Summer 8-15-2017

# Designing Metabolite Biosensors and Engineering Genetic Circuits to Regulate Metabolic Pathways

Di Liu
*Washington University in St. Louis*

WASHINGTON UNIVERSITY IN ST. LOUIS

School of Engineering & Applied Science
Department of Energy, Environmental and Chemical Engineering

Dissertation Examination Committee:
Fuzhong Zhang, Chair
Gautam Dantas
Bradley Evans
Tae Seok Moon
Himadri Pakrasi
Yinjie Tang

# Designing Metabolite Biosensors and Engineering Genetic Circuits to Regulate Metabolic Pathways

by

Di Liu

A dissertation presented to
The Graduate School
of Washington University in
partial fulfillment of the
requirements for the degree
of Doctor of Philosophy

August 2017
St. Louis, Missouri

# Table of Contents

# List of Figures

# List of Tables

# Acknowledgments

I would like to express my great appreciation to my advisor, Dr. Fuzhong Zhang, for his support and guidance throughout my research. He led me into the fantastic world of synthetic biology and taught me the artistry of critical thinking and rigorous reasoning. I share many memories with him: setting up the lab, running my very first molecular biology experiments, writing my first paper, practicing for my first conference talk, and, finally, writing this dissertation. I thank him for being critical, responsive and inspiring throughout my PhD studies. He has always been a great mentor, one who has largely influenced my attitudes towards life.

I would also like to thank Dr. Yinjie Tang for his support and encouragement during my PhD. I thank him for his trust, the collaboration opportunities with his group, and inviting me to the fun group activities. I am also grateful to my collaborators, Dr. Diego A. Oyarzún and Dr. Ahmad Mannan at Imperial College London. It was a great pleasure to work with them, and I learned a lot during the collaboration. I would also like to express my gratitude to Drs. Gautam Dantas, Bradley Evans, Tae Seok Moon, Himadri Pakrasi, and Yinjie Tang for kindly serving on my dissertation committee, giving their valuable time and precious suggestions in helping to improve my research.

I would also like to thank my colleagues at Washington University. In particular, I would like to thank Drs. Yi Xiao, Yanfang Jiang, Yi Yu, Arul Varman, and Lian He, who generously shared their

knowledge and experiences with me. I would also like to acknowledge my talented colleagues in the laboratory for the exchange of ideas and building an organized and fun environment in the lab. I also thank all the undergraduate students and rotation students who worked with me during my PhD for their help and inputs.

I thank all my friends for sharing joys and sorrows with me and cheering me up when I was down. Life was so much more colorful and enjoyable with their company. I also thank all the staff members in the department for their help. In addition, I thank the National Science Foundation, Human Frontiers Science Program, International Center for Advanced Renewable Energy and Sustainability, and Washington University for funding my Ph.D. projects.

Last but not least, I give my special thanks to my parents. Their unconditional understanding, unreserved support, and endless love have always been a source of power for me to take on new adventures.

Di Liu

*Washington University in St. Louis*

*August 2017*

ABSTRACT OF THE DISSERTATION

Designing Metabolite Biosensors and Engineering Genetic Circuits to

Regulate Metabolic Pathways

by

Di Liu

Doctor of Philosophy in Energy, Environmental and Chemical Engineering

Washington University in St. Louis, 2017

Professor Fuzhong Zhang, Chair

Microbial production of chemicals has provided an attractive alternative to chemical synthesis. A key to make this technology economically viable is to improve titers, productivities, and strain robustness. However, pathway productivities and yields are often limited by metabolic imbalances that inhibit cell growth and chemical production. In contrast, natural metabolic pathways are dynamically regulated according to cellular metabolic status. Dynamic regulation allows cells to adjust metabolite concentrations to optimal levels and avoid wasting carbon and energy. Inspired by nature, synthetic regulatory circuits have shown great promise in improving titers and productivities, because they can balance the metabolism by dynamically adjusting enzyme expression levels according to cellular metabolic status.

To engineer synthetic regulatory circuits to improve production, we must design and tune metabolite biosensors, and also understand the metabolic dynamics to identify the optimal regulatory architecture. The research presented here addresses both these key aspects and demonstrates an application of genetic circuits to improve pathway production. Specifically, we develop theories that predict and experimentally validate a coupling between dynamic range and response threshold in transcription factor-based biosensors, and provide design guidelines to orthogonally control the biosensor output and its sensitivity. Next, we develop a malonyl-CoA sensor-actuator and demonstrate its application to engineering a negative feedback circuit to improve fatty acid production. Finally, genetic circuits with various architectures are constructed to study metabolic dynamics, which reveal that negative feedback circuits can dramatically accelerate metabolic dynamics.

The findings of this dissertation provide rational design principles for transcription factor-based metabolite biosensors and a systematic understanding of metabolic dynamics under various regulation architectures. They provide valuable tools and knowledge to engineer metabolic circuits to regulate various metabolic pathways, increasing titers, productivities, and yields.

# Chapter 1 Introduction

Note: This chapter contains text and figures from the published paper (Liu, Di, et al. "Applications and advances of metabolite biosensors for metabolic engineering." *Metabolic engineering* 31 (2015): 35-43.) and the published book chapter (Liu, Di, et al. "Design of Dynamic Pathways. " *Biotechnology for Biofuel Production and Optimization*. Elsevier, 2016).

## 1.1 Background

### 1.1.1 Importance of metabolite biosensors

Advances in metabolic engineering have enabled microbial production of a wide variety of valuable compounds, providing alternative synthesis routes for chemicals including biofuels, pharmaceuticals, nutraceuticals, bulk chemicals, and materials. To produce these valuable compounds, efficient biosynthetic pathways must be constructed in appropriate hosts, which often requires extensive optimization to reach economically viable titers, yields, and productivities. The cycle of repeatedly tuning pathway parameters and evaluating production is laborious and time-consuming. Synthetic biology is a fast-growing field that develops new tools for biological engineering, fulfilling the need for efficient pathway optimization. It has proven effective in increasing process predictability and throughput as well as in creating new strategies to optimize biosynthetic pathways. Among these new tools, biosensors represent a significant contribution from synthetic biology and are widely used in metabolic engineering.

Biosensors are ubiquitous in nature and have evolved to detect both environmental signals (e.g.

temperature, pH, oxygen) and intra- and extracellular metabolites. These sensed signals are coupled with actuator outputs to modify the transcription, translation, and protein activities of cells. A crucial consideration in synthetic biology and metabolic engineering is that these natural biosensing machineries have evolved to maximize evolutionary fitness rather than biosynthetic production. However, detailed knowledge of the sensing mechanisms and endogenous functions of these biosensors serve as a valuable starting point to co-opt them for biosynthetic goals.

Here metabolite biosensors are referred to as genetically-encoded protein or RNA-based sensors that interact with a metabolite to generate an actuator output. The output domain of a metabolite biosensor generates detectable phenotypes through modulating transcription rates, translation rates, or post-translational parameters to control protein expression or activity. Over the past few decades, metabolite biosensors have drawn tremendous attention and have several applications in metabolic engineering (Figure 1). First, biosensors can be coupled to readable outputs such as fluorescence to semi-quantitatively report the concentration of a target compound. This approach is frequently used for high-throughput screening of high-producing strains and features distinct advantages over conventional methods such as gas chromatography (GC) and high-performance liquid chromatography (HPLC): 1) Biosensor-mediated quantification avoids time-consuming sample preparation and has much higher throughput than conventional chromatographic techniques; 2) metabolite biosensors are more suitable for detecting labile and low abundant metabolites such as acyl-phosphate, acyl-diphosphate, aldehyde, and acyl-CoAs, which are difficult to measure

accurately by conventional methods; 3) metabolite biosensors allow real-time monitoring of metabolite dynamics in living cells, which is impossible to study using chromatographic methods. These reporter outputs may also help coordinate complementary manipulations of the culture environment itself (mixing, nutrient addition, timing of harvest) to further improve production (Polizzi and Kontoravdi, 2015). Second, biosensors can be engineered to couple the sensing of a desirable product or intermediate metabolite with a fitness advantage for the cell by expressing a gene necessary for survival under selective conditions (Dietrich et al., 2013; Raman et al., 2014). The difference in cell growth allows direct enrichment of fast-growing cells from mutant libraries, which allows an easy selection for desirable production characteristics. Third, metabolite biosensors can also be used to control metabolic flux dynamically (Dahl et al., 2013; Liu et al., 2013; Zhang et al., 2012). The actuator can be designed to tune pathway enzyme expression or post-translational parameters in response to the level of the relevant metabolite, allowing for dynamic control of pathway activity based on the cellular metabolic state. As a result, the pathway is dynamically balanced, which not only reduces toxic intermediate accumulation but also saves carbon and energy that is otherwise diverted to synthesize unnecessary proteins or intermediates. Overall, the emerging tools to engineer biosensors and their applications towards metabolic engineering have greatly advanced microbial production of a variety of chemicals.

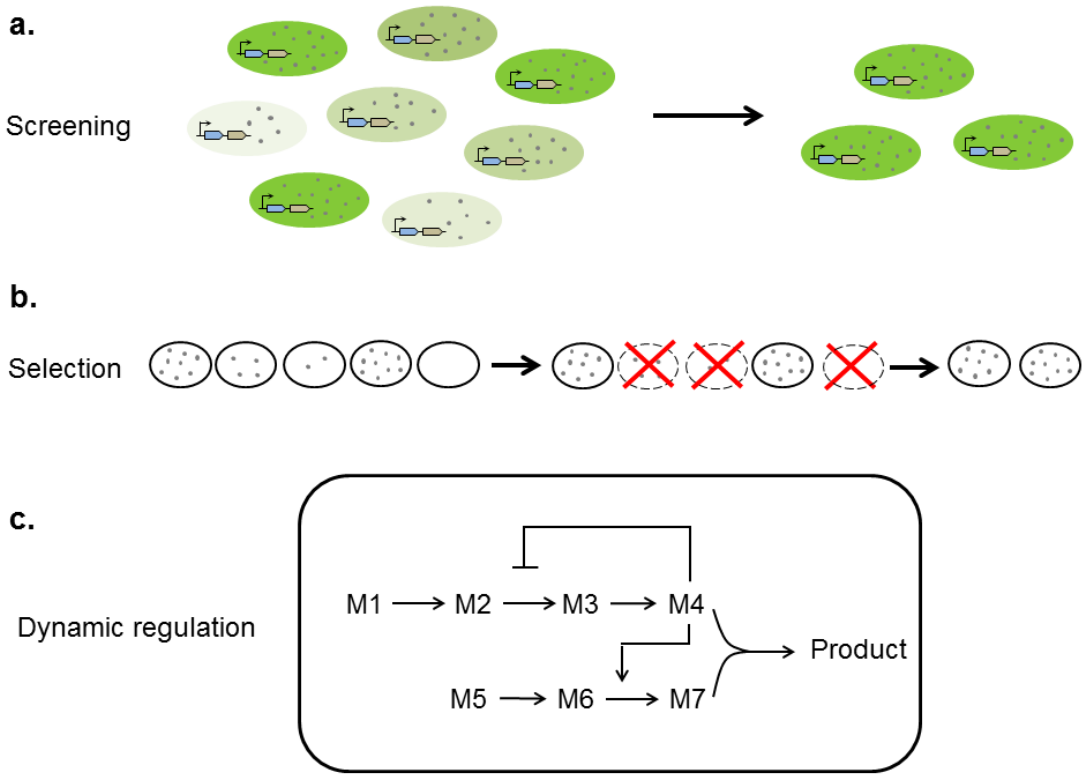



**Figure 1.1**. Applications of metabolite biosensors in metabolic engineering. a) Biosensors can be linked to a colorimetric output to report the concentration of metabolites, providing an easy way to screen for high-producing strains. b) Biosensors can be used to control an output associated with a fitness advantage under selective conditions. This allows direct enrichment and selection of high-producers. c) Biosensors can be used to control the activity of a metabolic pathway, which allows dynamic optimization of the pathway activity according to the level of the sensed metabolite.

### 1.1.2 Development of transcription-factor based metabolite biosensors and their applications in metabolic engineering

In nature, transcription factors regulate gene expression by specific binding to the chromosomal

DNA, blocking or promoting transcription by RNA polymerase. Among these, some transcription factors can be activated or deactivated by small molecules through ligand binding, phosphorylation, or interaction with other regulatory elements. Here we will focus on transcription factors that respond to metabolites.

Metabolite-responsive transcription factors (MRTFs) have been evolved to interact with various metabolites. *Escherichia coli*, for example, has more than 230 transcription factors (Binder et al., 2012), which sense a wide variety of metabolites, including sugars, sugar phosphates, amino acids, and lipids. Natural MRTFs haven been extensively explored to engineer biosensors for metabolic engineering applications. Typically, metabolite-responsive promoters with tunable output dynamic ranges can be engineered by inserting the cognate operator of a MRTF into a synthetic promoter to regulate genes of interest (Figure 2a, 2b). Using this strategy, biosensors that respond to a variety of metabolites have been created, including sensors for butanol (Dietrich et al., 2013), alkanes (Reed et al., 2012), malonyl-CoA (Liu et al., 2013; Xu et al., 2014), acyl-CoA (Zhang et al., 2012) and aromatic aldehyde (Fiorentino et al., 2009). The primary use of MRTF sensors is to screen for high-producing strains from a library of natural or engineered strains, as demonstrated in the production of several chemicals, including mevalonate (Tang and Cirino, 2011), L-lysine (Binder et al., 2012), and triacetic acid lactone (Tang et al., 2013). This approach becomes particularly powerful when coupled with fluorescence-activated cell sorting (FACS). In one example, an *eyfp* was cloned 3' of a *Corynebacterium glutamicum* promoter that is regulated by an endogenous

transcription factor Lrp, which can detect L-methionine and several branched-chain amino acids, including L-valine, L-leucine and L-isoleucine (Mustafi et al., 2012). Using chemical mutagens, random mutations were introduced to the *Corynebacterium glutamicum* strains, which carry the sensor plasmid. Cells were cultivated and screened by FACS, and the ones with enhanced fluorescence were isolated and re-cultivated to enrich the high-producing strains. Mutants that produce up to a total of 11 mM branched-chain amino acids were identified. In addition, MRTFs have also been used to control genes associated with cell growth/survival for selection (Dietrich et al., 2013; Raman et al., 2014). In a recent paper, Raman *et al* used MRTF-regulated promoters to control the expression of TolC, a protein that allows both positive and negative selections when supplemented with sodium dodecyl-sulfate (SDS) and colicin E1, respectively. While positive selections were needed to select for high-producing strains generated by multiplex automated genome engineering (MAGE), negative selections were used to eliminate the false positives caused by mutations that deactivate the sensor-selection system. By alternating between negative and positive cycles, the authors demonstrated enhanced production for both naringenin and glucaric acid (Raman et al., 2014). Overall, biosensor-mediated high-throughput screening and selection methods drastically shorten the time required to analyze mutant cells, enhancing the power of evolutionary approaches. In addition, MRTF-based biosensors have also been used to dynamically regulate metabolic flux. One of the earliest examples of using the MRTF sensor for dynamic regulation involves a fatty acyl-CoA biosensor FadR (Zhang et al., 2012). FadR naturally regulates several genes in *E. coli* fatty acid biosynthesis and degradation. In a fatty acid ethyl ester (FAEE)-

producing strain, the cognate actuator of FadR was engineered to upregulate acyl-CoA biosynthesis, ethanol production, and the expression of a wax ester synthase, which condenses ethanol and acyl-CoAs to FAEEs (Figure 2c). This allows the downstream pathway to be activated only when there is sufficient acyl-CoA, preventing the production of unnecessary proteins and ethanol at the early stage of fermentation. As a result, the FAEE titer was increased 3-fold, reaching 1.5 g/L. In addition, other studies constructed negative feedback circuits to dynamically regulate fatty acid biosynthesis using a malonyl-CoA–responsive MRTF (Liu et al., 2013; Xu et al., 2014). Malonyl-CoA is synthesized from acetyl-CoA by acetyl-CoA carboxylase (encoded by *acc*). Expression of *acc* improves fatty acid production, but *acc* overexpression also inhibits cell growth. To alleviate the toxicity from *acc* overexpression while maintaining high malonyl-CoA concentrations, malonyl-CoA biosensors were used to dynamically down-regulate *acc* expression when cells accumulate high malonyl-CoA levels. These sensor-enabled dynamic regulations enhanced fatty acid production dramatically, particularly when coupled with upregulation of fatty acid chain elongation genes.

So far, most MRTF-based biosensors have relied on naturally existing transcription factors, and these sensors are usually highly specific to their corresponding metabolites. To allow broader utilization, the specificity of an MRTF needs to be altered to detect a metabolite of interest to which no natural sensor exists. The specificity of a MRTF can be altered using various protein engineering approaches (rational design, evolution, etc.). For example, a biosensor for mevalonate

(a precursor of isoprenoid biosynthesis) was developed based on AraC, which naturally recognizes arabinose (Tang and Cirino, 2011). In this work, the specificity of AraC was modified by subjecting five residues of its metabolite-binding pocket to saturation mutagenesis. To select for mevalonate-responsive AraC variants, a *gfp* was placed under the control of a $P_{BAD}$ promoter. FACS was then used to isolate clones with higher GFP expression in the presence of mevalonate. This general approach is useful to develop sensors for a variety of compounds, as demonstrated for triacetic acid lactone (Tang et al., 2013) and D-arabinose (Tang et al., 2008).

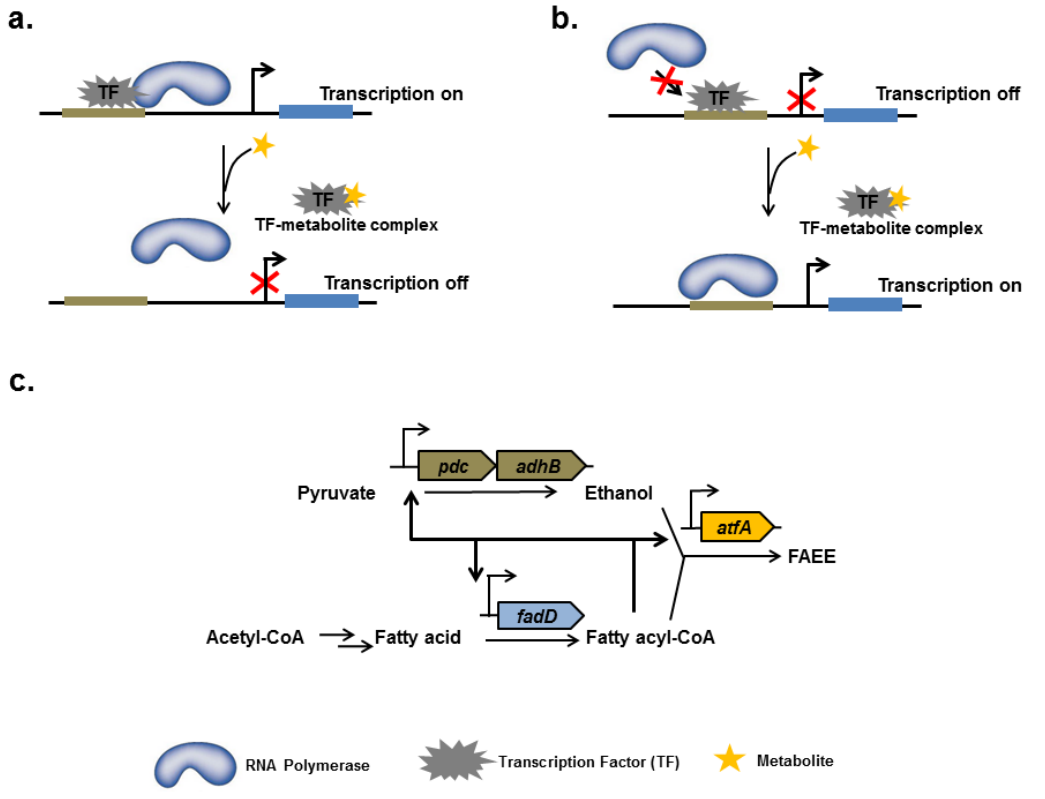



**Figure 1.2**. MRTF-based metabolite biosensors and examples of their applications in metabolic engineering. a) Metabolite binding to an MRTF represses transcription. b) Metabolite binding to

an MRTF enhances transcription. c) A dynamic pathway regulation system for FAEE production. Acyl-CoA biosensors were used to activate synthesis of FAEE and its precursors, ethanol and fatty acyl-CoA. This synthetic regulatory system dynamically turns on downstream enzyme expression only when the upstream pathway accumulates enough key intermediates, in this case acyl-CoA.

### 1.1.3 Importance of metabolic circuits and their benefits to natural and engineered metabolic pathways

In nature, organisms spanning from simple microbes to complex, multicellular animals rely on various levels of dynamic regulation to maintain metabolic homeostasis, basic cellular functions, and to adapt to ever-changing intracellular and extracellular environments. As a simple example, *Escherichia coli* utilizes a plethora of transcriptional regulatory networks: 34 coherent feedforward loops, 6 incoherent feedforward loops, 68 operons under single input module control, and 203 pairs of operons regulated by the same two transcription factors (Shen-Orr et al., 2002). These numbers, limited by current investigational techniques, are likely a vast underestimate of the actual abundance of transcription-level natural dynamic regulation in *E. coli.* The number of documented dynamic regulation suggests that this is an effective regulatory mechanism for maintaining cell homeostasis in response to naturally fluctuating environments. Many microorganisms other than *E. coli* use similar regulatory networks to regulate a variety of pathways. For example, seven out of eight genes in the *S. cerevisiae* leucine biosynthesis pathway are positively regulated by the transcription factor Leu3, which is in turn activated by a pathway intermediate – α-isopropyl-

malate (αIPM). As a result, leucine biosynthesis is positively controlled by the intermediate branching from the common precursor of branched-chain amino acid biosynthesis (Kohlhaw et al., 2003). Alternatively, over 100 diverse bacterial genomes contain a prototypical negative feedback loop in which the product, tryptophan, represses transcription of the *trp* operon (Xie et al., 2003).

