Metabolic engineering of microorganisms for the overproduction of fatty acids

by

Ting Wei Tee

A dissertation submitted to the graduate faculty

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Major: Chemical Engineering

Program of Study Committee:
Jacqueline V. Shanks, Major Professor
Laura R. Jarboe
R. Dennis Vigil
David J. Oliver
Marna D. Nelson

Iowa State University

Ames, Iowa

2013

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

I would like to thank my major professor, Dr. Jacqueline V. Shanks for her continuous guidance, support and inspiration for my research. She also serves as my mentor in my professional career and self development. Her constant motivation and advice builds my communication skills to network with peers and industrial folks.

I would also like to thanks my committee members: Dr. Laura Jarboe, Dr. Marna D. Nelson, Dr. David J. Oliver, and Dr. R. Dennis Vigil for serving on my Program of Study committee. Their guidance and support throughout the course of this research are much appreciated.

I am very thankful to my collaborators in my research works: Dr. Laura Jarboe, Dr. Zengyi Shao, Dr. Julie Dickerson, Dr. Ka-Yiu San, Dr, Suzanne Sandmeyer, Dr. Nancy Da Silva, Dr. Costa Maranas, and Dr. Ramon Gonzalez for valuable discussion. I want to also offer my appreciation to their students, especially to Ping Liu, Liam Royce, Dr. Ivan Chang, James Yu, Dr. Mai Li, Dr. Wei Li, Jesse Welsh, Erin Boggess, Al Fu, Dr. Sridhar Ranganathan, Anupam Chowhury, Chris Leber, and Javier Cardenas.

In addition, I would also like to thank my colleagues, friends and BRL 4th Floor folks for endless support and the great time together: Dr. Jong Moon Yoon, Fuyuan Jing, Le Zhao, Gwen Truong, Mark Brown, Jing Tao, Chunyu Lio, etc. I am also grateful to the department faculty and staff for making my time at Iowa State University a wonderful experience.

Finally, I would like to thank to my family and my girlfriend Capella for their endless love, support and encouragement.

ABSTRACT

Fatty acids naturally synthesized in many organisms are promising starting points for the catalytic production of industrial chemicals and diesel-like biofuels. However, bio-production of fatty acids in microbial hosts relies heavily on manipulating tightly regulated fatty acid biosynthetic pathways, thus complicating the engineering for higher yields. With the advent of systems metabolic engineering, we demonstrated an iterative metabolic engineering effort that integrates computationally driven predictions and metabolic flux analysis (MFA) was demonstrated to meet this challenge. With wild type *E. coli* fluxomic data, the OptForce procedure was employed to suggest genetic manipulations for fatty acid overproduction. In accordance with the OptForce prioritization of interventions, *fabZ* and acyl-ACP thioesterase were upregulated and *fadD* was deleted to arrive at a strain that produces 1.70 g/L and 0.14 g fatty acid/g glucose of $C_{14\text{-}16}$ fatty acid in minimal medium. However, OptForce does not infer gene regulation, enzyme inhibition and metabolic toxicity. Along with transcriptomics and metabolomics analysis, we re-deployed OptForce simulation using the redefined flux distribution as constraints to generate predictions for the second generation fatty acid-overproducing strain. MFA identified the up-regulation of the TCA cycle and down-regulation of pentose phosphate pathway under fatty acid overproduction to replenish the need of energy and reducing molecules. The elevation of intracellular metabolite levels in the TCA cycle complemented the flux findings. With re-defined flux boundary of the first generation strain, OptForce suggested the interruption of TCA cycle such as removal of succinate dehydrogenase as the most prioritized genetic intervention to further improve fatty acid production. Meanwhilem, the whole genome transcriptional analysis revealed acid stress response, membrane disruption, colanic acid and biofilm formation during fatty acid production, thus pinpointing the targets for future metabolic engineering effort. These results highlight the benefit of using computational strain design and system metabolic engineering tools in systematically guiding the strain design to produce free fatty acids. Nonetheless, *Saccharomyces cerevisiae* is another attractive host organism for the production of biochemicals and biofuels. However, *S. cerevisiae* is very susceptible to octanoic acid toxicity. Transcriptomics analysis revealed membrane stress and intracellular acidification during octanoic acid stress. MFA illustrated the increase of flux in the TCA cycle possibly to

facilitate the ATP-binding-cassette transporter activities. Further efforts can focus on improving membrane integrity or explore oleaginious yeasts to enhance the tolerance against fatty acids.

CHAPTER 1

INTRODUCTION

Introduction of chemical industry

The production of industrial chemicals is a trillion-dollar global enterprise that impacts all aspects of the society from personal care products to building materials. Traditionally, this industry is highly dependent on crude oil and natural gas as the feedstock (Jones, 2009). Hydrocarbons from crude oils are cracked thermally to form ethylene, propylene and benzene, as the primary chemical building blocks. These platform chemicals can be converted chemically to a wide range of industrial chemicals including polymers, textiles, paints, solvents, detergents and lubricants (Nikolau et al., 2008). High yields in hydrocarbon feedstock conversion, combined with highly optimized process flow, results in an efficient production of the constituent chemicals.

Nonetheless, the current production of petrochemicals via crude oils is not sustainable. Given the intrinsic need for carbon feedstock, renewable feedstock that can fulfill the long-term demand of the chemical industry is being researched. Biorefinery, which uses biomass to produce fuels, chemicals and materials, can replace the traditional fossil-based refinery, addressing environmental and energy concerns (Figure 1). We opt to utilize sunlight-derived crops as the raw materials for the biorefinery to produce a diverse array of final products (i.e. chemicals, energy and materials). Such biorefinery offers high investment return, and fulfills energy and economic goals simultaneously (Bozell and Petersen, 2010). Its diversity of products offers great flexibility to meet the changes in market demands, and allows the refinery to to secure feedstock from multiple sources.

The emergence of biorefinery requires the development of new platform biochemical, in which they will be transformed into a variety of chemical end-products. These platform biochemicals have unique properties and functionalities to serve as functional or direct replacements of currently used petrochemicals (Nikolau et al., 2008). Among the abundant platform chemical choices, the carboxylic acids and pyrones are chosen as the target chemicals for optimizations via metabolic engineering tools in *Escherichia coli* and *Saccharomyces cerevisiae*. The robustness, fast division, pH tolerance, simple nutrient requirement, completely

sequenced genome and long history as industrial workhost make *E. coli* and *S. cerevisiae* excellent biocatalysts to produce fatty acids and pyrones. Moreover, these microbes are proven to be excellent model organisms in metabolic engineering to knockout native genes and insert heterogeneous genes for the production of alcohols, amino acids, vitamins, organics acids, biofuels, etc.

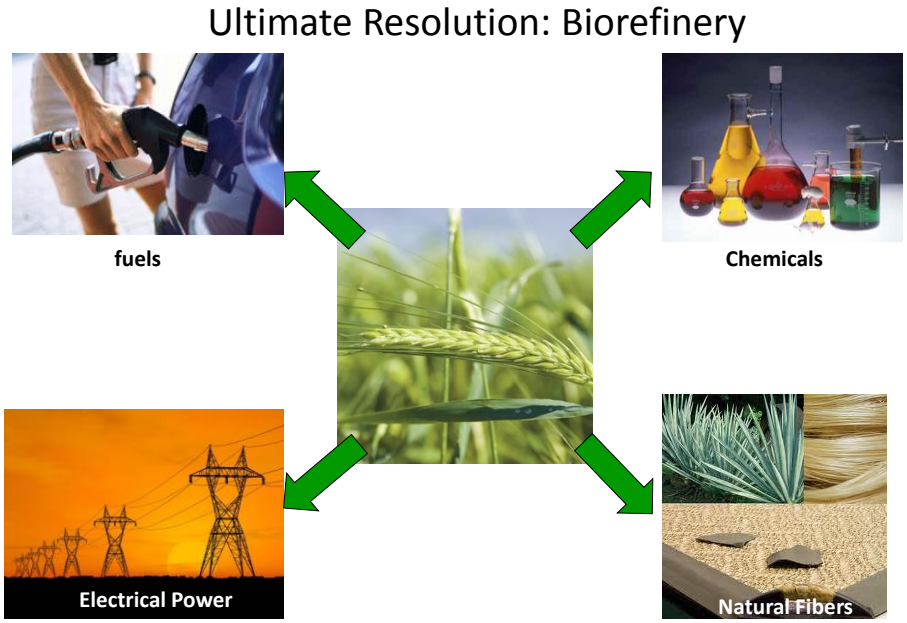


Figure 1. The concept of biorefinery using sunlight-derived crops to produce energies, chemicals and materials.

Carboxylic acids and pyrones

Carboxylic/ fatty acid recently gained extensive highlight in the global research community due to its potential as fuels and chemicals (Nikolau et al., 2008; Stephanopoulos, 2007). Acyl-ACP substrates of a particular chain length can be specifically hydrolyzed by plant thioesterases (Jing et al., 2011), thereby producing novel fatty acids. Medium chain fatty acids (C_6-C_{14}) can be applied as detergents, lubricants, cosmetics, and pharmaceuticals. Fatty acids could be catalytically deoxygenated via metal catalysts to produce α-olefins, which serve as the building blocks of important polymerization products. Carboxylic acids can be synthesized in a single metabolic pathway (i.e. polyketide/fatty acid biosynthesis) (Nikolau et al., 2008). Lennen et al. reported free fatty acid production in *E. coli* and decarboxylation of fatty acids into saturated alkanes via heterogeneous catalysis (Lennen et al., 2010). Odd-numbered mixture of

hydrocarbons (alkanes and alkenes) can be directly synthesized in engineered *E. coli* via fatty acyl-acyl carrier protein (ACP) with the expression of acyl ACP reductase and aldehyde decarbonylase from cyanobacteria (Dominguez de Maria, 2011). Steen and coworkers had successfully demonstrated the microbial production of structurally tailored fatty esters (biodiesel), fatty alcohols, and waxes branched from free fatty acids (Steen et al., 2010).

Triacetic acid lactone (TAL) can be converted into a wide array of valuable chemical intermediates and products including acetylacetone, bifunctional ketones, sorbic acid and γ-caprolactone through heterogeneous catalysts or thermal decomposition (Chia et al., 2012). TAL is the precursor for phloroglucinol production through multiple chemical steps via intermediacy of methyl esters. Phloroglucinol can then be converted into 1,3,5-triamino-2,4,6-trinitrobenzene and resorcinol for pharmaceutical applications (Achkar et al., 2005; Hansen and Frost, 2002; Xie et al., 2006; Zha et al., 2004). TAL synthesis involves the fatty acid synthase pathway, which condenses two malonyl-CoAs with an acetyl-CoA molecule and then undergoes cyclization to generate TAL (Xie et al., 2006). TAL has been discovered to be synthesized by *Gerbera hybrida* 2-pyrone synthase, mutated *Brevibacterium ammoniagenes* fatty acid synthase β, and mutated *Penicillium patulum* 6-methylsalicylic acid synthase (Xie et al., 2006). Heterologous expression of 2-pyrone synthase, 6-methylsalicylic acid synthase or other genetically modified polyketide synthases in genetic-modified *E. coli* and *S. cerevisiae* could yield up to 6% theoretical maximum with 1.8g/L TAL (Eckermann et al., 1998; Xie et al., 2006).

Metabolic engineering cycle

In order to make the microbial fermentation route economically feasible, the target chemicals (carboxylic acids and pyrones) must be produced at high yield, titer and productivity, which remain challenging (Jarboe et al., 2010). Every biological system is regulated by a hierarchical structure of information from genomics, transcriptomics, proteomics, metabolomics and fluxomics (Figure 2). To overcome these challenges, the role of metabolic engineering which involves an iterative process among strain development, evolution, characterization and bioinformatics tool, is crucial (Figure 3). To date, metabolic engineering has evolved to the Systems Metabolic Engineering by the integrated use of systems biology, synthetic biology, and evolutionary engineering (Lee et al., 2011). The latest development of high-throughput techniques for analyzing omics data, combined with computational tools, can predict gene targets for modification to enhance production of desired molecules (Lee et al., 2012). Integrations of

transcriptomic, proteomic, metabolomic and fluxomic approaches are increasingly applied to unravel the complexity of biological systems.

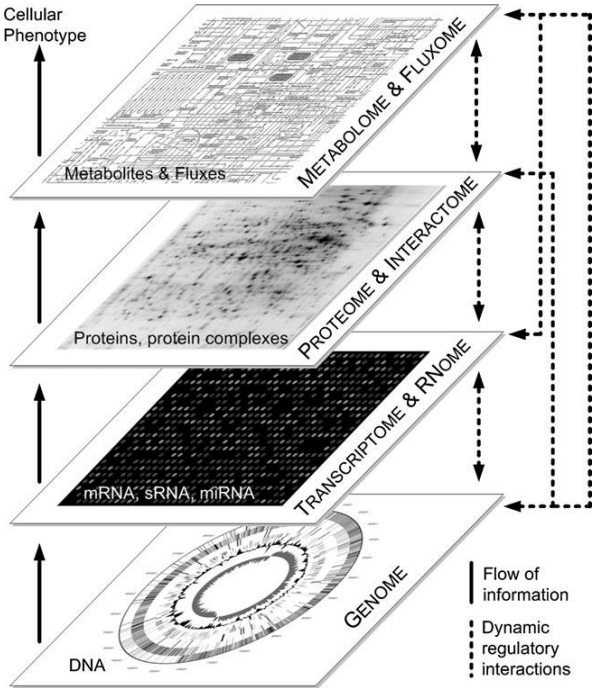


Figure 2. Illustration of different levels of information in a cellular system. Solid line represents flow of information; dashed line represents interaction between molecular species. (Kohlstedt, Becker and Wittmann 2010)

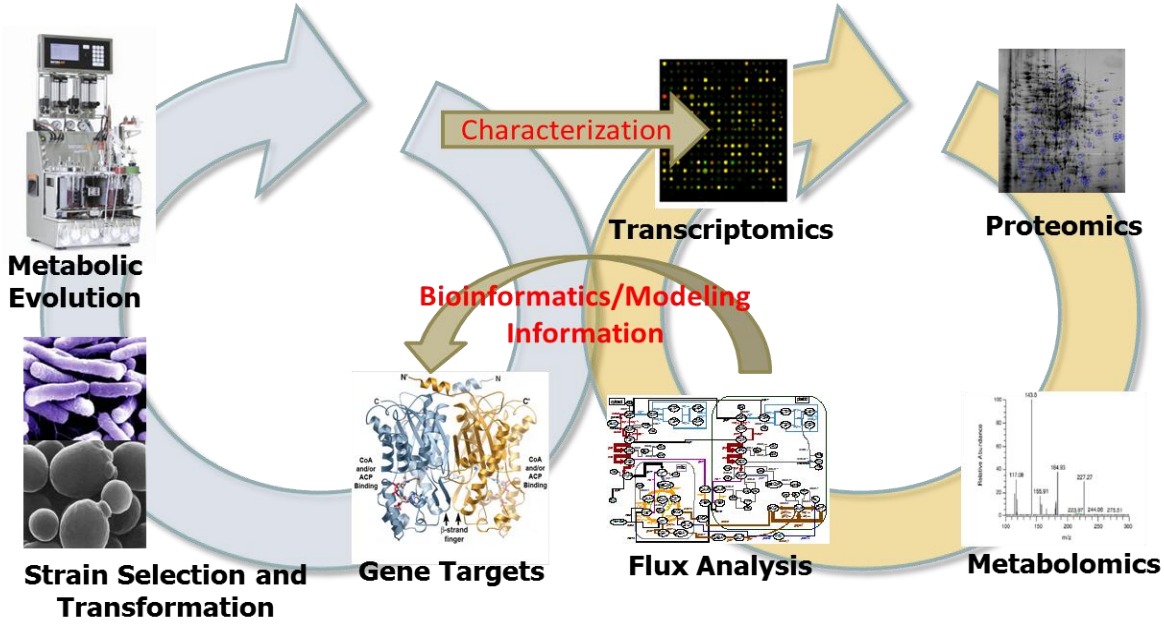


Figure 3 Cyclic iteration of metabolic engineering strategy using omics technologies: trascriptomic data from RNA-seq or from microarray; proteomics data from 2-D gel with mass spectrometer; metabolite profilings from GC/MS and LC/MS; ^{13}C metabolic flux analysis. Direct

evolution approach parallels with rational engineering approach. Bioinformatics and modeling information are integrated with omics data to predict the next round interventions. (Adapted from CBiRC)

Transcriptomics

The availability of complete genome sequences of a variety of microorganisms has boosted the avenue towards understanding cellular physiology at the system level (Park and Lee, 2008). Genomic information serves as the foundation for transcriptome and proteome profiling in which data analysis depends on the genome sequence. Transcriptome analysis, either by DNA microarray or sequencing-based quantification, allows examination of mRNA transcript levels for thousands of genes of multiple strains simultaneously (Duggan et al., 1999; Wang et al., 2009). We can understand cell physiology and regulatory mechanisms at the whole-cell transcript level by analyzing them under various genotypic and environmental conditions. Transcriptome analysis has proven to be a powerful tool in elucidation of useful metabolic genes, identification of novel target genes for improving strain performance, and identification of regulator controlling pathway for target molecules production (Hirasawa et al., 2010; Serrania et al., 2008; Yano et al., 2003). Jarboe et al. recently reviewed the use of transcriptome study to guide engineering of inhibitor tolerance for carboxylic acids, furfural and butanol productions (Jarboe et al., 2011).

For instance, Park et al. demonstrated the combination of genome engineering with transcriptome profiling and in-silico computation to design for an efficient L-valine producing *E. coli* (Park et al., 2007). Feedback inhibition and transcriptional attenuation were firstly removed by site-directed gene mutation. Competing pathways were blocked by deleting corresponding genes, and ilvBN operon involved in the first reaction of L-valine biosynthesis was overexpressed. By having the above genetic modifications, the engineered strain yielded 0.066 g L-valine/g glucose. Further improvement was achieved by performing transcriptomic analysis between the engineered and control strains. The *ilvCED*, *lrp* and *ygaZH* genes, identified by transcriptome profiling, were overexpressed, improving the yield to 0.152 g L-valine/g glucose. The authors incorporated *in-silico* gene knockout simulation to identify and delete *aceF*, *mdh*, and *pfkA* genes, resulting in 7.55g/L L-valine from batch culture with 0.378 g of L-valine/g glucose yield.

Proteomics

Proteins, the end products of genes, are major components for building the cellular structure, catalytic enzymes in metabolic pathways, and serve as signal transduction proteins in regulatory pathways. However, there is not always a one-to-one correspondence between gene transcript level and protein abundance (Vemuri and Aristidou, 2005). Proteome analysis could examine the levels of proteins and their changes under particular genetic and environmental conditions (Zhang et al., 2010). It can capture the regulation of cellular response beyond mRNAs, providing the information of complicated biological processes and post-translational modifications (Han and Lee, 2006).

One grand challenge of proteome analysis is to efficiently identify and separate a protein of interest from a cellular proteome, since proteins have similar physical attributes. Proteomics has relied primarily on two separation techniques for analysis: (i) 2D polyacrylamide gel electrophoresis (PAGE), in which proteins are separated by isoelectric point and mass coupled with mass spectrometry; (ii) gel-free liquid chromatography(LC) coupled with mass/tandem mass spectrometry. However, neither 2D PAGE nor LC-MS/MS method is close to saturating the identifications in small bacteria. Typically only 20-40% of the proteome can be identified without quantitative information (Vemuri and Aristidou, 2005). Numerous strategies for the identification of proteins and workflows to quantify protein abundance have also been developed (Han and Lee, 2006; Mukhopadhyay et al., 2008; Redding-Johanson et al., 2011). Stable isotope labeling-based isotope-coded affinity tags (ICAT), isobaric tag for relative and absolute quantification (iTRAQ) and label-free comparative quantitative proteomics are employed to quantitatively obtain proteome measurements. Even though it does not provide a true global picture owing to unknown proteins, it can be used to identify genes to be engineered to develop superior strains (Park and Lee, 2008). One successful proteomics example is illustrated by Redding-Johanson and coworkers by measuring protein abundance in engineered sesquiterpene-producing *E.coli* strains using a targeted proteomic approach via selected-reaction monitoring (SRM) mass spectrometry (Redding-Johanson et al., 2011). While the majority of the pathway proteins were detected at an appreciable level, the relative low abundance of mevalonatekinase (MK) and phosphomevalonatekinase (PMK) from *Saccharomyces cerevisiae* were identified as potential bottlenecks. The problem was overcome by codon-optimizing gene encoding MK and PMK and expressing stronger promoter, leading to over 3-fold improvement in the final sesquiterpene titer (4.5g/L).

Metabolomics

The rate of enzymatic reaction is controlled by the concentration of substrate and product by feedback regulation, and there is no clear correlation between genes and cellular metabolites. Thus, knowledge of metabolome presents a snapshot of the cell physiological state, which reflects the integrated output of a multitude of complex interactions (Ewald et al., 2009). The changes in intracellular metabolite concentrations unravel metabolic regulation such as allosteric control and metabolite-DNA binding (Vemuri and Aristidou, 2005). The goal of metabolomics analysis is to determine all the intracellular metabolites qualitatively and quantitatively. Nonetheless, quenching the cell activity instantly and extracting intracellular metabolites without cell leakage remain formidable challenges. Winder et al. and Canelas et al. reviewed and evaluated organism-specific standard operating procedures for quenching and extracting intracellular metabolites with minimum cell leakage from *E. coli* and *S. cerevisiae* respectively (Canelas et al., 2008; Canelas et al., 2009; Mashego et al., 2007; Winder et al., 2008). Even with advances in the analytical techniques, we cannot detect all the metabolites but only possibly identify thousands of metabolites and quantify ten to hundreds of the targeted metabolites (Mukhopadhyay et al., 2008). Gas chromatography time-of-flight mass spectrometry (GC-TOF), high-performance liquid chromatography mass spectrometry (LC-MS) and capillary electrophoresis mass spectrometry (CE-MS) instruments are typically adopted for metabolomics analysis (van der Werf et al., 2007; Zhang et al., 2010). Park et al. recently demonstrated the SRM LC-MS/MS method to identify some metabolites in the central carbon metabolism and measure the concentration change throughout the growth phase during batch culture (Park et al., 2011). Wu et al. proposed more accurate reliable way to quantify microbial metabolome using uniformly ^{13}C-labeled metabolites as internal standard in the metabolite extraction procedure, followed by liquid chromatography-electrospray ionization-tandem mass spectrometry (LC-ESI-MS/MS) analysis (Wu et al., 2005). One successful application of metabolomics is Wisselink et al. compared the intracellular metabolites concentrations between *S. cerevisiae* after evolution to ferment the pentose sugar and its non-evolved ancestor (Wisselink et al., 2010). Elevation of intracellular concentration of pentose phosphate pathway intermediates and up-regulation of transketolase and transaldolase isoenzymes (*TKL2* and *YGR043c*) suggested the important role of these genes in arabinose fermentation. This hypothesis was proven with 21% reduction of maximum specific growth rate on arabinose when *TKL2* and *YGR043c* gene were deleted.

Fluxomics

Among the different levels of information, metabolic flux distribution in central carbon metabolism is the most direct approach to understand the complex metabolic control mechanism of the whole cell (Matsuoka and Shimizu, 2010). Metabolic flux defined as in vivo enzymatic reaction rate is the manifestation of gene, protein expressions and the concentrations of intracellular metabolites. The goal of metabolic flux analysis (MFA) is to quantify intracellular metabolic fluxes and to scrutinize the functional aspects of metabolic network in depth (Kim et al., 2008). The principle of MFA is based on mass conservation around metabolites, wherein the intracellular fluxes can be calculated from measured specific rate with the incorporation of stoichiometry of metabolic reactions. The MFA can generally be divided into two categories: (i) constraints-based flux balance analysis (FBA); (ii) ^{13}C-MFA. Both these MFA methods have been widely applied for the following applications (Kohlstedt et al., 2010; Tang et al., 2009):

i. Discovery of novel pathways

   The rich isotopic information provides insights for actual cell metabolism, leading to distinct novel pathways. Tang et al. used ^{13}C flux experiment to reveal the absence of complete serine-isocitrate and the existence of a complete tricarboxylic acid (TCA) cycle pathway in *Shewanella oneidensis* under anaerobic conditions (Tang et al., 2007). Despite the common serine-isocitrate lyase pathway in methlylotrophic anaerobes, *Shewanella oneidensis* synthesizes glyoxylate via the isocitrate lyase reaction and converts to glycine. Glycine is indeed oxidized via a highly reversible degradation pathway along with the significant activity in anaplerotic pathways (malic enzyme and phosphoenolpyruvate carboxylase).

ii. Unraveling pathway function

   MFA has led to new understanding of the role of certain pathways beyond the classically attributed functions. For instance, pentose phosphate (PP) pathway was originally assumed to provide building blocks for biomass biosynthesis. MFA deciphers the flux through PPP far beyond the need for anabolism. In addition to biosynthesis requirement, PP pathway is also associated with catabolic breakdown and redox metabolism. The glucose-6-phosphate node at the central carbon metabolism acts to partition carbon fluxes between glycolysis and PP pathway. Blank

et al. suggested the flux through the PP pathway in yeasts is driven by the demand for NADPH (Blank et al., 2005).

iii. Pathway bottleneck identification to optimize biomass/ target molecule production

MFA drives rational engineering of cellular metabolism for target product biosynthesis. The systematic flux quantification in lysine-producing *Corynebacterium glutamicum* strains has revealed key pathways for genetic modification to optimize the production of desired molecules. Based on ^{13}C fluxomics, Becker et al. overexpressed *zwf* gene in the PP pathway to increase the supply of NADPH (which is required for lysine biosyntheisis), and amplified the expression of gluconeogenesis gene (fructose 1,6-bisphosphatase), successfully increasing lysine yield up to 70% (Becker et al., 2007). Moreover, downregulating isocitrate dehydrogenase in the TCA cycle to channel flux toward anaplerotic carboxylation also helps improve lysine production (Becker et al., 2009).

iv. Understanding rigidity of metabolic network

MFA is also manifested to unravel control mechanism in biological systems. We could study cellular response to maintain major growth characteristics and compensate external perturbation upon gene deletions. For instance, deletion of pyruvate kinase in *C. glutamicum* showed similar overall growth behaviors compared to those of the parent strain. However, there are slight decrease in lysine formation and slight increases in dihydroxyacetone and glycerol secretion. Hereby, local flux redirection of metabolic flux seems to be the mechanism to maintain major growth characteristics, fulfilling all anabolic and catabolic needs (Becker et al., 2008). The lack of pyruvate kinase is compensated by activation of phosphoenopyurvate (PEP) carboxylase, malate dehydrogenase and malic enzyme to create a by-pass from PEP to pyruvate, while other fluxes in the network remain rigid.

Optimization-based approaches

In-silico genome-scale metabolic models and associated simulation strategies can be applied to identify gene manipulation targets that would improve desired phenotypes (e.g. increased production yield, maximized growth) (Lee et al., 2012). Constraint based flux balance analysis (FBA) emerges as *in-silico* genome-scale optimization based simulation technique. This technique is based on mass balance stoichiometry around metabolites under pseudo-steady state

coupled with an objective function (e.g. maximization of cell growth rate). With the development of new algorithms, the simulation result becomes more accurate and reliable. Bi-level approach can be used to identify the best production strategy given a maximum number of genetic manipulations (Reed et al., 2010). OptKnock couples cell growth rate with product formation rate to predict gene deletion targets leading to overproduction of target chemicals by ensuring carbon, redox, potential and energy balance (Burgard et al., 2003). In addition to straightforward competing pathway deletion, OptKnock also suggests complex and non-intuitive mechanisms to compensate for the removed functionalities. For example, Yim et al. adapted OptKnock simulation to delete pathways in the common fermentation products lactate, formate and ethanol. A less intuitive manipulation by OptKnock suggested to engineer the oxidative branch of the TCA cycle to function under microaerobic cultivation conditions (Yim et al., 2011).

More recently, OptForce was developed to use bi-level approach to identify how metabolic fluxes must change to improve metabolite production, thereby identifying reaction manipulation targets (Ranganathan et al., 2010). OptForce overlays the *in-vivo* flux measurement of wild type strain with the metabolite overproducing condition to determine which reaction fluxes must change. It is independent on assumption about what objective functions are used to predict cellular behavior. Xu et al. implemented OptForce to predict the minimal intervention that cooperatively force carbon flux toward malonyl-CoA in *E. coli* (Xu et al., 2011). Strain construction with the identified knockout targets (Δ*fumC* and Δ*sucC*) and overexpression targets (ACC, PGK, GAPD and PDH), was reported to exhibit 4-fold increase in level of intracellular malonyl-CoA compared to the wild type strain.

Organization of dissertation

My dissertation is organized as described below:

Chapter 1 presents the overview of biorenewable chemicals and the roles of systems metabolic engineering.

Chapter 2 presents literature review of metabolic flux analysis.

Chapter 3 presents the case study of fatty acid production using systems metabolic engineering design.

Chapter 4 describes the integration of experimental and computational approaches to predict genetic intervention for the overproduction of fatty acids in *E. coli*.

Chapter 5 continues the engineering effort of chapter 4 by depicting the flux distributions of the first generation engineered *E. coli*. Along with transcriptome and metabolome analysis, we integrated the flux data with computational tool to predict the next generation of gene intervention, thus closing the metabolic engineering loop.

Chapter 6 describes the toxicity effect of octanoic acid in *S. cerevisiae* and the role of metabolic flux analysis and transcriptome analysis in unraveling the cell mechanism against octanoic acid toxicity.

Chapter 7 describes the phenotype and carbon flux distribution of oleaginous yeast *Yarrowia lipolytica* as the potential host strain for carboxylic acids and pyrone production.

Chapter 8 summarizes the dissertation and provides prospective for future works.

References

Achkar, J., M. Xian, H. Zhao, and J. W. Frost, 2005, Biosynthesis of Phloroglucinol: Journal of the American Chemical Society, v. 127, p. 5332-5333.

Becker, J., C. Klopprogge, A. Herold, O. Zelder, C. J. Bolten, and C. Wittmann, 2007, Metabolic flux engineering of l-lysine production in Corynebacterium glutamicum—over expression and modification of G6P dehydrogenase: Journal of Biotechnology, v. 132, p. 99-109.

Becker, J., C. Klopprogge, H. Schroeder, and C. Wittmann, 2009, Metabolic Engineering of the Tricarboxylic Acid Cycle for Improved Lysine Production by Corynebacterium glutamicum: Applied and Environmental Microbiology, v. 75.

Becker, J., C. Klopprogge, and C. Wittmann, 2008, Metabolic responses to pyruvate kinase deletion in lysine producing Corynebacterium glutamicum: Microbial Cell Factories, v. 7, p. 8.

Blank, L. M., F. Lehmbeck, and U. Sauer, 2005, Metabolic-flux and network analysis in fourteen hemiascomycetous yeasts: FEMS Yeast Research, v. 5, p. 545-558.

Bozell, J. J., and G. R. Petersen, 2010, Technology development for the production of biobased products from biorefinery carbohydrates-the US Department of Energy's "Top 10" revisited: Green Chemistry, v. 12.

Burgard, A. P., P. Pharkya, and C. D. Maranas, 2003, OptKnock: A bilevel programming framework for identifying gene knockout strategies for microbial strain optimization: Biotechnology and Bioengineering, v. 84, p. 647-657.

Canelas, A. B., C. Ras, A. ten Pierick, J. C. van Dam, J. J. Heijnen, and W. M. Van Gulik, 2008, Leakage-free rapid quenching technique for yeast metabolomics: Metabolomics, v. 4, p. 226-239.

Canelas, A. B., A. ten Pierick, C. Ras, R. M. Seifar, J. C. van Dam, W. M. van Gulik, and J. J. Heijnen, 2009, Quantitative Evaluation of Intracellular Metabolite Extraction Techniques for Yeast Metabolomics: Analytical Chemistry, v. 81, p. 7379-7389.

Chia, M., T. J. Schwartz, B. H. Shanks, and J. A. Dumesic, 2012, Triacetic acid lactone as a potential biorenewable platform chemical: Green Chemistry, v. 14, p. 1850-1853.

Dominguez de Maria, P., 2011, Recent developments in the biotechnological production of hydrocarbons: paving the way for bio-based platform chemicals: ChemSusChem, v. 4.

Duggan, D. J., M. Bittner, Y. D. Chen, P. Meltzer, and J. M. Trent, 1999, Expression profiling using cDNA microarrays: Nature Genetics, v. 21.

Eckermann, S., G. Schroder, J. Schmidt, D. Strack, R. A. Edrada, Y. Helariutta, P. Elomaa, M. Kotilainen, I. Kilpelainen, P. Proksch, T. H. Teeri, and J. Schroder, 1998, New pathway to polyketides in plants: Nature, v. 396.

Ewald, J. C., S. Heux, and N. Zamboni, 2009, High-Throughput Quantitative Metabolomics: Workflow for Cultivation, Quenching, and Analysis of Yeast in a Multiwell Format: Analytical Chemistry, v. 81.

Han, M. J., and S. Y. Lee, 2006, The Escherichia coli proteome: Past, present, and future prospects: Microbiology and Molecular Biology Reviews, v. 70, p. 362-+.

Hansen, C. A., and J. W. Frost, 2002, Deoxygenation of Polyhydroxybenzenes: An Alternative Strategy for the Benzene-Free Synthesis of Aromatic Chemicals: Journal of the American Chemical Society, v. 124, p. 5926-5927.

Hirasawa, T., C. Furusawa, and H. Shimizu, 2010, Saccharomyces cerevisiae and DNA microarray analyses: what did we learn from it for a better understanding and exploitation of yeast biotechnology?: Applied Microbiology and Biotechnology, v. 87.

Jarboe, L. R., P. Liu, and L. A. Royce, 2011, Engineering inhibitor tolerance for the production of biorenewable fuels and chemicals: Current Opinion in Chemical Engineering, v. 1, p. 38-42.

Jarboe, L. R., X. L. Zhang, X. Wang, J. C. Moore, K. T. Shanmugam, and L. O. Ingram, 2010, Metabolic Engineering for Production of Biorenewable Fuels and Chemicals: Contributions of Synthetic Biology: Journal of Biomedicine and Biotechnology, p. 18.

Jing, F., D. C. Cantu, J. Tvaruzkova, J. P. Chipman, B. J. Nikolau, M. D. Yandeau-Nelson, and P. J. Reilly, 2011, Phylogenetic and experimental characterization of an acyl-ACP

thioesterase family reveals significant diversity in enzymatic specificity and activity: Bmc Biochemistry, v. 12.

Jones, R. F., 2009, The Future of the US Chemical Industry: Future of the Chemical Industry, v. 1026.

Kim, H. U., T. Y. Kim, and S. Y. Lee, 2008, Metabolic flux analysis and metabolic engineering of microorganisms: Molecular BioSystems, v. 4, p. 113-120.

Kohlstedt, M., J. Becker, and C. Wittmann, 2010, Metabolic fluxes and beyond-systems biology understanding and engineering of microbial metabolism: Applied Microbiology and Biotechnology, v. 88.

Lee, J. W., T. Y. Kim, Y.-S. Jang, S. Choi, and S. Y. Lee, 2011, Systems metabolic engineering for chemicals and materials: Trends in Biotechnology, v. 29.

Lee, J. W., D. Na, J. M. Park, J. Lee, S. Choi, and S. Y. Lee, 2012, Systems metabolic engineering of microorganisms for natural and non-natural chemicals: Nature Chemical Biology, v. 8.

Lennen, R. M., D. J. Braden, R. M. West, J. A. Dumesic, and B. F. Pfleger, 2010, A Process for Microbial Hydrocarbon Synthesis: Overproduction of Fatty Acids in Escherichia coli and Catalytic Conversion to Alkanes: Biotechnology and Bioengineering, v. 106.

Mashego, M. R., K. Rumbold, M. De Mey, E. Vandamme, W. Soetaert, and J. J. Heijnen, 2007, Microbial metabolomics: past, present and future methodologies: Biotechnology Letters, v. 29.

Matsuoka, Y., and K. Shimizu, 2010, Current status of C-13-metabolic flux analysis and future perspectives: Process Biochemistry, v. 45, p. 1873-1881.

Mukhopadhyay, A., A. M. Redding, B. J. Rutherford, and J. D. Keasling, 2008, Importance of systems biology in engineering microbes for biofuel production: Current Opinion in Biotechnology, v. 19, p. 228-234.

Nikolau, B. J., M. A. D. N. Perera, L. Brachova, and B. Shanks, 2008, Platform biochemicals for a biorenewable chemical industry: Plant Journal, v. 54, p. 536-545.

Park, C., Y. J. Lee, S. Y. Lee, H. B. Oh, and J. Lee, 2011, Determination of the Intracellular Concentrations of Metabolites in Escherichia coli Collected during the Exponential and Stationary Growth Phases using Liquid Chromatography-Mass Spectrometry: Bulletin of the Korean Chemical Society, v. 32, p. 524-530.

Park, J. H., K. H. Lee, T. Y. Kim, and S. Y. Lee, 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, v. 104, p. 7797-7802.

Park, J. H., and S. Y. Lee, 2008, Towards systems metabolic engineering of microorganisms for amino acid production: Current Opinion in Biotechnology, v. 19.

Ranganathan, S., P. F. Suthers, and C. D. Maranas, 2010, OptForce: An Optimization Procedure for Identifying All Genetic Manipulations Leading to Targeted Overproductions: Plos Computational Biology, v. 6.

Redding-Johanson, A. M., T. S. Batth, R. Chan, R. Krupa, H. L. Szmidt, P. D. Adams, J. D. Keasling, T. S. Lee, A. Mukhopadhyay, and C. J. Petzold, 2011, Targeted proteomics for metabolic pathway optimization: Application to terpene production: Metabolic Engineering, v. 13, p. 194-203.

Reed, J. L., R. S. Senger, M. R. Antoniewicz, and J. D. Young, 2010, Computational Approaches in Metabolic Engineering: Journal of Biomedicine and Biotechnology.

Serrania, J., F.-J. Vorhoelter, K. Niehaus, A. Puehler, and A. Becker, 2008, Identification of Xanthomonas campestris pv. campestris galactose utilization genes from transcriptome data: Journal of Biotechnology, v. 135.

Steen, E. J., Y. Kang, G. Bokinsky, Z. Hu, A. Schirmer, A. McClure, S. B. del Cardayre, and J. D. Keasling, 2010, Microbial production of fatty-acid-derived fuels and chemicals from plant biomass: Nature, v. 463.

Stephanopoulos, G., 2007, Challenges in engineering microbes for biofuels production: Science, v. 315.

Tang, Y. J., H. G. Martin, S. Myers, S. Rodriguez, E. E. K. Baidoo, and J. D. Keasling, 2009, ADVANCES IN ANALYSIS OF MICROBIAL METABOLIC FLUXES VIA C-13 ISOTOPIC LABELING: Mass Spectrometry Reviews, v. 28.

Tang, Y. J., A. L. Meadows, J. Kirby, and J. D. Keasling, 2007, Anaerobic central metabolic pathways in Shewanella oneidensis MR-1 reinterpreted in the light of isotopic metabolite Labeling: Journal of Bacteriology, v. 189, p. 894-901.

van der Werf, M. J., K. M. Overkamp, B. Muilwijk, L. Coulier, and T. Hankemeier, 2007, Microbial metabolomics: Toward a platform with full metabolome coverage: Analytical Biochemistry, v. 370.

Vemuri, G. N., and A. A. Aristidou, 2005, Metabolic engineering in the -omics era: Elucidating and modulating regulatory networks: Microbiology and Molecular Biology Reviews, v. 69.

Wang, Z., M. Gerstein, and M. Snyder, 2009, RNA-Seq: a revolutionary tool for transcriptomics: Nature Reviews Genetics, v. 10.

Winder, C. L., W. B. Dunn, S. Schuler, D. Broadhurst, R. Jarvis, G. M. Stephens, and R. Goodacre, 2008, Global metabolic profiling of Escherichia coli cultures: An evaluation of methods for quenching and extraction of intracellular metabolites: Analytical Chemistry, v. 80.

Wisselink, H. W., C. Cipollina, B. Oud, B. Crimi, J. J. Heijnen, J. T. Pronk, and A. J. A. van Maris, 2010, Metabolome, transcriptome and metabolic flux analysis of arabinose fermentation by engineered Saccharomyces cerevisiae: Metabolic Engineering, v. 12.

Wu, L., M. R. Mashego, J. C. van Dam, A. M. Proell, J. L. Vinke, C. Ras, W. A. van Winden, W. M. van Gulik, and J. J. Heijnen, 2005, Quantitative analysis of the microbial metabolome by isotope dilution mass spectrometry using uniformly C-13-labeled cell extracts as internal standards: Analytical Biochemistry, v. 336, p. 164-171.

Xie, D. M., Z. Y. Shao, J. H. Achkar, W. J. Zha, J. W. Frost, and H. M. Zhao, 2006, Microbial synthesis of triacetic acid lactone: Biotechnology and Bioengineering, v. 93, p. 727-736.

Xu, P., S. Ranganathan, Z. L. Fowler, C. D. Maranas, and M. A. G. Koffas, 2011, Genome-scale metabolic network modeling results in minimal interventions that cooperatively force carbon flux towards malonyl-CoA: Metabolic Engineering, v. 13, p. 578-587.

Yano, S., T. Asano, N. Kurose, J. Hiramatsu, H. Shimoi, and K. Ito, 2003, Characterization of an alpha-ketoglutarate-resistant sake yeast mutant with high organic acid productivity: Journal of Bioscience and Bioengineering, v. 96, p. 332-336.

Yim, H., R. Haselbeck, W. Niu, C. Pujol-Baxley, A. Burgard, J. Boldt, J. Khandurina, J. D. Trawick, R. E. Osterhout, R. Stephen, J. Estadilla, S. Teisan, H. B. Schreyer, S. Andrae, T. H. Yang, S. Y. Lee, M. J. Burk, and S. Van Dien, 2011, Metabolic engineering of Escherichia coli for direct production of 1,4-butanediol: Nat Chem Biol, v. 7, p. 445-452.

Zha, W., Z. Shao, J. W. Frost, and H. Zhao, 2004, Rational Pathway Engineering of Type I Fatty Acid Synthase Allows the Biosynthesis of Triacetic Acid Lactone from d-Glucose in Vivo: Journal of the American Chemical Society, v. 126, p. 4534-4535.

Zhang, W. W., F. Li, and L. Nie, 2010, Integrating multiple 'omics' analysis for microbial biology: application and methodologies: Microbiology-Sgm, v. 156, p. 287-301.

CHAPTER 2

LITERATURE REVIEW

Metabolic flux analysis (MFA)

Metabolic Flux Analysis (MFA) is a fundamental metabolic engineering tool to understand *in vivo* cell physiology by integrating system response for gene-protein-metabolites interaction, i.e. quantification of all steady state intracellular metabolite fluxes through the central metabolic reaction network (Wiechert et al., 2001). The final outcome of MFA is a metabolic flux map that comprises of catabolic and anabolic reaction fluxes for certain species under specific growth conditions. By comparing the fluxes map of different strains under different conditions, we can assess the metabolic impacts of genetic modification and environmental perturbations, thus provide insights and targets for further rational strain design. Reaction bottleneck can be identified with the objective to optimize growth and desired product titer. Rational genetic manipulation of cellular metabolism can be performed to maximize product biosynthesis (Tang et al., 2009). In addition to the definition of metabolic phenotype, MFA provides invaluable information of new pathway identification, branch point control identification, nonmeasured extracellular fluxes calculation and maximum yield calculation (Stephanopoulos et al., 1998).

Conventional metabolic flux analysis (c-MFA)

The conventional MFA is the most basic approach of flux analysis, which solely depends on stoichiometry of biochemical network and measured specific rates, i.e. substrate consumption rate and product secretion rates (Varma and Palsson, 1994). To solve the unknown pathway fluxes, c-MFA was performed by writing mass balances for the intracellular metabolites, which results in a linear equation cascade. The assumption of pseudo steady state must be applied, in which all the intracellular metabolite pools do not change over the experimental time span (Wiechert et al., 2001).

In general, the number of measurable extracellular metabolite fluxes is limited in practice, therefore it usually leads to underdetermined algebraic system. Additional constraints can be implemented by incorporating objective function or cofactor balance, i.e. ATP, NADH,

and NADPH (Stephanopoulos et al., 1998). Appropriate objective functions can be introduced to optimize flux determination. The central carbon metabolism, which consists of anabolic and catabolic functions, provides cofactors, amino acids and energy (ATP) for growth requirement. Flux balance analysis (FBA) has been used extensively for *in-silico* genome scale network modeling to predict steady-state fluxes with the objective function to maximize cell growth (Matsuoka and Shimizu, 2010). *E. coli* was found to utilize carbon source almost optimally to achieve maximum growth rate, whereby it matches with the in-silico FBA predictions (Edwards et al., 2001; Ibarra et al., 2002). However, incomplete knowledge about the pathway cofactor balances and inappropriate choice of objective function can lead to erroneous flux estimation (Marx et al., 1996; Sauer and Bailey, 1999; Schuetz et al., 2007).

Unfortunately, c-MFA fails in case of parallel metabolic pathway, metabolic cycles, bidirectional and reversible reactions (Figure 1) (Wiechert et al., 2001). In a large complex compartmented metabolic network in most eukaryotes, there is always a lack of extracellular metabolite measurements to relate to intracellular flux distribution (Schmidt et al., 1999). Consequently, the limitations can be overcome by supplementing c-MFA with isotopic tracer tracking, thereby provides a rigorous alternative to c-MFA.

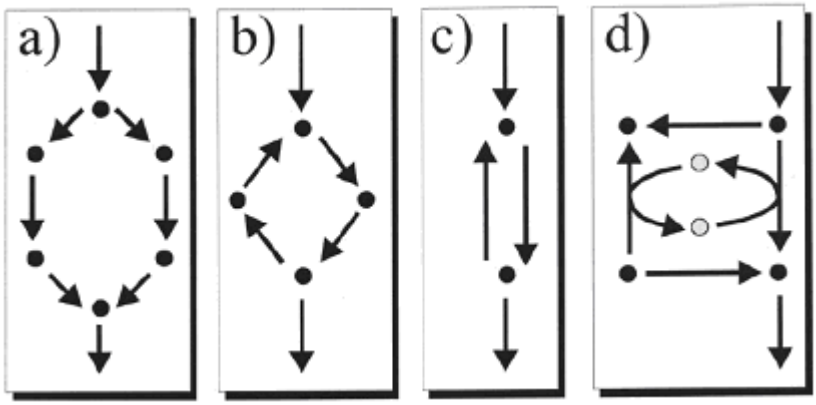


Figure 1. Typical situations in which conventional MFA fails (a) parallel pathways (b) metabolic cycles (c) bidirectional reactions (d) split pathways when cofactors are not balanced (Wiechert et al., 2001).

^{13}C metabolic flux analysis

^{13}C-MFA improves on the consistency and resolution of stoichiometric MFA by using experimental constraints derived from carbon labeling experiment (CLE) (Gomes and Simoes, 2012). CLE is based on feeding defined ^{13}C tracers into the biological system whereby all the carbon sources in the media are known. The most popular substrate is specific glucose with

unique labeling scheme (e.g. uniformly labeled glucose, 1-^{13}C labeled glucose). Labeled carbons will be distributed throughout the network, assimilated into metabolites and finally into proteins. Due to the scrambling and distinct cleavage in the carbon backbone of metabolites, CLE distinguishes the partitioning at certain metabolite nodes based on the positional labeling pattern of tracer atoms. Different labeling strategies are optimal for the resolution of relative pathway activities with different carbon-carbon bond modifications. For instance, Fischer et al. determined the combination of U-^{13}C glucose and 1-^{13}C glucose as the most optimal resolution for the entire metabolic central carbon metabolism in *E. coli* (Fischer et al., 2004). Exclusive use of 1-^{13}C glucose is particularly useful to resolve fluxes around oxidative pentose phosphate (PP) pathway and the Entner-Doudoroff (ED) pathway. Whereas, U-^{13}C glucose is optimal to resolve fluxes downstream of phosphoenopyruvate (PEP) and some exchange fluxes with C-C bond cleavage.

The greatest advantage of CLE is to avoid assumption of objective function and cofactor balance, since the acquired isotopomer data is adequate to determine the flux accurately (Gomes and Simoes, 2012). However, ^{13}C MFA is limited to elucidate flux distribution for small-scale reaction network (20-50 reactions) in central carbon metabolism, since carbon labeling data is not available for intermediate metabolites in the secondary reaction metabolism (Kohlstedt et al., 2010).

One challenge of CLE is to ensure the cell culture to reach isotopic and metabolic steady state in which all the intermediate concentration and fluxes are constant throughout the CLE as shown in Figure 2. Metabolic steady state is achieved at exponential growth phase in batch culture with constant growth rate. Continuous cultivation requires 3-5 volume changes to reach metabolic steady state (Zamboni et al., 2009). Nonetheless, isotopic steady state exists when the isotopomer labeling distributions remain unchanged over time. During CLE, the labeled tracer propagates through the network toward the metabolic products and biomass. However, due to the different turnover rate, the time necessary to reach isotopic steady state for each metabolite varies. In general, intracellular intermediate can reach isotopic steady state within seconds to minutes, whereas protein-bound amino acids require hours to reach isotopic steady state. Grotkjaer et al. studied the rate of ^{13}C substrate incorporated into biomass using a dynamic model detailing carbon transitions in the central carbon metabolism in *S. cerevisiae* (Grotkjær et al., 2004). The authors concluded the labeling of proteinogenic amino acids did not deviate after

3 residence times. To ensure ^{13}C isotopomer detection is at isotopic steady state, the cells are harvested after at least five generations at exponential growth phase in batch culture (Zamboni et al., 2009).

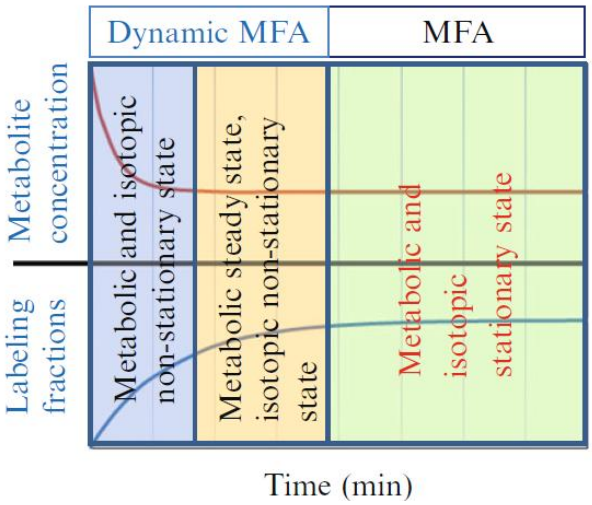


Figure 2. Metabolic and isotopic dynamics in cell metabolism (Antoniewicz et al., 2007a; Feng et al., 2012).

Isotopomer

The central concept of ^{13}C MFA lies on the isotopomer enrichment information in a given intracellular metabolite (Wiechert et al., 2001). Isotopomer represents a particular labeling state of a given metabolite. Each carbon in a given metabolite can be labeled or unlabeled; therefore there are 2^n isopotomers for n-carbon metabolite. For example, Figure 3 elucidates the 3-carbon metabolite has $2^3=8$ different isotopomers. The isotopomer distribution of metabolite is characterized by the isotopomer fraction (i.e. the percentage of each isotopomer in the metabolite), which sums up to 1. The metabolite fluxes at the branch points (pp pathway vs. glycolysis) and circular fluxes (TCA cycle) in the network can be determined from the isotopic labeling distribution of the metabolites.

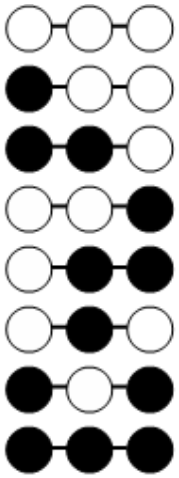


Figure 3. Eight isotopomer possibilities of three carbon metabolite. Filled and unfilled circles represent ^{13}C and ^{12}C atoms respectively.

The final isotopic enrichment in amino acids and intracellular metabolites can be assessed to reveal intracellular flux partitioning. Protein is a stable abundant source of labeling information, as the carbon backbones of central metabolism intermediate are conserved in protein (Zamboni et al., 2009). Only 16 out of the 20 amino acids can be accessed after protein hydrolysis. Besides the degradation of cysteine and tryptophan, glutamine and asparagines are converted to glutamate and aspartate respectively during acid hydrolysis (Tang et al., 2009). The amino acids provide isotopic labeling information for the precursor metabolites in the central carbon metabolism including pyruvate, phosphoenolpyruvate, acetyl-CoA, 3-phosphoglycerate, erythrose-4-phosphate, oxaloacetate, α-ketoglutrate, and ribose-5-phosphate (Maaheimo et al., 2001; Szyperski, 1995). Therefore, amino acids serve as major precursors to understand the central metabolic fluxes during steady-state growth as shown in Figure 4 (Fischer et al., 2004; Szyperski, 1995). In the contrary, free intracellular metabolites would provide richer information to resolve fluxes even beyond central metabolism and assess dynamic flux changes. Moreover, free intracellular metabolites reach isotopic steady state within seconds to minutes, much faster than proteinogenic amino acids. However, the shortfalls of metabolite-based MFA are their relatively low concentration, high turnover rate and diverse chemical nature (Sauer, 2006).

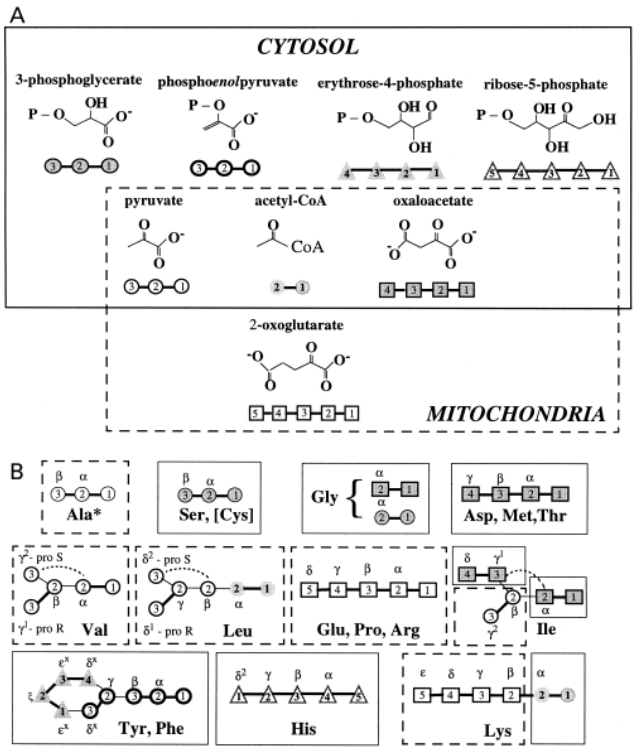


Figure 4. Carbon fragments originating from a single intermediate molecule present in proteinogenic amino acids (Maaheimo et al., 2001).

Nuclear magnetic resonance (NMR) and mass spectrocopy (MS) techniques

The most common techniques to detect isotopomer abundance are NMR and MS (Szyperski, 1995; Wittmann, 2002) to understand the intracellular labeling state of the system.

NMR: Proton NMR (^{1}H-NMR) was first extensively used in ^{13}C labeling experiment by Marx et al. to obtain labeling data from more than 25 NMR measurements. Each single protonated carbon atom position of certain metabolite pool can be distinguished. ^{1}H-NMR measures positional enrichment of each carbon atom position. However, isotopomer distribution can be resolved in more detail with ^{13}C-NMR, as a labeled carbon produces different splitting signals depending on the labeling state of the neighboring carbon. For instance, Figure 5 shows a 3-carbon molecule with different labeling states have different signal splitting patterns. If none of the neighboring carbon is labeled, a singlet peak emerges. With a neighboring carbon labeled, a doublet peak appears in which the splitting distance depends on the functional group on the neighboring carbon. If all three carbons are labeled, a double-doublet peak results, in which if the doublet is the same, a triplet peak emerges.

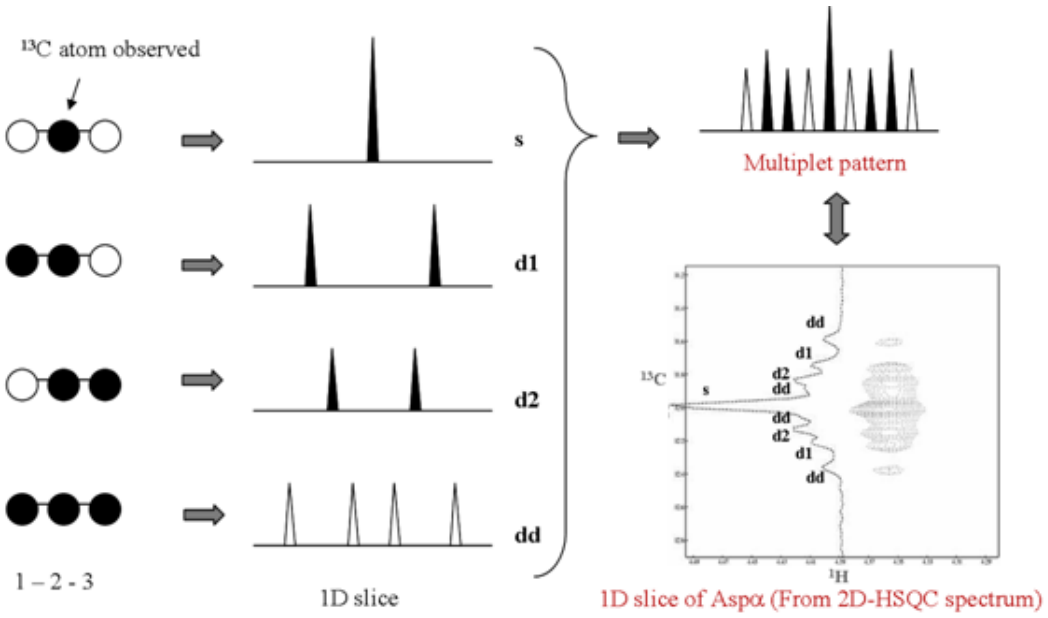


Figure 5. Relationship between isotopomer and multiplet pattern obtained from NMR measurement. Only 4 out of 8 isotopomers possible for three carbon metabolite are detected by NMR. If none of the neighboring carbon is labeled, a singlet peak emerges. With a neighboring carbon labeled, a doublet peak appears in which the splitting distance depends on the functional group on the neighboring carbon. If all three carbons are labeled, a double doublet results in which if the doublet is the same, a triplet peak emerges.

In 1D-NMR spectrum, overlapping peaks complicate the analysis. Therefore, separation is required to isolate metabolites from the hydrolysate for NMR quantification. A 2D-NMR with combination of 1H and 13C-NMR can resolve this problem, because metabolite peaks are separated in 2D-spectrum. Each spectrum can be evaluated without much interference (Sauer 2001). 2D-Correlation spectroscopy (COSY) and Heteronuclear Single Quantum Correlation (HSQC) spectroscopy are commonly applied. COSY detects the proton interaction on the carbon, while HSQC detects ^{13}C-^{13}C scalar coupling, i.e. the degree of coupling between adjacent carbon atoms. The indirect detection of carbon makes HSQC more sensitive than COSY. The HSQC technique therefore was used in this study to detect isotopic labeling of amino acids. NMR provides the most direct method to determine relative abundance of the isotopomers. Sriram et al. successfully used NMR technique to acquire amino acids isotopomer fraction, thereby generating flux map for soybean embryo and *Catharanthus roseus* (Sriram et al., 2006; Sriram et al., 2007).

MS: However, the overall sensitivity of NMR is significantly lower compared to MS technique. The former method requires at least 20 mg biomass, while the latter method requires

0.3-0.8 mg biomass. MS methods are extensively used in flux analysis (Wittmann, 2002). The MS instrument is coupled with either gas or liquid chromatography (GC or LC) to separate the compounds for ionization, fragmentation and finally analysis. The output of MS measurement is a total ion chromatogram shown in Figure 6. Assuming the compounds of interest are well separated, the mass spectra can be used for isotopic quantification. In reality, some amino acid fragments are rejected due to the choice of spectrum integration algorithm, concentration effect and overlapping fragments (Antoniewicz et al., 2007a; Feng et al., 2012).

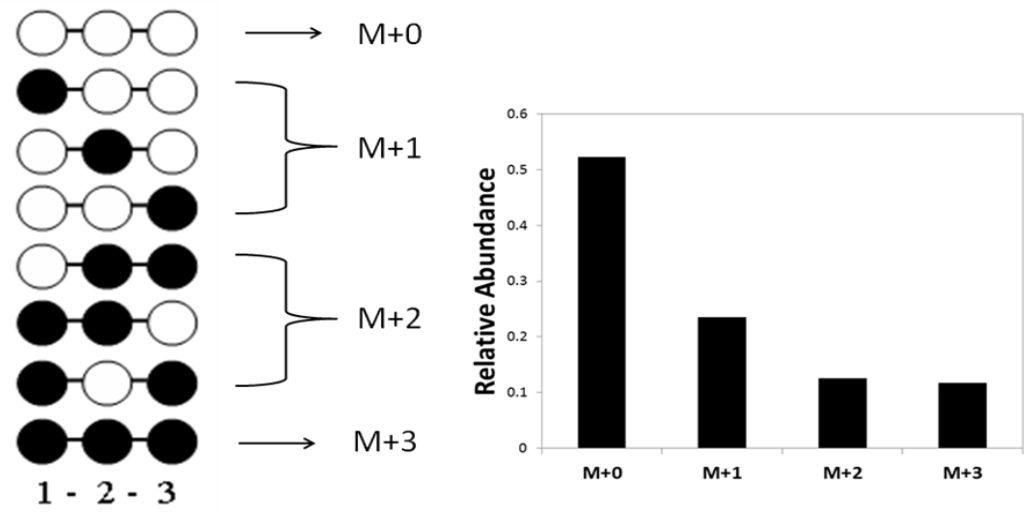


Figure 6. MS technique fails to distinguish among the exact isotopomers in M+1 and M+2.

For GC-MS measurement, the sample must be chemically derivatized to make it volatile to obtain chargable molecules. N-(tert-butyldimethylsilyl)-N-methyltrifluoroacetamide (TBDMS) is most commonly used to derivatize amino acids (Antoniewicz et al., 2007a). N,O-bis(trimethylsilyl)trifluoroacetaminde (BSTFA), a more gentle and sensitive derivatization agent, can be used to measure the extent of labeling in free organic acids, sugars and sugar phosphates (Ewald et al., 2009; Tang et al., 2009). In addition, the natural abundance of other atoms in analyte and derivatization residues needs to be corrected (i.e. ^{13}C, 1.13%; ^{18}O, 0.20%; ^{29}Si, 4.70% and ^{30}Si, 3.09%). Another pitfall of MS technique is the inability to distinguish exact labeled carbon position if one carbon of the 3-carbon molecule is labeled.

Free amino acids with fast turnover time open a new avenue for ^{13}C-MFA to investigate metabolism during various cell growth phases and also shorten the experimental duration by reaching isotopic steady state earlier. Iwatani et al. measured labeling information of intracellular free amino acids using LC MS/MS to profile flux change during fed-batch cultivation (Iwatani et

al., 2007). The study revealed the metabolic shift during different cell growth phases in a fed batch culture.

Besides intracellular amino acids, measurement of labeling pattern in central metabolites could extend the scope of ^{13}C-MFA and increase its accuracy. Kleijn et al. used LC-MS to analyze only intracellular metabolites in the glycolysis and PP pathway due to lack of understanding of metabolites localized in multiple compartments of yeast, therefore the flux model focused only on the glycolysis and the PP pathway (Kleijn et al., 2007). This approach was able to estimate the split ratio at the G6P node more accurately.

Overall, GC-MS, LC-MS and NMR had been combined for flux calculation by extracting the maximal isopotomer information from the intracellular metabolites and proteinogenic amino acids (Kleijn et al., 2007). The sensitivity of the fluxes around the metabolic nodes is highly dependent on the method of analysis. The combined information has been used to more accurately resolve flux distribution in *S. cerevisiae* with the compartmentation of cytosol and mitochondria.

Enumerating flux

The metabolite stoichiometric balances coupled with the carbon skeleton rearrangements are the fundamental principles to enumerate flux distribution. Extracellular measurements are required to account for complete carbon balance. Besides, carbon labeling data provides additional constraint to specify the split at network branch point. The basic workflow of ^{13}C-MFA comprising of experimental and computational steps is outlined in Figure 7.

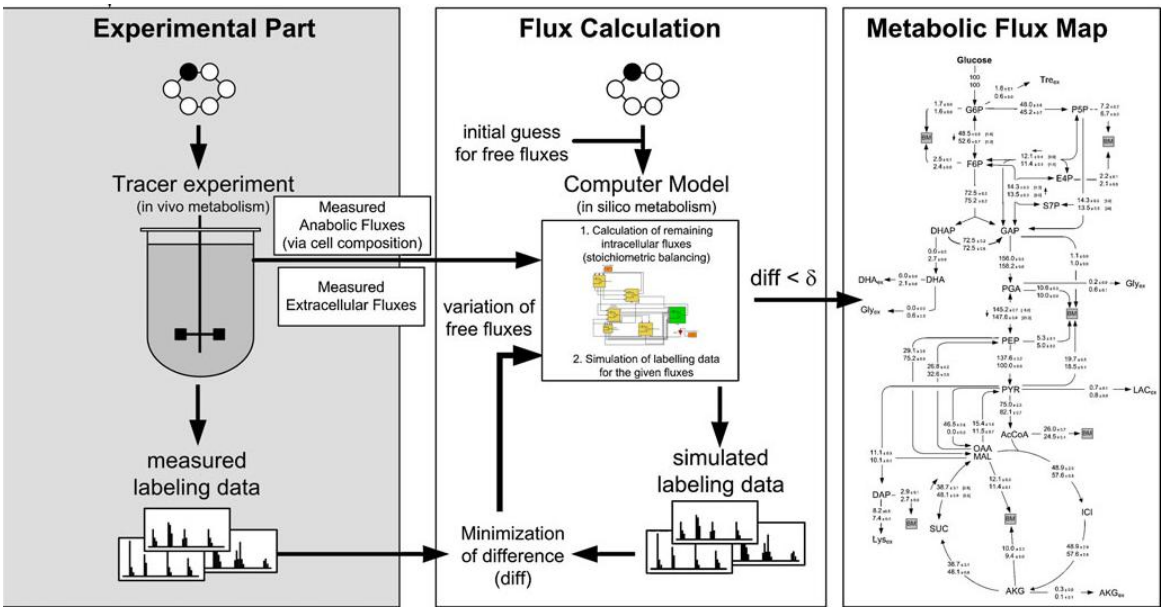


Figure 7. Schematic workflow for 13C-MFA comprising of experimental and computational routine (Kohlstedt et al., 2010).

Since the relationship between isotopomer fraction and intracellular flux distribution is non-linear, the computation procedure is complicated and analytical solution is practically impossible. Mathematics-involved numerical solutions were applied to solve isotopomer abundances for calculating intracellular fluxes (Schmidt et al., 1999; Wiechert et al., 2001; Zupke and Stephanopoulos, 1994) The development of computationally efficient algorithms for isotopomer balance such as cumomers (Wiechert et al., 2001) , elementary metabolite units (Antoniewicz et al., 2007b), and recently, Mathematical-involved (Srour et al., 2011) helps to reduce the computational expenses and increase the realiability of flux calculation. Generic softwares (such as 13C-Flux and NMR2Flux) using the concept of isotopomer balancing for flux analysis are also available (Sriram et al., 2006; Wiechert et al., 2001). METRAN and 13C-Flux2 software based on elementary metabolite units algorithm have also been developed for steady-state flux calculation (Antoniewicz et al., 2007b; Noack et al., 2011).

NMR2Flux program was developed in our group to evaluate a unique flux solution using global optimization routine (Sriram et al., 2006). An arbitrary set of fluxes is firstly chosen and verified for stoichiometric feasibility. The feasible fluxes are converted into isotopomer distribution through isotopomer and cumomer balances. Flux distributions are calculated by iteratively fitting simulated isotopomer distributions with the experimental measurements based on the chi-square error criteria. Simulated annealing algorithm is implemented to search for

global minimum without getting trapped in local minima. A statistical analysis is performed to account for isotopomer measurement errors and converts them to confidence intervals of the fluxes using Monte Carlo simulation.

Non-stationary flux analysis

The main prerequisite for ^{13}C-MFA is that the system must be in a metabolic and isotopic steady state, which constitutes a limitation to its application on non-growing cells without protein biosynthesis and higher cells with slow growth (Gomes and Simoes, 2012). The duration of CLE is significantly longer with the inverse of growth rate, thus leading to impractically long experiment. In addition, batch fermentation and fed-batch fermentation that are widely used in industry have the nature of non-stationary. Moreover, most of the target molecules are mostly produced at stationary phase, resulting in the need to investigate metabolic flux alternation with time series and metabolic phase shift. Therefore, in order to thwart these problems, the classical ^{13}C-MFA can be extended to isotopically non-stationary (INST) 13C-MFA (Noack et al., 2011; Noeh et al., 2007; Nöh et al., 2006). Instead of achieving isotopic equilibrium, non-stationary ^{13}C-MFA uses isotopic transient data to resolve fluxes within much shorter duration of CLE. Its application, however, raises new challenges in computations and analytics aspects. Noh et al. developed an extension of the EMU-based 13C-FLUX2 software to implement the simulation of non-stationary CLE, sensitivity analysis, data fitting, statistical identifiability analysis and optimal experimental design (Nöh et al., 2006).

In the range of seconds to minutes, ^{13}C tracers will be incorporated into the intracellular metabolites and free amino acids. With the fast turnover, ^{13}C incorporated metabolites would wash out the existing pool of metabolites, thereby dramatically shorten the experimental time. Thus, improvement in measuring the isotopic pattern for intracellular metabolites is necessary to keep track of changes in metabolite labeling and to complement the INST ^{13}C-MFA. CE-TOF-MS (Toya et al., 2007), LC-MS/MS (Iwatani et al., 2007; Noack et al., 2011), LC-MS (van Winden et al., 2005) and GC-MS (Ewald et al., 2009) have been used as robust tools to measure isotopomer information of the intracellular metabolites in the central carbon metabolisms instead of measuring the proteinogenic amino acids.

Reference

Antoniewicz, M. R., J. K. Kelleher, and G. Stephanopoulos, 2007a, Accurate Assessment of Amino Acid Mass Isotopomer Distributions for Metabolic Flux Analysis: Analytical Chemistry, v. 79, p. 7554-7559.

Antoniewicz, M. R., J. K. Kelleher, and G. Stephanopoulos, 2007b, Elementary metabolite units (EMU): A novel framework for modeling isotopic distributions: Metabolic Engineering, v. 9.

Edwards, J., R. Ibarra, and B. Palsson, 2001, In silico predictions of Escherichia coli metabolic capabilities are consistent with experimental data: Nat Biotech, v. 19, p. 125-130.

Ewald, J. C., S. Heux, and N. Zamboni, 2009, High-Throughput Quantitative Metabolomics: Workflow for Cultivation, Quenching, and Analysis of Yeast in a Multiwell Format: Analytical Chemistry, v. 81.

