOA	0.66667	8	Agr
252	C05851	C9H6O2	Coumarin	0.66667	7	Agr
253	C11383	C10H14O	(+)-(S)-Carvone	0.6	3	Agr
254	C16141	C15H24	(+)-Germacrene A	0.4	1	Agr

Appendix 2, continued

	KEGG ID	Formula	Common name	Maximal yield*	Minimal number of heterologous reaction required	Application‡
255	C16142	C15H24	(-)-Germacrene D	0.4	1	Agr
256	C00209	C2H2O4	Oxalate	3	1	Others
257	C00488	CH3NO	Formamide	6	2	Others
558	C00492	C18H32O16	Raffinose	0.33333	2	Others
259	C00530	C6H6O2	p-Benzenediol	1	1	Others
260	C00751	C30H50	Squalene	0.2	1	Others
261	C01108	C6H6O3	1,2,3-Trihydroxybenzene	1	6	Others
262	C01412	C4H8O	Butanal	1.5	1	Others
263	C01438	CH4	Methane	6	6	Others
264	C01548	C2H2	Acetylene	3	1	Others
265	C01659	C3H5NO	Acrylamide	2	5	Others
266	C01732	C5H6O4	Mesaconate	1.2	2	Others
267	C01750	C21H20O11	Quercitrin	0.28571	10	Others
268	C02048	C12H22O11	Laminaribiose	0.5	1	Others
269	C03044	C4H10O2	(R,R)-Butane-2,3-diol	1.5	2	Others
270	C03046	C4H10O2	(S,S)-Butane-2,3-diol	1.5	3	Others
271	C05552	C16H28N4O4S	Biocytin	0.19355	1	Others
272	C05688	C3H7NO2Se	L-Selenocysteine	2	1	Others
273	C05979	C3H8O	Propane-1-ol	2	2	Others
274	C06771	C6H15NO3	Triethanolamine	1	2	Others
275	C06772	C4H11NO2	Diethanolamine	1.5	1	Others
276	C07565	C8H9NO	Acetanilide	0.75	2	Others
277	C08620	C27H31O15	Cyanidin 3-O-rutinoside	0.22222	14	Others
278	C08639	C27H31O16	Cyanin	0.22222	10	Others
279	C09665	C15H24	alpha-Farnesene	0.4	1	Others
280	C09666	C15H24	beta-Farnesene	0.4	1	Others
281	C09699	C15H24	Longifolene	0.4	1	Others
282	C09814	C7H5N	Benzonitrile	0.85714	3	Others
283	C12627	C27H30O14	Rhoifolin	0.22222	10	Others
284	C16315	C27H31O16	Tulipanin	0.22222	12	Others

* The maximal yield is expressed as mmol of a target metabolite produced per mmol glucose.

‡ Application abreviations are Pharmeceuticals: Pha; Perfume and Cosmetics: Per and Cos; Building blocks: Bui; Food industry: Fod; Agriculture: Agr.