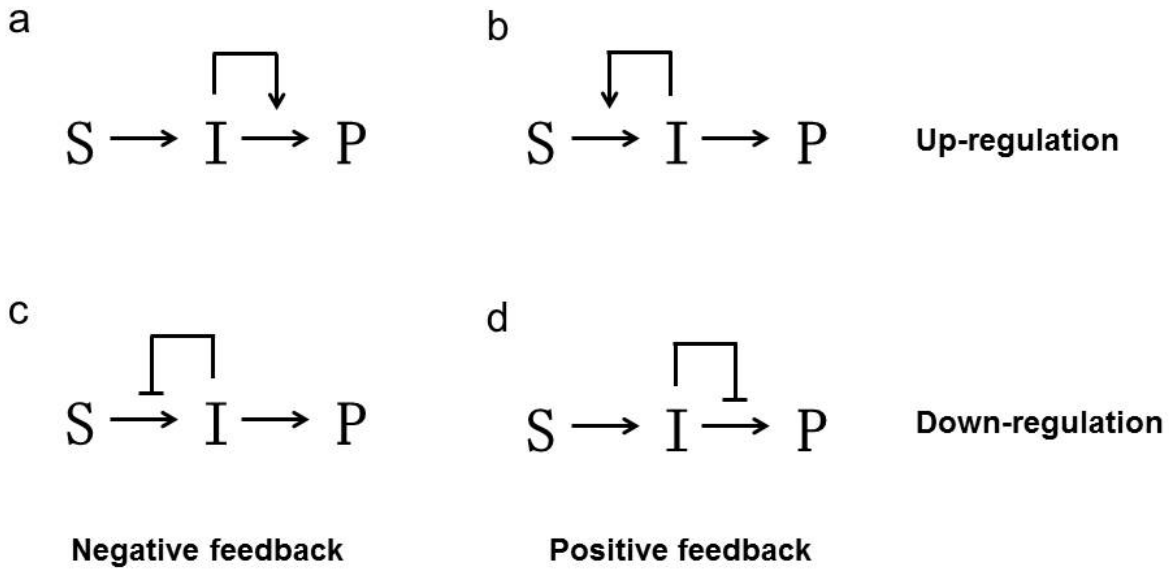



**Figure 1.3**. The intermediate-sensing regulation strategies for a simplified metabolic pathway. Substrate (S) is converted to product (P) via an intermediate (I), and the upstream or downstream gene of the intermediate can be either up-regulated or down-regulated. a & b show the up-regulation strategies that create either a negative feedback (a) or a positive feedback (b) circuit. c

& d show the down-regulation strategies that create either a negative feedback (c) or a positive feedback (d) circuit.

Figure 1.3 demonstrates some common naturally occurring regulatory topologies based on a simplified metabolic pathway, in which substrate (S) is converted to product (P) via the intermediate (I). The biosynthesis of I and the conversion of I to P can be either up-regulated (Fig. 1.3a & 1.3b) or down-regulated (Fig. 1.3c & 1.3d) by I. From the point of view of the intermediate metabolite, both up-regulation of a downstream pathway (Fig. 1.3a) and down-regulation of an upstream pathway (Fig. 1.3c) create negative feedback loop, since they both act to decrease the level of I. On the other hand, up-regulation of an upstream pathway (Fig. 1.3b) and down-regulation of a downstream pathway (Fig. 1.3d) lead to a positive feedback loop, since they both act to increase the level of I.

In practice, each type of regulatory topology has its specific benefits. Negative autoregulation, in which a transcription factor represses its own transcription, is a key regulatory mechanism used by almost half of all transcription factors in *E. coli.* In addition to limiting wasteful gene and transcription factor expression, negative autoregulation has been shown to increase the dynamic range of the input required to activate the target gene (Madar et al., 2011). This feature is desirable when building synthetic dynamic regulation systems, because it minimizes the need to tune the input dynamic range. Another important characteristic of negative autoregulation is that it can

decrease the response time of a genetic circuit (Rosenfeld et al., 2002). When a strong promoter is used, a fast initial rise in protein production is generated, and as the transcription factor level increases, it reaches the repression threshold for its own promoter, thus decreasing its own production rate. In this case, the negatively autoregulated transcription factor shuts off its transcription after the initial rise, quickly reaching a steady state (Rosenfeld et al., 2002). Furthermore, negative autoregulation can also reduce variations in protein levels across a cell population, another desired feature in an engineered dynamic regulatory system (Alon et al., 2007; Becskei et al., 2000).

Negative feedback loops, as shown in Figure 1.3a & 1.3c may also be of use in designing a dynamic circuit. In general, negative feedback can prevent cells from synthesizing unneeded proteins, thus conserving carbon and energy. Moreover, under some conditions, negative feedback loops may reduce noise in metabolic systems. This reduction in noise helps to avoid stochastic deviation from the expected metabolic state (Levine et al., 2007). In a metabolically engineered system, reduction of noise and improved stability of metabolic states helps achieve maximal product production (Oyzarun et al., 2015).

Another common regulation strategy in nature is a feedforward loop. Feedforward loops contain two transcription factors; the first transcription factor regulates the second transcription factor, and both the first and the second transcription factor regulate the same target gene. The activity of the

first transcription factor directly on the target gene is the direct loop of activation, while the activity of the second transcription on the target gene is the indirect loop of activation. Coherent loops, i.e., loops acting in concert, lead to product amplification due to the cooperative direction of the indirect and direct loops. Alternatively, incoherent feedforward loops, i.e., loops that are not acting in concert, contain indirect and direct loops oriented in opposite directions, leading to repression (Mangan et al., 2003).

Dynamic regulation is not only beneficial to natural systems; it can also be very powerful for engineered pathways and synthetic devices. One such example is a metabolic negative feedback loop based on a malonyl-CoA sensor-actuator in *E. coli*. When this dynamic regulatory circuit was used for the production of free fatty acids, there was a 34% improvement in free fatty acid titers obtained, compared to a strain lacking the dynamic control loop (Liu et al., 2013). Similarly, a dynamic regulatory circuit was utilized to improve fatty acid ethyl ester biodiesel formation three-fold (Zhang et al., 2012). These will be discussed in details in the following sections. The concept of harnessing transcription factors to dynamically regulate a metabolic pathway can be extended to a plethora of other pathways.

## 1.2 Dissertation Overview

In this dissertation, we focus on engineering genetic circuits to regulate engineered metabolic

pathways and develop design and tuning principles for genetic biosensors and metabolic circuits.

In Chapter 2, we examine the design constraints of transcription factor-based metabolite biosensors and identify a coupling between the response threshold and dynamic range when tuning the binding affinity between the transcription factor and the promoter. We further developed rational tuning principles that allow for orthogonal control over the biosensor output and its sensitivity.

In Chapter 3, we develop a malonyl-CoA sensor-actuator and apply it to construct a negative feedback circuit to regulate an engineered free fatty acid producing pathway. We demonstrate that the circuit recovers cell growth by dynamically controlling the expression of a key enzyme, thus alleviating its toxicity and improving free fatty acid titers and productivities.

In Chapter 4, metabolic dynamics are studied in various genetic circuits. Specifically, we construct negative feedback circuits with three different architectures, which represent three commonly found architectures in nature and engineered systems. Experimental measurements and mathematical simulations are combined to study the effects of regulation topologies and the associated biochemical parameters on the metabolic dynamics. The results reveal that negative feedback circuits can dramatically accelerate metabolic dynamics. The effects of control topologies and biochemical parameters on metabolic dynamics are also compared in these circuits.

Finally, in Chapter 5 the major conclusions and implications of the current work are summarized, with recommendations for future work.

# References


Alon, U. (2007) Network motifs: theory and experimental approaches, *Nat Rev Genet 8*, 450-461.

Becskei, A., and Serrano, L. (2000) Engineering stability in gene networks by autoregulation, *Nature 405*, 590-593.

Binder, S., Schendzielorz, G., Stabler, N., Krumbach, K., Hoffmann, K., Bott, M., Eggeling, L., 2012. A high-throughput approach to identify genomic variants of bacterial metabolite producers at the single-cell level. Genome Biol. 13.

Dahl, R. H., Zhang, F., Alonso-Gutierrez, J., Baidoo, E., Batth, T. S., Redding-Johanson, A. M., Petzold, C. J., Mukhopadhyay, A., Lee, T. S., Adams, P. D., Keasling, J. D., 2013. Engineering dynamic pathway regulation using stress-response promoters. Nat Biotechnol. 31**,** 1039-46.

Dietrich, J. A., Shis, D. L., Alikhani, A., Keasling, J. D., 2013. Transcription Factor-Based Screens and Synthetic Selections for Microbial Small-Molecule Biosynthesis. ACS synthetic biology. 2**,** 47-58.

Fiorentino, G., Ronca, R., Bartolucci, S., 2009. A novel E-coli biosensor for detecting aromatic aldehydes based on a responsive inducible archaeal promoter fused to the green fluorescent protein. Appl Microbiol Biot. 82**,** 67-77.

Kohlhaw, G. B. (2003) Leucine biosynthesis in fungi: Entering metabolism through the back door, *Microbiol Mol Biol R 67*, 1-+.

Levine, E., and Hwa, T. (2007) Stochastic fluctuations in metabolic pathways, *P Natl Acad Sci USA 104*, 9224-9229.

Liu, D., Xiao, Y., Evans, B., Zhang, F., 2013. Negative feedback regulation of fatty acid production based on a malonyl-CoA sensor-actuator. ACS synthetic biology.

Madar, D., Dekel, E., Bren, A., and Alon, U. (2011) Negative auto-regulation increases the input dynamic-range of the arabinose system of Escherichia coli, *Bmc Syst Biol 5*.

Mangan, S., and Alon, U. (2003) Structure and function of the feed-forward loop network motif, *P Natl Acad Sci USA 100*, 11980-11985.

Mustafi, N., Grunberger, A., Kohlheyer, D., Bott, M., Frunzke, J., 2012. The development and application of a single-cell biosensor for the detection of L-methionine and branched-chain amino acids. Metab Eng. 14**,** 449-457.

Oyarzun, D. A., Lugagne, J. B., and Stan, G. B. (2015) Noise propagation in synthetic gene circuits for metabolic control, *ACS synthetic biology 4*, 116-125.

Polizzi, K. M., Kontoravdi, C., 2015. Genetically-encoded biosensors for monitoring cellular stress in bioprocessing. Curr Opin Biotech. 31**,** 50-56.

Raman, S., Rogers, J. K., Taylor, N. D., Church, G. M., 2014. Evolution-guided optimization of biosynthetic pathways. P Natl Acad Sci USA. 111**,** 17803-17808.

Reed, B., Blazeck, J., Alper, H., 2012. Evolution of an alkane-inducible biosensor for increased responsiveness to short-chain alkanes. J Biotechnol. 158**,** 75-79.

Rosenfeld, N., Elowitz, M. B., and Alon, U. (2002) Negative autoregulation speeds the response times of transcription networks, *J Mol Biol 323*, 785-793.

Shen-Orr, S. S., Milo, R., Mangan, S., and Alon, U. (2002) Network motifs in the transcriptional regulation network of Escherichia coli, *Nat Genet 31*, 64-68.

Tang, S. Y., Cirino, P. C., 2011. Design and Application of a Mevalonate-Responsive Regulatory Protein. Angew Chem Int Edit. 50**,** 1084-1086.

Tang, S. Y., Fazelinia, H., Cirino, P. C., 2008. AraC regulatory protein mutants with altered effector specificity. J Am Chem Soc. 130**,** 5267-5271.

Tang, S. Y., Qian, S., Akinterinwa, O., Frei, C. S., Gredell, J. A., Cirino, P. C., 2013. Screening for Enhanced Triacetic Acid Lactone Production by Recombinant Escherichia coli Expressing a Designed Triacetic Acid Lactone Reporter. J Am Chem Soc. 135**,** 10099-10103.

Xie, G., Keyhani, N. O., Bonner, C. A., and Jensen, R. A. (2003) Ancient origin of the tryptophan operon and the dynamics of evolutionary change, *Microbiol Mol Biol R 67*, 303-+.

Xu, P., Wang, W. Y., Li, L. Y., Bhan, N., Zhang, F. M., Koffas, M. A. G., 2014. Design and Kinetic Analysis of a Hybrid Promoter-Regulator System for Malonyl-CoA Sensing in Escherichia coli. ACS chemical biology. 9**,** 451-458.

Zhang, F. Z., Carothers, J. M., Keasling, J. D., 2012. Design of a dynamic sensor-regulator system for production of chemicals and fuels derived from fatty acids. Nat Biotechnol. 30**,** 354-U166.

Zhang, F. Z., Keasling, J., 2011. Biosensors and their applications in microbial metabolic engineering. Trends Microbiol. 19**,** 323-329.

# Chapter 2 Fundamental Design Constraints of Transcription Factor-based Metabolite Biosensors

Note: This chapter contains text and figures from the published paper (Mannan, Ahmad A., et al. "Fundamental design principles for transcription-factor-based metabolite biosensors." *ACS Synthetic Biology*, 2017).

## Abstract

Metabolite biosensors are central to current efforts towards precision engineering of metabolism. Although most research has focused on building new biosensors, their tunability remains poorly understood and is a fundamental aspect for their broad applicability. Here we asked how genetic modifications shape the dose-response curve of biosensors based on metabolite-responsive transcription factors. Using the lac system in *Escherichia coli* as a model system, we built promoter libraries with variable operator sites that reveal interdependencies between biosensor dynamic range and response threshold. We developed a phenomenological theory to quantify such design constraints in biosensors with various architectures and tunable parameters. Our theory reveals a maximal biosensor dynamic range and exposes tunable parameters for orthogonal control of dynamic range and response threshold. Our work sheds light on fundamental limits of synthetic biology designs and provides quantitative guidelines for biosensor design in applications such as dynamic pathway control, strain optimization, and real-time monitoring of metabolism.

# 2.1 Introduction

A core principle in synthetic biology is the assembly of biological components into larger systems with predetermined functions. Metabolite biosensors, in particular, have received substantial attention because of their role in many applications at the interface of synthetic biology and metabolic engineering. Biosensors control gene expression in response to small molecules and provide a powerful tool to probe and control the metabolic state of a host. This makes them versatile for diverse applications, such as dynamic pathway engineering (1–4), high-throughput screening (5, 6) and complex genetic-metabolic circuitry (7).

A number of molecular mechanisms have been used for sensing intracellular metabolites, including e.g. RNA aptamers (8, 9) and metabolite-responsive transcription factors (TFs) (10, 11). The latter have become particularly popular because many organisms have evolved TFs that respond to native metabolites. In *E. coli*, for instance, about a third of TFs are known to respond to metabolites (12). Metabolite-responsive TFs can be re-purposed as biosensors in a different host (13) or re-engineered to respond to new ligands (14). The list of compounds for which biosensors have been developed is growing quickly (11) and includes precursors to biosynthetic pathways as well as products from secondary metabolism (15–19).

As illustrated in Figure 2.1A, for the purposes of biosensor design, metabolite-responsive TFs can be conceptualized as the composition of two modules: a sensing module for the interaction between the metabolite and the TF, and a regulation module where the TF controls the expression

of a target gene. Biosensors generally have one of four different architectures, depending on the type of interactions of the sensing and regulation module. Examples of these biosensor architectures can be found across diverse applications in metabolic engineering, see e.g. the reviews in (10, 11) or Table SF2 in the Supplementary File.

Most applications require biosensors to be tunable, so that designers can adjust biosensor output to the expected physiological concentration of a metabolite. Common strategies for biosensor tuning target the sensing and regulation modules separately, for example, via protein engineering to modify the binding kinetics between the metabolite and TF (14, 19), or promoter engineering to modify the transcriptional activity of the TF (20). Yet a major challenge for biosensor tuning is that their overall response compounds the effect of sensing and regulation, and thus changes to one component typically affect all parameters of the dose-response curve simultaneously (Figure 2.1B). As a result, biosensor design requires lengthy trial-and-error iterations between genetic modifications and strain characterization.

Previous studies have focused on the impact of transcriptional processes on the regulatory function of TFs (21, 22). Such studies have successfully used biophysical models to identify relations between parameters and the TF dose-response curve (23, 24). In the case of metabolite biosensors, however, their two-module architecture conflates the effect of metabolite sensing with the regulatory action of the TF. This makes it difficult to tease apart the impact of tunable parameters on the overall dose-response curve. This is especially relevant in metabolic engineering

applications, where biosensors are typically built with sensing and regulation modules taken from different sources, both of which can be tuned independently and have diverse molecular mechanisms (10, 11). As a result, biosensor design can benefit from system-level descriptions that abstract from mechanistic details and highlight the input-output dependencies among components.

Here we sought to characterize the interdependency between tunable parameters and the dose-response curves of metabolite biosensors. Combining phenomenological modeling and strain characterization, we provide a simple theory for the design of metabolite biosensors with various architectures and tunable parameters. Our results highlight fundamental constraints in biosensor design and expose tunable parameters that facilitate precise control of biosensor function.

# 2.2 Results

## 2.2.1 Design constraints in dose-response curves of metabolite biosensors

To study the relation between promoter tuning and biosensor function, we focused on metabolite-responsive TFs, the most widespread mechanism employed for sensing small molecules (10, 11, 25). In these biosensors, a convenient tunable parameter is the affinity of the TF to the promoter operator site, as it can be modified with rapid and cost-effective techniques such as random mutagenesis of promoter sequence or changes in operator copy number or location (25, 26).

As a model system for our investigation, we focused on the lactose inducible system in *Escherichia*

*coli*. We built eight *lacUV5*-based promoters with different mutations at the LacI-binding operator site (sequences in Figure 2.1C). We incorporated a red fluorescent protein (*rfp*) gene downstream of each promoter and measured the dose-response curve to varying concentrations of Isopropyl β-D-1-thiogalactopyranoside (IPTG), a non-metabolizable compound that mimics allolactose and induces the lac promoter by allosteric binding to the LacI repressor. We characterized the dose-response curves of each strain though steady state RFP fluorescence in mid-exponential growth. The resulting dose-response curves display significant differences (Figure 2.1C), with basal fluorescence outputs spanning two orders of magnitude and up to two-fold changes in maximal biosensor output.

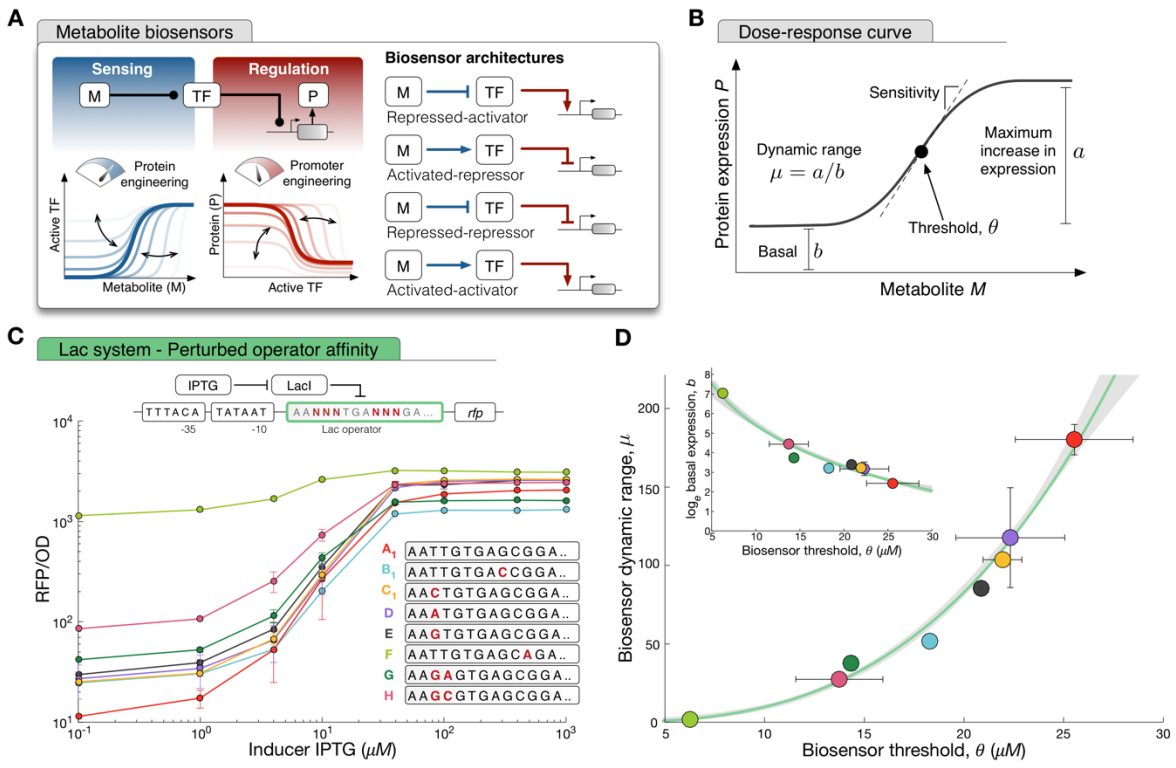



**Figure 2.1 Design constraints in dose-response curves of metabolite biosensors.**

**(A)** General architecture of biosensors based on metabolite-responsive transcription factors (TF). The metabolite (M) interacts with a TF that controls expression of a target protein (P). The sensing and regulation modules can be tuned with protein and promoter engineering. **(B)** Parameters that characterize a biosensor dose-response curve: basal output ($b$); response threshold ($\theta$), defined as the amount of metabolite required for 50% output expression relative to the baseline; dynamic range ($\mu$), defined as the maximal increase in expression relative to the basal output; response sensitivity, defined as the slope of the dose-response curve at the threshold. **(C)** Dose-response curves of LacI-based biosensors with variable operator sites, in response to IPTG induction. Error bars represent standard error from biological replicates; some error bars are too small to be observed. **(D)** Dose-response parameters of the strains in panel C, and fit of the phenomenological model in Eq. (1), green line. Error bars are the standard error of measured dose-response parameters across biological replicates; some error bars are too small to be observed. The gray band contains model predictions for 500 runs of the parameter estimation algorithm; the solid green line is the fit for a specific parameter set. Further details on the model fitting can be found Supplementary File 1.

To quantify the differences among the biosensors, we computed their dose-response parameters as defined in Figure 2.1B. We found a strong interdependency between the biosensor parameters for varying operator sites. In particular, when dynamic range ($\mu$) and threshold ($\theta$) are plotted against

each other (Figure 2.1D), we found a fundamental constraint for biosensor design. The constraint indicates that upon changes in operator affinity, biosensors with a broader dynamic range also display a larger response threshold. This suggests that tuning the promoter operator site can increase the dynamic range of the biosensor, but at the cost of simultaneously increasing the level of metabolite required to elicit a response.

To explain the observed constraint between threshold and dynamic range, we formulated a general mathematical model of the biosensors. Motivated by the modular description in Figure 2.1A, we used phenomenological models that describe the steady state of the sensing and regulation modules as Hill functions of their inputs. In the case of the lac system, which corresponds to the repressed-repressor architecture in Figure 2.1A, the model reads

$$f_1(M) = b_1 + \frac{a_1}{1 + (K_1 M)^{n_1}}, \qquad (1)$$
$$f_2(TF) = b_2 + \frac{a_2}{1 + (K_2 TF)^{n_2}},$$

where $f_1$ is the concentration of functional TF that can bind to the operator site as a function of the metabolite concentration (M), and $f_2$ is the expression level of the target gene (P) as a function of the concentration of repressor. The parameters $a_i$, $b_i$, $K_i$, and $n_i$ define the shape of the Hill curves for each module. In particular, the parameters $b_1$ and $b_2$ model the basal level of TF activity and promoter expression, respectively; parameters $a_1$ and $a_2$ are the maximum increase in TF activity and promoter expression, relative to their basal levels, respectively; parameter $K_1$ is the metabolite-TF affinity, and $K_2$ is the TF-operator affinity; and parameters $n_1$ and $n_2$ are proportional to the

sensitivity of metabolite-TF and TF-operator binding, respectively. Although specific molecular mechanisms can be well approximated by Hill functions similar to Eq. (1) (27–29), in this work we do not instance the models to specific mechanisms but rather focus on a phenomenological theory applicable to biosensors with various architectures. Furthermore, the model in Eq. (1) describes the steady state levels of the regulator (TF) and the target protein (P), and implicitly assumes that mRNA transcripts are also at steady state because mRNA half-lives are typically much shorter than protein lifetimes (30).