Feng, X., W.-Q. Zhuang, P. Colletti, and Y. J. Tang, 2012, Metabolic pathway determination and flux analysis in nonmodel microorganisms through 13C-isotope labeling: Methods in molecular biology (Clifton, N.J.), v. 881.

Fischer, E., N. Zamboni, and U. Sauer, 2004, High-throughput metabolic flux analysis based on gas chromatography–mass spectrometry derived 13C constraints: Analytical Biochemistry, v. 325, p. 308-316.

Gomes, L. C., and M. Simoes, 2012, C-13 Metabolic Flux Analysis: From the Principle to Recent Applications: Current Bioinformatics, v. 7, p. 77-86.

Grotkjær, T., M. Åkesson, B. Christensen, A. K. Gombert, and J. Nielsen, 2004, Impact of transamination reactions and protein turnover on labeling dynamics in 13C-labeling experiments: Biotechnology and Bioengineering, v. 86, p. 209-216.

Ibarra, R. U., J. S. Edwards, and B. O. Palsson, 2002, Escherichia coli K-12 undergoes adaptive evolution to achieve in silico predicted optimal growth: Nature, v. 420, p. 186-189.

Iwatani, S., S. Van Dien, K. Shimbo, K. Kubota, N. Kageyama, D. Iwahata, H. Miyano, K. Hirayama, Y. Usuda, K. Shimizu, and K. Matsui, 2007, Determination of metabolic flux changes during fed-batch cultivation from measurements of intracellular amino acids by LC-MS/MS: Journal of Biotechnology, v. 128.

Kleijn, R. J., J.-M. A. Geertman, B. K. Nfor, C. Ras, D. Schipper, J. T. Pronk, J. J. Heijnen, A. J. A. van Maris, and W. A. van Winden, 2007, Metabolic flux analysis of a glycerol-overproducing Saccharomyces cerevisiae strain based on GC-MS, LC-MS and NMR-derived C-13-labelling data: Fems Yeast Research, v. 7.

Kohlstedt, M., J. Becker, and C. Wittmann, 2010, Metabolic fluxes and beyond-systems biology understanding and engineering of microbial metabolism: Applied Microbiology and Biotechnology, v. 88.

Maaheimo, H., J. Fiaux, Z. P. Çakar, J. E. Bailey, U. Sauer, and T. Szyperski, 2001, Central carbon metabolism of Saccharomyces cerevisiae explored by biosynthetic fractional 13C labeling of common amino acids: European Journal of Biochemistry, v. 268, p. 2464-2479.

Marx, A., A. A. deGraaf, W. Wiechert, L. Eggeling, and H. Sahm, 1996, Determination of the fluxes in the central metabolism of Corynebacterium glutamicum by nuclear magnetic resonance spectroscopy combined with metabolite balancing: Biotechnology and Bioengineering, v. 49.

Matsuoka, Y., and K. Shimizu, 2010, Current status of C-13-metabolic flux analysis and future perspectives: Process Biochemistry, v. 45, p. 1873-1881.

Noack, S., K. Noeh, M. Moch, M. Oldiges, and W. Wiechert, 2011, Stationary versus non-stationary C-13-MFA: A comparison using a consistent dataset: Journal of Biotechnology, v. 154.

Noeh, K., K. Groenke, B. Luo, R. Takors, M. Oldiges, and W. Wiechert, 2007, Metabolic flux analysis at ultra short time scale: Isotopically non-stationary C-13 labeling experiments: Journal of Biotechnology, v. 129.

Nöh, K., A. Wahl, and W. Wiechert, 2006, Computational tools for isotopically instationary 13C labeling experiments under metabolic steady state conditions: Metabolic Engineering, v. 8, p. 554-577.

Sauer, U., 2006, Metabolic networks in motion: C-13-based flux analysis: Molecular Systems Biology, v. 2, p. 10.

Sauer, U., and J. E. Bailey, 1999, Estimation of P-to-O ratio in Bacillus subtilis and its influence on maximum riboflavin yield: Biotechnology and Bioengineering, v. 64.

Schmidt, K., L. C. Norregaard, B. Pedersen, A. Meissner, J. O. Duus, J. O. Nielsen, and J. Villadsen, 1999, Quantification of intracellular metabolic fluxes from fractional enrichment and 13C-13C coupling constraints on the isotopomer distribution in labeled biomass components: Metabolic engineering, v. 1.

Schuetz, R., L. Kuepfer, and U. Sauer, 2007, Systematic evaluation of objective functions for predicting intracellular fluxes in Escherichia coli: Mol Syst Biol, v. 3.

Sriram, G., D. B. Fulton, V. V. Iyer, J. M. Peterson, R. Zhou, M. E. Westgate, M. H. Spalding, and J. V. Shanks, 2006, Quantification of compartmented metabolic fluxes in developing soybean embryos by employing biosynthetically directed fractional C-13 labeling, two-dimensional C-13, H-1 nuclear magnetic resonance, and comprehensive isotopomer balancing (vol 136, pg 3043, 2004): Plant Physiology, v. 142.

Sriram, G., D. B. Fulton, and J. V. Shanks, 2007, Flux quantification in central carbon metabolism of Catharanthus roseus hairy roots by C-13 labeling and comprehensive bondomer balancing: Phytochemistry, v. 68.

Srour, O., J. D. Young, and Y. C. Eldar, 2011, Fluxomers: a new approach for C-13 metabolic flux analysis: Bmc Systems Biology, v. 5.

Stephanopoulos, G., A. A. Aristidou, J. H. Nielsen, and J. Nielsen, 1998, Metabolic Engineering: Principles and Methodologies, G. Stephanopoulos, Aristos A. Aristidou, Jens Høiriis Nielsen, Jens Nielsen.

Szyperski, T., 1995, Biosynthetically Directed Fractional 13C-labeling of Proteinogenic Amino Acids: European Journal of Biochemistry, v. 232, p. 433-448.

Tang, Y. J., H. G. Martin, S. Myers, S. Rodriguez, E. E. K. Baidoo, and J. D. Keasling, 2009, ADVANCES IN ANALYSIS OF MICROBIAL METABOLIC FLUXES VIA C-13 ISOTOPIC LABELING: Mass Spectrometry Reviews, v. 28.

Toya, Y., N. Ishii, T. Hirasawa, M. Naba, K. Hirai, K. Sugawara, S. Igarashi, K. Shimizu, M. Tomita, and T. Soga, 2007, Direct measurement of isotopomer of intracellular metabolites using capillary electrophoresis time-of-flight mass spectrometry for efficient metabolic flux analysis: Journal of Chromatography A, v. 1159, p. 134-141.

van Winden, W. A., J. C. van Dam, C. Ras, R. J. Kleijn, J. L. Vinke, W. M. van Gulik, and J. J. Heijnen, 2005, Metabolic-flux analysis of Saccharomyces cerevisiae CEN.PK113-7D based on mass isotopomer measurements of 13C-labeled primary metabolites: FEMS Yeast Research, v. 5, p. 559-568.

Varma, A., and B. O. Palsson, 1994, METABOLIC FLUX BALANCING - BASIC CONCEPTS, SCIENTIFIC AND PRACTICAL USE: Bio-Technology, v. 12.

Wiechert, W., M. Mollney, S. Petersen, and A. A. de Graaf, 2001, A universal framework for C-13 metabolic flux analysis: Metabolic Engineering, v. 3.

Wittmann, C., 2002, Metabolic flux analysis using mass spectrometry: Advances in Biochemical Engineering Biotechnology. Tools and applications of biochemical engineering science, v. 74.

Zamboni, N., S.-M. Fendt, M. Ruhl, and U. Sauer, 2009, 13C-based metabolic flux analysis: Nat. Protocols, v. 4, p. 878-892.

CHAPTER 3

SYSTEMS METABOLIC ENGINEERING DESIGN: FATTY ACID PRODUCTION AS EMERGING CASE STUDY

Manuscript submitted to Biotechnology and Bioengineering

Authors: Ting Wei Tee[1], Anupam Chowdhury[2], Costas D. Maranas[2] and Jacqueline V. Shanks[1]*

[1] Department of Chemical and Biological Engineering, Iowa State University, Ames, IA

[2] Department of Chemical Engineering, Pennsyvania State University, University Park, PA

Abstract

Increasing demand for petroleum has stimulated industry to develop sustainable production of chemicals and biofuels using microbial cell factories. Fatty acids of chain lengths from C6 to C16 are propitious intermediates for the catalytic synthesis of industrial chemicals and diesel-like biofuels. The abundance of genetic information available for *Escherichia coli* and specifically, fatty acid metabolism in *E. coli*, supports this bacterium as a promising host for engineering a biocatalyst for the microbial production of fatty acids. Recent successes rooted in different features of systems metabolic engineering in the strain design of high-yielding medium chain fatty acid producing *E. coli* strains provide an emerging case study of design methods for effective strain design. Classical metabolic engineering and synthetic biology approaches enabled different and distinct design paths towards a high-yielding strain. Here we highlight a rational strain design process in systems biology, an integrated computational and experimental approach for carboxylic acid production, as an alternative method. Additional challenges inherent in achieving an optimal strain for commercialization of medium chain-length fatty acids will likely require a collection of strategies from systems metabolic engineering. Not only will the continued advancement in systems metabolic engineering result in these highly productive strains more quickly, this knowledge will extend more rapidly the carboxylic acid platform to the microbial production of carboxylic acids with alternate chain-lengths and functionalities.

Introduction

Concerns regarding crude oil depletion and climate change have encouraged the development of renewable biochemicals and biofuels using carbohydrates as the feedstock (Demirbas 2009; Gabrielle 2008). Microbial biosynthesis of fatty acids (FAs) for biorenewable chemicals and biofuels has recently garnered extensive attention. Free FAs can be used as precursors for the production of alkanes by catalytic decarboxylation or transesterification (Lennen et al. 2010; Lu et al. 2008; Steen et al. 2010). Alternatively, FAs can be converted biologically to FA ethyl esters, which have high energy density and low water solubility (Steen et al. 2010). Medium chain FAs can be effectively used for industrial applications such as detergents, soaps, lubricants, cosmetics, and pharmaceuticals. FAs can also be catalytically deoxygenated via metal catalysts to produce α-olefins, the building blocks of polymerization.

The genetically suitable *Escherichia coli* is an excellent host for FA production, given its fully sequenced genome and well-studied FA metabolism. The first step in type II fatty acid biosynthesis (FAB) pathway in *E. coli* involves the ATP-dependent acetyl-CoA carboxylase (encoded by *accABCD*) where acetyl-CoA is converted into malonyl-CoA. Malonyl-CoA is further converted into malonyl-ACP by the enzyme malonyl-CoA ACP transacetylase (encoded by the *fabD*), which condenses with acetyl-CoA to synthesize a 4-carbon fatty acid acyl carrier protein (i.e. butyryl ACP). Subsequently, the cyclic chain elongation steps recruit two carbons in the backbone of the growing fatty acid ACP every turn of the cycle. Finally, the fatty acyl ACP is hydrolyzed into FA by thioesterases in a single step enzymatic conversion. Despite the intrinsic capability of synthesizing FAs for lipid and cell membrane biosynthesis, *E. coli* does not normally accumulate free FAs as intermediates. FA metabolism is tightly regulated at transcriptional and post-transcriptional levels by both the transcription factor and product inhibition, meaning that FA overproduction may require significant re-engineering of cellular metabolism. An excellent overview of FA biosynthesis and its regulation has been reviewed by Handke et al. (Handke et al. 2011).

The challenge then, is not only to create a microbial biocatalyst that can produce FAs at high yields, high rates and high product titers, but also to shorten the development time in the metabolic engineering design cycle, in order to compete effectively with petroleum-based processes. The metabolic engineering design process has evolved into a Systems Metabolic

Engineering design process, as shown in Figure 1. Systems Metabolic Engineering, which encompasses systems biology, synthetic biology, and evolutionary engineering at the system level, provides powerful techniques to design new biocatalysts (Lee et al. 2011a). The classical metabolic engineering procedures of constructing and screening strains, based on the collective wisdom of experience, are often complemented with one or more of the new tools to improve and/or fine-tune strain design. The design engineer is faced with a suite of choices in the design process, on whether to use methods in isolation or in combination, although a survey of the literature indicates that at combination of multiple approaches is still not very common to date (Lee et al, 2011). Fortunately, a plethora of engineering manipulations for free FA production in *E. coli* exist and have been reviewed in recent years (Huffer et al. 2012; Lennen and Pfleger 2012; Liu and Khosla 2010; Zhang et al. 2011a). Moreover, recent successes in construction of high-yielding medium chain fatty acid producing *E. coli* strains, rooted in different features of systems metabolic engineering, provide an emerging case study of design methods for effective strain design (Dellomonaco et al. 2011; San and Li 2012; Zhang et al. 2012b).

In this review, we focus mainly on medium-chain FA production in *E. coli* using different systems metabolic engineering approaches outside the scope of traditional metabolic engineering. In particular, we describe a classical metabolic engineering technique, an integrated experimental and computational strategy, and a synthetic engineering effort for enhancing fatty acid production in *E. coli*.

Classical metabolic engineering

Classical metabolic engineering involves an iterative process of synthesis and analysis, where increasingly refined strains are designed and constructed based on the past knowledge. Based on literature evidences and intuitive guesses, several strategies have been employed to improve FA production, as have been elucidated in Figure 2a. The adopted strategies include up-regulating the availability of precursors malonyl-CoA (Lee et al. 2011b; Lennen et al. 2010; Lu et al. 2008) and malonyl-ACP (Lee et al. 2011b; Zhang et al. 2012c) and elimination of the β-oxidation pathway genes *fadD* or *fadE* (Lennen et al. 2010; Lu et al. 2008; Steen et al. 2010) to prevent degradation of FAs. Overexpression of the chain-elongation genes *fabA*, *fabZ* and *fabG* encoding for the FAB pathway have also been performed (Yu et al. 2011). In addition, overexpression of native *E. coli* thioesterases *tesA* and *tesB* (Lu et al. 2008; Steen et al. 2010), as

well as heterologous plant thioesterases from *C. camphorum* (Liu et al. 2010) and *U. californica* (Lennen et al. 2010) has been identified to overproduce FAs with tailored carbon chain length. Optimal expression of plant thioesterases in *E. coli* guided by predictions of the ribosomal binding sites (Zhang et al. 2011b) as well as discoveries of new thioesterases, such as a recently identified *E. coli* thioesterase gene, *fadM*, involved in the β-oxidation pathway (Dellomonaco et al. 2011), was shown to improve medium-chain FA production. Removal of a competitive pathway towards acetate, however, did not increase the flux towards middle chain FA (Li et al. 2012). The synergy of the aforementioned positive interventions is often used to significantly boost FA production. For example, Steen et al. (2010) reported 1.2g/L FAs (14% of theoretical yield) by deletion of *fadD* and *fadE* β-oxidation gene with overexpression of cytosolic *tesA* thioesterase.

In an elegant example of a system-wide metabolic engineering approach, the existing biological system was redesigned by an engineered reversal of the β-oxidation pathway in *E. coli*, leading to a significant increase in the production yield of carboxylic acids (Dellomonaco et al. 2011). The cellular system was reprogrammed by the manipulation of global regulators. As such, mutations in *FadR* and *AtoC* regulon were introduced to express β-oxidation pathway enzymes in the absence of FAs. The native *crp* gene was replaced by a cAMP-independent mutant to alleviate the catabolite repression in the presence of glucose. *ArcA* gene was deleted to relieve *ArcA*-mediated repression induced by oxygen availability. In combination with the elimination of the native fermentation and the FA degradation pathway, and the overexpression of the selected terminal pathway, extracellular C_{10}-C_{18} FAs were produced at titer of ~7g/L in the bioreactor, with mineral salts medium with yield of 0.28g/g glucose (~80% theoretical yield). Thus, redesigning the native FA biosynthesis using a CoA-based functional reversal of β-oxidation provided an efficient platform for the production of FAs.

A classical "push and pull" concept was applied to enhance acetyl-CoA availability, minimize acetyl-CoA drains, eliminate competing pathways and overexpress product formation pathways, ultimately led to a strain with nearly ~100% maximum theoretical yield for medium chain-length FA production (San and Li 2012). Overexpression of *fabZ* encoding β-hydroxyacyl-ACP dehydratase increases FA titer and yield by pulling carbon flux toward FA elongation cycle (Figure 2b). Naturally occurring FA-sensing transcription factors coordinate and regulate the

synthesis and degradation of FA at transcription level. Whereby, *FabR* antagonizes FA synthesis by repressing *fabB* and *fabA* FAB genes, and vice versa for the *FadR* transcription factor. Indirect up-regulation of FA elongation reactions, by deletion of *fabR* and over expression of *fadR*, showed an increase in FA titer and yield. In addition to the terminal FAB pathway, limited focus has also been given on the central carbon metabolism manipulations for augmenting FA production, however, with less success. San et al. (2012) showed that redirection of TCA cycle flux (deletion of *sucC*, *fumAC* and *gltA*) towards fatty acid production improved middle chain FA production. Furthermore, gene interruption in the glycolytic pathway (*glk*, *ptsG*, *pfkA* and *pykF*) also was shown to be strategic genetic manipulation. Overall, the combination of the *fabZ* over-expression and *sucC* deletion in a *fadD* knockout strain boosted the production to 5.7g/L C_{14}-C_{16} FAs with yield of 0.38g/g glucose in rich media (~100% theoretical yield). The technology has been translated into industrial collaboration to produce synthetic diesel and lubricants from biomass (Williams 2013).

Integrated computational and experimental approach

Even though metabolic engineering has taken long strides in manipulating the metabolic network towards the overproduction of a desired chemical, the process is hampered by bottlenecks of time and accuracy. Recent advances made in genome sequencing have accelerated the construction of genome-scale metabolic networks, which in turn have led to the growth of several rationale-based strain optimization protocols (Burgard et al. 2003; Kim et al. 2011; Maia et al. 2012; Pharkya et al. 2004; Yang et al. 2011). Computational strain design protocols consider the complex interconnectivity of cellular metabolism including cofactor balances to identify key metabolic bottlenecks towards the production of a chemical, and predict (often non-intuitive) strategies to overcome them. Integrated with classical metabolic engineering techniques, these procedures have been successfully employed for the overproduction of several chemicals (Alper et al. 2005; Asadollahi et al. 2009; Bro et al. 2006; Park et al. 2007; Xu et al. 2011). We recently demonstrated the integrated approach of computationally-driven predictions and metabolic flux analysis techniques for the overproduction of FAs with different chain lengths (Ranganathan et al. 2012). The OptForce computational protocol (Ranganathan et al. 2010) was used for arriving at suggestions for strain redesign by identifying the minimal set of reactions that need to be actively manipulated to guarantee an imposed production yield. OptForce makes use of *in vivo* flux measurements to characterize the reference strain and then

solves a “worst-case” optimization problem to conservatively identify an exhaustive list of alternate intervention strategies required to meet a pre-specified yield of the desired chemical. It also provides a natural prioritization of results where the most important manipulations are identified first. We observed that the intervention strategies were mostly chain specific that optimized the utilization of the precursors, cofactors and energy equivalents required for the FA synthesis of a particular chain length. For palmitate production, the up-regulation of FA elongation cycle was suggested to pull acetyl-CoA towards FA synthesis, followed by a removal of the β-oxidation pathway to prevent FA degradation. In addition, OptForce identifies several non-intuitive manipulations distal to the terminal FAB pathway that channels metabolic flux towards palmitate (see Figure 3a). In particular, it suggests re-routing glycolytic flux towards ED pathway for the dual objectives of generating reduction cofactor NADPH required in the FA chain elongation and arresting cell growth by reduced production of ATP. In addition, OptForce identified down-regulation of TCA cycle and acetate production pathway as chronologically less prioritized interventions to prevent the drainage of acetyl-CoA away from FA synthesis. In accordance with OptForce prioritized suggestions, a strain with the over-expression of *fabZ* and acyl-ACP thioesterase (from *R. communis*), combined with the deletion of *fadD*, achieved 1.7 g/L and 0.14 g FA/g glucose of $C_{14\text{-}16}$ FA (~38% theoretical yield) in minimal medium. Interestingly, the prediction of FA biosynthesis up-regulation, TCA cycle interruption, and glycolysis interruption agreed with San et al. (2012), strengthening the robustness of the integrated approach of computational strain design and flux analysis tools. In addition, the intervention template suggested by OptForce for the overproduction of FA of individual chain lengths (see Figure 3b) can be used along with chain-specific thioesterases (Jing et al. 2011) to study the relatively unexplored area of short-chain FA production.

Synthetic biology

Transcriptional and post-transcriptional control in *E. coli* tightly regulates FA biosynthesis. Computational tools often do not provide any inference on gene regulation; however, synthetic biology plays a crucial role in modeling, understanding, and fine-tuning the core components in metabolic pathways. Engineering core pieces of metabolic pathways helps meet specified performance criteria, such as gaining desired phenotypes, once they are integrated into larger biological systems.

The synthetic biology approach fine-tunes the enzymatic pathways of a specific product, enabling the transfer of optimized systems to another chassis. Clomburg et al. (2012) used a bottom-up strategy to reconstruct a functional reversal of the β-oxidation cycle for production of carboxylic acids through the assembly of well-defined and self-contained enzymes composing the pathway (Fig. 4a). Functional reversal of the β-oxidation cycle comprises of thiolase (*AtoB*, *FadA*), 3-hydroxyacyl-CoA dehydrogenase (*FabB*), enoyl-CoA hydratase (*FabB*), and acyl-CoA dehydrogenase (*FadE*, *YdiO*, *egTER*). Each CoA intermediate in the cycle could be converted into carboxylic acids with thioesterase termination pathways. After in-vitro kinetic characterization, *AtoB*, *FabB* and *egTER* were assembled in-vivo in *E. coli* along with the native thioesterase termination pathway, resulting in 3.43g/L butyrate with 0.35g/g glycerol yield (~74% theoretical yield). In-vitro kinetic analysis revealed the capability of *FadA* thiolase on longer chain acyl-CoA. For the synthesis of longer chain carboxylic acids, functional reversal of the β-oxidation cycle could be operated multiple cycles through the integration of *AtoB*, *FadBA* and *egTER* into the host strain. The success in resembling self-contained enzyme units in the functional reversal of β-oxidation provides a paradigm for the efficient production of carboxylic acids using synthetic biology techniques.

Despite the advent in the genetic engineering, metabolic imbalance with low expression pathway genes becomes the bottleneck in biosynthetic pathways. Extremely high levels of gene expression divert cellular resources to unnecessary cell maintenance, instead of devoting the resource to produce the desired chemical. The dynamic sensor-regulator system (DSRS) was developed to dynamically control the synthesis of FAs and the derived biodiesels in *E. coli* (Zhang et al. 2012a). A FA/ acyl-CoA sensor was engineered based on the *FadR* protein and its associated regulator. Synthetic FA/ acyl-CoA-regulated promoters were designed to increase the limited dynamic ranges of native *FadR*-regulated promoters. The engineered biosensors responded primarily to acyl-CoA, which served as an indirect FA sensor (Fig.4b). With the insertion of this biosensor, the FA-producing *E. coli* strain with *tesA* expression and *fadE* deletion produced 3.8g/L FA (~56% theoretical yield). Furthermore, the biosensor concept was extended to the over-expression of *FadR* in the *E. coli* strain with *tesA* expression and *fadE* knockout, enhancing FA titer to 5.2g/L (73% theoretical yield) in minimal medium (Zhang et al. 2012b). *FadR* over-expression optimally tuned the expression levels of FA pathway genes for the

production of FAs. Thereby, the over-expression of an isolated gene in the FA synthesis pathway (*fabA*, *fabB* and *fabF)* did not increase FA titer as much as the *FadR* over-expression.

Synthetic biology enables the systematic investigation of pathway limitations and removes the metabolic bottlenecks that are tightly regulated. The customized expressions of enzymatic reactions could enhance carbon flux toward precursor and the corresponding product. The accumulation or depletion of intermediates could be avoided to prevent loss in cell viability and pathway productivity. Most recently, Koffas and coworkers applied modular synthetic biology strategy to optimize the transcription of fatty acid metabolic pathway, which consists of the modules of upstream acetyl-CoA formation, intermediary acetyl-CoA activation and fatty acid synthase (Xu et al. 2013). Modular pathway optimization by altering plasmid copy number led to a balance in the supply and consumption of fatty acid intermediates (acetyl-CoA and malonyl-ACP). Moreover, translation efficiency could be improved by customizing the ribosomal binding sites of fatty acid pathway modules, thus enhancing fatty acid production. The combination of these synthetic biology tools yielded 8.6g/L fatty acids (~22% theoretical yield) in fed-batch fermentation.

Synthetic biology can also be applied to identify and understand the controlling factors in directing carbon flux to the FA pathway (Liu and Khosla 2010). A cell-free system was developed to interrogate the regulation and synthesis of FAs in *E. coli* through manipulation of the substrate, cofactors, allosteric regulators and enzyme level. The study revealed high dependency of FA synthesis on the intracellular concentration of malonyl-CoA. Malonyl-CoA concentration was required to be increased by ten-folds of its reference value under FA overproduction conditions. The rate of FA synthesis was generally linearly correlated to ACC levels with respect to the selection of target ACC. In a subsequent *in-vitro* FA biosynthesis reconstitution study, *fabI* and *fabZ* were determined to enhance FA synthesis in a hyperbolic fashion (Yu et al. 2011). Nonetheless, *fabF* and *fabH* inhibited FA synthesis at concentration higher than 1μM. Thus, the optimization of FA biosynthesis genes expression is critical to further improve the strain for FA production.

Conclusion and future challenges

Classical metabolic engineering, integrated computational/experimental approach, and synthetic biology have contributed towards the improved production of FAs in *E. coli*, and could

be extended to the development of cell factories for specific chemical production. To further dissect the regulations in FA metabolism, system metabolic engineering can be employed to pinpoint beneficial key components in the complicated genetic circuit for strain optimization. Enzymatic bottlenecks could be accurately identified with the development of detailed kinetic models that include metabolic regulatory networks constrained by system biology findings. He *et al.* applied a combination of system biology approaches (i.e. fluxomics and transcriptomics) to gain metabolic insights into cellular metabolism under fatty acid production. It was found the reducing equivalent NADPH and ATP as the potential bottleneck for fatty acid production, guiding the direction for future strain development and process optimization to enhance fatty acid production (He et al. 2013). From the industrial standpoint, fermentation using minimal medium and efficient product separation process can lower operating costs and be competitive for the production of petroleum-based chemicals. Recently, medium optimization study showed phosphate limitation in continuous fermentation increased fatty acid yield and biomass-specific productivity compared to carbon-limited cultivation (Youngquist et al. 2013). It has also been noted that endogenous FA production reduced cell viability due to the loss of inner membrane integrity (Lennen et al. 2011). Secretion of endogenous FAs could possibly assuage the toxicity effect while reducing product extraction cost. Further investigation is warranted to address the challenges for promising commercialization.

Acknowledgement

This material is based upon work supported by the National Science Foundation under Award No. EEC-0813570.

References

Alper H, Jin YS, Moxley JF, Stephanopoulos G. 2005. Identifying gene targets for the metabolic engineering of lycopene biosynthesis in Escherichia coli. Metab Eng 7(3):155-64.

Asadollahi MA, Maury J, Patil KR, Schalk M, Clark A, Nielsen J. 2009. Enhancing sesquiterpene production in Saccharomyces cerevisiae through in silico driven metabolic engineering. Metab Eng 11(6):328-34.

Bro C, Regenberg B, Forster J, Nielsen J. 2006. In silico aided metabolic engineering of Saccharomyces cerevisiae for improved bioethanol production. Metab Eng 8(2):102-11.

Burgard AP, Pharkya P, Maranas CD. 2003. Optknock: a bilevel programming framework for identifying gene knockout strategies for microbial strain optimization. Biotechnol Bioeng 84(6):647-57.

Dellomonaco C, Clomburg JM, Miller EN, Gonzalez R. 2011. Engineered reversal of the beta-oxidation cycle for the synthesis of fuels and chemicals. Nature 476(7360).

Demirbas A. 2009. Political, economic and environmental impacts of biofuels: A review. Applied Energy 86:S108-S117.

Gabrielle B. 2008. [Significance and limitations of first generation biofuels]. J Soc Biol 202(3):161-5.

Handke P, Lynch SA, Gill RT. 2011. Application and engineering of fatty acid biosynthesis in Escherichia coli for advanced fuels and chemicals. Metabolic Engineering 13(1).

He L, Xiao Y, Gebreselassie N, Zhang F, Antoniewicz MR, Tang YJ, Peng L. 2013. Central metabolic responses to the overproduction of fatty acids in Escherichia coli based on 13C-metabolic flux analysis. Biotechnology and Bioengineering:n/a-n/a.

Huffer S, Roche CM, Blanch HW, Clark DS. 2012. Escherichia coli for biofuel production: bridging the gap from promise to practice. Trends in biotechnology 30(10).

Jing F, Cantu DC, Tvaruzkova J, Chipman JP, Nikolau BJ, Yandeau-Nelson MD, Reilly PJ. 2011. Phylogenetic and experimental characterization of an acyl-ACP thioesterase family reveals significant diversity in enzymatic specificity and activity. Bmc Biochemistry 12.

Kim J, Reed JL, Maravelias CT. 2011. Large-scale bi-level strain design approaches and mixed-integer programming solution techniques. PLoS One 6(9):e24162.

Lee JW, Kim TY, Jang Y-S, Choi S, Lee SY. 2011a. Systems metabolic engineering for chemicals and materials. Trends in Biotechnology 29(8).

Lee S, Jeon E, Yun HS, Lee J. 2011b. Improvement of Fatty Acid Biosynthesis by Engineered Recombinant Escherichia coli. Biotechnology and Bioprocess Engineering 16(4).

Lennen RM, Braden DJ, West RM, Dumesic JA, Pfleger BF. 2010. A Process for Microbial Hydrocarbon Synthesis: Overproduction of Fatty Acids in Escherichia coli and Catalytic Conversion to Alkanes. Biotechnology and Bioengineering 106(2).

Lennen RM, Kruziki MA, Kumar K, Zinkel RA, Burnum KE, Lipton MS, Hoover SW, Ranatunga DR, Wittkopp TM, Marner WD, II and others. 2011. Membrane Stresses Induced by Overproduction of Free Fatty Acids in Escherichia coli. Applied and Environmental Microbiology 77(22).

Lennen RM, Pfleger BF. 2012. Engineering Escherichia coli to synthesize free fatty acids. Trends in biotechnology 30(12):659-667.

Li M, Zhang X, Agrawal A, San K-Y. 2012. Effect of acetate formation pathway and long chain fatty acid CoA-ligase on the free fatty acid production in E. coli expressing acy-ACP thioesterase from Ricinus communis. Metabolic Engineering 14(4):380-387.

Liu T, Khosla C. 2010. Genetic Engineering of Escherichia coli for Biofuel Production. Annual Review of Genetics, Vol 44 44.

Liu T, Vora H, Khosla C. 2010. Quantitative analysis and engineering of fatty acid biosynthesis in E. coli. Metab Eng 12(4):378-86.

Lu X, Vora H, Khosla C. 2008. Overproduction of free fatty acids in E. coli: Implications for biodiesel production. Metabolic Engineering 10(6).

Maia P, Vilaca P, Rocha I, Pont M, Tomb JF, Rocha M. 2012. An integrated computational environment for elementary modes analysis of biochemical networks. Int J Data Min Bioinform 6(4):382-95.

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. Proc Natl Acad Sci U S A 104(19):7797-802.

Pharkya P, Burgard AP, Maranas CD. 2004. OptStrain: a computational framework for redesign of microbial production systems. Genome Res 14(11):2367-76.

Ranganathan S, Suthers PF, Maranas CD. 2010. OptForce: an optimization procedure for identifying all genetic manipulations leading to targeted overproductions. PLoS Comput Biol 6(4):e1000744.

Ranganathan S, Tee TW, Chowdhury A, Zomorrodi AR, Yoon JM, Fu Y, Shanks JV, Maranas CD. 2012. An integrated computational and experimental study for overproducing fatty acids in Escherichia coli. Metabolic Engineering 14(6):687-704.

San K-Y, Li M; 2012. Methods to produce fatty acids from renewable carbon sources. USA.

San K-Y, Li M, Zhang X; 2011. BACTERIA AND METHOD FOR SYNTHESIZING FATTY ACIDS USA.

Steen EJ, Kang Y, Bokinsky G, Hu Z, Schirmer A, McClure A, del Cardayre SB, Keasling JD. 2010. Microbial production of fatty-acid-derived fuels and chemicals from plant biomass. Nature 463(7280).

Williams M. 2013. Modified bacteria turn waste into fat for fuel.

Xu P, Gu Q, Wang W, Wong L, Bower AGW, Collins CH, Koffas MAG. 2013. Modular optimization of multi-gene pathways for fatty acids production in E. coli. Nat Commun 4:1409.

Xu P, Ranganathan S, Fowler ZL, Maranas CD, Koffas MA. 2011. Genome-scale metabolic network modeling results in minimal interventions that cooperatively force carbon flux towards malonyl-CoA. Metab Eng 13(5):578-87.

Yang L, Cluett WR, Mahadevan R. 2011. EMILiO: a fast algorithm for genome-scale strain design. Metab Eng 13(3):272-81.

Youngquist JT, Rose JP, Pfleger BF. 2013. Free fatty acid production in Escherichia coli under phosphate-limited conditions. Applied Microbiology and Biotechnology 97(11):5149-5159.

Yu X, Liu T, Zhu F, Khosla C. 2011. In vitro reconstitution and steady-state analysis of the fatty acid synthase from Escherichia coli. Proceedings of the National Academy of Sciences 108(46):18643-18648.

Zhang F, Carothers JM, Keasling JD. 2012a. Design of a dynamic sensor-regulator system for production of chemicals and fuels derived from fatty acids. Nature Biotechnology 30(4).

Zhang F, Ouellet M, Batth TS, Adams PD, Petzold CJ, Mukhopadhyay A, Keasling JD. 2012b. Enhancing fatty acid production by the expression of the regulatory transcription factor FadR. Metabolic Engineering 14(6):653-660.

Zhang F, Rodriguez S, Keasling JD. 2011a. Metabolic engineering of microbial pathways for advanced biofuels production. Current Opinion in Biotechnology 22(6).

Zhang X, Agrawal A, San K-Y. 2012c. Improving fatty acid production in escherichia coli through the overexpression of malonyl coA-Acyl carrier protein transacylase. Biotechnology Progress 28(1).

Zhang X, Li M, Agrawal A, San K-Y. 2011b. Efficient free fatty acid production in Escherichia coli using plant acyl-ACP thioesterases. Metabolic Engineering 13(6).

List of figures

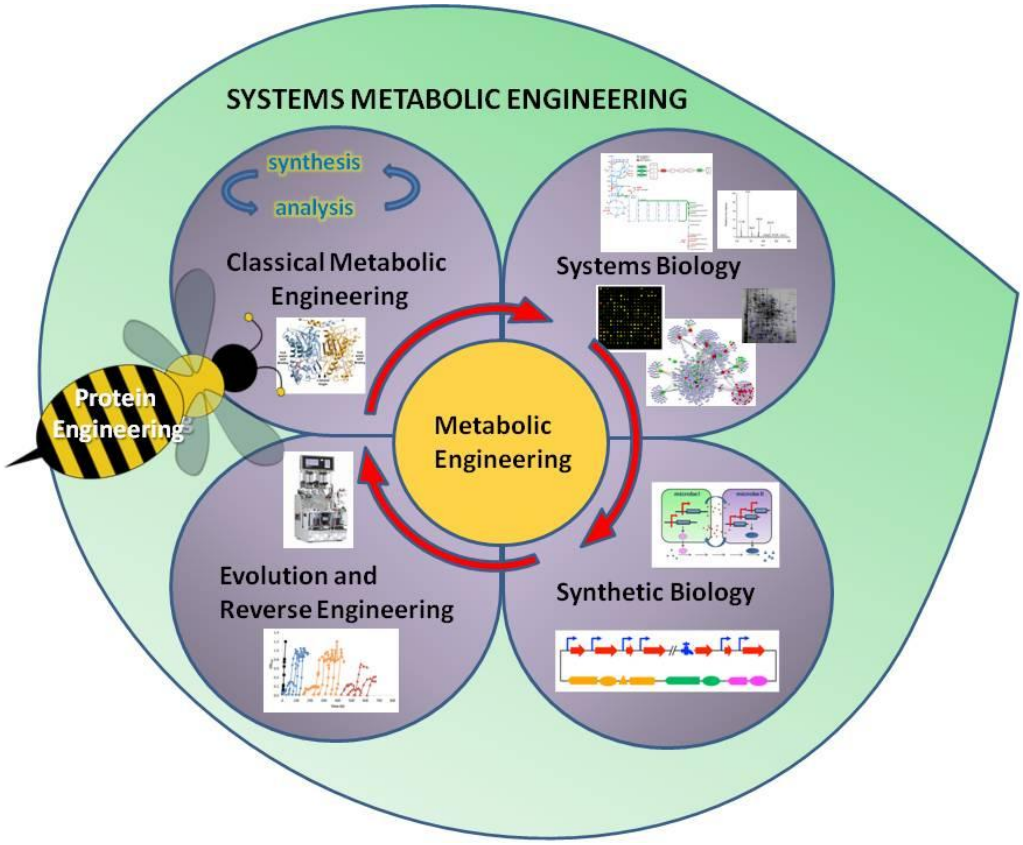

SYSTEMS METABOLIC ENGINEERING
synthesis
analysis
Classical Metabolic Engineering
Systems Biology
Protein Engineering
Metabolic Engineering
Evolution and Reverse Engineering
Synthetic Biology

Figure 1

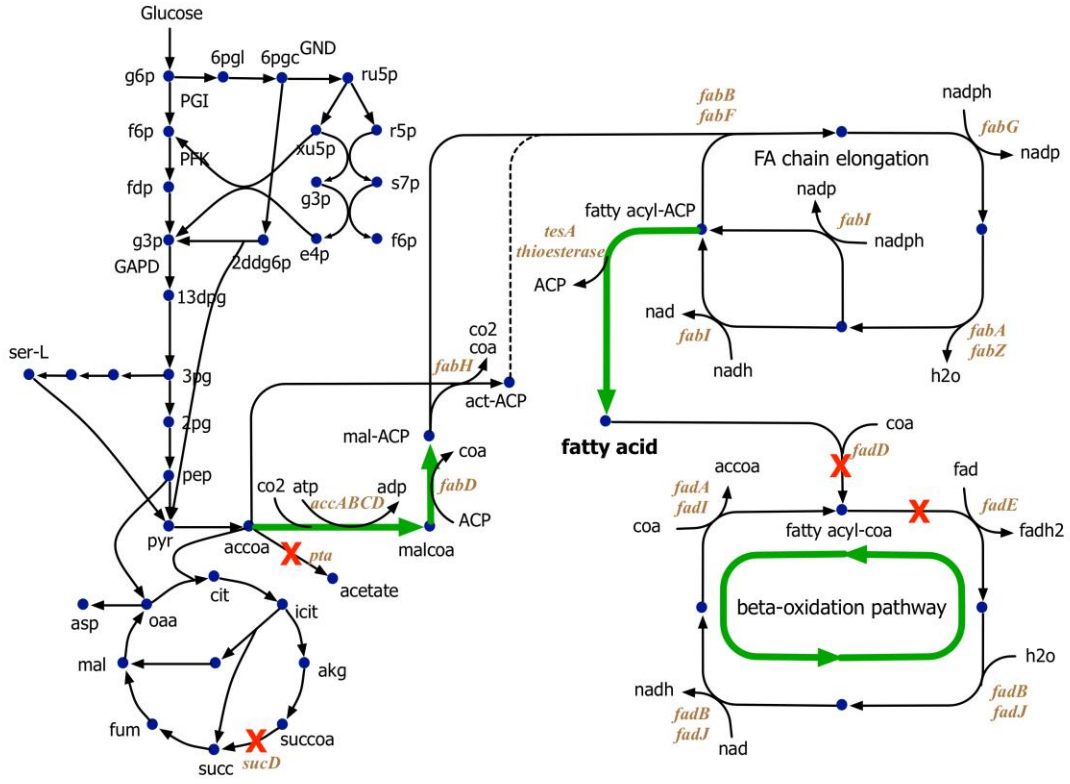


Figure 2a

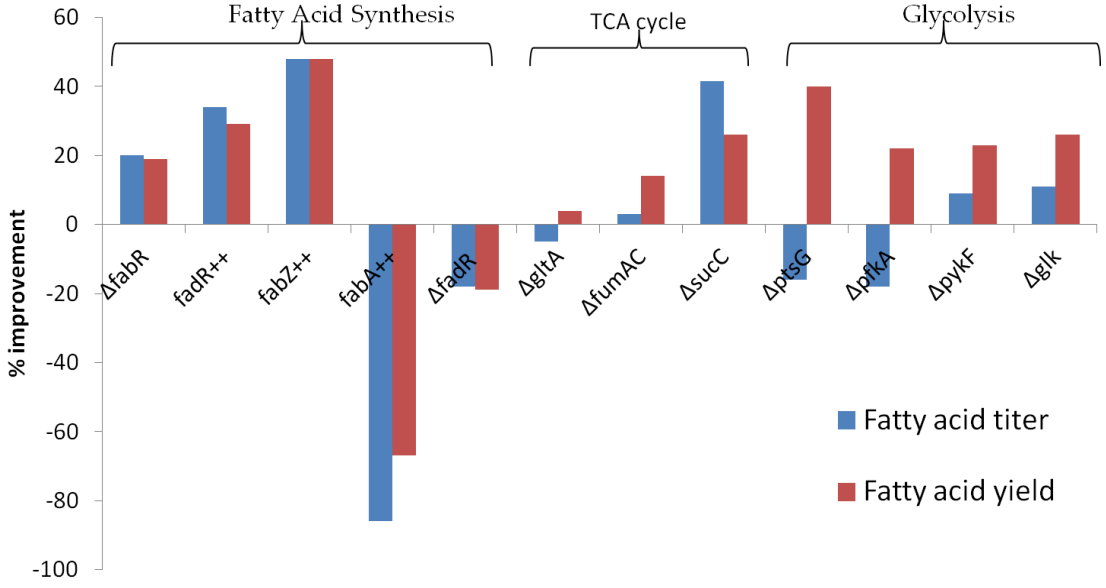


Figure 2b

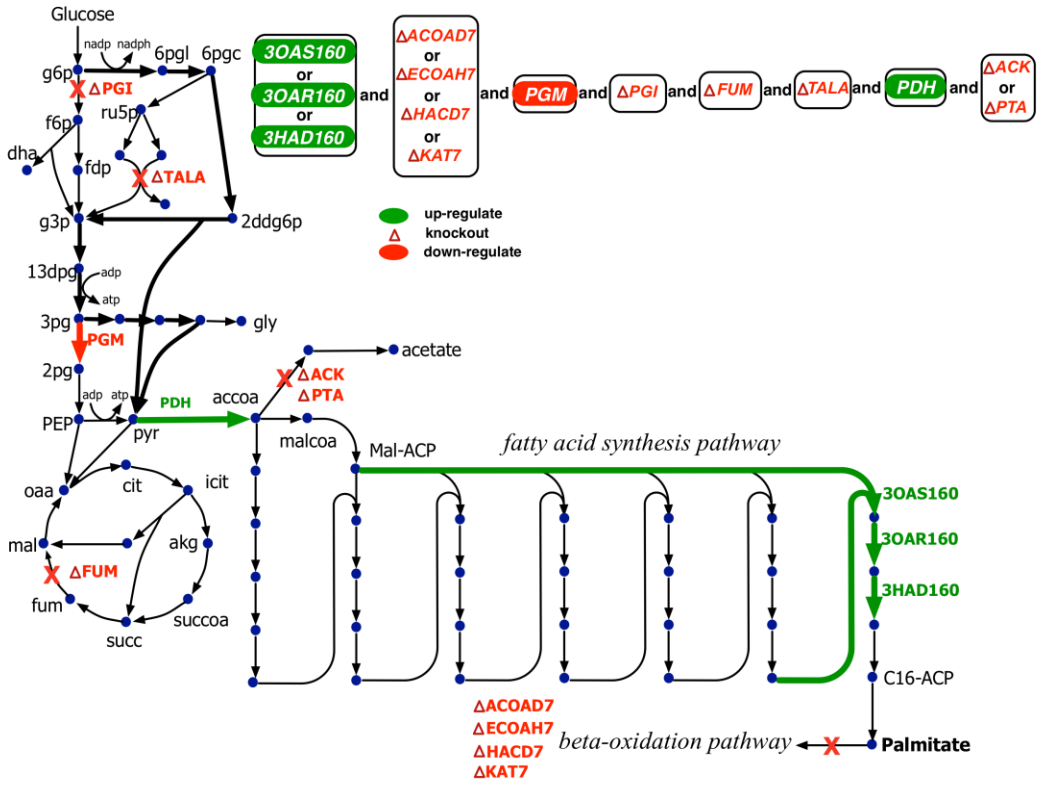

Glucose
nadp
nadph
6pgl
6pgc
g6p
ΔPGI
ru5p
f6p
dha
fdp
ΔTALA
g3p
2ddg6p
13dpg
adp
atp
3pg
gly
PGM
2pg
adp
atp
PEP
PDH
pyr
accoa
ΔACK
ΔPTA
acetate
malcoa
Mal-ACP
fatty acid synthesis pathway
oaa
cit
icit
mal
akg
ΔFUM
fum
succoa
succ
3OAS160
3OAR160
3HAD160
C16-ACP
ΔACOAD7
ΔECOAH7
ΔHACD7
ΔKAT7
beta-oxidation pathway
Palmitate
3OAS160
or
3OAR160
or
3HAD160
and
ΔACOAD7
or
ΔECOAH7
or
ΔHACD7
or
ΔKAT7
and
PGM
and
ΔPGI
and
ΔFUM
and
ΔTALA
and
PDH
and
ΔACK
or
ΔPTA
up-regulate
knockout
down-regulate

Figure 3a

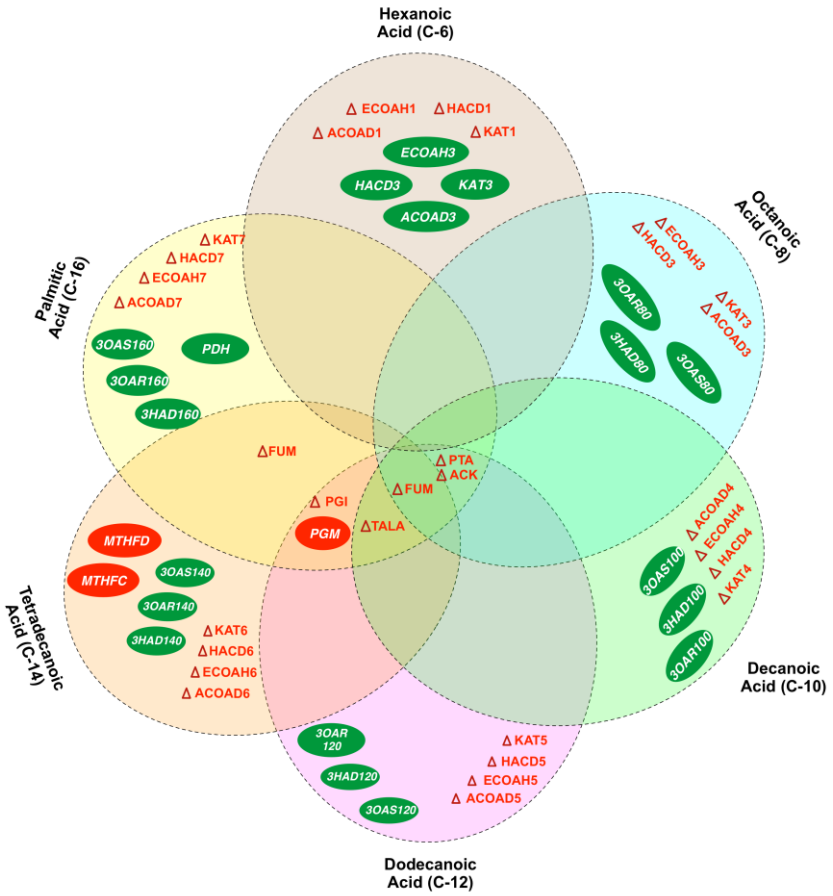


Figure 3b

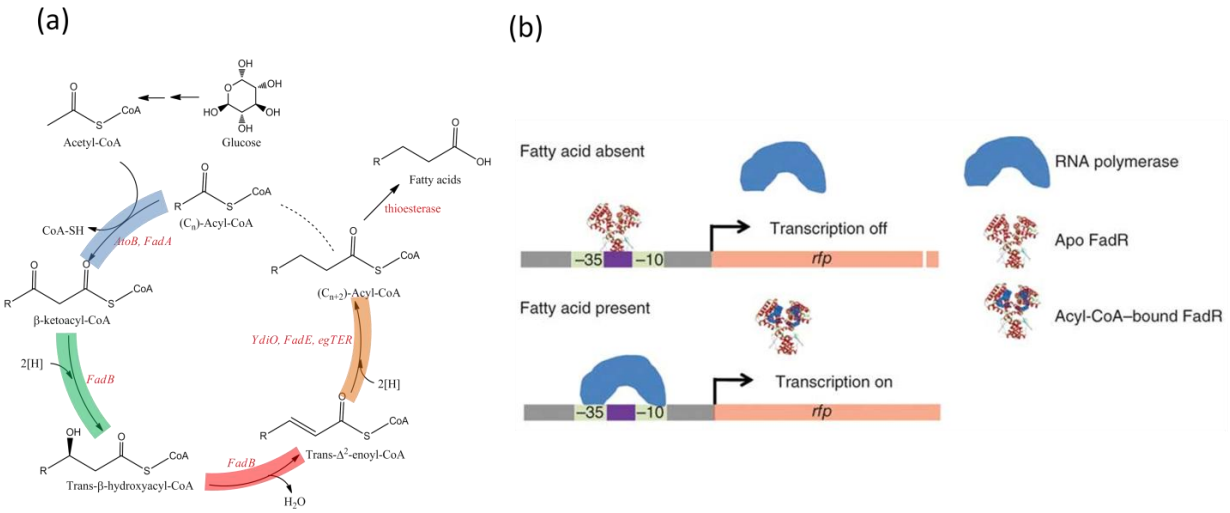


Figure 4

CHAPTER 4

AN INTEGRATED COMPUTATIONAL AND EXPERIMENTAL STUDY FOR OVERPRODUCING FATTY ACIDS IN *ESCHEICHIA COLI*

A paper published in Metabolic Engineering

Authors: Sridhar Ranganathan [1a], Ting Wei Tee [2a], Anupam Chowdhury[3b], Ali R. Zomorrodi[3b], Jong Moon Yoon[2], Yanfen Fu[2], Jacqueline V. Shanks[2], and Costas D. Maranas[3*]

[1] Huck Institutes of Life Sciences, Pennsylvania State University, University Park, PA
[2] Department of Chemical and Biological Engineering, Iowa State University, Ames, IA
[3] Department of Chemical Engineering, Pennsylvania State University, University Park, PA

a. Joint first authors, [b]Joint second authors

Abstract

Increasing demands for petroleum have stimulated sustainable ways to produce chemicals and biofuels. Specifically, fatty acids of varying chain lengths (C_6 – C_{16}) naturally synthesized in many organisms are promising starting points for the catalytic production of industrial chemicals and diesel-like biofuels. However, bio-production of fatty acids from plants and other microbial production hosts relies heavily on manipulating tightly regulated fatty acid biosynthetic pathways. In addition, precursors for fatty acids are used along other central metabolic pathways for the production of amino acids and biomass, which further complicates the engineering of microbial hosts for higher yields. Here, we demonstrate an iterative metabolic engineering effort that integrates computationally driven predictions and metabolic flux analysis techniques to meet this challenge. The OptForce procedure was used for suggesting and prioritizing genetic manipulations that overproduce fatty acids of different chain lengths from C_6 to C_{16} starting with wild-type *E. coli*. We identified some common but mostly chain-specific genetic interventions alluding to the possibility of fine-tuning overproduction for specific fatty acid chain lengths. In accordance with the OptForce prioritization of interventions, *fabZ* and acyl-ACP thioesterase were upregulated and *fadD* was deleted to arrive at a strain that produces 1.70 g/L and 0.14 g fatty acid/g glucose (~ 39% maximum theoretical yield) of $C_{14\text{-}16}$ fatty acid in minimal M9

medium. These results highlight the benefit of using computational strain design and flux analysis tools in the design of recombinant strains of *E. coli* to produce free fatty acids.

Keywords: Fatty acids overproduction, Computational strain design, Metabolic flux analysis

1. Background and introduction

The economical production of industrial chemicals (Nikolau *et al*., 2008) and transportation fuels (Stephanopoulos, 2007) from renewable resources is advancing but remains a grand challenge. In particular, microbial synthesis of free fatty acids for the production of biorenewable chemicals and fuels has garnered extensive interest recently ((Nikolau *et al*., 2008);(Steen *et al*., 2010); (Handke *et al*., 2011); (Liu et al., 2010b)). First generation biofuels, such as bio-ethanol (Fortman *et al*., 2008) produced from corn has relatively low energy density and water miscibility. New efforts are focused on longer chain alcohols such as 1-butanol (Gulevich *et al*., 2011; Lan and Liao, 2011; Shen and Liao, 2008), isobutanol (Atsumi *et al*., 2010) and 1,3-butanediol (Gonzalez *et al*., 2010) as gasoline bio-alternatives and fatty acids as promising intermediates for diesel bio-alternatives (Lu *et al*., 2008). Fatty acids produced during fermentation can be converted to alkanes by catalytic esterification or decarboxylation (Fjerbaek *et al*., 2009; Vasudevan and Briggs, 2008). Conversely, the host organism could be bioengineered to convert fatty acids towards fatty acid ethyl esters (FAEE) (Steen *et al*., 2010) which have high energy density and low water solubility (Atsumi *et al*., 2010). Medium chain fatty acids (C_6-C_{14}) find attractive industrial applications as sources for detergents, lubricants, cosmetics, and pharmaceuticals. Free fatty acids can be directly hydrogenated to form fatty alcohols (Voeste and Buchold, 1984). More recently, it has been shown that fatty acids could be catalytically deoxygenated via Pd or Rh catalysts (George Kraus, unpublished results) to produce α-olefins, which serve as building blocks of important polymerization products. In addition, existence of plant thioesterases that can specifically hydrolyze acyl-ACP substrates of a particular chain length (Jing *et al*., 2011) creates the opportunity to produce novel fatty acids.

Most bacteria are naturally equipped to produce fatty acids that form an important constituent of their cell envelopes (Magnuson *et al*., 1993), however, transcriptional and post-transcriptional control in *Escherichia coli* tightly regulates the metabolism of fatty acid biosynthesis. The two most important metabolic steps of fatty acid synthesis are the conversion of acetyl-CoA into malonyl-CoA by an ATP-dependent acetyl-CoA carboxylase (ACCOAC) and

the conversion of malonyl-CoA into fatty acid product bound to an acyl carrier protein (ACP) by a multi-subunit synthase (see Figure 1). Fatty acids produced from acetyl-CoA and malonyl-ACP are primarily used for phospholipid biosynthesis and synthesis of lipid A and coenzyme A (CoA) intermediates (Dellomonaco *et al*., 2011; Gulevich *et al*., 2011; Poirier *et al*., 2006) in *E. coli*. Overproduction of fatty acids in *E. coli* is however a difficult challenge because of tight transcriptional and post-transcriptional regulation of fatty acid biosynthesis (Fujita *et al*., 2007; Magnuson *et al*., 1993; White *et al*., 2005) including strong product inhibition (Gonzalez *et al*., 2010). For example, *fabH* and *fabI*, which encode the acyl-ACP synthesis and reduction of enoyl-ACP reaction steps (i.e., ketoacyl-ACP synthase (KAS15) and enoyl-ACP reductase), respectively, are also inhibited by long-chain fatty acyl-ACPs (Heath and Rock, 1996a; Heath and Rock, 1996b).

In spite of the difficulties, recent efforts have led to improvements in the synthesis of free fatty acids in *E. coli*; however, most of the reported engineering strategies rely on manipulating terminal pathways near the target fatty acid. This trend may be due to the complexity of metabolism and the difficulty of predicting the effect of manipulations at a systemic level. Genetic interventions and strategies reported to date include blocking fatty acid degradation through the β-oxidation pathway by knocking out *fadD* or *fadE* genes (Lu *et al*., 2008), heterologous expression of thioesterase genes from *U. californica* (Lu *et al*., 2008) or *C. camphorum* (Lu *et al*., 2008) to target fatty acids of specific chain lengths (Dehesh et al., 1996; Liu et al., 2010a; Nawabi et al., 2011), and augmenting the availability of precursors by overexpressing acetyl-CoA carboxylase (Lennen *et al*., 2010). Most of these approaches have resulted in *E. coli* strains that show selectivity towards the production of $C_{14\text{-}16}$ fatty acids. Alternatively, Dellomonaco *et al* (2011) recently reported on an engineered reversal of the β-oxidation pathway in *E. coli* leading to a significant increase in the production yield of 1-butanol as well as a number of long chain fatty acids. We note that there are a number of differences in biosynthesis of fatty acids using the chain elongation (type II fatty acid synthesis) pathway and reversed β-oxidation pathway. In particular, the type II fatty acid synthesis pathway involves addition of two carbons from malonyl-ACP to the acyl-ACP skeleton in each cycle, along with the release of one molecule of CO_2. In contrast, in the reversed β-oxidation pathway one mole of acetyl-CoA is recruited in each cycle to increase the fatty-acid chain length by two carbon units. In addition, the reduction steps in chain-elongation pathway require NADPH as the cofactor

while those for the reversed β-oxidation pathway utilize NADH and $FADH_2$. Finally, one ATP is needed to form each mole of malonyl-CoA in the type-II synthesis pathway whereas no ATP is involved in the reversed β-oxidation pathway. Nevertheless, the reported yields for middle chain-length fatty acids using these two pathways are quite similar. Despite recent improvements, the production yield of fatty acids remains far below that of alcohols (Magnuson *et al*., 1993). For example, the maximum yield of fatty acid production by the type-II synthesis pathway was recently reported to be 14% of the maximum theoretical yield (Steen *et al*., 2010), while that for alcohols is close to 70% (Magnuson *et al*., 1993).

In the past decade, a number of optimization procedures have been proposed to identify targets for gene knockouts (e.g. OptKnock (Burgard *et al*., 2003), RobustKnock (Tepper and Shlomi, 2010), OptORF (Kim and Reed, 2010), OptGene (Patil *et al*., 2005)), up-/down-regulations (e.g. OptReg (Pharkya and Maranas, 2006)) and knock-ins (i.e. OptStrain (Pharkya *et al*., 2004)) that lead to overproduction of specific biochemicals in microorganisms. In spite of their success stories, none of these procedures proactively make use of flux data to drive the strain design process. Metabolic flux analysis (MFA) has been increasingly been used to quantify internal metabolic fluxes (Wiechert et al. 2001) in strain engineering projects. Metabolic fluxes provide a unique description of cellular physiology and a starting point for pinpointing genetic manipulations (Stephanopoulos 1999, Peebles et al. 2010, Koffas and Stephanopoulos 2005). The OptForce procedure (Ranganathan, Suthers and Maranas 2010) was designed to make use of flux measurements available for the wild-type strain. OptForce first estimates the maximum range of flux variability for all the reactions in the metabolic network of the wild-type strain and overproducing network. By overlaying the two sets, a set of reactions is revealed whose flux must depart away from the wild-type range if the imposed overproduction target is to be met (i.e., MUST sets). This set of required changes in the network can be described with a single logic statement containing AND and OR operators linking reaction up/down manipulations. OptForce subsequently chooses from this subset of reactions a list of targets that must be actively engineered to guarantee a target yield for the desired product. The genetic interventions identified by OptForce are prioritized based on their impact on product yield improvement and alternate intervention scenarios can be explored. The OptForce paradigm was recently demonstrated by constructing a strain of *E. coli* that produces increased levels of intracellular malonyl-CoA (Xu et al. 2011), an important precursor for fatty acids.

In this study, we employed OptForce procedure to identify the most promising engineering interventions that lead to the overproduction of fatty acids C_6 through C_{16} in *E. coli*. We next report on the improved production yield of $C_{14\text{-}16}$ fatty acids following the implementation of some prioritized OptForce interventions (i.e., upregulation of *fabZ* and acyl-ACP thioesterase and *fadD* knockout). The computational results and experimental measurements presented in this paper demonstrate that *E. coli* metabolism can be reprogrammed for specific fatty acid chain lengths using an integrated computations and experimentation paradigm.

2. Materials and methods

2.1. Strains and plasmids:

All strains and plasmids used in this work are listed in Table 1 and were gifted by Ka-Yiu San's lab at Rice University.

2.2. Metabolic flux analysis experiments

Strains and culture conditions:

An *E. coli* MG1655 strain from glycerol frozen stock was streaked on Luria Broth (LB) plate and incubated overnight at 37^{o}C. A single colony from the plate was grown in 25 mL MOPS minimal medium (8.37g/L MOPS powder (Sigma-Alrich, Saint Louis, MO), 0.72g Tricine (Sigma Alrich, Saint Louis, MO), 2.92g/L NaCl, 0.51g/L NH_4Cl, 1.6g/L KOH, 0.215mg/L Na_2SeO_3, 0.303mg/L Na_2MoO_4, 0.17mg/L $ZnCl_2$, 50.3mg/L $MgCl_2$, 48.1mg/L K_2SO_4, 0.348g/L K_2HPO_4, and micronutrients containing 2.5mg/L $FeCl_2.4H_2O$, 92 µg/L $CaCl_2.2H_2O$, 31 µg/L H_3BO_3, 20 µg/L $MnCl_2.4H_2O$, 9 µg/L $CoCl_2.6H_2O$, 2 µg/L $CuCl_2.2H_2O$, and 48.4 µM HCl) supplemented with 1% glucose in an orbital shaker at 250 rpm until exponential phase. The culture then was centrifuged at 4000 rpm for 5 min at 4 ℃. The supernatant were discarded, and the pellet was re-suspended in fresh MOPS medium. The appropriate quantity of the washed cell suspension was used to inoculate 400 mL of MOPS medium in the 500 mL bioreactor (INFORS HT, Switzerland) to a starting OD_{550} of 0.03. For better identification of fluxes, a mixture of uniformly labeled [U-^{13}C], first carbon labeled [1-^{13}C] and natural glucose was used for ^{13}C flux analysis (Fischer *et al*., 2004). Specifically, 10% U-^{13}C glucose, 25% 1-^{13}C glucose, and 65% of naturally labeled glucose was used as the tracer to final medium

concentration of 1% glucose. About 500 µL of antifoam (Antifoam B Silicone Emulsion, J.T. Baker) aquatic solution (volume ratio antifoam: water = 1: 1) was added into the media to prevent foaming. The aerobic fermentation was conducted at 37 ℃, with a gas flow rate at 0.6 ml/min and agitation speed of 600 rpm. The pH was controlled at 7.0±0.05 by adding 1M potassium hydroxide. The dissolved oxygen level was maintained above 50% of saturated levels to ensure aerobic conditions. The cells were harvested at mid-exponential phase after at least 5 generations to ensure metabolic and isotopic steady state.

Due to reports of strains carrying the plant acyl-ACP thioesterase being unstable (Zhang *et al*., 2011), an additional metabolic flux experiment was conducted for *E. coli* but at a reduced temperature and bioreactor agitation. Batch fermentations were performed for ML103 strain (MG1655 *ΔfadD*) using minimal M9 (0.8 g/L NH_4Cl, 0.5 g/L NaCl, 7.52 g/L Na_2HPO_4, 3.0g/L KH_2PO_4, 0.24g/L $MgSO_4$, 11.1mg/L $CaCl_2$, 1ng/L thiamine HCl, and trace elements containing 166.7µg/L $FeCl_3.6H_2O$, 1.8µg/L $ZnSO_4.7H_2O$, 1.2µg/L $CuCl_2.2H_2O$, 1.2µg/L $MnSO_4.2H_2O$, 1.8µg/L $CoCl_2.6H_2O$, and 0.223mg/L $Na_2EDTA.2H_20$) medium supplemented with 1% glucose (20% U-^{13}C glucose and 80% 1-^{13}C glucose), 100 mg/L ampicillin, and antifoam. The fermentor was controlled at pH 7.0, 30 ^{o}C and 300 rpm agitation. As in the previous experiment, dissolved oxygen level was maintained above 50% of saturated levels and cells were harvested at mid-exponential phase after at least 5 generations to ensure isotopic steady state.

Analytical techniques:

Cell biomass dry weight was determined by measuring optical density OD_{550} using a spectrophotometer (Genesys 20, Madison, WI). Cell dry weight was estimated by the correlation: 1 OD_{550} =0.36 g cell dry weight/L (Choudhary *et al*., 2011). Biomass composition was determined based on literature data (John L. Ingraham, 1983). Media samples were taken during the exponential growth and filtered through 0.22 µm pore sized nylon filters (P.J. Cobert Associates, Saint Louis, MO) and kept at -20 ^{o}C for extracellular metabolite analysis. Glucose and acetate were measured using a Waters HPLC (Waters, Milford, MA) with 410 refractive index detector. The Aminex column (HPX-87H, Bio-Rad, Hercules, CA) was used at 30 ^{o}C with 0.3 mL/min of 5mM sulfuric acid as mobile phase.

Physiological parameters determination:

The substrate uptake rate and production secretion rate in batch culture are constant during exponential phase. The substrate uptake rate and product secretion rate are defined as the

coefficient of substrate/product concentration versus biomass divided by the growth rate. Acetate is the only product detected under aerobic batch cultivation.

Sample preparation for 2-Dimensional NMR analysis:

Cells were prepared as described previously (Choudhary *et al*., 2011). Briefly, cells are centrifuged, washed twice with saline water containing 0.9% NaCl, then hydrolysed with 6 N hydrochloric acid at 110^{o}C for 18-24 hours. Acids were evaporated, the residue reconstituted in nanopure water and filtered, then lyophilized. Finally, the sample was dissolved in deuterium oxide for NMR analysis.

NMR measurement:

2D [^{13}C, ^{1}H] Heteronuclear Single Quantum Correlation (HSQC) spectra were acquired on a Bruker Avance DRX 500 MHz spectrometer at 298 K and processed as described previously (Choudhary *et al*., 2011; Sriram *et al*., 2004). Nonoverlapping multiplets on the spectrum were quantified using NMRView (Johnson and Blevins, 1994). Overlapping multiplets (α amino acids) were analyzed using peak deconvolution software (Choudhary *et al*., 2011). The amino acids isotopomer abundances measured by 2D HSQC NMR are related to the precursor metabolites by using amino acids biosynthesis pathways as described by Szyperski (Szyperski, 1995). The resulting NMR intensities were used to calculate the isotopomer fractions as shown in Supplementary Table S4.

Metabolic network model for MFA:

A network model for *E. coli* metabolism was constructed based on existing literature, Ecocyc database and microarray data (see Table S1). The model includes glucose transport and phosphorylation pathway, Embden-Meyerhof-Parnas pathway, oxidative pentose phosphate branch, non-oxidative pentose phosphate branch, TCA cycle, anaplerotic pathways, metabolite exchange reactions, ED pathways, all amino acids biosynthesis pathways, and several amino acids transamination reactions (Fischer and Sauer, 2003a; Fischer *et al*., 2004; Sauer *et al*., 1999; Siddiquee *et al*., 2004; Toya *et al*., 2010).

Flux evaluation methodology:

Fluxes were quantified using NMR2Flux software developed by Sriram *et al* (2004). NMR2Flux employs isotopomer balancing and a global optimization routine to find stoichiometrically feasible fluxes set consistent with experimental measurements. Overall fluxes were estimated by minimizing the chi square error between experimentally measured and

simulated isotopomer fractions of amino acid. Errors in evaluated fluxes were estimated from errors in the extracellular fluxes, biomass growth rate, biomass synthesis fluxes, and isotopomer abundances by performing a bootstrap Monte Carlo statistical analysis as explained previously ((Sriram *et al*., 2004), Supplementary material IV).

2.3. Using OptForce for fatty acid overproduction:

The *i*AF1260 metabolic model of *Escherichia coli* (Feist *et al*., 2007) was used to perform the simulations with the OptForce procedure (Ranganathan and Maranas, 2010; Ranganathan *et al*., 2010; Xu *et al*., 2011b) for overproduction of fatty acids. Metabolic flux data for 35 reactions from the glycolytic, TCA and Pentose Phosphate pathway was used to define the phenotypic space of a base strain. All simulations were performed under aerobic minimal medium with glucose as the sole carbon source. Glucose minimal conditions were simulated by restricting the glucose uptake rate to 100 mmol gDW^{-1} h^{-1} and the oxygen uptake rate at 200 mmol gDW^{-1} h^{-1}. The lower bound for the remaining exchange fluxes corresponding to the metabolites present in the minimal medium was set to -1,000 and the non-growth associated ATP maintenance was fixed at 8.39 mmol gDW^{-1} h^{-1} (Feist *et al*., 2007). In addition, the biomass flux was fixed at the maximum achievable flux subject to the experimental flux measurements (i.e., 52% of the maximum theoretical). The upper bound for all other reactions was set to 1,000 whereas the lower bound was set to zero and -1,000 for irreversible and reversible reactions, respectively. All regulatory restrictions were imported from the *i*AF1260 model (Feist *et al*., 2007) except for the regulatory constraints repressing the β-oxidation pathway under aerobic minimal condition with glucose as the sole carbon source, which was excluded in this study. This is because previous studies have reported on a significant increase in fatty acids production yield upon removal of the β-oxidation pathway (Steen *et al*., 2010) implying its activity under this condition. No additional acyl-ACP thioesterase catalyzed reactions were included in the model since it already contained thiolase reactions hydrolyzing fatty-acyl ACPs to fatty acids. Also, it is important to note that no chain specificity of thioesterases is captured in the *i*AF1260 model. The phenotypic space of the wild type strain consistent with stoichiometry/regulation, uptake rates and flux measurements was constructed by successively maximizing and minimizing each reaction flux in the network subject to the network stoichiometry and all of the constraints mentioned above.

Similarly, the flux ranges consistent with a desired over-producing target for fatty acids of specific chain lengths were obtained by iteratively maximizing and minimizing each flux subject to the network stoichiometry, uptake and medium conditions, regulatory constraints and overproduction target. In this study, we imposed a minimum production yield of 90% of the theoretical maximum for all fatty acids of different lengths, while the biomass flux was constrained to be at least 10% of its theoretical maximum. The remaining parameter values were unchanged from the wild-type case. OptForce was subsequently used to identify the minimal set of reactions/genes that must be up-/down-regulated or knocked out so as to maximize the formation of targeted fatty acids. OptForce contrasts the maximal range of flux variability between the wild-type strain against the ones consistent for the overproducing phenotype designed to meet a pre-specified yield for hexanoate, octanoate, decanoate, dodecanoate, tetradecanoate and palmitate, respectively. As outlined in earlier efforts, by superimposing the flux ranges one-at-a-time, we first identify the fluxes that must depart from the original ranges in the face of overproduction ($MUST^{U}$, $MUST^{L}$, $MUST^{X}$ sets). You can extend this classification procedure by considering sums and differences of two fluxes ($MUST^{UU}$, $MUST^{UL}$, $MUST^{LL}$ sets) and arrive at a collective set of flux changes that must happen in the network for overproduction. In this study, we only considered up to MUST pairs as the available MFA data provided sufficient information to characterize the *wild-type* strain without having to consider higher order MUST sets. For example, we identified 193 MUST single reactions and 33 reactions participating in MUST doubles reactions (for C_8 fatty acid overproduction) providing a rich set of reaction alternatives to directly engineer. We subsequently extracted the minimal subset(s) of these reactions needed to guarantee the imposed bioengineering objective (i.e., FORCE sets). Here, we focused on elucidating the differences of the identified engineering strategies for different fatty acid lengths.