Bibliography

1. Gregory N. Stephanopoulos AAAaJN (1998) Metabolic Engineering: Principles and Methodologies. San Diego, CA: Academic Press.
2. Abatemarco J, Hill A, Alper HS (2013) Expanding the metabolic engineering toolbox with directed evolution. Biotechnology Journal 8: 1397-1410.
3. Park JH, Lee KH, Kim TY, Lee SY (2007) Metabolic engineering of Escherichia coli for the production of L-valine based on transcriptome analysis and in silico gene knockout simulation. Proceedings of the National Academy of Sciences of the United States of America 104: 7797-7802.
4. Jewett MC, Oliveira AP, Patil KR, Nielsen J (2005) The role of high-throughput transcriptome analysis in metabolic engineering. Biotechnology and Bioprocess Engineering 10: 385-399.
5. Tyo KE, Alper HS, Stephanopoulos GN (2007) Expanding the metabolic engineering toolbox: more options to engineer cells. Trends in Biotechnology 25: 132-137.
6. Alper H, Fischer C, Nevoigt E, Stephanopoulos G (2005) Tuning genetic control through promoter engineering. Proceedings of the National Academy of Sciences of the United States of America 102: 12678-12683.
7. Blazeck J, Alper H (2010) Systems metabolic engineering: Genome-scale models and beyond. Biotechnology Journal 5: 647-659.
8. Zomorrodi AR, Suthers PF, Ranganathan S, Maranas CD (2012) Mathematical optimization applications in metabolic networks. Metabolic Engineering 14: 672-686.
9. Sahm H, Eggeling L, Eikmanns B, Kramer R (1995) Metabolic Design in Amino-Acid Producing Bacterium Corynebacterium-Glutamicum. Fems Microbiology Reviews 16: 243-252.
10. Atsumi S, Hanai T, Liao JC (2008) Non-fermentative pathways for synthesis of branched-chain higher alcohols as biofuels. Nature 451: 86-U13.
11. Holatko J, Elisakova V, Prouza M, Sobotka M, Nesvera J, et al. (2009) Metabolic engineering of the L-valine biosynthesis pathway in Corynebacterium glutamicum using promoter activity modulation. Journal of Biotechnology 139: 203-210.
12. Tsuchida T, Momose H (1986) Improvement of an l-Leucine-Producing Mutant of Brevibacterium lactofermentum 2256 by Genetically Desensitizing It to alpha-Acetohydroxy Acid Synthetase. Appl Environ Microbiol 51: 1024-1027.
13. Shi F, Huan XJ, Wang XY, Ning JF (2012) Overexpression of NAD kinases improves the L-isoleucine biosynthesis in Corynebacterium glutamicum ssp lactofermentum. Enzyme and Microbial Technology 51: 73-80.
14. Causey TB, Shanmugam KT, Yomano LP, Ingram LO (2004) Engineering Escherichia coli for efficient conversion of glucose to pyruvate. Proceedings of the National Academy of Sciences of the United States of America 101: 2235-2240.
15. Sachs JL, Skophammer RG, Regus JU (2011) Evolutionary transitions in bacterial symbiosis. Proceedings of the National Academy of Sciences of the United States of America 108: 10800-10807.
16. Sachs JL, Mueller UG, Wilcox TP, Bull JJ (2004) The evolution of cooperation. Quarterly Review of Biology 79: 135-160.

17. Morris JJ, Lenski RE, Zinser ER (2012) The black queen hypothesis: evolution of dependencies through adaptive gene loss. Mbio 3(2):e00036-12.
18. Axelrod R, Hamilton WD (1981) The evolution of cooperation. Science 211: 1390-1396.
19. Nowak MA (2006) Five rules for the evolution of cooperation. Science 314: 1560-1563.
20. Kaeberlein T, Lewis K, Epstein SS (2002) Isolating "uncultivable" microorganisms in pure culture in a simulated natural environment. Science 296: 1127-1129.
21. Stoodley P, Cargo R, Rupp CJ, Wilson S, Klapper I (2002) Biofilm material properties as related to shear-induced deformation and detachment phenomena. Journal of Industrial Microbiology & Biotechnology 29: 361-367.
22. Shin CS, Kim HJ, Kim MJ, Ju JY (1998) Morphological change and enhanced pigment production of monascus when cocultured with *Saccharomyces cerevisiae* or *Aspergillus oryzae*. Biotechnology and Bioengineering 59: 576-581.
23. Kerner A, Park J, Williams A, Lin XXNN (2012) A programmable *Escherichia coli* consortium via tunable symbiosis. PLoS One 7: e34032.
24. Wintermute EH, Silver PA (2010) Emergent cooperation in microbial metabolism. Molecular Systems Biology 6: 407.
25. Hosoda K, Suzuki S, Yamauchi Y, Shiroguchi Y, Kashiwagi A, et al. (2011) Cooperative adaptation to establishment of a synthetic bacterial mutualism. PLoS One 6: e17105.
26. Pande S, Merker H, Bohl K, Reichelt M, Schuster Sea (2013) Fitness and stability of obligate cross-feeding interactions that emerge upon gene loss in bacteria. ISME J doi:10.1038/ismej.2013.211.
27. Harcombe W (2010) Novel cooperation experimentally evolved between species. Evolution 64: 2166-2172.
28. Hillesland KL, Stahl DA (2010) Rapid evolution of stability and productivity at the origin of a microbial mutualism. Proceedings of the National Academy of Sciences of the United States of America 107: 2124-2129.
29. Shou WY, Ram S, Vilar JMG (2007) Synthetic cooperation in engineered yeast populations. Proceedings of the National Academy of Sciences of the United States of America 104: 1877-1882.
30. Bull JJ, Harcombe WR (2009) Population dynamics constrain the cooperative evolution of cross-feeding. PLoS One 4:e4115.
31. Stolyar S, Van Dien S, Hillesland KL, Pinel N, Lie TJ, et al. (2007) Metabolic modeling of a mutualistic microbial community. Molecular Systems Biology 3:92.
32. Nagarajan H, Embree M, Rotaru AE, Shrestha PM, Feist AM, et al. (2013) Characterization and modelling of interspecies electron transfer mechanisms and microbial community dynamics of a syntrophic association. Nature Communications 4:2809.
33. Zomorrodi AR, Maranas CD (2012) OptCom: A Multi-Level Optimization Framework for the Metabolic Modeling and Analysis of Microbial Communities. Plos Computational Biology 8.
34. Brenner K, You LC, Arnold FH (2008) Engineering microbial consortia: a new frontier in synthetic biology. Trends in Biotechnology 26: 483-489.
35. Purnick PEM, Weiss R (2009) The second wave of synthetic biology: from modules to systems. Nature Reviews Molecular Cell Biology 10: 410-422.