Under the phenomenological model in Eq. (1), the overall response of the biosensor is the composite function $P = f_2(f_1(M))$. The resulting dose-response curve has all the mathematical properties of a sigmoidal curve, namely, it is monotonically increasing, it has a single inflection point, and it approaches a finite limit value for large values of $M$. These properties hold in all the four biosensor architectures in Figure 2.1A (see Supplementary File 2S1).

From biophysical considerations, it is typically assumed that mutations to the promoter operator site affect the TF-operator affinity (21, 26). In our phenomenological approach, we further assume that changes to operator affinity can be captured by perturbations to the $K_2$ parameter only. This simplification allows us to model each strain in our library with a different value of $K_2$. We numerically computed the $(\mu,\theta)$ parameters from the dose-response curves $P = f_2(f_1(M))$ for varying values of $K_2$, and fitted the model parameters to the $(b,\mu,\theta)$ triplets from our fluorescence data (see Methods). Despite its simplified nature, the results in Figure 2.1D suggests that our

phenomenological model qualitatively reproduces the observed relation between dynamic range and threshold, thus providing a simple method to map the impact of mutations to the operator site onto the biosensor dose-response curve.

### 2.2.2 Phenomenological theory for biosensor tuning

To further elucidate the constraints that underpin biosensor design, we obtained formulae for the dose-response parameters in terms of the tunable parameters**.** For the repressed-repressor architecture, described by the model in Eq. (1), the dose-response parameters are

$$b = b_2 + \frac{a_2}{1 + \left(K_2(b_1 + a_1)\right)^{n_2}}, \tag{2}$$

$$a = a_2 \cdot \frac{\left((b_1 + a_1)^{n_2} - b_1^{n_2}\right) \cdot K_2^{n_2}}{\left(1 + \left(K_2(b_1 + a_1)\right)^{n_2}\right) \cdot (1 + (b_1 K_2)^{n_2})}, \tag{3}$$

$$\theta = \frac{1}{K_1} \cdot \sqrt[n_1]{\frac{a_1 K_2}{\sqrt[n_2]{A - 1} - b_1 K_2} - 1}, \tag{4}$$

where $A$ is a function $a_1$, $b_1$, $K_2$, and $n_2$, shown in Supplementary File S1. For brevity, we report the computation of the biosensor sensitivity in Supplementary File S1. From Eqs. (2) and (3), we compute the biosensor dynamic range as

$$\mu = \frac{a}{b} = \mu_2 \cdot \frac{\left((b_1 + a_1)^{n_2} - b_1^{n_2}\right) K_2^{n_2}}{\left(1 + \left(K_2(b_1 + a_1)\right)^{n_2} + \mu_2\right) \cdot (1 + (b_1 K_2)^{n_2})}, \tag{5}$$

with $\mu_2 = a_2/b_2$ being the dynamic range of promoter expression. The formulae for the dose-

response parameters of the other biosensor architectures (Figure 2.1A) can be found in the Supplementary File S1. The results in Eqs. (2)-(5) reveal that the dose-response parameters are coupled to one another through the TF-operator affinity ($K_2$). Changes to the operator sequence cause simultaneous changes to the basal output ($b$), dynamic range ($\mu$) and response threshold ($\theta$), in accordance with the dependency observed in our data in Figure 2.1C.

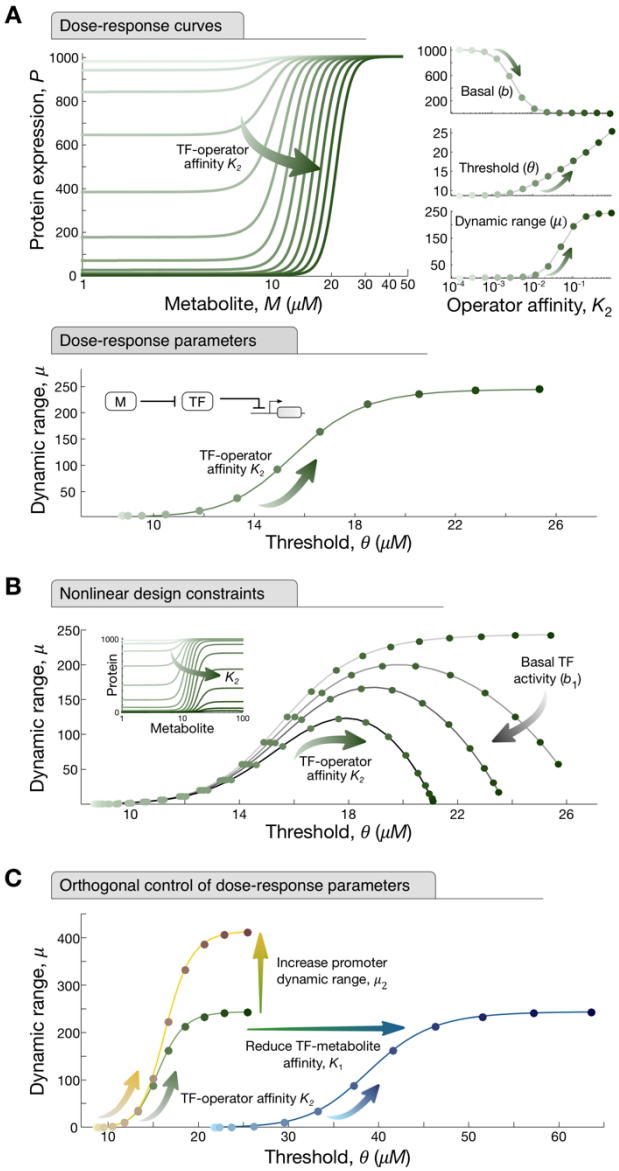



**Figure 2.2 Control of biosensor dose-response curves with tunable parameters.**

**(A)** Dose-response curves predicted from the model for biosensors with variable operator affinities. Our phenomenological theory predicts that tighter TF-operator binding leads to a lower basal output and a broader dynamic range, accompanied by an increased biosensor threshold. The bottom plot shows how this coupling between dynamic range ($\mu$) and threshold ($\theta$) constrains biosensor function, in agreement with the behavior observed from our data in Figure 2.1C-D. Parameter values are: $a_1 = 300$, $a_2 = 1000$, $b_1 = 0.01$, $b_2 = 4.1$, $n_1 = 6$, $n_2 = 2$, $K_1 = 0.1$, and $K_2$

values span the range $K_2 = 0.0005$ to $K_2 = 0.9$; all concentrations in $\mu$M. **(B)** Increased basal TF activity ($b_1$) leads to highly nonlinear constraints between dynamic range and threshold. Plots show increasing values of the $b_1$ parameter, from $b_1 = 0.01$ up to $b_1 = 8$. The inset shows the dose-response curves for increasing operator affinities ($K_2$), at a relatively high basal TF activity ($b_1 = 8$). **(C)** Two tunable parameters, the promoter dynamic range ($\mu_2$) and TF-metabolite affinity ($K_1$), provide orthogonal control and scale the dynamic range and threshold. Parameter values are the same as in panel A, except $a_2 = 1700$ (yellow curve) and $K_1 = 0.04$ (blue curve). In all panels, plots shown are for the repressed-repressor architecture, but the conclusions apply to all biosensor architectures in Figure 2.1A, see Supplementary File 2S1.

As shown in Figures 2.2A for a repressed-repressor architecture, for low TF-operator affinities ($K_2$) the biosensor produces an almost constitutive, metabolite-independent output, and thus displays low dynamic ranges. For increases in the $K_2$ parameter, the model predicts a decrease in basal biosensor output ($b$), with a relatively minor impact on the maximal output. This causes an increase in dynamic range and biosensor threshold, in agreement with what we observed in our lac promoter library (Figure 2.1D).

For high TF-operator affinities, however, our model predicts that the constraint between dynamic range and threshold depends strongly on the basal TF activity, modeled by the $b_1$ parameter in Eq. (1). In the case of the repressed-repressor architecture, the basal TF activity corresponds to the concentration of TF available for repression at maximum induction. When the $b_1$ parameter is nil,

we found a monotonic relationship between biosensor dynamic range and threshold, but an increased $b_1$ parameter produces a non-monotonic dependency between them (Figure 2.2B). This constraint appears in all biosensor architectures (see Supplementary File S1). In the case of the repressed-repressor architecture, this seemingly counterintuitive phenomenon arises because TFs with large basal activity (high $b_1$) will have some repressors available to bind to the operator site, even at full induction. An increased TF-operator affinity thus causes a stronger binding by these available repressors, decreasing protein expression, and in turn, lowering biosensor dynamic range, as shown in Figure 2.2B. In contrast, for TFs with negligible basal activity (low $b_1$) we do not observe a drop in dynamic range (Figure 2.2A), because at full induction there are so few repressors available that protein expression is insensitive to the TF-operator affinity.

A consequence of the constraint in Figure 2.2B is that changes to the TF-operator affinity can tune the biosensor threshold within a limited range only. In the case of the repressed-repressor architecture, the theoretical limits for the biosensor threshold are

$$\theta_{\min} = \frac{1}{K_1} \cdot \left( \frac{\mu_1}{2^{-\frac{1}{n_2}} \cdot (1 + (1+\mu_1)^{n_2})^{\frac{1}{n_2}} - 1} \right)^{\frac{1}{n_1}}, \tag{6}$$

$$\theta_{\max} = \frac{1}{K_1} \cdot \left( \frac{\mu_1}{2^{\frac{1}{n_2}} \cdot \left( \frac{(1+\mu_1)^{n_2}}{1 + (1+\mu_1)^{n_2}} \right)^{\frac{1}{n_2}} - 1} \right)^{\frac{1}{n_1}}, \tag{7}$$

where $\mu_1 = a_1/b_1$ is the dynamic range of TF activity. The limit thresholds for the other architectures

can be found in Supplementary File S1. From the formula in Eq. (5) we also computed the maximal dynamic range achievable with changes to the TF-operator affinity:

$$\mu_{\max} = \mu_2 \cdot \frac{((1+\mu_1)^{n_2} - 1)}{\left((1+\mu_1)^{\frac{n_2}{2}} + (1+\mu_2)^{\frac{1}{2}}\right)^2}, \tag{8}$$

As shown in Supplementary File S1, the formula for $\mu_{\max}$ applies to all four biosensor architectures in Figure 2.1A. Since Eq. (8) scales with both $\mu_1$ and $\mu_2$, it suggests that the maximal dynamic range can be controlled by tuning the TF expression level (through parameter $a_1$) or by adjusting the promoter strength (through parameter $a_2$). Detailed examination of the formula for $\mu_{\max}$, however, reveals that it has a more pronounced dependency on $\mu_2$, which is advantageous because according to Eqs. (2)-(4), parameter $\mu_1$ also affects all the other dose-response parameters simultaneously.

### 2.2.3 Orthogonal control of dynamic range and threshold

The results in Eqs. (4) and (5) reveal two tunable parameters, promoter dynamic range ($\mu_2$) and metabolite-TF affinity ($K_1$), that affect the biosensor dynamic range and threshold separately, whilst all remaining tunable parameters cause simultaneous changes in both. This means that $\mu_2$ and $K_1$ can be used for orthogonal control of dynamic range and threshold. In particular, the phenomenological model predicts that $\mu_2$ causes a vertical scaling in the ($\mu$,$\theta$)-curve, while $K_1$ scales it horizontally, as illustrated in Figure 2.2C. As shown in Supplementary file S1, we found that this behavior appears in all other biosensor architectures in Figure 2.1A, thus suggesting a

general principle for biosensor design.

To test the predicted orthogonal control in our lac system, we used two complementary strategies. First, we induced our strains in Figure 2.1C with methyl-1-thio-β-d-galactopyranoside (TMG), another gratuitous lac inducer with an affinity to LacI approximately 10 times lower than IPTG (31), which in our model corresponds to a reduced $K_1$ parameter. Second, we built six new lac promoters in *E. coli*. In addition to mutating the operator sites, the promoter strength was modified by replacing the -35 and -10 regions of the *lacUV5* promoter with those of the sequences from promoter $P_{A1}$ of phage T7. The $P_{A1}$ promoter has a higher binding affinity to RNA polymerase (32), and hence an increased promoter strength, which in our model corresponds to an increased value for $a_2$ and $\mu_2$.

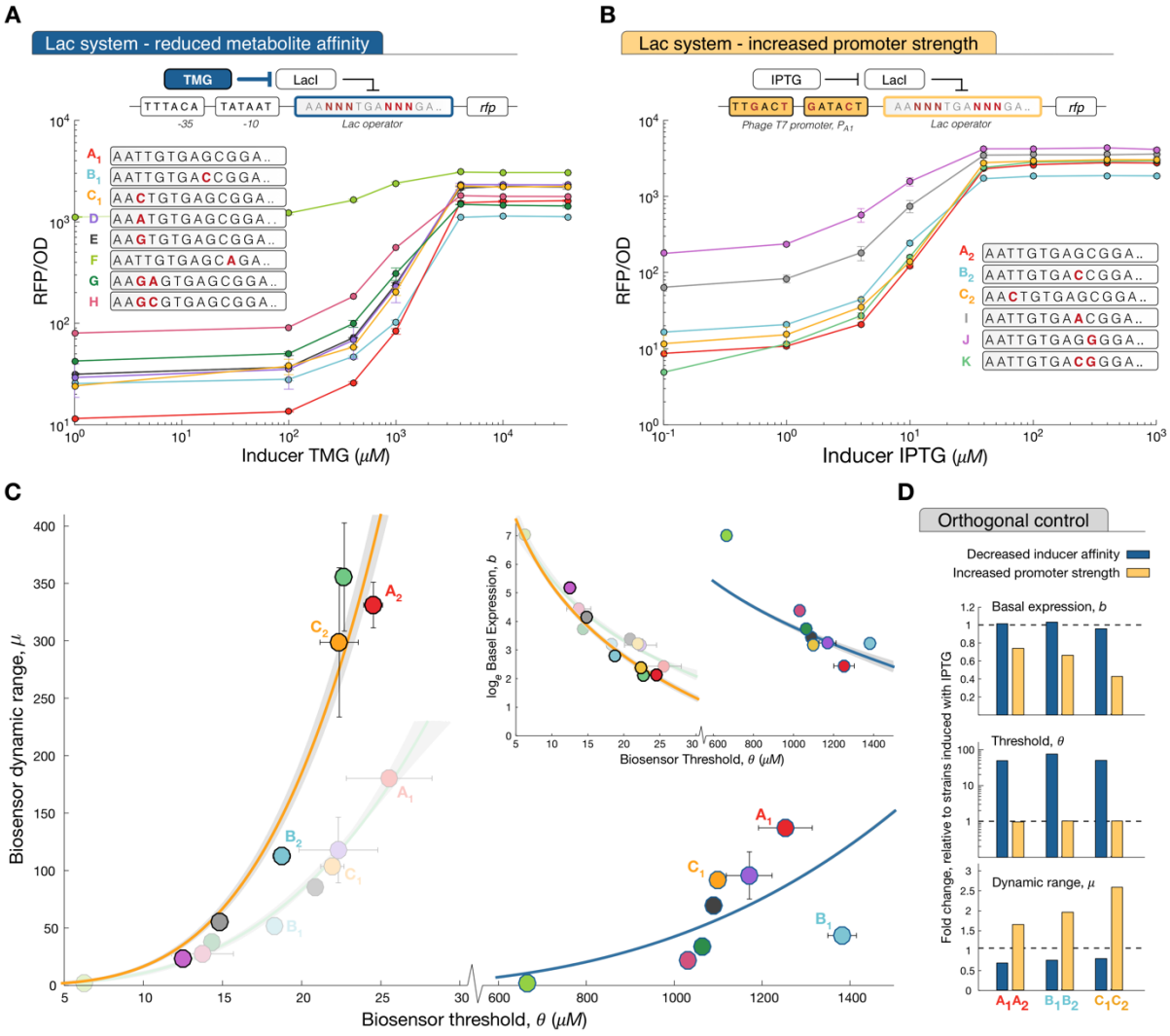



**Figure 2.3 Orthogonal control of biosensor dynamic range and threshold.**

**(A)** Dose-response curves of the lac strains from Figure 2.1C induced with TMG, which has a 10-fold lower affinity to LacI than IPTG. **(B)** Dose-response curves of strains with variable operator sites and increased promoter strength. Error bars represent standard error from biological replicates; some error bars are too small to be observed. **(C)** Dose-response parameters and model fits of strains in panel A (blue curve), and panel B (yellow curve). For comparison, the parameters of the IPTG-induced strains (Figure 2.1D) are in light gray. Error bars are the standard error of measured

dose-response parameters across biological replicates; some error bars are too small to be observed. The gray bands contain model predictions for 500 runs of the parameter estimation algorithm; the solid lines (blue, yellow) are fits for specific parameter sets. Further details on the model fitting can be found Supplementary File 1. **(D)** Validation of the predicted orthogonal control of dynamic range and threshold. We focused on three sets of strains ($A_i$, $B_i$, $C_i$) that are comparable across the three experiments. Strains within each set share the same operator sequence, but differ in their -35, -10 promoter region sequence. Bar plots show the fold change in dose-response parameters for a decreased inducer affinity or increased promoter strength, with respect to strains in Figure 2.1D.

We measured the dose-response curves of both sets of strains with RFP fluorescence, shown in Figure 2.3A-B, and quantified the dose-response parameters in Figure 2.3C. The results show a good qualitative agreement with our predictions. Strains with increased promoter strength display a larger dynamic range (Figure 2.3C, yellow), while a reduced metabolite-TF affinity caused an increase in response threshold (Figure 2.3C, blue), both with respect to our original strains in Figure 2.1C. We re-fitted the mathematical model for both sets of strains and observed that the model produces good fits in both cases. For the TMG-induced strains (blue curve, Figure 2.3C), we used the parameters from the IPTG experiments (Figure 2.1C) and re-fitted $K_1$ and $b_1$. Although we only expected a change in $K_1$, we were unable to produce good fits without also re-fitting parameter $b_1$. This is possibly because a lower affinity decreases the probability of TMG binding,

and so at full induction there are fewer repressors bound to TMG, as compared to those bound to IPTG, resulting in a higher $b_1$. For the strains with a promoter sequence perturbed at the -35 and -10 regions (yellow curve, Figure 2.3C), we used the parameters from the IPTG strains (Figure 2.1C) and re-fitted $a_2$, $b_2$ and $n_2$. We expected the $P_{A1}$ promoter sequence to affect parameter $a_2$ only, but we needed to re-fit $b_2$ and $n_2$ as well. This suggests biophysical couplings between tunable parameters that are not included in our model. Nevertheless, the results suggest that biosensor dynamic range increases with the $\mu_2 = a_2/b_2$ parameter, with negligible impact on the biosensor threshold, in agreement with our theoretical prediction.

To verify whether the changes in dynamic range and threshold were indeed orthogonal for changes in $K_1$ and $\mu_2$, we focused on three sets of strains, $A_1$ and $A_2$, $B_1$ and $B_2$, and $C_1$ and $C_2$, shown in Figure 2.3C. Strains within each set share the same operator sequence, but differ in the sequence of their -35 and -10 promoter regions. This enables us to survey the impact on the response parameters for changes in the two tunable parameters for each set of strains. Their response parameters, shown in Figure 2.3D, validate the predicted orthogonal control: a decrease in metabolite-TF affinity caused a large change in dose-response threshold and relatively small change in dynamic range, while a change in promoter strength caused a negligible change in threshold but a large shift in dynamic ranges.

## 2.3 Conclusions and Discussions

In this work we identified quantitative principles for the design of metabolite biosensors. Previous research on biosensors has focused primarily on expanding the repertoire of detectable metabolites (14, 33, 34). Most applications, however, require some degree of tunability on the biosensor dose-response curve, an aspect that remains poorly understood but is fundamental for their broad applicability. Given the substantial effort required to build new biosensors, a quantitative understanding of how tunable parameters shape biosensor function can help narrow down the design space, single out useful architectures, and determine the best experimental strategies to tune them.

Using the lac promoter in *E. coli* as a model system, we showed that mutations in the operator sequence simultaneously affect the basal output, dynamic range, and threshold of the dose-response curve. Such coupling between dose-response parameters makes it challenging to control biosensor function without a quantitative guideline. We quantified the parameter dependencies with a simple phenomenological model in which common tunable parameters such as TF-operator affinity, promoter strength and TF expression level can be readily incorporated, thus providing a widely applicable theory for biosensor design. Our theory revealed that upon changes in operator affinity, metabolite-responsive TFs are subject to design constraints between dynamic range and response threshold. These constraints become more severe for 'leaky' TFs that have a large basal activity, because they display a maximal achievable dynamic range that cannot be overcome by

changes to the operator site.

We also found that biosensor dynamic range and threshold can be controlled orthogonally with the promoter dynamic range and the TF-metabolite binding affinity. Our models predict that this principle holds in all considered biosensor architectures. Numerous promoter engineering techniques can be used to rapidly change promoter properties (20, 35, 36). Although changes to the TF-metabolite binding are significantly more challenging, recent progress in protein engineering have showcased the construction of metabolite-responsive TFs with perturbed affinities to their cognate ligands (19, 37), and even modified to bind to new molecules (14, 15). Our results suggest that promoter libraries with combinatorial designs for operator site and promoter strength cover a large portion of the design space for the biosensor dynamic range, whilst TF engineering can help to control the sensing threshold of leaky TFs.

Tunability of biosensors is essential for precision engineering of metabolism. In dynamic pathway engineering, biosensors control the expression of catalytic enzymes and are core components of feedback loops that re-route metabolic flux in response to pathway intermediates. For example, tuning the response curves of acyl-CoA and malonyl-CoA biosensors helped to increase production of fatty acid and fatty acid-derived fuels in *E. coli* (1, 2). With adequate tuning of the biosensor dose-response curve, feedback-regulated pathways can adapt their enzyme levels to the metabolic status of the host, prevent the accumulation of toxic intermediates, and control pathway variability (3, 4, 16, 38). The design constraints we have reported here thus highlight the need for

comprehensive characterization of biosensor libraries. The common approach of tuning biosensors based on dynamic range alone neglects potential knock-on effects on the sensing threshold, which may render a biosensor unresponsive to the physiological concentrations of a target metabolite.