In this study we slightly modified the original formulation presented in (Ranganathan *et al*., 2010) for identifying the FORCE set. In particular, we make use of a max-min bilevel optimization problem to identify alternative sets of *k* (pre-specified) engineering interventions that maximize the minimum product formation (worst-case scenario) in the network (see Supplementary Text S1). Modeled as a "worst-case" optimization problem, the OptForce procedure identifies metabolic interventions that guarantee an increase in the yield even when metabolic fluxes are allotted so as to directly counteract the desired overproduction. This

optimization problem is solved successively, starting with a low number of direct interventions (i.e., $k = 1$) and then considering more interventions (by increasing k) until the target yield is achieved. Given that the objective function of the outer problem is maximization of the product formation, manipulations with the highest impact on the product yield are identified first. By increasing the value of k, additional modifications that improve upon the previously identified ones are revealed, thereby providing a way of prioritizing the manipulations based on their impact on the product yield. Binary variables are used here to identify pertinent reactions from the MUST sets whose flux should be increased, decreased, or set to zero (i.e., removed) in order to maximize the minimal product formation yield. Removal of reactions associated with *in vivo* essential genes based on the KEIO collection (Baba *et al*., 2006; Feist *et al*., 2007) in minimal glucose growth medium under aerobic condition, was disallowed even for the ones that were not recognized as essential by the *i*AF1260 model. For example, removal of phosphofructokinase (PFK) and fructose-bisphosphate aldolase (FBA) in the glycolytic pathway (which appear in the MUST sets for C_{12}), and glutamate-5-semialdehyde dehydrogenase (G5SD) or glutamate 5-kinase (GLU5K) reactions in the arginine/proline metabolism, (which appear in the MUST sets for all fatty acids), were prevented. Likewise, the removal of reactions catalyzed by multiple isozymes whose simultaneous knockouts have been experimentally verified to be lethal to the organism (Suthers *et al*., 2009) was also disallowed. Reactions whose isozymes formed synthetic lethal pairs include phosphoglycerate mutase (PGM) in the glycolytic pathway, transketolase (TKT1 and TKT2) in the Pentose Phosphate Pathway, and aconitase (ACONT) in the TCA cycle. In addition, metabolic interventions in the reaction level inconsistent with gene-level manipulations were avoided in the FORCE sets. For example, the simultaneous up-regulation and down-regulation (or removal) of chain elongation reactions for two different fatty acids was prevented if they were encoded by the same gene(s). The binary variables corresponding to the reaction interventions that appear in all solutions, as well as those corresponding to trivial solutions (e.g., the up-regulation of the transport reaction corresponding to target product) were also fixed at one and zero respectively to reduce run time. A biomass flux of at least 10% of theoretical maximum was enforced in all OptForce simulations, along with other constraints mentioned before. The use of integer cuts allows for the identification of alternate optimal solutions that can serve as alternate genetic intervention choices. Notably, when the target yield is not achievable with interventions selected only from within the MUST (single and pair) sets

we allow for one (or more) interventions (knock out/up/down) from outside the MUST sets by addition of appropriate binary variables and constraints to the max-min optimization formulation. For example, this led to the identification of engineering strategies in the β-oxidation pathway, which did not appear in any of the MUST single or double sets (see Results). The termination criterion for the OptForce procedure was either meeting a production yield of at least 90% of theoretical maximum for each fatty acid, or exceeding the maximum allowable number of reaction interventions (i.e., eight). It is worth noting that the OptForce procedure operates at the reaction level and the set of manipulations at the gene level are subsequently identified manually by using gene-protein-reaction (GPR) associations presented in the model. In principle, we could have run OptForce at the gene level by appending the corresponding GPR constraints in the formulation. However, we have found that it is more instructive to first identify interventions at the reaction level to fathom the reasoning behind the identified interventions.

2.4. Metabolic interventions and fatty acid titer determinations

Fermentation procedure:

Each strain was freshly transformed and streaked on LB plate with 100 mg/L ampicillin overnight at 30 ^{o}C incubator. A single colony from the plate was grown in 5 mL M9 medium supplemented with 1.5% glucose and 100 mg/L ampicillin for 16-20 hours in orbital shaker at 30^{o}C and 250 rpm. The pre-culture was then inoculated into 250 mL flasks containing 40mL M9 medium with 1.5% glucose and 100 mg/L ampicillin. The expression of acyl-ACP thioestersase was induced by the addition of isopropyl-β-D-thiogalactopyranoside (IPTG) to final concentration of 1mM. Samples were taken at 24 and 48 hours for fatty acid and extracellular metabolite analysis.

Fatty acid analysis:

Cell cultures were harvested and prepared for fatty acid analysis as described previously (Zhang *et al*., 2011). Fatty acids in the broth were extracted using chloroform, methylated into methyl esters and recovered using hexane. Tridecanoic acid, pentadecanoic acid and heptadecanoic acid were added as internal standards in all samples before extraction. The fatty acid content was analyzed using an Agilent GC-FID/MS system. The GC system occupies single quadrupole mass spectrometer with an electron impact ionization source and FID detector. The DB-5MS column (30m, 0.25 mm i.d., 0.25μm, Agilent) was used to separate the fatty acids into

different chain lengths. The oven temperature was initially set at 50 ^{o}C for 1 min and raised to 140 ^{o}C with 20 ^{o}C/min ramping rate. The temperature was then increased to 220 ^{o}C with 4 ^{o}C/min ramping rate and then finally raised to 280 ^{o}C with 15 ^{o}C/min ramping rate. Helium was used as the carrier gas with flow rate of 1mL/min. Interface temperature and ion source temperature was set as 280 ^{o}C and 250 ^{o}C respectively. EI ionization was set at 0 kV relative to the tuning. Mass spectra were analyzed using the full scan method. Raw MS and FID data was integrated using Chemstation software. Compound peaks were assigned by running standards or referring to the mass fragmentation in the NIST library.

3. Results

3.1. Flux measurements:

In vivo metabolic flux analysis, based on the use of ^{13}C-labeled glucose followed by NMR analysis and isotopomer balancing, was used for the estimation of intracellular fluxes as it provides a more comprehensive description of the metabolic network operating under the physiological conditions. Media and temperature conditions for the production of free fatty acids by *E. coli* include LB rich medium (Zhang *et al*., 2012b); (Zhang *et al*., 2011); (Li *et al*., 2012) at 30 °C, M9 minimal medium at 37 °C (Zhang *et al*., 2012a), and MOPS minimal medium at 37 °C (Youngquist *et al*., 2012). A reduced temperature has been noted to help stabilize strains carrying the plant acyl-ACP thioesterase being unstable (Zhang *et al*., 2011). Thus, two sets of flux experiments were performed; one for *E. coli* MG1655 in MOPS medium at 37 °C and one for ML103 (MG1655 Δ*fadD*) in minimal M9 medium at 30 °C, both kept under fully aerobic growth in 400 ml batch reactors.

The flux maps for MG1655 under MOPS media and ML103 under M9 media are tabulated in Table S2 and shown in Figure S1 and S2 in the supplemental material. The flux values are all normalized to 100 mmol $gDW^{-1}\,h^{-1}$. Inspection of the normalized flux data in Table S2 indicates that the flux values are nearly identical for the two experiments. Most of the carbon flux (around 88%) is directed towards the glycolytic pathway, resulting in high activities of the lower glycolytic pathway. Around 10% of the carbon flux channels through pentose phosphate pathway to generate NADPH for reduction requirements. The ED pathway and glyoxylate pathway have negligible fluxes. The anapleurotic pathway of phosphoenolpyruvate (PEP)

carboxylase is active, which converts PEP to oxaloacetate (OAA) to refill the OAA pool for biosynthesis. These results are consistent with previous flux experiments showing active PEP carboxylase activity and inactive glyoxylate cycle under glucose aerobic batch culture (Fischer and Sauer, 2003a; Fischer and Sauer, 2003b). Acetate kinase is active to convert acetyl-CoA for acetate production. The TCA cycle operates at 45 mmol $gDW^{-1}\,h^{-1}$ (based on 100 mmol $gDW^{-1}\,h^{-1}$) to generate ATP and NAD(P)H for energy and reduction requirements for cell growth. Malic enzyme activity is not significant (Fischer and Sauer, 2003a).

3.2. OptForce results

Targeted pathway:

Figure 1 illustrates the pathways involved in fatty acid biosynthesis in *E. coli*. There are two important fatty acid pathways in *E. coli* metabolism (Fujita *et al*., 2007; Marrakchi *et al*., 2002; Schweizer and Hofmann, 2004). Type II or dissociated fatty acid biosynthesis (FAB) pathway involves the ATP-dependent acetyl-CoA carboxylase (encoded by *accABC*) as the first step. Acetyl-CoA is converted into malonyl-CoA which is further converted into malonyl-ACP by the enzyme malonyl-CoA: ACP transacetylase encoded by the gene *fabD*. The initiation of fatty acid biosynthesis starts with the C_4 chain where one mole of acetyl-CoA and one mole of malonyl-ACP synthesize a 4-carbon fatty acid acyl carrier protein (i.e. butyryl ACP). Butyryl ACP further elongates into the 6-carbon chain by recruiting one mole of malonyl-ACP to produce hexanoyl-ACP. This chain elongation step uses one mole of malonyl-ACP per cycle to form even-numbered fatty acid acyl carrier proteins. Fatty acid ACP is converted into a fatty acid by thioesterases in a single step enzymatic conversion. An alternative biosynthesis route for the production of fatty acids that has gained attention recently (Dellomonaco *et al*., 2011) is reversal of the β-oxidation pathway. While the native use of this pathway is to disassemble longer chain fatty acid ACP into smaller coenzyme-A derivatives, reversal of this pathway may lead to fatty acid synthesis as was recently demonstrated by Dellomonaco *et al* (2011) for overproduction of 1-butanol and a number of long chain fatty acids. The calculated maximum theoretical yield for both pathways is similar; however, the type II fatty acid biosynthesis is dependent on the availability of ATP and the specificity of the termination enzyme (i.e. thioesterase) ((Dehesh *et al*., 1996); (Lennen *et al*., 2010)). In this paper, we use OptForce for identifying genetic

interventions in *E. coli* that result in increased production of free fatty acids using the type II fatty acid pathways.

Identification of MUST sets:

Figure 2 summarizes the set of fluxes in the network that MUST change when overproduction objectives for fatty acids of specific lengths are imposed. Results are presented in Figure 2 so as to highlight the conservation of MUST changes as the fatty acid chain length increases (Figure 2a) or decreases (Figure 2b). For example, in Figure 2a the additional reactions that enter the MUST set are shown within the growing ellipses as the fatty acid length changes from 6 to 16 carbons. In contrast, in Figure 2b the reactions added to the MUST set are listed as the fatty acid length decreases from 16 carbons down to 6. Figure 2 provides a pictorial view of the conservation patterns of the required changes (i.e., MUST sets) in the metabolic networks as the fatty acid chain length increases or decreases, respectively. For example, increase in the flux for any of the chain elongation reactions 3-oxy-acyl-ACP synthase (3OAS40/60/80), 3-oxo-acyl-ACP reductase (3OAR40/60/80) and 3-hydroxy-acyl dehydratase (3HAD40/60/80) ensure higher flow of carbon through the fatty acid synthesis pathway, and augment fatty acid production. These chain elongation reactions appear in the MUST sets of fatty acids up to C_{12}, as is shown in the C_{12} ellipse on the right. Even though these reactions do not appear in the (single or double) MUST sets of C_{14} and C_{16} fatty acids, it is likely that they appear in higher order MUST sets (i.e. Must Triples, Quadruples). Elimination of acetate kinase (ACK) and phosphotransacetylase (PTA) appear in MUST Sets of all fatty acids as shown in the C_6 ellipse on the left, as they prevent degradation of pyruvate towards fermentation byproducts. Similarly, elimination of malate dehydrogenase (MDH) and downregulation of citrate synthase (CS) redirects metabolic flux towards fatty acid synthesis by reducing consumption of acetyl-CoA in the TCA cycle and are universally found for all fatty acid lengths.

In addition to these universal changes, a number of network modifications need to take place for a given chain length or higher (see left panel) or a given chain length or lower (see right panel). For example, up-regulation for enolase (ENO) in glycolytic pathway appear only for C_{10} and longer chains, which is consistent with increased requirements of carbon flux towards fatty acid synthesis for longer chain lengths. Furthermore, up-regulation of pentose phosphate (PP) reaction 6-phosphogluconolactonase (PGL) for fatty acids C_{12} and longer, and eliminations of

glycolytic pathway reactions phosphofructokinase (PFK) and fructose-bisphosphate aldolase (FBA) are indicative of the increased need of rerouting flux through the PP pathway to produce reducing agents NADPH required in fatty acid chain elongation.

Notably, reactions along the β-oxidation pathway for the C_8 fatty acid were classified in the $MUST^U$ set of reactions only for hexanoic acid. This is because the *i*AF1260 metabolic model does not contain a fatty acid-acyl-ACP hydrolase that can directly convert hexanoyl-ACP into hexanoic acid. Instead, the favored pathway involves chain elongation into the C_8 chain and subsequently, octanoyl-ACP is reduced into hexanoyl-CoA via the β-oxidation pathway. Eventually, hexanoyl-CoA is converted into hexanoic acid by the thioesterase. However, for the higher chain fatty acids (C_8 or higher), the β-oxidation pathway is not the favored synthesis route because the hydrolase that directly converts ACP-bound end products to the corresponding acids is included in the *i*AF1260 model. It is interesting to note that in contrast to C_6 results (see Figure 3), reaction removals in the β-oxidation pathway are required for overproduction of C_8 (see Figure 4).

Upregulation of pyruvate dehydrogenase (PDH), which appears in MUST sets of all fatty acids, leads to the production of acetyl-CoA which is converted to malonyl-ACP (mal-ACP) fueling the chain elongation reactions. In addition, eliminations in the TCA cycle and pathways branching out from glycolysis, such as glycine hydroxymethyltransferase (GHMT), prevent leaking of the glycolytic flux thus ensuring maximum carbon flow towards fatty acid synthesis chain.

Identification of FORCE sets:

Using as candidates the reactions that populate the MUST sets, OptForce max-min optimization formulation (see Methods) is next used to identify minimal sets of engineering modifications for each specific fatty acid chain length. As noted earlier, the termination criterion for the OptForce procedure was either meeting a production yield of at least 90% of theoretical maximum or exceeding the maximum allowable number of reaction interventions (i.e., eight). The identified FORCE sets for each fatty acid chain length are discussed in the subsequent sections.

Hexanoic acid

Figure 3 depicts the genetic engineering strategies for overproducing hexanoic acid identified by OptForce using the MUST sets as candidate interventions. Results for hexanoic acid, as well as for all higher chain acids revealed that no non-zero minimal yield of product could be guaranteed by using only one intervention. By allowing up to two interventions OptForce predicted the upregulation of any of the reactions of β-oxidation pathway along the C_8 chain (i.e., octanoyl-CoA dehydrogenase (ACOAD3), 3-oxooctanoyl-CoA dehydrogenase (HACD3) or 3-ketoacyl-CoA thiolase (KAT3)) coupled with the removal of any of the β-oxidation reactions along the C_4 chain (i.e., ACOAD, HACD1 or KAT1). The up-regulation of the β-oxidation reactions along C_8 chain by at least two times the maximum achievable wild-type flux (i.e., from 28 to 54 mmol gDW^{-1} h^{-1}) causes degradation of longer-chain fatty acids to hexanoic acid while the elimination of reactions along the C_4 chain prevents further degradation of any hexanoate formed. It is worth noting that most fatty acid degradation steps corresponding to different chain-lengths in the β-oxidation pathway are encoded by the same gene(s). For example, *fad*E encodes the acyl-CoA dehydrogenase (ACOAD) reactions for all chain lengths while *fad*A does the same for all thiolases (KAT). It is therefore impossible to simultaneously up-regulate the degradation step for a higher chain-length and down-regulate (or knock-out) the one for a shorter chain using interventions at the gene level. However, the enzyme catalyzing reaction ACOAD1 (EC 1.3.8.1) differs from the ones catalyzing the same reaction in longer chain acids (EC 1.3.99.3) thereby providing a feasible route for an independent manipulation. OptForce predicted that the aforementioned two interventions would be enough to achieve a theoretical yield of 90% for hexanoic acid and hence, we did not explore additional genetic manipulations.

Octanoic and decanoic acid

Metabolic interventions predicted by OptForce for octanoic (C_8) and decanoic (C_{10}) acids are quite similar relying on the strict redirection of carbon flux from glycolytic pathways to fatty acid biosynthesis (see Figures 4 and 5). OptForce predicts that at least four and five interventions respectively are required to achieve a theoretical yield of ~90% for both octanoic and decanoic acid. The primary interventions which account for approximately 86% of the yield increase include up-regulation of any of the chain-elongation reactions (i.e. 3-oxy-acyl-ACP synthase (3OAS), 3-oxo-acyl-ACP reductase (3OAR) or 3-hydroxy-acyl dehydratase (3HAD)) by at least

two times of the maximum achievable flux in the wild type (i.e., from 28 to 54 mmol $gDW^{-1} h^{-1}$) followed by reaction removals in the β-oxidation pathway. OptForce suggests up-regulating any one of the chain-elongation reactions along C_8 chain that directly leads to synthesis of octanoate. Reaction removal in the β-oxidation pathway along the C_8 chain (i.e., ACOAD3, ECOAH3i, HACD3 or KAT3) prevents further degradation of the end product. This same holds true for interventions along the C_{10} chain.

OptForce suggests the elimination of fumarase (FUM) in the TCA cycle to maintain a high pool of acetyl-CoA and redirect flux towards the fatty acid elongation chain. This is because the demand of mal-ACP, which serves as the primary building block of fatty acids, increases proportionally with the chain length. For example, each molecule of octanoate requires 3 molecules of mal-ACP whereas one molecule of decanoate requires four. Mal-ACP is produced through the carboxylation of acetyl-CoA. In addition, OptForce suggests the elimination of acetate kinase (ACK) or phosphotransacetylase (PTA) possibly to prevent utilization of acetyl-CoA towards production of acetate. Removal of transaldolase (TALA) in the Pentose Phosphate pathway was also suggested to further redirect glycolytic flux towards decanoate overproduction.

Notably, OptForce does not suggest up-regulation of acetyl-CoA carboxylase as a potential intervention for any of the fatty acids. This is consistent with previous reports where the up-regulation of acetyl-CoA carboxylase (ACCOAC) to increase the pool of malonyl-ACP (Lennen *et al*., 2010; Lu *et al*., 2008), did not lead to any significant increase in production of fatty acids. A possible reason for this may be that in the absence of a sink that consumes excess malonyl-CoA (e.g., fatty acid biosynthesis); it is mostly diverted towards biomass component formation.

C_{12} and longer chain fatty acids

Figures 6-8 show the genetic manipulations suggested by OptForce for the overproduction of dodecanoate (C_{12}), tetradecanoate (C_{14}) and palmitate (C_{16}) in *E. coli*. Consistent with the trends observed for Octanoate and Decanoate, up-regulations of one of the chain elongation reactions for fatty acid pathways (matching the desired chain length) as well as removal of the corresponding β-oxidation pathway are predicted. Interestingly, the specific set of interventions for C_{12} and longer chain-length fatty acids includes redirecting glycolytic flux

through the oxidative phase of the Entner–Doudoroff (ED) pathway leading to additional NADPH at the expense of ATP production.

Figure 6 represents the metabolic interventions suggested for C_{12}. Similar to the previous cases, OptForce predicts chain-specific manipulations for reactions in the elongation step of fatty acid biosynthesis and β-oxidation pathways along the C_{12} chain. Interestingly, OptForce also requires an at least eight-fold reduction in the phosphoglycerate mutase (PGM) flux along with the removal of glucose-6-phosphate isomerase (PGI) to bypass the lower and upper glycolytic pathway and instead redirect the metabolic flux towards pyruvate through serine metabolism and Entner–Doudoroff (ED) pathway, respectively. This drastic rewiring of metabolism was suggested by OptForce in order to utilize a less energy efficient route towards production of pyruvate by bypassing the pyruvate kinase (PYK) reaction which generates one mole of ATP and redirecting the metabolic flux through ED pathway, which generates less ATP compared to glycolytic pathway. A possible reason for the preference of a low energy efficient pathway is to arrest the cell growth and channel more metabolic flux towards pyruvate and fatty acid biosynthesis, similar to what has been observed for ethanol production in *Zymomonas mobilis* (Zhang *et al*., 1995). We computationally explored the validity of this hypothesis by artificially decreasing the energy efficiency of the glycolysis pathway through reducing the stoichiometric coefficient of ATP (and ADP) in phosphoglucokinase (PGK). In particular, upon reducing the stoichiometric coefficient of ATP (and ADP) from one to 0.96, we observed a 1.6% increase in the production yield of C_{12} as well as a 1.9% decrease in the maximum biomass formation in the network supporting the put forth hypothesis. Using a less energy efficient pathway leads to the decreased availability of ATP for further conversion of dodecanoate towards tetradecanoate, as manifested by a 95% decrease in tetradecanoate formation. Notably, reducing the stoichiometric coefficient of ATP in PGK beyond 0.96 renders the optimization problem infeasible, as the imposed constraint on the minimum biomass formation in the network (i.e. 10% of theoretical maximum) cannot be satisfied.

In addition, as fatty acid chain length increases, the NADPH demand for the reduction reactions in chain elongation steps also increases. Hence, OptForce suggests utilization of a more efficient NADPH producing pathway by redirecting the glycolysis flux towards the oxidative phase of PP Pathway in order to gain one additional mole of NADPH per mole of glucose (through glucose 6-phosphate dehydrogenase), which could be supplied to 3-oxo-acyl-ACP

reductase (3OAR) and enoyl-ACP reductase (EAR). A yield of ~66% of theoretical maximum was obtained after seven interventions as described above. Interventions for the overproduction of the C_{14} and C_{16} fatty acid (see Figures 7 and 8) follow the same pattern observed for C_{12}. For example, OptForce predicts downregulation of PGM flux and removal of PGI to the lower and upper glycolysis, respectively. In addition, OptForce suggests a four-fold reduction of either methenyltetrahydrofolate cyclohydrolase (MTHFC) or methylenetetrahydrofolate dehydrogenase (MTHFD) to reduce metabolic flux from being diverted towards folate metabolism. Additional interventions for C_{16} fatty acid (see Figure 8), include up-regulation of pyruvate dehydrogenase (PDH) by at least 1.5 times its wild-type maximum value, thus directly enhancing the acetyl-CoA pool for fatty acid production. After seven and eight interventions respectively, OptForce predicted a yield of ~65% of theoretical maximum for both tetradecanoate and palmitate. Allowing for further interventions did not lead any appreciable increase in the guaranteed yield for dodecanoate (C_{12}) or tetradecanoate (C_{14}).

The genetic manipulations required in *E. coli* for the overproduction of palmitate (C_{16} fatty acid) and corresponding impact on the yield are shown in Figure 9. As seen in the figure, the up-regulation of one of the elongation reactions in the C_{16} chain conjunction with a reaction removal in the β-oxidation pathway, along with redirection of the glycolytic flux leads to an increase in the yield of about 32% of the theoretical maximum. Additional deletions and knock-downs result in improving the yield close to 66% of theoretical maximum. Notably, OptForce suggests that as we move towards longer chain fatty acids, the number of genetic interventions required in central metabolism increases as the carbon flow re-direction becomes more pronounced.

3.3. Experimental characterization of metabolic interventions

We chose to test OptForce predictions for the production of medium chain-length fatty acids as we had access to plasmids with acyl-ACP thioesterase gene from *Ricinus communis*, which produces a mixture of C_{14} and C_{16} fatty acids. To the best of our knowledge, a thioesterase that strictly produces C_{14} or C_{16} alone has not been identified yet. The first set of prioritized interventions suggested by OptForce for overproduction of C_{14} and C_{16} fatty acid include up-regulation of one of the chain elongation reactions for fatty acid as well as removal of the corresponding β-oxidation pathway (Figures 7 and 8). We thus implemented these two

interventions by overexpresing *fabZ*, which encodes a β-hydroxyacyl-ACP dehydratase, and eliminating the β-oxidation pathway through the deletion of *fadD*.

Strain ML103 (MG1655 *ΔfadD)* with plasmid pXZ18, which carries the gene for the C_{14-16} thioesterase, serves as the reference strain for comparison since it was shown to produce free fatty acid titer and yield similar to those from strain MG1655 pXZ18 in rich media conditions (Li *et al*., 2012). Since OptForce predictions are based on flux data using defined minimal medium, experiments were performed using minimal M9 medium with 1.5% glucose as shown in Figure 10. With the overexpression of *fabZ*, the fatty acid titer increased 3.5 fold (from 0.6 g/L for the base strain to 1.7 g/L total fatty acids) after 48 hours cultivation, whereas the yield was increased from 0.04 g fatty acid/g glucose for the base strain (~ 11% maximum theoretical yield) to 0.14 g fatty acid/g glucose (~ 39% maximum theoretical yield) (Figure S3). These results are consistent with the OptForce predicted minimum yield of 0.12 g C_{16} /g glucose after first set of interventions (Figure 9). Overexpression of fatty acid elongation reaction pulls the carbon fluxes from acetyl-CoA in the central carbon metabolism to form malonyl-CoA as the precursor for fatty acid synthesis. This "pull" is in addition to the one provided by expression of the heterologous thioesterase, as the expression of pXZ18 alone already lowered acetate yields from 0.5-0.65 mol acetate/mol glucose in MG1655 and ML103 (Table S3) to ~ 0.08 mol acetate/mol glucose in ML103 pXZ18 and ML103 pXZ18z (Figure S4 and S5). It is important to note that OptForce predictions are based upon network stoichiometry alone, and do not involve reaction kinetics. However, the above results suggest that the heterologous thioesterase is not rate limiting in fatty acid production, and that additional interventions in addition to *fabZ* overexpression should help increase the yield.

Notably, another intervention predicted by OptForce includes downregulation of the TCA cycle to maintain high pool of acetyl-CoA and channel the carbon fluxes towards fatty acid synthesis. Nonetheless, OptForce suggests that the fatty acid titer and yield improvement upon this intervention are not as significant as those achievable with overexpression of *fabZ* (see Figure 9). To test this prediction, we took advantage of an already available strain containing *sucC* deletion in the TCA cycle (MLK163 pXZ18) and it was observed that this intervention increased the fatty acid titer and yield in M9 medium by only 2.4 fold to 1.3 g/L after 48 hours, corresponding to a yield of 0.12 g fatty acid/g glucose (see Figure 10 and Figure S3), which are

lower than those achieved with overexpression of *fabZ* thereby corroborating the OptForce predictions.

The relative composition of the fatty acids produced by the strains ML103 pXZ18 (reference strain), ML103 pXZ18z (*fabZ*$^{++}$) and MLK163 pXZ18 (Δ*sucC*) at 24 and 48 hours are shown in Figure 11, which reveals an abundance of mostly C_{14} and C_{16} straight chain lengths. The fractional composition of C_{14} fatty acid increased over time at the expense of $C_{16:1}$ mon-unsaturated fatty acid. The composition of saturated $C_{16:0}$ fatty acid did not change significantly. Changes in the composition of different fatty acid chain lengths over time can possibly be explained due to the changes in the cellular physiology at different phases of growth (Zhang *et al*., 2011).

4. Summary and discussion

In this paper, we described computationally derived predictions followed by experimental characterization of strategies for overproducing fatty acids in *E. coli*. Suggested modifications include not only straightforward up-regulations of terminal pathways but also many modifications distant to the fatty acid target that prune away competing pathways, up-regulate pathways to accommodate increased precursor flows or increase the availability of relevant cofactors. Contrary to many existing strategies (Davis *et al*., 2000; James and Cronan, 2004; Subrahmanyam and Cronan, 1998) that rely on augmenting acetyl-CoA and malonyl-CoA pools, OptForce does not suggest the overexpression of acetyl-CoA carboxylase (*accABC*). As noted earlier, this is to avoid diverting resources towards biomass formation as malonyl-CoA is a key precursor for many biomass constituents. This is in agreement with a recent study (Xu *et al*., 2011a), that observed experimentally that an increase in the intracellular levels of malonyl-CoA leads to significant cell growth increase. By overexpressing fatty acid enzymes, malonyl-CoA is diverted from biomass formation towards fatty acid biosynthesis. Interestingly, OptForce suggested the up-regulation of one of the four reactions in the β-oxidation pathway for overproducing hexanoate. In the *i*AF1260 *E. coli* metabolic model, acyl-ACP thioesterase enzyme that catalyzes the conversion of hexanoyl-ACP into hexanoic acid is absent. Hence, the desired route to produce hexanoic acid is by elongating the chain further into the C_8 chain and degrading octanoyl-ACP through the β-oxidation cycle.

The chain-dependent nature of the OptForce interventions is reflected in the Venn diagram shown in Figure 12. No universal engineering strategy was predicted for overproduction of all fatty acids, indicating the chain length specificity of each of the strategies. We observe that the up-regulation of fatty acid reactions is completely chain specific. For example, for overproducing C_{10} fatty acid the up-regulation of only the C_{10} pathway with elimination in the C_{10} β-oxidation pathway is needed. As we move along in a clock-wise direction (see Figure 12), genetic manipulations that reduce the activity of TCA cycle and reductive part of the Pentose Phosphate Pathway and increase the carbon flow towards the lower part of glycolysis start to emerge. Interventions for fatty acids of longer chain lengths (i.e. C_{12}, C_{14} and C_{16}) also necessitate diversion of glycolytic flux to the serine metabolism and ED pathway which are less efficient ATP but more efficient NADPH producing pathways. This redirects flux from cell growth towards satisfying the increased NADPH demand for the reduction steps of the chain elongation pathways. This suggests that upon targeting a fatty acid of higher chain length, a stricter redirection of central metabolic carbon flow towards the precursors becomes progressively more important.

Computational predictions from OptForce were validated for the C_{14}-C_{16} chain length by examining the fatty acid production of a large number of engineered strains carrying a designed medium chain (i.e., C_{14}-C_{16}) thioesterase from *R. communis* in minimal (M9) medium. As mentioned earlier, the common thread from OptForce indicates an upregulation in the fatty acid chain elongation pathway (except for C_6 fatty acid, which lacks a thiolase to hydrolyze C_6 fatty acyl-ACP) as the intervention of primal importance in the overproduction of fatty acids. Transcriptional and post-transcriptional control in *Escherichia coli* tightly regulates the metabolism of fatty acid biosynthesis, making it difficult to decide on specific genetic manipulations. OptForce can only predict the enzymatic steps which require alteration, but cannot provide any inference about regulatory genes. Therefore, several host strains and plasmids were constructed to test our current understanding of regulation in the fatty acid biosynthesis. In accordance with the OptForce prioritization of interventions, *fabZ* and *fadD* were upregulated and deleted, respectively (in a strain carrying upregulated C14-16 Acyl-ACP thioesterase) to arrive at a strain that produces 1.7 g/L of $C_{14\text{-}16}$ fatty acids and 0.14 g fatty acid/g glucose (~ 39% maximum theoretical yield) in M9 medium.

We cannot, however, rule out the possibility that other manipulations may increase fatty acid production further. Independent from this work, San and co-workers recently used a classical "push and pull" concept in metabolic engineering, in which acetyl CoA supply is enhanced, acetyl CoA drains are minimized, by-product pathways are eliminated, and product formation pathways are enhanced. A large number of engineered strains carrying a designed medium chain (C_{14}-C_{16}) thioesterase, pXZ18, were screened in their ability to increase fatty acid yield and titer in LB medium at 30 °C. The results of their study are summarized in Figure 13 (San *et al*., 2011). The increases/decreases shown are in reference to ML 103 (pXZ18), which had a titer of 3.1 g/L and yield of 0.17 g fatty acid/g glucose at 48 hours. Although these strains were grown in LB medium and the manipulations were performed independently, the results are in general agreement with OptForce predictions. For example, overexpression of *fabZ* (in a strain carrying *fadD* knockout) leads to the highest yield and titer in rich medium. Manipulation of transcription factors (*fabR* and *fadR*) in fatty acid biosynthesis, the deletion of genes in the TCA cycle (*sucC, fumAC* and *gltA*) and the deletion of genes in glycolytic pathway (*glk, ptsG, pfkA* and *pykF*) also improve fatty acid yield in rich media, but not as much as that obtained from *fabZ* overexpression (San *et al*., 2011). Note that the order of improvements in fatty acid titer was: $fabZ^{++}$ > $fadR^{++}$ > Δ*sucC* > Δ*fabR*> Δ*glk* > Δ*pykF* > Δ*fumAC*. This independent classical genetic intervention study not only reinforces OptForce predictions to upregulate fatty acid biosynthesis, downregulate TCA cycle and redirect glycolysis flux towards fatty acid production, but also closely emulates the prioritization of interventions suggested by OptForce.

Even though both *fabA* and *fabZ* are genes responsible for dehydration of β-hydroxyacyl-acyl carrier protein (ACP) in *E.coli*, overexpression of *fabA* gene in fatty acid elongation cycle led to conflicting results on fatty acid titer and yield (Fig. 13) (San *et al*., 2011). The decrease in fatty acid production may be explained by the high complexity of fatty acid synthesis regulation. Indirect upregulation of fatty acid chain elongation reactions by deletion of *fabR* and overexpression of *fadR* was shown to increase fatty acid titer and yield (San *et al*., 2011). *FabR* and *fadR* transcription factors regulate fatty acid biosynthesis in *E.coli* simultaneously. *FabR* represses the expression of the fatty acid synthesis gene *fabB* and *fabA*. On the other hand, *fadR* acts as a repressor to regulate fatty acid degradation and it also activates *fabA* and *fabB* (Campbell and Cronan, 2001). Even though OptForce does not incorporate gene regulatory networks in its framework, the overall idea of upregulating fatty acid chain elongation steps can

pinpoint some target transcription factor to be engineered. Interestingly, an *in vitro* kinetic analysis carried out by Yu *et al* (2011) revealed that a specific combinatorial overexpression of all the enzymes (i.e. *ACP, fabH, fabB, fabG, fabZ, fabI* and *tesA*) in the fatty acid synthesis pathway is necessary for optimizing the productivity of fatty acids. Their observations infer that further research is required to fully understand the complexity of the fatty acid synthesis as a collective activity of all the enzymes, rather than as a sequence of individual steps as considered at present.

Removal of fumarase (FUM) reaction suggested by OptForce was tested independently by San *et al* (2011) using a strain carrying *fumAC* knockout, where only small increase of fatty acid titer and yield was achieved, which is in agreement with OptForce prioritization of interventions. An alternative for fumarase knockout to downregulate TCA cycle flux and redirect carbon flux to malonyl CoA pool was *sucC* knockout which was also tested by San *et al* (2011). It was observed that *sucC* deletion outperforms deletion of *fumAC*, however the reason for this is not clear yet. Overall, downregulation of TCA cycle is proven to be an effective strategy to reroute carbon fluxes towards fatty acid synthesis.

OptForce also suggests downregulation of phosphoglycerate mutase (PGM) flux and removal of glucose-6-phosphate isomerase (PGI) reaction in the glycolytic pathway to redirect carbon fluxes through Pentose Phosphate Pathway for medium chain fatty acid production. Even though these interventions were not performed experimentally in this study, previous efforts have already indicated the impact of these removals on improving fatty acids production yield. For example, San *et al* (2011) manipulated upper and lower glycolysis with deletion of *glk, ptsG, pfkA* and *pykF* (Fig. 13). More than 20% of improvement in fatty acid yield was achieved among all the strains. Deletion of *pfkF* and *pykF* genes showed slight improvement in fatty acid titer, whereas deletion of *ptsG* and *pfkA* showed inferior productivity. A possible explanation is the biological burden on the cells causing slower growth and lower glucose uptake rate. However, *R.communis* thioesterase effectively pulls the carbons towards fatty acid synthesis even with lower glucose uptake, leading to improved fatty acid yield.

In another study Steen and coworkers (2010) reported cytosolic expression of *E.coli* thioesterase *TesA* with deletion of β-oxidation pathway genes *fadD* and *fadE* to obtain ~0.7g/L and ~1.1g/L free fatty acid titer respectively. Since fatty acid production is highly regulated,

Zhang *et al* (2012a) subsequently developed a dynamic sensor-regulator system to produce fatty acid-derived products in *E.coli*. Fatty acid/acyl CoA biosensor was engineered based on *fadR* transcription factor and transformed into an *E.coli* strain carrying cytosolic *TesA* thioesterase. This strain was reported to produce 3.8g/L fatty acid after 3 day cultivation under M9 medium with 2% glucose supplemented with MOPS, mineral and micronutrients. The fatty acid titer was lower in M9 media than in LB media due to nutrient limitation. We anticipate higher fatty acid production for ML103 pXZ18z (Δ*fadD*, $fabZ^{++}$) can be achieved by using richer medium, higher glucose supplement and longer culturing period. Further effort can be explored by incorporating fatty acid/acyl CoA biosensor (engineered *fadR*) into *E.coli* strain ML103 pXZ18z (Δ*fadD*, $fabZ^{++}$) to increase fatty acid titer. Interestingly, deletion of *fadD/fadE* and overexpression of *fadR* as biosensor agree with OptForce suggestion to upregulate fatty acid synthesis and to eliminate β-oxidation of fatty acid for higher production.

While many earlier studies are based on expert intuition and the use of rich media, a systematic quantitative approach (i.e. OptForce supplemented with experimental flux results) with the employment of defined minimal media employed in this study provide a paradigm to a shorter turnover for strain development and cost saving from an industrial standpoint.

Authors' contributions

JVS and CDM conceived the project. TWT, JMY and YF designed and performed the experiments and the metabolic flux analysis. SR AC and ARZ designed and performed the OptForce simulations SR, AC, ARZ, TWT, JMY, JVS and CDM analyzed the data and wrote the paper. All authors have read and confirmed the manuscript.

Acknowledgement

This material is based upon work supported by the National Science Foundation under Award No. EEC-0813570

References

Atsumi, S., Wu, T. Y., Eckl, E. M., Hawkins, S. D., Buelter, T., Liao, J. C., 2010. Engineering the isobutanol biosynthetic pathway in Escherichia coli by comparison of three aldehyde reductase/alcohol dehydrogenase genes. Appl Microbiol Biotechnol. 85, 651-7.

Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M., Datsenko, K. A., Tomita, M., Wanner, B. L., Mori, H., 2006. Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. Mol Syst Biol. 2, 2006 0008.

Burgard, A. P., Pharkya, P., Maranas, C. D., 2003. Optknock: a bilevel programming framework for identifying gene knockout strategies for microbial strain optimization. Biotechnol Bioeng. 84, 647-57.

Campbell, J. W., Cronan, J. E., Jr., 2001. Escherichia coli FadR positively regulates transcription of the fabB fatty acid biosynthetic gene. J Bacteriol. 183, 5982-90.

Choudhary, M. K., Yoon, J. M., Gonzalez, R., Shanks, J. V., 2011. Re-examination of metabolic fluxes in Escherichia coli during anaerobic fermentation of glucose using (13)C labeling experiments and 2-dimensional nuclear magnetic resonance (NMR) spectroscopy. Biotechnology and Bioprocess Engineering. 16, 419-437.

Davis, M. S., Solbiati, J., Cronan, J. E., Jr., 2000. Overproduction of acetyl-CoA carboxylase activity increases the rate of fatty acid biosynthesis in Escherichia coli. J Biol Chem. 275, 28593-8.

Dehesh, K., Jones, A., Knutzon, D. S., Voelker, T. A., 1996. Production of high levels of 8:0 and 10:0 fatty acids in transgenic canola by overexpression of Ch FatB2, a thioesterase cDNA from Cuphea hookeriana. Plant J. 9, 167-72.

Dellomonaco, C., Clomburg, J. M., Miller, E. N., Gonzalez, R., 2011. Engineered reversal of the beta-oxidation cycle for the synthesis of fuels and chemicals. Nature. 476, 355-9.

Feist, A. M., Henry, C. S., Reed, J. L., Krummenacker, M., Joyce, A. R., Karp, P. D., Broadbelt, L. J., Hatzimanikatis, V., Palsson, B. O., 2007. A genome-scale metabolic reconstruction for Escherichia coli K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. Mol Syst Biol. 3, 121.

Fischer, E., Sauer, U., 2003a. Metabolic flux profiling of Escherichia coli mutants in central carbon metabolism using GC-MS. European Journal of Biochemistry. 270, 880-891.

Fischer, E., Sauer, U., 2003b. A novel metabolic cycle catalyzes glucose oxidation and anaplerosis in hungry Escherichia coli. Journal of Biological Chemistry. 278, 46446-46451.

Fischer, E., Zamboni, N., Sauer, U., 2004. High-throughput metabolic flux analysis based on gas chromatography-mass spectrometry derived C-13 constraints. Analytical Biochemistry. 325, 308-316.

Fjerbaek, L., Christensen, K. V., Norddahl, B., 2009. A review of the current state of biodiesel production using enzymatic transesterification. Biotechnol Bioeng. 102, 1298-315.

Fortman, J. L., Chhabra, S., Mukhopadhyay, A., Chou, H., Lee, T. S., Steen, E., Keasling, J. D., 2008. Biofuel alternatives to ethanol: pumping the microbial well. Trends Biotechnol. 26, 375-81.

Fujita, Y., Matsuoka, H., Hirooka, K., 2007. Regulation of fatty acid metabolism in bacteria. Mol Microbiol. 66, 829-39.

Gonzalez, E., Fernandez, M. R., Marco, D., Calam, E., Sumoy, L., Pares, X., Dequin, S., Biosca, J. A., 2010. Role of Saccharomyces cerevisiae oxidoreductases Bdh1p and Ara1p in the metabolism of acetoin and 2,3-butanediol. Appl Environ Microbiol. 76, 670-9.

Gulevich, A. Y., Skorokhodova, A. Y., Sukhozhenko, A. V., Shakulov, R. S., Debabov, V. G., 2011. Metabolic engineering of Escherichia coli for 1-butanol biosynthesis through the inverted aerobic fatty acid beta-oxidation pathway. Biotechnol Lett.

Handke, P., Lynch, S. A., Gill, R. T., 2011. Application and engineering of fatty acid biosynthesis in Escherichia coli for advanced fuels and chemicals. Metabolic Engineering. 13, 28-37.

Heath, R. J., Rock, C. O., 1996a. Inhibition of beta-ketoacyl-acyl carrier protein synthase III (FabH) by acyl-acyl carrier protein in Escherichia coli. J Biol Chem. 271, 10996-1000.

Heath, R. J., Rock, C. O., 1996b. Regulation of fatty acid elongation and initiation by acyl-acyl carrier protein in Escherichia coli. J Biol Chem. 271, 1833-6.

James, E. S., Cronan, J. E., 2004. Expression of two Escherichia coli acetyl-CoA carboxylase subunits is autoregulated. J Biol Chem. 279, 2520-7.

Jing, F., Cantu, D. C., Tvaruzkova, J., Chipman, J. P., Nikolau, B. J., Yandeau-Nelson, M. D., Reilly, P. J., 2011. Phylogenetic and experimental characterization of an acyl-ACP thioesterase family reveals significant diversity in enzymatic specificity and activity. Bmc Biochemistry. 12.

John L. Ingraham, O. M., Frederick C. Neidhardt, 1983. Growth of the Bacterial Cell. Sinauer Associates, Inc.

Johnson, B. A., Blevins, R. A., 1994. NMR VIEW - A COMPUTER-PROGRAM FOR THE VISUALIZATION AND ANALYSIS OF NMR DATA. Journal of Biomolecular Nmr. 4, 603-614.

Kim, J., Reed, J. L., 2010. OptORF: Optimal metabolic and regulatory perturbations for metabolic engineering of microbial strains. BMC Syst Biol. 4, 53.

Koffas, M., Stephanopoulos, G., 2005. Strain improvement by metabolic engineering: lysine production as a case study for systems biology. Curr Opin Biotechnol. 16, 361-6.

Lan, E. I., Liao, J. C., 2011. Metabolic engineering of cyanobacteria for 1-butanol production from carbon dioxide. Metab Eng. 13, 353-63.

Lennen, R. M., Braden, D. J., West, R. A., Dumesic, J. A., Pfleger, B. F., 2010. A process for microbial hydrocarbon synthesis: Overproduction of fatty acids in Escherichia coli and catalytic conversion to alkanes. Biotechnol Bioeng. 106, 193-202.

Li, M., Zhang, X., Agrawal, A., San, K.-Y., 2012. Effect of acetate formation pathway and long chain fatty acid CoA-ligase on the free fatty acid production in E. coli expressing acy-ACP thioesterase from Ricinus communis. Metabolic Engineering.

Liu, T., Vora, H., Khosla, C., 2010a. Quantitative analysis and engineering of fatty acid biosynthesis in E. coli. Metab Eng. 12, 378-86.

Liu, T., Vora, H., Khosla, C., 2010b. Quantitative analysis and engineering of fatty acid biosynthesis in E. coli. Metabolic Engineering. 12, 378-386.

Lu, X., Vora, H., Khosla, C., 2008. Overproduction of free fatty acids in E. coli: implications for biodiesel production. Metab Eng. 10, 333-9.

Magnuson, K., Jackowski, S., Rock, C. O., Cronan, J. E., Jr., 1993. Regulation of fatty acid biosynthesis in Escherichia coli. Microbiol Rev. 57, 522-42.

Marrakchi, H., Zhang, Y. M., Rock, C. O., 2002. Mechanistic diversity and regulation of Type II fatty acid synthesis. Biochem Soc Trans. 30, 1050-5.

Nawabi, P., Bauer, S., Kyrpides, N., Lykidis, A., 2011. Engineering Escherichia coli for biodiesel production utilizing a bacterial fatty acid methyltransferase. Appl Environ Microbiol. 77, 8052-61.

Nikolau, B. J., Perera, M. A. D. N., Brachova, L., Shanks, B., 2008. Platform biochemicals for a biorenewable chemical industry. Plant Journal. 54, 536-545.

Patil, K. R., Rocha, I., Forster, J., Nielsen, J., 2005. Evolutionary programming as a platform for in silico metabolic engineering. BMC Bioinformatics. 6, 308.

Peebles, C. A., Sander, G. W., Hughes, E. H., Peacock, R., Shanks, J. V., San, K. Y., 2010. The expression of 1-deoxy-D-xylulose synthase and geraniol-10-hydroxylase or anthranilate synthase increases terpenoid indole alkaloid accumulation in Catharanthus roseus hairy roots. Metab Eng. 13, 234-40.

Pharkya, P., Burgard, A. P., Maranas, C. D., 2004. OptStrain: a computational framework for redesign of microbial production systems. Genome Res. 14, 2367-76.

Pharkya, P., Maranas, C. D., 2006. An optimization framework for identifying reaction activation/inhibition or elimination candidates for overproduction in microbial systems. Metab Eng. 8, 1-13.

Poirier, Y., Antonenkov, V. D., Glumoff, T., Hiltunen, J. K., 2006. Peroxisomal beta-oxidation--a metabolic pathway with multiple functions. Biochim Biophys Acta. 1763, 1413-26.

Ranganathan, S., Maranas, C. D., 2010. Microbial 1-butanol production: Identification of non-native production routes and in silico engineering interventions. Biotechnol J. 5, 716-25.

Ranganathan, S., Suthers, P. F., Maranas, C. D., 2010. OptForce: an optimization procedure for identifying all genetic manipulations leading to targeted overproductions. PLoS Comput Biol. 6, e1000744.

San, K.-Y., Li, M., Zhang, X., BACTERIA AND METHOD FOR SYNTHESIZING FATTY ACIDS Vol. WO/2011/116279 , USA, 2011.

Sauer, U., Lasko, D. R., Fiaux, J., Hochuli, M., Glaser, R., Szyperski, T., Wuthrich, K., Bailey, J. E., 1999. Metabolic flux ratio analysis of genetic and environmental modulations of Escherichia coli central carbon metabolism. Journal of Bacteriology. 181, 6679-6688.

Schweizer, E., Hofmann, J., 2004. Microbial type I fatty acid synthases (FAS): major players in a network of cellular FAS systems. Microbiol Mol Biol Rev. 68, 501-17.

Segre, D., Vitkup, D., Church, G. M., 2002. Analysis of optimality in natural and perturbed metabolic networks. Proc Natl Acad Sci U S A. 99, 15112-7.

Shen, C. R., Liao, J. C., 2008. Metabolic engineering of Escherichia coli for 1-butanol and 1-propanol production via the keto-acid pathways. Metab Eng. 10, 312-20.

Siddiquee, K. A., Arauzo-Bravo, M. J., Shimizu, K., 2004. Metabolic flux analysis of pykF gene knockout Escherichia coli based on C-13-labeling experiments together with measurements of enzyme activities and intracellular metabolite concentrations. Applied Microbiology and Biotechnology. 63, 407-417.

Sriram, G., Fulton, D. B., Iyer, V. V., Peterson, J. M., Zhou, R. L., Westgate, M. E., Spalding, M. H., Shanks, J. V., 2004. Quantification of compartmented metabolic fluxes in developing soybean embryos by employing Biosynthetic ally directed fractional C-13 labeling, C-13, H-1 two-dimensional nuclear magnetic resonance, and comprehensive isotopomer balancing. Plant Physiology. 136, 3043-3057.

Steen, E. J., Kang, Y., Bokinsky, G., Hu, Z., Schirmer, A., McClure, A., del Cardayre, S. B., Keasling, J. D., 2010. Microbial production of fatty-acid-derived fuels and chemicals from plant biomass. Nature. 463, 559-U182.

Stephanopoulos, G., 1999. Metabolic fluxes and metabolic engineering. Metab Eng. 1, 1-11.

Stephanopoulos, G., 2007. Challenges in engineering microbes for biofuels production. Science. 315, 801-4.

Subrahmanyam, S., Cronan, J. E., Jr., 1998. Overproduction of a functional fatty acid biosynthetic enzyme blocks fatty acid synthesis in Escherichia coli. J Bacteriol. 180, 4596-602.

Suthers, P. F., Zomorrodi, A., Maranas, C. D., 2009. Genome-scale gene/reaction essentiality and synthetic lethality analysis. Mol Syst Biol. 5, 301.

Szyperski, T., 1995. Biosynthetically directed fractional 13C-labeling of proteinogenic amino acids. An efficient analytical tool to investigate intermediary metabolism. Eur J Biochem. 232, 433 - 448.

Tepper, N., Shlomi, T., 2010. Predicting metabolic engineering knockout strategies for chemical production: accounting for competing pathways. Bioinformatics. 26, 536-43.

Toya, Y., Ishii, N., Nakahigashi, K., Hirasawa, T., Soga, T., Tomita, M., Shimizu, K., 2010. (13)C-Metabolic Flux Analysis for Batch Culture of Escherichia coli and Its pyk and pgi Gene Knockout Mutants Based on Mass Isotopomer Distribution of Intracellular Metabolites. Biotechnology Progress. 26, 975-992.

Vasudevan, P. T., Briggs, M., 2008. Biodiesel production--current state of the art and challenges. J Ind Microbiol Biotechnol. 35, 421-30.

Voeste, T., Buchold, H., 1984. PRODUCTION OF FATTY ALCOHOLS FROM FATTY-ACIDS. Journal of the American Oil Chemists Society. 61, 350-352.

White, S. W., Zheng, J., Zhang, Y. M., Rock, 2005. The structural biology of type II fatty acid biosynthesis. Annu Rev Biochem. 74, 791-831.

Wiechert, W., Mollney, M., Petersen, S., de Graaf, A. A., 2001. A universal framework for 13C metabolic flux analysis. Metab Eng. 3, 265-83.

Xu, P., Ranganathan, S., Fowler, Z. L., Maranas, C. D., Koffas, M. A., 2011a. Genome-scale metabolic network modeling results in minimal interventions that cooperatively force carbon flux towards malonyl-CoA. Metab Eng. 13, 578-87.

Xu, P., Ranganathan, S., Fowler, Z. L., Maranas, C. D., Koffas, M. A., 2011b. Genome-scale metabolic network modeling results in minimal interventions that cooperatively force carbon flux towards malonyl-CoA. Metab Eng.

Youngquist, J. T., Lennen, R. M., Ranatunga, D. R., Bothfeld, W. H., Ii, W. D. M., Pfleger, B. F., 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.

Yu, X., Liu, T., Zhu, F., Khosla, C., 2011. In vitro reconstitution and steady-state analysis of the fatty acid synthase from Escherichia coli. Proc Natl Acad Sci U S A. 108, 18643-8.

Zhang, F., Carothers, J. M., Keasling, J. D., 2012a. Design of a dynamic sensor-regulator system for production of chemicals and fuels derived from fatty acids. Nat Biotech. 30, 354-359.

Zhang, M., Eddy, C., Deanda, K., Finkelstein, M., Picataggio, S., 1995. Metabolic Engineering of a Pentose Metabolism Pathway in Ethanologenic Zymomonas mobilis. Science. 267, 240-3.

Zhang, X., Agrawal, A., San, K.-Y., 2012b. Improving fatty acid production in escherichia coli through the overexpression of malonyl coA-Acyl carrier protein transacylase. Biotechnology Progress. 28, 60-65.

Zhang, X., Li, M., Agrawal, A., San, K.-Y., 2011. Efficient free fatty acid production in Escherichia coli using plant acyl-ACP thioesterases. Metabolic Engineering. 13, 713-722.

List of figures

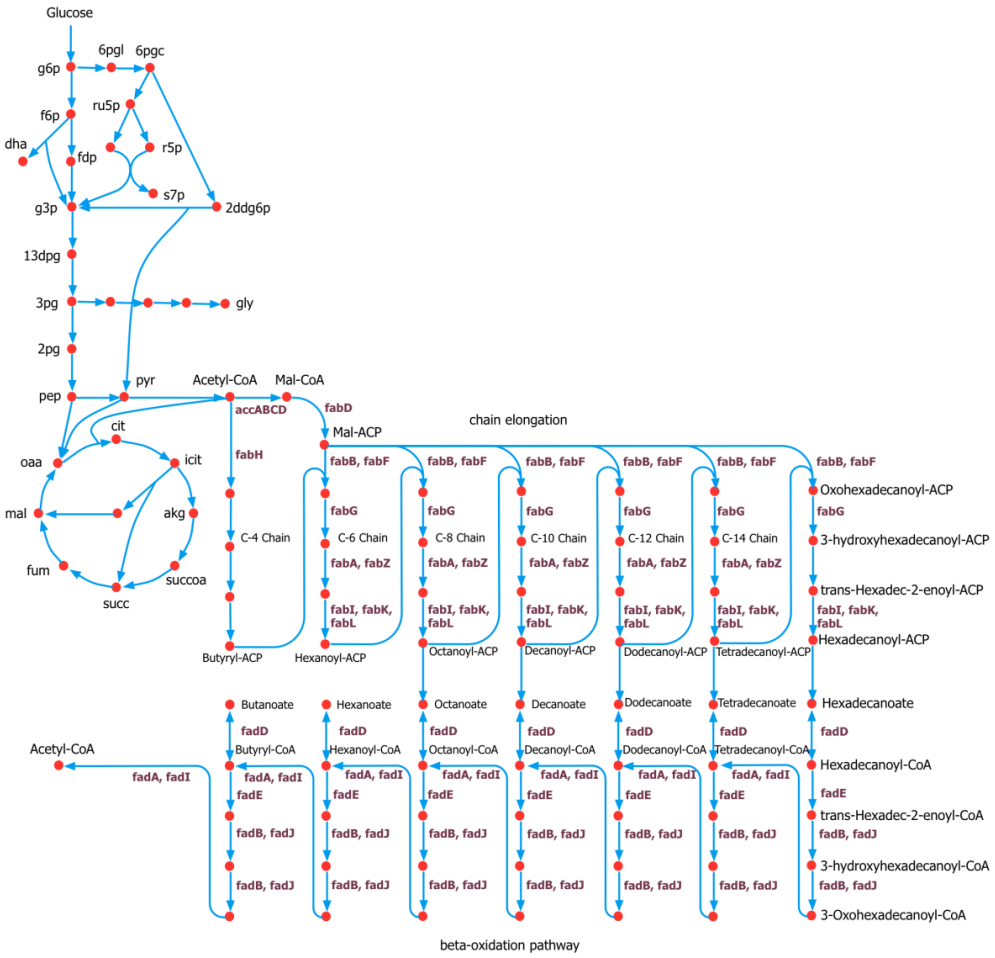


Figure 1. Organization of fatty acid biosynthetic pathways in *E. coli* metabolism according to *i*AF1260 model.

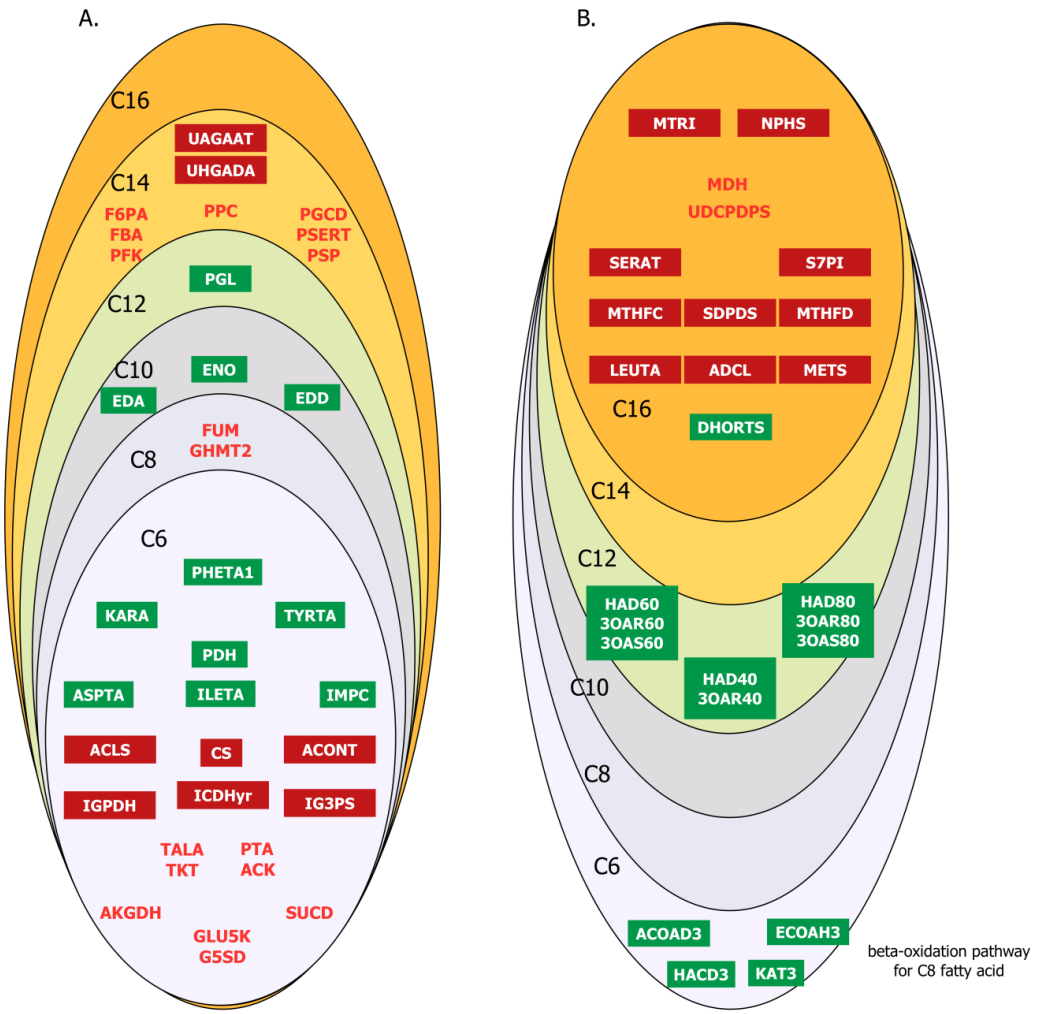


Figure 2. Venn diagram representing the changes that MUST happen in the network when the overproduction objectives are imposed for fatty acids C_6 through C_{16}.

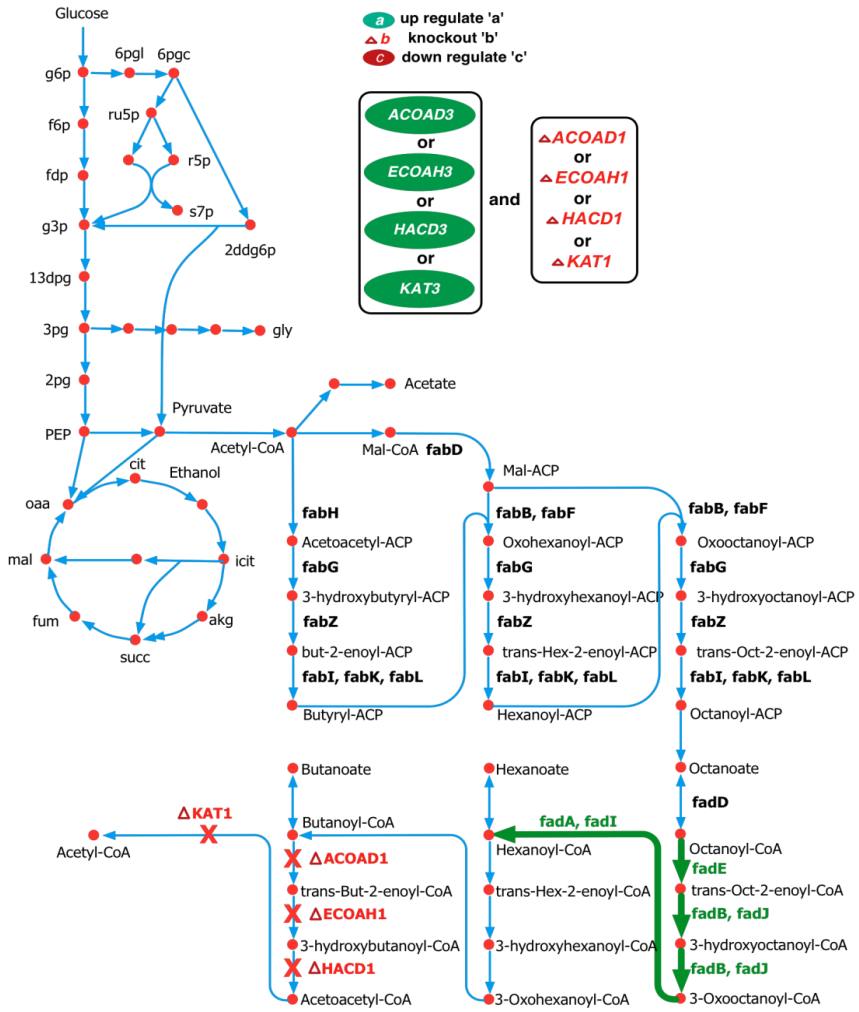


Figure 3. OptForce interventions for the overproduction of hexanoic acid in *E. coli*.

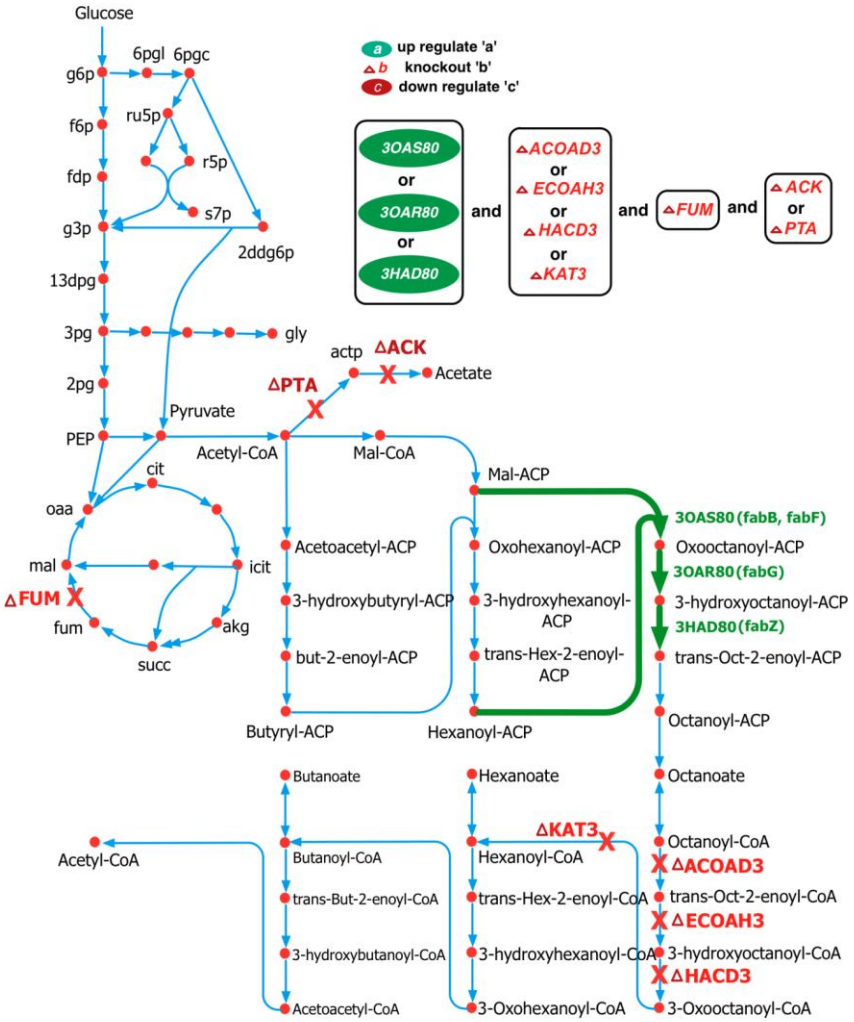


Figure 4. OptForce interventions for the overproduction of octanoic acid in *E. coli*.

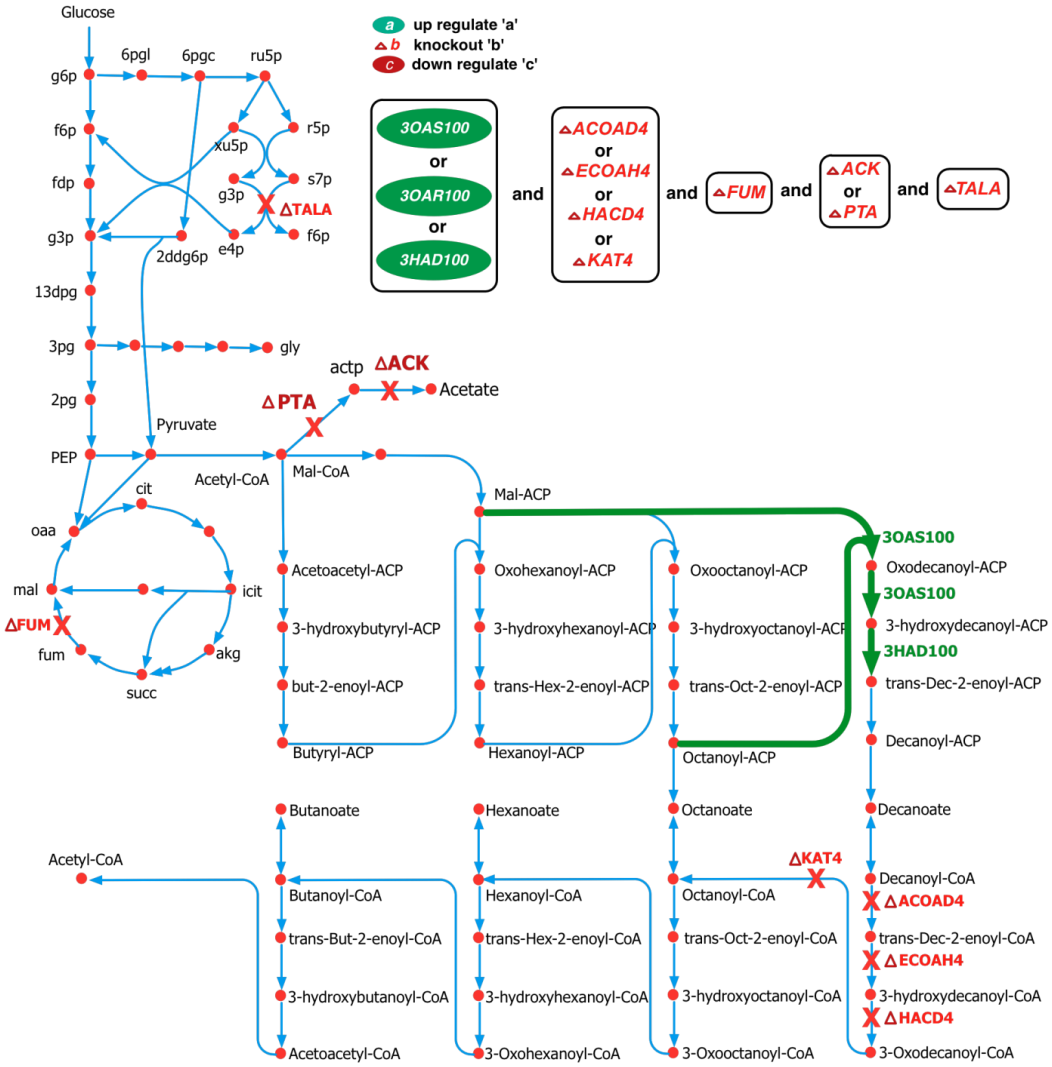


Figure 5. OptForce interventions for the overproduction of decanoic acid in *E. coli*.