36. Rawlings DE, Johnson DB (2007) The microbiology of biomining: development and optimization of mineral-oxidizing microbial consortia. Microbiology-Sgm 153: 315-324.
37. Waks Z, Silver PA (2009) Engineering a synthetic dual-organism system for hydrogen production. Applied and Environmental Microbiology 75: 1867-1875.
38. Charusanti P, Fong NL, Nagarajan H, Pereira AR, Li HJ, et al. (2012) Exploiting adaptive laboratory evolution of streptomyces clavuligerus for antibiotic discovery and overproduction. PLoS One 7: e33727.
39. Minty JJ, Singer ME, Scholz SA, Bae CH, Ahn JH, et al. (2013) Design and characterization of synthetic fungal-bacterial consortia for direct production of isobutanol from cellulosic biomass. Proceedings of the National Academy of Sciences of the United States of America 110: 14592-14597.
40. Wintermute EH, Silver PA (2010) Dynamics in the mixed microbial concourse. Genes & Development 24: 2603-2614.
41. Datsenko KA, Wanner BL (2000) One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. Proceedings of the National Academy of Sciences of the United States of America 97: 6640-6645.
42. Sternberg N, Hoess R (1983) The molecular-genetics of bacteriophage-P1. Annual Review of Genetics 17: 123-154.
43. Christensen JE, Steele JL (2003) Impaired growth rates in milk of *Lactobacillus helveticus* peptidase mutants can be overcome by use of amino acid supplements. Journal of Bacteriology 185: 3297-3306.
44. Baumler DJ, Peplinski RG, Reed JL, Glasner JD, Perna NT (2011) The evolution of metabolic networks of *E. coli*. Bmc Systems Biology 5:182.
45. Harris DC (2003) Quantitative chemical analysis 6th edition. New York: W.H. Freeman and Company. Chapter 5. 80-98 p.
46. Feist AM, Henry CS, Reed JL, Krummenacker M, Joyce AR, et al. (2007) A genome-scale metabolic reconstruction for *Escherichia coli* K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. Molecular Systems Biology 3: 121.
47. Varma A, Palsson BO (1994) Stoichiometric flux balance models quantitatively predict growth and metabolic by-product secretion in wild-type *Escherichia coli* W3110. Applied and Environmental Microbiology 60: 3724-3731.
48. Neidhardt FC, Umbarger EH (1996) Chemical Composition of *Escherichia coli*. In: Neidhardt FC, editor. *Escherichia coli* and *Salmonella*: Cellular and Molecular Biology. 2nd ed. Washington DC: ASM Press. pp. 13-16.
49. Fong SS, Joyce AR, Palsson BO (2005) Parallel adaptive evolution cultures of *Escherichia coli* lead to convergent growth phenotypes with different gene expression states. Genome Research 15: 1365-1372.
50. Fong SS, Palsson BO (2004) Metabolic gene-deletion strains of *Escherichia coli* evolve to computationally predicted growth phenotypes. Nature Genetics 36: 1056-1058.
51. Davis BD (1952) Biosynthetic interrelations of lysine, diaminopimelic acid, and threonine in mutants of *Escherichia coli*. Nature (Lond): 534.
52. CASIDA L (1956) Preparation of diaminopimelic acid and lysine. US 2771396A.
53. Rosen BP (1971) Basic amino acid transport in *Escherichia coli*. Journal of Biological Chemistry 246: 3653-3662.