Our theory has exposed design constraints applicable to biosensors with any of the four architectures in Figure 2.1A. Biosensors with various architectures have been already developed for a number of applications (10, 11). Our results provide a quantitative framework for orthogonal control of dynamic range and threshold in these biosensors, beyond the repressed-repressor architecture we have studied here. For example, the activated-activator architecture can be found in biosensors based on the TyrR regulator (5), while the repressed-activator and activated-repressor architectures can be found in biosensors based on SoxR and BetI, respectively (19, 39); see Table SF2 in the Supplementary File for more examples. Further, here we focused on design constraints under a variable operator affinity, as this is one of the most common tunable parameters. Our theory can be extended to uncover constraints for other tunable parameters accessible in specific applications, such as variable promoter strengths ($a_2$ parameter) or promoter sensitivity ($n_2$ parameter), which can be implemented with promoter libraries with variable RNA polymerase binding sites (36) or operator copy number (2). We also expect the use of other tunable parameters, such as the TF expression level, to produce more drastic changes in dose-response curves than those we observed here, for example by affecting the basal and maximal output simultaneously.

In this work we have deliberately used phenomenological models because they provide a versatile

tool to explore the parameter space for various biosensor architectures. The drawback of this simplification is that it overlooks the specific mechanisms for metabolite-TF binding and TF regulation (25). Our models also inherently assume that the tunable parameters are independent from one another, yet in reality they are coupled through the biophysical interactions between the TF, promoter and other transcriptional components such as RNA polymerases and $\sigma$-factors, which may produce further design constraints (21, 22). Although our data showed some effects that cannot be fully explained by our models, overall we found that our phenomenological theory provides a good first approximation to link design parameters with biosensor dose-response curves. Our approach revealed constraints in dose-response curves for common tunable parameters, providing a quantitative basis to identify useful biosensor architectures and to determine suitable experimental strategies for biosensor tuning.

Unlike most other engineering disciplines, synthetic biology suffers from a limited availability of sensing devices. Our work has uncovered fundamental design principles for metabolite biosensors, which in light of the tremendous progress in DNA, RNA and protein engineering, are essential to bring precision metabolic engineering closer to reality.

# 2.4 Methods

## 2.4.1 Mathematical modeling and curve fitting

The phenomenological models were built through the composite function $P = f_2(f_1(\mathrm{M}))$, where $f_1$

and $f_2$ are increasing or decreasing Hill functions. The function $P(M)$ is monotonic in $M$, has a single inflection point, and reaches its extremal values at $M = 0$ and $M \to \infty$, thus resembling a sigmoid function. We obtained the biosensor parameters $a$ and $b$ from the definitions

$$a = \max f_2\big(f_1(M)\big) - \min f_2\big(f_1(M)\big)\,,$$

$$b = \min f_2\big(f_1(M)\big)\,,$$

for $M \geq 0$, from where the dynamic range is $\mu = a/b$. We computed the response threshold $\theta$ as the solution of the equation

$$f_2\big(f_1(\theta)\big) = b + \frac{a}{2},$$

which represents the metabolite concentration for 50% output, relative to the basal. The computation of the dose-response sensitivity is explained in Supplementary File 1. The upper and lower bounds on $\mu$ and $\theta$ were computed by differentiation with respect to TF-operator affinity $K_2$ (details in Supplementary File S1). Equations (2)-(7) are valid for the repressed-repressor architecture, while Eq. (8) for the maximal dynamic range is valid for all architectures. Detailed calculations of dose-response parameters in all architectures are given in Supplementary File S1.

For the parameter fitting in Figures 2.1D and 2.3C, we obtained the dose-response parameters from the promoter characterization data, $\mu_{i,E}$, $\theta_{i,E}$ and $b_{i,E}$, of the $i^{\text{th}}$ strain for each of the three sets of experiments. The dynamic range ($\mu_{i,E}$) and basal expression level ($b_{i,E}$) were calculated from the data at zero and full induction, and the threshold ($\theta_{i,E}$) from fits of a Hill function to the data. The model predictions ($\mu_E, b_E, \theta_E$) were generated from equations (2)-(5) evaluated over a fixed

range of $K_2$ from $K_2$=3×10$^{-5}$ to $K_2$=10$^3$, with the other seven parameters $p = (a_1, b_1, K_1, n_1, a_2, b_2, n_2)$ fitted to the data triplets ($\mu_{i,E}$, $\theta_{i,E}, b_{i,E}$) via nonlinear least-squares. We solved the optimization problem

$$\min_p \sum_{i=1}^{\#\,\text{Strains}} \left(\frac{\mu_{i,E} - \mu_{i,M}(K_{2,i}, p)}{\hat{\mu}_E}\right)^2 + \left(\frac{\theta_{i,E} - \theta_{i,M}(K_{2,i}, p)}{\hat{\theta}_E}\right)^2 + \left(\frac{\log_{10} b_{i,E} - \log_{10} b_{i,M}(K_{2,i}, p)}{\log_{10} \hat{b}_E}\right)^2,$$

where $\mu_{i,M}$, $\theta_{i,M}$ and $b_{i,M}$ are computed from equations (2)-(5) and $K_{2,i}$ corresponding to the value of $K_2$ where the model prediction is closest to the $i^{\text{th}}$ data point. Each term of the objective is normalized by the maximum measured value, denoted $\hat{\mu}_E$, $\hat{\theta}_E$ or $\hat{b}_E$. In Figure 2.1D we fitted all model parameters. The green line is a model fit with parameters reported in Table SF1 in Supplementary File 1. In Figure 2.3C (blue) we used the fitted parameters from Figure 2.1D (green) and re-fitted $K_1$ and $b_1$. In Figure 2.3C (yellow), we used the parameters from Figure 2.1D (green) and refitted $a_2$, $b_2$ and $n_2$. Further details on the parameter fitting can be found in Supplementary File 1. The parameter fitting was done through 500 runs of the fmincon solver in MATLAB Global Optimization toolbox.

### 2.4.2 Construction and characterization of promoters

The lac promoter libraries were constructed by introducing mutations to the lac operator site in a Biobrick plasmid pBbB5k-RFP (40). The $P_{A1}$ promoter library (increased $\mu_2$) was cloned by switching the -35 and -10 regions of the LacUV5 promoter to those of $P_{A1}$ promoter from phage

T7 (sequence shown in Figure 2.3B), yielding plasmid pBbB5pgk-RFP. To vary the $K_2$ parameter, two lac promoter libraries were created with operator sequences AATTGTGANNNGATAACAATT and AANNNTGAGCGGATAACAAT (Figure 2.1C, 2.3A & 2.3B), generating a strain library with the size of 128 sequences. The lac promoter libraries were constructed using a one-step Golden-Gate DNA assembly method and were then transformed into MG1655 cells. The promoter libraries were pre-screened from a random selection of colonies from the whole library, and the ones with relatively high and distinct dynamic ranges were selected for further characterization.

Cell growth and fluorescence were recorded on an Infinite F200PRO (Tecan) plate reader. Strains were first cultivated overnight in Luria-Bertani (LB) medium supplemented with 50 mg/L kanamycin. The overnight LB cultures of the lac promoter strains were inoculated 2% v/v into M9 medium supplemented with 1% glycerol, 50 mg/L kanamycine, and amino acids, composed as for the EZ-rich medium (41) for adaptation. The overnight culture was inoculated 2% v/v into the same medium and grew to an $OD_{600}$ of 0.6. Cell cultures were then diluted by 30-fold in the same medium and induced with varying IPTG (0.1, 1, 4, 10, 40, 100, 400, and 1000 μM) and TMG (1, 100, 400, 1000, 2000, 4000, 10000, and 40000 μM) concentrations.

Cell density ($OD_{600}$) and red fluorescence (excitation: 535 ± 9 nm; emission: 620 ± 20 nm) were recorded every 1000 s until the cell culture reached the stationary phase. Fluorescence from a wild-type *E. coli* MG1655 cell culture was used as the background, and was subtracted from all

fluorescence measurements. The background-corrected fluorescence was later normalized by cell density as measured at $OD_{600}$. Cells were maintained in the exponential growth phase for 5-6 cell cycles until the normalized fluorescence reached to the steady state, and the steady state fluorescence were used to generate the dose-response curves. We extracted the biosensor parameters ($b$, $\mu$, $\theta$) from Hill functions fitted to the measured dose-response curves. Standard errors of the dose-response parameters were calculated from the fitted parameters of the response curves for biological replicates, for each strain.

# Acknowledgements

This work was funded by the Human Frontier Science Program through a Young Investigator Grant awarded to D.O and F.Z. (Grant no. RGY0076-2015) and the US National Science Foundation (MCB1453147) to F.Z.

# References


1. Zhang F, Carothers JM, Keasling JD (2012) Design of a dynamic sensor-regulator system for production of chemicals and fuels derived from fatty acids. *Nat Biotechnol* 30(4):354–9.
2. Xu P, Li L, Zhang F, Stephanopoulos G, Koffas M (2014) Improving fatty acids production by engineering dynamic pathway regulation and metabolic control. *Proc Natl Acad Sci* 111(31):11299–11304.
3. Oyarzún DA, Stan G-B V (2013) Synthetic gene circuits for metabolic control: design trade-offs and constraints. *J R Soc Interface* 10(78). doi:10.1098/rsif.2012.0671.
4. Dunlop MJ, Keasling JD, Mukhopadhyay A (2010) A model for improving microbial biofuel production using a synthetic feedback loop. *Syst Synth Biol* 4(2):95–104.
5. Xiao Y, Bowen CH, Liu D, Zhang F (2016) Exploiting nongenetic cell-to-cell variation for enhanced biosynthesis. *Nat Chem Biol* 12(5):339–344.
6. Rogers JK, et al. (2015) Synthetic biosensors for precise gene control and real-time monitoring of metabolites. *Nucleic Acids Res* 43(15):7648–7660.
7. Oyarzún DA, Chaves M (2015) Design of a bistable switch to control cellular uptake. *J R Soc Interface* 12(20150618). doi:doi: 10.1098/rsif.2015.0618.
8. Beisel CL, Smolke CD (2009) Design principles for riboswitch function. *PLoS Comput Biol* 5(4):e1000363.
9. Berens C, Suess B (2015) Riboswitch engineering - making the all-important second and third steps. *Curr Opin Biotechnol* 31:10–15.
10. Liu D, Evans T, Zhang F (2015) Applications and advances of metabolite biosensors for metabolic engineering. *Metab Eng* 31:35–43.
11. Mahr R, Frunzke J (2016) Transcription factor-based biosensors in biotechnology: current state and future prospects. *Appl Microbiol Biotechnol* 100(1):79–90.
12. Chubukov V, Gerosa L, Kochanowski K, Sauer U (2014) Coordination of microbial metabolism. *Nat Rev Microbiol* 12(5):327–40.
13. Skjoedt ML, et al. (2016) Engineering prokaryotic transcriptional activators as metabolite biosensors in yeast. *Nat Chem Biol* 12(11):951–958.
14. Taylor ND, et al. (2016) Engineering an allosteric transcription factor to respond to new ligands. *Nat Methods* 13(2):177–183.
15. Tang S-Y, Cirino PC (2011) Design and application of a mevalonate-responsive regulatory protein. *Angew Chem Int Ed Engl* 50(5):1084–1086.
16. Liu D, Xiao Y, Evans BS, Zhang F (2015) Negative feedback regulation of fatty acid production based on a malonyl-CoA sensor-actuator. *ACS Synth Biol* 4(2):132–140.
17. Xu P, et al. (2014) Design and Kinetic Analysis of a Hybrid Promoter–Regulator System for Malonyl-CoA Sensing in Escherichia coli. *ACS Chem Biol* 9(2):451–458.
18. Chou HH, Keasling JD (2013) Programming adaptive control to evolve increased

metabolite production. *Nat Commun* 4:2595.
19. Saeki K, Tominaga M, Kawai-Noma S, Saito K, Umeno D (2016) The rapid diversification of BetI-based transcriptional switches for the control of biosynthetic pathways and genetic circuits. *ACS Synth Biol*. doi:10.1021/acssynbio.5b00230.
20. Blazeck J, Alper HS (2013) Promoter engineering: recent advances in controlling transcription at the most fundamental level. *Biotechnol J* 8(1):46–58.
21. von Hippel PH, Berg OG (1986) On the specificity of DNA-protein interactions. *Proc Natl Acad Sci U S A* 83(6):1608–12.
22. Bintu L, et al. (2005) Transcriptional regulation by the numbers: Models. *Curr Opin Genet Dev* 15(2):116–124.
23. Martins BMC, Swain PS (2011) Trade-Offs and Constraints in Allosteric Sensing. 7(11):1–13.
24. Ang J, Harris E, Hussey BJ, Kil R, McMillen DR (2013) Tuning response curves for synthetic biology. *ACS Synth Biol* 2(10):547–67.
25. de Paepe B, Peters G, Coussement P, Maertens J, de Mey M (2016) Tailor-made transcriptional biosensors for optimizing microbial cell factories. *J Ind Microbiol Biotechnol*:1–23.
26. Garcia HG, et al. (2012) Operator sequence alters gene expression independently of transcription factor occupancy in bacteria. *Cell Rep* 2(1):150–161.
27. Yagil G, Yagil E (1971) On the Relation between Effector Concentration and the Rate of Induced Enzyme Synthesis. *Biophys J* 11(1):11–27.
28. Buchler NE, Cross FR (2009) Protein sequestration generates a flexible ultrasensitive response in a genetic network. *Mol Syst Biol* 5:272.
29. Bintu L, et al. (2005) Transcriptional regulation by the numbers: Applications. *Curr Opin Genet Dev* 15(2):125–135.
30. Bremer H, Dennis P (1996) Modulation of chemical composition and other parameters of the cell by growth rate. In Escherichia coli and Salmonella typhimurium. *Neidhardt, F (ed) Washington, DC Am Soc Microbiol Press* (122):1553.
31. Marbach A, Bettenbrock K (2012) lac operon induction in Escherichia coli: Systematic comparison of IPTG and TMG induction and influence of the transacetylase LacA. *J Biotechnol* 157(1):82–88.
32. Lanzer M, Bujard H (1988) Promoters largely determine the efficiency of repressor action. *Proc Natl Acad Sci U S A* 85(23):8973–7.
33. Younger AKD, Dalvie NC, Rottinghaus AG, Leonard JN (2017) Engineering Modular Biosensors to Confer Metabolite-Responsive Regulation of Transcription. *ACS Synth Biol* 6(2):311–325.
34. Nadler DC, Morgan S-A, Flamholz A, Kortright KE, Savage DF (2016) Rapid construction of metabolite biosensors using domain-insertion profiling. *Nat Commun* 7:12266.
35. De Mey M, et al. (2010) Promoter knock-in: a novel rational method for the fine tuning of

genes. *BMC Biotechnol* 10:26.
36. Brewster RC, Jones DL, Phillips R (2012) Tuning promoter strength through RNA polymerase binding site design in Escherichia coli. *PLoS Comput Biol* 8(12):e1002811.
37. Richards DH, Meyer S, Wilson CJ (2017) Fourteen Ways to Reroute Cooperative Communication in the Lactose Repressor: Engineering Regulatory Proteins with Alternate Repressive Functions. *ACS Synth Biol* 6(1):6–12.
38. Oyarzún DA, Lugagne J-B, Stan G-B (2014) Noise propagation in synthetic gene circuits for metabolic control. *ACS Synth Biol*. doi:10.1021/sb400126a.
39. Siedler S, et al. (2014) SoxR as a single-cell biosensor for NADPH-consuming enzymes in Escherichia coli. *ACS Synth Biol* 3(1):41–47.
40. Mason G, et al. (2002) BglBrick vectors and datasheets: A synthetic biology platform for gene expression. *BMC Biotechnol* 2(1):20.
41. Neidhardt FC, Bloch PL, Smith DF (1974) Culture Medium for Enterobacteria. *J Bacteriol* 119(3):736–747.

# Appendix: Supplementary Information for Chapter 2

**Supplementary Material**

**Fundamental design principles for transcription-factor-based metabolite biosensors**

Ahmad A. Mannan[1], Di Liu[2], Fuzhong Zhang[2*] and Diego A. Oyarzún[1*].

[1] Department of Mathamatics, Imperial College London, UK.
[2] Department of Energy, Environmental & Chemical Engineering, Washington University in St Louis, USA.
* Corresponding authors: fzhang@seas.wustl.edu, d.oyarzun@imperial.ac.uk

**Summary.** Here we present the phenomenological models used in the paper, together with the calculations of the dose-response parameters and further details on the parameter fitting.

## 1 Phenomenological model for metabolite biosensors

As shown in Figure 1A of the main text, we model metabolite biosensors as a cascade of two modules: a sensing module, which describes the interaction between the metabolite ($M$) and transcription factor ($TF$), and a regulation module, which describes the interaction between the transcription factor and protein expression ($P$). We describe the dose-response curve $P(M)$ as the composition of two Hill functions, $P = f_2(f_1(M))$, where $TF = f_1(M)$ is the activity level of the transcription factor for a given amount of metabolite $M$, and $P = f_2(TF)$ is the output expression for a given concentration of transcription factor. Depending on whether the interactions are inhibitory or activatory, we use the following Hill functions

$$\text{Inhibitory} \quad : \quad f_i(x) = b_i + \frac{a_i}{1 + (K_i \cdot x)^{n_i}}, \text{ for } i = \{1, 2\}, \tag{1}$$

$$\text{Acivatory} \quad : \quad f_i(x) = b_i + \frac{a_i \cdot (K_i \cdot x)^{n_i}}{1 + (K_i \cdot x)^{n_i}}, \text{ for } i = \{1, 2\}, \tag{2}$$

where the parameters describe different aspects of the sensing and regulation modules (see main text). The dose-response curves for the four architectures in Figure 1A are thus

- Activated-Repressor (AR)

$$f_2(f_1(M)) = b_2 + \frac{a_2}{1 + \left(K_2 \cdot \left(b_1 + \frac{a_1 \cdot (K_1 \cdot M)^{n_1}}{1 + (K_1 \cdot M)^{n_1}}\right)\right)^{n_2}}, \tag{3}$$

- Repressed-Activator (RA)

$$f_2(f_1(M)) = b_2 + \frac{a_2 \cdot \left(K_2 \cdot \left(b_1 + \frac{a_1}{1 + (K_1 \cdot M)^{n_1}}\right)\right)^{n_2}}{1 + \left(K_2 \cdot \left(b_1 + \frac{a_1}{1 + (K_1 \cdot M)^{n_1}}\right)\right)^{n_2}}, \tag{4}$$

- Repressed-Repressor (RR)

$$f_2(f_1(M)) = b_2 + \frac{a_2}{1 + \left(K_2 \cdot \left(b_1 + \frac{a_1}{1 + (K_1 \cdot M)^{n_1}}\right)\right)^{n_2}}, \tag{5}$$

- Activated-Activator (AA)

$$f_2(f_1(M)) = b_2 + \frac{a_2 \cdot \left(K_2 \cdot \left(b_1 + \frac{a_1 \cdot (K_1 \cdot M)^{n_1}}{1 + (K_1 \cdot M)^{n_1}}\right)\right)^{n_2}}{1 + \left(K_2 \cdot \left(b_1 + \frac{a_1 \cdot (K_1 \cdot M)^{n_1}}{1 + (K_1 \cdot M)^{n_1}}\right)\right)^{n_2}}. \tag{6}$$

It can be shown that the composite functions in (3)–(6) have the following properties: they have one minimum and one maximum, they are monotonically decreasing (AR and RA) or increasing (AA and RR), and they have a single inflection point. As a result, the composite functions are qualitatively similar to sigmoid functions, and hence we can precisely define the dose-response parameters, as shown next.

## 2 Formulas for the dose-response parameters

**Basal output ($b$).** The basal expression, defined as the minimum value of the dose-response curve, can be computed by evaluating the expressions in (3)–(6) for $M = 0$ or $M \to \infty$:

$$b = \begin{cases} f_2(f_1(0)) \,, & \text{for AA and RR architectures,} \\ \\ f_2(f_1(\infty)) \,, & \text{for RA and AR architectures,} \end{cases} \tag{7}$$

which leads to

$$b = \begin{cases} b_2 + \dfrac{a_2}{1 + K_2^{n_2}(b_1 + a_1)^{n_2}} \,, & \text{for AR and RR architectures,} \\ \\ b_2 + \dfrac{a_2 \cdot (K_2 b_1)^{n_2}}{1 + (K_2 b_1)^{n_2}} \,, & \text{for RA and AA architectures.} \end{cases} \tag{8}$$

**Maximum change in output ($a$).** The maximum change in output is the difference between the maximum and minimum value of the dose-response curve. Similarly as for the basal expression ($b$), we derive the formula for $a$ from the definition:

$$a = \begin{cases} f_2(f_1(\infty)) - f_2(f_1(0)) \,, & \text{for AA and RR architectures,} \\ \\ f_2(f_1(0)) - f_2(f_1(\infty)) \,, & \text{for RA and AR architectures,} \end{cases} \tag{9}$$

which leads to a formula for $a$ that is identical for all four biosensor architectures

$$a = a_2 \cdot K_2^{n_2} \cdot \left( \frac{(b_1 + a_1)^{n_2} - b_1^{n_2}}{(1 + K_2^{n_2}(b_1 + a_1)^{n_2}) \cdot (1 + (K_2 b_1)^{n_2})} \right). \tag{10}$$

**Biosensor dynamic range ($\mu$).** From the previous definitions, the dynamic range can be simply computed as $\mu = a/b$, with $a$ and $b$ as given in equations (8) and (10). The resulting expressions are

$$\mu = \begin{cases} \mu_2 \cdot \frac{((b_1+a_1)^{n_2} - b_1^{n_2}) K_2^{n_2}}{(1+\mu_2+(K_2(b_1+a_1))^{n_2}) \cdot (1+(b_1 K_2)^{n_2})} \,, & \text{for AR and RR architectures,} \\ \\ \mu_2 \cdot \frac{((b_1+a_1)^{n_2} - b_1^{n_2}) K_2^{n_2}}{(1+(K_2(b_1+a_1))^{n_2}) \cdot (1+(1+\mu_2)(b_1 K_2)^{n_2})} \,, & \text{for RA and AA architectures.} \end{cases} \tag{11}$$

The formula for $\mu$ in the repressed-repressor architecture in equation (11) is the same as equation (5) in the main text.