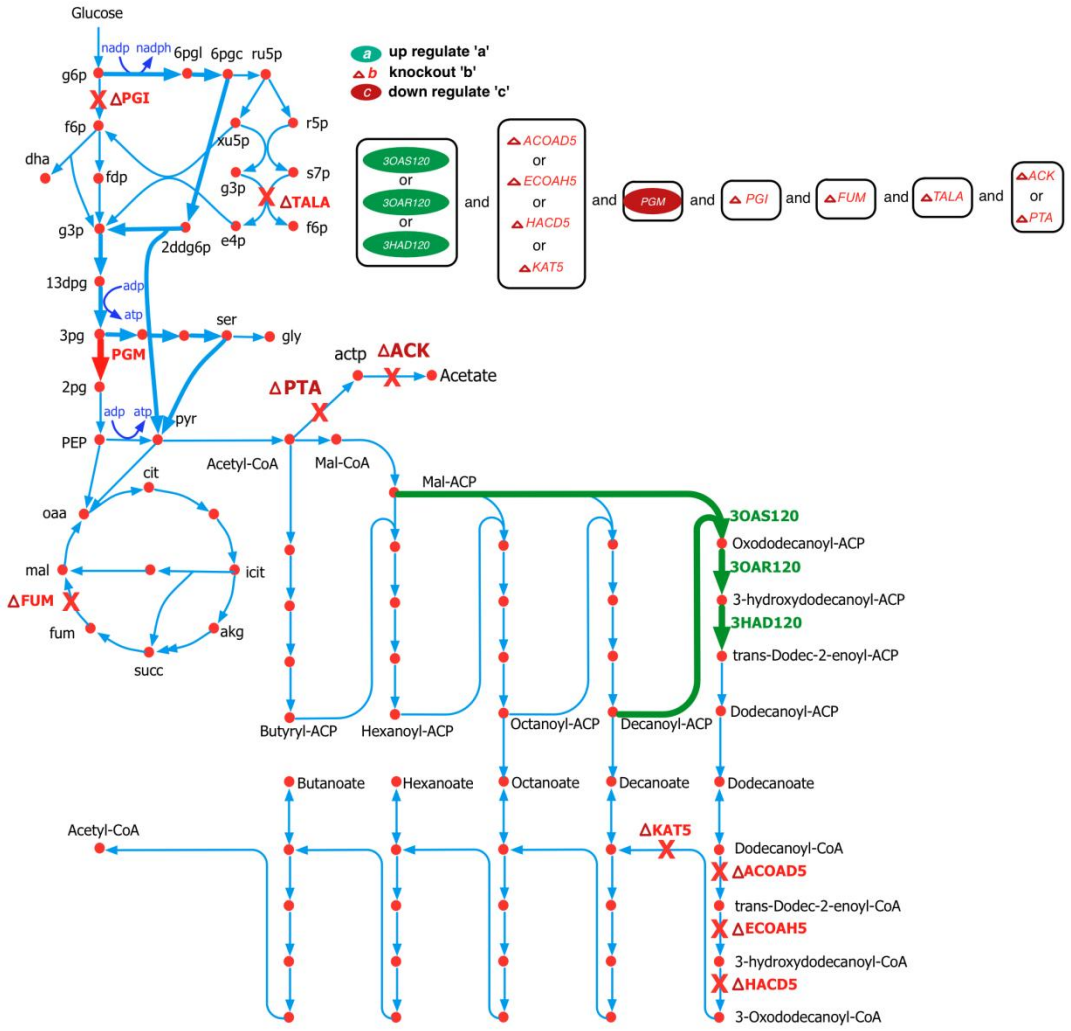


Figure 6. OptForce interventions for the overproduction of dodecanoic acid in *E. coli*.

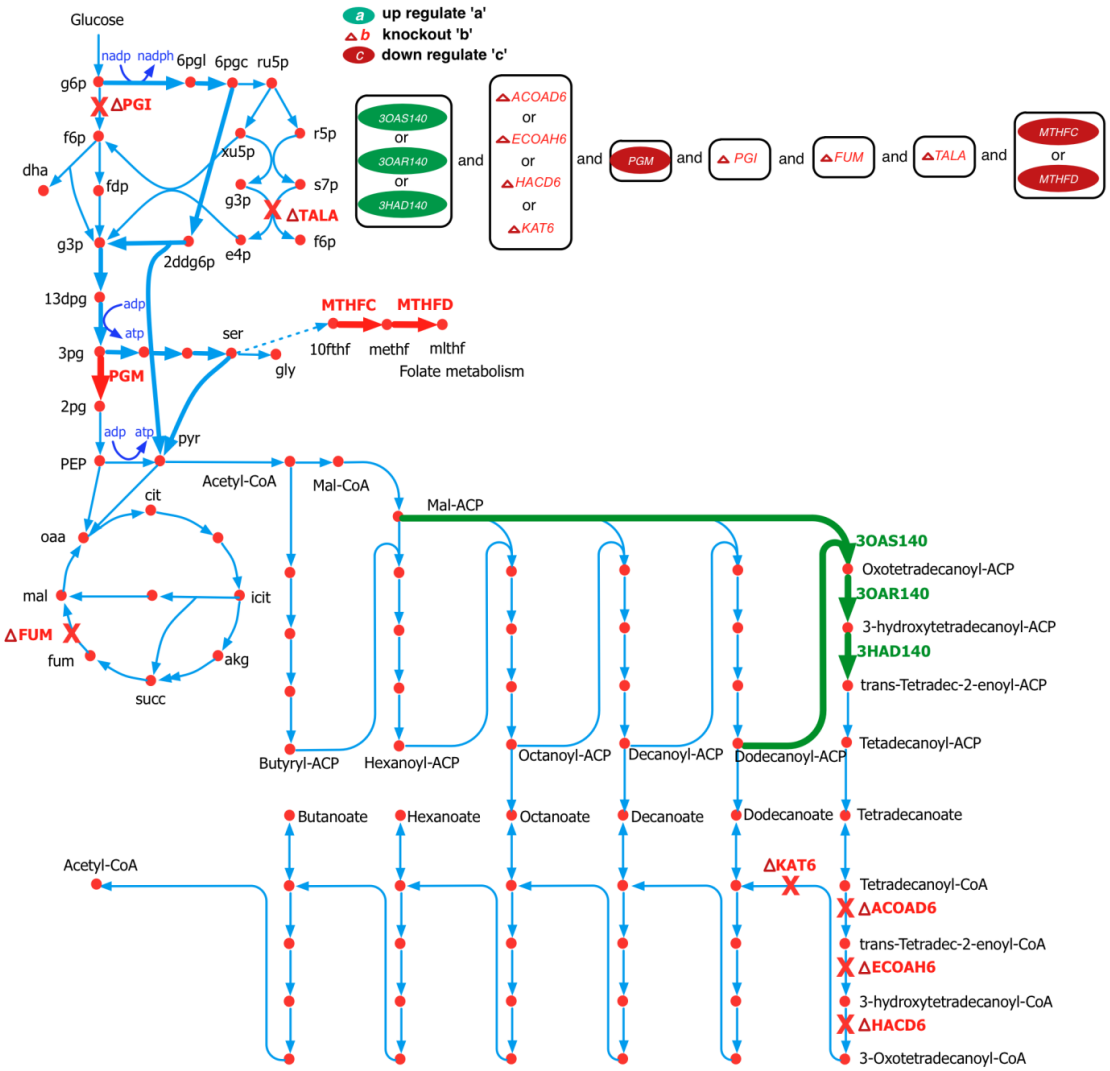


Figure 7. OptForce interventions for the overproduction of tetradecanoic acid in *E. coli*.

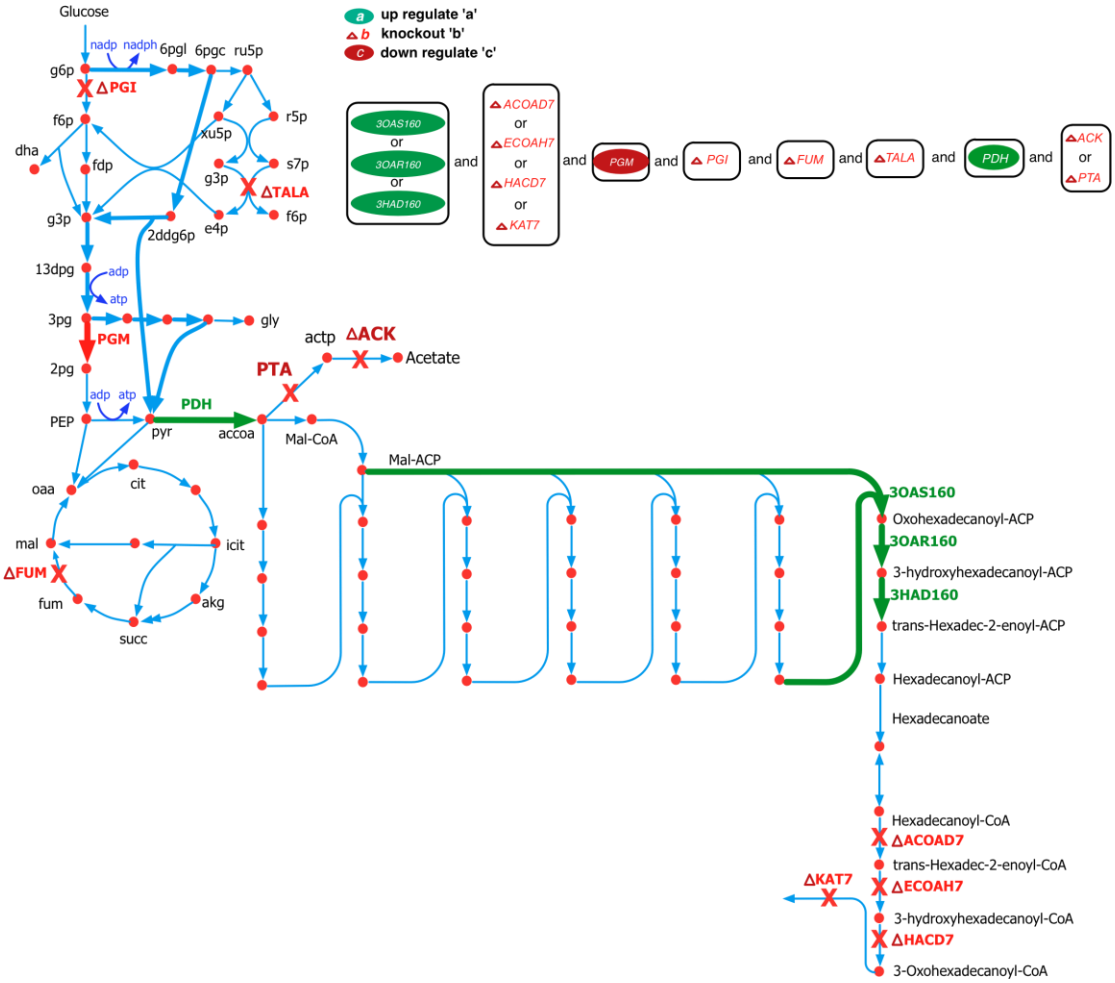


Figure 8. OptForce interventions for the overproduction of palmitic acid in *E. coli*.

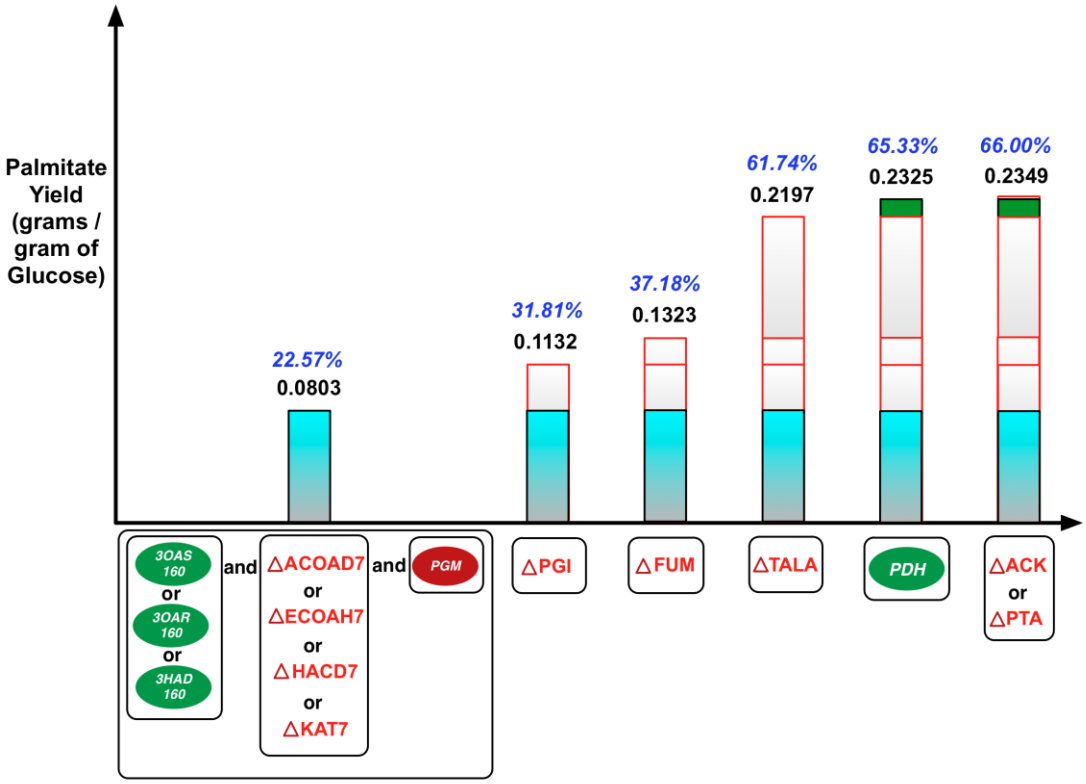


Figure 9. Impact of each genetic intervention predicted by OptForce on the yield of palmitic acid.

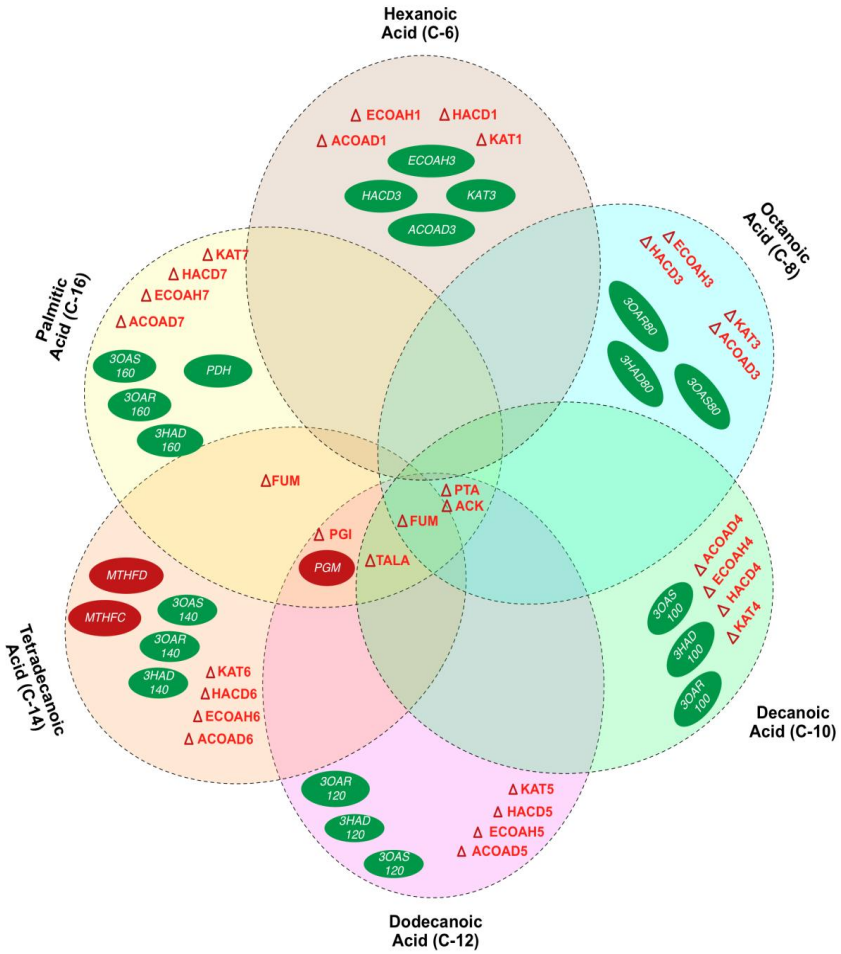


Figure 12. Venn diagram representing the shared genetic interventions predicted by OptForce for fatty acids of chain length C_6 to C_{16}.

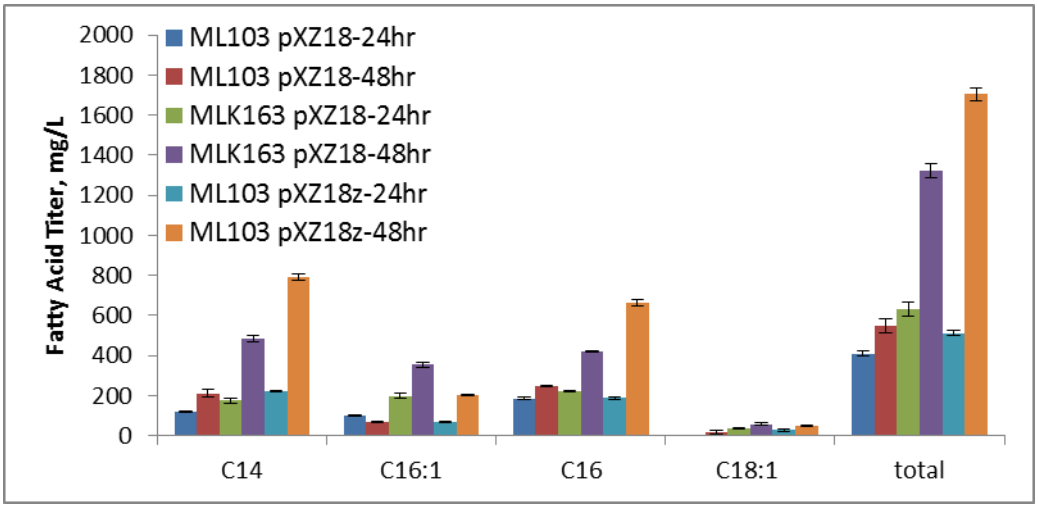


Figure 10. Accumulation of free fatty acids by ML103 pXZ18 (ΔfadD), MLK163 pXZ18 (ΔfadD, ΔsucC) and ML103pXZ18z (ΔfadD, fabZ$^+$) at 24 and 48 hours

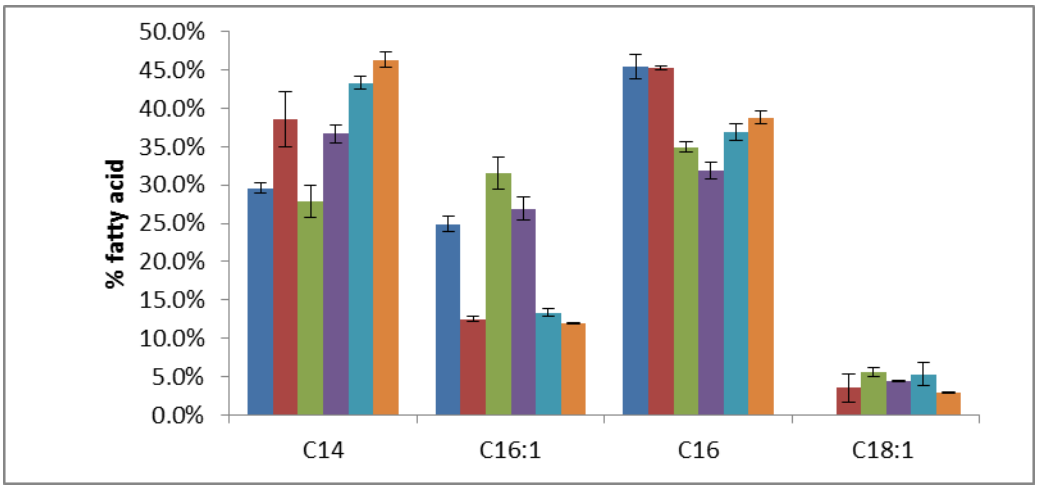


Figure 11. Free fatty acid composition of ML103 pXZ18(ΔfadD), MLK163 pXZ18 (ΔfadD, ΔsucC) and ML103pXZ18z (ΔfadD, fabZ^{+}) at 24 and 48 hours

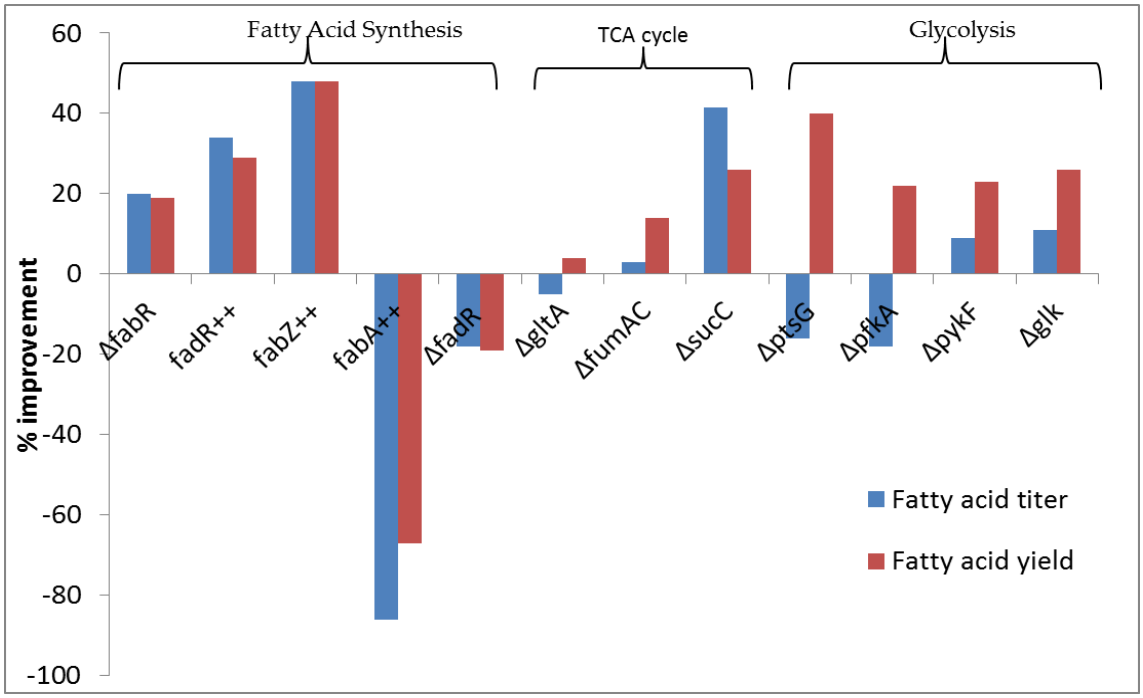


Figure 13. Effect of different genetic modifications on the improvement of fatty acid titer and yield reported by San *et al* (2011)

Tables

Table 1. List of strains and plasmids used

	Relative Genotype	Source or Reference
Strain Name		
MG1655	F^- lamda$^-$ *ilvG*$^-$ *rfb*$^-$ *rph*$^-$	ATCC 47076
ML103	MG1655 (Δ*fadD*)	(Li *et al*., 2012)
MLK163	MG1655 (Δ*fadD, Δ sucC*)	(San *et al*., 2011)
Plasmid name		
pTrc99a	pTrc99a, cloning vector	Amersham Pharmacia
PXZ18	pTrc99a carrying an acyl thioesterase *R. communis*	(Zhang *et al*., 2011)
PXZ18z	pTrc99a carrying an acyl thioesterase *R. communis* and overexpressed *fabZ*	(San *et al*., 2011)

Supporting information

Text S1. Modified OptForce procedure identification of prioritized interventions.

Figure S1. Metabolic flux map of E. coli MG1655 grown in MOPs media with 1% glucose (25% 1-13C glucose, 10% U-13C glucose, 65% unlabeled glucose).

Figure S2. Metabolic flux map of E. coli ML103 (*fadD*) strain grown in minimal M9 media with 1% glucose (20% U-13C glucose and 80% 1-13C glucose).

Figure S3. Free fatty acid yield of strains in minimal M9 medium

Figure S4. Glucose consumption of different strains in minimal M9 medium

Figure S5. Acetate accumulation of different strains in minimal M9 medium

Table S1. Metabolic pathways represented by stoichiometry, atom transition, and involved genes

Table S2. Comparison of flux values between E coli MG1655 and ML103 strains

Table S3. Fermentation data of MG 1665 and ML103 in 13-C flux experiments

Table S4. Measured and simulated isotopomer distributions of proteinogenic amino acids from 2D-NMR

Table S1. Metabolic pathways represented by stoichiometry, atom transition, and involved genes from Ecocyc database (http://ecocyc.org).

Reaction name	Stoichiometry chemistry	Atom transition	Gene
Glycolysis pathway			
pts	Glu+PEP → G6P+PYR	abcdef+ABC → abcdef+ABC	ptsG,manZ,ptsH,ptsP
pgi	G6P → F6P	abcdef → abcdef	Pgi
fbp	F6P → FBP	abcdef → abcdef	pfkA,pfkB
fba	FBP → T3P+T3P	abcdef → cba+def	fbaA,fbaB
tpi	T3P → G3P	abc → abc	tpiA
pgk	G3P → 3PG	abc → abc	pgk
eno	3PG → PEP	abc → abc	eno
pyk	PEP → PYR	abc → abc	pykF/pykA
Entner Doudoroff pathway			
eda	6PG → PYR+T3P	abcdef → abc+def	eda
Pentose phosphate pathway			
zwf	G6P → 6PG	abcdef → abcdef	zwf
rpi	6PG → R5P+CO2	abcdef → bcdef+a	gnd,rpe,rpiA,rpiB
tkt	R5P+R5P → S7P+T3P	abcde+ABCDE → abABCDE+cde	tktA,tktB
tktAB	R5P+E4P → F6P+T3P	abcde+ABCD → abABCD+cde	tktA,tktB
talf	S7P+T3P → F6P+E4P	abcdefg+ABC → abcABC+defg	talA,talB
TCA cycle			
ace	PYR → ACCOA+CO2	abc → bc+a	lpd,aceF,aceE
icd	ACCOA+OAA → AKG+CO2	ab+ABCD → DCBba+A	icdB,icd
suc	AKG → SUCC+CO2	abcde → bcde+a	lpd,sucA,sucB
frd	SUCC → MAL	abcd → abcd	sdhA,sdhB,sdhC,sdhD fumA,fumB,fumC
mdh	MAL → OAA	abcd → abcd	mqo,mdh

Table S1 continued

Acetate production			
ackf	ACCOA → AC	ab → ab	pta,ackA,acs
Anaplerotic pathway			
ppc	PEP+CO2 → OAA	abc+A → abcA	ppc
ana	Mal → PYR+CO2	abcd → abc+d	maeA,maeB
aceA	ACCOA+OAA → GOx+SUCC	ab+ABCD → DC+ABba	aceA
aceB	ACCOA+GOX → MAL	ab+AB → Abba	aceB
C1 metabolism			
ser	3PG → Ser	abc → abc	serA,serC,serB
gly	Ser → Gly+C1	abc → ab+c	glyA
Amino acid biosynthesis and metabolic pathway			
SdaRf	Ser → PYR	abc → abc	sdaA,sdaB,tdcB,tdcG
thr	OAA → Thr	abcd → abcd	thrC
thrgly	Thr → Gly+ACCOA	abcd → ab+cd	ItaE
GluDy	AKG → Glut	abcde → abcde	gdhA
GlnDy	Glut → Gln	abcde → abcde	glnA
GLUSy	AKG+Gln → Glut+Glut	abcde+ABCDE → abcde+ABCDE	gltB,gltD
ALATA	AKG+Ala → Glut+PYR	abcde+ABC → abcde+ABC	alaA,alaB,alaC
ASPTA	AKG+Asp → Glut+OAA	abcde+ABC → abcde+ABC	aspC
Ala	PYR → Ala	abc → abc	dadA
Asp	OAA → Asp	abcd → abcd	aspC
Arg	Glut → Arg	abcde → abcde	argH
Pro	Glut → Pro	abcde → abcde	proC

Table S2. Comparison of flux values between E coli MG1655 and ML103 strains. The flux values are based on 100 mol/g/hr glucose uptake rate.

reaction name	stoichiometry chemistry	E coli MG1655		E coli ML103	
		Average	SD	Average	SD
Glycolysis pathways					
pts	Glu+PEP → G6P+PYR	100.00	0.00	100.00	0.00
pgi	G6P → F6P	88.20	0.54	87.70	0.90
	reversibility	0.99	0.00	0.06	0.08
fba	F6P → T3P+T3P	89.24	0.74	87.75	0.73
tpi	T3P → G3P	179.82	0.84	177.71	1.10
	reversibility	0.62	0.30	0.15	0.28
pgk	G3P → 3PG	179.82	0.84	177.71	1.10
eno	3PG → PEP	173.04	0.98	167.97	1.57
pyk	PEP → PYR	50.33	1.32	44.72	3.08
Entner Doudoroff pathway					
eda	6PG → PYR+T3P	2.46	1.05	3.74	0.93
Pentose phosphate pathway					
zwf	G6P → 6PG	10.64	0.50	10.98	0.92
rpi	6PG → R5P+CO2	8.19	0.71	7.24	1.49
tkt	R5P+R5P → S7P+T3P	1.81	0.16	1.22	0.50
	reversibility	0.79	0.02	0.90	0.02
tktAB	R5P+E4P → F6P+T3P	-0.40	0.18	-0.73	0.58
	reversibility	0.57	0.24	0.69	0.17
talf	S7P+T3P → F6P+E4P	1.81	0.16	1.22	0.50
	reversibility	0.99	0.00	0.09	0.11
TCA cycle					
ace	PYR → ACCOA+CO2	139.98	1.65	143.07	1.88
icd	ACCOA+OAA → AKG+CO2	50.50	2.51	49.16	3.75
suc	AKG → SUCC+CO2	44.13	2.46	42.21	3.76
succ	FUM → MAL	45.71	2.01	46.07	3.34
	reversibility	0.80	0.08	0.50	0.28
	scrambling	0.53	0.25	0.67	0.22
mdh	MAL → OAA	47.15	2.05	44.16	4.09
	reversibility	0.98	0.01	0.64	0.15
Acetate production					
ackf	ACCOA → AC	65.29	0.63	67.66	1.22
ac	AC → ACout	65.29	0.63	67.66	1.22
Anaplerotic pathway					
ppc	PEP+CO2 → OAA	19.68	1.18	20.21	2.77
	reversibility	0.10	0.06	0.39	0.11
ana	Mal → PYR+CO2	0.14	0.33	5.77	2.77

Table S2 continued

aceA	ACCOA+OAA → GOx+SUCC	1.58	0.92	3.86	0.83
aceB	ACCOA+GOX → MAL	1.58	0.92	3.86	0.83
C1 metabolism					
ser	3PG → Ser	3.74	0.58	6.26	1.11
gly	Ser → Gly+C1	0.30	0.01	0.28	0.03
	reversibility	0.94	0.01	0.98	0.00
Amino acid biosynthesis pathway					
SdaRf	Ser → PYR	1.43	0.54	3.47	1.11
thr	OAA → Thr	4.02	0.14	3.49	0.17
thrgly	Thr → Gly+ACCOA	2.64	0.01	2.28	0.04
GluDy	AKG → Glut	4.09	1.18	0.47	1.72
	reversibility	0.66	0.29	0.00	0.05
GlnDy	Glut → Gln	3.76	0.50	5.03	0.91
	reversibility	0.95	0.02	0.99	0.00
GLUSy	AKG+Gln → Glut+Glut	2.19	0.49	3.39	0.91
ALATA	AKG+Ala → Glut+PYR	0.01	0.04	2.37	1.00
	reversibility	0.24	0.33	0.33	0.20
ASPTA	AKG+Asp → Glut+OAA	1.30	0.91	0.45	0.78
	reversibility	0.55	0.26	0.05	0.15
Ala	PYR → Ala	2.71	0.31	5.53	1.02
	reversibility	0.34	0.29	0.56	0.32
Asp	OAA → Asp	2.76	0.90	1.69	0.81
	reversibility	0.45	0.30	0.50	0.27
Arg	Glut → Arg	1.54	0.03	1.84	0.02
	reversibility	0.44	0.33	0.48	0.26
Pro	Glut → Pro	1.16	0.02	1.37	0.03
	reversibility	0.40	0.31	0.46	0.29
Transport pathway					
co2	CO2 → CO2out	223.26	5.22	227.23	8.13
Biomass synthesis pathway					
G6pb	G6P → biomass	1.16	0.11	1.32	0.09
R5pb	R5P → biomass	4.97	0.51	5.52	0.40
E4pb	E4P → biomass	2.21	0.13	1.96	0.23
T3pb	T3P → biomass	0.71	0.08	0.80	0.10
PEPb	PEP → biomass	3.03	0.25	3.04	0.48
PYRb	PYR → biomass	11.69	0.78	11.47	1.23
ACCOAb	ACCOA → biomass	23.66	0.97	20.81	2.18
AKGb	AKG → biomass	0.49	0.00	0.27	0.02
OAAb	OAA → biomass	9.27	0.62	6.64	0.76
3PGb	3PG → biomass	3.05	0.34	3.48	0.07
F6Pb	F6P → biomass	0.37	0.02	0.44	0.05

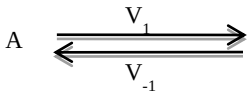
Table S2 continued

C1b	C1 → biomass	0.30	0.01	0.28	0.03
serb	Ser → biomass	2.01	0.16	2.51	0.00
glyb	Gly → biomass	2.94	0.00	2.55	0.02
thrb	Thr → biomass	1.38	0.13	1.21	0.15
glutb	Glut → biomass	1.61	0.04	1.82	0.00
glunb	Glu → biomass	1.57	0.11	1.64	0.00
alab	Ala → biomass	2.69	0.30	3.16	0.14
aspb	Asp → biomass	1.46	0.07	1.23	0.18
argb	Arg → biomass	1.54	0.03	1.84	0.02
prob	Pro → biomass	1.16	0.02	1.37	0.03

(1) Abbreviation: Glu, Glucose; G6P, Glucose-6-phosphate; F6P, Fructose-6-phosphate; T3P, Triose-3-phosphate; 3PG, 3-phosphoglycerate; PEP, Phosphoenolpyruvate; PYR, Pyruvate; ACCOA, Acetate-CoA; AKG, -Ketoglutarate; SUCC, Succinate; MAL, Malate; OAA, Oxaloacetate; 6PG, 6-P-gluconate; R5P, Ribose-5-phosphate; S7P, Sedoheptulose-7-phosphate; E4P, Erythronate-4-phosphate; GOX, Glyoxylate; AC, Acetate; CO2, carbon dioxide; Ser, Serine; Gly, Glycine; Arg, Arginine; Pro, Proline; Asp, Aspartate; Thr, Threonine; Glut, Glutamate; Gln, Glutamine.

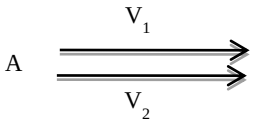
(2) Reversibility (R) and the extent of scrambling (S) are defined as follows.

Reversible Reaction
V_1: forward reaction
V_{-1}:reverse reaction



$$V_{net} = V_1 - V_{-1}, R = \frac{\min(V_1, V_{-1})}{\max(V_1, V_{-1})}$$

Scrambling Reaction
V_1 and V_2 has exactly the same stoichiometry, but different carbon skeletal rearrangement



$$V_{net} = V_1 + V_2, S = \frac{V_2}{V_{net}}$$

Table S3. Fermentation data: growth rate, glucose uptake rate, and acetate production rate of MG 1665 and ML103 in 13-C flux experiments. They were grown in MOPs and M9, respectively.

	E coli MG1655 in MOPs	E coli ML103 in M9
Growth rate (hr^{-1})	0.772	0.389
Glucose uptake rate (mmol/g/hr)	13.402	7.108
Acetate production rate (mmole/g/hr)	6.794	4.512
Acetate yield (mol/mol)	0.507	0.650

Table S4. Measured and simulated isotopomer distributions of proteinogenic amino acids from 2D-NMR for E. coli MG 1655 grown in MOPS and ML103 strain grown in M9 media.

Cross peak	MG 1655		ML103			
(multiplet)	Experimental	Simulated	Experimental	Simulated	Precursor	Isotopomer
Ala α (s)	0.174	0.121	0.063	0.074	Pyr	[1**2**3]
Ala α (d1)	0.040	0.056	0.068	0.092	Pyr	[**12**3]
Ala α (d2)	0.115	0.028	0.051	0.076	Pyr	[1**23**]
Ala α (dd)	0.671	0.795	0.818	0.759	Pyr	[**123**]
Ala β (s)	0.496	0.562	0.629	0.647	Pyr	[x2**3**]
Ala β (d)	0.504	0.438	0.371	0.353	Pyr	[x**23**]
Arg β (s)	0.482	0.442	0.225	0.205	AKG	[x2**3**4x]
Arg β (d)	0.461	0.466	0.487	0.496	AKG	[x**23**4x] + [x2**34**x]
Arg β (t)	0.057	0.092	0.288	0.299	AKG	[x**234**x]
Arg δ (s)	0.143	0.185	0.129	0.181	AKG	[xxx4**5**]
Arg δ (d)	0.857	0.815	0.871	0.819	AKG	[xxx**45**]
Asp α (s)	0.341	0.379	0.276	0.255	OAA	[1**2**3x]
Asp α (d1)	0.204	0.180	0.192	0.194	OAA	[**12**3x]
Asp α (d2)	0.178	0.160	0.196	0.228	OAA	[1**23**x]
Asp α (dd)	0.278	0.281	0.336	0.323	OAA	[**123**x]
Asp β (s)	0.414	0.446	0.330	0.344	OAA	[x2**3**4]
Asp β (d1)	0.258	0.208	0.231	0.253	OAA	[x**23**4]
Asp β (d2)	0.230	0.195	0.214	0.212	OAA	[x2**34**]
Asp β (dd)	0.099	0.152	0.225	0.190	OAA	[x**234**]
Glu β (s)	0.450	0.442	0.172	0.205	AKG	[x2**3**4x]
Glu β (d)	0.496	0.466	0.486	0.496	AKG	[x**23**4x] + [x2**34**x]
Glu β (t)	0.054	0.092	0.315	0.299	AKG	[x**234**x]
Gly α (s)	0.245	0.274	0.153	0.243	Gly	[1**2**x]
Gly α (d)	0.755	0.726	0.847	0.757	Gly	[**12**x]
His δ2 (s)	0.522	0.513	0.461	0.498	R5P	[**1**2xxx]
His δ2 (d)	0.478	0.487	0.539	0.503	R5P	[**12**xxx]
Ile α (s)	0.483	0.478	0.287	0.372	OAA/Pyr	[1**2**xx] {x2x]

Table S4 continued

Ile α (d1)	0.383	0.409	0.353	0.398	OAA/Pyr	[**12**xx] {x2x]
Ile α (d2)	0.096	0.061	0.198	0.111	OAA/Pyr	[12xx] {x**2**x]
Ile α (dd)	0.038	0.052	0.162	0.119	OAA/Pyr	[**12**xx] {x**2**x]
Ile γ1(s)	0.577	0.580	0.484	0.460	Pyr/OAA	[x2x] {xx**3**4]
Ile γ1(d)	0.423	0.381	0.416	0.448	Pyr/OAA	[x**2**x] {xx**3**4] +[x2x] {xx**34**]
Ile γ1(t)	0.000	0.039	0.100	0.093	Pyr/OAA	[x**2**x] {xx**34**]
Ile δ (s)	0.570	0.582	0.605	0.647	OAA	[xx3**4**]
Ile δ (d)	0.430	0.418	0.395	0.353	OAA	[xx**34**]
Ile γ2 (s)	0.527	0.562	0.449	0.428	Pyr	[x2**3**]
Ile γ2 (d)	0.473	0.438	0.551	0.572	Pyr	[x**23**]
Leu α (s)	0.433	0.499	0.415	0.496	ACoA/Pyr	[1**2**] {x2x]
Leu α (d1)	0.396	0.389	0.270	0.274	ACoA/Pyr	[**12**] {x2x]
Leu α (d2)	0.134	0.063	0.199	0.149	ACoA/Pyr	[1**2**] {x**2**x]
Leu α (dd)	0.037	0.049	0.116	0.082	ACoA/Pyr	[**12**] {x**2**x]
Leu (s)	0.704	0.703	0.383	0.352	ACoA/Pyr	[x2] {x**2**x].[x2x]
Leu (d)	0.296	0.274	0.465	0.523	ACoA/Pyr	[x**2**] {x**2**x].[x2x]+[x2] {x**2**x].[x**2**x]
Leu (t)	0.000	0.005	0.152	0.125	ACoA/Pyr	[x**2**] {x**2**x].[x**2**x]
Leu δ1 (s)	0.638	0.643	0.667	0.647	Pyr	[x2**3**]
Leu δ1 (d)	0.362	0.367	0.333	0.353	Pyr	[x**23**]
Leu δ2 (s)	0.854	0.859	0.780	0.770	Pyr	[x2x] {xx**3**]
Leu δ2 (d)	0.146	0.151	0.220	0.231	Pyr	[x**2**x] {xx**3**]
Lys β (s)	0.393	0.495	0.401	0.372	OAA/Pyr	½{[x2**3**4] + [x2**3**] {xxx4]}
Lys β (d)	0.468	0.434	0.426	0.466	OAA/Pyr	½{[x**23**4] + [x2**34**] + [x**23**] {xxx4] + [x2**3**] {xxx**4**]}
Lys β (t)	0.129	0.071	0.172	0.162	OAA/Pyr	½{[x**234**] + [x**23**] {xxx**4**]}
Lys γ (s)	0.344	0.459	0.174	0.195	OAA/Pyr	[xx3**4**] {xx3]
Lys γ (d)	0.503	0.453	0.520	0.494	OAA/Pyr	[xx**34**] {xx3] + [xx3**4**] {xx**3**]
Lys γ (t)	0.153	0.088	0.305	0.312	OAA/Pyr	[xx**34**] {xx**3**]

Table S4 continued

Lys δ (s)	0.425	0.495	0.344	0.372	OAA/Pyr	½{[x2**3**4] + [x2**3**] {xxx4]}
Lys δ (d)	0.492	0.434	0.446	0.466	OAA/Pyr	½{[x**23**4] + [x2**34**] + [x**23**] {xxx4] + [x2**3**] {xxx**4**]}
Lys δ (t)	0.083	0.071	0.211	0.162	OAA/Pyr	½{[x**234**] + [x**23**] {xxx**4**]}
Phe α (s)	0.120	0.111	0.083	0.061	PEP	[1**2**3]
Phe α (d1)	0.011	0.034	0.057	0.073	PEP	[**12**3]
Phe α (d2)	0.039	0.028	0.075	0.065	PEP	[1**23**]
Phe α (dd)	0.829	0.827	0.784	0.800	PEP	[**123**]
Pro (s)	0.385	0.442	0.146	0.205	AKG	[x2**3**4x]
Pro (d)	0.469	0.466	0.514	0.496	AKG	[x**23**4x] + [x2**34**x]
Pro (t)	0.146	0.092	0.339	0.299	AKG	[x**234**x]
Pro γ(s)	0.370	0.519	0.349	0.363	AKG	[xx3**4**5]
Pro γ(d)	0.503	0.448	0.486	0.482	AKG	[xx**34**5] + [x23**45**]
Pro γ(t)	0.127	0.034	0.165	0.155	AKG	[xx**345**]
Pro δ (s)	0.137	0.185	0.150	0.181	AKG	[xxx4**5**]
Pro δ (d)	0.863	0.815	0.850	0.819	AKG	[xxx**45**]
Ser α (s)	0.113	0.139	0.033	0.078	Ser	[1**2**3]
Ser α (d1)	0.194	0.162	0.120	0.148	Ser	[**12**3]
Ser α (d2)	0.033	0.035	0.075	0.086	Ser	[1**23**]
Ser α (dd)	0.661	0.665	0.772	0.689	Ser	[**123**]
Ser β (s)	0.613	0.641	0.689	0.667	Ser	[x2**3**]
Ser β (d)	0.387	0.359	0.311	0.333	Ser	[x**23**]
Thr α (s)	0.356	0.357	0.260	0.255	OAA	[1**2**3x]
Thr α (d1)	0.252	0.256	0.165	0.194	OAA	[**12**3x]
Thr α (d2)	0.172	0.139	0.233	0.228	OAA	[1**23**x]
Thr α (dd)	0.220	0.249	0.342	0.323	OAA	[**123**x]
Thr γ2 (s)	0.520	0.576	0.418	0.428	OAA	[xx3**4**]
Thr γ2 (d)	0.480	0.424	0.582	0.572	OAA	[xx**34**]
Tyr α (s)	0.029	0.111	0.050	0.061	PEP	[1**2**3]
Tyr α (d1)	0.162	0.034	0.054	0.073	PEP	[**12**3]

Table S4 continued

Tyr α (d2)	0.045	0.028	0.121	0.065	PEP	[1**23**]
Tyr α (dd)	0.805	0.827	0.775	0.800	PEP	[**123**]
Val α (s)	0.112	0.132	0.103	0.115	Pyr	[1**2**x] {x2x]
Val α (d1)	0.650	0.755	0.552	0.655	Pyr	[**12**x] {x2x]
Val α (d2)	0.225	0.017	0.076	0.034	Pyr	[1**2**x] {x**2**x]
Val α (dd)	0.012	0.096	0.270	0.196	Pyr	[**12**x] {x**2**x]
Val γ1 (s)	0.530	0.562	0.630	0.647	Pyr	[x2**3**]
Val γ1 (d)	0.471	0.438	0.370	0.353	Pyr	[x**23**]
Val γ2 (s)	0.880	0.888	0.750	0.770	Pyr	[x2x] {xx**3**]
Val γ2 (d)	0.120	0.112	0.250	0.231	Pyr	[x**2**x] {xx**3**]

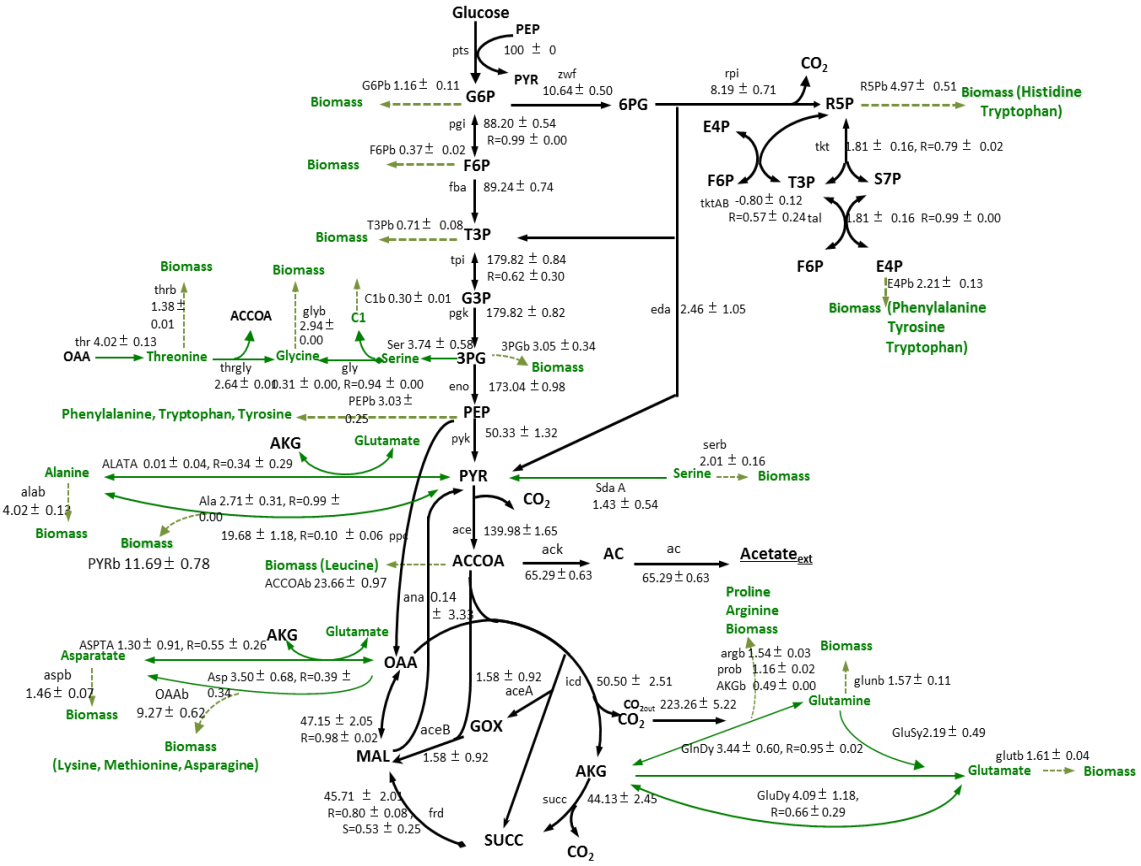


Figure S1. Metabolic flux map of E. coli MG1655 grown in MOPs media with 1% glucose (25% 1-13C glucose, 10% U-13C glucose, 65% unlabeled glucose) at 37 °C.

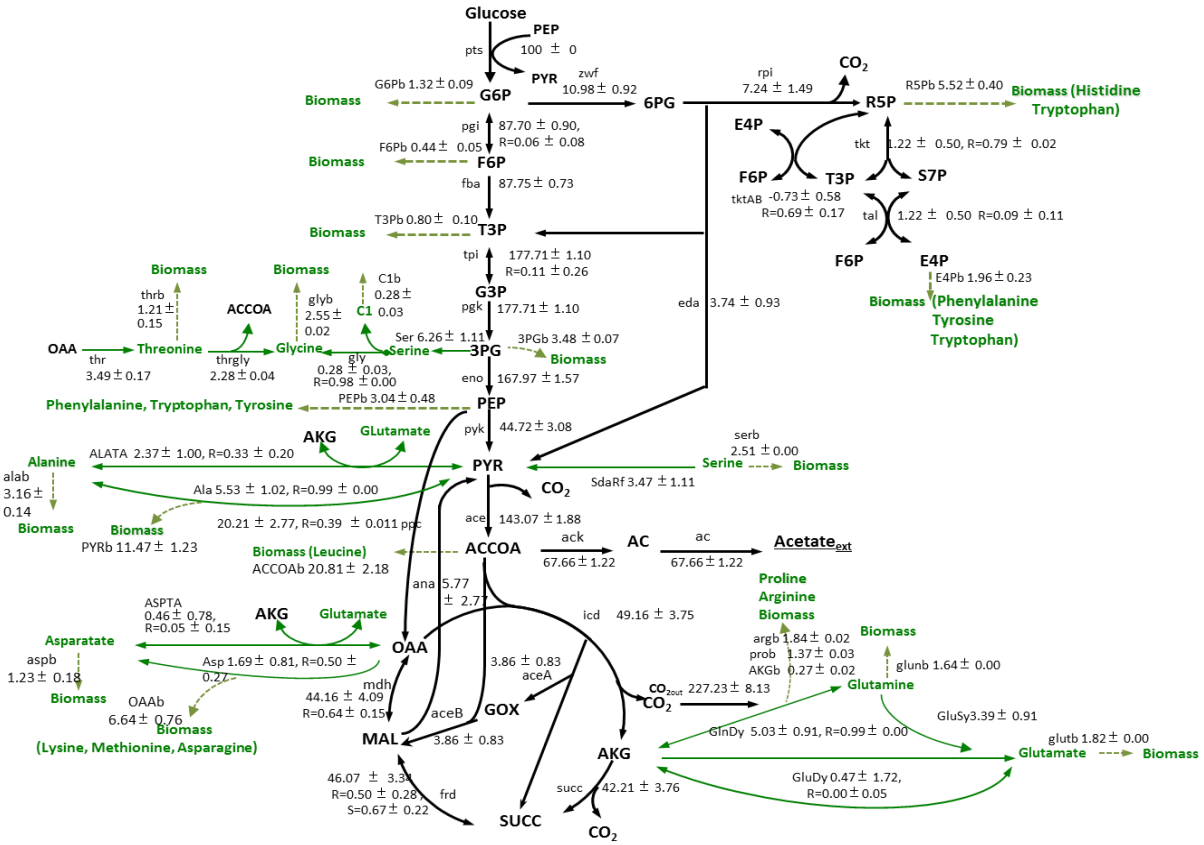


Figure S2. Metabolic flux map of E. coli ML103 (*fadD*) strain grown in minimal M9 media with 1% glucose (20% U-13C glucose and 80% 1-13C glucose) at 30°C.

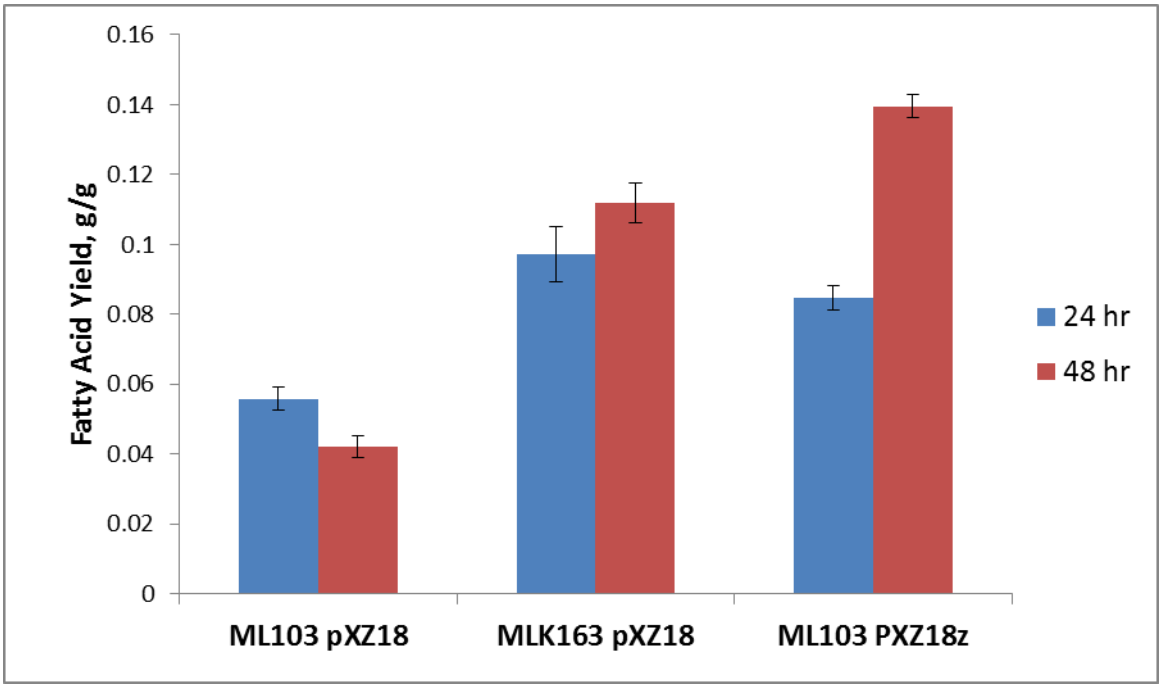


Figure S3. Free fatty acid yield per gram of glucose consumed of ML103 pXZ18(Δ*fadD*), MLK163 pXZ18 (Δ*fadD*,Δ*sucC*) and ML103pXZ18z (Δ*fadD*,*fabZ*$^{++}$) at 24 and 48 hour. The data shown are means ± standard deviation of at least 3 replicates.

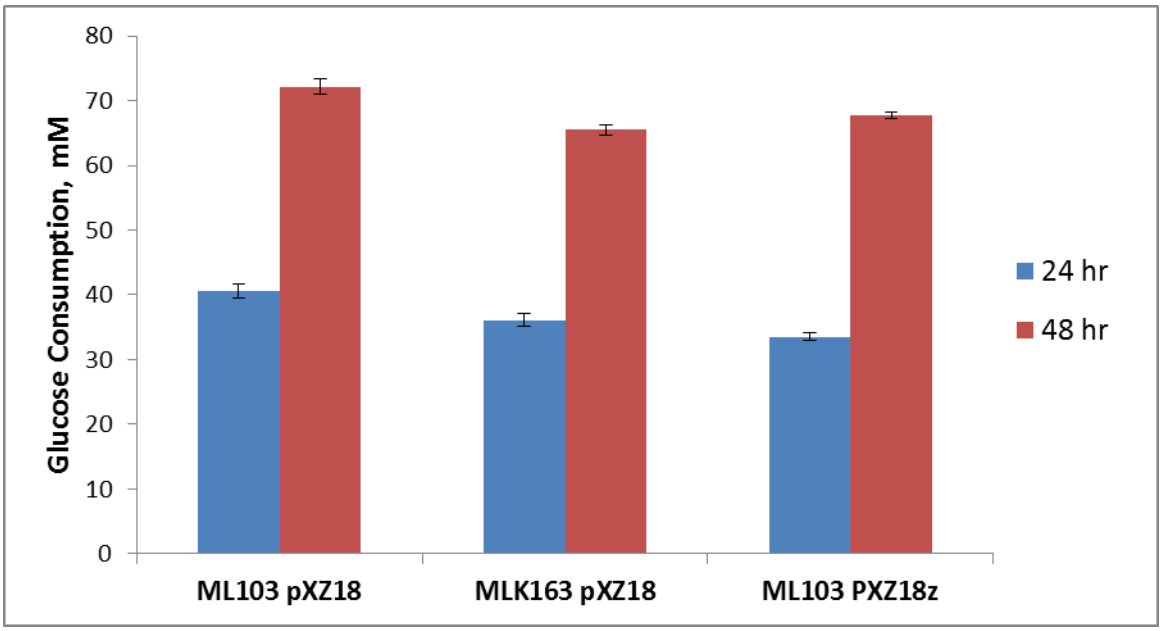


Figure S4 Glucose utilization of ML103 pXZ18(Δ*fadD*), MLK163 pXZ18 (Δ*fadD*,Δ*sucC*) and ML103pXZ18z (Δ*fadD*,*fabZ*$^{++}$) at 24 and 48 hour. The data shown are means ± standard deviation of at least 3 replicates.

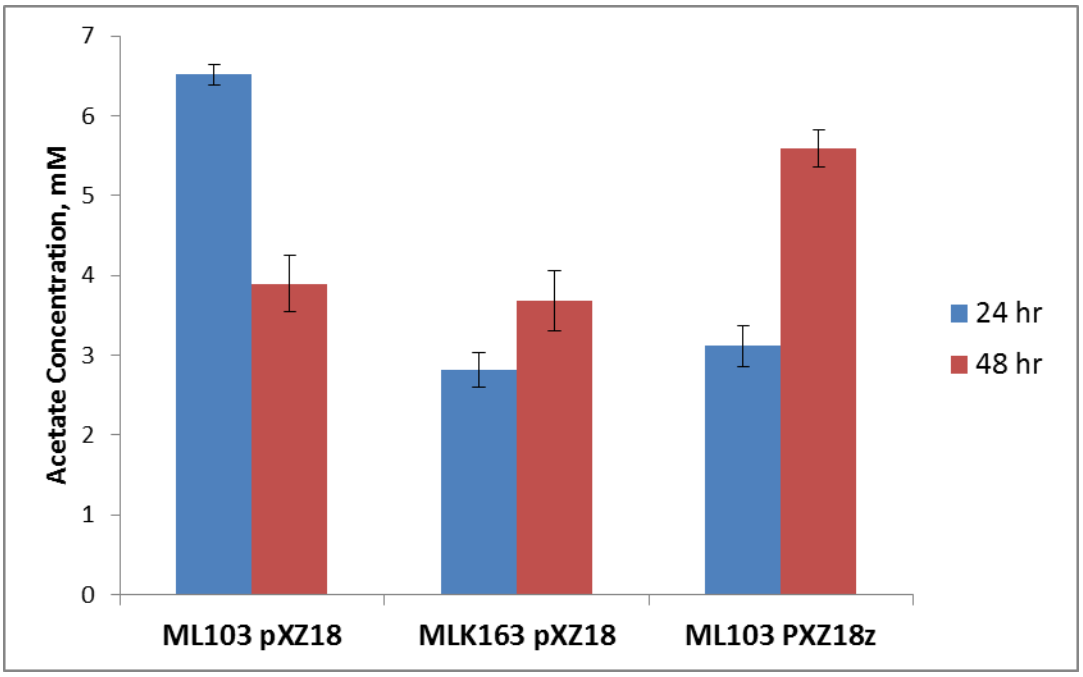


Figure S5 Acetate accumulation of ML103 pXZ18(Δ*fadD*), MLK163 pXZ18 (Δ*fadD*,Δ*sucC*) and ML103pXZ18z (Δ*fadD*,*fabZ*$^{++}$) at 24 and 48 hour. The data shown are means ± standard deviation of at least 3 replicates.

CHAPTER 5

CLOSING METABOLIC ENGINEERING CYCLE BY INTEGRATED APPROACH OF COMPUTATIONAL AND OMICS TOOLS FOR OVERPRODUCTION OF FATTY ACIDS

A draft to be submitted to Metabolic Engineering

Authors: Ting Wei Tee[1], Anupam Chowdhury[2], Erin Boggess[3], Ping Liu[1], Jong Moon Yoon[1], Julie Dickerson[3], Costa D. Maranas[2], Laura R. Jarboe[1] and Jacqueline V. Shanks[1]

[1] Department of Chemical and Biological Engineering, Iowa State University, Ames, IA

[2] Department of Chemical Engineering, Pennsylvania State University, University Park, PA

[3] Department of Electrical and Computer Engineering, Iowa State University, Ames, IA

Abstract

Fatty acids naturally synthesized in many organisms are promising platform intermediates for the catalytic production of industrial chemicals and biofuels. The tight transcriptional and translational regulation in fatty acid biosynthetic pathways complicates the engineering of microbial hosts for higher yields. In this study, we characterized the microbial physiology of the first generation engineered strain ML103 pXZ18Z (deletion of β-oxidation, over-expression of fatty acid elongation and thioesterase) under the control and fatty acid producing condition, along with transcriptomics and metabolomics analysis. We then redeployed an integrated approach of metabolic flux analysis and computational techniques for the overproduction of fatty acid. The OptForce procedure, a computational strain design algorithm, was employed to suggest and prioritize genetic manipulations to overproduce fatty acids. Metabolic flux analysis identified up-regulation of the TCA cycle to replenish the need of energy and reducing molecules for the biosynthesis of fatty acids under fatty acid overproduction. The elevation of metabolite level in the TCA cycle complements the flux findings. As a consequence, updated OptForce simulation suggested interruption of TCA cycle and enhancing acetyl-CoA pool. In accordance to OptForce suggestion, succinate dehydrogenase was deleted but does not improve the fatty acid production due to by-product accumulation. Using the whole genome transcriptional analysis by RNAseq, we observed acid stress response and membrane disruption during fatty acid

production as the result of product toxicity. This pinpoints the future effort to incorporate metabolic toxicity and genetic regulatory effects into the next generation OptForce to improve the accuracy of genetic predictions. Our study reinforces the advantage of integrating computational, experimental and omics tools for the design and engineering of microbial strains to overproduce value-added chemicals or biofuels.

1. Background and introduction

Concerns about crude oil depletion and climate changes have encouraged the development of processes that produce renewable biochemicals and biofuels from biomass-derived feedstocks. The microbial production of fatty acids as biorenewable chemicals and biofuels has recently garnered extensive attention (Lennen and Pfleger, 2012). Free fatty acids can be used as precursors for the production of alkanes/ alkenes (Fjerbaek et al., 2009; Lennen et al., 2010), fatty esters (Steen et al., 2010), fatty alcohols (Dellomonaco et al., 2011) and alkyl ketones (Goh et al., 2012). Interestingly, medium chain fatty acids (C_6-C_{14}) can be used as sources for industrial applications such as detergents, soaps, lubricants, cosmetics, and pharmaceuticals (Steen et al., 2010). As fatty acid chain length can directly affect the chemical properties, it is desirable to be able to have predictable and tight control over this product property. Plant thioesterases, that can hydrolyze acyl-ACP substrates with specific chain length specificity, such as those recently described (Jing et al., 2011), open new avenue for the production of novel fatty acids.

Even though bacteria are intrinsically capable of synthesizing fatty acids for lipid and cell membrane biosynthesis, *Escherichia coli* normally does not accumulate free fatty acids during the metabolic process. Wherein, fatty acid metabolism (i.e. consumption and production) is regulated at transcriptional levels by FadR and FabR transcription factors in *E. coli* (Fujita et al., 2007). In addition, strong product inhibition tightly controls the rate of fatty acid synthesis in *E. coli* (Heath and Rock, 1996a, b). Furthermore, the toxicity of fatty acids reduces cell viability and causes loss of membrane integration, further hindering the production of fatty acids (Jarboe et al., 2011; Lennen et al., 2011; Royce et al., 2013). Therefore, strain design for fatty acid biological production remains a huge challenge and possibly requires significant reengineering of cellular metabolism.

Recent metabolic engineering efforts have led to improvements in the synthesis of free fatty acids by *E. coli*. As the first step of fatty acid biosynthesis, acetyl-CoA is converted into

malonyl-CoA by an ATP-dependent acetyl-CoA carboxylase (ACC), followed by its conversion to malonyl-ACP through malonyl-CoA-ACP transacylase. This reaction has been previously identified as the initial rate limiting step of fatty acid production (Davis et al., 2000). To overcome this bottleneck, the availability of precursors was augmented by overexpressing ACC (Lennen et al., 2010) and malonyl-CoA-ACP transacylase (Zhang et al., 2012c). Acetoacyl-ACP, which is generated from the condensation of acetyl-CoA and malonyl-ACP, enters the elongation cycle. The elongation cycle fully reduces the β-ketoacyl-ACP to an acyl-ACP, followed by the condensation with malonyl-ACP. The accumulation of long chain fatty acyl-ACP inhibits multiple enzymes in fatty acid biosynthesis (Davis and Cronan, 2001). To alleviate the feedback inhibition, there are reported strategies to increase expression of *E.coli* native thioesterase *tesA* gene (Steen et al., 2010) and heterologous acyl-ACP thioesterases from plant (Liu et al., 2010; Lu et al., 2008; Zhang et al., 2011). In addition, fatty acid degradation through the β-oxidation pathway can be blocked by deleting *fadD* (Li et al., 2012; Lu et al., 2008; Steen et al., 2010) or *fadE* (Steen et al., 2010) gene. Acetate is the major byproduct in aerobic fermentation for fatty acid production, however deletion of acetate formation pathway does not increase fatty acid yield (Li et al., 2012). More recently, Zhang *et al*. developed a dynamic sensor-regulator system to produce fatty acid-derived products in *E. coli* by using FadR transcription factor as fatty acid biosensor (Zhang et al., 2012a). Most of these genetic interventions are based on expert intuition, and lack of systematic quantitative approach and omics-based system level metabolic engineering tools. Rational engineering complemented with omics approaches could unravel the underlying cellular response during fatty acid overproduction, providing insights for future strain engineering. This could potentially shorten the turnover time for strain development.

Omics tools have been widely used to understand and engineer the molecular system. The latest development of high-throughput techniques for deciphering genomes, transcriptomes, proteomes, metabolomes and fluxomes, together with computational tools, can predict gene targets to enhance production of desired molecules (Lee et al., 2012). Transcriptome analysis is robust in elucidation of useful metabolic genes, identification of novel target genes for improving strain performance, and identification of regulators controlling pathway for target molecules production (Hirasawa et al., 2010; Jarboe et al., 2011; Serrania et al., 2008; Yano et al., 2003). Proteome analysis quantifies the abundance of proteins and their changes under particular genetic and environmental conditions (Zhang et al., 2010). Hence, it can capture regulation of

cellular response beyond the mRNAs, providing the information of complicated biological processes and posttranslational modifications (Han and Lee, 2006). The rate of an enzymatic reaction is controlled by the concentration of the substrate and enzyme abundance; often feedback regulation by the product also occurs. Thus, metabolomic data serves as a snapshot of the physiological state of the cell, which reflects the integrated output of multiple complex interactions (Ewald et al., 2009). Among the different levels of information, metabolic flux distribution in central carbon metabolism is the most direct approach to understand the complex metabolic control mechanism of the whole cell (Matsuoka and Shimizu, 2010). The goal of metabolic flux analysis (MFA) is to quantify intracellular metabolic fluxes and to scrutinize the functional aspects of metabolic network in depth (Kim et al., 2008; Sriram et al., 2004; Sriram et al., 2007). Combinations of omics investigations (i.e, transcriptomic, proteomic, metabolomic and fluxomic) are increasingly applied with the goal of gaining a comprehensive understanding of biological systems. However, to our knowledge, there are only few reported combinations of various omics techniques to study fatty acid production by *E. coli*.

Integration of experimental and computational approaches has proven to be a successful systematic method for strain design (Ranganathan et al., 2012). The OptForce algorithm (Ranganathan et al., 2010) is an optimization procedure for identifying genetic manipulations to improve the production of useful chemicals in an organism. It uses MFA data to characterize the reference state, and subsequently identifies a minimum set of manipulations that could redirect the metabolic flux towards the desired product chemical. The OptForce procedure prioritizes these interventions in accordance of their impact on the yield of target chemical, as well as suggesting various genetic strategies for achieving the same target production. We have successfully employed OptForce to predict interventions for the production of fatty acids of specific chain-lengths in *E. coli* (Ranganathan et al., 2012). In accordance with the OptForce prioritization of interventions for medium-chain fatty acids, we implemented the first two suggested manipulations by up-regulating *fabZ* (β-hydroxyacyl-ACP dehydratases), and deleting *fadD* (fatty acyl-CoA synthetase) genes. A heterologous thioesterase from *R. communis* was expressed to restrict the chain-elongation specificity to medium-chain fatty acids. The engineered strain produced fatty acid primarily of C_{14} and C_{16} chain lengths with a yield of 39% of its theoretical maximum. The interventions suggested by OptForce were based on the metabolic characterization of the reference phenotype of the strain. As we implement each intervention,

regulatory effects, metabolic toxicity, and other factors beyond the purview of OptForce may increasingly affect the redistribution of metabolic fluxes, eventually causing significant departures between network phenotype predictions and experimental results. These factors are particularly significant for fatty acid production, since fatty acids are toxic to the host strain (Lennen et al., 2011).

In this study, we propose an iterative strategy to re-deploy OptForce after short intervals in strain construction by taking advantage of multiple MFA measurements (Figure 1). We re-characterized the first generation engineered strain, and subsequently re-deployed the OptForce procedure on the "new" reference strain to identify a new set of genetic manipulations. This process could be repeated until desired strain performance is achieved. In this set of new interventions, some of the old interventions remain unaltered, while entirely new ones, in accordance with the flux re-distribution, appear. In addition, the prioritization of the interventions is also significantly affected. Meanwhile, we conducted transcriptomics and metabolomics analysis to compare the physiology of first generation engineered strain under normal growth and fatty acid production conditions to unravel underlying cellular physiology and guide the genetic manipulation for the overproduction of fatty acids. We also reported on the production yield of fatty acids following the implementation of OptForce intervention, thus closing the metabolic engineering cycle.

2. Material and methods

2.1. Strains and plasmids

E. coli strain ML103 (MG1655, Δ*fadD*) carrying plasmid pXZ18Z (*fabZ*+, thioesterase from *R. communis*) was used as the reference strain for this study. ML103 pXZ18Z was the first generation of $C_{14\text{-}16}$ fatty acid-overproducing strain developed from our previous study (Ranganathan et al., 2012). In accordance to Optforce prediction, knockout of the succinate dehydrogenase (*sdhABCD*) in strain ML103 was performed following Datsenko and Wanner's method (Datsenko and Wanner, 2000). The upstream 600 bp homologous region was amplified with primers sdh-UpF and sdh-UpR. The downstream 540 bp homologous region was amplified with primers sdh-DnF and sdh-DnR. The chloramphenicol resistant gene (cat) was amplified from plasmid pKD3 using primers sdh-F and sdh-R, which each have a 50 bp overlap with the upstream and downstream homologous region. Overlap extension PCR was carried out to assemble the upstream homologous sequence, chloramphenicol resistant gene, and downstream

homologous sequence, resulting in a 2180 bp PCR product which was used to knockout the sdh genes. To verify the knockout, two primers P1 and P4 were used to amplify the chloramphenicol resistant gene in the genome. PCR product was sequenced to confirm the correct sequence. All the primer sequences are listed in the Supplementary Table S1.