54. Rahmania.M, Claus DR, Oxender DL (1973) Multiplicity of leucine transport-systems in *Escherichia coli* K-12. Journal of Bacteriology 116: 1258-1266.
55. Connor RC (1995) The benefits of mutualism - a conceptual-framework. Biological Reviews of the Cambridge Philosophical Society 70: 427-457.
56. Cooper VS, Schneider D, Blot M, Lenski RE (2001) Mechanisms causing rapid and parallel losses of ribose catabolism in evolving populations of *Escherichia coli* B. Journal of Bacteriology 183: 2834-2841.
57. Mongold JA, Bennett AF, Lenski RE (1996) Evolutionary adaptation to temperature .IV. Adaptation of *Escherichia coli* at a niche boundary. Evolution 50: 35-43.
58. Keasling JD (2010) Manufacturing Molecules Through Metabolic Engineering. Science 330: 1355-1358.
59. Nielsen J (2001) Metabolic engineering. Applied Microbiology and Biotechnology 55: 263-283.
60. Ingram LO, Conway T, Clark DP, Sewell GW, Preston JF (1987) Genetic-Engineering of Ethanol-Production in Escherichia-Coli. Applied and Environmental Microbiology 53: 2420-2425.
61. Steen EJ, Kang YS, Bokinsky G, Hu ZH, Schirmer A, et al. (2010) Microbial production of fatty-acid-derived fuels and chemicals from plant biomass. Nature 463: 559-U182.
62. Beller HR, Goh EB, Keasling JD (2010) Genes Involved in Long-Chain Alkene Biosynthesis in Micrococcus luteus. Applied and Environmental Microbiology 76: 1212-1223.
63. Schirmer A, Rude MA, Li XZ, Popova E, del Cardayre SB (2010) Microbial Biosynthesis of Alkanes. Science 329: 559-562.
64. Hawkins KM, Smolke CD (2008) Production of benzylisoquinoline alkaloids in Saccharomyces cerevisiae. Nature Chemical Biology 4: 564-573.
65. Pfeifer BA, Admiraal SJ, Gramajo H, Cane DE, Khosla C (2001) Biosynthesis of complex polyketides in a metabolically engineered strain of E-coli. Science 291: 1790-1792.
66. Siewers V, San-Bento R, Nielsen J (2010) Implementation of Communication-Mediating Domains for Non-Ribosomal Peptide Production in Saccharomyces cerevisiae. Biotechnology and Bioengineering 106: 841-844.
67. Ro DK, Paradise EM, Ouellet M, Fisher KJ, Newman KL, et al. (2006) Production of the antimalarial drug precursor artemisinic acid in engineered yeast. Nature 440: 940-943.
68. Leonard E, Ajikumar PK, Thayer K, Xiao WH, Mo JD, et al. (2010) Combining metabolic and protein engineering of a terpenoid biosynthetic pathway for overproduction and selectivity control. Proceedings of the National Academy of Sciences of the United States of America 107: 13654-13659.
69. Asadollahi MA, Maury J, Patil KR, Schalk M, Clark A, et al. (2009) Enhancing sesquiterpene production in Saccharomyces cerevisiae through in silico driven metabolic engineering. Metabolic Engineering 11: 328-334.
70. Fong SS, Burgard AP, Herring CD, Knight EM, Blattner FR, et al. (2005) In silico design and adaptive evolution of Escherichia coli for production of lactic acid. Biotechnology and Bioengineering 91: 643-648.
71. Zhang XL, Jantama K, Moore JC, Jarboe LR, Shanmugam KT, et al. (2009) Metabolic evolution of energy-conserving pathways for succinate production in Escherichia