**Biosensor threshold ($\theta$).** We define the threshold as the concentration of metabolite for which the level of biosensor output expression is 50% of the maximum, relative to the basal. This definition corresponds to solving the following equation for $\theta$,

$$f_2(f_1(\theta)) = b + \frac{a}{2}, \tag{12}$$

where $b$ and $a$ are given in equations (8) and (10), respectively. Solving equation (12) for each architecture in (3)–(6), we get

$$\theta = \begin{cases} \dfrac{1}{K_1} \cdot \sqrt[n_1]{\dfrac{a_1 K_2}{\sqrt[n_2]{A-1} - b_1 K_2} - 1}\,, & \text{for RA and RR architectures,} \\ \dfrac{1}{K_1} \cdot \sqrt[n_1]{\dfrac{1}{\dfrac{a_1 K_2}{\sqrt[n_2]{A-1} - b_1 K_2} - 1}}\,, & \text{for AR and AA architectures,} \end{cases} \tag{13}$$

where $A$ is a function of the model parameters

$$A = 2 \cdot \frac{(1 + (K_2(b_1 + a_1))^{n_2}) \cdot (1 + (b_1 K_2)^{n_2})}{2 + ((b_1 + a_1)^{n_2} + b_1^{n_2}) K_2^{n_2}}\,. \tag{14}$$

**Biosensor sensitivity ($n$).** We define the sensitivity as the slope of the normalized dose-response curve

$$n = 4\theta \cdot \left| \frac{\mathrm{d}}{\mathrm{d}M} \left( \frac{f_2(f_1(M)) - b}{a} \right)_{M=\theta} \right|, \tag{15}$$

where $|\cdot|$ is the absolute value, and $(a, b, \theta)$ are the dose-response parameters computed in equations (8), (10) and (13). The definition in (15) applies to the four biosensor architectures and is inspired by the observation that in a Hill equation of the form $f(x) = (x/\theta)^n / (1 + (x/\theta)^n)$, it can be shown that $n = 4\theta\, \mathrm{d}f/\mathrm{d}x|_{x=\theta}$. The definition in (15) leads to explicit formulas for the sensitivity as a function of tunable parameters, but the resulting expressions are involved and we omit them for brevity.

## 3 Bounds for dose-response parameters

### 3.1 Constraints between dynamic range and threshold

In this work we consider the TF-operator affinity ($K_2$) as the main tuneable parameter to adjust biosensor function. Thus in this section we consider the dose-response parameters as functions of $K_2$, that is, from the expressions in equations (8)–(13) we write $b = b(K_2)$, $\mu = \mu(K_2)$ and $\theta = \theta(K_2)$. The formulas for the dose-response parameter reveal that dynamic range and threshold are inherently coupled to each other through the $K_2$ parameter. In Figure SF1 we plot $\mu(K_2)$ vs $\theta(K_2)$ and observe that the shape of the curve depends strongly on the $b_1$ parameter, which represents the basal level of TF activity. Numerical exploration of the parameter space suggests that for $b_1 = 0$, the $(\mu, \theta)$-curve is monotonic, while a non-zero $b_1$ value leads to a non-monotonic relationship between threshold and dynamic range. This phenomenon appears in all four biosensor architectures.

From the expression for the basal output in (8), it can be shown that $b(K_2)$ is a decreasing function of $K_2$ for the AR and RR architectures, and an increasing function of $K_2$ for the RA and AA architectures. In contrast, we found that the function $a(K_2)$ defined in (10) is non-monotonic, and reaches a maximum

$$a_{\max} = \frac{a_2 \cdot ((1 + \mu_1)^{n_2} - 1)}{\left((1 + \mu_1)^{\frac{n_2}{2}} + 1\right)^2}, \tag{16}$$

for an optimal operator affinity

$$K_2^{\max,a} = \frac{1}{b_1} \cdot \frac{1}{\sqrt{1 + \mu_1}}, \tag{17}$$

where $\mu_1 = a_1/b_1$. Note that when $b_1 = 0$, the optimal operator affinity $K_2^{\max,a} \to \infty$, and thus $a$ becomes a monotonic function of $K_2$.

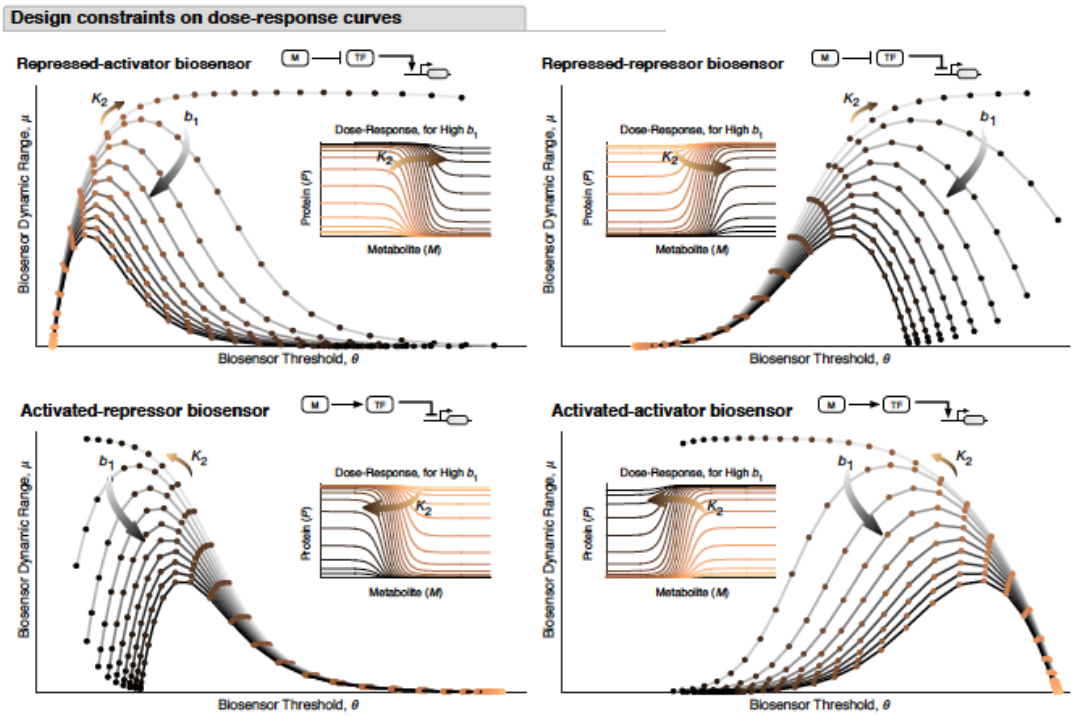



Figure SF 1: Plots of the $(\mu(K_2), \theta(K_2))$-curves for each biosensor architecture and increasing values of the $b_1$ parameter. The insets show the dose-response curves for different $K_2$ values for a non-zero value of $b_1$.

### 3.2 Maximal dynamic range

To compute the operator affinity that maximizes the dynamic range, we solved the equation $\mathrm{d}\mu(K_2)/\mathrm{d}K_2 = 0$ with $\mu(K_2) = a(K_2)/b(K_2)$ and $a(K_2)$ and $b(K_2)$ given in (8) and (10), respectively. We found that the dynamic range reaches a maximum at

$$K_2^{\max,\mu} = \begin{cases} \dfrac{1}{b_1} \cdot \dfrac{1}{\sqrt{1+\mu_1}} \cdot (1+\mu_2)^{\frac{1}{2n_2}} & \text{for AR and RR architectures,} \\ \dfrac{1}{b_1} \cdot \dfrac{1}{\sqrt{1+\mu_1}} \cdot \dfrac{1}{(1+\mu_2)^{\frac{1}{2n_2}}} & \text{for RA and AA architectures.} \end{cases} \tag{18}$$

The maximal dynamic range is then

$$\mu_{\max} = \mu_2 \cdot \frac{(1+\mu_1)^{n_2} - 1}{\left((1+\mu_1)^{\frac{n_2}{2}} + (1+\mu_2)^{\frac{1}{2}}\right)^2}, \tag{19}$$

where $\mu_2 = a_2/b_2$. The expression in (19) is valid for all four biosensor architectures. Importantly, we also note that in the limit $b_1 \to 0$, the optimal affinity $K_2^{\max,\mu} \to \infty$, which means that dynamic range becomes a monotonically increasing function of $K_2$, in agreement with the numerical observations in Figure SF1.

### 3.3 Bounds for the dose-response threshold

From the expression for $\theta$ in equation (13), we note that $\theta$ has minimal and maximal values depending on the extremal values of the function $A$ in (14). It can be shown that $\mathrm{d}A/\mathrm{d}K_2 < 0$ for all $K_2 > 0$ and

hence $A$ is a monotonically decreasing function of $K_2$, with a maximum at $K_2 = 0$ and a minimum at $K_2 \to \infty$. Assuming all parameters other than $K_2$ are fixed, positive real values, we evaluated $A(K_2)$ for $K_2 = 0$ and $K_2 \to \infty$, to get the expressions for $A_{\max}$ and $A_{\min}$, respectively, as

$$A_{\max} = A(0) = 2^{\frac{-1}{n_2}} \cdot b_1 \cdot (1 + (1+\mu_1)^{n_2})^{\frac{1}{n_2}}, \tag{20}$$

$$A_{\min} = A(\infty) = 2^{\frac{1}{n_2}} \cdot (b_1 + a_1) \cdot \frac{1}{(1 + (1+\mu_1)^{n_2})^{\frac{1}{n_2}}}. \tag{21}$$

We then computed the maximum and minimum bounds of the response threshold, $\theta_{\min}$ and $\theta_{\max}$, by substituting $A_{\max}$ and $A_{\min}$ into expression for the threshold $\theta$ in (13). The bounds for the threshold are

- For RA and RR architectures:

$$\theta_{\min} = \frac{1}{K_1} \cdot \sqrt[n_1]{\frac{\mu_1}{2^{-\frac{1}{n_2}} \cdot (1 + (1+\mu_1)^{n_2})^{\frac{1}{n_2}} - 1} - 1}, \tag{22}$$

$$\theta_{\max} = \frac{1}{K_1} \cdot \sqrt[n_1]{\frac{\mu_1}{2^{\frac{1}{n_2}} \cdot \left( \frac{(1+\mu_1)}{(1 + (1+\mu_1)^{n_2})^{\frac{1}{n_2}}} \right) - 1} - 1}. \tag{23}$$

- For AR and AA architectures:

$$\theta_{\min} = \frac{1}{K_1} \cdot \sqrt[n_1]{\frac{1}{\frac{\mu_1}{2^{\frac{1}{n_2}} \cdot \left( \frac{(1+\mu_1)}{(1 + (1+\mu_1)^{n_2})^{\frac{1}{n_2}}} \right)^{-1} - 1} - 1}}, \tag{24}$$

$$\theta_{\max} = \frac{1}{K_1} \cdot \sqrt[n_1]{\frac{1}{\frac{\mu_1}{2^{-\frac{1}{n_2}} \cdot (1 + (1+\mu_1)^{n_2})^{\frac{1}{n_2}} - 1} - 1}}. \tag{25}$$

## 4 Parameter fitting

The dose-response parameters in Figures 1D and 3C of the main text were obtained from the promoter characterization data. For each induction curve, we computed the dynamic range ($\mu$) and basal expression level ($b$) from the data at zero and full induction with IPTG or TMG. The response threshold was determined from Hill functions fitted to the data using the `Solver` routine in MS Excel for Mac 2016.

The model fits in Figures 1D (green curve) and 3C (blue and yellow curves) were produced as explained in the Methods section. We used the optimizer `fmincon` from the Global Optimization Toolbox in Matlab 2016a. We ran the optimization 500 times with varying initial guesses determined by the `multistart` routine in Matlab 2016a. The convergence of the objective function, shown in Fig. SF2, suggests the optimal solutions are global minimizers in all three parameter fittings. In Table SF1, we report the parametric constraints employed and the ensemble average across all the runs of the optimization routine.

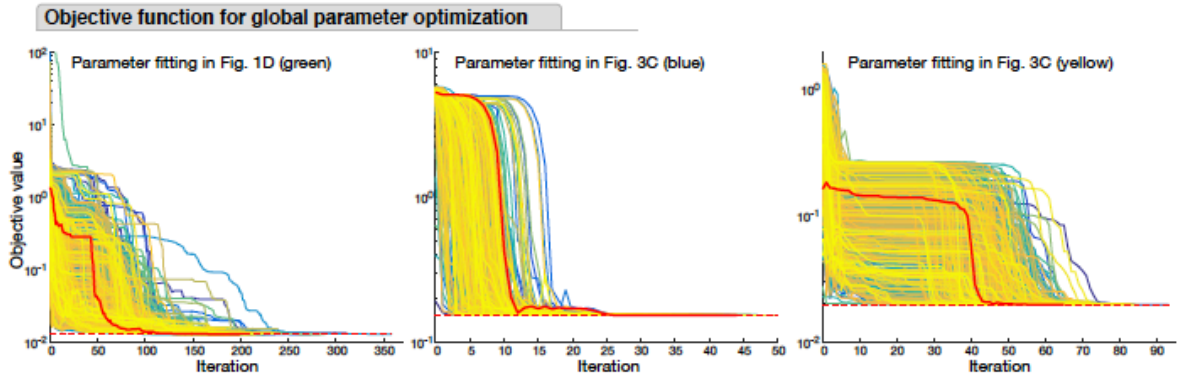



Figure SF 2: Evolution of objective function as a function of the iteration number of the optimization routine. The results show 500 runs of the optimization solver for each model in Fig. 1D and 3C in the main text, with varying initial conditions. The trace highlighted in red is of the convergence to the optimal solution with the lowest objective value, the corresponding parameters of which are reported in the last column of Table SF1. These parameters were used to generate the solid curves in Figs. 1D (green curve) and 3C (blue and yellow curves).

Table SF 1: Results of parameter fitting. Lower and upper bounds used for parameter optimization, plus their mean and standard error across 500 runs of the optimization algorithm. The last column are the parameter values corresponding to the optimal solution with the lowest objective value, and were used to generate the solid curves in Figs. 1D (green curve) and 3C (blue and yellow curves). All values are reported to 4 significant figures.

| Parameter | Lower bound | Upper bound | Mean ± SE | Reported |
|---|---|---|---|---|
| **Model in Figure 1D (green)** | | | | |
| $b_1$ | 0 | 0.1 | $3.339{\cdot}10^{-2} \pm 1.961{\cdot}10^{-3}$ | $2.588 \cdot 10^{-8}$ |
| $a_1$ | 0 | 500 | $348.4 \pm 6.448$ | 62.63 |
| $K_1$ | $10^{-4}$ | 1 | $1.804{\cdot}10^{-1} \pm 1.228{\cdot}10^{-3}$ | $2.112 \cdot 10^{-1}$ |
| $n_1$ | 1 | 5 | $2.436 \pm 2.132{\cdot}10^{-2}$ | 3.039 |
| $b_2$ | 0 | 20 | $1.231 \pm 5.216{\cdot}10^{-2}$ | $4.132 \cdot 10^{-6}$ |
| $a_2$ | 0 | $3 \cdot 10^3$ | $2149 \pm 3.933{\cdot}10^{-1}$ | 2147 |
| $n_2$ | 1 | 3 | $1.526 \pm 2.409{\cdot}10^{-2}$ | 1.008 |
| **Model in Figure 3C (blue)** | | | | |
| $b_1$ | 0 | 1 | $9.482{\cdot}10^{-2} \pm 8.352{\cdot}10^{-4}$ | $7.931{\cdot}10^{-2}$ |
| $K_1$ | $10^{-6}$ | 1 | $3.454{\cdot}10^{-3} \pm 4.001{\cdot}10^{-8}$ | $3.454{\cdot}10^{-3}$ |
| **Model in Figure 3C (yellow)** | | | | |
| $b_2$ | 0 | 5 | $4.065{\cdot}10^{-6} \pm 7.199{\cdot}10^{-7}$ | $3.110{\cdot}10^{-9}$ |
| $a_2$ | $10^3$ | $10^4$ | $3002 \pm 4.071{\cdot}10^{-1}$ | 3005 |
| $n_2$ | 1 | 3 | $1.189 \pm 2.535{\cdot}10^{-4}$ | 1.189 |

## 5 Biosensor architectures

Table SF 2: Some examples of engineered biosensors based on metabolite-responsive transcription factors. Each biosensor can be abstracted into one of the four architectures shown in Fig. 1A. [1]TyrR is a dual-regulator that has been used as an activator and repressor of gene expression in response to tyrosine. [2]BetI has been re-engineered to respond positively or negatively to choline.

| Architecture | Metabolite (M) | Transcription Factor (TF) |
|---|---|---|
| Repressed-Activator | Isopentenyl diphosphate | AraC-Idi [1] |
| | NADPH | SoxR [8] |
| Activated-Repressor | Tyrosine | TyrR[1] [1] |
| | Choline | BetI[2] [7] |
| Repressed-Repressor | Acyl-CoA | FadR [11] |
| | Choline | BetI[2] [7] |
| | IPTG | LacI [6] |
| | Malonyl-CoA | FapR [3] |
| Activated-Activator | Tyrosine | TyrR[1] [11] |
| | Mevalonate | AraC [9] |
| | Alkanes | AlkS [5] |
| | Benzoate | BenR [10] |
| | L-methionine | Lrp [4] |
| | o-acetyl-L-(homo)serine | CysR [2] |

## References


[1] Chou, H. H., and Keasling, J. D. Programming adaptive control to evolve increased metabolite production. *Nature Communications 4* (2013), 1–8.

[2] Hoffmann, K., Grünberger, A., Lausberg, F., Bott, M., and Eggeling, L. Visualization of imbalances in sulfur assimilation and synthesis of sulfur-containing amino acids at the single-cell level. *Applied and Environmental Microbiology 79*, 21 (2013), 6730–6736.

[3] Liu, D., Xiao, Y., Evans, B. S., and Zhang, F. Negative Feedback Regulation of Fatty Acid Production Based on a Malonyl-CoA SensorActuator. *ACS Synthetic Biology 4*, 2 (2013), 132–140.

[4] Mustafi, N., Grünberger, A., Kohlheyer, D., Bott, M., and Frunzke, J. The development and application of a single-cell biosensor for the detection of l-methionine and branched-chain amino acids. *Metabolic Engineering 14*, 4 (2012), 449–457.

[5] Reed, B., Blazeck, J., and Alper, H. Evolution of an alkane-inducible biosensor for increased responsiveness to short-chain alkanes. *Journal of Biotechnology 158*, 3 (2012), 75–79.

[6] Richards, D. H., Meyer, S., and Wilson, C. J. Fourteen Ways to Reroute Cooperative Communication in the Lactose Repressor: Engineering Regulatory Proteins with Alternate Repressive Functions. *ACS Synthetic Biology 6*, 1 (2017), 6–12.

[7] Saeki, K., Tominaga, M., Kawai-noma, S., Saito, K., and Umeno, D. The rapid diversification of BetI-based transcriptional switches for the 2 control of biosynthetic pathways and genetic circuits. 3. *ACS Synthetic Biology 5*, 11 (2015), 1201–1210.

[8] Siedler, S., Schendzielorz, G., Binder, S., Eggeling, L., Bringer, S., and Bott, M. SoxR as a single-cell biosensor for NADPH-consuming enzymes in Escherichia coli. *ACS Synthetic Biology 3*, 1 (2014), 41–47.

[9] Tang, S.-Y., and Cirino, P. C. Design and application of a mevalonate-responsive regulatory protein. *Angewandte Chemie Int. Ed Engl. 50*, 5 (2011), 1084–1086.

[10] Uchiyama, T., and Miyazaki, K. Product-induced gene expression, a product-responsive reporter assay used to screen metagenomic libraries for enzyme-encoding genes. *Applied and Environmental Microbiology 76*, 21 (2010), 7029–7035.

[11] Xiao, Y., Bowen, C. H., Liu, D., and Zhang, F. Exploiting non-genetic, cell-to-cell variation for enhanced biosynthesis. *Nature Chemical Biology 12*, 5 (2016), 339–344.

# Chapter 3 Negative Feedback Regulation of Fatty Acid Production Based on a Malonyl-CoA Sensor-Actuator

Note: This chapter contains text and figures from the published paper (Liu, Di, et al. "Negative feedback regulation of fatty acid production based on a malonyl-CoA sensor–actuator." *ACS synthetic biology* 4.2 (2014): 132-140).

## Abstract

Engineering metabolic biosynthetic pathways has enabled the microbial production of many useful chemicals. However, pathway productivities and yields are often limited by metabolic imbalances. Synthetic regulatory circuits have been shown to be able to balance engineered pathways, improving titers and productivities. Here we developed a negative feedback regulatory circuit based on a malonyl-CoA-based sensor-actuator. Malonyl-CoA is biosynthesized from acetyl-CoA by the acetyl-CoA carboxylase, which is the rate-limiting step for fatty acid biosynthesis. Overexpression of acetyl-CoA carboxylase improves fatty acid production, but slows down cell growth. We have devised a malonyl-CoA sensor-actuator that controls gene expression levels based on intracellular malonyl-CoA concentrations. This sensor-actuator was used to construct a negative feedback circuit to regulate the expression of acetyl-CoA carboxylase. The negative feedback circuit was able to turn up acetyl-CoA carboxylase expression when the malonyl-CoA concentration is low and turn down acetyl-CoA carboxylase expression when excess amounts of malonyl-CoA is accumulated. We have shown that the regulatory circuit effectively alleviated the

toxicity associated with acetyl-CoA carboxylase overexpression. When used to regulate the fatty acid pathway, the feedback circuit increased fatty acid titer and productivity by 34% and 33%, respectively.

## 3.1 Introduction

Engineered microbes have shown great potential as cell-factories for the production of fuels[1, 2], chemicals[3], pharmaceuticals[4, 5], nutraceuticals[6], and materials[7] etc. To make these technologies economically viable, it is important to obtain high yields and productivities. Common engineering strategies to improve yields and productivities include overexpressing bottleneck enzymes, bypassing native regulations, blocking competing pathways, genome-scale optimization[8-11], etc. However, due to various reasons, such as imbalanced enzyme activities, engineered pathways often have imbalanced metabolism. Under these circumstances, enzymes or intermediates can accumulate to toxic levels, which inhibit cell growth and decrease production[8, 12].

An engineered metabolic pathway can potentially be balanced by manipulating the intracellular concentration of pathway enzymes. Enzyme concentrations can be controlled (1) at the DNA level by tuning gene copy numbers[13], (2) at the transcription level by tuning the strength of promoters[14, 15], and (3) at the translation level by choosing a suitable ribosome binding site (RBS)[16].

These strategies allow static control of enzymes levels, which are mostly constant during chemical production.