2.2. Metabolic flux analysis experiments

Strains and culture conditions:

The ^{13}C labeling experiment was performed in duplicate as described by Ranganathan et al. (Ranganathan et al., 2012). In brief, a freshly transformed *E. coli* ML103pXZ18Z was precultured in 25 mL minimal M9 medium supplemented with 1% glucose and 100 mg/L ampicillin in an orbital shaker at 250 rpm until exponential phase (OD ~2). The cells were inoculated into the bioreactor (INFORS HT, Switzerland) to a starting OD_{550} of 0.02. 20% U-^{13}C glucose and 80% 1-^{13}C glucose were used as the tracers to the final medium concentration of 1% glucose. 1mM isopropyl-β-D-thiogalactopyranoside (IPTG) was added to induce fatty acid production. The aerobic fermentation was conducted at 30 ℃, pH 7.0, and 0.6 ml/min air flow rate. The pH of culture medium was maintained at 7.0 using 0.5M KOH. The dissolved oxygen level was maintained above 50% of saturated levels to ensure aerobic conditions by varying agitation speed. The cells were harvested at mid-exponential phase (OD ~2.0) after at least 5 generations to ensure metabolic and isotopic steady state. Parallel experiments were conducted using natural glucose as biological replicates for transcriptomics analysis.

Analytical techniques:

Cell dry weight was determined by measuring optical density OD_{550} using a spectrophotometer (Genesys 20, Madison, WI). Media samples were taken during the exponential growth and filtered through 0.22 μm pore sized nylon filters (P.J. Cobert Associates, Saint Louis, MO) and kept at -20^{o}C for extracellular metabolite analysis. Glucose and organic acids were measured using a Waters HPLC (Waters, Milford, MA) with 410 refractive index detector. The Aminex column (HPX-87H, Bio-Rad, Hercules, CA) was used at 30^{o}C with 0.3 mL/min of 5mM sulfuric acid as mobile phase. After the experiment, cell biomass were harvested and hydrolyzed into amino acids. The isotopomer labeling fractions were measured using 2D [^{13}C, ^{1}H] Heteronuclear Single Quantum Correlation (HSQC) spectra were acquired on a Bruker Avance DRX 500MHz spectrometer at 298 K. The isotopomer fractions are listed in Table S3.

Flux evaluation methodology:

A metabolic network model for *E. coli* was constructed based on Ecocyc database, literature and transcriptome data (Ranganathan et al., 2012). Fluxes were computed using NMR2Flux software (Sriram et al., 2004) by cumomer and isotopomer balancing. A global optimization routine was applied to find stoichiometrically feasible fluxes set consistent with experimental measurements. Optimal fluxes were estimated by minimizing the chi square error between measured and simulated isotopomer fractions of amino acid. Bootstrap Monte Carlo statistical analysis was then used to evaluate the standard deviations of the flux set.

2.3 Using OptForce for fatty acid overproduction

The *i*AF1260 metabolic model of *E. coli* (Feist *et al*., 2007) was used to perform the simulations with the OptForce procedure for the overproduction of fatty acids. In our previous study, no alternation in the stoichiometric network of *i*AF1260 was required since the wild-type *E. coli* MG1655 strain was our reference strain. However, as a result of the genetic manipulations required to construct the engineered strain, the base model of *i*AF1260 is no longer consistent with the metabolic map of the new reference strain. Knock-out/in of genes, as well as up/down-regulation in gene expressions must be captured accordingly to accurately characterize the present state of the strain phenotype. Hence, we identified a set of general rules to incorporate the effect of genetic manipulations in the model of new reference strain.

i. Gene knock-in/deletion: When a heterologous gene is expressed in the strain, reaction(s) encoded by the gene are added to the model, with appropriate reaction directionality. Likewise, reaction(s) encoded by genes that have been knocked out are set to carry zero flux. These genes could encode for a single reaction, multiple-reactions, or part of an enzyme-complex. For example, the *fadD* gene encodes for all the fatty-acid-CoA ligase reactions in the β-oxidation pathway that converts fatty acids to their respective fatty acyl-CoA in *E. coli*. We set all the fluxes of these reactions to zero in order to implement the removal of *fadD* in our first-generation mutant.
ii. Gene over-expression/down-regulation: Implementing the effect of gene up/down-regulation is not as straightforward as gene knock-in/out. If we have direct flux measurements (through MFA) of a reaction(s) that is up-regulated due to increased expression of a gene, we set a lower bound for the flux equal to the new reference flux for the engineered mutant. In cases when direct measurement of the flux is not available, the

lower bound is set on the nearest measured flux in the pathway that directly affects the up-regulated reaction. Alternately, if such an up-regulation leads towards increase in the production of an external metabolite, we can set as a lower bound the experimentally measured exchange flux of the secreted metabolite. For example, up-regulation of *fabZ* gene increased the flux through fatty acid synthesis pathway. We implemented the intervention by setting a lower bound on the exchange flux for fatty acids equal to their experimental yield in the first-generation engineered strain. The changes to implement the down-regulation of genes are similar as above, with the difference being we set upper bounds on the fluxes of intracellular/exchange reactions instead of lower bounds.

In addition, some manipulations could have metabolic effects that cannot be captured by the above simple rules, and should be dealt with on a case-by-case basis. For example, we expressed a heterologous thioesterase from *R. communis* in wild-type *E. coli* to produce fatty acids of C_{14}-C_{16} specific chain length. GC-MS studies showed that the fatty acid secreted by the engineered mutant was composed of primarily C_{14} and C_{16} with traces of C_{18}. Fatty acids of other chain-lengths were negligible. We used a constraint which fixes the flux ratios for the exchange reactions towards specific fatty acid chain-lengths in the values observed experimentally. These general rules for implementing the effect of interventions are appended for each intervention as additional constraints in all of the steps of the OptForce procedure (i.e. characterization of phenotypes, computation of MUST sets and the final FORCE sets).

The operational strategy for re-deployment of the OptForce protocol on the first-generation mutant overproducing fatty acid is structurally similar to the procedure followed in our previous study (Ranganathan et al., 2012). All simulations, using the *i*AF1260 model with added modifications, were performed in aerobic minimal conditions with glucose as the sole source of carbon in the medium. The upper and lower limits for all reactions in the model including those for uptake of glucose, oxygen, and other exchange nutrients were set at the same values as specified in our earlier study. Likewise, all regulatory restrictions were imported from our previous study. The new reference phenotype was characterized to be consistent with the redistribution of metabolic fluxes in the first generation engineered strain, as revealed from the ^{13}C-isotopomer study (see Figure 1). The MFA data for 33 reactions from glycolysis, pentose phosphate pathway, TCA cycle and anaplerotic reactions were used for the characterization of the new reference phenotype. The biomass flux was set at the maximum achievable flux subject

to the experimental flux measurements. This value decreased from 52% of maximum theoretical yield (wild-type = base strain) to 22% (double mutant = base strain) in the second generation strain design. This indicates that more flux has been diverted away from biomass towards fatty acid production. It is important to note that this reduction in biomass flux is in agreement with experimental observations that showed that the specific growth rate is halved in the first generation mutant as compared to that of wild-type *E. coli* (see Figure S1). The overproducing phenotype was characterized to be consistent with 90% production of medium-chain fatty acid. The maximum theoretical production of medium-chain fatty acid was determined by maximizing the production of the sum of C_{14}, C_{16} and C_{18} chain-length fatty acid in the same ratio of production as obtained in our previous study.

After characterization of the reference and the desired phenotypes, the genetic interventions were identified implementing the same procedure as mentioned in our previous work (Ranganathan et al., 2012). Subsequently, we identified the minimum set of interventions that predicts a guaranteed yield of fatty acids (i.e. FORCE sets) by implementing the bilevel optimization. As in our previous study, we prevented the removal of reactions associated with *in vivo* essential genes (based on the KEIO collection (Baba et al., 2006; Feist et al., 2007)) and experimentally verified synthetic lethal genes (Suthers et al., 2009), even for the ones that were not recognized as essential by the *i*AF1260 model. In addition, any further manipulations on reactions that have already been manipulated (i.e. reactions in fatty acid chain-elongation pathway and the β-oxidation pathway) were disallowed.

2.4 Intracellular metabolite profiling:

E. coli ML103 carrying pXZ18Z plasmid was grown in shake flask (30^{o}C, 250rpm) for metabolomics analysis between the control and fatty acid overproduction condition. Samples were harvested at 24, 48 and 72 hours. After harvesting, 1mL of samples were kept on ice and centrifuged for 10 minutes at 8000x g and 4^{o}C. With the supernatant discarded, 5ng norleucine was added as internal standard to the cell pellet. Cell pellets were then reconstituted in 1mL 60% cold methanol to quench the cell and extract the intracellular metabolites. Cell debris was removed by centrifugation (8000x g, 4^{o}C, 5 minutes). The extracts were transferred into fresh tubes and stored at -80^{o}C for future analysis. The samples were dried completely in vacuum centrifuge and derivatized twice as described by (Ewald et al., 2009). In the first derivatization, 20μL methoxyamine solution (20mg/mL methoxyamine hydrochloride (Sigma-Aldrich) in

pyridine (analytical grade, Sigma-Aldrich)) was added and incubated at 37^{o}C and 250rpm for 90 minutes. 7.5μL aliquots were transferred into GC glass vial and derivatized for the second time with 15μL N-methyl-N-(trimethylsilyl)-trifluoroacetamide with 1% tert-butyl-dimethylchlorosilane (Fluka). The mixture was then incubated at 60^{o}C for an hour. The intracellular metabolite content was analyzed using an Agilent GC-MS system with the DB-5MS column (30m, 0.25mm i.d., 0.25μm, Agilent). The oven temperature was initially set at 70^{o}C for 1 min and raised to 140^{o}C with 20^{o}C/min ramping rate. The temperature was then increased to 300^{o}C with 5^{o}C/min ramping rate and held for a minute. Raw MS data was integrated using Chemstation software. Compound peaks were identified by running pure standards and quantified by normalizing to the internal standard.

2.5 RNAseq transcriptome profiling

Samples for RNA isolation were collected from the bioreactor at mid-log phase (OD ~2.0) and immediately cooled in the dry ice-ethanol bath. After centrifuging for 10 minutes (4^{o}C, 8000x g), the supernatant was discarded and cell pellets were resuspended in 2mL RNA Later solution (Qiagen) and kept at -80^{o}C. Total RNA was purified using Qiagen RNeasy mini kit according to the product handbook with application of DNase to remove genome DNA. The mRNA was then purified using MICROBExpress Kit (Life Technology, CA) following the manufacturer's protocol.

The high-throughput mRNA sequencing (RNA-seq) for three control and three IPTG treatment samples was performed at Iowa State University DNA Facility using the Illumina HiSeq 2000 platform. The short reads obtained were paired-end and 100 base pairs (bp) in length. Differential gene and transcript expression analysis was performed using the Tuxedo protocol (Trapnell et al., 2012). The *E. coli* K-12 MG1655 genome (Blattner et al., 1997) was used as a reference genome and a General Feature Format (GFF) file was generated from EcoCyc data to provide transcriptome annotations.

In accordance with the Tuxedo protocol, TopHat (version 2.0.3) (Trapnell et al., 2012) was run to map reads from each sample to the reference genome. The underlying alignment algorithm was Bowtie2 (version 2.0.0-beta5) (Langmead and Salzberg, 2012). Next, Cuffdiff, from the Cufflinks RNA-Seq analysis tools (version 2.0.0), was used to estimate transcript abundance and test for differential expression. The three control and three treatment samples were each pooled for this analysis. Sets of genes with significant increases or decreases in

abundance were examined for common biological processes as described in associated Gene Ontology (Ashburner et al., 2000) annotations. BiNGO (Maere et al., 2005), a plugin for Cytoscape (Smoot et al., 2011), was used to identify overrepresented terms and generate a visualization of the ontology. Additionally, Network Component Analysis (NCA) (Kao et al., 2004) was used to predict transcription factor activities based on fold changes of genes with a significant False Discovery Rate (FDR) corrected p-value.

2.6 Metabolic interventions and fatty acid titer determinations

Fermentation procedure:

Strain evaluation was conducted in shake flask system with higher glucose content. Each strain was freshly transformed and streaked on LB plate with 100 mg/L ampicillin overnight at 30^{o}C incubator. A single colony from the plate was grown in 5 mL M9 medium supplemented with 1.5% glucose and 100 mg/L ampicillin for 16-20 hours in orbital shaker at 30^{o}C and 250 rpm. The pre-culture was then inoculated into 250mL flasks containing 40mL M9 medium with 1.5% glucose and 100 mg/L ampicillin. The expression of acyl-ACP thioestersase was induced by the addition of IPTG to final concentration of 1mM. Samples were taken at 24 and 48 hours for measurement of fatty acids and extracellular metabolites.

Fatty acid analysis:

Cell cultures were harvested and prepared for fatty acid analysis as described previously (Ranganathan et al., 2012). Fatty acids in the broth were extracted using chloroform, methylated into methyl esters and recovered using hexane. Tridecanoic acid, pentadecanoic acid and heptadecanoic acid were added as internal standards in all samples before extraction. The fatty acid content was analyzed using an Agilent GC-FID/MS system. Raw MS and FID data was integrated using Chemstation software. Compound peaks were assigned by running standards or referring to the mass fragmentation in the NIST library.

3 Results

3.1 Flux measurements:

Phenotype analysis of ML103 pXZ18Z (acyl-ACP thioesterase, *fabZ* overexpression and *fadD* knockout) between control and fatty acid producing condition was compared among four biological replicates, in which the carbon source is either natural glucose or mixture of ^{13}C labeled glucose. There was no difference in macroscopic growth characteristics using ^{13}C

glucose and natural ^{12}C glucose (Leighty and Antoniewicz, 2012). The experiments were performed in M9 media under aerobic growth in batch bioreactor. The expression of acyl-ACP thioestersase and β-hyroxyacyl-ACP dehydratase was induced by IPTG. We observed similar macroscopic characteristics (growth rate and acetate yield) of ML103 pXZ18Z in the un-induced normal growth (control) condition, compared with the wild type *E. coli* MG1655 and ML103 (Δ*fadD*) (Ranganathan et al., 2012). When fatty acid production is induced, the cell growth rate and biomass yield both decrease by ~50% relative to the un-induced condition, though the glucose uptake rate remains unchanged (Figure S1). Fatty acid yield increases 2-fold over the control during the exponential growth with acetate as by-product. While, acetate yield during fatty acid overproduction was ~43% lower than the control. Increased expression of fatty acid elongation pathway and heterologous thioesterase pulls the carbon fluxes from acetyl-CoA in the central carbon metabolism to form malonyl-CoA as the precursor for fatty acid synthesis, instead of channeling carbon fluxes to amino acid biosynthesis and acetate formation.

In vivo MFA quantifies the metabolic network operating under the physiological conditions. Flux experiments were performed to compare ML103 pXZ18Z under the control and fatty acid producing condition. The flux distributions from the latter experiment were then incorporated into OptForce to predict the second round of genetic manipulations. The flux maps for *E. coli* ML103 pXZ18Z under the control and fatty acid producing condition are tabulated in Table S2 and shown in Figure 2. The flux values are all normalized to 100 mmol $gDW^{-1}\,h^{-1}$ glucose uptake rate and an average of duplicate experiments. Overall, most of the carbon flux (around 90%) is directed toward the glycolytic pathway, resulting in ~7% of the carbon flux being channeled through pentose phosphate (PP) and Enter-Doudoroff (ED) pathways, presumably to generate NADPH for reduction requirements. The anaplerotic pathway of phosphoenolpyruvate (PEP) carboxylase is active, which converts PEP to oxaloacetate (OAA) to refill the OAA pool for biosynthesis. Acetate was formed as the major product at ~60 mmol $gDW^{-1}\,h^{-1}$. The TCA cycle operates at 55 mmol $gDW^{-1}\,h^{-1}$ presumably to generate biomass, ATP and NAD(P)H for cell growth.

Under fatty acid production, the combined effect of the heterologous thioesterase expression and *fabZ* up-regulation pulls more flux in the central metabolism towards acetyl-CoA and fatty acid synthesis. As a result, we observe a 20% increase in flux in lower glycolysis as compared to the control, while the flux through the PP and ED pathways decreases ~60%. In

addition, the flux through the TCA cycle increases 40%. We also observe 50% increase of carbon flux through pyruvate kinase, instead of the PEP carboxylase (ppc) pathway. With the decrease in cell growth rate and biomass yield during fatty acid production, the amino acid biosynthesis reactions are generally lower than the control. The increase of flux from acetyl-coA to fatty acid synthesis is complemented with the decrease of flux toward acetate (byproduct formation). In addition, the ED pathway and glyoxylate pathway have negligible flux values. For fatty acid overproduction, carbon flux at the acetyl-CoA node should be ideally channeled to fatty acid biosynthesis pathways, as well as minimizing the acetyl-CoA drain toward the TCA cycle and acetate formation. From the *in-vivo* flux distribution, we could hypothesize the TCA cycle should be down-regulated in order to increase carbon fluxes toward fatty acid synthesis. However, it should also be noted that the TCA cycle is a source of metabolites that are critical to biosynthesis, such as α-ketoglutarate.

Even though cofactor balance was not included in MFA, we derive the net cofactor production from the *in-vivo* flux distribution. Figure S2 illustrates the use and production of NADH, NADPH and ATP in the central carbon metabolism. For example, NADPH is produced by the oxidative PP pathway, TCA cycle and malic enzyme. Interestingly, malic dehyrogenases are encoded by *maeA* (Yamaguchi, 1979) and *maeB* (Iwakura et al., 1979) genes, which are specific to NAD+ and NADP+ as precursor respectively. The malic dehydrogenase catalyzed by these enzymes cannot be distinguished by ^{13}C-MFA, which is constrained by reaction carbon stoichiometry. The malic enzyme is thus considered to produce either NADH or NADPH, and treated as the measurement errors in cofactor production. The net productions of ATP, NADH and NADPH per glucose consumed were increased 12-20% in the fatty acid production condition (Figure S2).

Overall, ^{13}C-MFA reveals flux increase in pyruvate kinase and the TCA cycle to channel pyruvate generated mostly through glycolysis towards acetyl-coA and the TCA cycle during fatty acid production. In addition, cofactor balancing estimated from the MFA deciphers the importance of NADPH and ATP in *E. coli* fatty acid synthesis.

3.2. OptForce results

The reconfiguration in flux distribution in the central carbon metabolism of the first-generation engineered strain has an impact on the updated set of OptForce predictions, as illustrated in Figure 3. All the interventions were not from the fatty-acid synthesis pathway and

concentrated around central carbon metabolism, since we had already engineered the fatty acid synthesis pathway in the first-generation mutant. While some of the original interventions were identified again, entirely new interventions, such as the deletion of C-acyltransferase (GLYAT) and acetaldehyde dehydrogenase (ACALD), replaced others in the light of new MFA data. In addition, the priority of the interventions was dependent on whether the flux redirection was consistent with the worst-case predictions of OptForce.

The combined effect of the heterologous thioesterase expression and *fabZ* up-regulation pulls more flux in the central metabolism towards acetyl-CoA and fatty acid synthesis, leading to 40% increases in the TCA cycle activity. As a consequence, the updated OptForce predictions suggest that down-regulation of TCA cycle activity and up-regulation of flux towards acetyl-CoA as the most important interventions for improving medium-chain fatty acid production. Removal of succinate dehydrogenase (SUCD) reaction severs the TCA cycle into two separate branches. The additional flux can now be routed towards fatty acid synthesis. OptForce suggests at-least one-and-half fold up-regulation in pyruvate dehydrogenase (PDH) flux from its reference flux to further increase the availability of the precursor acetyl-CoA towards fatty acid. It is worth observing that even though these two strategies (i.e. down-regulation of TCA activity and up-regulation in PDH flux) were identified in the original set of interventions (see Figure 3), the present strategy assigns them higher priority in accordance with the new flux distribution. As the next intervention, OptForce identifies the removal of either phosphotransacetylase (PTA) or acetate kinase (ACK) reactions to prevent the conversion of acetyl-coA towards acetate. The lower priority for this intervention when compared with SUCD removal/ PDH up-regulation can be explained in terms of the observed redistribution of metabolic fluxes. Acetate production for the first generation mutant decreases by 40% relative to the control, without having to impose any direct interventions. This indicates that the OptForce imposed safeguards against such a "worst-case" re-direction in flux are not warranted and therefore received lower priority in the revised predictions. For the final set of interventions, OptForce suggests the removal of glycine C-acyltransferase (GLYAT) and acetaldehyde dehydrogenase (ACALD), which was absent in the original set of interventions. Removal of GLYAT and ACALD reactions prevents the conversion of acetyl-CoA towards biomass precursor threonine and acetaldehyde respectively, channeling the flux towards fatty acid synthesis. Further interventions did not provide significant improvement in medium-chain fatty acid yield. OptForce predicted a minimum guaranteed yield

of 0.24 gm medium-chain fatty acid/gm of glucose after implementation of five interventions, which corresponds to 66% of the maximum theoretical yield of the product (see Figure 3). Even though the new manipulations do not improve on the yield predicted after the original set of interventions, OptForce suggests it can be achieved by one fewer intervention compared to the previous strategy.

The comparison of the new set of suggestions with the old strategy also reveals the absence of interventions which were inconsistent with the new flux distribution. The original strategy suggested re-routing of glycolytic flux through the ED pathway through down-regulation of phosphoglycerate mutase (PGM), removal of glucose-6-phosphate isomerase (PGI) in glycolysis, and removal of transaldolase (TALA) in the PP pathway. This re-direction was suggested to arrest excess ATP generation in the organism that will lead to higher biomass production, as well as to generate cofactor NADPH for increased fatty acid pathway activity. However, contrary to expectations, the MFA data of the first-generation mutant showed that both the PP pathway and ED pathway show 65% decrease in activity, and the gluocse flux was being routed through glycolysis instead. A possible reason for this could be the strong pulling force of the thioesterase and FabZ, which pulls glucose through the shortest route towards fatty acid synthesis, which is the main glycolytic pathway. Consistent with such observations, OptForce refines its predictions by not identifying any of these interventions in the updated set of manipulations.

3.3 Intracellular metabolite profiling results

Metabolomics has complemented genomic, transcriptomic and fluxomic study by characterizing the rapid cell response to metabolic flux changes via allosteric or feedback inhibition regulations (Mashego et al., 2007). We investigated the relative change in concentrations of intracellular metabolites (Figure 4) and amino acids (Figure S3) of our first generation *E. coli* fatty acid producer under the control and fatty acid production condition. The strain was grown in shake flasks in M9 minimal media with 1.5% glucose. The strain produces up to 2.3 g/L fatty acids after 72 hour cultivation. Under fatty acid production, most of the intracellular metabolites in the TCA cycle (i.e. succinate, fumarate and malate) exhibit increase in concentration at 48 and 72 hours compared to the control condition. However, citrate abundance during fatty acid production is lower than the control. Notably, the increase in the concentrations of the TCA cycle metabolites positively correlates with the increase in the TCA

cycle fluxes. This accumulation of intracellular metabolites in the TCA cycle diverts carbon away from fatty acid production. As a consequence, metabolomics study suggests that down-regulation of the TCA cycle activity could alleviate the accumulation of intracellular TCA cycle metabolites and increase carbon flux toward fatty acid biosynthesis.

In addition, free amino acid analysis reveals an overall increase trend in concentrations of isoleucine, valine and alanine during fatty acid overproduction. Valine and alanine are synthesized from pyruvate, while pyruvate and oxaloacetate are the precursor for isoleucine formation. Remarkably, the fluxes in central carbon metabolism around pyruvate (pyruvate kinase and pyruvate dehydrogenase) and oxaloacetate nodes (malate dehydrogenase and citrate synthase) are also increased during fatty acid production. On the contrary, the concentration of leucine, which originates from pyruvate and acetyl-CoA, decreases under fatty acid production condition. Despite the elevated concentration of free amino acids derived from pyruvate, this may suggest low level of acetyl-CoA as the possible bottleneck for fatty acid production. Notably, pyruvate dehydrogenase (PDH) is suggested by OptForce to increase the availability of acetyl-CoA pool as the precursor for fatty acid synthesis. In addition, glycine concentration is lower under fatty acid production, complementing the decrease in biomass fluxes from 3-phophosglycerate toward serine and glycine. Proline concentration however decreases at 72 hour under fatty acid production despite up-regulation in the TCA cycle fluxes.

3.4 Experimental characterization of metabolic interventions

The updated OptForce interventions for palmitic acid (medium-chain fatty acid) include deletion of succinate dehydrogenase and up-regulation of pyruvate dehydrogenase. We thus implemented the prioritized intervention by deleting *sdhABCD*, which encodes succinate dehydrogenase, to redirect carbon flux toward fatty acid biosynthesis instead of the TCA cycle. The fatty acid titer experiments were performed in minimal M9 medium with 1.5% glucose using ML103 pXZ18Z as the reference strain. Surprisingly, the fatty acid titer of TJ103 pXZ18Z (*sdhABCD* deletion mutant) decreases 25% compared to ML103 pXZ18Z after 48 hours cultivation, along with 25% decrease in fatty acid yield (Figure 5). Interruption of the TCA cycle will ideally channel most of the acetyl-CoA flux toward fatty acid synthesis. However, deletion of succinate dehydrogenase could lead to an incomplete TCA cycle, in which succinate could not be oxidized and would accumulate. Extracellular metabolite analysis revealed the accumulation of succinate (20mM), acetate (16mM) and pyruvate (5mM) after 48-hour cultivation

(Supplementary Figure S4). So instead of the desired effect of diverting carbon to fatty acid biosynthesis, it seems that deletion of succinate dehydrogenase resulted in increased carbon flux to fermentation by-products. Notably, the acetate accumulation is 3-fold higher for the succinate dehydrogenase knockout strain after 48-hour cultivation. This could suggest the by-product accumulations might have allosteric regulation on the thioesterase and fatty acid pathway, which is not considered in OptForce simulation.

3.5 RNAseq analysis

Transcriptome analysis was performed to explore the effect of *fabZ* overexpression. Using a transcriptome of 4,498 annotations generated from EcoCyc gene data, 4,114 genes were successfully tested using Cuffdiff. Of these, 146 genes have FDR corrected p-values < 0.05 and are determined to exhibit significant change between the control and treatment samples. 116 genes had positive fold changes that range from 2.45 to 10.72, and 30 genes have negative fold changes that range from -2.23 to -4.53. Significantly perturbed transcripts are listed in Table 1, while the full transcript data set is listed in Electronic Supplementary Material. NCA was performed using connectivity matrices with predicted transcription factor-gene links (Table S4). Transcription factors ArgR, CRP, H-NS, NtrC, and RcsAB are predicted to have increased activity, while FlhDC, FNR, and LexA are predicted to exhibit decreased activity.

We first validated the RNAseq result with specific gene markers. No *fadD* transcripts were detected, consistent with our deletion of the *fadD* gene. Under IPTG induction, *fabZ* and *tesB* gene are expressed about 2.0-2.6 fold higher than the control (Table 1). Keasling and coworkers reported over-expression of FadR transcription factor led to the increase in fatty acid production and gene expression in fatty acid pathway (*fabB*, *fabF* and *accA*) (Zhang et al., 2012b). However, our result shows no significant perturbation in genes involved in fatty acid pathway except *fabZ*. This could be due to the difference in strain genetic engineering, as FadR is the global regulator to tune the expression of genes involved in fatty acid biosynthesis, degradation and membrane transport (Zhang et al., 2012b). Nonetheless, the over-expression of *fabZ* gene and FadR transcription factor are distinct strategies to increase the expression of specific genes in fatty acid biosynthesis pathway, enhancing the carbon flux toward fatty acid production.

Stress response

Perturbations in gene expression related to oxidative hydrogen peroxide stress and acid stress were observed under fatty acid production. Genes responding to oxidative stress were up-regulated, including *ydeI*, *ychH*, *ygiW*, and *yodD*. These oxidative stress-related genes were previously discovered to be altered under the presence of carboxylic acids in *E. coli* (Lennen et al., 2011) and *S. cerevisiae* (Legras et al., 2010). On the other hand, the glutamate-dependent acid resistance genes are significantly increased conferring resistance to extreme acid condition. These genes include *gadA*, *gadB*, *gadC*, *gadE*, *gadW* and *gadX* (table 1). The expression of *hdeAB*, which is a periplasmic acid stress chaperone, was up-regulated (>2-fold). *hdeAB* gene is identified to defend against acid stress (Kern et al., 2007).

E. coli contains several inducible acid resistance systems to survive extreme acid environments (Richard and Foster, 2004). For instance, acid resistance system 2 (AR2) is induced by extracellular glutamate during growth at acidic minimal medium. AR2 comprises of isoforms of glutamate decarboxylase (*gadA* and *gadB*), and glutamate-aminobutyric acid (GABA) antiporter (*gadC*). Meanwhile, acid resistance system 3 (AR3) requires extracellular arginine and is composed of the acid-inducible arginine decarboxylase *adiA* and the *adiC* antiporter, which exchanges extracellular arginine for the intracellular agmatine. AR2 and AR3 protect the cells from acid stress by consuming intracellular protons through amino acid decarboxylation to maintain an internal pH compatible with cell viability (Figure 6a). The NCA result revealed the up-regulation in ArgR transcription factor, which represses the transcription of genes involved in biosynthesis and transport of arginine. This scenario might cause arginine to be the limiting substrate for AR3 system.

Membrane disruption

Decreased membrane integrity had been identified as a problem during exposure to our production of fatty acids (Lennen et al., 2011; Liu et al., 2013; Royce et al., 2013). We observed significantly increased expression of several membrane protein genes during production of fatty acids (e.g. *wza*, *yneM*, *yohC*, *ymgE*, *yjiY*, *yaiY*, *yibI*, *yibH*, *yqaE*, *yfeN* and *yncL*). However, the *ompF* and *ompW*, which encode outer membrane proteins, showed decreased gene expression. On the contrary, NCA result showed up-regulation of Fur, a transcriptional activator for ferric uptake regulation. Liu et al. (2013) showed that activation of iron starvation pathways might be a symptom of membrane disruption in *S. cerevisiae* (Liu et al., 2013). Given the perturbed activities in membrane protein and iron uptake regulation followed by the lack of improved fatty

acid titer despite supplementation (Figure S5), we hypothesize the membrane properties (leakage and fluidity) might be disrupted during fatty acid overproduction. Royce and coworkers recently studied the membrane fluidity and leakage using *E. coli* ML103 pZX18Z in MOPS medium in bioreactor (Royce et al., 2013). They observed that fatty acid production is associated with an increase in membrane leakage; no change in membrane fluidity was observed. Membrane leakage indicates the porosity of the cellular membrane, in which membrane can selectively exclude harmful compound and retain valuable metabolites in cell. Membrane leakage thus might cause metal ions and cofactors to leak out of the cells and could not maintain optimum environment for cell homeostasis.

4 Summary and discussion

In this paper, we performed ^{13}C metabolic flux analysis on the engineered *E. coli* strain ML103 pXZ18Z which is derived from the first generation integrated approach of OptForce and experimental flux analysis of wild type *E. coli* (Ranganathan et al., 2012) . Incorporated with omics tool, we then extended to second round of computationally derived predictions followed by characterization of strategies for overproducing fatty acids in *E. coli,* thus closing the metabolic engineering loop. From phenotype analysis, the fatty acid-producing *E. coli* illustrates reduced cell growth with similar glucose consumption rate, despite the increase in fatty acid yield. IPTG was applied to induce the cells to express *fabZ* and *R. communis* thioesterase, leading to increase fluxes toward fatty acid synthesis instead of maximizing the cell growth and acetate formation. MFA revealed up-regulation of TCA cycle and down-regulation of PP pathway when the cells are producing fatty acids. Fatty acid biosynthesis is known as an energy intensive process and also requires reducing equivalents for its elongation cycle. The TCA cycle is a key component of metabolic pathway for ATP generation under aerobic condition. Similarly, each complete TCA cycle can generate 2 NADHs and a NADPH. We interpreted the up-regulation of the TCA cycle functions as the sources for ATP, NADH and NADPH production for fatty acids biosynthesis. Despite the role to provide NADPHs for cell growth, the PP pathway was down-regulated during fatty acid production. However, the increase in the TCA cycle offset its decrease to replenish the NADPHs by isocitrate dehydrogenase. Therefore, the overall availability of ATP and NAD(P)H were higher under fatty acid producing condition. The flux changes were complemented by the increase in TCA cycle metabolite concentrations. In accordance of increasing carbon fluxes through the TCA cycle, we observed accumulation of

succinate, fumarate and malate in the TCA cycle. MFA and metabolomics studies depicted the increase of TCA cycle to possibly replenish energy and reducing powers for fatty acid biosynthesis. On the contrary, the increase of acetyl-CoA fluxes toward the TCA cycle would trade off the fluxes toward fatty acid synthesis. Independent to any computational prediction, down-regulation of TCA cycle is hypothesized to redirect more carbon fluxes toward fatty acid biosynthesis.

We updated our original OptForce predictions on medium-chain fatty acid production in *E. coli* using MFA information on the first-generation engineered strain. Consistent with observations from the new flux distribution, down-regulation of TCA activity and up-regulation of the flux towards acetyl-CoA were identified as the most important interventions for increasing fatty acid productions. In contrast, the strategy to re-direct of flux through the ED pathway was omitted since it was inconsistent with the new flux redirection.

Independent from this work, San and co-workers used a classical "push and pull" concept in metabolic engineering to enhance acetyl-CoA supply toward fatty acid pathway by amplifying *fabZ,* and to minimize the acetyl-CoA drains by deleting *sucC* in the TCA cycle (San and Li, 2012). The beneficial interventions are combined to derive the best strain with highest fatty acid titer and yield. The results are depicted in Figure 9, in which the combination of *fabZ* over-expression and *sucC* deletion led to the highest titer and yield (~100% theoretical yield). However, the study was conducted in LB rich medium. Interestingly, the interruption of the TCA cycle, combined with the up-regulation of fatty acid pathway and deletion of beta oxidation, agrees with updated OptForce suggestions. Moreover, the down-regulation of phosphoglucomutase *(pgm)* predicted by the original OptForce simulation shows detrimental effect on fatty acid titer and yield (Figure 9), reinforcing the advantage of re-deploying OptForce.

This study highlights the advantages of re-deploying the OptForce procedure in suitable intervals of strain construction to adjust with unanticipated changes in flux due to metabolic toxicity, and other regulatory effects in the phenotype. The new set of measured fluxes re-calibrates the phenotype a closer point of reference to the ultimate desired phenotype, which improves the accuracy of OptForce predictions. More importantly, we gain valuable insight on whether the flux re-distribution is in accordance with the worst-case simulations of OptForce. This information aids in weeding out incorrect strategies, as well as re-prioritizing the original interventions. Our results show that this iterative method of re-deployment of OptForce results in

more accurate prediction of genetic manipulations than a “once-only” implementation of the procedure. Furthermore, this method provides a first-of-its-kind protocol for incorporation of multiple MFA measurements for constructing an engineered strain. In our present study, we measured the fluxes after implementing the first two implementations suggested by OptForce. Ideally, it is desirable to re-deploy OptForce after implementing each of the suggested interventions. However, such a procedure would require a large number of flux re-measurements, which is expensive, time-consuming and impractical. In addition, the metabolic phenotype may not change much with a single intervention, rendering such a procedure sub-optimal. Alternatively, interventions located in series along a pathway, or having a common impact on the network, could be implemented before re-characterization of the phenotype. For example, we measured the fluxes again after implementing both the interventions located in the fatty acid synthesis pathway, which had the similar effect of pulling flux towards fatty acid synthesis. However, with further experience with this strategy, we expect to gain more insight as to the optimum number of interventions for re-deploying OptForce that would yield the best results.

Transcriptomics analysis reveals significant perturbed gene expression profiles under overproduction of fatty acid. We identified acid and oxidative stress genes response that might be induced by endogenous fatty acid production. Elevation of membrane protein gene and iron uptake regulation shed insight on the membrane stress, in which the membrane leakage enhanced with increasing fatty acid titer. Lennen et al. performed a thorough transcriptomic study on fatty acid-inducible stresses in an engineered fatty acid-producing *E. coli* strain (Lennen et al., 2011). Some changes described in Lennen’s work were observed in our data (Table 1). For example, phage shock proteins are induced by the exposure to cell envelope stress (exposure to organic solvent, heat and osmotic shock) and by endogenous fatty acids (Lennen et al., 2010). Over-expression of phage shock proteins stabilizes the membranes (Brissette et al., 1990) and increases cell viability under fatty acid overproduction (Lennen et al., 2011). This study also exhibited the increase (>1.5 fold) in the transcription of phage shock proteins encoded in *pspABCDE* and *pspG* during mid-log phase of fatty acid production. In addition, genes in the marA/rob/soxS regulon were activated by the endogenously produced fatty acid (Brissette et al., 1990). These genes might involve in resistance to oxidative stress, antibiotics and organic solvents (Zhang et al., 2012b). Our data only showed few genes (*spy*, *hdeA*, *hdeB* and *poxB*)

were up-regulated under fatty acid production (Table 1). Lennen *et al.* also reported the increase in genes involved in energy metabolism, which are in the *nuo* and *cyo* operon (Lennen et al., 2011). However, these genes were not perturbed significantly in our data. Whereas, some genes involved in amino acid ATP-binding cassette membrane transporter were up-regulated significantly (Table 1). The deviation in observation might be due to differences in growth phase, fermentation condition and strain genetic manipulation. The transcriptomic analysis unraveled the importance of cell membrane and acid stress as the target for future metabolic engineering effort to enhance the robustness of biocatalyst in fatty acid production. Moreover, genes related to colanic acid biosynthesis and biofilm formation are significantly up-regulated under fatty acid overproduction. These phenomena depicted by transcriptomics analysis will be investigated in our future study. Overall, the transcriptome profiling elucidates toxicity effects of endogenous fatty acid which is not captured by OptForce. Omics-based data and gene regulation could be incorporated into the next-generation OptForce to improve accuracy of predictions. Integration of omics data with computational tools can help in the engineering of robust biocatalysts for biorenewable chemicals production, and provide a paradigm to a shorter turnover for strain development and cost saving from an industrial standpoint.

Authors' contributions

JVS, CDM, LRJ and JAD conceived the project. TWT and JMY designed and performed the fermentation, metabolic flux analysis and metabolomics. EB, TWT, PL and JAD analyzed the transcriptomics analysis. AC and ARZ designed and performed the OptForce simulations. TWT, AC, EB and JVS analyzed the data and wrote the paper. All authors read and confirmed the manuscript.

Acknowledgement

This material is based upon work supported by the National Science Foundation under Award No. EEC-0813570

References

Ashburner, M., C. A. Ball, J. A. Blake, D. Botstein, H. Butler, J. M. Cherry, A. P. Davis, K. Dolinski, S. S. Dwight, J. T. Eppig, M. A. Harris, D. P. Hill, L. Issel-Tarver, A. Kasarskis, S. Lewis, J. C. Matese, J. E. Richardson, M. Ringwald, G. M. Rubin, G. Sherlock, and C. Gene Ontology, 2000, Gene Ontology: tool for the unification of biology: Nature Genetics, v. 25, p. 25-29.

Baba, T., T. Ara, M. Hasegawa, Y. Takai, Y. Okumura, M. Baba, K. A. Datsenko, M. Tomita, B. L. Wanner, and H. Mori, 2006, Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection: Mol Syst Biol, v. 2, p. 2006 0008.

Blattner, F. R., G. Plunkett, C. A. Bloch, N. T. Perna, V. Burland, M. Riley, J. ColladoVides, J. D. Glasner, C. K. Rode, G. F. Mayhew, J. Gregor, N. W. Davis, H. A. Kirkpatrick, M. A. Goeden, D. J. Rose, B. Mau, and Y. Shao, 1997, The complete genome sequence of Escherichia coli K-12: Science, v. 277, p. 1453-&.

Brissette, J. L., M. Russel, L. Weiner, and P. Model, 1990, PHAGE SHOCK PROTEIN, A STRESS PROTEIN OF ESCHERICHIA-COLI: Proceedings of the National Academy of Sciences of the United States of America, v. 87, p. 862-866.

Datsenko, K. A., and B. L. Wanner, 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, v. 97, p. 6640-6645.

Davis, M. S., and J. E. Cronan, 2001, Inhibition of Escherichia coli acetyl coenzyme A carboxylase by acyl-acyl carrier protein: Journal of Bacteriology, v. 183.

Davis, M. S., J. Solbiati, and J. E. Cronan, 2000, Overproduction of acetyl-CoA carboxylase activity increases the rate of fatty acid biosynthesis in Escherichia coli: Journal of Biological Chemistry, v. 275.

Dellomonaco, C., J. M. Clomburg, E. N. Miller, and R. Gonzalez, 2011, Engineered reversal of the beta-oxidation cycle for the synthesis of fuels and chemicals: Nature, v. 476.

Ewald, J. C., S. Heux, and N. Zamboni, 2009, High-Throughput Quantitative Metabolomics: Workflow for Cultivation, Quenching, and Analysis of Yeast in a Multiwell Format: Analytical Chemistry, v. 81.

Feist, A. M., C. S. Henry, J. L. Reed, M. Krummenacker, A. R. Joyce, P. D. Karp, L. J. Broadbelt, V. Hatzimanikatis, and B. O. Palsson, 2007, A genome-scale metabolic reconstruction for Escherichia coli K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information: Mol Syst Biol, v. 3, p. 121.

Fjerbaek, L., K. V. Christensen, and B. Norddahl, 2009, A Review of the Current State of Biodiesel Production Using Enzymatic Transesterification: Biotechnology and Bioengineering, v. 102.

Fujita, Y., H. Matsuoka, and K. Hirooka, 2007, Regulation of fatty acid metabolism in bacteria: Molecular Microbiology, v. 66, p. 829-839.

Goh, E.-B., E. E. K. Baidoo, J. D. Keasling, and H. R. Beller, 2012, Engineering of Bacterial Methyl Ketone Synthesis for Biofuels: Applied and Environmental Microbiology, v. 78, p. 70-80.

Han, M. J., and S. Y. Lee, 2006, The Escherichia coli proteome: Past, present, and future prospects: Microbiology and Molecular Biology Reviews, v. 70, p. 362-+.

Heath, R. J., and C. O. Rock, 1996a, Inhibition of beta-ketoacyl-acyl carrier protein synthase III (FabH) by acyl-acyl carrier protein in Escherichia coli: J Biol Chem, v. 271, p. 10996-1000.

Heath, R. J., and C. O. Rock, 1996b, Regulation of fatty acid elongation and initiation by acyl-acyl carrier protein in Escherichia coli: J Biol Chem, v. 271, p. 1833-6.

Hirasawa, T., C. Furusawa, and H. Shimizu, 2010, Saccharomyces cerevisiae and DNA microarray analyses: what did we learn from it for a better understanding and exploitation of yeast biotechnology?: Applied Microbiology and Biotechnology, v. 87.

Iwakura, M., J. Hattori, Y. Arita, M. Tokushige, and H. Katsuki, 1979, STUDIES ON REGULATORY FUNCTIONS OF MALIC ENZYMES .6. PURIFICATION AND MOLECULAR-PROPERTIES OF NADP-LINKED MALIC ENZYME FROM ESCHERICHIA-COLI W: Journal of Biochemistry, v. 85.

Jarboe, L. R., P. Liu, and L. A. Royce, 2011, Engineering inhibitor tolerance for the production of biorenewable fuels and chemicals: Current Opinion in Chemical Engineering, v. 1, p. 38-42.

Jing, F., D. C. Cantu, J. Tvaruzkova, J. P. Chipman, B. J. Nikolau, M. D. Yandeau-Nelson, and P. J. Reilly, 2011, Phylogenetic and experimental characterization of an acyl-ACP thioesterase family reveals significant diversity in enzymatic specificity and activity: Bmc Biochemistry, v. 12.

Kao, K. C., Y. L. Yang, R. Boscolo, C. Sabatti, V. Roychowdhury, and J. C. Liao, 2004, Transcriptome-based determination of multiple transcription regulator activities in Escherichia coli by using network component analysis: Proceedings of the National Academy of Sciences of the United States of America, v. 101, p. 641-646.

Kern, R., A. Malki, J. Abdallah, J. Tagourti, and G. Richarme, 2007, Escherichia coli HdeB is an acid stress chaperone: Journal of Bacteriology, v. 189, p. 603-610.

Kim, H. U., T. Y. Kim, and S. Y. Lee, 2008, Metabolic flux analysis and metabolic engineering of microorganisms: Molecular BioSystems, v. 4, p. 113-120.

Langmead, B., and S. L. Salzberg, 2012, Fast gapped-read alignment with Bowtie 2: Nature Methods, v. 9, p. 357-U54.

Lee, J. W., D. Na, J. M. Park, J. Lee, S. Choi, and S. Y. Lee, 2012, Systems metabolic engineering of microorganisms for natural and non-natural chemicals: Nature Chemical Biology, v. 8.

Legras, J. L., C. Erny, C. Le Jeune, M. Lollier, Y. Adolphe, C. Demuyter, P. Delobel, B. Blondin, and F. Karst, 2010, Activation of Two Different Resistance Mechanisms in Saccharomyces cerevisiae upon Exposure to Octanoic and Decanoic Acids: Applied and Environmental Microbiology, v. 76, p. 7526-7535.

Leighty, R. W., and M. R. Antoniewicz, 2012, Parallel labeling experiments with U-(13)C glucose validate E. coli metabolic network model for (13)C metabolic flux analysis: Metabolic engineering, v. 14.

Lennen, R. M., D. J. Braden, R. M. West, J. A. Dumesic, and B. F. Pfleger, 2010, A Process for Microbial Hydrocarbon Synthesis: Overproduction of Fatty Acids in Escherichia coli and Catalytic Conversion to Alkanes: Biotechnology and Bioengineering, v. 106.

Lennen, R. M., M. A. Kruziki, K. Kumar, R. A. Zinkel, K. E. Burnum, M. S. Lipton, S. W. Hoover, D. R. Ranatunga, T. M. Wittkopp, W. D. Marner, II, and B. F. Pfleger, 2011, Membrane Stresses Induced by Overproduction of Free Fatty Acids in Escherichia coli: Applied and Environmental Microbiology, v. 77.

Lennen, R. M., and B. F. Pfleger, 2012, Engineering Escherichia coli to synthesize free fatty acids: Trends in biotechnology, v. 30, p. 659-667.

Li, M., X. Zhang, A. Agrawal, and K.-Y. San, 2012, Effect of acetate formation pathway and long chain fatty acid CoA-ligase on the free fatty acid production in E. coli expressing acy-ACP thioesterase from Ricinus communis: Metabolic Engineering, v. 14, p. 380-387.

Liu, P., A. Chernyshov, T. Najdi, Y. Fu, J. Dickerson, S. Sandmeyer, and L. Jarboe, 2013, Membrane stress caused by octanoic acid in Saccharomyces cerevisiae: Applied Microbiology and Biotechnology, v. 97, p. 3239-3251.

Liu, T., H. Vora, and C. Khosla, 2010, Quantitative analysis and engineering of fatty acid biosynthesis in E. coli: Metab Eng, v. 12, p. 378-86.

Lu, X., H. Vora, and C. Khosla, 2008, Overproduction of free fatty acids in E. coli: Implications for biodiesel production: Metabolic Engineering, v. 10.

Maere, S., K. Heymans, and M. Kuiper, 2005, BiNGO: a Cytoscape plugin to assess overrepresentation of Gene Ontology categories in Biological Networks: Bioinformatics, v. 21, p. 3448-3449.

Mashego, M. R., K. Rumbold, M. De Mey, E. Vandamme, W. Soetaert, and J. J. Heijnen, 2007, Microbial metabolomics: past, present and future methodologies: Biotechnology Letters, v. 29.

Matsuoka, Y., and K. Shimizu, 2010, Current status of C-13-metabolic flux analysis and future perspectives: Process Biochemistry, v. 45, p. 1873-1881.

Ranganathan, S., P. F. Suthers, and C. D. Maranas, 2010, OptForce: an optimization procedure for identifying all genetic manipulations leading to targeted overproductions: PLoS Comput Biol, v. 6, p. e1000744.

Ranganathan, S., T. W. Tee, A. Chowdhury, A. R. Zomorrodi, J. M. Yoon, Y. Fu, J. V. Shanks, and C. D. Maranas, 2012, An integrated computational and experimental study for overproducing fatty acids in Escherichia coli: Metabolic Engineering, v. 14, p. 687-704.

Richard, H., and J. W. Foster, 2004, Escherichia coli glutamate- and arginine-dependent acid resistance systems increase internal pH and reverse transmembrane potential: Journal of Bacteriology, v. 186, p. 6032-6041.

Royce, L. A., P. Liu, M. J. Stebbins, B. C. Hanson, and L. R. Jarboe, 2013, The damaging effects of short chain fatty acids on Escherichia coli membranes: Applied Microbiology and Biotechnology, v. 97, p. 8317-8327.

San, K.-Y., and M. Li, 2012, Methods to produce fatty acids from renewable carbon sources, USA.

Serrania, J., F.-J. Vorhoelter, K. Niehaus, A. Puehler, and A. Becker, 2008, Identification of Xanthomonas campestris pv. campestris galactose utilization genes from transcriptome data: Journal of Biotechnology, v. 135.

Smoot, M. E., K. Ono, J. Ruscheinski, P.-L. Wang, and T. Ideker, 2011, Cytoscape 2.8: new features for data integration and network visualization: Bioinformatics, v. 27, p. 431-432.

Sriram, G., D. B. Fulton, V. V. Iyer, J. M. Peterson, R. L. Zhou, M. E. Westgate, M. H. Spalding, and J. V. Shanks, 2004, Quantification of compartmented metabolic fluxes in developing soybean embryos by employing Biosynthetic ally directed fractional C-13 labeling, C-13, H-1 two-dimensional nuclear magnetic resonance, and comprehensive isotopomer balancing: Plant Physiology, v. 136.

Sriram, G., D. B. Fulton, and J. V. Shanks, 2007, Flux quantification in central carbon metabolism of Catharanthus roseus hairy roots by C-13 labeling and comprehensive bondomer balancing: Phytochemistry, v. 68.

Steen, E. J., Y. Kang, G. Bokinsky, Z. Hu, A. Schirmer, A. McClure, S. B. del Cardayre, and J. D. Keasling, 2010, Microbial production of fatty-acid-derived fuels and chemicals from plant biomass: Nature, v. 463.

Suthers, P. F., A. Zomorrodi, and C. D. Maranas, 2009, Genome-scale gene/reaction essentiality and synthetic lethality analysis: Mol Syst Biol, v. 5, p. 301.

Trapnell, C., A. Roberts, L. Goff, G. Pertea, D. Kim, D. R. Kelley, H. Pimentel, S. L. Salzberg, J. L. Rinn, and L. Pachter, 2012, Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks: Nature Protocols, v. 7, p. 562-578.

Yamaguchi, M., 1979, STUDIES ON REGULATORY FUNCTIONS OF MALIC ENZYMES .4. EFFECTS OF SULFHYDRYL-GROUP MODIFICATION ON THE CATALYTIC FUNCTION OF NAD-LINKED MALIC ENZYME FROM ESCHERICHIA-COLI: Journal of Biochemistry, v. 86.

Yano, S., T. Asano, N. Kurose, J. Hiramatsu, H. Shimoi, and K. Ito, 2003, Characterization of an alpha-ketoglutarate-resistant sake yeast mutant with high organic acid productivity: Journal of Bioscience and Bioengineering, v. 96, p. 332-336.

Zhang, F., J. M. Carothers, and J. D. Keasling, 2012a, Design of a dynamic sensor-regulator system for production of chemicals and fuels derived from fatty acids: Nature Biotechnology, v. 30.

Zhang, F., M. Ouellet, T. S. Batth, P. D. Adams, C. J. Petzold, A. Mukhopadhyay, and J. D. Keasling, 2012b, Enhancing fatty acid production by the expression of the regulatory transcription factor FadR: Metabolic Engineering, v. 14, p. 653-660.

Zhang, W. W., F. Li, and L. Nie, 2010, Integrating multiple 'omics' analysis for microbial biology: application and methodologies: Microbiology-Sgm, v. 156, p. 287-301.

Zhang, X., A. Agrawal, and K.-Y. San, 2012c, Improving fatty acid production in escherichia coli through the overexpression of malonyl coA-Acyl carrier protein transacylase: Biotechnology Progress, v. 28.

Zhang, X., M. Li, A. Agrawal, and K.-Y. San, 2011, Efficient free fatty acid production in Escherichia coli using plant acyl-ACP thioesterases: Metabolic Engineering, v. 13.

List of figures and tables

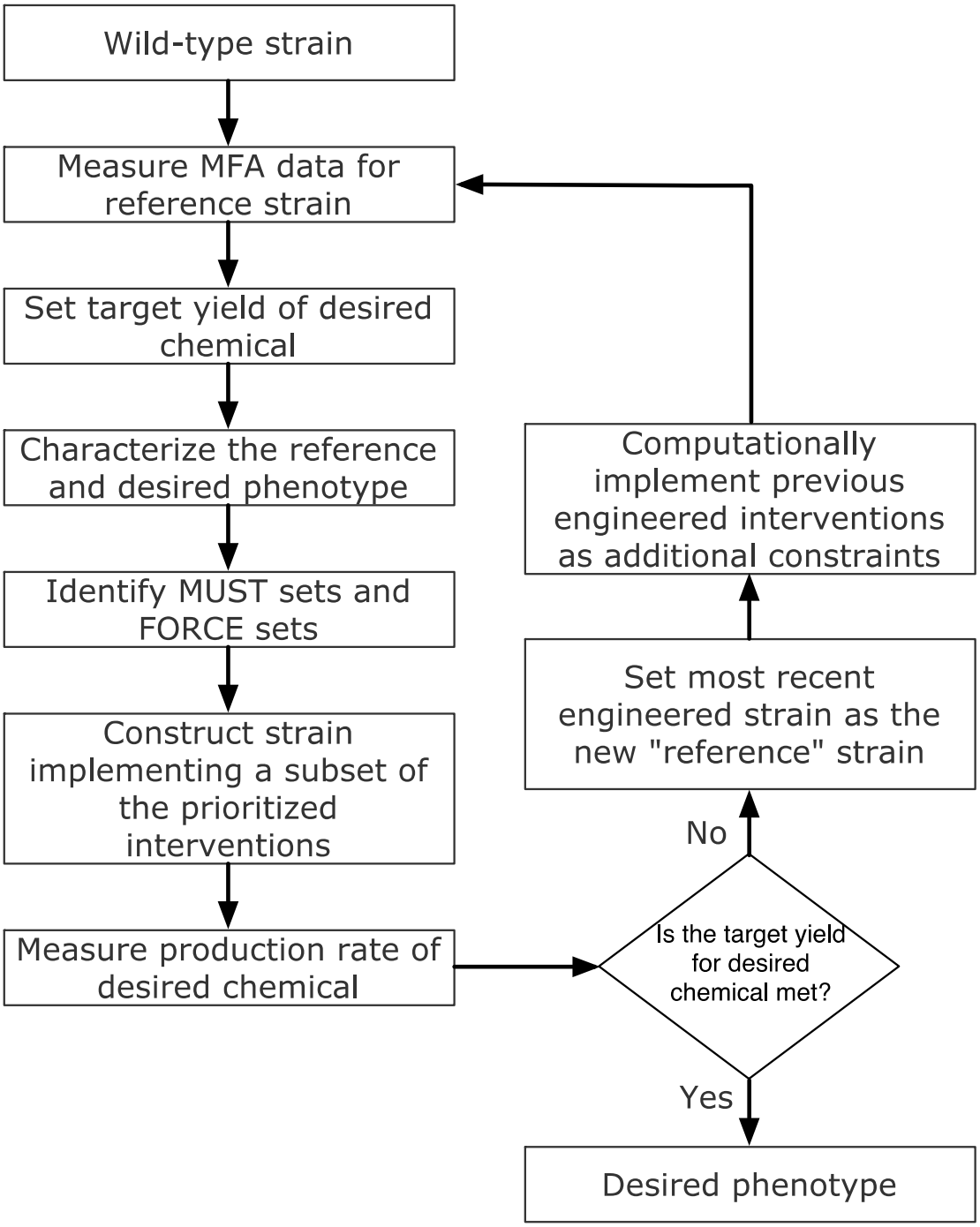


Figure 1. Workflow for iterative re-deployment of OptForce for improving strain characterization and accuracy of OptForce predictions.

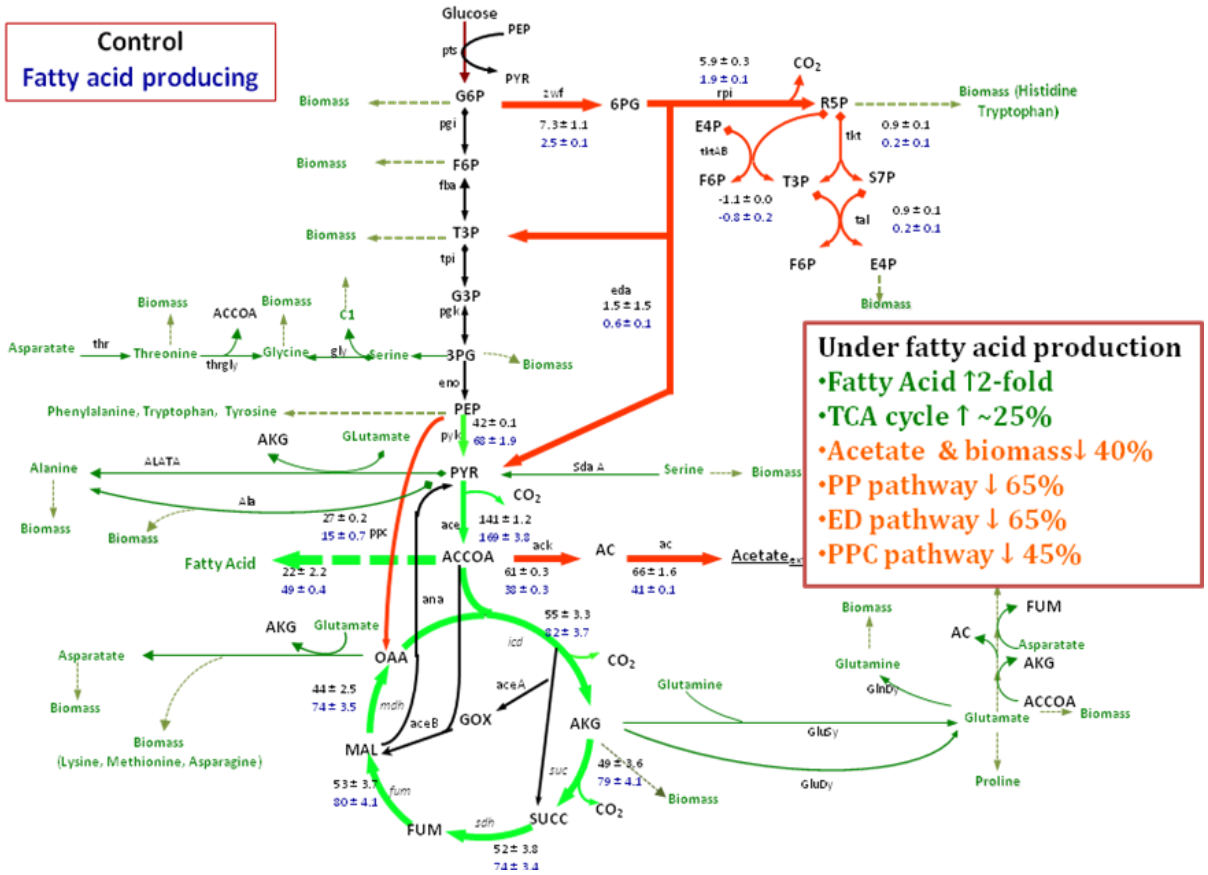


Figure 2. *In vivo* metabolic flux distribution for *E. coli* ML103 pXZ18Z under control condition (black font) and fatty acid producing condition (blue font) as calculated via isotopomer balancing using NMR2Flux software. Estimated fluxes are normalized to 100 mmol/g DCW.hr based on the average specific glucose uptake rates. The colored arrows indicate a significant difference (>20%) between the control and fatty acid production condition. The flux values shown are average from two replicates ± standard deviation.

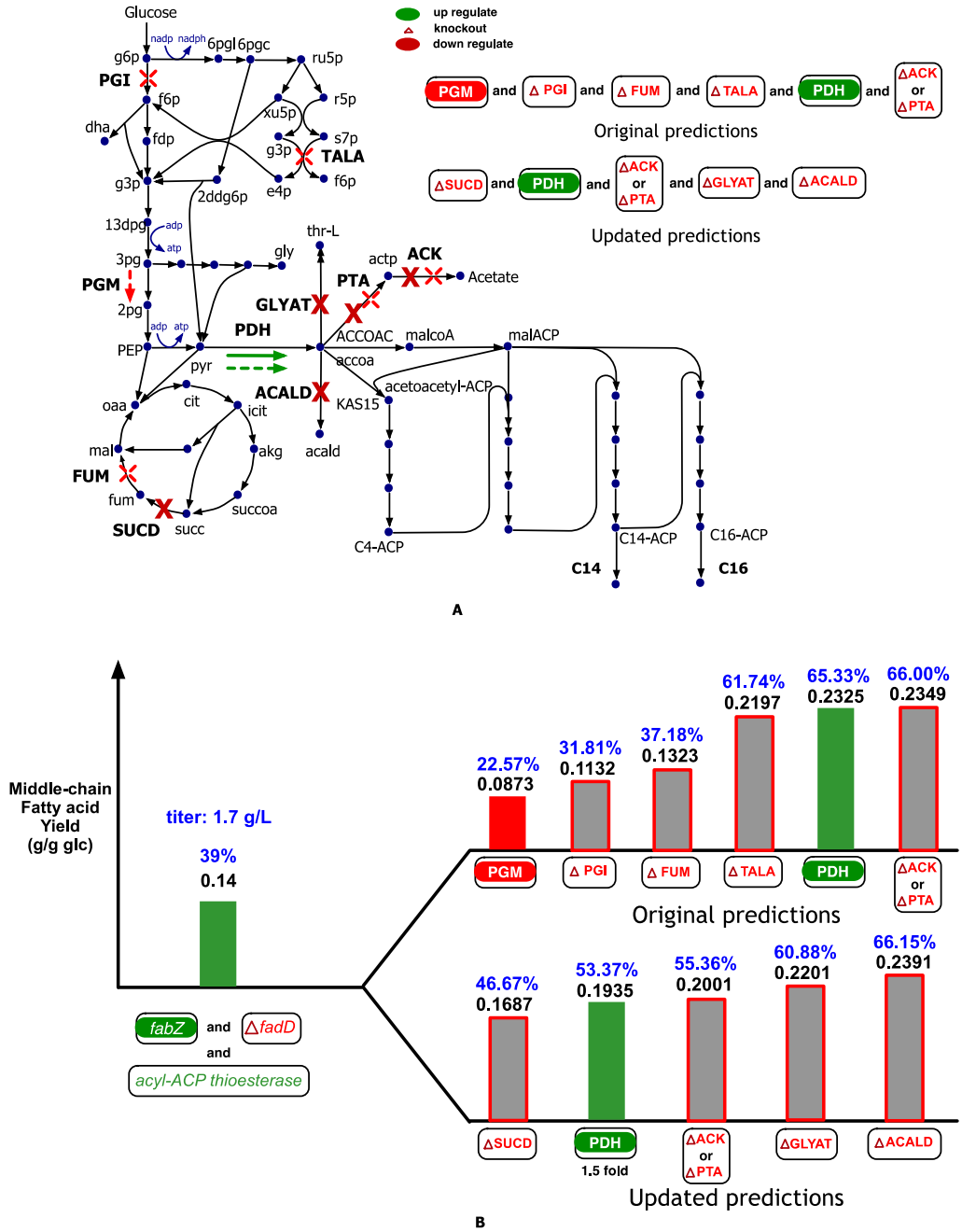


Figure 3. Original and updated OptForce interventions and their impact on product yield for overproduction of medium-chain fatty acids in *E. coli*. A) Original and updated interventions as predicted by OptForce. The dotted and solid arrows indicate previous and updated interventions respectively. B) Impact of each intervention of medium-chain fatty acid yield. The interventions on top and bottom represent original and updated OptForce predictions respectively.

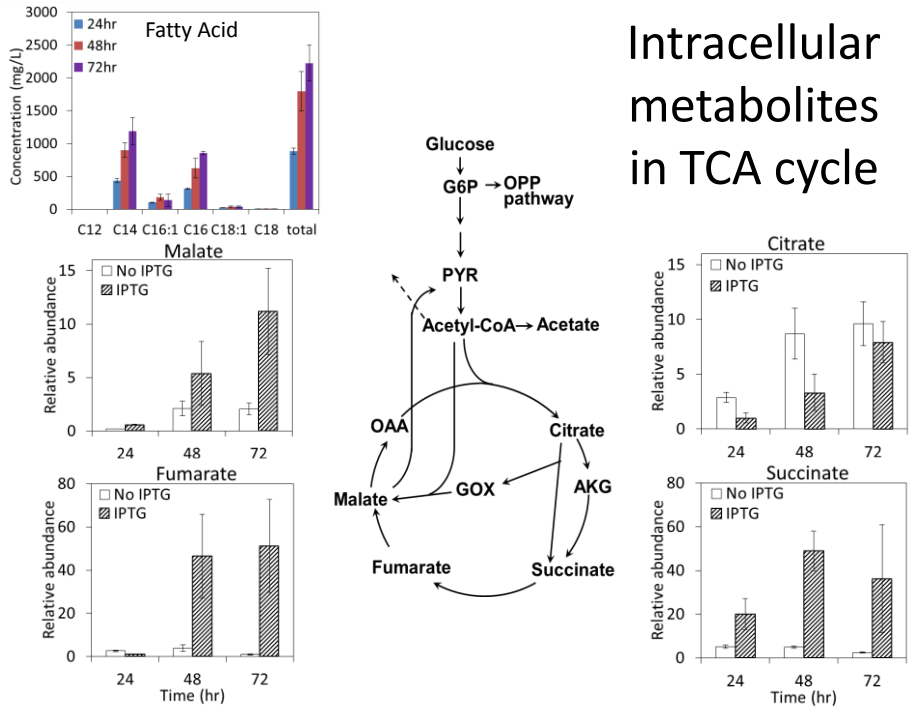
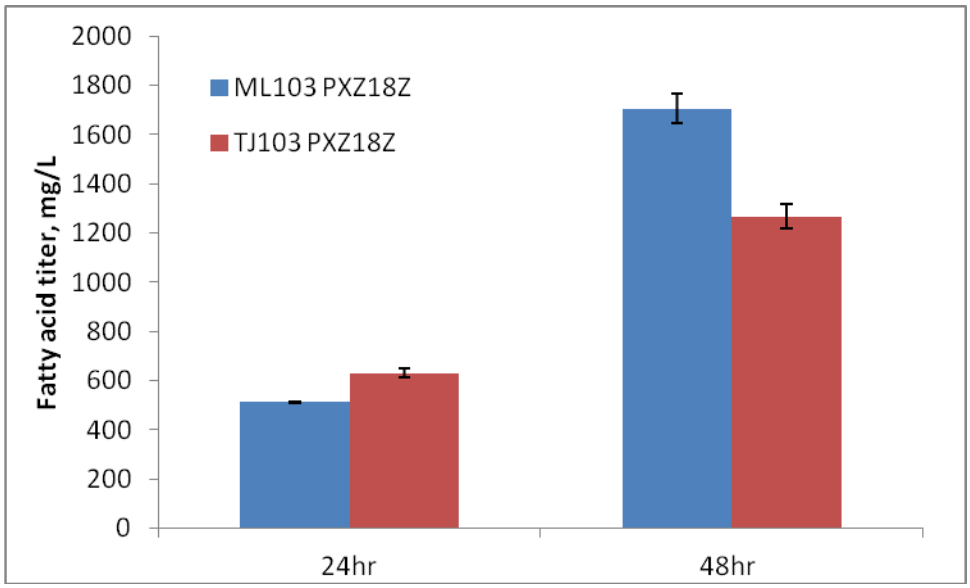


Figure 4. (a) The total free fatty acid accumulation of ML103 pXZ18Z (Δ*fadD*, *fabZ*$^{++}$, thioesterase^{++}) under IPTG induction. The total fatty acids in uninduced strain were always less than 300mg/L. (b) The relative concentration of intracellular metabolites in the TCA cycle of *E. coli* ML103 pXZ18Z (Δ*fadD*, *fabZ*$^{++}$, thioesterase^{++}) under control condition and fatty acid producing (induced by IPTG) condition. The concentration of intracellular metabolites are relative to the internal standard norleucine (5ng/mL). The strains were grown in shake flasks using M9 minimal medium with 1.5% glucose at 30^{o}C. Error bars represent standard deviation of at least triplicate cultures.

(a)



(b)

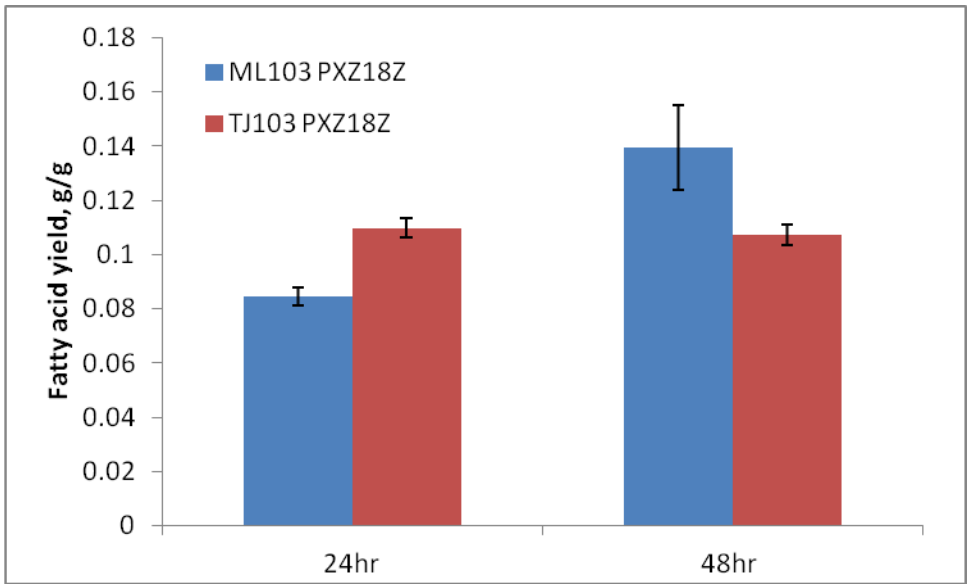


Figure 5. (a) Accumulation of free fatty acids and (b) fatty acid yield over glucose by ML103pXZ18z (Δ*fadD*, *fabZ*$^{+}$) and TJ103 PXZ18Z (Δ*fadD*, *fabZ*$^{++}$, Δ*sdhABCD*). The strains were grown in shake flasks in M9 minimal medium with 1.5% glucose at 30^{o}C and 250rpm, and sampled at 24 and 48 hours. Error bars represent standard deviation of at least triplicate cultures.