coli. Proceedings of the National Academy of Sciences of the United States of America 106: 20180-20185.
72. Nakamura CE, Whited GM (2003) Metabolic engineering for the microbial production of 1,3-propanediol. Current Opinion in Biotechnology 14: 454-459.
73. Wierckx NJP, Ballerstedt H, de Bont JAM, Wery J (2005) Engineering of solvent-tolerant Pseudomonas putida S12 for bioproduction of phenol from glucose. Applied and Environmental Microbiology 71: 8221-8227.
74. Zha WJ, Rubin-Pitel SB, Shao ZY, Zhao HM (2009) Improving cellular malonyl-CoA level in Escherichia coli via metabolic engineering. Metabolic Engineering 11: 192-198.
75. Mark A. Eiteman EA (2012) Microbial production of pyruvate and pyruvate derivatives.
76. Ravi R. Gokarn MAE, Elliot Altman (1999) Pyruvate carboxylase overexpression for enhanced production of oxaloacetate-derived biochemicals in microbial cells.
77. Ondrey G (2001) Chem Eng: 21.
78. Zhu YH, Eiteman MA, Altman R, Altman E (2008) High Glycolytic Flux Improves Pyruvate Production by a Metabolically Engineered Escherichia coli Strain. Applied and Environmental Microbiology 74: 6649-6655.
79. Tomar A, Eiteman MA, Altman E (2003) The effect of acetate pathway mutations on the production of pyruvate in Escherichia coli. Applied Microbiology and Biotechnology 62: 76-82.
80. Xu GQ, Hua Q, Duan NJ, Liu LM, Chen J (2012) Regulation of thiamine synthesis in Saccharomyces cerevisiae for improved pyruvate production. Yeast 29: 209-217.
81. Wang ZK, Gao CJ, Wang Q, Liang QF, Qi QS (2012) Production of pyruvate in Saccharomyces cerevisiae through adaptive evolution and rational cofactor metabolic engineering. Biochemical Engineering Journal 67: 126-131.
82. Wieschalka S, Blombach B, Eikmanns BJ (2012) Engineering Corynebacterium glutamicum for the production of pyruvate. Applied Microbiology and Biotechnology 94: 449-459.
83. Burgard AP, Pharkya P, Maranas CD (2003) OptKnock: A bilevel programming framework for identifying gene knockout strategies for microbial strain optimization. Biotechnology and Bioengineering 84: 647-657.
84. Yim H, Haselbeck R, Niu W, Pujol-Baxley C, Burgard A, et al. (2011) Metabolic engineering of Escherichia coli for direct production of 1,4-butanediol. Nature Chemical Biology 7: 445-452.
85. Kim J, Reed JL (2010) OptORF: Optimal metabolic and regulatory perturbations for metabolic engineering of microbial strains. Bmc Systems Biology 4.
86. JH M (1992) A short course in bacterial genetics. A laboratory manual and handbook for Escherichia coli and related bacteria: Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.
87. Schwalbach MS, Keating DH, Tremaine M, Marner WD, Zhang YP, et al. (2012) Complex Physiology and Compound Stress Responses during Fermentation of Alkali-Pretreated Corn Stover Hydrolysate by an Escherichia coli Ethanologen. Applied and Environmental Microbiology 78: 3442-3457.
88. Youngquist JT, Lennen RM, Ranatunga DR, Bothfeld WH, Marner WD, et al. (2012) Kinetic modeling of free fatty acid production in Escherichia coli based on continuous cultivation of a plasmid free strain. Biotechnology and Bioengineering 109: 1518-1527.