In contrast to the static controls, natural metabolic pathways are dynamically regulated according to cell metabolic status[17-19]. Some pathways are regulated by positive feedback loops. As an example, seven out of eight genes in the *Saccharomyces cerevisiae* leucine biosynthesis pathway are positively regulated by the transcription factor Leu3, which is in turn activated by a pathway intermediate – alpha-isopropyl-malate (αIPM). As a result, leucine biosynthesis is positively controlled by the intermediate branching from the common precursor of branched-chain amino acid biosynthesis[20]. Similarly, negative feedback regulation that controls the concentration of certain metabolites is also found in nature[21]. For example, all the eight genes involved in *S. cerevisiae* arginine biosynthesis pathway are negatively regulated by cellular level of arginine through a transcription factor ArgR[17]. These regulation systems allow cells to adjust metabolite concentrations at desirable levels, balancing the pathway for optimal cell growth[22]. The balanced pathway also prevents the biosynthesis of unnecessary RNAs, proteins or metabolites, increasing the efficiency of energy and carbon usage[23]. Based on this concept, a synthetic regulatory system that detects key metabolic intermediates and regulates pathway gene expression according to intermediate concentrations could be used to balance its metabolism and improve product titers and yields[22, 24].

With the development of synthetic biology, several strategies now exist to create synthetic control circuits. One of the earliest metabolic regulatory circuits was constructed by harnessing an acetyl phosphate sensor to regulate lycopene biosynthesis[25]. In a recent study, a dynamic sensor-regulator system was built for an engineered pathway, in which expression levels of several heterologous genes were dynamically controlled by the concentration of a key metabolite, acyl-CoA[23]. In both systems the cells were able to dynamically modulate their metabolic flux in response to intracellular physiology, and therefore dramatically improved production. Besides these experimental efforts, *in-silico* simulations also predicted that dynamic pathway regulation could improve production[26-28].

Here we report the development of a malonyl-CoA-based negative feedback system in *Escherichia coli* to improve fatty acid titer and productivity. Malonyl-CoA is a common building block for the biosynthesis of several types of compounds, including fatty acids, 3-hydroxypropionic acid, polyketides, and flavonoids[29, 30]. These compounds can be used as or converted to biofuels, commodity chemicals, fine chemicals, and drugs. Bioproduction of fatty acids is particularly important because fatty acid derivatives provide renewable transportation fuels[31, 32]. In *E. coli*, malonyl-CoA is biosynthesized from acetyl-CoA by acetyl-CoA carboxylase encoded by *accABCD* (*acc*). The *E. coli* acetyl-CoA carboxylase consists of four subunits: a biotin carboxyl carrier protein, a biotin carboxylase, and two carboxyltransferase subunits. This is the first step of fatty acid biosynthesis and is believed to be the rate-limiting step[33, 34]. Several studies

have shown that overexpression of *acc* genes improved both fatty acid and flavonoid production[35-38]. On the other hand, overexpression of *acc* is toxic to cells[34, 36]. The exact molecular mechanism of the toxicity is currently unclear. Possible reasons include the depletion of free CoA pool by malonyl-CoA overproduction, or the interference with biotin-utilizing enzymes by the imbalanced expression of the biotin carboxylase subunit[34]. To enhance malonyl-CoA supply while alleviating the toxicity caused by *acc* overexpression, we have devised a malonyl-CoA sensor-actuator and used it to negatively regulate *acc* expression. We have shown that the malonyl-CoA-based negative control system can alleviate the toxicity caused by *acc* overexpression. When the system was applied to fatty acid pathway, it improved both fatty acid titer and productivity by 34% and 33%, respectively.

## 3.2 Results and Discussions

### 3.2.1 Design of a malonyl-CoA sensor-actuator

We designed the malonyl-CoA sensor-actuator based on a naturally-occurring malonyl-CoA-responsive transcription factor, FapR, from the Gram-positive bacteria *Bacillus subtilis*[39]. FapR specifically binds to a 17-bp DNA sequence and negatively regulates fatty acid and phospholipid metabolism in *B. subtilis*[40]. The binding of malonyl-CoA to FapR triggers a conformation change to the FapR, causing FapR-DNA complex to dissociate[39]. To our knowledge, such a malonyl-CoA-responsive transcription factor is absent in *E. coli*. To create malonyl-CoA biosensors in *E. coli*,

we cloned the *B. subtilis fapR* gene into *E. coli* using a low copy number plasmid under the control of a $P_{BAD}$ promoter (pA8c-fapR, Table 3.1). Meanwhile, we constructed a FapR-regulated synthetic promoter ($P_{FR1}$, the actuator) by inserting the 17-bp FapR-binding sequence into two regions flanking the -10 region of a phage $P_{A1}$ promoter[41] (Table 3.3). The actuator was later cloned to control the expression of a red fluorescence protein gene (*rfp*), resulting in plasmid pBFR1k-rfp. In the absence of malonyl-CoA, FapR is expected to bind to the 17-bp DNA sequence, which poses steric hindrance to RNA polymerase binding and inhibits RFP transcription. When malonyl-CoA is present, the binding of malonyl-CoA to FapR is expected to release FapR from the promoter ($P_{FR1}$), enhancing RFP transcription (Figure 3.1).

**Table 3.1** Plasmids used in this study

| Plasmids | Replication Origin | Overexpressed Operon | Resistance | References |
|---|---|---|---|---|
| pBFR1k-RFP | BBR1 | $P_{FR1}$-*rfp* | $Kan^R$ | This study |
| pA8c-FapR | p15A | $P_{BAD}$-*fapR* (*B. subtilis*) | $Cm^R$ | This study |
| pE7a-acc | ColE1 | $P_{T7}$-*acc* (*E. coli*) | $Amp^R$ | This study |
| pE7a-yACC | ColE1 | $P_{T7}$-*acc* (*S. cerevisiae*) | $Amp^R$ | This study |
| pBFR#k-RFP (#: 2-6) | BBR1 | $P_{FR\#}$-*rfp* | $Kan^R$ | This study |
| pA8c-0 | p15A | $P_{BAD}$-*none* | $Amp^R$ | This study |
| pE7a-0 | ColE1 | $P_{T7}$-*none* | $Amp^R$ | This study |

| | | | | |
|---|---|---|---|---|
| pBFR1k-lacI-8FapR | BBR1 | $P_{FR1}$-*lacI*, $P_{BAD}$-*fapR* | $Kan^R$ | This study |
| pBFR1k-lacI-8MFapR | BBR1 | $P_{FR1}$-*lacI*, $P_{BAD}$-*mfapR* | $Kan^R$ | This study |
| pBFR1k-RFP-8FapR | BBR1 | $P_{FR1}$-*rfp*,$P_{BAD}$-*fapR* | $Kan^R$ | This study |
| pBFR1k-RFP-8MFapR | BBR1 | $P_{FR1}$-*rfp*,$P_{BAD}$-*mfapR* | $Kan^R$ | This study |
| pA2c-tesA | p15A | $P_{tet}$-*'tesA* (*E. coli*) | $Cm^R$ | This study |

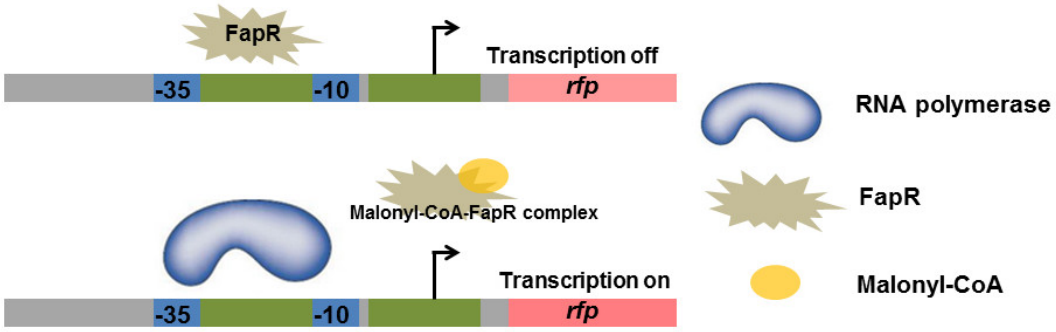



**Figure 3.1** Malonyl-CoA sensor-actuator design. The FapR-binding sites are inserted flanking the -10 region of the promoter (colored green). The presence of malonyl-CoA antagonizes the DNA binding activity of FapR and releases FapR from the engineered promoter, initiating RFP transcription.

To evaluate the malonyl-CoA sensor-actuator, we varied the intracellular malonyl-CoA concentrations using a malonyl-CoA-accumulating plasmid, pE7a-acc. Specifically, pE7a-acc contained an additional copy of the *E. coli acc* genes under the control of two T7 promoters so that

the intracellular malonyl-CoA concentration could be varied by titrating the media with IPTG (Figure 3.2a). We co-transformed pE7a-acc with the sensor-actuator plasmids, pA8c-FapR and pBFR1k-RFP, and cultivated the strain (SR1, Table 3.2) in LB medium. Cell culture fluorescence (normalized by cell density) at various IPTG concentrations was measured. As expected, cell culture fluorescence was increased by 4-fold with increasing IPTG concentrations (Figure 3.2b).

**Table 3.2** Strains used in this study.

| Strains | Relevant genotype | References |
|---|---|---|
| SAcc1 | *E. coli* BL21 (DE3): pE7a-acc | This study |
| SAcc2 | *E. coli* BL21 (DE3): pE7a-yACC | This study |
| SR1 | *E. coli* BL21 (DE3): pBFR1k-RFP, pA8c-FapR, pE7a-acc | This study |
| SR2 | *E. coli* BL21 (DE3): pBFR2k-RFP, pA8c-FapR, pE7a-acc | This study |
| SR3 | *E. coli* BL21 (DE3): pBFR3k-RFP, pA8c-FapR, pE7a-acc | This study |
| SR4 | *E. coli* BL21 (DE3): pBFR4k-RFP, pA8c-FapR, pE7a-acc | This study |
| SR5 | *E. coli* BL21 (DE3): pBFR5k-RFP, pA8c-FapR, pE7a-acc | This study |
| SRCtl1 | *E. coli* BL21 (DE3): pBFR1k-RFP, pA8c-FapR, pE7a-0 | This study |
| SRCtl2 | *E. coli* BL21 (DE3): pBFR1k-RFP, pA8c-0, pE7a-acc | This study |
| SRCtl3 | *E. coli* BL21 (DE3): pBFR6k-RFP, pA8c-FapR, pE7a-acc | This study |
| FA1 | *E. coli* BL21 (DE3): pA2c-tesA | This study |
| FA2 | *E. coli* BL21 (DE3): pBFR1k-RFP-8FapR, pE7a-acc, pA2c-tesA | This study |
| FA3 | *E. coli* BL21 (DE3): pBFR1k-lacI-8FapR, pE7a-acc, pA2c-tesA | This study |
| MFA2 | *E. coli* BL21 (DE3): pBFR1k-RFP-8MFapR, pE7a-acc, pA2c-tesA | This study |
| MFA3 | *E. coli* BL21 (DE3): pBFR1k-lacI-8MFapR, pE7a-acc, pA2c-tesA | This study |

We next verified that the sensor-actuator worked according to the mechanism we designed by individually removing Acc, FapR, and the FapR operator site (FapRO) from the system, yielding three control strains (SRCtl1, SRCtl2, SRCtl3, Table 3.2). The cell culture fluorescence of these strains was measured with or without induction for malonyl-CoA accumulation (Figure 3.2c). In the strain lacking the Acc plasmid (SRCtl1, see Methods) , cell culture fluorescence remained at basal levels, regardless of the inducer, indicating that malonyl-CoA synthesized from the genomic enzymes was not sufficient to turn on the sensor. In fact, fluorescence of strain SRCtl1 (normalized fluorescence 13.3 ± 0.8 a.u.) was lower than that of strain SR1 under the non-inducible condition (0 IPTG, normalized fluorescence 177.8 ± 0.3 a.u.), suggesting that in the latter case, there was leaky expression of *acc* from the T7 promoter, causing a slight increase in malonyl-CoA concentration. When either FapR or FapRO was removed, the promoter constantly remained at high expression levels, confirming our design that the promoter was repressed by FapR at the FapRO sites. As compared with these two non-repressive strains (normalized fluorescence ~4500 a.u.), the SR1 strain exhibited a maximal fluorescence of 743 ± 0.5 a.u. (with 1 mM IPTG induction), indicating that the malonyl-CoA accumulated under this condition in LB medium was not sufficient to fully turn on the promoter. Overall, we have shown that the malonyl-CoA sensor-actuator responded specifically to cellular malonyl-CoA concentrations through FapR and FapRO.

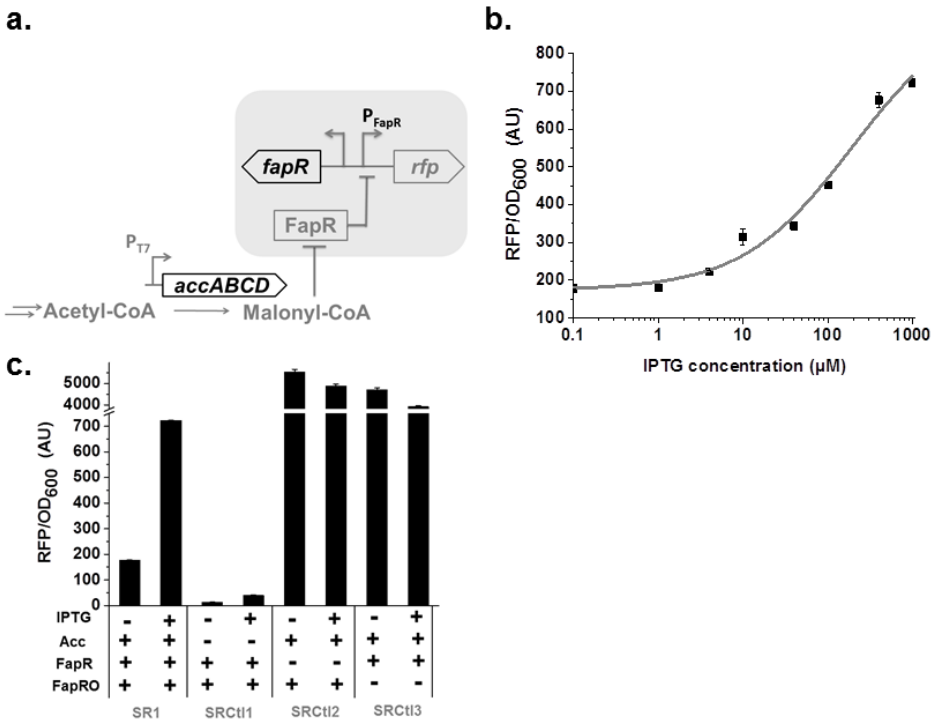



**Figure 3.2** Characterization of malonyl-CoA sensor-actuator. a) The malonyl-CoA sensor-actuator consists of a constantly expressed FapR, an engineered promoter-reporter system, and the *acc* under $P_{T7}$. b) Response of malonyl-CoA sensor-actuator to IPTG. IPTG concentration was increased from 0 μM (plotted as 0.1 μM) to 1 mM to induce *acc* expression to accumulate varying cellular malonyl-CoA levels. Cell culture fluorescence was measured after 24hr and normalized by $OD_{600}$. c) Response of strains with individual *acc*, FapR, FapRO knockout. Strains with FapR/FapRO knockout exhibited high RFP expression levels, demonstrating the sensor-actuator was repressed by FapR at the FapRO site.

Next we tested the behavior of the sensor-actuator in M9 medium containing 2% glucose, the medium usually used for chemical production. This would allow us to learn the behavior of the sensor-actuator for pathway regulation during production. Due to catabolic repression of the $P_{BAD}$ promoter in the glucose-rich M9 medium, less FapR was expressed compared to that in LB medium, leading to enhanced fluorescence (Figure 3.3a). To obtain optimal sensor-actuator behavior, we tuned the FapR expression level by titrating with various amounts of arabinose and examined the responses of the sensor-actuator strain. When a high arabinose concentration was used (0.1%), a large amount of FapR was expressed, leading to no malonyl-CoA activation even with 1 mM of IPTG induction (Figure 3.3a). By contrast, strains with low arabinose concentration produced an insufficient amount of FapR, resulting in an *rfp* expression that was not responsive to malonyl-CoA. The desired malonyl-CoA response was obtained with the addition of 0.01% arabinose, indicating that the amount of FapR produced under this condition was optimal for regulation. Although the current strain requires the addition of 0.01% arabinose to achieve an optimal FapR level in the M9/glucose medium, the need for inducer can be potentially eliminated by using a constitutive promoter with the proper strength to drive FapR expression.

To quantitatively evaluate the sensor-actuator, we used LC-MS to measure the accumulated malonyl-CoA concentrations under various induction levels in M9/glucose medium. With the plasmid pE7a-acc, cellular malonyl-CoA concentrations of strain SAcc1 (Table 3.2) increased from 0.14 ± 0.1 μM to 24.4 ± 0.9 μM when IPTG concentration was increased from 0 to 1 mM

(Figure 3.3b). This data allowed us to obtain a fluorescence/malonyl-CoA concentration curve, assuming strains SAcc1 and SR1 accumulated the same amount of malonyl-CoA under the same induction conditions (Figure 3.3c). Next, we fitted the data to a thermodynamic model modified from previous studies involved in inducible systems (Supplementary material)[42]. Although we had only a few experimental data points due to the difficulties in accumulating malonyl-CoA to desirable levels and the challenges associated with malonyl-CoA quantification, our fitted binding constant for FapR-malonyl-CoA interaction, 6.3 µM, was in the same order of magnitude to the value determined by *in vitro* isothermal titration calorimetry (2.4 µM)[39].

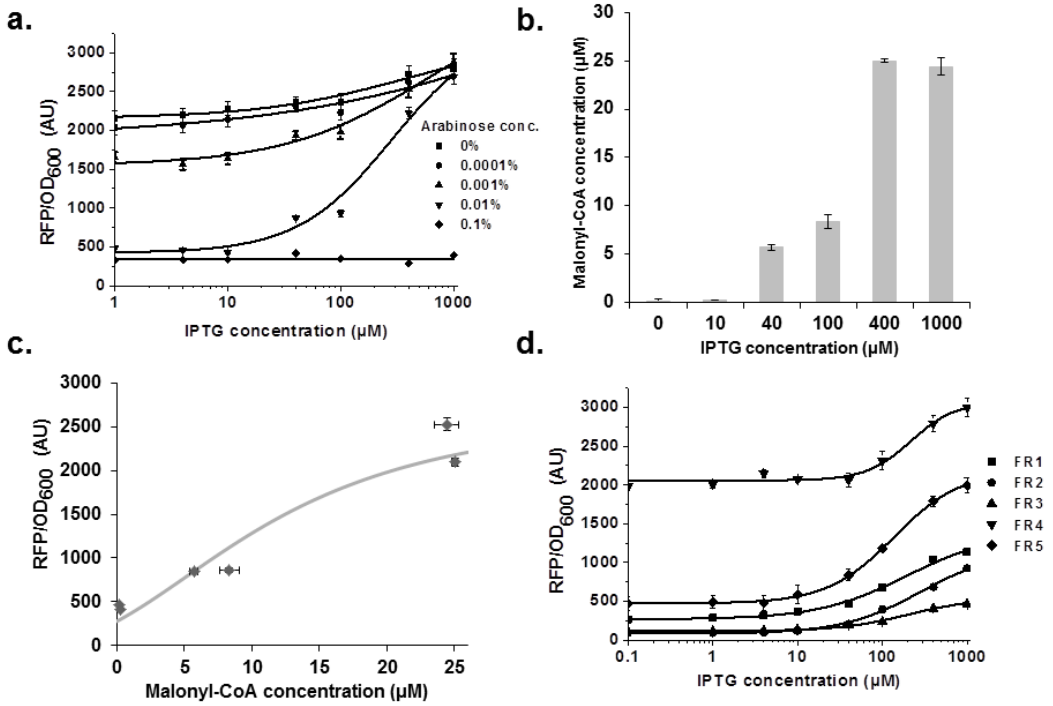



**Figure 3.3**. Characterization of malonyl-CoA sensor-actuator in minimal medium containing 2% glucose and its tuning to expand the output range. a) Tuning of sensor-actuator in minimal medium

by titrating arabinose (0% to 0.1%) to change the amount of FapR. IPTG was increased from 0 (plotted as 1 μM) to 1mM to vary cellular malonyl-CoA concentrations. At a high arabinose level, excess FapR fully repressed the PFR1, leading to no malonyl-CoA activation. At low arabinose levels, an insufficient amount of FapR resulted in high sensor-actuator background levels, and subjected it non-responsive to malonyl-CoA. The optimal sensor-actuator behavior was obtained with 0.01% arabinose. b) HPLC-MS quantification of malonyl-CoA at various acc induction levels. c) Response of the sensor-actuator to malonyl-CoA in minimal medium. The results were fitted to a thermodynamic model. The fitted Kd for malonyl-CoA - FapR binding was 6.3 μM. d) Tuning of malonyl-CoA sensor-actuator in LB medium. Mutations were introduced to the FapR binding site and the -10 and -35 region of the promoter $P_{FR1}$. The sensor-actuator variants exhibited a broad range of expression levels.

We also sought to expand the capability of the malonyl-CoA sensor-actuator by tuning its output range. Based on the thermodynamic model, the behavior of the sensor-actuator is determined by the interaction between the promoter with FapR and the RNA polymerase. Both of these interactions can be tuned by changing the promoter sequence. We modified the $P_{FR1}$ by introducing mutations to the FapR binding site and to the -35 and -10 region of the promoter, leading to a series of $P_{FR1}$ variants ($P_{FR2}$ - $P_{FR5}$, Table 3.3). We cloned the *rfp* 3' of each promoter and characterized them (Strains SR2 – SR5, Table 3.2) using cell culture fluorescence. These promoters positively responded to increasing *acc* induction levels and generated different RFP intensities, exhibiting

varied strengths and dynamic ranges (Figure 3.3d). Overall, these promoters allow the expression levels of the malonyl-CoA-regulated genes to be tuned across a broad range.

In general, the fluorescence based malonyl-CoA sensor-actuator provides a comparative method for measuring intracellular malonyl-CoA concentrations. It avoids the labor-intense extraction procedures using analytical quantification and the experimental errors caused by hydrolysis during sample preparation. Furthermore, the fluorescence-based measurement is quick and easy, allowing high throughput analysis of a large number of samples. More importantly, the malonyl-CoA-based sensor-actuators can be used to regulate malonyl-CoA-involved metabolic pathways.