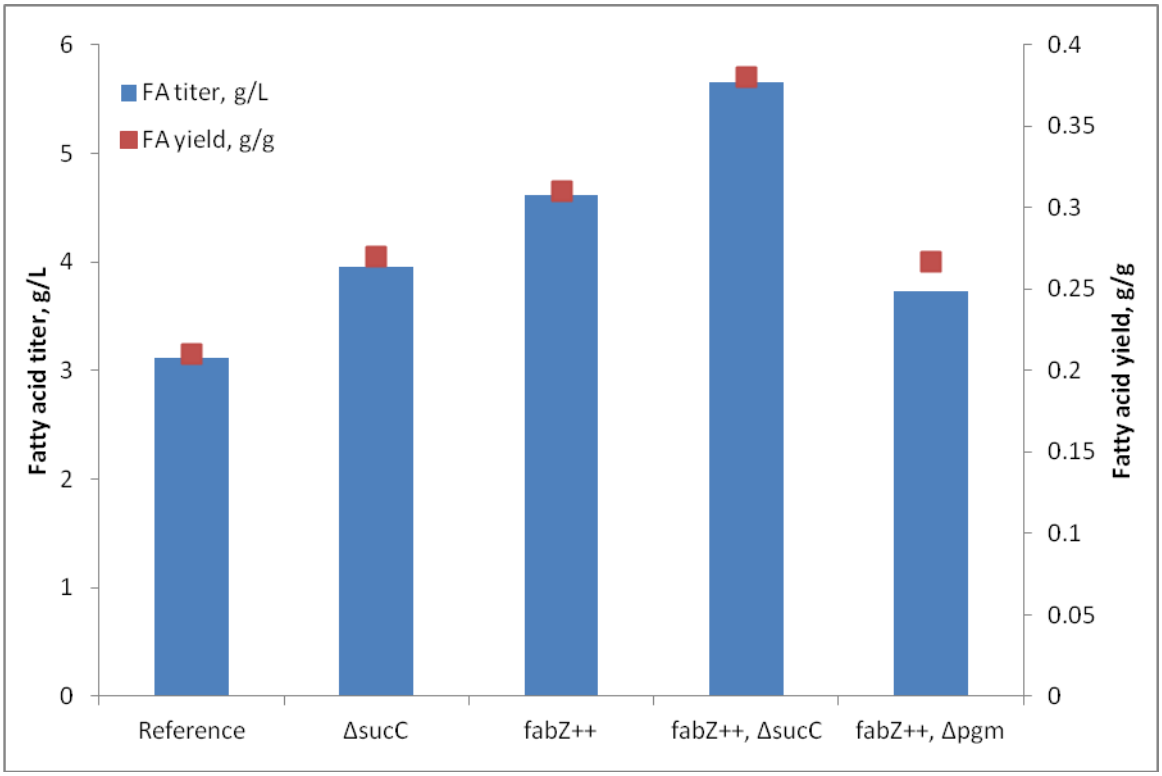


Figure 6. Effect of different genetic modifications on the fatty acid titer and yield reported by (San and Li, 2012). All the genetic modifications were carried out in *E. coli* strain ML103 (Δ*fadD*). An acyl-ACP thioesterase (pXZ18) was overexpressed in engineered strains to test the effect of the gene knockout (Δ) or overexpression (++). The strains were cultured in LB media with 1.5% glucose and sampled at 48 hours.

Table 1. Relative gene transcript changes measured by RNAseq for ML103 pXZ18Z under the control versus fatty acid over-production condition. The genes were selected based on their function group, which was referred from Ecocyc data base. The magnitude of transcriptional fold change was color-coded, in which red and green represent an increase and a decrease in gene expression respectively. q-value represents False Discovery Rate (FDR) corrected p-value. The strain was cultured in M9 minimal medium with 1.5% glucose in fermentors.

gene	b-num	log2(fold_change)	p-value	q-value
Colanic Acid Production				
wcaE	b2055	6.68	0.000	0.000
wcaF	b2054	5.82	0.000	0.000
wcaD	b2056	5.55	0.000	0.001
wcaI	b2050	5.53	0.000	0.000
wcaB	b2058	5.40	0.010	0.154
wcaJ	b2047	5.40	0.000	0.000
wcaC	b2057	5.37	0.000	0.002
wcaA	b2059	5.27	0.000	0.000
wcaK	b2045	4.91	0.000	0.002
wcaL	b2044	3.55	0.001	0.037
wcaM	b2043	3.14	0.000	0.002
wzb	b2061	5.74	0.000	0.000
wzc	b2060	5.42	0.000	0.000
wza	b2062	5.42	0.000	0.000
wzxC	b2046	4.92	0.000	0.000
gmm	b2051	6.66	0.000	0.000
gmd	b2053	6.17	0.000	0.000
cpsG	b2048	5.82	0.000	0.000
cpsB	b2049	5.25	0.000	0.000
fcl	b2052	6.57	0.000	0.000
mcbA	b0806	3.74	0.000	0.001
Biofilm Formation				
yjbE	b4026	8.34	0.000	0.000
bdm	b1481	5.28	0.000	0.000
ydeI	b1536	4.00	0.000	0.000
ychH	b1205	3.34	0.000	0.009
ycfJ	b1110	3.09	0.000	0.013
ariR	b1166	3.14	0.000	0.019
hha	b0460	1.72	0.029	0.286
tomB	b0461	2.26	0.015	0.196
bssR	b0836	3.68	0.001	0.035
bssS	b1060	2.34	0.022	0.248

Table 1 continued

Stress Response				
ydeI	b1536	4.00	0.000	0.000
ychH	b1205	3.34	0.000	0.009
ygiW	b3024	3.09	0.001	0.033
yodD	b1953	3.07	0.001	0.038
gadA	b3517	3.32	0.004	0.088
gadB	b1493	3.27	0.007	0.126
gadC	b1492	3.22	0.005	0.108
gadE	b3512	2.88	0.003	0.064
gadW	b3515	1.75	0.059	0.423
gadY	b4452	1.65	0.106	0.550
gadX	b3516	2.12	0.026	0.266
hdeD	b3511	2.43	0.008	0.141
hdeA	b3510	2.30	0.031	0.297
hdeB	b3509	2.28	0.027	0.273
Membrane Protein				
wza	b2062	5.42	0.000	0.000
yneM	b4599	4.42	0.000	0.000
yohC	b2135	4.38	0.000	0.001
ymgE	b1195	3.65	0.000	0.002
yjiY	b4354	3.37	0.000	0.002
yaiY	b0379	3.16	0.000	0.005
yibI	b3598	3.01	0.000	0.010
yibH	b3597	2.78	0.001	0.028
yqaE	b2666	2.74	0.001	0.025
yfeN	b2408	2.52	0.001	0.036
yncL	b4598	2.45	0.001	0.035
b1256	ompW	-3.07	0.000	0.002
b0929	ompF	-3.26	0.007	0.120
Fatty Acid Biosynthesis				
b0180	fabZ	2.62	0.015	0.195
b0452	tesB	2.00	0.011	0.164
Phage Shock Protein				
pspA	b1304	2.69	0.009	0.143
pspB	b1305	2.42	0.092	0.517
pspC	b1306	2.62	0.037	0.316
pspD	b1307	2.29	0.004	0.085
pspE	b1308	1.64	0.033	0.302
pspG	b4050	1.63	0.034	0.305
MarA/Rob/SoxS regulon				

Table 1 continued

hdeB	b3509	2.27	0.027	0.272
hdeA	b3510	2.30	0.032	0.295
poxB	b0871	2.05	0.046	0.364
spy	b1743	4.00	0.000	0.000
Energy Metabolism/Transporter				
ddpA	b1487	4.54	0.000	0.000
glnH	b0811	4.19	0.000	0.007
argT	b2310	4.11	0.000	0.005
ydcS	b1440	3.77	0.000	0.007
ydcT	b1441	3.21	0.000	0.011
yhdW	b3268	5.22	0.000	0.000
yhdX	b3269	3.14	0.000	0.001
yhdY	b3270	2.57	0.000	0.021
yhdZ	b3271	2.04	0.004	0.086
rbsA	b3749	-2.51	0.001	0.033
artJ	b0860	-3.06	0.001	0.022

Table 2. Transcription factors significantly perturbed by IPTG induction for fatty acid production as determined by network component analysis

TF	TFA125	TFA381	Description
ArgR	2.78	2.78	Arginine repressor
CRP	3.05	3.05	cAMP receptor protein/catabolite gene activator protein
FlhDC	-2.51	-2.51	FlhDC DNA-binding transcriptional dual regulator
FNR	-3.23	-3.23	FNR DNA-binding transcriptional dual regulator
Fur	3.08	3.08	Ferric Uptake Regulation
H-NS	-2.42	-2.42	Histone-like nucleoid structuring protein
LexA	4.44	4.22	LexA DNA-binding transcriptional repressor
NtrC	4.99	4.61	NtrC transcriptional dual regulator
RcsAB	2.78	2.78	Regulator capsule synthesis B

Supporting information

Figure S1. Selected physiological states comparison of *E. coli* ML103 pXZ18Z (Δ*fadD*, *fabZ*$^{++}$ and thioesterase^{++}) under control condition and fatty acid producing (induced by 1mM IPTG) condition. The strains were grown aerobically in fementor in M9 minimal medium with 1% glucose at 30^{o}C. Error bars represent standard deviation of at least triplicate cultures. An asterisk indicates significantly different values from the control (p-value <0.05).

Figure S2. Cofactor (NADH, NADPH) and energy (ATP) production in central carbon metabolism. The red triangle represents NADH, the purple triangle represents NADPH and the green rectangular represents ATP. The total production in central carbon metabolism for control and fatty acid producing conditions is show in the graph. Malic enzyme activity involves production of both NADH and NADPH, the overall reducing power production can be captured by NAD(P)H.

Figure S3. Relative concentration of intracellular free amino acids of *E. coli* ML103 pXZ18Z (Δ*fadD*, *fabZ*$^{++}$, thioesterase^{++}) under control condition and fatty acid producing (induced by 1mM IPTG) condition. The strains were grown in shaker flasks using M9 minimal medium with 1.5% glucose at 30^{o}C. Error bars represent standard deviation of at least triplicate cultures.

Figure S4. Accumulation of acetate, succinate and pyruvate by TJ103 pXZ18Z (Δ*sdhABCD*, Δ*fadD*, *fabZ*$^{++}$, thioesterase^{++}). The strains were grown in shake flasks in M9 minimal medium with 1.5% glucose at 30^{o}C and 250rpm, and sampled at 24 and 48 hours. Error bars represent standard deviation of at least triplicate cultures. Asterisk represents p –value <0.05

Table S1. The primers used for TJ103 strain construction with succinate dehydrogenase (*sdhABCD)* and acyl-CoA synthetase (*fadD)* knockout

Table S2. Comparison of flux values of *E.coli* ML103 pXZ18Z between the control and fatty acid producing condition

Table S3. Measured and simulated isotopomer distributions of proteinogenic amino acids from 2D-NMR

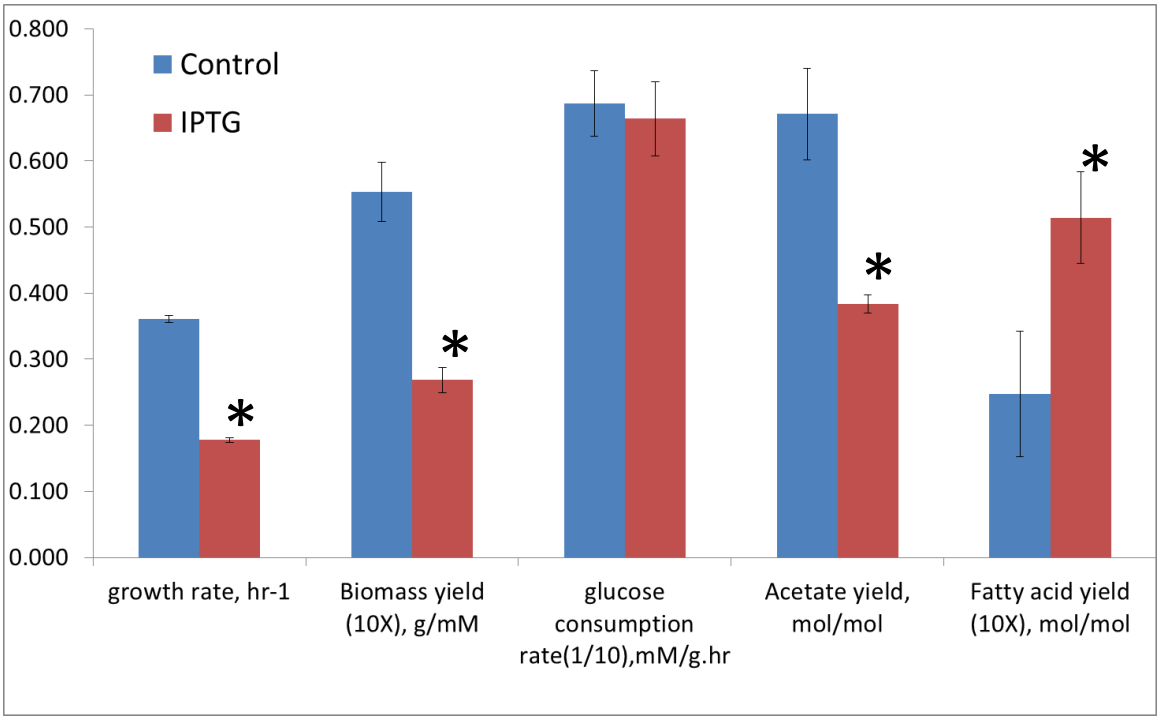


Figure S1. Selected physiological states comparison of *E. coli* ML103 pXZ18Z (Δ*fadD*, *fabZ*$^{++}$ and thioesterase^{++}) under control condition and fatty acid producing (induced by 1mM IPTG) condition. The strains were grown aerobically in fementor in M9 minimal medium with 1% glucose at 30^{o}C. Error bars represent standard deviation of at least triplicate cultures. An asterisk indicates significantly different values from the control (p-value <0.05).

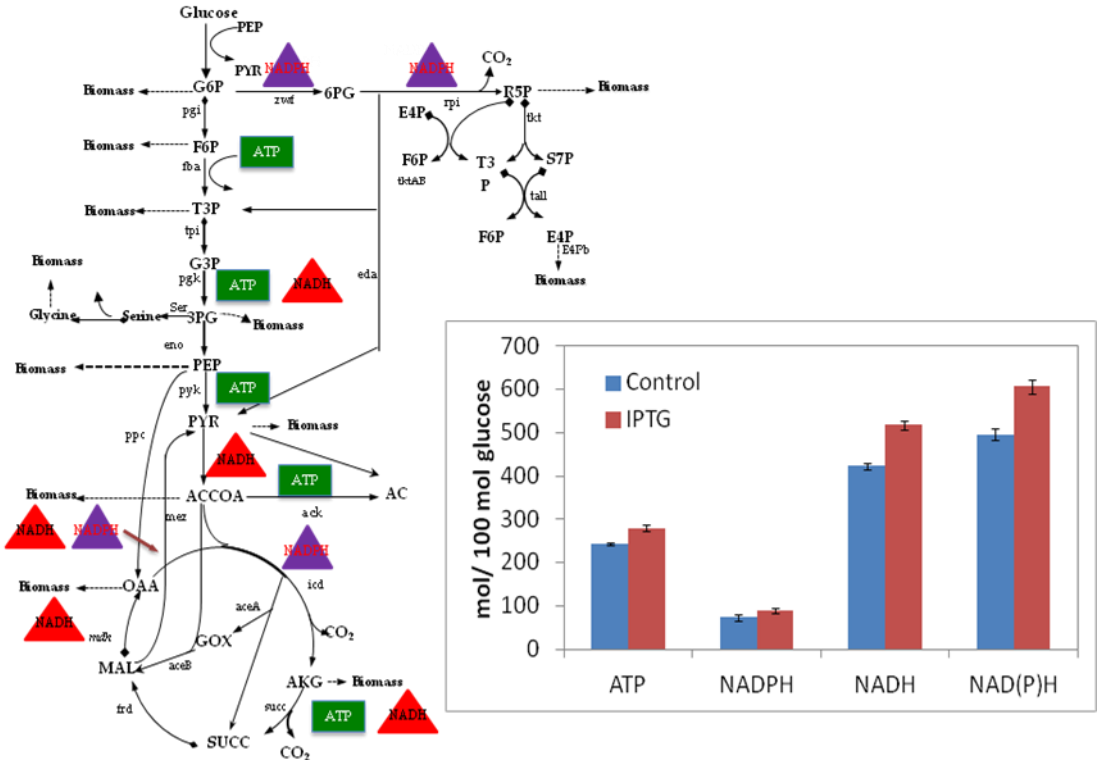


Figure S2 Cofactor (NADH, NADPH) and energy (ATP) production in central carbon metabolism. The red triangle represents NADH, the purple triangle represents NADPH and the green rectangular represents ATP. The total production in central carbon metabolism for control and fatty acid producing conditions is show in the graph. Malic enzyme activity involves production of both NADH and NADPH, the overall reducing power production can be captured by NAD(P)H.

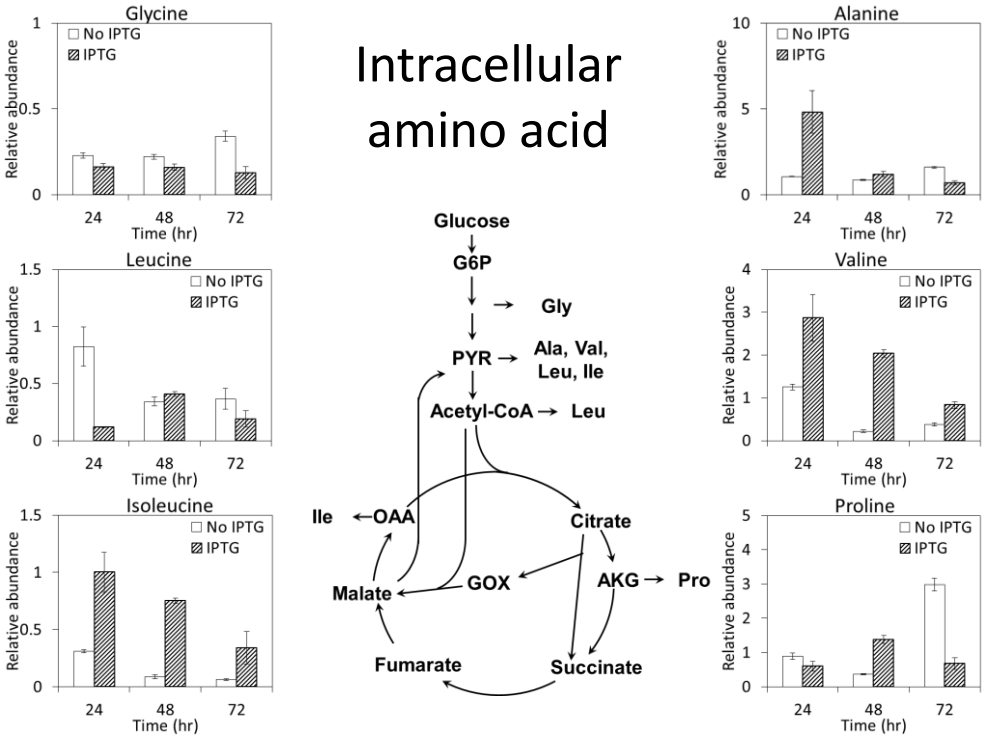


Figure S3. Relative concentration of intracellular free amino acids of *E. coli* ML103 pXZ18Z (Δ*fadD, fabZ*$^{++}$, thioesterase^{++}) under control condition and fatty acid producing (induced by 1mM IPTG) condition. The strains were grown in shaker flasks using M9 minimal medium with 1.5% glucose at 30^{o}C. Error bars represent standard deviation of at least triplicate cultures.

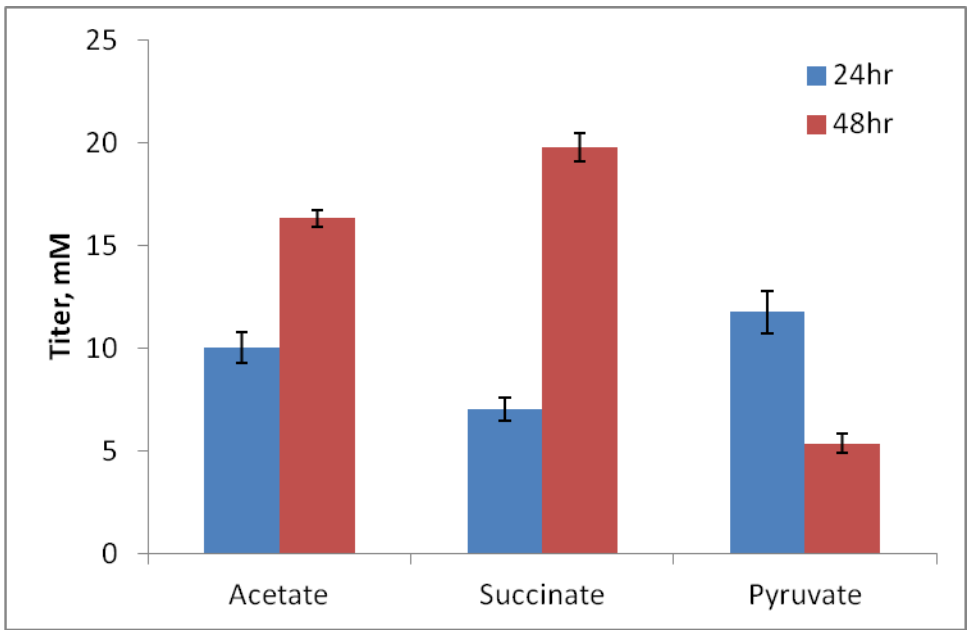


Figure S4. Accumulation of acetate, succinate and pyruvate by TJ103 pXZ18Z (Δ*sdhABCD*, Δ*fadD*, *fabZ*$^{++}$, thioesterase^{++}). The strains were grown in shake flasks in M9 minimal medium with 1.5% glucose at 30^{o}C and 250rpm, and sampled at 24 and 48 hours. Error bars represent standard deviation of at least triplicate cultures. Asterisk represents p –value <0.05.

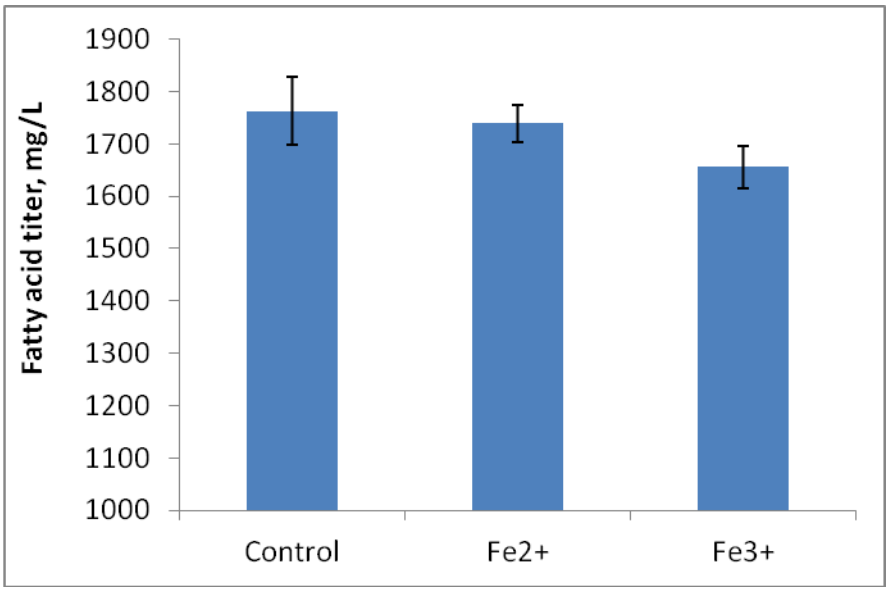


Figure S5. Accumulation of free fatty acids by ML103 pXZ18z (ΔfadD, fabZ^{+}) with the supplementation of 0.5mM ferrous chloride (Fe2+) and 0.5mM ferrous chloride (FE3+). The strains were grown in shake flasks in M9 minimal medium with 1.5% glucose at 30^{o}C and 250rpm, and sampled at 24 and 48 hours. Error bars represent standard deviation of at least triplicate cultures. Asterisk represents p –value <0.05.

Table S1. The primers used for TJ103 strain construction with succinate dehydrogenase (*sdhABCD)* and acyl-CoA synthetase (*fadD)* knockout

Primer	Sequence
sdh-UpF	(5'-AAGAGGGGAAAACCTGGGTA-3')
Sdh-UpR	(5'-GAATAACGCCCACATGCTGT-3')
Sdh-DnF	(5'- AAGCGTCGCATCAGGCAAC-3')
Sdh-DnR	(5'-GGTCGGAGATCGTTGAAGAG-3')
Sdh-F	(5'-TGTGCCCGTAGTCCCCAGGGAATAATAAGAAC AGCATGTGGGGTTATTCATGGGAATTAGCCATG GTCC-3')
Sdh-R	(5'-CGGCACTGGTTGCCTGATGCGACGCTTGCGCG TCTTATCAGGCCTACGGTGTGTAGGCTGGAGCT GCTTC-3')
P1	(5'-ACCCAGCGTTGTAACGTGTC-3')
P4	(5'- ACTTCGCCGTGGATACTACCA-3')

Table S2. Metabolic pathways represented by stoichiometry, atom transition, and involved genes from Ecocyc database (http://ecocyc.org).

Reaction name	Stoichiometry chemistry	Atom transition	Gene
Glycolysis pathway			
pts	Glu+PEP → G6P+PYR	abcdef+ABC → abcdef+ABC	ptsG,manZ,ptsH,ptsP
pgi	G6P → F6P	abcdef → abcdef	Pgi
fbp	F6P → FBP	abcdef → abcdef	pfkA,pfkB
fba	FBP → T3P+T3P	abcdef → cba+def	fbaA,fbaB
tpi	T3P → G3P	abc → abc	tpiA
pgk	G3P → 3PG	abc → abc	pgk
eno	3PG → PEP	abc → abc	eno
pyk	PEP → PYR	abc → abc	pykF/pykA
Entner Doudoroff pathway			
eda	6PG → PYR+T3P	abcdef → abc+def	eda
Pentose phosphate pathway			
zwf	G6P → 6PG	abcdef → abcdef	zwf
rpi	6PG → R5P+CO2	abcdef → bcdef+a	gnd,rpe,rpiA,rpiB
tkt	R5P+R5P → S7P+T3P	abcde+ABCDE → abABCDE+cde	tktA,tktB
tktAB	R5P+E4P → F6P+T3P	abcde+ABCD → abABCD+cde	tktA,tktB
talf	S7P+T3P → F6P+E4P	abcdefg+ABC → abcABC+defg	talA,talB
TCA cycle			
ace	PYR → ACCOA+CO2	abc → bc+a	lpd,aceF,aceE
icd	ACCOA+OAA → AKG+CO2	ab+ABCD → DCBba+A	icdB,icd
suc	AKG → SUCC+CO2	abcde → bcde+a	lpd,sucA,sucB
frd	SUCC → FUM	abcd → abcd	sdhA,sdhB,sdhC,sdhD
fum	FUM → MAL	abcd → abcd	fumA,fumB,fumC
mdh	MAL → OAA	abcd → abcd	mqo,mdh

Table S2 continued

Acetate production			
ackf	ACCOA → AC	ab → ab	pta,ackA,acs
Pox	PYR → AC + CO2	abc →bc +a	poxB
Anaplerotic pathway			
ppc	PEP+CO2 → OAA	abc+A → abcA	ppc
aceA	ACCOA+OAA → GOx+SUCC	ab+ABCD → DC+ABba	aceA
aceB	ACCOA+GOX → MAL	ab+AB → Abba	aceB
ana	Mal → PYR+CO2	abcd → abc+d	maeA,maeB
C1 metabolism			
ser	3PG → Ser	abc → abc	serA,serC,serB
gly	Ser → Gly+C1	abc → ab+c	glyA
Amino acid biosynthesis and metabolic pathway			
SdaRf	Ser → PYR	abc → abc	sdaA,sdaB,tdcB,tdcG
thr	Asp → Thr	abcd → abcd	thrC
thrgly	Thr → Gly+ACCOA	abcd → ab+cd	ItaE
GluDy	AKG → Glut	abcde → abcde	gdhA
GlnDy	Glut → Gln	abcde → abcde	glnA
GLUSy	AKG+Gln → Glut+Glut	abcde+ABCDE → abcde+ABCDE	gltB,gltD
ALATA	AKG+Ala → Glut+PYR	abcde+ABC → abcde+ABC	alaA,alaB,alaC
ASPTA	AKG+Asp → Glut+OAA	abcde+ABCD → abcde+ABCD	aspC
Ala	Ala → PYR	abc → abc	dadA
Pro	Glut → Pro	abcde → abcde	proC
argA	Glut+ACCOA →acGlut	abcde+AB → abcdeAB	argA
argD	acGlut+Glut→acOrn+AKG	abcdefg+ABCDE →abcdefg+ABCDE	argD
argE	acOrn→Orn+AC	abcdefg →abcde+fg	argE
argH	Orn+Asp→Arg+Fum	abcde+ABCD →abcde+ABCD	argH

Table S3. Comparison of flux values of *E. coli* ML103 pXZ18Z (ΔfadD, fabZ+, thioesterase+) under control and fatty acid producing (induced by 1mM IPTG) condition. The flux values are based on 100 mol/g/hr glucose uptake rate. The average and standard deviation is obtained from a duplicate of 13C flux experiments

reaction name	stoichiometry chemistry	Control		1mM IPTG	
		Average	SD	Average	SD
Glycolysis pathways					
pts	Glu+PEP → G6P+PYR	100.00	0.00	100.00	0.00
pgi	G6P → F6P	91.60	1.11	96.93	0.02
	reversibility	0.27	0.15	0.14	0.11
fbp	F6P → FBP	90.98	1.09	96.18	0.26
	reversibility	0.46	0.20	0.57	0.30
fba	FBP → T3P+T3P	90.98	1.09	96.18	0.26
	reversibility	0.40	0.24	0.57	0.22
tpi	T3P → G3P	181.62	0.74	191.77	0.84
	reversibility	0.37	0.08	0.51	0.09
pgk	G3P → 3PG	181.62	0.74	191.77	0.84
	reversibility	0.37	0.08	0.51	0.09
eno	3PG → PEP	172.34	0.08	184.83	1.05
	reversibility	0.43	0.05	0.35	0.09
pyk	PEP → PYR	42.74	0.05	67.85	1.88
	reversibility	0.76	0.03	0.77	0.04
Entner Doudoroff pathway					
eda	6PG → PYR+T3P	1.48	1.45	0.59	0.14
Pentose phosphate pathway					
zwf	G6P → 6PG	7.34	1.14	2.47	0.05
rpi	6PG → R5P+CO2	5.86	0.31	1.89	0.10
tkt	R5P+R5P → S7P+T3P	0.88	0.04	0.22	0.05
	reversibility	0.71	0.32	0.94	0.01
tktAB	R5P+E4P → F6P+T3P	-1.14	0.04	-0.79	0.15
	reversibility	0.47	0.49	0.14	0.04
talf	S7P+T3P → F6P+E4P	0.88	0.04	0.22	0.05
	reversibility	0.31	0.39	0.33	0.22
TCA cycle					
ace	PYR → ACCOA+CO2	141.47	1.17	169.65	3.81
icd	ACCOA+OAA → AKG+CO2	55.30	3.34	81.80	3.71
suc	AKG → SUCC+CO2	49.44	3.63	78.71	4.12
frd	SUCC →FUM	51.80	3.75	79.45	4.18
	reversibility	0.88	0.03	0.54	0.09
	scrambling	0.88	0.04	0.35	0.01
fum	FUM→MAL	53.30	3.73	80.25	4.10
	reversibility	0.68	0.08	0.36	0.01
	scrambling	0.824	0.073	0.10	0.01

Table S3 continued

mdh	MAL → OAA	44.83	2.50	74.22	3.46
	reversibility	0.19	0.06	0.13	0.01
Acetate production					
ackf	ACCOA → AC	60.52	0.29	37.65	0.28
	reversibility	0.56	0.20	0.73	0.02
ac	AC → ACout	4.16	1.95	2.50	0.09
Anaplerotic pathway					
ppc	PEP+CO2 → OAA	26.85	0.16	15.43	0.67
	reversibility	0.15	0.12	0.49	0.07
ana	Mal → PYR+CO2	2.36	0.12	0.74	0.05
aceA	ACCOA+OAA →GOx+SUCC	2.36	0.12	0.74	0.05
aceB	ACCOA+GOX → MAL	10.82	1.35	6.77	0.70
C1 metabolism					
ser	3PG → Ser	6.48	0.69	5.46	0.00
gly	Ser → Gly+C1	0.35	0.03	0.14	0.01
	reversibility	0.98	0.01	0.99	0.00
Amino acid biosynthesis pathway					
SdaRf	Ser → PYR	3.74	0.35	4.18	0.14
	reversibility	0.93	0.00	0.59	0.12
thr	Asp→ Thr	4.06	0.18	1.87	0.29
thrgly	Thr → Gly+ACCOA	2.75	0.10	1.21	0.20
	reversibility	0.05	0.01	0.04	0.00
GluDy	AKG → Glut	23.42	1.09	13.52	1.25
	reversibility	0.54	0.03	0.57	0.03
GlnDy	Glut → Gln	1.14	0.11	0.24	0.28
	reversibility	0.53	0.01	0.48	0.05
GLUSy	AKG+Gln → Glut+Glut	-0.22	0.03	-0.47	0.18
ALATA	AKG+Ala → Glut+PYR	-3.96	0.10	-3.21	0.10
	reversibility	0.43	0.01	0.46	0.03
ASPTA	AKG+Asp → Glut+OAA	-12.20	0.72	-6.13	0.86
	reversibility	0.95	0.05	0.69	0.02
Ala	Ala → PYR	1.26	0.10	1.81	0.10
Pro	Glut → Pro	1.18	0.11	0.61	0.09
	reversibility	0.53	0.01	0.50	0.04
argA	Glut+ACCOA→acGlut	1.50	0.02	0.80	0.07
argD	acGlut+Glut→acOrn+AKG+AKG	1.50	0.80	0.07	0.26
	reversibility	0.50	0.03	0.50	0.05
argE	acOrn->Orn +AC	1.50	0.02	0.80	0.07
argH	Orn+Asp->ARG+FUM	1.50	0.80	0.07	0.29
	reversibility	0.52	0.10	0.20	0.01
	scrambling	0.384	0.028	0.23	0.05

Table S3 continued

Transport pathway					
co2	CO2 → CO2out	240.20	10.96	325.89	12.82
Biomass synthesis pathway					
G6pb	G6P -> biomass	1.06	0.03	0.59	0.07
R5pb	R5P -> biomass	5.24	0.37	2.23	0.34
E4pb	E4P ->biomass	2.02	0.08	1.01	0.10
T3pb	T3P ->biomass	0.67	0.03	0.38	0.04
PEPb	PEP ->biomass	2.75	0.13	1.55	0.16
PYRb	PYR ->biomass	10.45	0.42	5.84	0.65
ACCOAb	ACCOA ->biomass	22.19	2.20	49.12	0.41
AKGb	AKG -> biomass	0.32	0.01	0.17	0.02
OAAb	OAA -> biomass	1.82	0.08	0.97	0.12
3PGb	3PG -> biomass	2.80	0.13	1.47	0.20
F6Pb	F6P ->biomass	0.36	0.01	0.19	0.03
C1b	C1 ->biomass	0.35	0.03	0.14	0.01
serb	Ser ->biomass	2.39	0.31	1.14	0.13
glyb	Gly ->biomass	3.11	0.14	1.35	0.21
thrb	Thr ->biomass	1.30	0.08	0.66	0.10
glutb	Glut ->biomass	1.50	0.06	0.80	0.12
glunb	Glu ->biomass	1.37	0.08	0.71	0.11
alab	Ala ->biomass	2.70	0.19	1.40	0.20
aspb	Asp->biomass	1.32	0.13	0.63	0.09
argb	Arg -> biomass	1.50	0.02	0.80	0.07
prob	Pro ->biomass	1.18	0.11	0.61	0.09

(1) Abbreviation: Glu, Glucose; G6P, Glucose-6-phosphate; F6P, Fructose-6-phosphate; T3P, Triose-3-phosphate; 3PG, 3-phosphoglycerate; PEP, Phosphoenolpyruvate; PYR, Pyruvate; ACCOA, Acetate-CoA; AKG, Ketoglutarate; SUCC, Succinate; MAL, Malate; OAA, Oxaloacetate; 6PG, 6-P-gluconate; R5P, Ribose-5-phosphate; S7P, Sedoheptulose-7-phosphate; E4P, Erythronate-4-phosphate; GOX, Glyoxylate; AC, Acetate; acGlut,acetyl glutamate; acOrn, acetyl ornithine; Orn, ornithine; CO2, carbon dioxide; Ser, Serine; Gly, Glycine; Arg, Arginine; Pro, Proline; Asp, Aspartate; Thr, Threonine; Glut, Glutamate; Gln, Glutamine; Arg, Arginine.
(2) Reversibility (R) and the extent of scrambling.

Table S4. Measured and simulated isotopomer distributions of proteinogenic amino acids from 2D-NMR for *E.coli* ML103 pXZ18Z strain grown in M9 media under the control and fatty acid producing condition.

Cross peak	**Control 1**		**Control 2**		**IPTG 1**		**IPTG 2**			
(multiplet)	**Expt**	**Sim**	**Expt**	**Sim**	**Expt**	**Sim**	**Expt**	**Sim**	**Precursor**	**Isotopomer**
Ala α (s)	0.143	0.070	0.051	0.074	0.063	0.087	0.086	0.079	Pyr	[123]
Ala α (d1)	0.104	0.079	0.078	0.069	0.070	0.058	0.093	0.051	Pyr	[123]
Ala α (d2)	0.038	0.051	0.091	0.040	0.085	0.052	0.057	0.045	Pyr	[123]
Ala α (dd)	0.715	0.801	0.780	0.817	0.782	0.804	0.763	0.825	Pyr	[123]
Ala β (s)	0.635	0.666	0.671	0.667	0.669	0.661	0.661	0.663	Pyr	[x23]
Ala β (d)	0.365	0.334	0.329	0.333	0.331	0.339	0.339	0.337	Pyr	[x23]
Arg β (s)	0.233	0.209	0.222	0.193	0.193	0.179	0.205	0.178	AKG	[x234x]
Arg β (d)	0.489	0.497	0.492	0.493	0.481	0.488	0.483	0.488	AKG	[x234x] + [x234x]
Arg β (t)	0.278	0.294	0.286	0.315	0.326	0.333	0.312	0.334	AKG	[x234x]
Arg δ (s)	0.104	0.172	0.098	0.169	0.083	0.151	0.079	0.140	AKG	[xxx45]
Arg δ (d)	0.896	0.828	0.902	0.831	0.917	0.849	0.921	0.860	AKG	[xxx45]
Asp α (s)	0.277	0.267	0.268	0.264	0.277	0.243	0.271	0.244	OAA	[12**3**x]
Asp α (d1)	0.186	0.188	0.203	0.180	0.153	0.193	0.155	0.191	OAA	[**12**3x]
Asp α (d2)	0.189	0.244	0.145	0.218	0.186	0.220	0.186	0.225	OAA	[1**23**x]
Asp α (dd)	0.347	0.302	0.384	0.338	0.385	0.344	0.388	0.340	OAA	[**123**x]
Asp β (s)	0.332	0.349	0.334	0.381	0.367	0.385	0.368	0.381	OAA	[x2**3**4]
Asp β (d1)	0.212	0.267	0.227	0.256	0.230	0.251	0.224	0.255	OAA	[x**23**4]
Asp β (d2)	0.212	0.198	0.215	0.195	0.186	0.207	0.185	0.204	OAA	[x2**34**]
Asp β (dd)	0.243	0.187	0.223	0.168	0.217	0.157	0.224	0.160	OAA	[x**234**]
Glu β (s)	0.178	0.209	0.198	0.193	0.124	0.179	0.140	0.178	Glut	[x2**3**4x]
Glu β (d)	0.472	0.497	0.516	0.493	0.441	0.488	0.438	0.488	Glut	[x**23**4x] + [x2**34**x]
Glu β (t)	0.296	0.294	0.287	0.315	0.360	0.333	0.336	0.334	Glut	[x**234**x]
Glu γ (s)	0.388	0.356	0.367	0.389	0.439	0.404	0.445	0.399	Glut	[xx3**4**5]
Glu γ (d1)	0.244	0.307	0.195	0.275	0.192	0.256	0.199	0.262	Glut	[xx**34**5]
Glu γ (d2)	0.223	0.181	0.282	0.197	0.249	0.208	0.240	0.204	Glut	[xx3**45**]
Glu γ (dd)	0.145	0.156	0.157	0.139	0.120	0.132	0.116	0.134	Glut	[xx**345**]
Gly α (s)	0.166	0.214	0.161	0.214	0.191	0.193	0.189	0.203	Gly	[12x]
Gly α (d)	0.834	0.786	0.839	0.786	0.809	0.807	0.811	0.797	Gly	[**12**x]

Table S4 continued

His β (s)	0.081	0.066	0.083	0.045	0.078	0.047	0.078	0.044	R5P	[x2**3**4x]
His β (d1)	0.473	0.489	0.492	0.513	0.527	0.531	0.504	0.517	R5P	[x2**34**x]
His β (d2)	0.018	0.016	0.012	0.009	0.001	0.009	0.014	0.008	R5P	[x**23**4x]
His β (dd)	0.428	0.429	0.414	0.434	0.393	0.413	0.404	0.431	R5P	[x**234**x]
Ile α (s)	0.341	0.399	0.344	0.371	0.305	0.352	0.320	0.359	OAA/Pyr	[1**2**xx] {x2x]
Ile α (d1)	0.369	0.389	0.395	0.409	0.362	0.415	0.399	0.411	OAA/Pyr	[**12**xx] {x2x]
Ile α (d2)	0.170	0.107	0.147	0.105	0.133	0.107	0.104	0.107	OAA/Pyr	[1**2**xx] {x**2**x]
Ile α (dd)	0.120	0.104	0.114	0.115	0.200	0.126	0.177	0.122	OAA/Pyr	[**12**xx] {x**2**x]
Ile γ1(s)	0.479	0.488	0.487	0.499	0.478	0.490	0.464	0.490	Pyr/OAA	[x2x] {xx**3**4]
Ile γ1(d)	0.416	0.432	0.413	0.422	0.414	0.426	0.413	0.426	Pyr/OAA	[x**2**x] {xx**3**4] +[x2x] {xx**34**]
Ile γ1(t)	0.105	0.081	0.100	0.079	0.107	0.084	0.123	0.083	Pyr/OAA	[x**2**x] {xx**34**]
Ile δ (s)	0.456	0.426	0.451	0.431	0.445	0.389	0.461	0.387	OAA	[xx3**4**]
Ile δ (d)	0.544	0.574	0.549	0.569	0.555	0.611	0.539	0.613	OAA	[xx**34**]
Ile γ2 (s)	0.618	0.666	0.608	0.667	0.631	0.661	0.633	0.663	Pyr	[x2**3**]
Ile γ2 (d)	0.382	0.334	0.392	0.333	0.369	0.339	0.367	0.337	Pyr	[x**23**]
Leu α (s)	0.465	0.523	0.462	0.518	0.437	0.506	0.458	0.510	ACoA/Pyr	[1**2**] {x2x]
Leu α (d1)	0.287	0.266	0.277	0.262	0.234	0.261	0.262	0.261	ACoA/Pyr	[**12**] {x2x]
Leu α (d2)	0.143	0.140	0.158	0.146	0.213	0.154	0.162	0.152	ACoA/Pyr	[1**2**] {x**2**x]
Leu α (dd)	0.106	0.071	0.104	0.074	0.116	0.079	0.119	0.078	ACoA/Pyr	[**12**] {x**2**x]
Leu (s)	0.382	0.366	0.379	0.343	0.356	0.317	0.375	0.316	ACoA/Pyr	[x2] {x**2**x].[x2x]
Leu (d)	0.472	0.521	0.486	0.534	0.506	0.546	0.495	0.548	ACoA/Pyr	[x**2**] {x**2**x].[x2x]+[x2] {x**2**x].[x**2**x]
Leu (t)	0.146	0.113	0.135	0.123	0.138	0.137	0.130	0.135	ACoA/Pyr	[x2] {x2x].[x2x]
Leu δ1 (s)	0.709	0.666	0.688	0.667	0.688	0.661	0.709	0.663	Pyr	[x23]
Leu δ1 (d)	0.291	0.334	0.312	0.333	0.312	0.339	0.291	0.337	Pyr	[x23]
Leu δ2 (s)	0.813	0.789	0.810	0.780	0.802	0.767	0.829	0.770	Pyr	[x2x] {xx3]
Leu δ2 (d)	0.187	0.211	0.190	0.220	0.199	0.233	0.171	0.230	Pyr	[x**2**x] {xx**3**]
Lys β (s)	0.369	0.384	0.388	0.410	0.425	0.419	0.426	0.417	OAA/Pyr	½{[x2**3**4] + [x2**3**] {xxx4]}
Lys β (d)	0.440	0.461	0.434	0.450	0.414	0.450	0.410	0.450	OAA/Pyr	½{[x**23**4] + [x2**34**] + [x**23**] {xxx4] + [x2**3**] {xxx**4**]}
Lys β (t)	0.192	0.155	0.178	0.140	0.161	0.131	0.164	0.133	OAA/Pyr	½{[x**234**] + [x**23**] {xxx**4**]}
Lys γ (s)	0.155	0.197	0.182	0.187	0.153	0.160	0.110	0.158	OAA/Pyr	[xx3**4**] {xx3]

Table S4 continued

Lys γ (d)	0.525	0.494	0.515	0.491	0.522	0.480	0.529	0.479	OAA/Pyr	[xx**34**] {xx3] + [xx3**4**] {xx**3**]
Lys γ (t)	0.320	0.310	0.303	0.322	0.326	0.359	0.360	0.363	OAA/Pyr	[xx**34**] {xx**3**]
Lys δ (s)	0.364	0.384	0.394	0.410	0.412	0.419	0.396	0.417	OAA/Pyr	½{[x2**3**4] + [x2**3**] {xxx4]}
Lys δ (d)	0.443	0.461	0.428	0.450	0.408	0.450	0.420	0.450	OAA/Pyr	½{[x2**3**4] + [x2**34**] + [x**23**] {xxx4] + [x2**3**] {xxx**4**]}
Lys δ (t)	0.193	0.155	0.178	0.140	0.180	0.131	0.184	0.133	OAA/Pyr	½{[x**234**] + [x**23**] {xxx**4**]}
Phe α (s)	0.062	0.065	0.050	0.068	0.050	0.081	0.056	0.074	PEP	[1**2**3]
Phe α (d1)	0.028	0.062	0.020	0.056	0.038	0.050	0.063	0.043	PEP	[**12**3]
Phe α (d2)	0.093	0.045	0.083	0.031	0.082	0.045	0.074	0.038	PEP	[1**23**]
Phe α (dd)	0.818	0.828	0.847	0.845	0.831	0.824	0.807	0.846	PEP	[**123**]
Phe (s)	0.507	0.524	0.481	0.521	0.450	0.509	0.475	0.513	PEP	[x2**3**].[2x]
Phe (d1)	0.268	0.264	0.294	0.261	0.292	0.261	0.282	0.260	PEP	[x**23**].[2x]
Phe (d2)	0.133	0.141	0.124	0.145	0.181	0.152	0.171	0.150	PEP	[x2**3**].[**2**x]
Phe (dd)	0.092	0.071	0.102	0.073	0.077	0.078	0.073	0.076	PEP	[x**23**].[**2**x]
Pro (s)	0.145	0.209	0.146	0.193	0.128	0.179	0.123	0.178	PRO	[x2**3**4x]
Pro (d)	0.508	0.497	0.514	0.493	0.486	0.488	0.493	0.488	PRO	[x**23**4x] + [x2**34**x]
Pro (t)	0.347	0.294	0.340	0.315	0.386	0.333	0.384	0.334	PRO	[x**234**x]
Pro γ(s)	0.360	0.356	0.336	0.389	0.396	0.404	0.397	0.399	PRO	[xx3**4**5]
Pro γ(d)	0.472	0.488	0.499	0.472	0.441	0.464	0.438	0.467	PRO	[xx**34**5] + [x23**45**]
Pro γ(t)	0.169	0.156	0.165	0.139	0.162	0.132	0.165	0.134	PRO	[xx**345**]
Pro δ (s)	0.211	0.172	0.094	0.169	0.168	0.151	0.212	0.140	PRO	[xxx4**5**]
Pro δ (d)	0.789	0.828	0.906	0.831	0.832	0.849	0.788	0.860	PRO	[xxx**45**]
Ser α (s)	0.052	0.080	0.050	0.078	0.064	0.074	0.056	0.076	Ser	[1**2**3]
Ser α (d1)	0.116	0.183	0.120	0.139	0.191	0.192	0.184	0.173	Ser	[**12**3]
Ser α (d2)	0.135	0.075	0.082	0.060	0.060	0.076	0.081	0.075	Ser	[1**23**]
Ser α (dd)	0.697	0.662	0.748	0.723	0.686	0.659	0.679	0.677	Ser	[**123**]
Ser β (s)	0.715	0.697	0.743	0.686	0.729	0.707	0.684	0.700	Ser	[x2**3**]
Ser β (d)	0.285	0.303	0.257	0.314	0.271	0.293	0.317	0.300	Ser	[x**23**]
Thr α (s)	0.262	0.264	0.306	0.263	0.288	0.241	0.281	0.243	THR	[1**2**3x]
Thr α (d1)	0.198	0.198	0.156	0.188	0.139	0.201	0.144	0.199	THR	[**12**3x]
Thr α (d2)	0.251	0.238	0.223	0.215	0.199	0.217	0.191	0.222	THR	[1**23**x]

Table S4 continued

Thr α (dd)	0.289	0.299	0.315	0.335	0.374	0.341	0.384	0.337	THR	[**123**x]
Thr γ2 (s)	0.435	0.442	0.450	0.442	0.429	0.403	0.419	0.399	THR	[xx3**4**]
Thr γ2 (d)	0.565	0.558	0.550	0.558	0.571	0.597	0.581	0.601	THR	[xx**34**]
Tyr α (s)	0.068	0.065	0.060	0.068	0.075	0.081	0.073	0.074	PEP	[1**2**3]
Tyr α (d1)	0.089	0.062	0.068	0.056	0.076	0.050	0.071	0.043	PEP	[**12**3]
Tyr α (d2)	0.090	0.045	0.081	0.031	0.077	0.045	0.075	0.038	PEP	[1**23**]
Tyr α (dd)	0.753	0.828	0.792	0.845	0.772	0.824	0.782	0.846	PEP	[**123**]
Tyr β (s)	0.486	0.524	0.578	0.521	0.521	0.509	0.554	0.513	PEP	[x2**3**].[2x]
Tyr β (d)	0.385	0.405	0.378	0.406	0.367	0.413	0.334	0.410	PEP	[x**23**].[2x]+ [x2**3**].[**2**x]
Tyr β (t)	0.129	0.071	0.044	0.073	0.113	0.078	0.112	0.076	PEP	[x**23**].[**2**x]
Tyr δ (s)	0.499	0.474	0.536	0.464	0.474	0.464	0.516	0.486	PEP/E4P	[xxx].[2**3**].[1x]+[xxx].[2].[**4**3x]
Tyr δ (d)	0.372	0.441	0.398	0.448	0.388	0.447	0.363	0.431	PEP/E4P	[xxx].[**23**].[1x]+[xxx].[**2**].[**4**3x]+ [xxx].[2**3**].[**1**x]+[xxx].[2].[**43**x]
Tyr δ (t)	0.130	0.085	0.066	0.088	0.138	0.090	0.122	0.083	PEP/E4P	[xxx].[**23**].[**1**x]+[xxx].[**2**].[**43**x]
Tyr ε (s)	0.073	0.128	0.033	0.089	0.059	0.094	0.063	0.123	PEP/E4P	[xxx].[x3].[**1**2]+[xxx].[x].[4**3**2]
Tyr ε (d)	0.290	0.288	0.297	0.302	0.308	0.286	0.302	0.282	PEP/E4P	[xxx].[x**3**].[**1**2]+[xxx].[x].[**43**2]+ [xxx].[x3].[**12**]+[xxx].[x].[4**32**]
Tyr ε (t)	0.637	0.584	0.670	0.609	0.634	0.620	0.635	0.595	PEP/E4P	[xxx].[x**3**].[**12**]+[xxx].[x].[**432**]
Val α (s)	0.121	0.095	0.084	0.089	0.090	0.106	0.103	0.096	Pyr	[1**2**x] {x2x]
Val α (d1)	0.602	0.694	0.640	0.692	0.626	0.661	0.642	0.675	Pyr	[**12**x] {x2x]
Val α (d2)	0.040	0.025	0.024	0.025	0.036	0.032	0.027	0.029	Pyr	[1**2**x] {x**2**x]
Val α (dd)	0.238	0.186	0.252	0.194	0.247	0.201	0.228	0.201	Pyr	[**12**x] {x**2**x]
Val γ1 (s)	0.648	0.666	0.636	0.667	0.652	0.661	0.646	0.663	Pyr	[x2**3**]
Val γ1 (d)	0.352	0.334	0.365	0.333	0.348	0.339	0.354	0.337	Pyr	[x**23**]
Val γ2 (s)	0.806	0.789	0.803	0.780	0.791	0.767	0.779	0.770	Pyr	[x2x] {xx**3**]
Val γ2 (d)	0.194	0.211	0.197	0.220	0.209	0.233	0.221	0.230	Pyr	[x**2**x] {xx**3**]

CHAPTER 6

UNRAVELING OCTANOIC ACID TOXICITY IN *SACCHAROMYCES CEREVISIAE* USING OMICS TOOLS

Authors: Ting Wei Tee, Jong Moon Yoon, Ping Liu, Laura Jarboe and Jacqueline V. Shanks

Department of Chemical and Biological Engineering, Iowa State University, Ames, IA

Abstract

Short-chained fatty acids synthesized via fermentation from biorenewable feedstocks are a potential source of platform chemicals, and thus could help replace the traditional petrochemical's dependence on crude oil. However, toxicity of fatty acids is an obstacle in the high titer of fatty acid production and it remains a key challenge in metabolic engineering. Metabolic flux analysis (MFA), the quantification of fluxes in metabolic pathways, is an integral tool for the development of strategies for genetic modification and the identification of metabolic regulation, by comparing fluxes under different environments. Transcriptome analysis allows examination of mRNA transcript levels for thousands of genes of multiple strains simultaneously to understand cell physiology and regulatory mechanism at the whole-cell transcript level by analysis them under various genotypic and environmental conditions. We used *Saccharomyces cerevisiae* as a model system to study the effect of toxicity of octanoic acid. The exposure of octanoic acid to yeast caused significant growth inhibition. We elucidated the metabolic flux differences in central metabolism between control and octanoic acid inhibition by conducting ^{13}C labeling experiments using fermentors. The yeast cultures were fed with a mixture of uniformly ^{13}C labeled glucose and 1-^{13}C positional labeled glucose. We quantified glucose uptake rate and fermentation product secretion rate using HPLC. The amino acid isotopomer fractions were measured using 2D [^{13}C, ^{2}H] HSQC NMR. Flux distributions were computed from simulating isotopomer distribution and then fitting it to the experimental measurements. We found distinctions in central metabolism flux distribution between control and treatment, especially in

the TCA cycle. Interestingly, transcriptome analysis revealed membrane stress and intracellular acidification based on the perturbation of membrane protein genes and the up-regulation of plasma membrane ATP-required transporter. Coupled with the elevation of active efflux transporter, yeast cells under octanoic acid stress tend to generate more energy and reducing power to counteract the effect of membrane stress and acidification through the plasma membrane transporter activity. Further efforts can focus on improving membrane integrity to enhance the tolerance against octanoic acid.

1. Background and introduction

With the depletion of fossil fuel supplies and intensifying environmental concern, the demand for production of fuels and chemicals from biorenewable sources becomes increasingly apparent for sustainability. Tremendous research efforts have been conducted to engineer microbes for efficient synthesis of biofuels and biochemicals (Atsumi et al., 2008; Steen et al., 2010; Yim et al., 2011). Advances in metabolic engineering expand the portfolio of fuels and commodity chemicals that can be synthesized biologically (Bozell and Petersen, 2010; Nikolau et al., 2008). Carboxylic acids recently gains extensive attention in the global research community due to its potential as fuels and chemicals (Nikolau et al., 2008; Stephanopoulos, 2007). Diversity of plant thioesterases, that can cleave acyl-ACP substrates at specific chain length, presents the opportunity for novel fatty acid production (Jing et al., 2011). As platform molecules, carboxylic acids naturally synthesize through fermentation can be transformed into building blocks for industrial chemicals such as fatty alcohols (Dellomonaco et al., 2011), fatty esters (Steen et al., 2010), and alkyl ketones (Goh et al., 2012). It has been recently shown that short-chained fatty acids could be catalytically deoxygenated via heterogeneous metal catalysts to produce α-olefins, which serve as building blocks of important polymerization products (Alonso et al., 2010; Shanks, 2010).

S. cerevisiae which grows well at low pH is a promising biocatalyst for carboxylic acid production. Their robustness, fast division, pH tolerance, simple nutrient requirement, completely sequenced genome and long history as industrial workhorse make it an excellent microbial cell factory to produce carboxylic acids (Abbott et al., 2009). In order to produce fuels and chemicals at high titers and yields, product inhibition to the biocatalysts becomes an increasingly serious problem that needs to be resolved (Jarboe et al., 2011). For example,

isobutanol is inhibitory for cell growth of the microbial host system when the production titer exceeds the toxicity level (Atsumi et al., 2010). The toxicity effect can be mitigated by in-situ product removal from the broth in bioreactor to further enhance product titer (Baez et al., 2011). However, this method does not work for all compounds and involves additional process cost. Another common strategy to attenuate product toxicity is by selection-based approach from metabolic evolution and random mutagenesis of certain key enzyme/ transcription machinery (Atsumi et al., 2010; Minty et al., 2011). However this black box model does not explain the mechanism of tolerance and requires extensive reverse engineering to unveil it. With the advances in DNA recombinant technology, we can use system metabolic engineering tools to understand the underlying inhibitory mechanism for rational engineering to mitigate the product toxicity.

Carboxylics acids such as octanoic and decanoic acids have been reported to inhibit cell growth of *S. cerevisiae* by reducing the cell viability and even cause death (Cabral et al., 2001; Viegas et al., 1989). Medium-chain fatty acids penetrate inside the cell by passive diffusion in a non-ionized form and dissociate in the cytosol at higher internal pH, leading to a decrease of the intracellular pH and accumulation of toxic anion (Alexandre et al., 1996; Cabral et al., 2001; Legras et al., 2010). Decanoic acid stress increased membrane fluidity in yeast and changed the membrane lipid composition (Alexandre et al., 1996). Furthermore, Abbott et al. (2007) investigated the effect of weak acids (benzoate, sorbate, acetate and propionate) using microarray based transcriptome analysis and revealed consistent up-regulation trend of genes related to cell wall, mitochondrial superoxide removal and DNA synthesis/repair. Recently, another transcriptome study, which investigated yeast response to the exposure to octanoic and decanoic acids, unraveled activation of oxidative stress response (Legras et al., 2010). Notably, membrane transporter genes were activated under octanoic acid stress.

Despite the above efforts to understand the medium-chained acid stress on yeast, *in-vivo* flux distribution is not studied. Metabolic flux analysis (MFA) quantifies internal metabolic fluxes and provides a snapshot of cellular physiology to pinpoint genetic manipulations (Koffas and Stephanopoulos, 2005; Peebles et al., 2010; Stephanopoulos, 1999). Due to a hierarchy of information in organism, MFA could determine the phenotype by quantifying the integration system response for gene-protein-metabolites interactions (Gregory N. Stephanopoulos, 1998; Sauer, 2006; van Rijsewijk et al., 2011). In this study, we employed MFA along with

transcriptome study to understand *Saccharomyces cerevisiae* response to octanoic acid stress, thus providing the platform for rational engineering of product-tolerant biocatalyst for production of short-chained fatty acids.

2. Material and methods

2.1. Strains and plasmids: The wild type *Sacchromyce cerevisiae* BY4741 (*MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0*) was used throughout this work. The strain was a gift from Dr. Nancy Da Silva's lab.

2.2. Metabolic flux analysis experiment

Strains and culture conditions:

The BY4741 strain from glycerol frozen stock was streaked on Yeast Peptone Dextrose (YPD) plate and incubated overnight at 30^{o}C. The colonies from YPD plate was transferred in 250mL baffled flasks containing 50mL of the following synthetic dextrose (SD) minimal medium: 10 g/L glucose, 6.7 g/L yeast nitrogen base without amino acid, 100 mg/L uracil, 100 mg/L histidine, 100 mg/L methionine and 300mg/L leucine. The medium pH was adjusted to 5.0. After preculturing for about 24 hours on the rotatory shaker at 30^{o}C and 150 rpm, cells were used to inoculate 50mL medium in baffled flasks to optical density (OD) at 600nm of 0.05 for hexanoic acid and octanoic acid toxicity studies.

For ^{13}C labeling experiment (CLE), cells were centrifuged at 4000 rpm for 5 min at 4 ℃ and re-suspended in fresh SD minimal medium to inoculate 400 mL of SD minimal medium in the 500 mL bioreactor (INFORS HT, Switzerland) to a starting OD_{600} of 0.02. For better identification of fluxes, 20% U-^{13}C glucose and 80% 1-^{13}C glucose was used as the tracer to final medium concentration of 1% glucose (Fischer et al., 2004). The aerobic fermentation was conducted at 30 ℃, with a gas flow rate at 0.4 ml/min and agitation speed of 600 rpm. The pH was controlled at 5.0±0.05 by adding 1M potassium hydroxide. The dissolved oxygen level was maintained above 50% of saturated levels to ensure aerobic conditions. The cells were harvested at mid-exponential phase after at least 5 generations to ensure metabolic and isotopic steady state. Parallel experiment was conducted as biological replicate.

Analytical techniques:

Cell biomass dry weight was determined by measuring optical density OD_{550} using a spectrophotometer (Genesys 20, Madison, WI). Cell dry weight was estimated by the correlation: 1 OD_{600} =0.56 g cell dry weight/L. Biomass composition was determined based on literature data (Forster et al., 2003). Media samples were taken during the exponential growth and filtered through 0.22 μm pore sized nylon filters (P.J. Cobert Associates, Saint Louis, MO) and kept at -20 ^{o}C for extracellular metabolite analysis. Glucose, glycerol, ethanol and acetate were measured using a Waters HPLC (Waters, Milford, MA) with 410 refractive index detector. The Aminex column (HPX-87H, Bio-Rad, Hercules, CA) was used at 30 ^{o}C with 0.3 mL/min of 5mM sulfuric acid as mobile phase. The substrate uptake rate and production secretion rate in batch culture are constant during exponential phase. The substrate uptake rate and product secretion rate are defined as the coefficient of substrate/product concentration versus biomass divided by the growth rate.

Amino acids in the medium were analyzed using HPLC with Waters AccQ Tag column and UV detector. The samples were derivatized and the HPLC was programmed according to the manufacturer protocols. α-aminobutyric acid was used as the internal standards for quantification.

Octanoic acid in the medium was quantified as described by Jing and coworkers (Jing et al., 2011). Octanoic acid samples were centrifuged and the supernatant was acidified with HCl. The samples were extracted using chloroform-methanol mixture and then concentrated to ~100uL. Samples were analyzed on an Agilent Technologies 6890 Series gas chromatograph system with an Agilent CP-Wax 58 FFAP CB column (25 m × 0.15 mm × 0.39 mm). 10μg heptanoic acid was used as the internal standard for quantification using AMDIS software.

2-Dimensional NMR analysis:

Cells were harvested and hydrolyzed into amino acids as described previously (Choudhary et al., 2011; Ranganathan et al., 2012). The sample was dissolved in deuterium oxide for NMR analysis. 2D [^{13}C, ^{1}H] Heteronuclear Single Quantum Correlation spectra were acquired on a Bruker Advance DRX 500 MHz spectrometer at 298 K and processed as described previously (Choudhary et al., 2011; Sriram et al., 2004). The amino acids isotopomer abundances are related to the precursor metabolites with the reference to amino acids biosynthesis pathways

(Maaheimo et al., 2001). The resulting isotopomer fractions are shown in Supplementary Table S4.

Metabolic network model for MFA:

A network model for *E. coli* metabolism was constructed based on existing literature and yeastgenome database (see Table S2). The model includes glucose transport and phosphorylation pathway, Embden-Meyerhof-Parnas pathway, oxidative pentose phosphate branch, non-oxidative pentose phosphate branch, TCA cycle, anaplerotic pathways, metabolite exchange reactions, all amino acids biosynthesis pathways. Eukaryotic compartmentation into mitochondrial and cytosolic subsystems was considered by distinguishing distinct pools of pyruvate, oxaloacetate, and acetyl-coA in both compartments (Fiaux et al., 2003). The pyruvate transport into the mitochondria is considered unidirectional, driven by the proton motive force. Glyoxylate cycle (Gancedo, 1998) and phosphoenolpyruvate carboxylase kinase (Haarasilta and Oura, 1975; Yin et al., 1996) are repressed under the presence of glucose, thus not included in our network.

Flux evaluation methodology:

Fluxes were computed using NMR2Flux software developed by Sriram et al (2004). NMR2Flux employs isotopomer balancing and a global optimization routine to find stoichiometrically feasible fluxes set consistent with experimental measurements. Overall fluxes were estimated by minimizing the chi square error between experimentally measured and simulated isotopomer fractions of amino acid. Bootstrap Monte Carlo statistical analysis was employed to evaluate the standard deviations of the fluxes.

2.3 Microarray

Microarray experiment was performed to examine the whole-gene transcript perturbation for insights on the cell physiology and regulatory mechanism under octanoic acid stress. Independent from CLE, *S. cerevisiae* BY4741 was grown in SDC medium with 2% glucose at 30^{o}C with orbital shaking at 150rpm. SDC medium contained 0.67% (w/v) yeast nitrogen base without amino acids, 0.5% (w/v) casamino acids, and 20mg/L uracil. The pH was adjusted with HCl to 5.0 before inoculation. 0.3mM octanoic acid was added in the treatment cultures. Cells were harvested during mid-log growth. The triplicate samples were processed and analyzed as described previously (Liu et al., 2013). Log2 ratio of transcription factor activities between

octanoic acid stress and the control were estimated based on Network Component Analysis (NCA) algorithm (Chang et al., 2008). The transcriptome data has been reported by Jarboe and coworker (Liu et al., 2013).

3. Results:

3.1 Toxicity of medium-chain fatty acids in yeast cells

S.cerevisiae was grown in SD minimal medium with 1% glucose under exposure to various concentrations of hexanoic and octanoic acid to characterize the growth behavior. This analysis showed that the growth of *S. cerevisiae* BY4741 can only tolerate haxanoic and octanoic acid concentrations of less than 5mM in the shake flask systems (Supplementary Figure S1). The growth inhibition escalates with increasing concentration of fatty acids. In addition, the inhibitory effect is more severe with increasing carbon chain length of fatty acids. This finding agrees with the supplementation of C6-C10 fatty acid in rich SDC medium with higher glucose concentration (Liu et al., 2013) even though the overall cell growth rates are higher in rich medium.

The inhibition effects of octanoic acid in the fermentor experiment were alleviated compared to shake flask experiments (Supplementary Table S1). This is probably because acid inhibition is dependent upon the rate of acid penetration, the subsequent decrease in internal pH and the effect of pH on specific enzyme systems. Lower medium pH during fermentation in the shake flasks also leads to the accumulation of undissociated form of fatty acid, which is toxic (Liu et al., 2013). Therefore we conducted ^{13}C labeling experiment in fermentor system with controlled pH, agitation and aeration rate to minimize the environmental variations.

3.2 Metabolic flux analysis results

Phenotype comparison between control and C8 stress

S. cerevisiae was grown in minimal SD medium with 1% glucose in the bioreactor at $30^{o}C$, pH 5.0 and dissolved oxygen above 50% saturation to characterize the effects of octanoic acid on cell growth. Phenotype differences were observed when the *S. cerevisiae* cells were under 0.4mM octanoic acid during aerobic batch fermentation as shown in Figure 1. The cells exhibited respiratory-fermentative metabolism with secretion of ethanol, acetate and glycerol. Under octanoic acid exposure, cells grew slower with 25% lower specific growth rate and

slightly higher specific glucose uptake rate. Ethanol and acetate yield increased by 10% and 2-fold respectively, coupled with the decrease in biomass yield of 41% under C8 fatty acid inhibition. Interestingly, glycerol yield decreased 7-fold under octanoic acid inhibition. Notably, glycerol plays important roles in yeast physiological processes such as combating osmotic stress, managing cytosolic phosphate levels and maintaining the NAD+/NADH redox balance (Blomberg, 1997; Wang et al., 2001).

Since *S. cerevisiae* BY4741 strain is auxotrophic for histidine, leucine and methionine, we quantified the residue concentrations of the amino acids to obtain amino acid uptake rate (Supplementary Figure S2). Among the amino acids, leucine was the most consumed under the control condition (0.7mM/g.hr). Compared to the glucose uptake rate of ~17mM/g.hr, leucine uptake is just about 4% of the total carbon substrate, thus negligible to affect the flux distributions. Octanoic acid was also quantified using GC-MS, but we found no significant evidence of octanoic acid oxidation that is being degraded and catabolized by the cells (data not shown).

In vivo metabolic flux analysis, a robust tool to estimate intracellular fluxes, was performed to unravel the metabolic network under the physiological conditions (i.e. 0.4mM octanoic acid exposure). The flux maps for *S. cerevisiae* under the control and octanoic acid stress are tabulated in Supplementary Table S2 and shown in Figure 2. The flux values are all normalized to 100 mmol $gDW^{-1}\,h^{-1}$ and an average of the two replicate experiments. Most of the carbon flux (around 80%) is directed towards the glycolytic pathway, resulting in high activities of the lower glycolytic pathway in respiro-fermenting cells. Around 20% of the carbon flux channels through pentose phosphate (PP) pathway to generate NADPH for reduction requirements. These results are consistent with previous flux experiments showing active PP pathway under glucose aerobic batch culture (Fiaux et al., 2003; Gombert et al., 2001). The fluxes through the TCA cycle which operated as oxidative cycle, showed low activity under aerobic batch fermentation to generate ATP and NAD(P)H for energy and reduction requirements for cell growth. Gombert *et al* (2001) reported the same observation of low TCA activity under batch aerobic condition. Nonetheless, most of the carbon fluxes are drained towards formation of ethanol (150 mmol/gDW.h) through pyruvate decarboxylase and regenerated NAD^{+}. Glycerol and acetate are the other products in yeast aerobic fermentation

with 15 mmol/gDW.hr and 2 mmol/gDW.hr respectively. The flux through malic enzyme which reroutes malate to pyruvate is negligible.