89. Feist AM, Zielinski DC, Orth JD, Schellenberger J, Herrgard MJ, et al. (2010) Model-driven evaluation of the production potential for growth-coupled products of Escherichia coli. Metabolic Engineering 12: 173-186.
90. Reed JL, Vo TD, Schilling CH, Palsson BO (2003) An expanded genome-scale model of Escherichia coli K-12 (iJR904 GSM/GPR). Genome Biology 4.
91. Peng L, Shimizu K (2003) Global metabolic regulation analysis for Escherichia coli K12 based on protein expression by 2-dimensional electrophoresis and enzyme activity measurement. Applied Microbiology and Biotechnology 61: 163-178.
92. Sawers G, Watson G (1998) A glycyl radical solution: oxygen-dependent interconversion of pyruvate formate-lyase. Molecular Microbiology 29: 945-954.
93. Bologna FP, Campos-Bermudez VA, Saavedra DD, Andreo CS, Drincovich MF (2010) Characterization of Escherichia coli EutD: a Phosphotransacetylase of the Ethanolamine Operon. Journal of Microbiology 48: 629-636.
94. Zhou L, Zuo ZR, Chen XZ, Niu DD, Tian KM, et al. (2011) Evaluation of Genetic Manipulation Strategies on d-Lactate Production by Escherichia coli. Current Microbiology 62: 981-989.
95. Tarmy EM, Kaplan NO (1968) Kinetics of Escherichia Coli B D-Lactate Dehydrogenase and Evidence for Pyruvate-Controlled Change in Conformation. Journal of Biological Chemistry 243: 2587-&.
96. Sawers G, Hesslinger C, Muller N, Kaiser M (1998) The glycyl radical enzyme TdcE can replace pyruvate formate-lyase in glucose fermentation. Journal of Bacteriology 180: 3509-3516.
97. Nagy PL, Marolewski A, Benkovic SJ, Zalkin H (1995) Formyltetrahydrofolate Hydrolase, a Regulatory Enzyme That Functions to Balance Pools of Tetrahydrofolate and One-Carbon Tetrahydrofolate Adducts in Escherichia-Coli. Journal of Bacteriology 177: 1292-1298.
98. Baba T, Ara T, Hasegawa M, Takai Y, Okumura Y, et al. (2006) Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. Molecular Systems Biology 2.
99. Neidhardt FC, John L. Ingraham, and Moselio Schaechter (1990) Physiology of the bacterial cell: a molecular approach. Sunderland, Mass: Sinauer Associates.
100. Toya Y, Nakahigashi K, Tomita M, Shimizu K (2012) Metabolic regulation analysis of wild-type and arcA mutant Escherichia coli under nitrate conditions using different levels of omics data. Mol Biosyst 8: 2593-2604.
101. Kanehisa M (2002) The KEGG database. In Silico Simulation of Biological Processes 247: 91-103.
102. Orth JD, Palsson BO (2012) Gap-filling analysis of the iJO1366 Escherichia coli metabolic network reconstruction for discovery of metabolic functions. Bmc Systems Biology 6.
103. Kumar VS, Maranas CD (2009) GrowMatch: An Automated Method for Reconciling In Silico/In Vivo Growth Predictions. Plos Computational Biology 5.
104. Kim J, Reed JL, Maravelias CT (2011) Large-Scale Bi-Level Strain Design Approaches and Mixed-Integer Programming Solution Techniques. Plos One 6.
105. Pharkya P, Burgard AP, Maranas CD (2004) OptStrain: A computational framework for redesign of microbial production systems. Genome Research 14: 2367-2376.

106. Hatzimanikatis V, Li CH, Ionita JA, Henry CS, Jankowski MD, et al. (2005) Exploring the diversity of complex metabolic networks. Bioinformatics 21: 1603-1609.
107. Henry CS, Broadbelt LJ, Hatzimanikatis V (2010) Discovery and Analysis of Novel Metabolic Pathways for the Biosynthesis of Industrial Chemicals: 3-Hydroxypropanoate. Biotechnology and Bioengineering 106: 462-473.
108. Finley SD, Broadbelt LJ, Hatzimanikatis V (2010) In silico feasibility of novel biodegradation pathways for 1,2,4-trichlorobenzene. Bmc Systems Biology 4.
109. Ma HW, Zeng AP (2003) Reconstruction of metabolic networks from genome data and analysis of their global structure for various organisms. Bioinformatics 19: 270-277.
110. Pey J, Prada J, Beasley JE, Planes FJ (2011) Path finding methods accounting for stoichiometry in metabolic networks. Genome Biology 12.
111. Reed CJTaJL (2014) MapMaker and PathTracer: Automated Carbon Map Generation and Reaction Pathway Determination in Genome-Scale Networks. In preparation.
112. Orth JD, Conrad TM, Na J, Lerman JA, Nam H, et al. (2011) A comprehensive genome-scale reconstruction of Escherichia coli metabolism-2011. Molecular Systems Biology 7.
113. Lian JZ, Chao R, Zhao HM (2014) Metabolic engineering of a Saccharomyces cerevisiae strain capable of simultaneously utilizing glucose and galactose to produce enantiopure (2R,3R)-butanediol. Metabolic Engineering 23: 92-99.
114. Ng CY, Jung MY, Lee J, Oh MK (2012) Production of 2,3-butanediol in Saccharomyces cerevisiae by in silico aided metabolic engineering. Microbial Cell Factories 11.
115. Shin HD, Yoon SH, Wu J, Rutter C, Kim SW, et al. (2012) High-yield production of meso-2,3-butanediol from cellodextrin by engineered E-coli biocatalysts. Bioresource Technology 118: 367-373.
116. Nielsen DR, Yoon SH, Yuan CJ, Prather KLJ (2010) Metabolic engineering of acetoin and meso-2,3-butanediol biosynthesis in E. coli. Biotechnology Journal 5: 274-284.
117. Oliver JWK, Machado IMP, Yoneda H, Atsumi S (2013) Cyanobacterial conversion of carbon dioxide to 2,3-butanediol. Proceedings of the National Academy of Sciences of the United States of America 110: 1249-1254.
118. Ji XJ, Nie ZK, Huang H, Ren LJ, Peng C, et al. (2011) Elimination of carbon catabolite repression in Klebsiella oxytoca for efficient 2,3-butanediol production from glucose-xylose mixtures. Applied Microbiology and Biotechnology 89: 1119-1125.
119. Shen CR, Liao JC (2008) Metabolic engineering of Escherichia coli for 1-butanol and 1-propanol production via the keto-acid pathways. Metabolic Engineering 10: 312-320.
120. Shen CR, Liao JC (2013) Synergy as design principle for metabolic engineering of 1-propanol production in Escherichia coli. Metabolic Engineering 17: 12-22.
121. Choi YJ, Park JH, Kim TY, Lee SY (2012) Metabolic engineering of Escherichia coli for the production of 1-propanol. Metabolic Engineering 14: 477-486.
122. Jain R, Yan YJ (2011) Dehydratase mediated 1-propanol production in metabolically engineered Escherichia coli. Microbial Cell Factories 10.
123. Deng Y, Fong SS (2011) Metabolic engineering of Thermobifida fusca for direct aerobic bioconversion of untreated lignocellulosic biomass to 1-propanol. Metabolic Engineering 13: 570-577.