### 3.2.2 Design of a negative feedback circuit to regulate an engineered free fatty acid biosynthetic pathway

It has been previously shown in several studies that efforts to enhance cellular malonyl-CoA pool by overexpressing acetyl-CoA carboxylase led to reduced cell growth due to toxicity[36, 43]. To evaluate the toxicity, we overexpressed acetyl-CoA carboxylases from both *E. coli* and *S. cerevisiae* to different expression levels. Cell growth was measured, and the stationary phase cell density ($OD_{600}$ at 20 hours after IPTG induction) of different cultures was plotted against the different induction levels (Supplementary Figure 3S1). Our data showed that overexpression of both acetyl-CoA carboxylase enzymes were toxic. We also evaluated the effects of *acc* overexpression on fatty acid production. Free fatty acids can be produced in *E. coli* by the

overexpression of a cytosolic thioesterase (encoded by a leader sequence deleted *tesA*), which hydrolyzes acyl-ACPs and releases free fatty acids. Previous expression of *tesA* in a *fadE* gene knockout strain led to the production of fatty acid at 1.2 g/L after three days cultivation in M9 medium containing 2% glucose[44]. To control the expression of *tesA* and *acc* separately, we placed the *tesA* under the control of an aTc-inducible promoter $P_{Tet}$. As compared to the strain that only overexpressed the thioesterase (strain FA1), overexpression of *tesA* and *acc* (strain FA2) increased the fatty acid titer by 84% (Figure 3.4b), although it grew slower (Figure 4c). Both the increased free fatty acid titer and decreased cell growth are consistent with previous studies[29, 34, 45].

Next we sought to develop a negative feedback circuit to tightly control the expression of acetyl-CoA carboxylase by sensing the cellular malonyl-CoA concentration. Ideally, we need the system to express *acc* when the malonyl-CoA concentration is low and to turn the *acc* expression down when the malonyl-CoA concentration is too high. Since the sensor-actuator positively responds to malonyl-CoA (higher malonyl-CoA level leads to higher expression), an invertor was needed to complete a negative feedback circuit. To complete the circuit, we placed the *acc* under the control of a LacI-repressive T7 promoter, $P_{T7}$, and placed the *lacI* under the control of $P_{FR1}$ (Figure 3.4a; Strain FA3, Table 3.2). With this design, IPTG induction initiates *acc* expression, which produces malonyl-CoA. When excess malonyl-CoA is accumulated in this strain (FA3), the malonyl-CoA sensor-actuator will turn on *lacI* expression, which in turn down-regulates *acc*, decreasing the

malonyl-CoA synthesis rate. The control strain (FA2) has the *lacI* replaced by the *rfp*, so that its *acc* is controlled by $P_{T7}$, not the malonyl-CoA-regulated promoter.

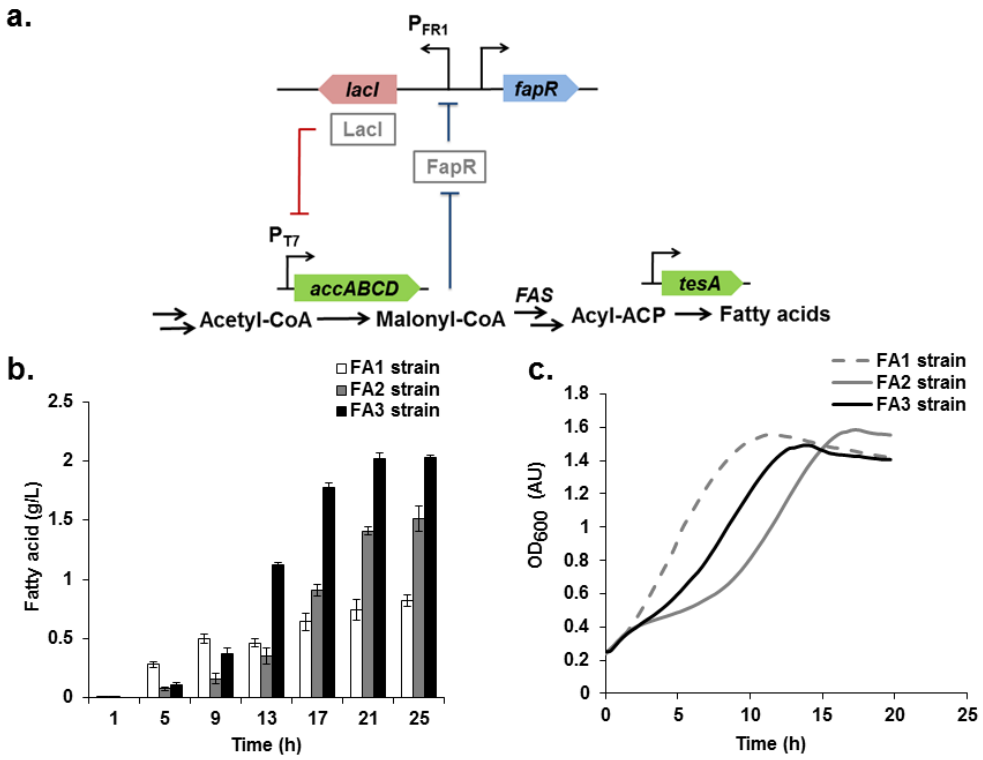



**Figure 3.4** Construction of a negative feedback regulatory circuit and its effects on improving fatty acid titer and cell growth. a) The negative feedback circuit regulated fatty acid-producing strain. A cytoplasmic thioesterase (encoded by *tesA*) was controlled by $P_{Tet}$ to produce fatty acids. The *acc* genes were controlled by a LacI-repressed $P_{T7}$ promoter. LacI expression is controlled by the $P_{FR1}$ that is repressed by FapR, which is further controlled by malonyl-CoA. When excess malonyl-CoA is accumulated, the biosensor will turn on the expression of lacI, which down-regulates *acc* expression, alleviating toxicity caused by *acc* overexpression. b) Time course analysis of fatty acid

titers from strain FA1 (white column), FA2 (grey column), and FA3 (black column). Strains were cultured in minimal medium with 2% glucose at 37°C in a shaking flask. Fatty acid production titers were analyzed as described in Methods. c) Cell growth of strains FA1 (dotted grey curve), FA2 (solid grey curve), and FA3 (black curve) were monitored in a plate reader until the stationary phase.

We examined whether the negative feedback regulation circuit could alleviate the toxicity caused by *acc* overexpression. We monitored cell growth under fatty acid production conditions. FA2 and FA3 strains were cultured in minimal medium and induced for both fatty acid production and *acc* expression (using various amounts of IPTG). Cell growth was monitored continuously on the plate reader until stationary phase was reached (Figure 3.5a and 3.5b). Compared to the FA2 strain, the FA3 strain had less growth inhibition, consistent with our design that the negative regulation circuit was able to control the *acc* expression level and to alleviate the toxicity, resulting in improved cell growth (Supplementary Figure 3S2). Cell growth inhibition was observed at an IPTG concentration higher than 40 μM, which corresponded to 5.7 ± 0.3 μM of malonyl-CoA. It is important to note that this malonyl-CoA concentration matches the fitted $K_d$ (6.3μM) of the sensor-regulator, which suggests that the designed sensor-actuator could respond to malonyl-CoA at a metabolically relevant concentration.

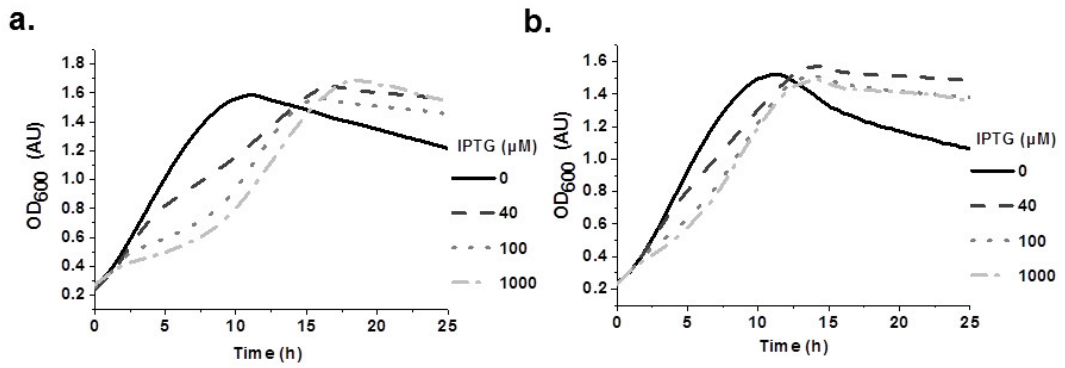



**Figure 3.5.** Effects of negative feedback regulatory circuit to alleviate toxicity caused by *acc* overexpression. a) Cell growth was monitored for fatty acid-producing strains with (b, FA3 strain) and without (a, FA2 strain) the regulatory circuit under different *acc* induction levels. Cells were cultivated in minimal medium containing 2% glucose. Growth curves were monitored on a plate reader (see Methods).

At last, we tested the effect of the negative feedback regulation circuit on fatty acid production. Fatty acid titers were measured at several time points during the exponential and early stationary phases (the first 25 hours). As shown above, overexpression of *acc* from an inducible promoter increased fatty acid titer (strain FA2 compared to strain FA1). When the negative regulation circuit was used, the FA3 strain further increased the fatty acid titer (Figure 3.4b). After 25 hours, the FA3 strain produced 2.03 ± 0.1 g/L of fatty acids, being 34% higher than that of the FA2 strain.

Furthermore, the FA3 strain had a fatty acid productivity of 4.2 ± 0.2 g/(L•d), being 33% higher than that of the FA2 strain, which was 3.2 ± 0.1 g/(L•d). Due to the alleviated toxicity effect, the negative-regulated strain entered the exponential growth phase earlier. This is particularly important for industrial applications, as minimizing fermentation time and maximizing productivity are pursued. Overall, our data indicates that the negative regulatory system can not only alleviate the toxicity that inhibits cell growth, but also improve fatty acid titer and productivity.

To verify that the improved fatty acid titer was caused by the negative feedback circuit, we replaced the FapR of the FA2 and FA3 strains with a FapR double mutant (R106A and G107V), resulting in two additional control strains MFA2 and MFA3 (Table 3.2). Previous study has shown that the FapR double mutant could bind to the FapRO site with the same affinity as that of the wild-type FapR, but could not interact with malonyl-CoA effectively[46]. Thus, the resulting control strain, MFA3, only differs from FA3 by two amino acids in FapR, but does not have the feedback regulation due to the disrupted malonyl-CoA/FapR interaction. As a result, LacI expression is expected to be constantly repressed in MFA3 strain, leading to dramatic *acc* overexpression when induced with IPTG. Indeed, the reduced cell growth inhibition observed in FA3 strain was not observed in MFA3 strain (Supplementary Figure 3S3). Furthermore, fatty acid titers of both MFA2 and MFA3 strains were similar to that of FA2 strain, but less than that of the negative feedback regulated FA3 strain (Supplementary Figure 3S4). Overall, our result demonstrated that strains

with disrupted feedback circuit could neither reduce cell growth inhibition nor enhance fatty acid titer.

To further verify that the improved fatty acid titer is caused by the negative feedback regulation, rather than use of different genetic constructs, we titrated the *acc* expression levels of the FA2 strain that used the inducible system with a broad range of IPTG concentrations (from 0 to 1mM, Supplementary Figure S4). Under all conditions, the maximal fatty acid titer produced by the inducible system was 1.59 ± 0.21 g/L, being 22% lower than that of the negative regulation strain, FA3. In addition, further optimization of the negative feedback circuit, especially by tuning the concentration of the LacI repressor using the $P_{FR1}$ variants ($P_{FR2}$ – $P_{FR5}$), might further increase titers and productivities.

Overall, we have demonstrated that the negative feedback circuit has alleviated growth inhibition caused by either acetyl-CoA carboxylase overexpression or malonyl-CoA accumulation, improving fatty acid titers and productivities. This method can be readily extended to the production of other chemicals that use malonyl-CoA as a precursor, such as flavonoids and polyketides, to improve cell growth and enhance productivities. Similar sensor-actuators can be designed to sense other critical cell metabolites. At last, the negative feedback circuit design strategy can be implemented to other systems, alleviating the toxicity caused by protein

overexpression or metabolite intermediate accumulation, to enhance the titer and productivity of various chemicals.

# 3.3 Methods

## 3.3.1 Materials

Phusion DNA polymerase was purchased from New England Biolabs (Beverly, MA, USA). Restriction enzymes, T4 ligase, gel purification kit and plasmid miniprep kit were purchased from Thermo Scientific (Waltham, Massachusetts, USA). All primers were synthesized by Integrated DNA Technologies™ (Coralville, IA, USA). All reagents were purchased from Sigma Aldrich (St. Louis, MO, USA). *E. coli* DH10B was used for cloning purposes, and *E. coli* BL21 (DE3) was used for fluorescence characterization and fatty acid production.

## 3.3.2 Plasmids and strains

Plasmid pBFR1k-RFP was constructed from phage $P_{A1}$ promoter by placing two FapRO sites flanking the -10 region of the promoter using a one-step Golden-Gate DNA assembly method[47]. Acc overexpression plasmid pE7a-acc was constructed by cloning *accDA* and *accBC* from *E.coli* genome 3' of two separate T7 promoters. Gene *fapR* was amplified from *B. subtilis* genomic DNA and cloned 3' of the $P_{BAD}$ promoter in a BioBrick plasmid pBbA8c-RFP[48], yielding pA8c-FapR. Plasmids pE7a-0 and pA8c-0 were constructed from pE7a-acc and pA8c-FapR by restriction

digestion using BamHI/BglII to remove the corresponding genes. The empty vectors were later purified and ligated. To create fatty acid producing strain, a cytosolic thioesterase gene *tesA* (*'tesA*; leader sequence deleted) was cloned under the control of $P_{Tet}$, giving pA2c-tesA. To create pBFR1k-RFP-8FapR, the $P_{BAD}$-FapR operon in pA8c-FapR was amplified and inserted to pBFR1k-RFP at 5' of the $P_{FR1}$-RFP operon using the Golden-Gate assembly method. To create pBFR1k-lacI-8FapR, *lacI* was first amplified from the *E. coli* genome to construct plasmid pBFR1k-lacI. The $P_{BAD}$-FapR operon from pA8c-FapR was then amplified and inserted to pBFR1k-lacI at 5' of the $P_{FR1}$-lacI operon. Plasmids pBFR1k-RFP-8MFapR and pBFR1k-lacI-8MFapR were constructed by site-directed mutagenesis (R106A and G107V) of pBFR1k-RFP-8FapR and pBFR1k-lacI-8FapR. Strains were created by transforming the corresponding plasmids into BL21 (DE3) competent cells (Table 2) by electroporation.

### 3.3.3 Cell growth and fluorescence assay

Cell growth curves and cell culture fluorescence were recorded on an Infinite F200PRO (TECAN) plate reader. Strains were first cultivated overnight in Luria−Bertani (LB) medium (220 rpm, 37 °C) supplemented with appropriate antibiotics (50 mg/L ampicillin, 50 mg/L kanamycin, and 30 mg/L chloramphenicol). The overnight LB cultures were inoculated 2% v/v into fresh LB medium. Cells were induced at $OD_{600}$ of 0.6 with varied IPTG concentrations (0, 1, 4, 10, 40, 100, 400, 1000 μM). Cells were incubated in a 96-well plate inside the plate reader with shaking (218.3 rpm, 37 °C).

Relative cell density (in arbitrary units) was measured by monitoring absorption at 600 nm, and fluorescence was recorded using an excitation wavelength of 535 ± 9 nm and an emission wavelength of 620 ± 20 nm. Data were taken every 1000 s until cell culture reached stationary phase. Fluorescence from the wild-type *E. coli* BL21 (DE3) cell culture was used as background, and was subtracted from all fluorescence measurements. The background corrected fluorescence was later normalized by $OD_{600}$ and reported.

For measurement in minimal medium, the LB overnight culture was used to inoculate minimal medium (M9 medium supplemented with 75 mM MOPS, 2 mM $MgSO_4$, 1 mg/l thiamine, 10 nM $FeSO_4$, 0.1 mM $CaCl_2$ and micronutrients including 3 μM $(NH_4)_6Mo_7O_{24}$, 0.4 mM boric acid, 30 μM $CoCl_2$, 15 μM $CuSO_4$, 80 μM $MnCl_2$ and 10 μM $ZnSO_4$) with 2% glucose and appropriate antibiotics and incubated at 37 °C for overnight. The overnight culture in minimal medium was then used to inoculate a fresh minimal medium with an initial OD of 0.08, and grown to an $OD_{600}$ of 0.6 for induction. Cells were induced with varied amounts of arabinose (0%, 0.0001%, 0.001%, 0.01%, 0.1%) and IPTG (0, 4, 10, 40, 100, 400, 1000 μM) and incubated in a 96-well plate inside the plate reader. Data were recorded using the same method as described above.

### 3.3.4 Malonyl-CoA quantification

To quantify cellular malonyl-CoA levels, overnight culture of strain SAcc1 was used to inoculate 5 mL of fresh minimal medium with 2% glucose and appropriate antibiotics for adaptation. The overnight culture was then inoculated into minimal medium with an initial $OD_{600}$ of 0.08. The cultures were induced with 0, 10, 40, 100, 400, 1000 μM of IPTG when $OD_{600}$ reached 0.6. After 10 hours, 2 mL cultures were rapidly collected and centrifuged at 14,000 rpm for 30 s at 4 °C. The supernatant was immediately removed. The pellets were flash frozen using liquid nitrogen and stored at -80 °C until LC-MS analysis. To quantify malonyl-CoA concentrations, each sample was extracted with 100 μL of a solution containing 90% acetonitrile (ACN), 10% formic acid with 1 μM $^{13}C_3$ malonyl-CoA and zirconia beads by vortexing for 3 minutes. The insoluble material was removed by centrifugation; then the supernatant was filtered before LC-MS. Eight microliters were injected on a 2.1 x 50 mm Onyx C18 column (Phenomenex). The metabolites were separated using a linear gradient from 100% A (10 mM $NH_4HCO_3$) to 75% B (90% ACN 10 mM $NH_4HCO_3$) over 14 minutes followed by re-equilibration at initial conditions for 6 minutes using a 1200 LC system (Agilent). Mass spectra were recorded using a wide SIM scan (*m*/*z* 845-865) in a negative profile mode at a resolution of 60,000 on an LTQ-Velos Pro Orbitrap (Thermo-Fisher Scientific). Data were analyzed manually using the QualBrowser application of Xcalibur (Thermo-Fisher Scientific). Chromatograms were extracted for both the natural abundance malonyl-CoA and the $^{13}C_3$ malonyl-CoA. The amount of natural abundance malonyl-CoA was determined by calculating the ratio of the peak area of the natural abundance peak to the peak area of the $^{13}C_3$ peak.

### 3.3.5 Fatty acid production analysis

The FA1, FA2 and FA3 strains were inoculated into LB medium with appropriate antibiotics. The overnight culture was inoculated 2% v/v into minimal medium containing appropriate antibiotics for adaptation. The overnight cultures in minimal medium were used to inoculate 25mL of fresh minimal medium with an initial $OD_{600}$ of 0.08. Cells were induced when $OD_{600}$ reached 0.6. The large culture volume ensured that a lot of culture remained at the end, thus avoiding inaccuracies caused by evaporation and different oxygen-transfer rates. All strains were induced with 200 nM of aTc for thioesterase expression. Strains FA2 and FA3 were induced with 0.01% arabinose and 0.1mM IPTG. Samples at several time points were collected, and cell growth was monitored using the above-mentioned method. The collected samples were stored at -80 °C until fatty acid quantification.

### 3.3.6 Fatty acid quantification

Fatty acid titers were quantified using a previously published method[23]. Specifically, 0.5 ml of cell culture was acidified with 50 µl of concentrated HCl (6 N). The fatty acids were extracted twice with 0.5 ml ethyl acetate, which was spiked with 50 µg/ml of $C_{19:0}$ fatty acid methyl ester as an internal standard. The extracted fatty acids were derivatized to fatty acid methyl esters (FAME) by adding 10 µl concentrated HCl, 90 µl methanol and 100 µl of TMS-diazomethane, and incubated at room temperature for 15 min. FAME was then analyzed by a GC-MS (Hewlett Packard model 7890A, Agilent Technologies) equipped with a DB5-MS column (J&W Scientific) and a mass

spectrometer (5975C, Agilent Technologies). For each sample, the column was equilibrated at 80 °C for 1 min, followed by a ramp to 280 °C at 30 °C/min, and was then held at this temperature for 3 min. Final FAME concentration was analyzed based on the FAME standard curve obtained from standard FAME mix (GLC-20 & GLC-30, Sigma Aldrich).

# Acknowledgements

This work was supported by Washington University in St. Louis, the International Center for Advanced Renewable Energy and Sustainability (I-CARES), and the Defense Advanced Research Projects Agency (D13AP00038). The authors thank Keasling lab for providing the Biobrick plasmids, the *E. coli* thioesterase, and the *S. cerevisiae* acetyl-CoA carboxylase genes.

# References


1. Peralta-Yahya, P. P., Zhang, F., del Cardayre, S. B., and Keasling, J. D. (2012) Microbial engineering for the production of advanced biofuels, *Nature 488*, 320-328.
2. Gronenberg, L. S., Marcheschi, R. J., and Liao, J. C. (2013) Next generation biofuel engineering in prokaryotes, *Current opinion in chemical biology 17*, 462-471.
3. Curran, K. A., and Alper, H. S. (2012) Expanding the chemical palate of cells by combining systems biology and metabolic engineering, *Metabolic engineering 14*, 289-297.
4. Ajikumar, P. K., Xiao, W. H., Tyo, K. E., Wang, Y., Simeon, F., Leonard, E., Mucha, O., Phon, T. H., Pfeifer, B., and Stephanopoulos, G. (2010) Isoprenoid pathway optimization for Taxol precursor overproduction in Escherichia coli, *Science 330*, 70-74.
5. Ro, D. K., Paradise, E. M., Ouellet, M., Fisher, K. J., Newman, K. L., Ndungu, J. M., Ho, K. A., Eachus, R. A., Ham, T. S., Kirby, J., Chang, M. C., Withers, S. T., Shiba, Y., Sarpong, R., and Keasling, J. D. (2006) Production of the antimalarial drug precursor artemisinic acid in engineered yeast, *Nature 440*, 940-943.
6. Xue, Z., Sharpe, P. L., Hong, S. P., Yadav, N. S., Xie, D., Short, D. R., Damude, H. G., Rupert, R. A., Seip, J. E., Wang, J., Pollak, D. W., Bostick, M. W., Bosak, M. D., Macool, D. J., Hollerbach, D. H., Zhang, H., Arcilla, D. M., Bledsoe, S. A., Croker, K., McCord, E. F., Tyreus, B. D., Jackson, E. N., and Zhu, Q. (2013) Production of omega-3 eicosapentaenoic acid by metabolic engineering of Yarrowia lipolytica, *Nature biotechnology 31*, 734-740.
7. Agnew, D. E., Stevermer, A. K., Youngquist, J. T., and Pfleger, B. F. (2012) Engineering Escherichia coli for production of C(1)(2)-C(1)(4) polyhydroxyalkanoate from glucose, *Metabolic engineering 14*, 705-713.
8. Martin, V. J., Pitera, D. J., Withers, S. T., Newman, J. D., and Keasling, J. D. (2003) Engineering a mevalonate pathway in Escherichia coli for production of terpenoids, *Nature biotechnology 21*, 796-802.
9. Keasling, J. D. (2010) Manufacturing molecules through metabolic engineering, *Science 330*, 1355-1358.
10. Zhang, F., Ouellet, M., Batth, T. S., Adams, P. D., Petzold, C. J., Mukhopadhyay, A., and Keasling, J. D. (2012) Enhancing fatty acid production by the expression of the regulatory transcription factor FadR, *Metabolic engineering 14*, 653-660.
11. Woodruff, L. B., Boyle, N. R., and Gill, R. T. (2013) Engineering improved ethanol production in Escherichia coli with a genome-wide approach, *Metabolic engineering 17*, 1-11.
12. Kurland, C. G., and Dong, H. (1996) Bacterial growth inhibition by overproduction of protein, *Molecular microbiology 21*, 1-4.
13. Tyo, K. E., Ajikumar, P. K., and Stephanopoulos, G. (2009) Stabilized gene duplication enables long-term selection-free heterologous pathway expression, *Nature biotechnology 27*, 760-765.