When cells were exposed to 0.4mM octanoic acid stress, the oxidative TCA cycle activities are observed to be up-regulated 12 fold while reactions in the PP pathway remain rigid. The glycerol secretion decreases by 85%, leading to 10% increase in fluxes flowing through lower glycolysis pathway. At the pyruvate node, pyruvate dehydrogenase converting pyruvate to acetyl-CoA increases 11-fold, while pyruvate carboxylase that converts pyruvate to oxaloacetate is inhibited by 40% under octanoic acid stress. With the increase in ethanol and acetate yield, carbon fluxes through pyruvate decarboxylase slightly increases ~6%. In addition, along the decrease of cell growth rate during fatty acid stress, the amino acid biosynthesis reactions are generally lower than the control.

Even though cofactor balance is not included in MFA, the net cofactor production is derived from the *in vivo* flux distribution. Figure 3 illustrates the role of NADH, NADPH and ATP in the central carbon metabolism. For example, NADPH is produced in the oxidative PP pathway, TCA cycle and malic enzyme activities. Interestingly, malate dehyrogenases can use both NAD^+ and $NADP^+$ as reaction precursor (Boles et al., 1998; Yamaguchi, 1979). Aldehyde dehydrogenase converts aldehyde to acetate through reduction of NAD^+ and $NADP^+$ (Bostian and Betts, 1978). The malic dehydrogenase and aldehyde dehydrogenase catalyzed by different enzymes cannot be distinguished by ^{13}C-MFA, which is constrained by reaction carbon stoichiometry. The malic enzyme and aldehyde dehydrogenase are thus considered to produce either NADH or NADPH and treated as the measurement errors in cofactor production. The net productions of ATP per glucose consumed are up-regulated 20% under fatty acid stress (Figure 3). The net production of NADH and NADPH are 5-fold higher and 13% lower respectively, leading to ~2.8 fold increase in the overall reducing power NAD(P)H under octanoic acid inhibition.

3.3 Microarray analysis

Liu et al. performed transcriptomics analysis of *S. cerevisiae* BY4741 during mid-log growth in SDC media to compare gene transcriptional perturbation under the control and exposure to 0.3mM octanoic acid, which was sufficient to decrease the specific growth rate by 25%. This analysis identified 937 genes with significantly ($p<0.01$) perturbed expression under

octanoic acid stress relative to the control; 136 of these genes have expression that is perturbed more than 2-fold ((Liu et al., 2013), supplementary materials). Despite the difference in culture medium and environment, 0.3mM octanoic acid exposure led to the similar extent of cell inhibition. This transcriptomics study thus serves as the reference to compare with *in vivo* flux study to unravel inhibition mechanism of octanoic acid. We mapped the gene transcriptional perturbations under octanoic acid stress on the central carbon metabolism as shown in Supplementary Figure S4. However, most of the genes in the central carbon metabolism did not change significant except *TKT1* in the oxidative PP pathway and *MAE1* in the malic enzyme activity. The *TKT1* and *MAE1* genes were down-regulated 13.8 fold and 2.1 fold respectively under octanoic acid inhibition, but the corresponding reaction fluxes of transketolase and malate dehydrogenase did not perturb. This might be due to unknown post translational regulation or indirect interaction among gene, protein, metabolite and flux levels.

Interestingly, the genes related to plasma membrane ATP-binding cassette (ABC) transporter such as *PDR3*, *PDR12*, *PDR15* and *PDR16* were up-regulated significantly (>2-fold) under octanoic acid stress (Supplementary Table S5). In addition, the expression of *PDR5* and *YOR1* increased >1.5 fold complementing the up-regulation of ABC plasma membrane proteins. ABC membrane transporter mediates cellular detoxification by effluxing the harmful xenobiotics (Jungwirth and Kuchler, 2006). *PDR* plasma membrane transporter ABC superfamily is the first line defense to be involved in active expulsion of weak lipophilic acids (Mira et al., 2010b). The increase of PDR gene transcription might imply octanoic acid stress acidified the intracellular pH of the cytosol and mitochondrion, and then disrupt the integrity of cell membrane. The low pH would activate the plasma membrane transporter to repel the toxin out of the cells, at the consumption of energy.

The microarray data also depicted the gene perturbation of membrane proteins and metal ion uptake (Supplementary Table S5). Most of the membrane proteins were down-regulated under octanoic acid exposure, probably interrupting the membrane function to maintain cell homeostasis. This could suggest the potential loss of membrane integrity resulting from octanoic acid toxicity. We also observed up-regulation of the gene expression related to uptake of metal ions (iron, zinc, copper, calcium and ammonia). This finding was in agreement with the literature (Hazelwood et al., 2010; Lawrence et al., 2004), suggesting these metal ions might be crucial for tolerance against octanoic acid. Lipophilic acid could chelate metal cations and then limit metal

availability to affect the cellular metal homeostasis (Abbott et al., 2008; Hazelwood et al., 2010). On the other hand, it might be the symptom of membrane disintegrity to perturb the intracellular metal ions availability (Liu et al., 2013). Given the transcriptional change in the membrane protein and metal uptake, we hypothesized the presence of octanoic acid disrupted the membrane integrity and thus inhibit the cell growth. Jarboe and coworkers quantitatively confirmed the membrane leakage problem exacerbates with increasing concentration and chain length of C_{6-10} fatty acids.

Network component analysis was performed with and without the new regulatory links from the reconstruction of gene regulatory networks. The results exhibited the significant perturbation of Hap5, Haa1,and Stb5 transcription factor in all analyses ((Liu et al., 2013), supplementary material). Moreover, War1, Pdr3 and Pdr1 were predicted to be significantly changed based on either one or more regulatory network in the NCA analysis. Hap5 was reported to be involved in regulation of iron homeostasis (Liu et al., 2013). Haa1 regulates membrane stress proteins and adaptation to weak acid stress, while Stb5 regulates multidrug resistance and oxidative stress. War1 regulates *Pdr12p* for resistance to weak lipophilic acid (Legras et al., 2010). Pdr1 and Pdr3 transcription factors also play roles in regulating pleiotroplic drug resistance in yeast (Mira et al., 2010b). Based on the NCA results, it suggested the presence of octanoic acid induces acid stress response and triggers multidrug efflux and ABC membrane plasma transporter activities. Membrane stress response was also detected indirectly from the perturbation in iron uptake regulation (Liu et al., 2013).

4. Summary and discussion

Short-chained fatty acids are potential platform molecules synthesized biologically for biorenewable fuels and chemicals. In order to make the microbial production economically viable, we have to overcome increasingly vexing biocatalyst inhibition by the products and achieve high product yield and titer. In the present work, we performed ^{13}C metabolic flux analysis on *S. cerevisiae* BY4741 along with the reported transcriptomics analysis to investigate cell response under octanoic acid inhibition. From phenotype analysis, the yeast cells under octanoic acid stress illustrate reduced cell growth with slight increase in glucose consumption rate. The glycerol yield significantly decreases 7-fold, while the yield of acetate and ethanol increase 2-fold and 10% respectively. Extracellular octanoic acid might cause osmotic stress and

disrupt the function of glycerol production. We therefore deleted some key genes in glycerol formation (*GDP1*, *GPP1*) to test the tolerance towards octanoic acid toxicity in the SDC media. However, neither improvement nor deterioration in cell growth was found under octanoic acid stress. This led to proposition that octanoic acid might not lead to the increase of osmotic stress but the acidification of intracellular cytosol resulting in membrane stress.

The plasma membrane is a highly specialized organelle in yeasts to selectively transport a multitude of molecules (Spira et al., 2012). It also serves as a platform for various signaling complexes (Spira et al., 2012). Organic solvents could impair the function cell membrane, leading to retarded growth and even death (Ramos et al., 2002; Segura et al., 2012). Figure 4 explains the mechanism of acidification caused by octanoic acid exposure and the activation of ATP-binding cassette transporter for remediation. Octanoic acid is known to be toxic in undissociated form (Viegas et al., 1989) and could permeate the plasma membrane by diffusion. In the nearly-neutral cytosol, the dissociation of octanoic acid would cause intracellular acidification and the accumulation of the protons and anions. In addition, the spatial composition of membrane plasma is likely disrupted, resulting in the function disruption of membrane embedded proteins (Mira et al., 2010b). This would increase nonspecific cell permeability to ions and small metabolites. This would stimulate protons to diffuse passively into the cytosol, further reducing the intracellular pH. Intracellular acidification would inhibit metabolic activity and dissipate the electrochemical potential maintained across the plasma membrane (Mira et al., 2010b). When yeasts are exposed to lipophilic acid, the activity of plasma membrane H^+-ATPase is induced to generate the transmembrane proton gradient and regulate intracellular pH, thus maintaining pH homeostasis (Cabral et al., 2001; Legras et al., 2010; Viegas et al., 1998). The activation of H^+-ATPase couples with ATP hydrolysis to proton extrusion. The up-regulation of plasma membrane ABC transporter *PDR* gene expression strengthens the acidification of cytosol by external octanoic acid leading to cell inhibition. *Pdr12p* gene, which is regulated by WAR1 transcription factor and facilitates ATP-dependent efflux of moderately lipophilic short-chain acid anions, has been identified as a key determinant in organic acid tolerance (Legras et al., 2010; Piper et al., 1998).

The MFA data revealed up-regulation of the TCA cycle and glycolysis, when the cells are exposed to octanoic acid, along with the up-regulation of plasma membrane ATP-binding transporter genes indicated by the transcriptome analysis. The activation energy-intensive

defense mechanism (H^+-ATPase ABC transporter) would enhance ATP depletion in the cells under octanoic acid stress. The increase in TCA cycle and glycolysis activities generates more ATPs for H^+ ATPase activity to efflux the protons and anions through plasma membrane ATP binding cassette transporter. Independent with this study, yeasts under lipophilic acid stress exhibited up-regulation of genes and proteins encoding the glycolysis and the TCA cycle to compensate the depletion of ATP (Abbott et al., 2007; Abbott et al., 2008; Almeida et al., 2009; Mira et al., 2010a; Mira et al., 2009; Mira et al., 2010b).

The gene transcript perturbation of membrane proteins and metal uptake regulation lead to the hypothesis of loss membrane integrity, in which Jarboe and coworkers quantitatively confirmed the membrane leakage problem under octanoic acid exposure. Decreased integrity of the plasma membrane can cause leakage of intracellular materials, leading to the decreased availability of important cofactors for biological process and then growth inhibition (Osman and Ingram, 1985). To mitigate the membrane leakage, yeast cells might tightly regulate the fluxes through central carbon metabolism to replenish the cofactors (ATPs and NAD(P)Hs) for cell survival. Notably, MFA elucidated the overall energy and reducing powers productions are higher under octanoic acid stress to compensate for the effect of membrane leakage. In addition, the up-regulation of genes related to ion uptake is crucial to maintain metal homeostasis for survival under octanoic acid stress. The above findings suggest the ability to control membrane composition as the crucial strategy to enhance tolerance against organic acid as illustrated by Jarboe and coworkers (Liu et al., 2013).

In conclusion, we have integrated the metabolic flux analysis and transcriptome analysis reported by Jarboe *et al.* to identify the up-regulation of the TCA cycle fluxes and gene expression related to plasma membrane ATP-dependent transporter, membrane proteins and metal ion uptake as the key determinant in octanoic acid stress to yeast. Membrane integrity is disrupted under octanoic acid inhibition, thereby modifying the membrane composition, replenishing cofactor supplies and over-expressing efflux transporter may further enhance the tolerance against octanoic acid.

Authors' contributions

JVS and LJ conceived the project. TWT and JMY designed and performed the experiments and the metabolic flux analysis. PL designed and performed the microarray experiment. TWT drafted the manuscript.

Acknowledgments

This work was supported by the National Science Foundation Engineering Research Center for Biorenewable Chemicals, NSF award number EEC-0813570

References

Abbott, D. A., T. A. Knijnenburg, L. M. I. de Poorter, M. J. T. Reinders, J. T. Pronk, and A. J. A. van Maris, 2007, Generic and specific transcriptional responses to different weak organic acids in anaerobic chemostat cultures of Saccharomyces cerevisiae: Fems Yeast Research, v. 7.

Abbott, D. A., E. Suir, A. J. A. van Maris, and J. T. Pronk, 2008, Physiological and transcriptional responses to high concentrations of lactic acid in anaerobic chemostat cultures of Saccharomyces cerevisiae: Applied and Environmental Microbiology, v. 74, p. 5759-5768.

Abbott, D. A., R. M. Zelle, J. T. Pronk, and A. J. A. van Maris, 2009, Metabolic engineering of Saccharomyces cerevisiae for production of carboxylic acids: current status and challenges: Fems Yeast Research, v. 9, p. 1123-1136.

Alexandre, H., B. Mathieu, and C. Charpentier, 1996, Alteration in membrane fluidity and lipid composition, and modulation of H+-ATPase activity in Saccharomyces cerevisiae caused by decanoic acid: Microbiology-Sgm, v. 142, p. 469-475.

Almeida, B., S. Ohlmeier, A. J. Almeida, F. Madeo, C. Leao, F. Rodrigues, and P. Ludovico, 2009, Yeast protein expression profile during acetic acid-induced apoptosis indicates causal involvement of the TOR pathway: Proteomics, v. 9, p. 720-732.

Alonso, D. M., J. Q. Bond, and J. A. Dumesic, 2010, Catalytic conversion of biomass to biofuels: Green Chemistry, v. 12, p. 1493-1513.

Atsumi, S., T. Hanai, and J. C. Liao, 2008, Non-fermentative pathways for synthesis of branched-chain higher alcohols as biofuels: Nature, v. 451.

Atsumi, S., T.-Y. Wu, I. M. P. Machado, W.-C. Huang, P.-Y. Chen, M. Pellegrini, and J. C. Liao, 2010, Evolution, genomic analysis, and reconstruction of isobutanol tolerance in Escherichia coli: Molecular Systems Biology, v. 6.

Baez, A., K.-M. Cho, and J. C. Liao, 2011, High-flux isobutanol production using engineered Escherichia coli: a bioreactor study with in situ product removal: Applied Microbiology and Biotechnology, v. 90.

Blomberg, A., 1997, Osmoresponsive proteins and functional assessment strategies in Saccharomyces cerevisiae: Electrophoresis, v. 18, p. 1429-1440.

Boles, E., P. de Jong-Gubbels, and J. T. Pronk, 1998, Identification and characterization of MAE1, the Saccharomyces cerevisiae structural gene encoding mitochondrial malic enzyme: Journal of Bacteriology, v. 180.

Bostian, K. A., and G. F. Betts, 1978, KINETICS AND REACTION-MECHANISM OF POTASSIUM-ACTIVATED ALDEHYDE DEHYDROGENASE FROM SACCHAROMYCES-CEREVISIAE: Biochemical Journal, v. 173.

Bozell, J. J., and G. R. Petersen, 2010, Technology development for the production of biobased products from biorefinery carbohydrates-the US Department of Energy's "Top 10" revisited: Green Chemistry, v. 12.

Cabral, M. G., C. A. Viegas, and I. Sa-Correia, 2001, Mechanisms underlying the acquisition of resistance to octanoic-acid-induced-death following exposure of Saccharomyces cerevisiae to mild stress imposed by octanoic acid or ethanol: Archives of Microbiology, v. 175.

Chang, C., Z. Ding, Y. S. Hung, and P. C. W. Fung, 2008, Fast network component analysis (FastNCA) for gene regulatory network reconstruction from microarray data: Bioinformatics, v. 24, p. 1349-1358.

Choudhary, M. K., J. M. Yoon, R. Gonzalez, and J. V. Shanks, 2011, Re-examination of metabolic fluxes in Escherichia coli during anaerobic fermentation of glucose using (13)C labeling experiments and 2-dimensional nuclear magnetic resonance (NMR) spectroscopy: Biotechnology and Bioprocess Engineering, v. 16, p. 419-437.

Dellomonaco, C., J. M. Clomburg, E. N. Miller, and R. Gonzalez, 2011, Engineered reversal of the beta-oxidation cycle for the synthesis of fuels and chemicals: Nature, v. 476.

Fiaux, J., Z. P. Cakar, M. Sonderegger, K. Wuthrich, T. Szyperski, and U. Sauer, 2003, Metabolic-flux profiling of the yeasts Saccharomyces cerevisiae and Pichia stipitis: Eukaryotic Cell, v. 2.

Fischer, E., N. Zamboni, and U. Sauer, 2004, High-throughput metabolic flux analysis based on gas chromatography-mass spectrometry derived C-13 constraints: Analytical Biochemistry, v. 325, p. 308-316.

Forster, J., I. Famili, P. Fu, B. O. Palsson, and J. Nielsen, 2003, Genome-scale reconstruction of the Saccharomyces cerevisiae metabolic network: Genome Research, v. 13.

Gancedo, J. M., 1998, Yeast carbon catabolite repression: Microbiology and Molecular Biology Reviews, v. 62.

Goh, E.-B., E. E. K. Baidoo, J. D. Keasling, and H. R. Beller, 2012, Engineering of Bacterial Methyl Ketone Synthesis for Biofuels: Applied and Environmental Microbiology, v. 78, p. 70-80.

Gombert, A. K., M. M. dos Santos, B. Christensen, and J. Nielsen, 2001, Network identification and flux quantification in the central metabolism of Saccharomyces cerevisiae under different conditions of glucose repression: Journal of Bacteriology, v. 183.

Gregory N. Stephanopoulos, A. A. A., Jens Nielsen, 1998, Metabolic Engineering: Principles and Methodologies.

Haarasilta, S., and E. Oura, 1975, ACTIVITY AND REGULATION OF ANAPLEROTIC AND GLUCONEOGENETIC ENZYMES DURING GROWTH PROCESS OF BAKERS-YEAST - BIPHASIC GROWTH: European Journal of Biochemistry, v. 52.

Hazelwood, L. A., M. C. Walsh, J. T. Pronk, and J.-M. Daran, 2010, Involvement of Vacuolar Sequestration and Active Transport in Tolerance of Saccharomyces cerevisiae to Hop Iso-alpha-Acids: Applied and Environmental Microbiology, v. 76, p. 318-328.

Jarboe, L. R., P. Liu, and L. A. Royce, 2011, Engineering inhibitor tolerance for the production of biorenewable fuels and chemicals: Current Opinion in Chemical Engineering, v. 1, p. 38-42.

Jing, F., D. C. Cantu, J. Tvaruzkova, J. P. Chipman, B. J. Nikolau, M. D. Yandeau-Nelson, and P. J. Reilly, 2011, Phylogenetic and experimental characterization of an acyl-ACP thioesterase family reveals significant diversity in enzymatic specificity and activity: Bmc Biochemistry, v. 12.

Jungwirth, H., and K. Kuchler, 2006, Yeast ABC transporters - A tale of sex, stress, drugs and aging: Febs Letters, v. 580, p. 1131-1138.

Koffas, M., and G. Stephanopoulos, 2005, Strain improvement by metabolic engineering: lysine production as a case study for systems biology: Curr Opin Biotechnol, v. 16, p. 361-6.

Lawrence, C. L., C. H. Botting, R. Antrobus, and P. J. Coote, 2004, Evidence of a new role for the high-osmolarity glycerol mitogen-activated protein kinase pathway in yeast: Regulating adaptation to citric acid stress: Molecular and Cellular Biology, v. 24, p. 3307-3323.

Legras, J. L., C. Erny, C. Le Jeune, M. Lollier, Y. Adolphe, C. Demuyter, P. Delobel, B. Blondin, and F. Karst, 2010, Activation of Two Different Resistance Mechanisms in Saccharomyces cerevisiae upon Exposure to Octanoic and Decanoic Acids: Applied and Environmental Microbiology, v. 76, p. 7526-7535.

Liu, P., A. Chernyshov, T. Najdi, Y. Fu, J. Dickerson, S. Sandmeyer, and L. Jarboe, 2013, Membrane stress caused by octanoic acid in Saccharomyces cerevisiae: Applied Microbiology and Biotechnology, v. 97, p. 3239-3251.

Maaheimo, H., J. Fiaux, Z. P. Çakar, J. E. Bailey, U. Sauer, and T. Szyperski, 2001, Central carbon metabolism of Saccharomyces cerevisiae explored by biosynthetic fractional 13C labeling of common amino acids: European Journal of Biochemistry, v. 268, p. 2464-2479.

Minty, J. J., A. A. Lesnefsky, F. M. Lin, Y. Chen, T. A. Zaroff, A. B. Veloso, B. Xie, C. A. McConnell, R. J. Ward, D. R. Schwartz, J. M. Rouillard, Y. A. Gao, E. Gulari, and X. N. Lin, 2011, Evolution combined with genomic study elucidates genetic bases of isobutanol tolerance in Escherichia coli: Microbial Cell Factories, v. 10.

Mira, N. P., J. D. Becker, and I. Sa-Correia, 2010a, Genomic Expression Program Involving the Haa1p-Regulon in Saccharomyces cerevisiae Response to Acetic Acid: Omics-a Journal of Integrative Biology, v. 14, p. 587-601.

Mira, N. P., A. B. Lourenco, A. R. Fernandes, J. D. Becker, and I. Sa-Correia, 2009, The RIM101 pathway has a role in Saccharomyces cerevisiae adaptive response and resistance to propionic acid and other weak acids: Fems Yeast Research, v. 9, p. 202-216.

Mira, N. P., M. C. Teixeira, and I. Sa-Correia, 2010b, Adaptive Response and Tolerance to Weak Acids in Saccharomyces cerevisiae: A Genome-Wide View: Omics-a Journal of Integrative Biology, v. 14, p. 525-540.

Nikolau, B. J., M. A. D. N. Perera, L. Brachova, and B. Shanks, 2008, Platform biochemicals for a biorenewable chemical industry: Plant Journal, v. 54, p. 536-545.

Osman, Y. A., and L. O. Ingram, 1985, MECHANISM OF ETHANOL INHIBITION OF FERMENTATION IN ZYMOMONAS-MOBILIS CP4: Journal of Bacteriology, v. 164, p. 173-180.

Peebles, C. A., G. W. Sander, E. H. Hughes, R. Peacock, J. V. Shanks, and K. Y. San, 2010, The expression of 1-deoxy-D-xylulose synthase and geraniol-10-hydroxylase or anthranilate synthase increases terpenoid indole alkaloid accumulation in Catharanthus roseus hairy roots: Metab Eng, v. 13, p. 234-40.

Piper, P., Y. Mahe, S. Thompson, R. Pandjaitan, C. Holyoak, R. Egner, M. Muhlbauer, P. Coote, and K. Kuchler, 1998, The Pdr12 ABC transporter is required for the development of weak organic acid resistance in yeast: Embo Journal, v. 17.

Ramos, J. L., E. Duque, M. T. Gallegos, P. Godoy, M. I. Ramos-Gonzalez, A. Rojas, W. Teran, and A. Segura, 2002, Mechanisms of solvent tolerance in gram-negative bacteria: Annual Review of Microbiology, v. 56, p. 743-768.

Ranganathan, S., T. W. Tee, A. Chowdhury, A. R. Zomorrodi, J. M. Yoon, Y. Fu, J. V. Shanks, and C. D. Maranas, 2012, An integrated computational and experimental study for overproducing fatty acids in Escherichia coli: Metabolic Engineering, v. 14, p. 687-704.

Sauer, U., 2006, Metabolic networks in motion: 13C-based flux analysis: Molecular Systems Biology, v. 2.

Segura, A., L. Molina, S. Fillet, T. Krell, P. Bernal, J. Munoz-Rojas, and J. L. Ramos, 2012, Solvent tolerance in Gram-negative bacteria: Current Opinion in Biotechnology, v. 23, p. 415-421.

Shanks, B. H., 2010, Conversion of Biorenewable Feedstocks: New Challenges in Heterogeneous Catalysis: Industrial & Engineering Chemistry Research, v. 49, p. 10212-10217.

Spira, F., N. S. Mueller, G. Beck, P. von Olshausen, J. Beig, and R. Wedlich-Soldner, 2012, Patchwork organization of the yeast plasma membrane into numerous coexisting domains: Nature Cell Biology, v. 14, p. 640-+.

Sriram, G., D. B. Fulton, V. V. Iyer, J. M. Peterson, R. L. Zhou, M. E. Westgate, M. H. Spalding, and J. V. Shanks, 2004, Quantification of compartmented metabolic fluxes in developing soybean embryos by employing Biosynthetic ally directed fractional C-13 labeling, C-13, H-1 two-dimensional nuclear magnetic resonance, and comprehensive isotopomer balancing: Plant Physiology, v. 136, p. 3043-3057.

Steen, E. J., Y. Kang, G. Bokinsky, Z. Hu, A. Schirmer, A. McClure, S. B. del Cardayre, and J. D. Keasling, 2010, Microbial production of fatty-acid-derived fuels and chemicals from plant biomass: Nature, v. 463.

Stephanopoulos, G., 1999, Metabolic fluxes and metabolic engineering: Metab Eng, v. 1, p. 1-11.

Stephanopoulos, G., 2007, Challenges in engineering microbes for biofuels production: Science, v. 315.

van Rijsewijk, B. R. B. H., A. Nanchen, S. Nallet, R. J. Kleijn, and U. Sauer, 2011, Large-scale C-13-flux analysis reveals distinct transcriptional control of respiratory and fermentative metabolism in Escherichia coli: Molecular Systems Biology, v. 7, p. -.

Viegas, C. A., P. F. Almeida, M. Cavaco, and I. Sa-Correia, 1998, The H+-ATPase in the plasma membrane of Saccharomyces cerevisiae is activated during growth latency in octanoic acid-supplemented medium accompanying the decrease in intracellular pH and cell viability: Applied and Environmental Microbiology, v. 64.

Viegas, C. A., M. F. Rosa, I. Sacorreia, and J. M. Novais, 1989, INHIBITION OF YEAST GROWTH BY OCTANOIC AND DECANOIC ACIDS PRODUCED DURING ETHANOLIC FERMENTATION: Applied and Environmental Microbiology, v. 55.

Wang, Z., J. Zhuge, H. Fang, and B. A. Prior, 2001, Glycerol production by microbial fermentation: A review: Biotechnology Advances, v. 19, p. 201-223.

Yamaguchi, M., 1979, STUDIES ON REGULATORY FUNCTIONS OF MALIC ENZYMES .4. EFFECTS OF SULFHYDRYL-GROUP MODIFICATION ON THE CATALYTIC FUNCTION OF NAD-LINKED MALIC ENZYME FROM ESCHERICHIA-COLI: Journal of Biochemistry, v. 86.

Yim, H., R. Haselbeck, W. Niu, C. Pujol-Baxley, A. Burgard, J. Boldt, J. Khandurina, J. D. Trawick, R. E. Osterhout, R. Stephen, J. Estadilla, S. Teisan, H. B. Schreyer, S. Andrae, T. H. Yang, S. Y. Lee, M. J. Burk, and S. Van Dien, 2011, Metabolic engineering of Escherichia coli for direct production of 1,4-butanediol: Nat Chem Biol, v. 7, p. 445-452.

Yin, Z. K., R. J. Smith, and A. J. P. Brown, 1996, Multiple signalling pathways trigger the exquisite sensitivity of yeast gluconeogenic mRNAs to glucose: Molecular Microbiology, v. 20.

List of figures

Figure 1. Phenotype analysis of *S.cerevisiae* under 0.4mM C8 fatty acid inhibition. Specific glucose uptake rate and ethanol yield were scaled down 100x and 10x respectively. Unit: specific cell growth rate, hr-1; specific glucose uptake rate, mmol/g DCW.hr; biomass yield, g DCW/g glucose; fermentation product yields, mol/mol glucose

Figure 2 *In vivo* metabolic flux distribution for *S.cerevisiae* BY4741 under control condition (red font) and 0.4mM octanoic acid stress (blue font) as calculated via isotopomer balancing using NMR2Flux software. Estimated fluxes are normalized to 100 mmol/g DCW.hr based on the average specific glucose uptake rates. The colored arrows indicate a significant difference (>20%) between the control and fatty acid production condition. The flux values shown are average from two replicates ± standard deviation.

Figure 3 Cofactor (NADH, NADPH) and energy (ATP) production in central carbon metabolism. The red triangle represents NADH, the purple triangle represents NADPH and the green rectangular represents ATP. The total production in central carbon metabolism for control and fatty acid producing conditions is show in the graph. Malic enzyme activity and aldehyde dehydrogenase involve both NADH and NADPH, the overall reducing power production can be captured by NAD(P)H.

Figure 4. Mechanistic model for the yeast response to octanoic acid stress. The H^+ATPase ATP Binding Cassette transporter is stimulated in the membrane plasma to recover the intracellular pH.

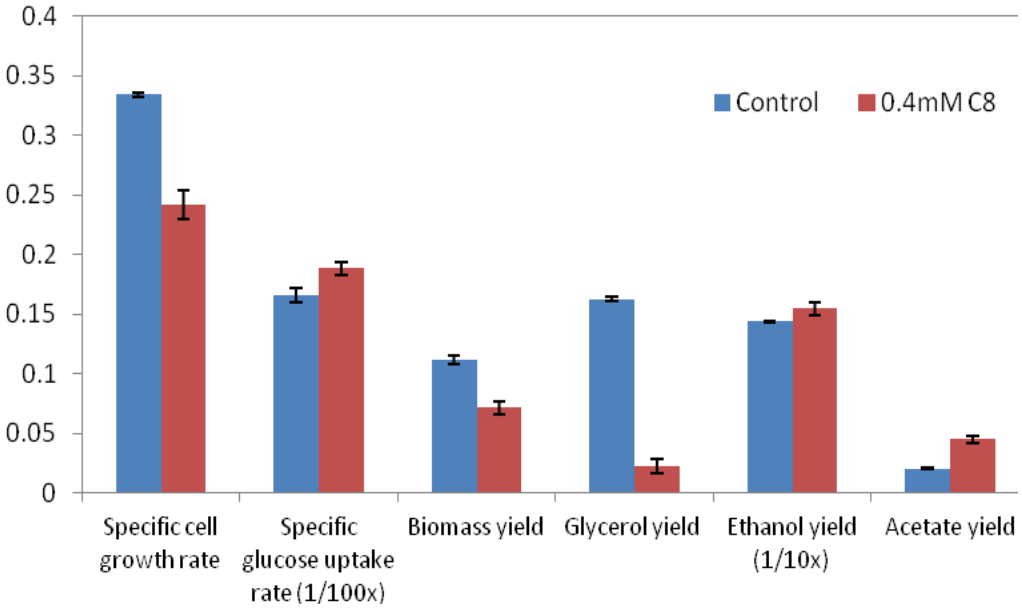


Figure 1. Phenotype analysis of *S.cerevisiae* under 0.4mM C8 fatty acid inhibition. Specific glucose uptake rate and ethanol yield were scaled down 100x and 10x respectively. Unit: specific cell growth rate, hr-1; specific glucose uptake rate, mmol/g DCW.hr; biomass yield, g DCW/g glucose; fermentation product yields, mol/mol glucose

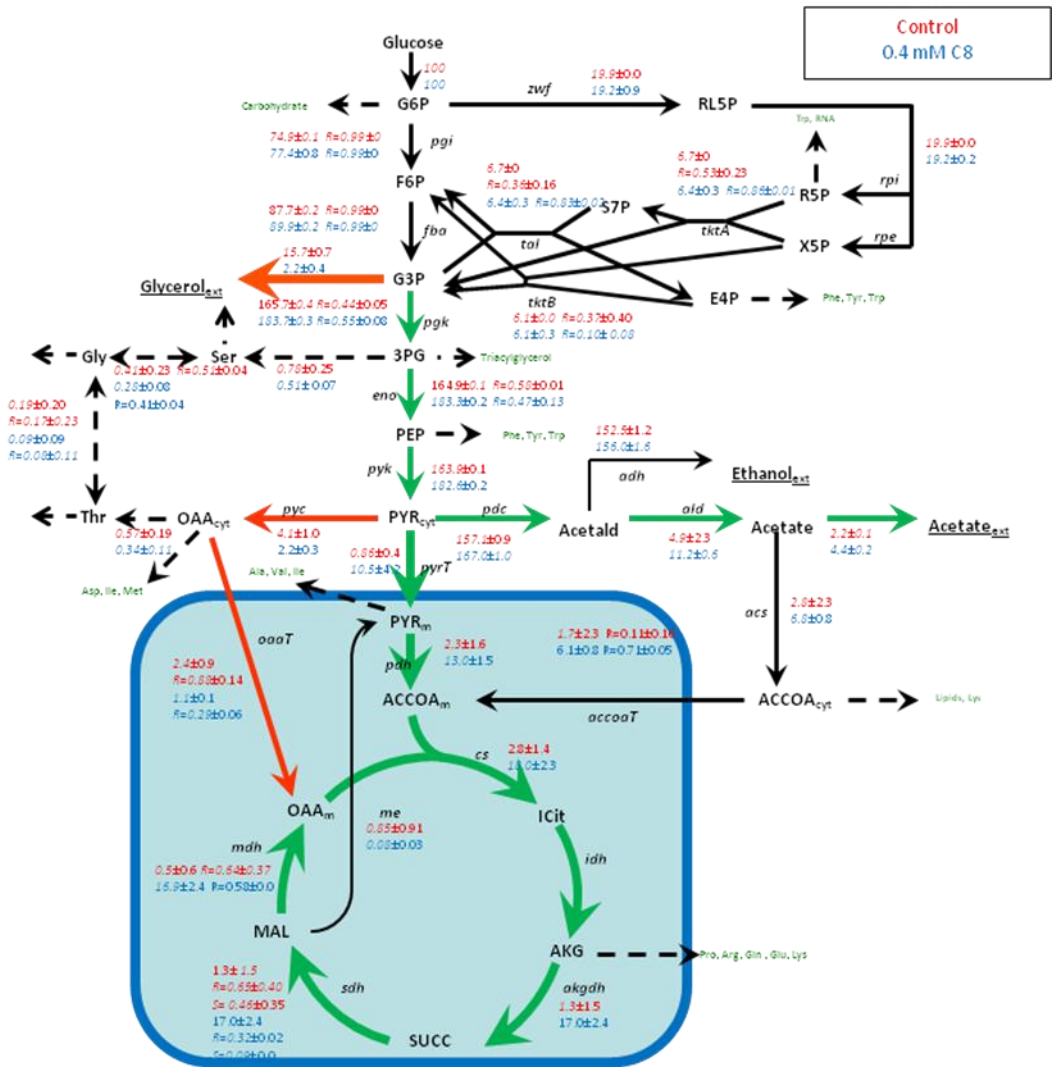


Figure 2 *In vivo* metabolic flux distribution for *S.cerevisiae* BY4741 under control condition (red font) and 0.4mM octanoic acid stress (blue font). Estimated fluxes are normalized to 100 mmol/g DCW.hr based on the average specific glucose uptake rates. The colored arrows indicate a significant difference (>20%) between the control and fatty acid production condition. The flux values shown are average from two replicates ± standard deviation.

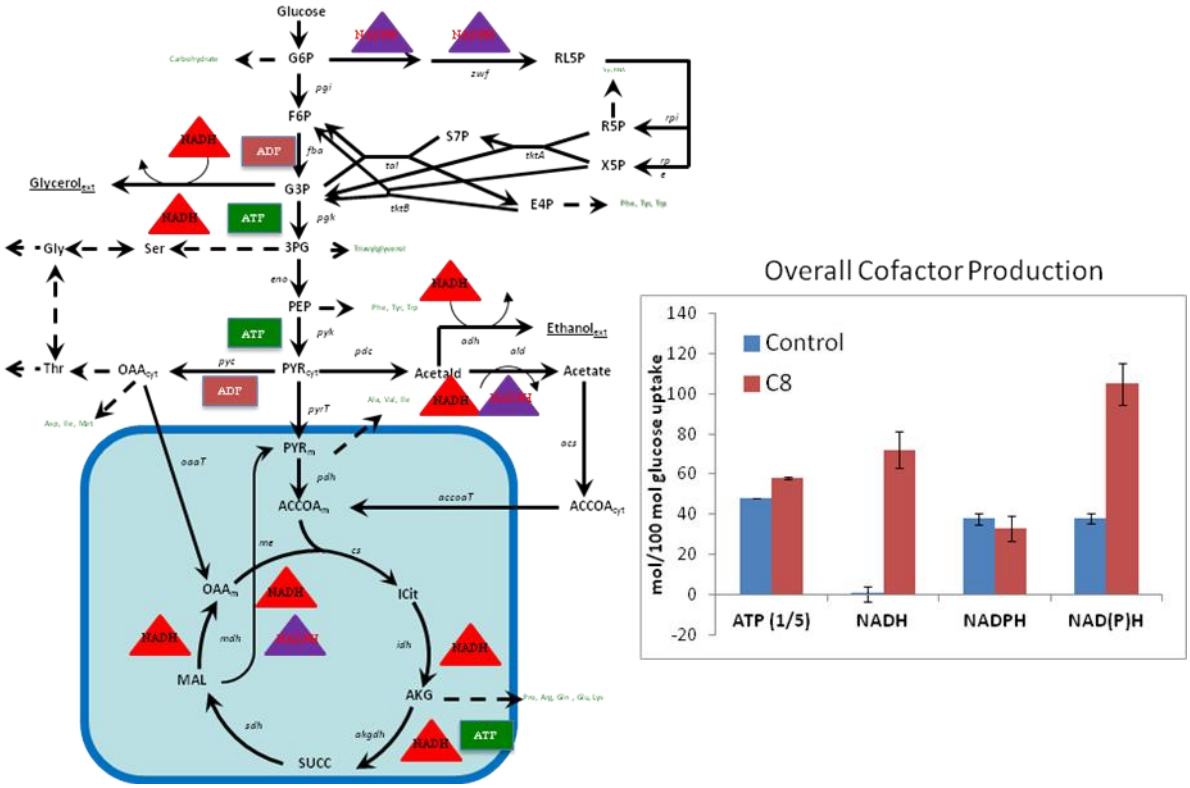


Figure 3 Cofactor (NADH, NADPH) and energy (ATP) production in central carbon metabolism. The red triangle represents NADH, the purple triangle represents NADPH and the green rectangular represents ATP. The total production in central carbon metabolism for control and fatty acid producing conditions is show in the graph. Malic enzyme activity and aldehyde dehydrogenase involve both NADH and NADPH, the overall reducing power production can be captured by NAD(P)H.

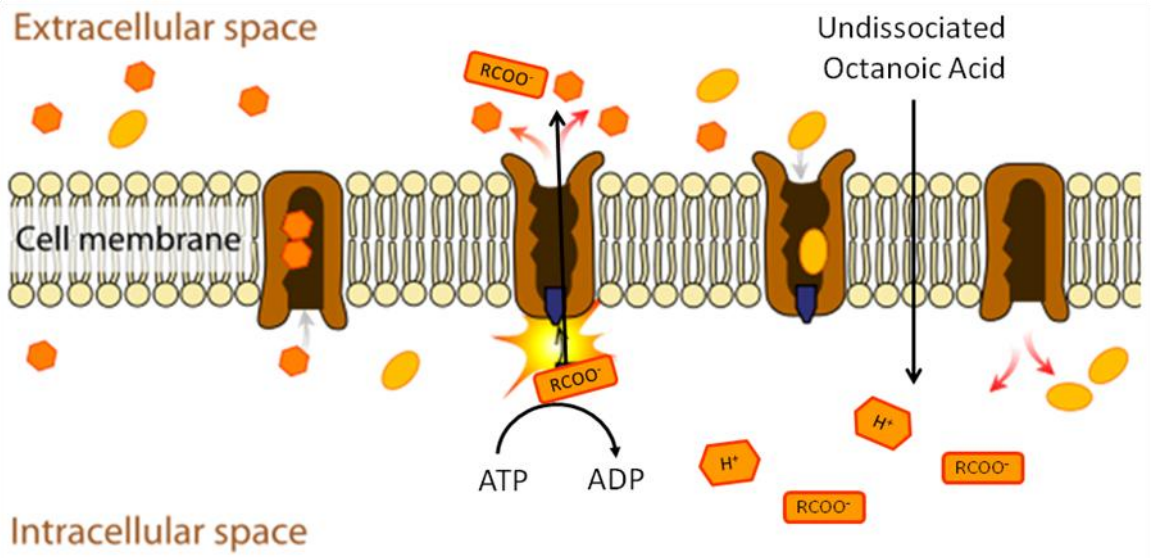


Figure 4. Mechanistic model for the yeast response to octanoic acid stress. The H^+ATPase ABC transporter is stimulated in the membrane plasma to recover the intracellular pH.

Supplementary information

Figure S1 Short-chain fatty acids hexanoic acid (C6) and octanoic acid (C8) inhibit yeast growth. The specific growth rate of *S. cerevisiae* BY4741 was measured in optimized SD minimal media, pH 5.0, 30^0C, 150 rpm with various concentrations of the indicated acids.

Figure S2. Amino acid uptake rate during the exponential growth of *S.cerevisiae* BY4741. The strain was grown aerobically in the fermentor at 30^oC, 600rpm and pH 5.0.

Figure S3. Transcript level perturbation in the central carbon metabolism under 0.3mM octanoic acid inhibition along with the flux difference from another independent experiment under 0.4mM octanoic acid stress. The transcriptomic experiment was conducted in shaker flasks using SDC medium under 0.3mM octanoic, the gene perturbations were reported as the ratio to the control condition shown in the parenthesis. Color and green reaction arrows representation downregulation and upregulation of fluxes respectively under 0.4mM octanoic acid exposure in the fermentor under SD minimal medium.

Table S1: Specific growth rate of S.cerevisiae BY 4741 under octanoic acid exposure in shake flasks and fermentor system.

Table S2. Metabolic pathways represented by stoichiometry, atom transition, and involved genes.

Table S3. Comparison of flux values of *S.cerevisiae* BY4741 between the control and octanoic acid inhibition condition.

Table S4. Measured and simulated isotopomer distributions of proteinogenic amino acids from 2D-NMR

Table S5: Gene transcripts significantly pretreated by 0.3mM octanoic acid stress(Liu et al., 2013)

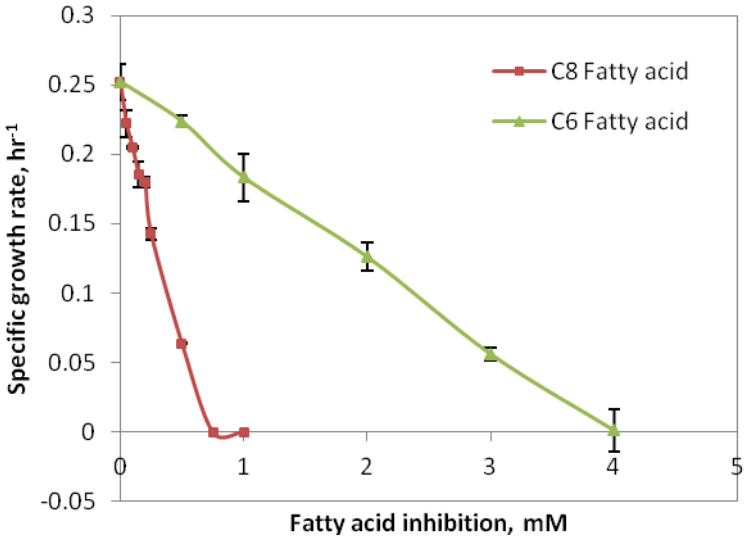


Figure S1 Short-chain fatty acids hexanoic acid (C6) and octanoic acid (C8) inhibit yeast growth. The specific growth rate of *S. cerevisiae* BY4741 was measured in optimized SD minimal media, pH 5.0, 30^0C, 150 rpm with various concentrations of the indicated acids.

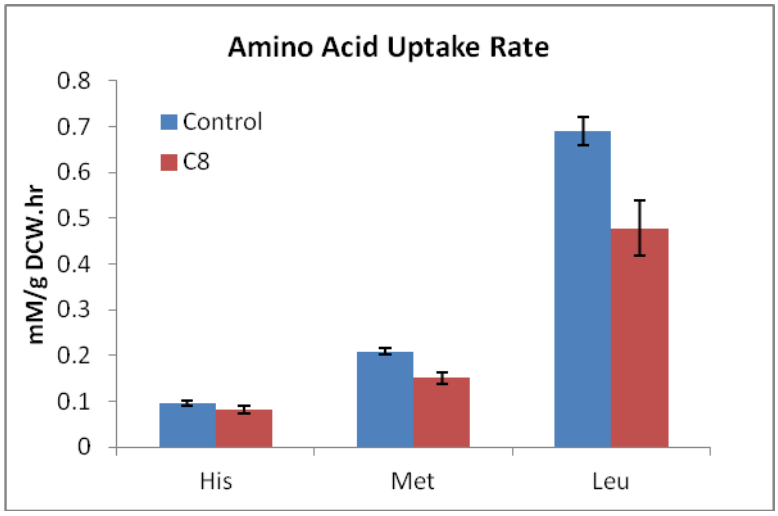


Figure S2. Amino acid uptake rate during the exponential growth of *S.cerevisiae* BY4741. The strain was grown aerobically in the fermentor at 30^oC, 600rpm and pH 5.0.

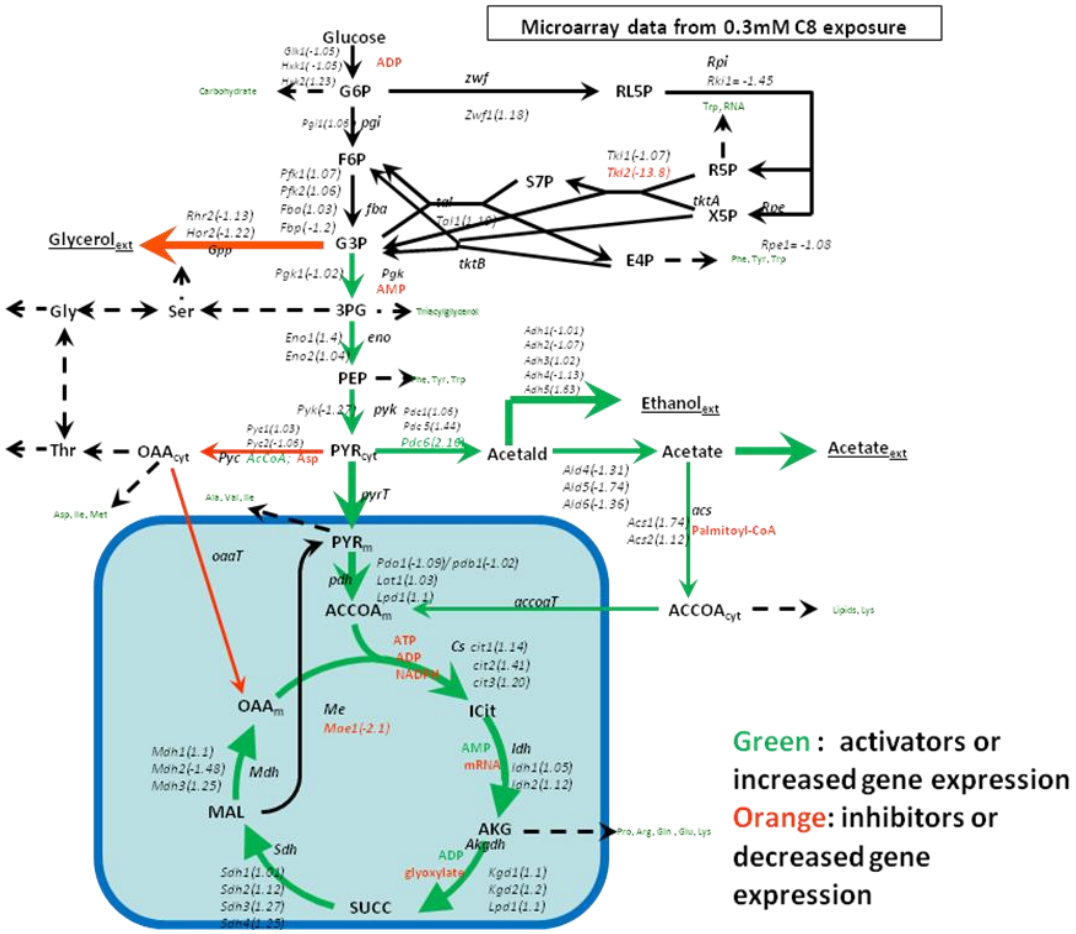


Figure S3. Transcript level perturbation in the central carbon metabolism under 0.3mM octanoic acid inhibition along with the flux difference from another independent experiment under 0.4mM octanoic acid stress. The transcriptomic experiment was conducted in shaker flasks using SDC medium under 0.3mM octanoic, the gene perturbations were reported as the ratio to the control condition shown in the parenthesis. Color and green reaction arrows representation downregulation and upregulation of fluxes respectively under 0.4mM octanoic acid exposure in the fermentor under SD minimal medium.

Table S1: Specific growth rate of *S.cerevisiae* BY 4741 under octanoic acid exposure in shake flasks and fermentor system.

	Specific growth rate, hr^{-1}	
Condition	Flask	Fermentor
EtOH Control	0.252 ±0.013	0.338
0.15mM C8	0.186 ±0.009	0.303
0.25mM C8	0.143 ±0.004	0.261

Table S2. Metabolic pathways represented by stoichiometry, atom transition, and involved genes from yeastgenome database.

Reaction name	Stoichiometry chemistry	Atom transition	Gene
Glycolysis pathway			
gluin	Glu → G6P	abcdef → abcdef	glk1, hxk1, hxk2
pgi	G6P → F6P	abcdef → abcdef	pgi
fba	FBP → T3P+T3P	abcdef → cba+def	pfk1, pfk2, fba1
pgk	T3P → 3PG	abc → abc	tdh1, tdh2, tdh3, pgk1
glyc	T3P → glycerol	abc → abc	gpd1, gpd2, rhr2, hor2
eno	3PG → PEP	abc → abc	gpm1,eno1,eno2
pyk	PEP → PYR	abc → abc	pyk2, cdc19
Pentose phosphate pathway			
zwf	G6P → 6PG	abcdef → abcdef	zwf
rpi	6PG → R5P+CO2	abcdef → bcdef+a	sol3,sol4,gnd1,gnd2
tkt	R5P+R5P → S7P+T3P	abcde+ABCDE → abABCDE+cde	tkt1,tkt2
tktAB	R5P+E4P → F6P+T3P	abcde+ABCD → abABCD+cde	tkt1
talf	S7P+T3P → F6P+E4P	abcdefg+ABC → abcABC+defg	tal1
TCA cycle			
mdh	MALm → OAAm	abcd → abcd	mdh1
frd	SUCCm→MALm	abcd → abcd	sdh1,sdh2,sdh3,sdh4,fum1
icd	ACCOAm+OAAm→AKGm+CO2	ab+ABCD→DCBba+A	cit1,cit3,aco1,aco3,idh1,idh2
succ	AKGm→SUCCm+CO2	abcde→bcde+a	lpd1,kgd1,kgd2,lsc1,lsc2
Acetyl-CoA biosynthesis and fermentative pathway			
ace	PYRm→ACCOAm+CO2	abc→bc+a	pdb1, pda1, lat1
pyc	PYR+CO2→ OAA	abc+d→abcd	pyc1,pyc2
pdc	PYR→ACETALD+CO2	abc→bc+a	pdc6.pdc5,pdc1
adh	ACETALD→EtOH	ab→ab	adh1,adh2,adh3,adh4,adh5

Table S2 continued

ald	ACETALD→ACE	ab→ab	ald2,ald5,ald4
acs	ACE+CO2→ACCOA	ab→ab	acs1,acs2
Anaplerotic pathway			
ana	Mal → PYR+CO2	abcd → abc+d	mae1
C1 metabolism			
ser	3PG → Ser	abc → abc	ser1,ser2,ser3,ser33
gly	Ser → Gly+C1	abc → ab+c	shm1,shm3
Amino acid biosynthesis and metabolic pathway			
thr	OAA→Thr	abcd→abcd	thr1,thr4
thrgly	Thr→Gly+ACCOA	abcd→ab+cd	gly1
Transport pathway			
ac	ACE→ACEout		
etoh	EtOH→EtOHout		
glyc	GLYC→GLYCout		
co2	CO2 →CO2out		
pyrt	PYR →PYRm		
oaat	OAA →OAAm		
acct	ACCOA→ACCOAm		

Table S3. Comparison of flux values of *S.cerevisiae* BY4741under the control and 0.4mM octanoic acid stress. The flux values are based on 100 mol/g/hr glucose uptake rate. The average and standard deviation is obtained from a duplicate of 13C flux experiments

reaction name	stoichiometry chemistry	Control		0.4mM C8	
		Average	SD	Average	SD
Glycolysis pathways					
gluin	Glu → G6P	100.00	0.00	100.00	0.00
pgi	G6P → F6P	74.88	0.13	77.44	0.80
	reversibility	0.99	0.00	0.99	0.01
fba	FBP → T3P+T3P	87.48	0.30	89.95	0.19
	reversibility	0.99	0.00	0.99	0.00
pgk	T3P → 3PG	165.70	0.35	183.77	0.29
	reversibility	0.44	0.05	0.55	0.08
glyc	T3P → glycerol	15.70	0.67	2.16	0.36
eno	3PG → PEP	164.92	0.10	183.26	0.22
	reversibility	0.58	0.01	0.48	0.13
pyk	PEP → PYR	163.92	0.07	182.62	0.21
Pentose phosphate pathway					
zwf	G6P → 6PG	20.08	0.05	19.22	0.89
rpi	6PG → R5P+CO2	20.08	0.05	19.22	0.89
tkt	R5P+R5P → S7P+T3P	6.66	0.01	6.43	0.30
	reversibility	0.53	0.23	0.86	0.01
tktAB	R5P+E4P → F6P+T3P	6.13	0.02	6.43	0.30
	reversibility	0.37	0.40	0.83	0.02
talf	S7P+T3P → F6P+E4P	6.66	0.01	6.08	0.31
	reversibility	0.36	0.16	0.10	0.08
TCA cycle					
mdh	MALm → OAAm	0.46	0.55	16.94	2.39
	reversibility	0.64	0.37	0.58	0.00
frd	SUCCm→MALm	1.32	1.46	17.03	2.36
	reversibility	0.65	0.40	0.32	0.02
	scrambling	0.46	0.35	0.09	0.00
icd	ACCOAm+OAAm→AKGm+CO2	2.83	1.41	18.03	2.28
succ	AKGm→SUCCm+CO2	1.32	1.46	17.03	2.36
Acetyl-CoA biosynthesis and fermentative pathway					
ace	PYRm→ACCOAm+CO2	1.07	0.91	11.94	1.45
pyc	PYR+CO2→ OAA	4.15	0.97	2.20	0.26
pdc	PYR→ACETALD+CO2	157.14	0.87	167.05	1.03
adh	ACETALD→EtOH	152.45	1.24	155.95	1.67
ald	ACETALD→ACE	4.87	2.31	11.19	0.55
acs	ACE+CO2→ACCOA	2.76	2.29	6.76	0.75

Table S3 continued

Anaplerotic pathway					
ana	Mal → PYR+CO2	0.85	0.91	0.09	0.03
C1 metabolism					
ser	3PG → Ser	0.78	0.25	0.52	0.07
gly	Ser → Gly+C1	0.41	0.23	0.28	0.08
	reversibility	0.51	0.04	0.41	0.04
Amino acid biosynthesis pathway					
thr	OAA→Thr	0.57	0.19	0.34	0.11
thrgly	Thr → Gly+ACCOA	0.19	0.20	0.09	0.09
	reversibility	0.17	0.23	0.08	0.11
Transport pathway					
ac	ACE→ACEout	2.09	0.18	4.43	0.20
etoh	EtOH→EtOHout	153.33	1.52	155.95	1.67
glyc	GLYC→GLYCout	15.23	1.34	2.16	0.36
co2	CO2 →CO2out	177.00	1.61	231.16	6.18
pyrt	PYR →PYRm	2.33	1.55	13.03	1.54
oaat	OAA →OAAm	2.36	0.86	1.09	0.12
	Reversibility	0.88	0.14	0.29	0.06
acct	ACCOA→ACCOAm	1.76	2.32	6.10	0.83
	Reversibility	0.12	0.17	0.71	0.05
Biomass synthesis pathway					
G6pb	G6P -> biomass	5.32	0.30	3.34	0.09
R5pb	R5P -> biomass	0.46	0.03	0.29	0.02
E4pb	E4P ->biomass	0.53	0.03	0.35	0.01
PEPb	PEP ->biomass	1.02	0.06	0.64	0.01
PYRb	PYR ->biomass	1.49	0.09	0.94	0.01
ACCOAb	ACCOA ->biomass	1.02	0.06	0.66	0.08
AKGb	AKG -> biomass	1.55	0.08	1.00	0.09
OAAb	OAA -> biomass	1.27	0.07	0.77	0.03
3PGb	3PG -> biomass	0.07	0.00	0.04	0.00
C1b	C1 ->biomass	0.57	0.04	0.28	0.08
serb	Ser ->biomass	0.39	0.02	0.39	0.02
glyb	Gly ->biomass	0.62	0.04	0.38	0.02
thrb	Thr ->biomass	0.39	0.02	0.25	0.02
alab	Ala->biomass	0.99	0.06	0.58	0.00

(1) Abbreviation: Glu, Glucose; G6P, Glucose-6-phosphate; F6P, Fructose-6-phosphate; T3P, Triose-3-phosphate; 3PG, 3-phosphoglycerate; PEP, Phosphoenolpyruvate; PYR, Pyruvate; ACCOA, Acetate-CoA; AKG, Ketoglutarate; SUCC, Succinate; MAL, Malate; OAA, Oxaloacetate; 6PG, 6-P-gluconate; R5P, Ribose-5-phosphate; S7P, Sedoheptulose-7-phosphate; E4P, Erythronate-4-phosphate; GOX, Glyoxylate; AC, Acetate; GLYC, glycerol;EtOH, ethanol; CO2, carbon dioxide; Ser, Serine; Gly, Glycine; Ala; alanine.

Table S4. Measured and simulated isotopomer distributions of proteinogenic amino acids from 2D-NMR for *S.cerevisiae* BY4741 strain grown in SD minimal media under the control and octanoic acid inhibtion.

Cross peak	Control 1		Control 2		C8-1		C8-2			
(multiplet)	Expt	Sim	Expt	Sim	Expt	Sim	Expt	Sim	Precursor	Isotopomer
Ala β (s)	0.700	0.675	0.671	0.646	0.728	0.653	0.695	0.653	Pyr	[x23]
Ala β (d)	0.300	0.325	0.329	0.354	0.273	0.347	0.305	0.347	Pyr	[x23]
Arg β (s)	0.063	0.082	0.081	0.084	0.153	0.187	0.161	0.192	AKGm	[x2**3**4x]
Arg β (d)	0.467	0.456	0.462	0.459	0.499	0.491	0.518	0.492	AKGm	[x**23**4x] + [x2**34**x]
Arg β (t)	0.471	0.462	0.457	0.457	0.348	0.322	0.384	0.316	AKGm	[x**234**x]
Arg γ (s)	0.419	0.500	0.456	0.485	0.345	0.383	0.356	0.366	AKGm	[xx3**4**5]
Arg γ (d)	0.413	0.419	0.396	0.427	0.435	0.473	0.442	0.481	AKGm	[xx**34**5] + [xx3**45**]
Arg γ (t)	0.168	0.081	0.148	0.088	0.220	0.143	0.202	0.152	AKGm	[xx**345**]
Arg δ (s)	0.082	0.079	0.094	0.076	0.090	0.060	0.079	0.068	AKGm	[xxx4**5**]
Arg δ (d)	0.918	0.921	0.906	0.924	0.910	0.940	0.921	0.932	AKGm	[xxx**45**]
Asp α (s)	0.127	0.101	0.144	0.099	0.143	0.084	0.165	0.107	OAA	[1**2**3x]
Asp α (d1)	0.096	0.085	0.035	0.042	0.076	0.063	0.083	0.089	OAA	[**12**3x]
Asp α (d2)	0.074	0.086	0.124	0.114	0.102	0.117	0.146	0.142	OAA	[1**23**x]
Asp α (dd)	0.702	0.728	0.698	0.744	0.679	0.736	0.607	0.662	OAA	[**123**x]
Asp β (s)	0.498	0.477	0.464	0.471	0.484	0.463	0.471	0.449	OAA	[x2**3**4]
Asp β (d1)	0.303	0.242	0.243	0.257	0.241	0.255	0.228	0.257	OAA	[x**23**4]
Asp β (d2)	0.112	0.183	0.158	0.167	0.166	0.180	0.165	0.184	OAA	[x2**34**]
Asp β (dd)	0.088	0.097	0.135	0.105	0.110	0.102	0.136	0.110	OAA	[x**234**]
Glu β (s)	0.034	0.082	0.055	0.084	0.127	0.187	0.152	0.192	AKGm	[x2**3**4x]
Glu β (d)	0.413	0.456	0.430	0.459	0.451	0.491	0.485	0.492	AKGm	[x**23**4x] + [x2**34**x]
Glu β (t)	0.483	0.462	0.467	0.457	0.347	0.322	0.324	0.316	AKGm	[x**234**x]
Glu γ (s)	0.500	0.500	0.482	0.485	0.419	0.383	0.416	0.366	AKGm	[xx3**4**5]
Glu γ (d1)	0.167	0.152	0.152	0.158	0.279	0.270	0.282	0.286	AKGm	[xx**34**5]
Glu γ (d2)	0.252	0.267	0.258	0.269	0.207	0.204	0.207	0.195	AKGm	[xx3**45**]
Glu γ (dd)	0.081	0.081	0.108	0.088	0.095	0.143	0.095	0.152	AKGm	[xx**345**]
Gly α (s)	0.156	0.143	0.140	0.130	0.162	0.151	0.148	0.144	Gly	[1**2**x]
Gly α (d)	0.844	0.857	0.860	0.870	0.838	0.849	0.852	0.856	Gly	[**12**x]

Table S4 continued

Ile α (s)	0.130	0.137	0.147	0.167	0.126	0.158	0.161	0.196	OAA/Pyrm	[12xx] {x2x]
Ile α (d1)	0.514	0.597	0.552	0.614	0.616	0.630	0.540	0.593	OAA/Pyrm	[12xx] {x2x]
Ile α (d2)	0.091	0.050	0.058	0.047	0.054	0.043	0.070	0.052	OAA/Pyrm	[12xx] {x2x]
Ile α (dd)	0.266	0.216	0.244	0.173	0.204	0.169	0.228	0.159	OAA/Pyrm	[12xx] {x2x]
Ile γ1(s)	0.584	0.528	0.620	0.568	0.627	0.566	0.615	0.557	Pyrm/OAA	[x2x] {xx34]
Ile γ1(d)	0.330	0.397	0.302	0.372	0.301	0.374	0.304	0.381	Pyrm/OAA	[x2x] {xx34] +[x2x] {xx34]
Ile γ1(t)	0.086	0.074	0.078	0.060	0.072	0.060	0.081	0.062	Pyrm/OAA	[x2x] {xx34]
Ile δ (s)	0.460	0.410	0.439	0.420	0.444	0.400	0.430	0.392	OAA	[xx34]
Ile δ (d)	0.540	0.590	0.562	0.580	0.557	0.600	0.570	0.608	OAA	[xx34]
Ile γ2 (s)	0.635	0.677	0.638	0.643	0.677	0.652	0.682	0.652	Pyrm	[x23]
Ile γ2 (d)	0.365	0.323	0.362	0.357	0.323	0.348	0.318	0.348	Pyrm	[x23]
Lys β (s)	0.279	0.281	0.258	0.269	0.227	0.249	0.200	0.242	AccoA/AKGm	[x2] [23xx]
Lys β (d)	0.527	0.524	0.518	0.520	0.529	0.511	0.517	0.507	AccoA/AKGm	[x2] [23xx]+ [x2] [23xx]
Lys β (t)	0.194	0.195	0.224	0.211	0.245	0.240	0.282	0.251	AccoA/AKGm	[x2] [23xx]
Lys γ (s)	0.034	0.082	0.058	0.084	0.119	0.187	0.133	0.192	AccoA/AKGm	[xx] [234x]
Lys γ (d)	0.464	0.456	0.472	0.459	0.519	0.491	0.541	0.492	AccoA/AKGm	[xx] [234x]+ [xx] [234x]
Lys γ (t)	0.502	0.462	0.470	0.457	0.362	0.322	0.327	0.316	AccoA/AKGm	[xx] [234x]
Lys δ (s)	0.498	0.500	0.512	0.485	0.399	0.383	0.438	0.366	AccoA/AKGm	[xx] [x345]
Lys δ (d)	0.418	0.419	0.422	0.427	0.440	0.473	0.468	0.481	AccoA/AKGm	[xx] [x345]+ [xx] [x345]
Lys δ (t)	0.084	0.081	0.067	0.088	0.161	0.143	0.104	0.152	AccoA/AKGm	[xx] [x345]
Lys ε (s)	0.098	0.079			0.100	0.060	0.107	0.068	AccoA/AKGm	[xx] [xx45]
Lys ε (d)	0.902	0.921			0.900	0.940	0.893	0.932	AccoA/AKGm	[xx] [xx45]
Phe α (s)	0.032	0.030	0.040	0.027	0.032	0.044	0.043	0.051	PEP	[123]
Phe α (d1)	0.021	0.003	0.028	0.003	0.031	0.011	0.061	0.015	PEP	[123]
Phe α (d2)	0.045	0.068	0.062	0.094	0.036	0.076	0.034	0.084	PEP	[123]
Phe α (dd)	0.902	0.899	0.870	0.875	0.901	0.869	0.862	0.851	PEP	[123]
Pheβ (s)	0.524	0.514	0.439	0.512	0.503	0.516	0.484	0.516	PEP	[x23].[2x]
Pheβ (d1)	0.298	0.277	0.334	0.279	0.318	0.274	0.310	0.274	PEP	[x23].[2x]
Pheβ (d2)	0.121	0.136	0.149	0.135	0.126	0.137	0.163	0.137	PEP	[x23].[2x]
Phe β (dd)	0.057	0.073	0.078	0.074	0.053	0.073	0.044	0.073	PEP	[x23].[2x]
Pro β (s)	0.041	0.082	0.059	0.084	0.132	0.187	0.151	0.192	AKGm	[x234x]

Table S4 continued

Pro β (d)	0.421	0.456	0.443	0.459	0.513	0.491	0.487	0.492	AKGm	[x**23**4x] + [x2**34**x]
Pro β (t)	0.538	0.462	0.498	0.457	0.354	0.322	0.361	0.316	AKGm	[x**234**x]
Pro γ(s)	0.467	0.500	0.450	0.485	0.400	0.383	0.364	0.366	AKGm	[xx3**45**]
Pro γ(d)	0.413	0.419	0.430	0.427	0.451	0.473	0.485	0.481	AKGm	[xx**345**] + [x23**45**]
Pro γ(t)	0.120	0.081	0.120	0.088	0.148	0.143	0.152	0.152	AKGm	[xx**345**]
Ser α (s)	0.033	0.040	0.038	0.037	0.023	0.050	0.040	0.054	Ser	[**12**3]
Ser α (d1)	0.100	0.120	0.120	0.134	0.119	0.111	0.138	0.110	Ser	[**12**3]
Ser α (d2)	0.077	0.073	0.026	0.087	0.077	0.079	0.040	0.083	Ser	[**12**3]
Ser α (dd)	0.790	0.768	0.816	0.741	0.781	0.760	0.782	0.752	Ser	[**123**]
Ser β (s)	0.728	0.690	0.718	0.697	0.728	0.688	0.690	0.688	Ser	[x**23**]
Ser β (d)	0.272	0.310	0.282	0.303	0.272	0.312	0.310	0.312	Ser	[x**23**]
Thr α (s)	0.126	0.101	0.130	0.099	0.141	0.084	0.159	0.107	THR	[**12**3x]
Thr α (d1)	0.081	0.087	0.093	0.074	0.039	0.064	0.119	0.097	THR	[**12**3x]
Thr α (d2)	0.144	0.086	0.105	0.110	0.103	0.116	0.067	0.140	THR	[**12**3x]
Thr α (dd)	0.649	0.726	0.673	0.717	0.718	0.735	0.655	0.656	THR	[**123**x]
Thr β (s)	0.566	0.477	0.476	0.462	0.544	0.463	0.496	0.447	THR	[x2**34**]
Thr β (d)	0.352	0.426	0.419	0.431	0.365	0.435	0.405	0.443	THR	[x**234**]+ [x**23**4]
Thr β (t)	0.082	0.097	0.105	0.107	0.091	0.102	0.099	0.110	THR	[x**234**]
Thr γ2 (s)	0.426	0.412	0.380	0.445	0.408	0.401	0.399	0.401	THR	[xx3**4**]
Thr γ2 (d)	0.574	0.588	0.620	0.555	0.592	0.599	0.601	0.599	THR	[xx**34**]
Tyr α (s)	0.033	0.030	0.033	0.027	0.031	0.044	0.045	0.051	PEP	[**12**3]
Tyr α (d1)	0.037	0.003	0.024	0.003	0.020	0.011	0.064	0.015	PEP	[**12**3]
Tyr α (d2)	0.034	0.068	0.083	0.094	0.059	0.076	0.064	0.084	PEP	[**12**3]
Tyr α (dd)	0.897	0.899	0.860	0.875	0.891	0.869	0.828	0.851	PEP	[**123**]
Tyr β (s)	0.515	0.514	0.519	0.512	0.519	0.516	0.554	0.516	PEP	[x**23**].[2x]
Tyr β (d)	0.346	0.413	0.328	0.415	0.363	0.411	0.326	0.411	PEP	[x**23**].[2x]+ [x**23**].[**2**x]
Tyr β (t)	0.139	0.073	0.153	0.074	0.118	0.073	0.120	0.073	PEP	[x**23**].[**2**x]
Tyr δ (s)	0.555	0.512	0.592	0.509	0.545	0.514	0.603	0.512	PEP/E4P	[xxx].[**23**].[1x]+[xxx].[2].[**43**x]
Tyr δ (d)	0.379	0.414	0.351	0.416	0.389	0.412	0.344	0.413	PEP/E4P	[xxx].[**23**].[1x]+[xxx].[**2**].[**43**x]+ [xxx].[2**3**].[**1**x]+[xxx].[2].[**43**x]
Tyr δ (t)	0.066	0.074	0.058	0.075	0.066	0.074	0.054	0.074	PEP/E4P	[xxx].[**23**].[**1**x]+[xxx].[**2**].[**43**x]

Table S4 continued

Val α (s)	0.185	0.210	0.173	0.128	0.093	0.099	0.113	0.109	Pyrm	[1**2**x] {x2x]
Val α (d1)	0.535	0.524	0.585	0.652	0.663	0.689	0.637	0.680	Pyrm	[**12**x] {x2x]
Val α (d2)	0.084	0.076	0.058	0.036	0.029	0.027	0.036	0.029	Pyrm	[1**2**x] {x**2**x]
Val α (dd)	0.197	0.190	0.184	0.184	0.215	0.185	0.214	0.182	Pyrm	[**12**x] {x**2**x]
Val γ1 (s)	0.668	0.677	0.661	0.643	0.703	0.652	0.709	0.652	Pyrm	[x2**3**]
Val γ1 (d)	0.332	0.323	0.339	0.357	0.297	0.348	0.291	0.348	Pyrm	[x**23**]
Val γ2 (s)	0.747	0.734	0.757	0.780	0.807	0.788	0.785	0.789	Pyrm	[x2x] {xx**3**]
Val γ2 (d)	0.253	0.266	0.243	0.220	0.193	0.212	0.215	0.211	Pyrm	[x**2**x] {xx**3**]

Table S5: Gene transcripts significantly pretreated by 0.3mM octanoic acid stress (Liu et al., 2013)

Name	Y-ID	Fold	p
Plasma membrane ABC proteins			
PDR16	YNL231c	2.01	0.001
PDR5	YOR153w	1.60	0.000
PDR15	YDR406w	4.51	0.000
PDR12	YPL058c	4.18	0.000
YOR1	YGR281w	1.63	0.005
PDR3	YBL005w	3.42	0.000
Membrane proteins			
HSP12	YFL014w	-4.81	0.000
AQR1	YNL065w	-5.04	0.014
DIC1	YLR348c	2.03	0.000
SUR7	YML052w	-2.02	0.001
FMP45	YDL222c	-10.34	0.072
HSP12	YFL014w	-4.81	0.000
SUR7	YML052w	-2.02	0.001
AQR1	YNL065w	-5.04	0.014
SPI1	YER150w	-3.67	0.009
MMP1	YLL061w	-2.32	0.007
Uptake of iron, zinc, copper, potassium, calcium and ammonia			
IZH4	YOL101c	2.02	0.001
IZH1	YDR492w	-2.11	0.008
ATX1	YNL259c	2.39	0.001
CCC2	YDR270w	2.02	0.000
FIT2	YOR382w	12.60	0.000
FIT3	YOR383c	5.19	0.000
SIT1	YEL065w	2.87	0.000
QDR2	YIL121w	2.20	0.000
FIG1	YBR040w	3.72	0.000
ATO2	YNR002c	2.22	0.015
GRE2	YOL151w	2.18	0.001

CHAPTER 7

CHARACTERISTIC OF YARROWIA LIPOLYTICA FOR LIPID PRODUCTION

Authors: Ting Wei Tee[1], Ivan Chang[2], James Yu[2], Suzanne Sandmeyer[2] and Jacqueline V. Shanks[1]

[1] Department of Chemical and Biological Engineering, Iowa State University, Ames, IA

[2] Department of Biological Chemistry, University of California, Irvine, CA

Abstract

Fatty acids synthesized via fermentation from biorenewable feedstocks are a potential source of platform chemicals, and thus could help addressing the increasing environmental impacts and insecurity of future energy. However, toxicity of fatty acids remains an obstacle for high titer production of fatty acid in *Sacchromyces cerevisiae.* The commercialization of fatty acids using *S. cerevisiae* is not feasible and uneconomical unless all the toxicity issue is resolved in metabolic engineering cycle. Oleaginous yeasts, which store energy as lipids, may have enhanced potential as host organism for the production of industrially chemicals through polyketide pathways. *Yarrowia lipolytica* is the most extensively studied oleaginous yeast with complete genome sequenced, thus could become an attractive platform organism. However, the metabolism of *Y. lipolytica* is less known compared to the conventional yeast. In this study, we characterized phenotype and morphology of *Y. lipolytica* under nitrogen and carbon-limited environment. We observed higher fatty acid production under nitrogen starvation and possible correlation between filamentous cells and lipid accumulation. Glycerol 3-phosphate dehydrogenase (*gut2*), which oxidizes glycerol-3-phosphate to dihydroxyacetone, was deleted to increase fatty acid production. Metabolic flux analysis (MFA) was applied to quantify fluxes in metabolic pathways and identify metabolic regulation by comparing fluxes under different environments. We elucidated the metabolic flux differences in central metabolism between *Y. lipolytica* wild type and *gut2*Δ mutant by conducting ^{13}C labeling experiments in shake flasks with a mixture of uniformly ^{13}C labeled glucose and 1-^{13}C positional labeled glucose. No significant distinction in central metabolism flux distribution between the wild type and *gut2*Δ

mutant was observed under carbon-limited environment. Interestingly, the TCA cycle and pentose phosphate pathway were the major pathway for generations of NADPHs and ATPs for cellular requirement. Fluxomic data will be integrated with RNASeq transcriptomic data of *Y. lipolytica* into the hybrid kinetics-flux balance analysis (FBA) framework. The hybrid model, which includes dynamics regulation of gene and enzymes, will be incorporated into *in-silico* computational tools to guide the strain design.