124. Ammar EM, Wang ZQ, Yang ST (2013) Metabolic engineering of Propionibacterium freudenreichii for n-propanol production. Applied Microbiology and Biotechnology 97: 4677-4690.
125. ltd MRC (2014) Acrylic Acid: 2014 World Market Outlook and Forecast up to 2018.
126. Danner H, Urmos M, Gartner M, Braun R (1998) Biotechnological production of acrylic acid from biomass. Applied Biochemistry and Biotechnology 70-2: 887-894.
127. Obrien DJ, Panzer CC, Eisele WP (1990) Biological Production of Acrylic-Acid from Cheese Whey by Resting Cells of Clostridium-Propionicum. Biotechnology Progress 6: 237-242.
128. Akedo M, Cooney CL, Sinskey AJ (1983) Direct Demonstration of Lactate-Acrylate Interconversion in Clostridium-Propionicum. Bio-Technology 1: 791-794.
129. Straathof AJJ, Sie S, Franco TT, van der Wielen LAM (2005) Feasibility of acrylic acid production by fermentation. Applied Microbiology and Biotechnology 67: 727-734.
130. Werpy T, Petersen G (2004) Top Value Added Chemicals from Biomass: Volume I -- Results of Screening for Potential Candidates from Sugars and Synthesis Gas. DOE/GO-102004-1992; TRN: US200427%%671 United States10.2172/15008859TRN: US200427%%671Tue Feb 12 21:52:38 EST 2008NRELEnglish DOE/GO-102004-1992; TRN: US200427%%671 United States10.2172/15008859TRN: US200427%%671Tue Feb 12 21:52:38 EST 2008NRELEnglish. Medium: ED; Size: 76 pp. pages p.
131. Kumar A, Suthers PF, Maranas CD (2012) MetRxn: a knowledgebase of metabolites and reactions spanning metabolic models and databases. Bmc Bioinformatics 13.
132. Oh M, Yamada T, Hattori M, Goto S, Kanehisa M (2007) Systematic analysis of enzyme-catalyzed reaction patterns and prediction of microbial biodegradation pathways. Journal of Chemical Information and Modeling 47: 1702-1712.
133. Hou BK, Ellis LBM, Wackett LP (2004) Encoding microbial metabolic logic: predicting biodegradation. Journal of Industrial Microbiology & Biotechnology 31: 261-272.
134. Lee M, Smith GM, Eiteman MA, Altman E (2004) Aerobic production of alanine by Escherichia coli aceF ldhA mutants expressing the Bacillus sphaericus alaD gene. Applied Microbiology and Biotechnology 65: 56-60.