1. Introductions

The demands for sustainable production of fuels and chemicals have substantially increased with detrimental environmental impact of fossil-based industry. Microbial fermentation using renewable feedstock has potential to replace the traditional fossil-based production and thus address environmental and energy concerns. Carboxylic acids derived from microbial fermentation could be the platform biochemicals to generate an array of industrial chemicals (Chia et al., 2012; Nikolau et al., 2008). For instance, free fatty acid synthesized from biocatalysts could be converted to alkanes, fatty esters, and fatty alcohols (Lennen et al., 2010; Steen et al., 2010). The synthesis of these biochemicals involved polyketide pathway, particularly acetyl-CoA and malonyl-CoA. However, the growth of *Sacchromyces cerevisiae* is known to be inhibited by the presence of lipophilic acids (Mira et al., 2010). Jarboe and coworkers reported dose-dependent and chain length-dependent toxicity effects on S. cerevisiae (Liu et al., 2013). The cell growth was completely inhibited by less than 5 mM of $C_{6\text{-}10}$ fatty acids due to the loss of membrane integrity (Liu et al., 2013). The product toxicity problem renders the strain to be infeasible and uneconomical for future commercialization. To overcome this menace, directed evolution and genetic engineering might root in success (Jarboe et al., 2011). However, we could select other host organisms that can naturally accumulate fatty acid without much product toxicity issue.

Yarrowia lipolytica, an oleaginous yeast, can accumulate up to 70% of its biomass in lipids, making it an attractive testbed organism for strain development (Beopoulos et al., 2009). Unlike *Saccharomyces* which stores its energy as polysaccharide, *o*leaginous yeasts store energy as lipids and may therefore have enhanced potential for production of industrially desirable compounds (Beopoulos et al., 2009). *Y. lipolytica* is an obligate aerobic yeast that can survive under hydrophobic substrates (alkane, fatty acid and hydrocarbons). This oleaginous yeast was

known to accumulate fatty acids as lipid bodies under nitrogen starvation (Beopoulos et al., 2009). During nitrogen exhaustion, the organism could not continue to proliferate through the synthesis of proteins and nucleic acids. The carbon source is assimilated to synthesize fatty acids with the resulting accumulation of triglycerides as lipid bodies. The complete genome of *Y. lipolytica* has been sequenced (Dujon et al., 2004) and was found to be distantly related to the conventional *S. cerevisiae.* Similarly, genome-scale metabolic network of *Y. lipolytica* has been recently released to guide the metabolic engineering and conceptualize high-throughput data (Loira et al., 2012; Pan and Hua, 2012). The development of genetic tools for *Y. lipolytica* establishes the technology platform to engineer the strain. For instance, Dupont has demonstrated the production of omega-3 and omega-6 polyunsaturated fatty acids in *Y. lipolytica* by expressing heterogeneous genes encoding the ω-3/ω-6 biosynthetic pathway (Beopoulos et al., 2010). Thus, *Y. lipolytica* is an excellent organism for polyketide pathway-based biochemical products.

However, oleaginous yeast metabolism is not as well understood as that of *S. cerevisiae* so that improvements in characterization of oleaginous strains would be useful for strain development. ^{13}C metabolic flux analysis (MFA) provides a realistic *in vivo* profile of metabolic flux distributions (fluxomics) of an organism by relying on the ^{13}C carbon tracing patterns as additional constraints rather than a growth objective function typical in flux balance analysis (FBA) (Chen et al., 2011). ^{13}C MFA using GC-MS has been implemented to obtain the fluxomics of the yeasts under various conditions (Kleijn et al., 2007). With the availability of complete genome sequence, transcriptomics analysis could be performed to understand cell physiology and regulatory mechanism at the whole-cell transcript level by analyzing them under various genotypic and environmental conditions simultaneously (Wang et al., 2009). In addition, computational tool can be applied with the availability of genome-scale metabolic model to identify genetic manipulation to enhance product yield (Lee et al., 2012). However, most of the in-silico tools do not include gene regulation (Burgard et al., 2003; Ranganathan et al., 2010). Integration of fluxomic data and existing RNASeq transcriptomic data of wild type *Y. lipolytica* into the hybrid kinetics-FBA framework of the genome-scale reconstructed model of *Y. lipolytica* could overcome this limitation by accounting for the dynamics range of enzymes.

In this study, we investigated the growth properties, the respiration rate, and the morphology of Y. *lipolytica* under aerobic fermentation with glucose as feedstock. We performed ^{13}C-MFA to compare carbon flux distribution of wild type *Y.lipolytica* and its

engineered mutant strain. Truncated metabolic network was constructed based on the genome-scale metabolic model and applied for ^{13}C- MFA simulations.

2. Material and Methods

Strains and plasmids:

The wild type *Y. lipolytica* and ΔGUT2 mutant were used throughout this work. The strains were gifts from Dr. Suzanne Sandmeyer's lab.

Strains and culture conditions:

The *Y. lipolytica* strains from glycerol frozen stock was streaked on Yeast Peptone Dextrose (YPD) plate and incubated overnight at 30^{o}C. The colonies from YPD plate was transferred in 50mL tubes containing 10mL of the following synthetic dextrose (SD) minimal medium: 10 g/L glucose and 6.7 g/L yeast nitrogen base without amino acid. The medium pH was adjusted to 5.5. The preculture cells were centrifuged at 4000 rpm for 5 min at 4℃. The supernatant were discarded, and the pellet was re-suspended in fresh SD minimal medium. The appropriate quantity of the washed cell suspension was used to inoculate 50 mL of SD minimal medium in the 250mL flasks to a starting OD_{600} of 0.02. For better identification of fluxes, a mixture of uniformly labeled [U-^{13}C], first carbon labeled [1-^{13}C] and natural glucose was used for ^{13}C flux analysis (Fischer et al., 2004). Specifically, 20% U-^{13}C glucose and 80% 1-^{13}C glucose was used as the tracer to final medium concentration of 1% glucose. The aerobic fermentation was conducted at 25 ℃ and 250 rpm. The cells were harvested at mid-exponential phase after at least 5 generations to ensure metabolic and isotopic steady state. 3 parallel experiments were conducted as biological replicates.

Analytical techniques:

Cell biomass dry weight was determined by measuring optical density OD_{600} using a spectrophotometer. Cell dry weight was estimated by drying centrifuged cell pellets (20mL) in the 80^{o}C oven. The increase in tube weight was measured and correlated with the sample's OD. Biomass composition was determined based on literature data (Pan and Hua, 2012). Media samples were taken over the times and filtered through 0.22 μm pore sized nylon filters (P.J. Cobert Associates, Saint Louis, MO) and kept at -80 ^{o}C for extracellular metabolite analysis. Only glucose was detected using Waters HPLC (Waters, Milford, MA) with 410 refractive index

detector. The Aminex column (HPX-87H, Bio-Rad, Hercules, CA) was used at 30 ^{o}C with 0.3 mL/min of 5mM sulfuric acid as mobile phase. Glucose was analyzed using enzymatic assay kit (Sigma, St Louis, MO) according the manufacturer protocol. The substrate uptake rate in batch culture is constant during exponential phase.

The morphology of the oleaginous yeasts was observed under confocal microscope where the cells were dyed with Nile red to stain the lipid bodies. The oxygen consumption rate was measured by Hansetech oxygen probe according to the manufacturer's protocol.

Fatty acid quantification:

1-2 mL samples were collected and processed for fatty acid analysis as described by Ranganathan et al. (Ranganathan et al., 2012). The fatty acids were extracted using chloroform and methylated into fatty acid methyl esters for GC-MS/FID analysis. The fatty acids were also analyzed using sulfo-phospho-vanilin assay (Knight et al., 1972).

Proteinegenic amino acid analysis:

Cells were prepared as described by Zamboni et al. (Zamboni et al., 2009). Briefly, cells are centrifuged, washed twice with saline water containing 0.9% NaCl, and then hydrolysed with 6 N hydrochloric acid at 110^{o}C for 18-24 hours. The samples were filtered and the acids were evaporated. Finally, the sample was dissolved in DMF and derivitized with TBDMS for GC-MS analysis. The isotopomer peak integrations were performed in Chemstation software. Natural abundance of the amino acids were corrected using FiatFlux software (Zamboni et al., 2005).

Metabolic network model for MFA:

A network model for *Y. lipolytica* metabolism was constructed based on the available genome-scale network models (see Table S2) (Loira et al., 2012; Pan and Hua, 2012). The model includes glucose transport and phosphorylation pathway, Embden-Meyerhof-Parnas pathway, oxidative pentose phosphate branch, non-oxidative pentose phosphate branch, TCA cycle, anaplerotic pathways, glyoxylate cycle, metabolite exchange reactions, all amino acids biosynthesis pathways. Eukaryotic compartmentation into mitochondrial and cytosolic subsystems was included in the model by considering distinct pools of pyruvate, oxaloacetate, citrate and acetyl-CoA in both compartments (Fiaux et al., 2003). The pyruvate transport into the mitochondria is

unidirectional, driven by the proton motive force. Phosphoenolpyruvate carboxylase kinase (Yin et al., 1996) was repressed under the presence of glucose, thus not included in the network.

Flux evaluation methodology:

Fluxes were quantified using 13C-FLUX2 software developed by Wiechert et al. (Weitzel et al., 2013). 13C-FLUX2 employs cumomer/ elementary metabolite unit (EMU) balancing and a global optimization routine to find stoichiometrically feasible fluxes set consistent with experimental measurements. Overall fluxes were estimated by minimizing the chi-square error between experimentally measured and simulated isotopomer fractions. Statiscal analysis was performed using a bootstrap Monte Carlo or linearized statistical analysis.

3. Results and discussions

Phenotype characterization of *Y. lipolytica*

We characterized the growth properties and lipid production of *Y. lipolytica* under carbon and nitrogen limitation conditions. A growth curve for both of these conditions was constructed (Figure 1a). Lipid production was measured for cells in log, late log, early stationary, and late stationary. In late stationary, nitrogen-limited cultures produced four-fold greater lipid than carbon-limited cells. The major fatty acid was C18:1, with second highest C18:2 in both carbon- and nitrogen-limited growth. This reflected greater than 0.02 g fatty acid per gram wet weight in the nitrogen-limited sample (Figure 1b). The fatty acids produced under nitrogen-limited condition were significantly higher (>5 fold) than carbon-limited environment.

The mitochondrial glycerol-3-phosphate dehydrogenase (g*ut2)* is involved in glycerol degradation to oxidize glycerol-3-phosphate to dihydroxyacetone in mitochondria. Glycerol-3-phosphate is then transported back to cytosol to enter glycolysis or gluconeogenesis. *Gut2* gene was deleted to prevent the degradation of glycerol, in which glycerol is the precursor for triglyceride biosynthesis. *Gut2* was deleted using *URA3*, flanked upstream and downstream by 1000bp of *gut2* flanking sequence, selected on ura- medium, screened on glycerol for no growth, and confirmed by PCR (data not shown). This strain was grown under nitrogen-limiting conditions and unsaturated lipids were evaluated using the phospho-vanillin assay. The unsaturated fatty acid contents of *gut2*Δ mutant was higher than the control (data not shown), but the whole fatty acid profile would need to analyzed using GC-MS.

We characterized a cellular morphology difference between the wild type and *gut2*Δ *Y. lipolytica* in stationary phase. The *gut2Δ* mutant maintains a high percentage of filamentous cells compared to the budding cells in the wild type *Y. lipolytica* (Figure 2). The increase of filamentous behavior might be linked with nutrient limitation and the elevation of lipid production. In addition, we investigated the effect of oxygen consumption on the filamentous percentage during mid-logarithm and stationary growth stage (Figure 3). With glucose as the carbon source, we observed the percentage of filamentous cells was higher during mid-log growth than during the stationary phase. Similarly, oxygen consumption was higher during mid-log phase. This might imply respiration (which requires oxygen) is correlated to the tendency of cells to be filamentous, indirectly linked to lipid production.

Metabolic flux analysis

^{13}C MFA experiments were performed on both wild type and *gut2*Δ *Y. lipolytica* using a combination of 20% uniformly labeled ^{13}C glucose and 80% 1-^{13}C labeled glucose as the only carbon tracers. The phenotype analysis revealed similar growth rate, glucose consumption rate and biomass yield between wild type and mutant (Figure 4a, 4b). The growth rate of wild type and *gut2*Δ *Y. lipolytica* were very similar (0.21 hr^{-1}) during growth phase. With higher glucose consumption, the biomass yield of *gut2*Δ mutant decreased 25% compared to that of wild type. However, Sauer and coworkers reported higher growth rate and glucose consumption rate in their study (Christen and Sauer, 2011). The deviations might be possibly due to differences in cultivation temperature and medium. No by-product formation was detected by HPLC analysis. The isotopomer fraction of proteinogenic amino acids quantified using GC-MS depicted high similarity between wild type and *gut2*Δ mutant (Figure 5).

We have successfully constructed a data consistent metabolic model of *Y. lipolytica* for ^{13}C MFA via multiple rounds of iterative modifications to the metabolic model of the previous *S. cerevisiae* ^{13}C MFA (Gombert et al., 2001) (Figure 6). *S. cerevisiae* network model was used as the scaffold. *Y. lipolytica* specific reactions from genome scale reaction models were incorporated incrementally (Loira et al., 2012). Finally, cross-validation of the resulting simulated fluxes between the general net flux ratio approach and the detailed global optimization approach was used as the fitness function to determine next iteration. The resulting fluxomics from the carbon tracing simulations via 13CFlux2 are then mapped to the model using Omix (Figure 7). The flux distributions

between *Y. lipolytica* and *gut2*Δ mutant were similar at carbon-limited condition at exponential growth phase. Most of the carbon fluxes (around 83%) channel through pentose phosphate pathway to generate NADPH for reduction requirements. The remaining carbon flux (around 13%) is directed towards the glycolytic pathway, resulting in low activities of the lower glycolytic pathway. The TCA cycle operates at 57 mmol $gDW^{-1}\,h^{-1}$ (based on 100 mmol $gDW^{-1}\,h^{-1}$) to generate ATP and NAD(P)H for energy and reduction requirements for cell growth. The glyoxylate cycle is active with 15 mmol $gDW^{-1}\,h^{-1}$ to bypass the TCA cycle. Anaplerotic reactions (phosphoenolpyruvate carboxylkinase and malic enzyme reactions) are not significant (Fischer and Sauer, 2003)(Fischer and Sauer, 2003)(Fischer and Sauer, 2003)(Fischer and Sauer, 2003)(Fischer and Sauer, 2003)(Fischer and Sauer, 2003)(Fischer and Sauer, 2003).

Contrast with *S. cerevisiae* (chapter 6), glucose was mostly catabolized through glycolysis to form ethanol, glycerol and acetate. The TCA cycle flux was negligible. However, pentose pathway pathway and the TCA cycle oleaginous *Y. lipolytica* became the major pathways for sugar catabolism and cofactor generations. These findings were in agreement with Christen et al. (Christen and Sauer, 2011) that aerobic *Yarrowia* species exhibited high respiratory fluxes. High similarity in flux distribution between wild type and *gut2*Δ *Y. lipolytica* might be due to the similar phenotype and fatty acid production under carbon-limited condition. Difference in flux distributions might be observed under nitrogen limited environment where *gut2*Δ mutant produced more lipids than the wild type, which is yet to be investigated.

4. Summary and discussion

Due to the membrane disruption, *S.cerevisiae* can only tolerate less than 5mM octanoic acid which is not feasible for commercialization (Liu et al., 2013). Unlike *S. cerevisiae* which stores its energy as polysaccharide, *o*leaginous yeasts store energy as lipids and may have enhanced potential for production of industrially desirable compounds and tolerance against lipid inhibition. Thus we characterized the growth and lipid production of *Y. lipolytica* under carbon- and nitrogen-limited conditions. Nitrogen deprivation promoted *Y. lipolytica* to accumulate lipid bodies. Glycerol-3-phophaste dehydrogense deletion led to higher unsaturated fatty acid production correlated with filamentous Yarrowia cells formation. Respiration rate of wild type *Y. lipolytica* was observed to be higher during growth phase compared to stationary stage. Metabolic flux analysis was performed on *Y. lipolytica* wild type strain and its *gut2*Δ mutant under carbon-limited environment. The TCA cycle

and pentose phosphate pathway fluxes were the major pathways to generate ATPs and NADPHs. RNAseq experiments would be conducted at University of California, Irvine to identify the key metabolic differences between conditions and mutant isolates which favor and disfavor lipid synthesis. The fluxomic data will be integrated with RNASeq transcriptomic data of *Y. lipolytica* into the hybrid kinetics-flux balance analysis (FBA) framework. The hybrid kinetics-FBA model will include gene and enzyme dynamic regulation in order to predict the range of possible *in-vivo* fluxes in each enzymatic reaction. Finally, the hybrid-FBA could be incorporated into OptForce (Ranganathan et al., 2010) to predict for minimal genetic manipulation for yield improvement of desired chemicals.

References

Beopoulos, A., T. Chardot, and J.-M. Nicaud, 2009, Yarrowia lipolytica: A model and a tool to understand the mechanisms implicated in lipid accumulation: Biochimie, v. 91, p. 692-696.

Beopoulos, A., T. Desfougéres, J. Sabirova, and J. M. Nicaud, 2010, Yarrowia lipolytica as a Cell Factory for Oleochemical Biotechnology, *in* K. Timmis, ed., Handbook of Hydrocarbon and Lipid Microbiology, Springer Berlin Heidelberg, p. 3003-3010.

Burgard, A. P., P. Pharkya, and C. D. Maranas, 2003, OptKnock: A bilevel programming framework for identifying gene knockout strategies for microbial strain optimization: Biotechnology and Bioengineering, v. 84, p. 647-657.

Chen, X., A. P. Alonso, D. K. Allen, J. L. Reed, and Y. Shachar-Hill, 2011, Synergy between C-13-metabolic flux analysis and flux balance analysis for understanding metabolic adaption to anaerobiosis in E. coli: Metabolic Engineering, v. 13, p. 38-48.

Chia, M., T. J. Schwartz, B. H. Shanks, and J. A. Dumesic, 2012, Triacetic acid lactone as a potential biorenewable platform chemical: Green Chemistry, v. 14, p. 1850-1853.

Christen, S., and U. Sauer, 2011, Intracellular characterization of aerobic glucose metabolism in seven yeast species by 13C flux analysis and metabolomics: Fems Yeast Research, v. 11, p. 263-272.

Dujon, B., D. Sherman, G. Fischer, P. Durrens, S. Casaregola, I. Lafontaine, J. de Montigny, C. Marck, C. Neuveglise, E. Talla, N. Goffard, L. Frangeul, M. Aigle, V. Anthouard, A. Babour, V. Barbe, S. Barnay, S. Blanchin, J. M. Beckerich, E. Beyne, C. Bleykasten, A. Boisrame, J. Boyer, L. Cattolico, F. Confanioleri, A. de Daruvar, L. Despons, E. Fabre, C. Fairhead, H. Ferry-Dumazet, A. Groppi, F. Hantraye, C. Hennequin, N. Jauniaux, P. Joyet, R. Kachouri, A. Kerrest, R. Koszul, M. Lemaire, I. Lesur, L. Ma, H. Muller, J. M. Nicaud, M. Nikolski, S. Oztas, O. Ozier-Kalogeropoulos, S. Pellenz, S. Potier, G. F. Richard, M. L. Straub, A. Suleau, D. Swennen, F. Tekaia, M. Wesolowski-Louvel, E. Westhof, B. Wirth, M. Zeniou-Meyer, I. Zivanovic, M. Bolotin-Fukuhara, A. Thierry, C.

Bouchier, B. Caudron, C. Scarpelli, C. Gaillardin, J. Weissenbach, P. Wincker, and J. L. Souciet, 2004, Genome evolution in yeasts: Nature, v. 430, p. 35-44.

Fiaux, J., Z. P. Cakar, M. Sonderegger, K. Wuthrich, T. Szyperski, and U. Sauer, 2003, Metabolic-flux profiling of the yeasts Saccharomyces cerevisiae and Pichia stipitis: Eukaryotic Cell, v. 2.

Fischer, E., and U. Sauer, 2003, Metabolic flux profiling of Escherichia coli mutants in central carbon metabolism using GC-MS: European Journal of Biochemistry, v. 270, p. 880-891.

Fischer, E., N. Zamboni, and U. Sauer, 2004, High-throughput metabolic flux analysis based on gas chromatography-mass spectrometry derived C-13 constraints: Analytical Biochemistry, v. 325, p. 308-316.

Gombert, A. K., M. M. dos Santos, B. Christensen, and J. Nielsen, 2001, Network identification and flux quantification in the central metabolism of Saccharomyces cerevisiae under different conditions of glucose repression: Journal of Bacteriology, v. 183.

Jarboe, L. R., P. Liu, and L. A. Royce, 2011, Engineering inhibitor tolerance for the production of biorenewable fuels and chemicals: Current Opinion in Chemical Engineering, v. 1, p. 38-42.

Kleijn, R. J., J.-M. A. Geertman, B. K. Nfor, C. Ras, D. Schipper, J. T. Pronk, J. J. Heijnen, A. J. A. van Maris, and W. A. van Winden, 2007, Metabolic flux analysis of a glycerol-overproducing Saccharomyces cerevisiae strain based on GC-MS, LC-MS and NMR-derived C-13-labelling data: Fems Yeast Research, v. 7.

Knight, J. A., J. M. Rawle, and S. Anderson, 1972, CHEMICAL BASIS OF SULFO-PHOSPHO-VANILLIN REACTION FOR ESTIMATING TOTAL SERUM-LIPIDS: Clinical Chemistry, v. 18, p. 199-&.

Lee, J. W., D. Na, J. M. Park, J. Lee, S. Choi, and S. Y. Lee, 2012, Systems metabolic engineering of microorganisms for natural and non-natural chemicals: Nature Chemical Biology, v. 8.

Lennen, R. M., D. J. Braden, R. M. West, J. A. Dumesic, and B. F. Pfleger, 2010, A Process for Microbial Hydrocarbon Synthesis: Overproduction of Fatty Acids in Escherichia coli and Catalytic Conversion to Alkanes: Biotechnology and Bioengineering, v. 106.

Liu, P., A. Chernyshov, T. Najdi, Y. Fu, J. Dickerson, S. Sandmeyer, and L. Jarboe, 2013, Membrane stress caused by octanoic acid in Saccharomyces cerevisiae: Applied Microbiology and Biotechnology, v. 97, p. 3239-3251.

Loira, N., T. Dulermo, J.-M. Nicaud, and D. J. Sherman, 2012, A genome-scale metabolic model of the lipid-accumulating yeast Yarrowia lipolytica: Bmc Systems Biology, v. 6.

Mira, N. P., M. C. Teixeira, and I. Sa-Correia, 2010, Adaptive Response and Tolerance to Weak Acids in Saccharomyces cerevisiae: A Genome-Wide View: Omics-a Journal of Integrative Biology, v. 14, p. 525-540.

Nikolau, B. J., M. A. D. N. Perera, L. Brachova, and B. Shanks, 2008, Platform biochemicals for a biorenewable chemical industry: Plant Journal, v. 54, p. 536-545.

Pan, P., and Q. Hua, 2012, Reconstruction and In Silico Analysis of Metabolic Network for an Oleaginous Yeast, Yarrowia lipolytica: Plos One, v. 7.

Ranganathan, S., P. F. Suthers, and C. D. Maranas, 2010, OptForce: An Optimization Procedure for Identifying All Genetic Manipulations Leading to Targeted Overproductions: Plos Computational Biology, v. 6.

Ranganathan, S., T. W. Tee, A. Chowdhury, A. R. Zomorrodi, J. M. Yoon, Y. Fu, J. V. Shanks, and C. D. Maranas, 2012, An integrated computational and experimental study for overproducing fatty acids in Escherichia coli: Metabolic Engineering, v. 14, p. 687-704.

Steen, E. J., Y. Kang, G. Bokinsky, Z. Hu, A. Schirmer, A. McClure, S. B. del Cardayre, and J. D. Keasling, 2010, Microbial production of fatty-acid-derived fuels and chemicals from plant biomass: Nature, v. 463.

Wang, Z., M. Gerstein, and M. Snyder, 2009, RNA-Seq: a revolutionary tool for transcriptomics: Nature Reviews Genetics, v. 10.

Weitzel, M., K. Noeh, T. Dalman, S. Niedenfuehr, B. Stute, and W. Wiechert, 2013, 13CFLUX2-high-performance software suite for C-13-metabolic flux analysis: Bioinformatics, v. 29, p. 143-145.

Yin, Z. K., R. J. Smith, and A. J. P. Brown, 1996, Multiple signalling pathways trigger the exquisite sensitivity of yeast gluconeogenic mRNAs to glucose: Molecular Microbiology, v. 20.

Zamboni, N., S.-M. Fendt, M. Ruhl, and U. Sauer, 2009, 13C-based metabolic flux analysis: Nat. Protocols, v. 4, p. 878-892.

Zamboni, N., E. Fischer, and U. Sauer, 2005, FiatFlux - a software for metabolic flux analysis from C-13-glucose experiments: Bmc Bioinformatics, v. 6.

List of figures

(a) (b)

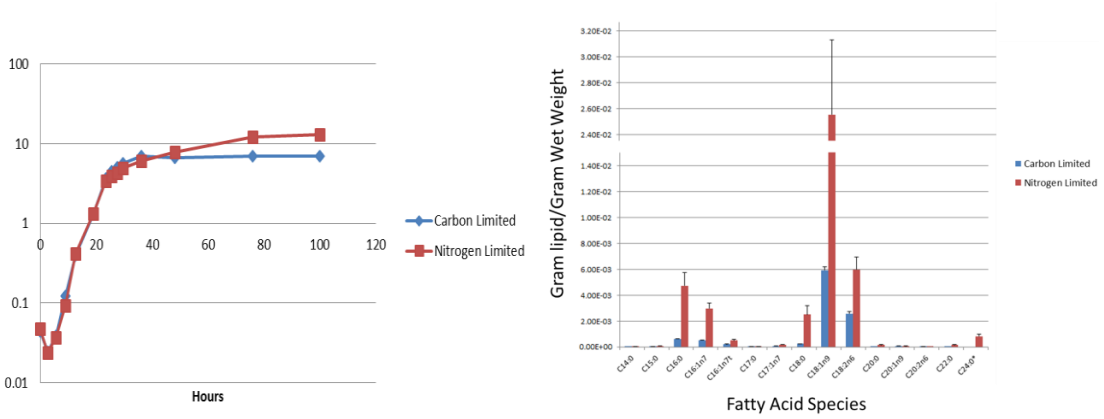


Figure 1 (a) The growth curve and (b) lipid productions of of *Y. lipolytica* under carbon-limited and nitrogen-limited condition.

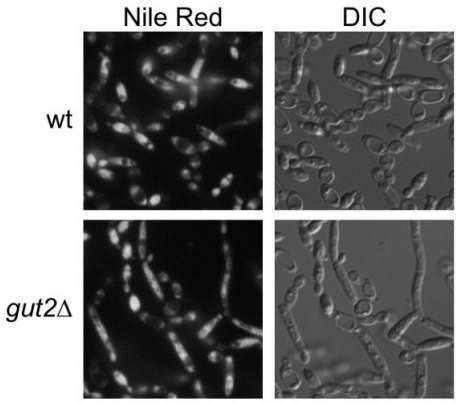


Figure 2. Dimorphism in *Y. lipolytica*. Morphology of WT vs *gut2Δ* in early stationary phase via differential interference contrast microscopy. Nile red dye stains for intracellular lipid droplets. Images provided by Virginia Bilanc of the Sandmeyer lab.

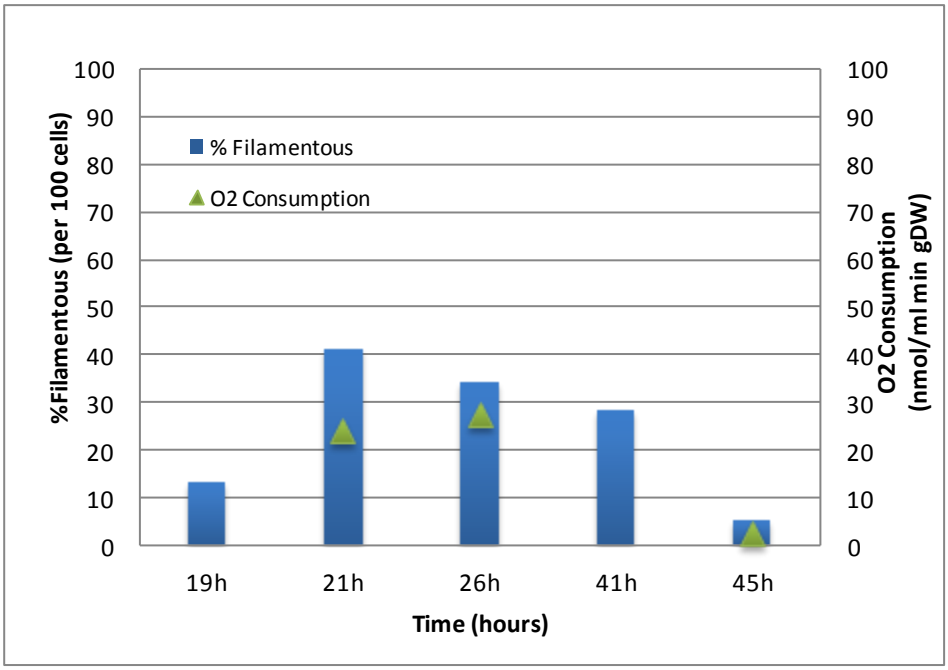


Figure 3. The filamentous percentage and oxygen consumption of *Y. lipolytica* wild type strain during mid-log growth (19, 21 and 26 hours) and stationary phase (41 and 45 hours). The fermentation was performed in shake flasks at 25°C and 250 rpm with glucose as the carbon source.

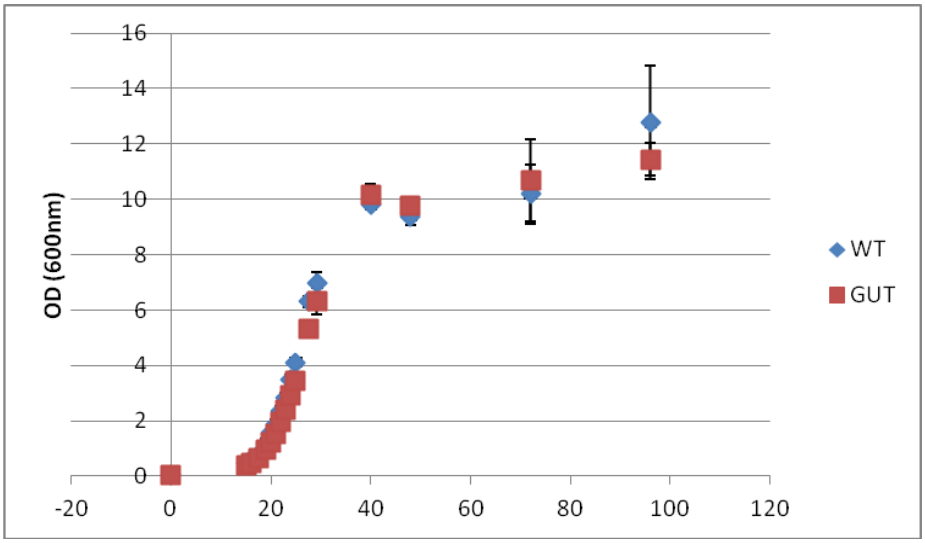


Figure 4a. The growth profile of *Y. lipolytica* wild type (WT) and *gut2*Δ (GUT) mutant during ^{13}C labeling experiment with 1% glucose (80% uniformly ^{13}C labeled and 20% 1-^{13}C labeled). The strain was grown aerobically in the shake flasks at 25°C and 250 rpm. The average and standard deviation are based on at least 3 replicates.

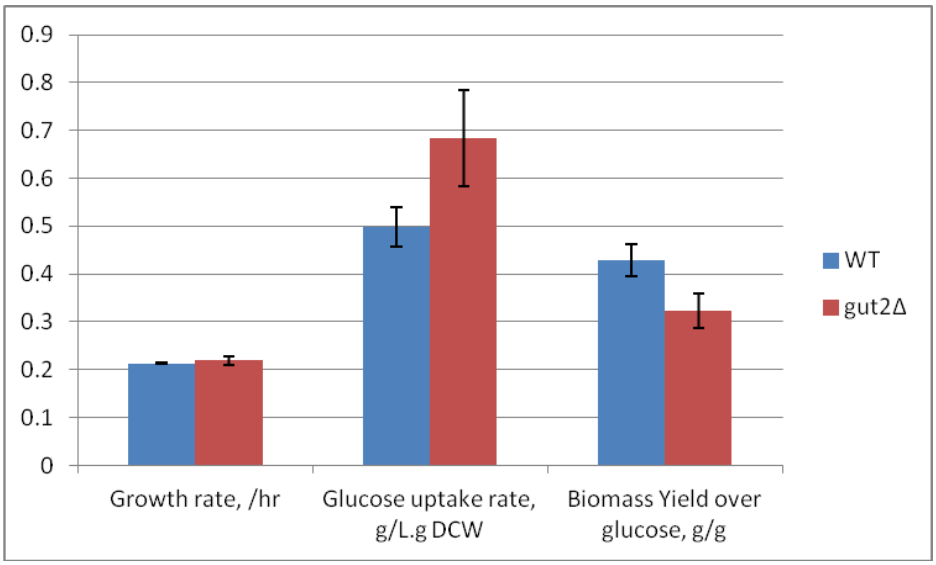


Figure 4b. Phenotype analysis of *Y. lipolytica* wild type and *gut2*Δ mutant. The strains were grown using 1% glucose (80% uniformly ^{13}C labeled and 20% 1-^{13}C labeled). The strain was grown aerobically in the shake flasks at 25^{o}C and 250 rpm. The average and standard deviation are based on at least 3 replicates.

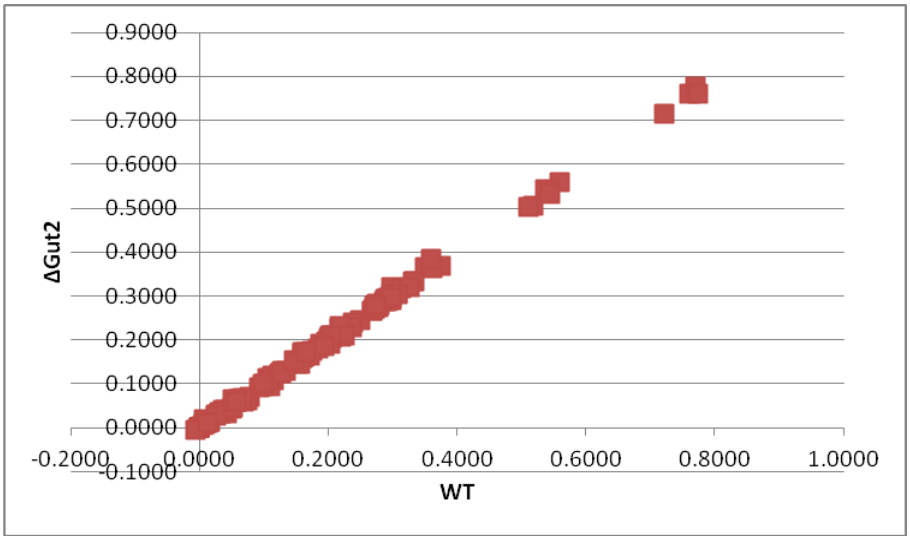


Figure 5. Isotopomer fraction of *Y. lipolytica* wild type and *gut2*Δ mutant

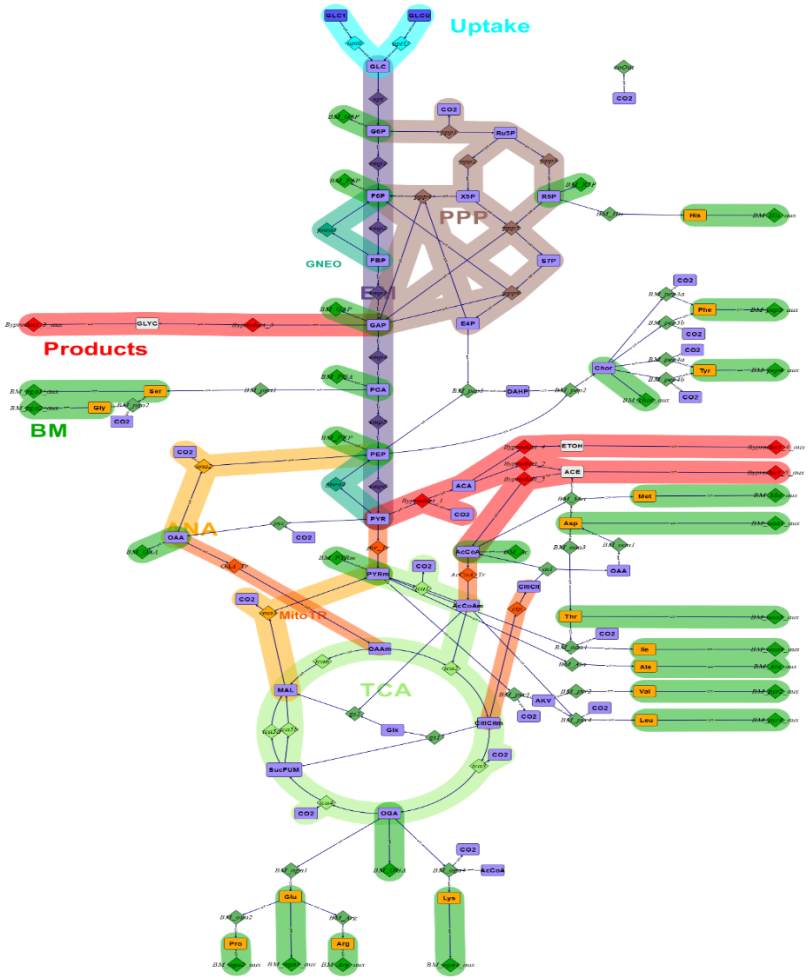


Figure 6. Metabolic pathway model of *Y. lipolytica*. Model was constructed in *Omix* for use in 13C MFA carbon tracing simulation. Metabolites are shown as rectangular boxes, with blue representing carbon source, purple - intermediate metabolites, white - potential extracellular metabolite products, and orange - amino acids. Reactions are shown as diamond shapes, and are grouped into sub-pathways, with cyan representing substrate uptake, purple – glycolysis pathway, aqua green – gluconeogenesis, brown – pentose phosphate pathway, light orange – anaplerotic pathway, red orange – transport to mitochondria, light green – TCA cycle pathway, green – reaction going into biomass, and red – exchange with media.

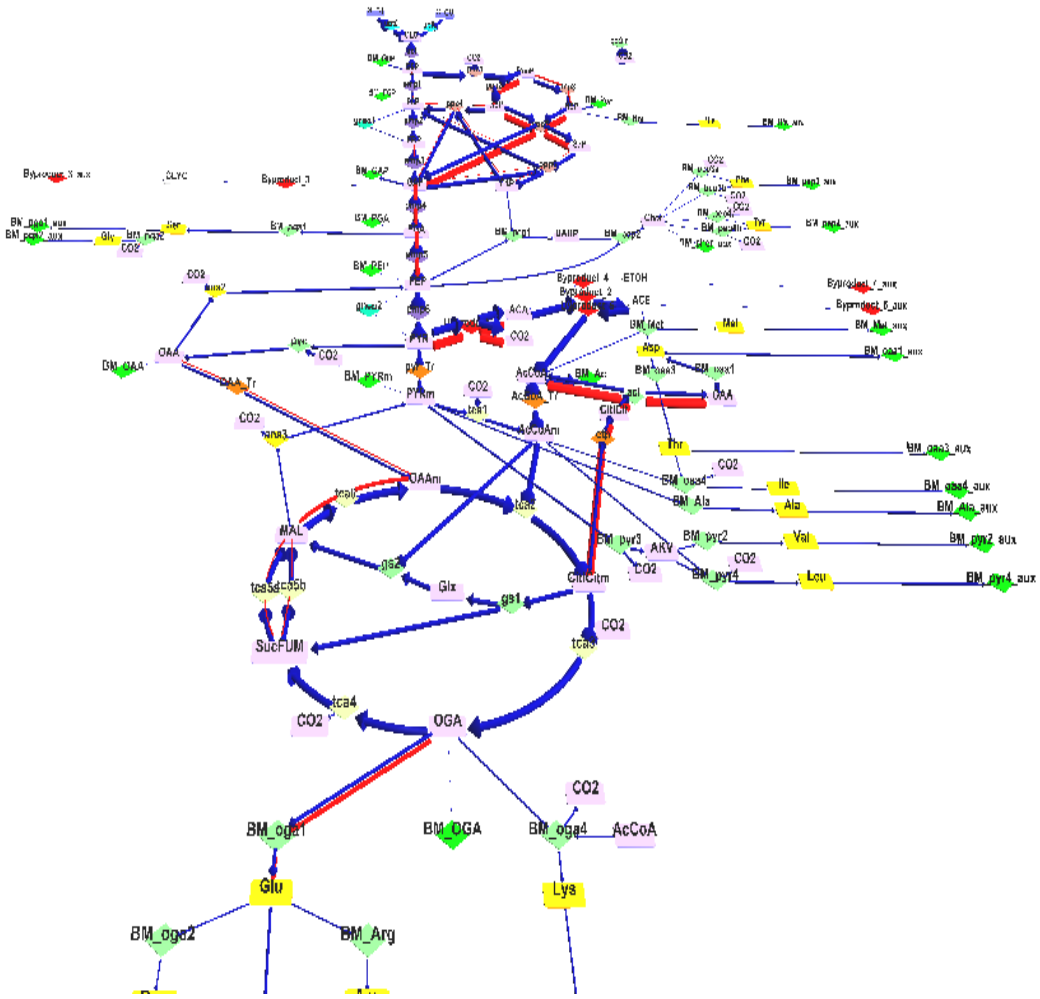


Figure 7. Fluxomics of wild type *Y. lipolytica* (displayed in 3D via *Omix*). Fluxomics representing the optimal simulation of WT *Y. lipolytica* metabolic fluxes as constrained by the mass fraction labeling patterns derived from the 13C carbon tracing. Blue arrow edges representing forward fluxes, while red arrow edges representing reverse fluxes. The size of the edges are displayed in log scale.

CHAPTER 8

CONCLUSIONS AND FUTURE PERSPECTIVES

Fatty acids of varying chain lengths (C_6 – C_{16}) naturally synthesized in many organisms are promising starting points for the catalytic production of industrial chemicals and diesel-like biofuels. However, bio-production of fatty acids from plants and other microbial production hosts relies heavily on manipulating tightly regulated fatty acid biosynthetic pathways. In addition, precursors for fatty acids are used along in other central metabolic pathways for the production of amino acids and biomass, which further complicates the engineering of microbial hosts for higher yields.

With the advent of systems metabolic engineering, we demonstrated an iterative metabolic engineering effort that integrates computationally driven predictions and metabolic flux analysis techniques to meet this challenge. Metabolic flux analysis (MFA), the quantification of fluxes in metabolic pathways, is an integral tool for the development of strategies for genetic modification and the identification of metabolic regulation, by comparing fluxes under different environments. With wild type *E. coli* fluxomic, the OptForce procedure was used to suggest and prioritize genetic manipulations that overproduce fatty acids of different chain lengths from C_6 to C_{16}. We identified some common but mostly chain-specific genetic interventions alluding to the possibility of fine-tuning overproduction of specific fatty acid chain lengths. In accordance with the OptForce prioritization of interventions, *fabZ* and acyl-ACP thioesterase were upregulated and *fadD* was deleted to arrive at a strain that produces 1.70 g/L and 0.14 g fatty acid/g glucose (~ 39% maximum theoretical yield) of $C_{14\text{-}16}$ fatty acid in minimal medium. These results highlight the benefit of using computational strain design and flux analysis tools to systematically guide the strain design to produce free fatty acids.

However, OptForce simulation does not infer gene regulation and enzyme inhibition. We extended the above study to close the metabolic engineering cycle by fine-tuning the phenotype and flux distribution of the first generation engineered strain. Along with transcriptomics and metabolomics analysis, we performed second round of OptForce simulation using the redefined flux distribution as constraints to predict genetic manipulations for second generation fatty acid-

overproducing strain. Metabolic flux analysis identified the up-regulation of the TCA cycle and down-regulation of pentose phosphate pathway to replenish the need of energy and reducing molecules for biosynthesis or fatty acids. The elevation of metabolite level in the TCA cycle complemented the flux findings. Using the whole genome transcriptional analysis, we observed acid stress response, membrane disruption, colanic acid and biofilm formation during fatty acid production, thus pinpointing the targets for future metabolic engineering effort. In accordance to OptForce suggestion, succinate dehydrogenase was deleted to improve the fatty acid titer and yield further 7-10% in M9 minimal media. Our study reinforces the advantages of integrating computational, experimental and omics tools for the design and engineering of microbial strains to overproduce value-added chemicals.

Product toxicity remains a tangible challenge that needs to be overcome in order to enhance the strain robustness. *Saccharomyces cerevisiae* can grow well at low pH with simple nutrient requirement. The robustness and extensive knowledge of genetics and physiology make *S. cerevisiae* a potential biocatalyst for fatty acid production. We elucidated the metabolic flux differences in central metabolism between the control and octanoic acid inhibition by conducting ^{13}C labeling experiments in fermentors. Distinctions in central metabolism flux distribution between control and octanoic acid stress were observed, especially in the TCA cycle. Interestingly, transcriptome analysis revealed membrane stress and intracellular acidification based on the perturbation of membrane protein genes and the up-regulation of plasma membrane ATP-required transporter. Coupled with the elevation of active efflux transporter, yeast cells under octanoic acid stress tend to generate more energy and reducing power to counteract the effect of membrane stress and acidification through the plasma membrane transporter activity. Further efforts can focus on improving membrane integrity to enhance the tolerance against octanoic acid.

Nonetheless, *S. cerevisiae* can only resist to less than 1mM fatty acids, thus not feasible to be the industrial host for fatty acid production. Oleaginous yeast (*Yarrowia lipolytica)*, which stores energy as lipids, may have potential to be host organism for fatty acids production. However, the metabolism of *Y. lipolytica* is less known compared to the conventional yeast. We observed higher fatty acid production under nitrogen starvation. The morphology study depicted filamentous cells formation might correlate with lipid accumulation. Glycerol-3-phosphate dehydrogenase (*gut2*), which oxidizes glycerol 3-phosphate to dihydroxyacetone, was deleted to

increase fatty acid production. MFA elucidated no significant distinction in the central metabolism flux distribution between the wild type and *gut2*Δ mutant. Interestingly, the TCA cycle and pentose phosphate pathway were the major pathways for generations of energy and reducing equivalents for growth requirement. Fluxomic data will be integrated with RNASeq transcriptomic data of *Y. lipolytica* into the hybrid kinetics-flux balance analysis (FBA) framework. The hybrid model, which includes dynamics regulation of gene and enzymes, will be incorporated into *in-silico* computational tools to guide the strain design.

Future perspectives

In order to produce commercial fatty acid via biological processes, we would need to achieve higher product yield and productivity, while increasing the product tolerance in the organisms. These objectives could be attained via systems metabolic engineering efforts and fermentation process development strategies.

Systems metabolic engineering aspect

Metabolic flux analysis (MFA) has been demonstrated to be a powerful tool to diagnose cell physiology during exponential growth phase. However, MFA requires the attainment of isotopic and metabolic steady states. This limits the application of MFA to have a snapshot of cell physiology during stationary or late stationary stage. The synthesis of desired products in engineered organism mostly peaks during stationary phase after sufficient cell growth and protein expression during mid-logarithm growth. Moreover, it remains a challenge to apply MFA during fed-batch fermentation which is broadly used in industry. To overcome these challenges, nonstationary ^{13}C-MFA uses short-time isotopic transient data to resolve fluxes within much shorter labeling experiments (Nöh et al., 2006). Its application, however, raises new challenges in computation and analytics aspects. For organism with slow growth rate, it will take prolonged duration to reach isotopic and metabolic steady states, leading to impractical long experiment. ^{13}C tracers will be incorporated into the intracellular metabolites and free amino acids in the range of seconds to minutes. With the fast turnover, ^{13}C incorporated metabolites would wash out the existing pool of metabolites. Hence, MFA, which is based on intracellular metabolite isotopomer, can dramatically shorten the experimental time.

Systems biology, which consists of genomic, transcriptomic, proteomic, metabolomic and fluxomic, could elucidate the underlying cellular network and reveal interactions and

regulations between hierarchy structure of information (Lee et al., 2012). However, there is a lack of integration of these omics tools to study the biological problem. Each method has its advantages and limitations under certain circumstances. Integration of omics tools, therefore, can pinpoint the bottleneck of the biological system collectively to increase product yield and tolerance against product inhibition. Solving biological problems from a multiple integrated perspectives would be more effective and reliable.

In-silico genome-scale metabolic simulations had been proven for its robustness in guidance of strain design. It accounts for all redox potentials balance and cofactor requirements. However, most of the computational tools do not include regulation network of gene and enzyme. Thus, *in-silico* based simulations could include omics inputs and gene/enzyme regulatory network to define the dynamics range of enzymatic reactions in the metabolic networks. The incorporation of these omics constraints will improve the reliability of the simulated genetic interventions, thereby is an effective strategy for engineering metabolism at the systems level for the production of desired chemicals.

Fatty acid toxicity was proven to hinder the cell viability and the productivity of fatty acid synthesis. Omics analysis revealed the possible culprits are the loss of membrane integrity and possible intracellular acidification (Lennen et al., 2011; Liu et al., 2013). Further metabolic engineering efforts to enhance tolerance could be focused on engineering transporter mechanism to pump out the toxic products, or strengthening membrane cell wall to defend the membrane stress. Synthetic biology tools could be applied to facilitate the strain construction processes.

Fermentation process engineering aspect

The advances in systems metabolic engineering, which comprises of classical metabolic engineering, system biology, synthetic biology and directed revolution, could root in success in strain design with shorter turnover time. Besides that, a sustainable bacterial fermentation route to produce carboxylic acids is required for commercialization. Most of the microbiology labs still opt to shake flasks and culture tubes to evaluate the strain performance preliminarily. Fermentation under well-controlled bioreactors is required to further characterize the strain and enhance the strain robustness. Fermentation in bioreactors aims to increase cell growth and maximize the conversion of substrate to desired products. In order to commercialize a biological process, the scale-up, separation and recovery processes play important roles to minimize the

overall production cost. Therefore, fermentation process development is crucial to optimize the strain performance in term of pH, temperature, dissolved oxygen, agitation rate and substrate concentration. In this context, bioreactor cultivation could be conducted in batch, fed-batch and continuous mode under a wide range of environmental conditions.

The fed-batch fermentation is the most widely used fermentation mode to achieve high cell density, high product titer/yield. The feed profile and oxygen uptake rate are the most intriguing parameter to be controlled in designing fed-batch fermentation. Therefore, the first generation fatty acid-overproducing strain (ML103 pXZ18Z) (Ranganathan et al., 2012) was grown in M9 minimal medium using batch and fed-batch mode in the bioreactors. Figure 1 and 2 illustrate the cell density, glucose, acetate and fatty acid profile throughout the fermentation period. During the batch fermentation, we observed acetate production till the depletion of glucose. When glucose was depleted, acetate was assimilated back into the metabolism to form fatty acids. On the other hand, the fed-batch fermentation shared the same characteristics till glucose was fed into the system. Unlike the batch fermentation, acetate production slowed down when glucose concentration was low, and continued to increase after glucose was fed. We observed the glucose consumption and fatty acid production were slower after the addition of glucose. Given the decrease in cell density, endogenous fatty acid (4.8g/L) might cause toxicity effect by affecting the cell growth, glucose consumption and fatty acid production. Fed-batch fermentation produced higher titers of fatty acids and acetate by-product. Interestingly, the fatty acid yield of fed-batch culture (0.18 g/g) was significant higher than the batch culture (0.14 g/g). Manipulation of glucose level could possibly increase the productivity of fatty acid. The fed-batch fermentation could be extended to constant, linear and exponential feeding strategies. Feedback control based on dissolved oxygen and pH could be applied to control the glucose feeding and avoid overfed scenario. Process development of fed-batch fermentation will help increase the productivity and the yield of desired products, thus aiding the technological transition to the scale-up in the industrial fermentation.

Separation and recovery remain one of the major costs for the downstream process of fermentation products. Fatty acids could not be vaporized through flash tank or extracted through gas-stripping. Distillation and solvent extraction appear to be potential strategies to recover fatty acids from the fermentation broths. However, these extraction methods may involve harsh chemical treatment and require excessive energy input. Based on the density and solubility of

fatty acids, we investigated the feasibility of centrifugation and the temperature effects on the separation efficiency. Figure 3 depicts the separations of fatty acids (white layers above the supernatants) after centrifugation for 30 minutes. Low temperature (4^{o}C) was believed to further assist fatty acid separations. Further engineering efforts for fermentation process development and downstream product recovery, are required to reduce the operating cost.

References

Lee, J. W., D. Na, J. M. Park, J. Lee, S. Choi, and S. Y. Lee, 2012, Systems metabolic engineering of microorganisms for natural and non-natural chemicals: Nature Chemical Biology, v. 8.

Lennen, R. M., M. A. Kruziki, K. Kumar, R. A. Zinkel, K. E. Burnum, M. S. Lipton, S. W. Hoover, D. R. Ranatunga, T. M. Wittkopp, W. D. Marner, II, and B. F. Pfleger, 2011, Membrane Stresses Induced by Overproduction of Free Fatty Acids in Escherichia coli: Applied and Environmental Microbiology, v. 77.

Liu, P., A. Chernyshov, T. Najdi, Y. Fu, J. Dickerson, S. Sandmeyer, and L. Jarboe, 2013, Membrane stress caused by octanoic acid in Saccharomyces cerevisiae: Applied Microbiology and Biotechnology, v. 97, p. 3239-3251.

Nöh, K., A. Wahl, and W. Wiechert, 2006, Computational tools for isotopically instationary 13C labeling experiments under metabolic steady state conditions: Metabolic Engineering, v. 8, p. 554-577.

Ranganathan, S., T. W. Tee, A. Chowdhury, A. R. Zomorrodi, J. M. Yoon, Y. Fu, J. V. Shanks, and C. D. Maranas, 2012, An integrated computational and experimental study for overproducing fatty acids in Escherichia coli: Metabolic Engineering, v. 14, p. 687-704.

List of figures

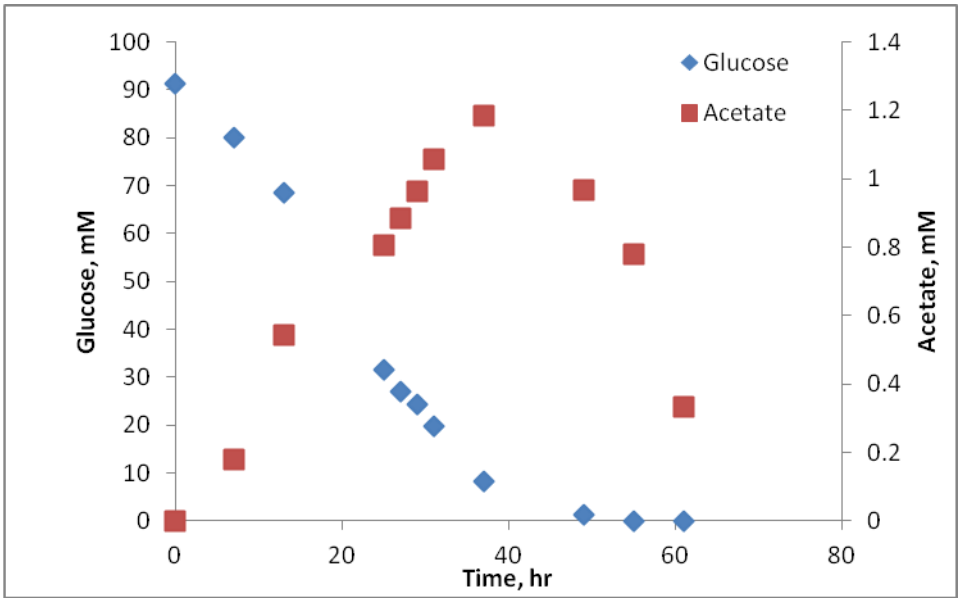


Figure 1a. Glucose and acetate profile of *E. coli* ML103 pXZ18Z under batch fermentation in bioreactor in M9 minimal media with 1.5% glucose.

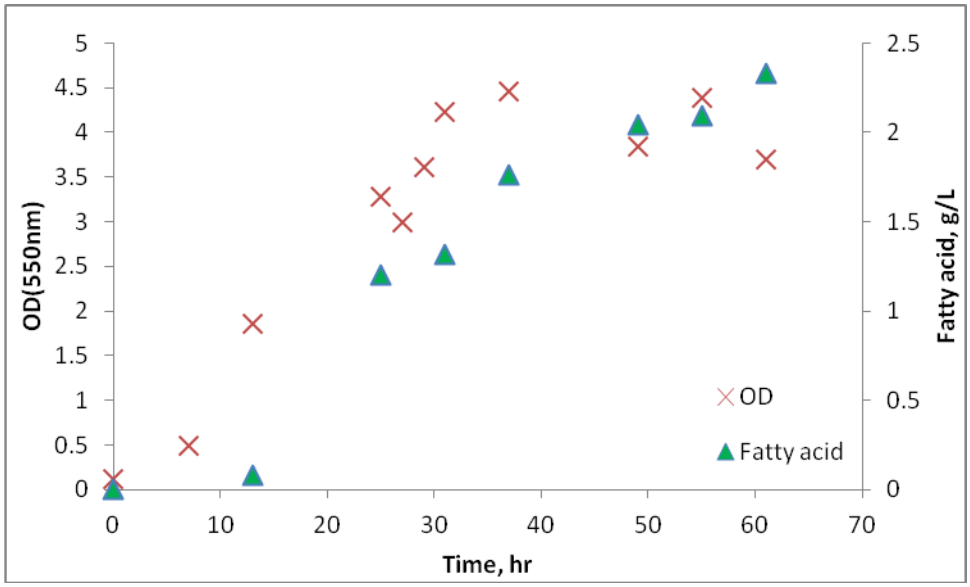


Figure 1b. Cell density and fatty acid production of *E. coli* ML103 pXZ18Z under batch fermentation in bioreactor in M9 minimal media with 1.5% glucose.

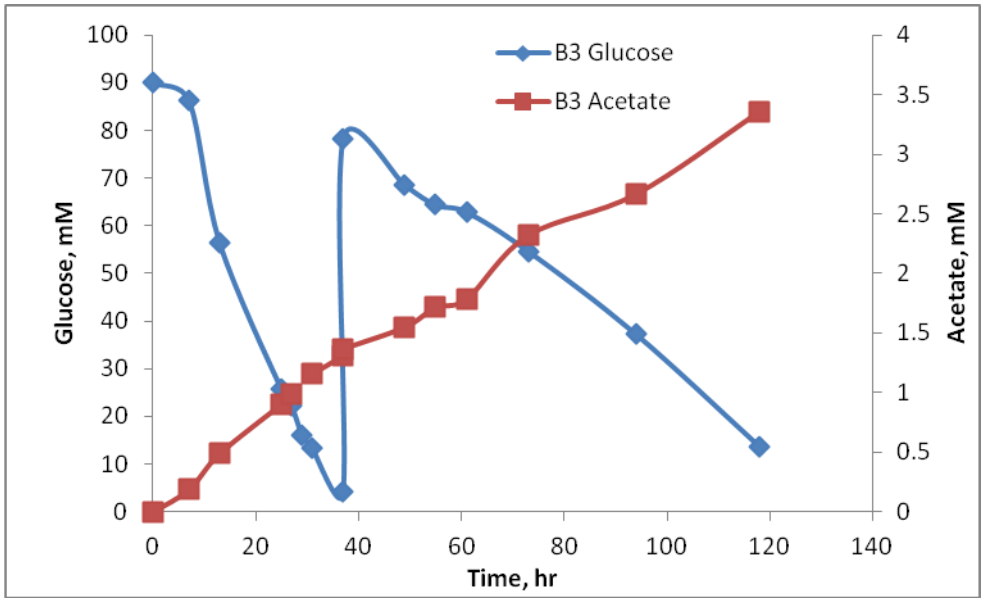


Figure 2a. Glucose and acetate profile of *E. coli* ML103 pXZ18Z under fed-batch fermentation in bioreactor in M9 minimal media with 1.5% glucose. Glucose was pulsed into the system at 37 hour.

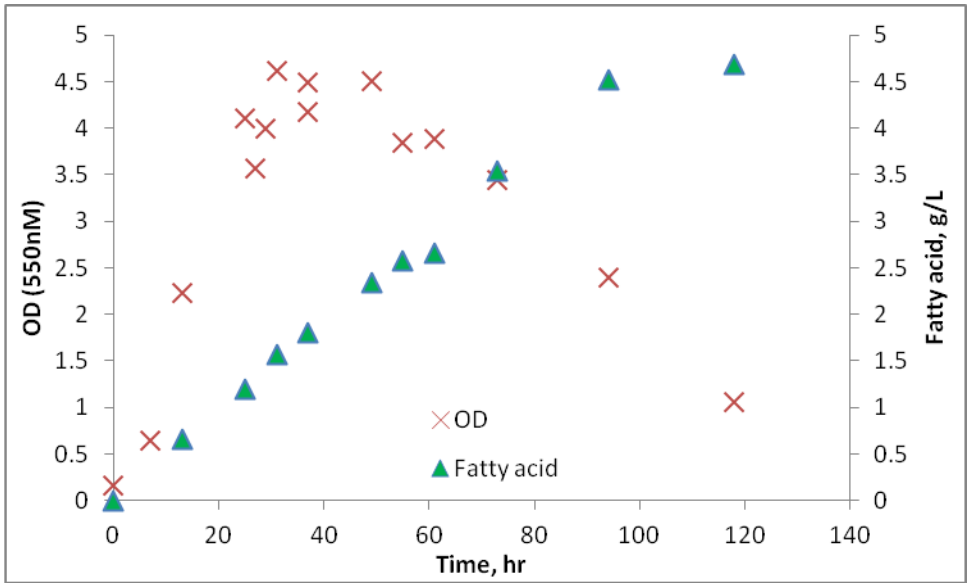


Figure 2b. Cell density and fatty acid production of *E. coli* ML103 pXZ18Z under fed-batch fermentation in bioreactor in M9 minimal media with 1.5% glucose. . Glucose was pulsed into the system at 37 hour.

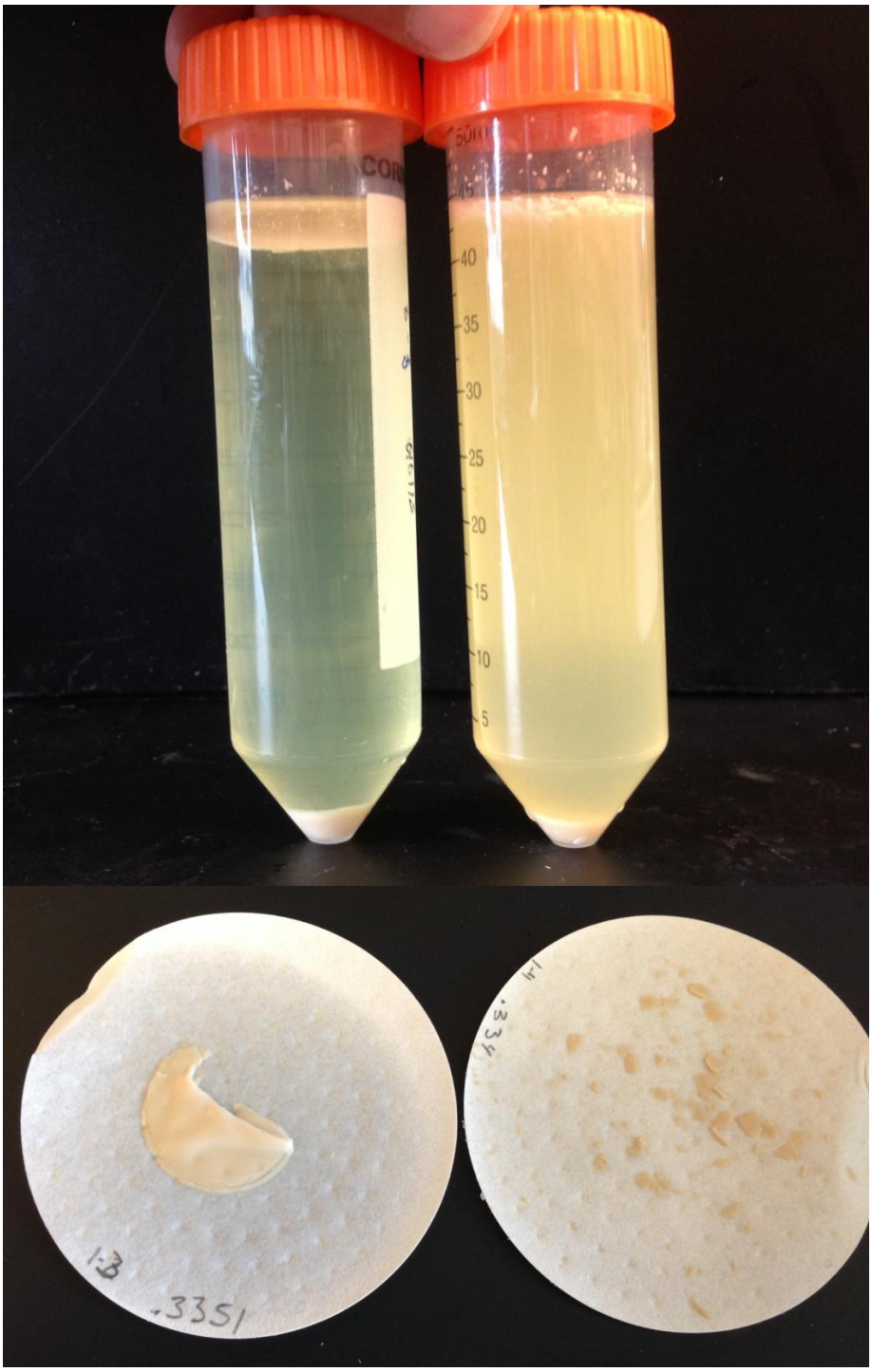


Figure 2. The fatty acid fermentation broth after centrifugation for 30 minutes at 4^{o}C (left) and room temperature (right). The broth was vacuum-filtered and fatty acid solid layers remained on the filter papers.