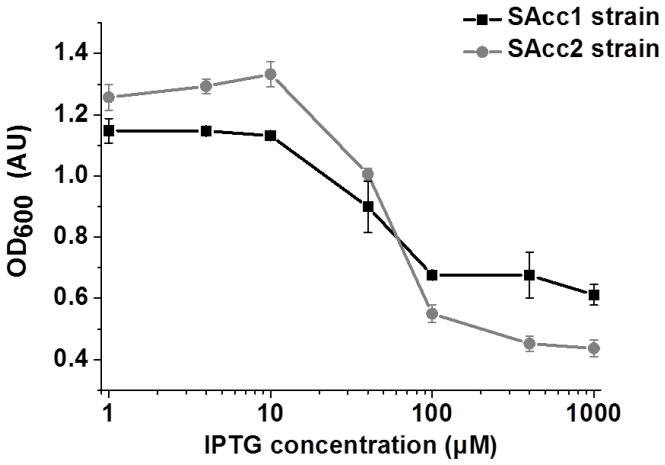
14. De Mey, M., Maertens, J., Lequeux, G. J., Soetaert, W. K., and Vandamme, E. J. (2007) Construction and model-based analysis of a promoter library for E-coli: an indispensable tool for metabolic engineering, *Bmc Biotechnol 7*.
15. Alper, H., Fischer, C., Nevoigt, E., and Stephanopoulos, G. (2005) Tuning genetic control through promoter engineering, *P Natl Acad Sci USA 102*, 12678-12683.
16. Salis, H. M., Mirsky, E. A., and Voigt, C. A. (2009) Automated design of synthetic ribosome binding sites to control protein expression, *Nat Biotechnol 27*, 946-U112.
17. Chubukov, V., Zuleta, I. A., and Li, H. (2012) Regulatory architecture determines optimal regulation of gene expression in metabolic pathways, *P Natl Acad Sci USA 109*, 5127-5132.
18. Kohlhaw, G. B. (2003) Leucine biosynthesis in fungi: Entering metabolism through the back door, *Microbiol Mol Biol R 67*, 1-+.
19. Shen-Orr, S. S., Milo, R., Mangan, S., and Alon, U. (2002) Network motifs in the transcriptional regulation network of Escherichia coli, *Nat Genet 31*, 64-68.
20. Chin, C. S., Chubukov, V., Jolly, E. R., DeRisi, J., and Li, H. (2008) Dynamics and design principles of a basic regulatory architecture controlling metabolic pathways, *Plos Biol 6*, 1343-1356.
21. Davis, M. S., and Cronan, J. E., Jr. (2001) Inhibition of Escherichia coli acetyl coenzyme A carboxylase by acyl-acyl carrier protein, *J Bacteriol 183*, 1499-1503.
22. Holtz, W. J., and Keasling, J. D. (2010) Engineering Static and Dynamic Control of Synthetic Pathways, *Cell 140*, 19-23.
23. Zhang, F. Z., Carothers, J. M., and Keasling, J. D. (2012) Design of a dynamic sensor-regulator system for production of chemicals and fuels derived from fatty acids, *Nat Biotechnol 30*, 354-U166.
24. Zhang, F., and Keasling, J. (2011) Biosensors and their applications in microbial metabolic engineering, *Trends Microbiol 19*, 323-329.
25. Farmer, W. R., and Liao, J. C. (2000) Improving lycopene production in Escherichia coli by engineering metabolic control, *Nat Biotechnol 18*, 533-537.
26. Anesiadis, N., Cluett, W. R., and Mahadevan, R. (2008) Dynamic metabolic engineering for increasing bioprocess productivity, *Metab Eng 10*, 255-266.
27. Gadkar, K. G., Doyle, F. J., Edwards, J. S., and Mahadevan, R. (2005) Estimating optimal profiles of genetic alterations using constraint-based models, *Biotechnol Bioeng 89*, 243-251.
28. Oyarzun, D. A., and Stan, G. B. (2012) Synthetic gene circuits for metabolic control: design trade-offs and constraints, *Journal of the Royal Society, Interface / the Royal Society*.
29. Fowler, Z. L., Gikandi, W. W., and Koffas, M. A. (2009) Increased malonyl coenzyme A biosynthesis by tuning the Escherichia coli metabolic network and its application to flavanone production, *Applied and environmental microbiology 75*, 5831-5839.
30. Xu, P., Ranganathan, S., Fowler, Z. L., Maranas, C. D., and Koffas, M. A. (2011) Genome-scale metabolic network modeling results in minimal interventions that cooperatively force carbon flux towards malonyl-CoA, *Metabolic engineering 13*, 578-587.

31. Lennen, R. M., and Pfleger, B. F. (2013) Microbial production of fatty acid-derived fuels and chemicals, *Current opinion in biotechnology*.
32. Zhang, F., Rodriguez, S., and Keasling, J. D. (2011) Metabolic engineering of microbial pathways for advanced biofuels production, *Current opinion in biotechnology 22*, 775-783.
33. Handke, P., Lynch, S. A., and Gill, R. T. (2011) Application and engineering of fatty acid biosynthesis in Escherichia coli for advanced fuels and chemicals, *Metab Eng 13*, 28-37.
34. Davis, M. S., Solbiati, J., and Cronan, J. E., Jr. (2000) Overproduction of acetyl-CoA carboxylase activity increases the rate of fatty acid biosynthesis in Escherichia coli, *The Journal of biological chemistry 275*, 28593-28598.
35. Lu, X., Vora, H., and Khosla, C. (2008) Overproduction of free fatty acids in E. coli: implications for biodiesel production, *Metabolic engineering 10*, 333-339.
36. Zha, W. J., Rubin-Pitel, S. B., Shao, Z. Y., and Zhao, H. M. (2009) Improving cellular malonyl-CoA level in Escherichia coli via metabolic engineering, *Metab Eng 11*, 192-198.
37. Miyahisa, I., Kaneko, M., Funa, N., Kawasaki, H., Kojima, H., Ohnishi, Y., and Horinouchi, S. (2005) Efficient production of (2S)-flavanones by Escherichia coli containing an artificial biosynthetic gene cluster, *Appl Microbiol Biot 68*, 498-504.
38. Leonard, E., Lim, K. H., Saw, P. N., and Koffas, M. A. G. (2007) Engineering central metabolic pathways for high-level flavonoid production in Escherichia coli, *Appl Environ Microb 73*, 3877-3886.
39. Schujman, G. E., Guerin, M., Buschiazzo, A., Schaeffer, F., Llarrull, L. I., Reh, G., Vila, A. J., Alzari, P. M., and de Mendoza, D. (2006) Structural basis of lipid biosynthesis regulation in Gram-positive bacteria, *The EMBO journal 25*, 4074-4083.
40. Schujman, G. E., Paoletti, L., Grossman, A. D., and de Mendoza, D. (2003) FapR, a bacterial transcription factor involved in global regulation of membrane lipid biosynthesis, *Dev Cell 4*, 663-672.
41. Lutz, R., and Bujard, H. (1997) Independent and tight regulation of transcriptional units in Escherichia coli via the LacR/O, the TetR/O and AraC/I1-I2 regulatory elements, *Nucleic acids research 25*, 1203-1210.
42. Moon, T. S., Lou, C., Tamsir, A., Stanton, B. C., and Voigt, C. A. (2012) Genetic programs constructed from layered logic gates in single cells, *Nature 491*, 249-253.
43. Davis, M. S., Solbiati, J., and Cronan, J. E. (2000) Overproduction of acetyl-CoA carboxylase activity increases the rate of fatty acid biosynthesis in Escherichia coli, *J Biol Chem 275*, 28593-28598.
44. Steen, E. J., Kang, Y. S., Bokinsky, G., Hu, Z. H., Schirmer, A., McClure, A., del Cardayre, S. B., and Keasling, J. D. (2010) Microbial production of fatty-acid-derived fuels and chemicals from plant biomass, *Nature 463*, 559-U182.
45. Lennen, R. M., Braden, D. J., West, R. A., Dumesic, J. A., and Pfleger, B. F. (2010) A process for microbial hydrocarbon synthesis: Overproduction of fatty acids in Escherichia coli and catalytic conversion to alkanes, *Biotechnology and bioengineering 106*, 193-202.

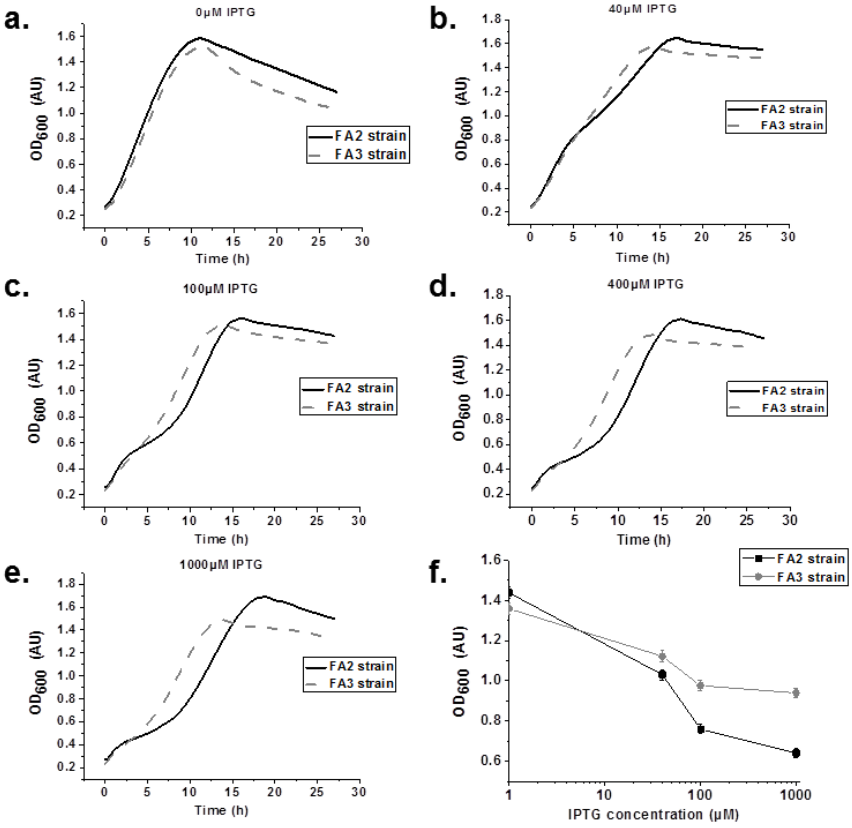
46. Schujman, G. E., Guerin, M., Buschiazzo, A., Schaeffer, F., Llarrull, L. I., Reh, G., Vila, A. J., Alzari, P. M., and de Mendoza, D. (2006) Structural basis of lipid biosynthesis regulation in Gram-positive bacteria, *Embo J 25*, 4074-4083.
47. Engler, C., Kandzia, R., and Marillonnet, S. (2008) A One Pot, One Step, Precision Cloning Method with High Throughput Capability, *Plos One 3*.
48. Lee, T. S., Krupa, R. A., Zhang, F., Hajimorad, M., Holtz, W. J., Prasad, N., Lee, S. K., and Keasling, J. D. (2011) BglBrick vectors and datasheets: A synthetic biology platform for gene expression, *Journal of biological engineering 5*, 12.

# Appendix: Supplementary Information for Chapter 3

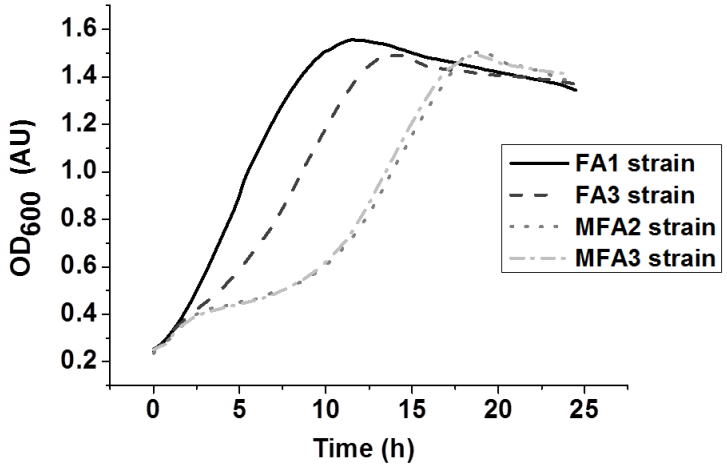
**Supplementary Figure 3S1.** Growth inhibition from the acetyl-CoA carboxylase overexpression. SAcc1 and SAcc2 strains (Table 2) were cultured in minimal medium with 2% glucose and cell growth was monitored continuously in a plate reader until stationary phase was reached. Cell growth at 20 hours after induction was plotted against different IPTG induction levels.



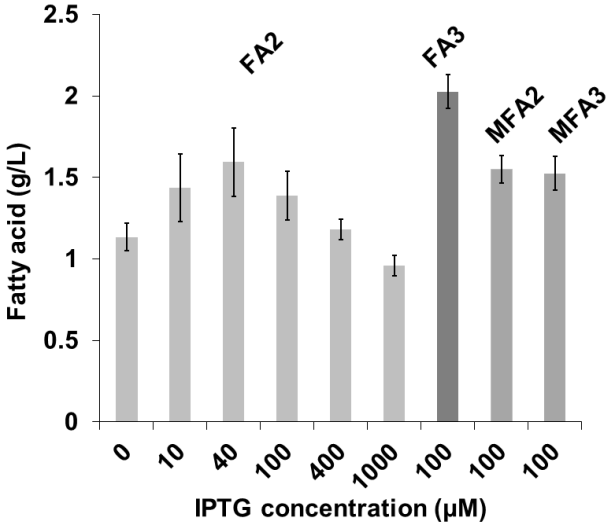
**Supplementary Figure 3S2.** Comparison of FA2 and FA3 cell growth under various IPTG induction levels. FA2 and FA3 strains were cultured in minimal medium with 2% glucose and 0.01% arabinose, and were induced with 200nM aTc and a) 0 μM (plotted as 1 μM), b) 40 μM, c) 100 μM, d) 400 μM, e) 1000 μM IPTG. Cell growth was monitored continuously on a plate reader (Method). f) $OD_{600}$ of FA2 and FA3 strains at 8 hours post induction.



**Supplementary Figure 3S3.** Effect of negative feedback circuit on cell growth. All the strains were cultured in minimal medium with 2% glucose and 0.01% arabinose, and were induced with 200 nM aTc and 0.1mM of IPTG. Cell growth was monitored continuously on a plate reader.



**Supplementary Figure 3S4.** Comparison of fatty acid production of the FA2, FA3, MFA2 and MFA3 strains. FA2 strain was induced with 200 nM aTc and various amounts of IPTG, and cultured in minimal medium with 2% glucose and 0.01% arabinose for 25 hours. FA3, MFA2 and MFA3 strains were cultured under the same condition and induced with 200 nM aTc and 0.1 mM IPTG induction. Fatty acid titers were analyzed as described in Methods.



***Mathematical model of Malonyl-CoA inducible System***

At equilibrium, malonyl-CoA binding to FapR is described by the equation (1)

$$FapR_f = \frac{K_d}{Mal+K_d} FapR_t \quad (1)$$

where $FapR_f$ is the concentration of free FapR, Mal is the concentration of free malonyl-CoA, $K_d$ is the dissociation constant of malonyl-CoA binding to FapR, $FapR_t$ is the total intracellular concentration of FapR.

The activity of a malonyl-CoA-responsive promoter $P_{mal}$, as represented by the normalized cell culture fluorescence, can be described by equation (2) (based on previously established models[1, 2]):

$$P_{mal} = P_{max} \frac{K_p P}{1+K_p P+2K_1 FapR_f+{K_1}^2 {FapR_f}^2} \quad (2)$$

where $P_{max}$ is the fully activated promoter activity (represented by the normalized cell culture fluorescence in the absence of the repressor). $K_p$ is the binding constant of the RNA polymerase to the promoter, and P is the concentration of the RNA polymerase. $K_1$ is the binding constant of FapR to the promoter.

Combining equations (1) and (2):

$$P_{mal} = P_{max} \frac{K_p P}{1 + K_p P + 2K_1 \frac{K_d}{Mal + K_d} FapR_t + K_1{}^2 \left(\frac{K_d}{Mal + K_d} FapR_t\right)^2} \quad (3)$$

Equation (3) was fitted to the fluorescence/malonyl-CoA data using malonyl-CoA concentrations quantified from LC-MS and the normalized cell culture fluorescence measured from plate-reader at various induction levels. Parameters were fitted using Matlab.

# References


1. Moon, T. S., Lou, C. B., Tamsir, A., Stanton, B. C., and Voigt, C. A. (2012) Genetic programs constructed from layered logic gates in single cells, *Nature 491*, 249-253.
2. Bintu, L., Buchler, N. E., Garcia, H. G., Gerland, U., Hwa, T., Kondev, J., and Phillips, R. (2005) Transcriptional regulation by the numbers: models, *Curr Opin Genet Dev 15*, 116-124.

# Chapter 4 Negative Feedback Circuits Provide Rapid Control Over Metabolite Dynamics


## Abstract

Metabolism constitutes the basis of life, and the dynamics of metabolism dictate various cellular processes. However, exactly how metabolite dynamics are controlled remains poorly understood. By studying an engineered fatty acid-producing pathway as a model, we found that a metabolic product from an unregulated pathway requires seven cell cycles to reach to its steady state level, with the speed mostly limited by enzyme expression dynamics. To overcome this limit, we designed metabolic feedback circuits (MFCs) with three different architectures, and experimentally measured and modeled their metabolite dynamics. Our engineered MFCs could dramatically shorten the rise-time of metabolites, decreasing it by as much as 12-fold. The findings of this study provide a systematic understanding of metabolite dynamics in different architectures of MFCs and have potentially immense applications in designing synthetic circuits to improve the productivities of engineered metabolic pathways.

# 4.1 Introduction

In nature, metabolism supports most cellular activities by providing building blocks, energy, cofactors and proper redox environments. Engineering microbial metabolic pathways, on the other hand, allows the production of chemicals, biofuels, and pharmaceuticals(Peralta-Yahya *et al*, 2012; Paddon & Keasling, 2014; Peralta-Yahya *et al*, 2011). As a primary goal of systems biology, understanding the complex regulatory networks of metabolism not only informs us how nature allocates limited cellular resources to perform various cellular activities in changing environments(Kotte *et al*, 2010; Berthoumieux *et al*, 2013; Shen-Orr *et al*, 2002; Mukherji & van Oudenaarden, 2009), but also provides design principles for synthetic biologists to better control metabolism for a variety of applications(Zhang *et al*, 2012; Liu *et al*, 2013; Dahl *et al*, 2013; Gupta *et al*, 2017; Xiao *et al*, 2016; Xu *et al*, 2014).

Microbial metabolism is mostly regulated via two types of controls: transcriptional regulation by controlling enzyme expression levels, which in turn affect metabolite concentration, and post-translational regulation by directly modifying enzyme activities via allosteric inhibition or activation. Post-translational metabolic regulation is common in central carbon metabolism, fatty acid biosynthetic pathway, and amino acid biosynthetic pathways, to rapidly adjust the cellular metabolite level within seconds(Pisithkul *et al*, 2015; Link *et al*, 2013). On the other hand, transcriptional metabolic regulation is widely used in nutrient uptake, the tricarboxylic acid (TCA)

cycle, amino acid biosynthesis, and energy production, with the primary goal to optimize protein expression levels and avoid overproduction of unnecessary proteins, which waste cellular resources(Chin *et al*, 2008; Chubukov *et al*, 2012; Zaslaver *et al*, 2004). Due to the involvement of transcription, translation, protein folding, and catalysis, transcriptional regulation changes the concentration of under-regulated metabolites much more slowly than post-translational regulation. Thus the dynamics of metabolite changes during the course of slow transcriptional regulation becomes important to cellular activities, particularly in those processes involved in cell survival and growth, such as nutrient uptake, building block biosynthesis, and quorum sensing.

From a control systems standpoint, an ideal metabolic control would both optimize protein expression level and offer rapid and precise control over metabolite concentrations. These criteria position transcriptional regulation as an ideal control strategy. Indeed, transcriptional regulation is regarded as the most common way for microbes to control their metabolism(Kochanowski *et al*, 2013). Previous studies have revealed that it takes five cell cycles for proteins to reach their steady state levels in an unregulated pathway, and the speed can be accelerated to as short as two cell cycles with negative auto-regulation. However, negative auto-regulation is mostly found to control transcription factors in nature and thus does not directly affect the dynamics of metabolism. Due to the delay in enzyme catalysis, the dynamics of metabolite will be even slower (more than five cell cycles) than proteins in unregulated pathways, and thus postpone cellular response to environmental changes and could lead to sub-optimal cell growth. In addition, from an engineering

perspective, the capability to accelerate metabolite dynamics is also preferred for rapid signal detection. Thus, design of synthetic regulations to speed up metabolite dynamics is crucial to both 1) understand how the regulatory architectures and the associated biochemical parameters affect metabolism in natural regulatory networks and 2) provide guidelines to re-program metabolite dynamics for applications such as environmental bio-sensing, metabolite oscillators, and quorum sensing mediated population control(Xiao *et al*, 2017; Bennett, 2015).

In this work, we employed an engineered fatty acid biosynthetic pathway that has defined interactions with other cellular processes to illustrate the interplay between individual transcriptional regulatory circuits and metabolite dynamics. Strikingly, we found that free fatty acid (FFA), a metabolic product synthetized from an unregulated, one-step enzymatic pathway, requires seven cell doublings to reach its steady state level once enzyme expression is turned on. Our modeling showed that slow metabolite dynamics are not affected by enzymatic catalytic parameters. We then constructed three metabolic feedback loops with different architectures to study metabolite dynamics under the control of each feedback regulation, using a combination of experimental and modelling methods. Our results show that negative feedback circuits can dramatically speed up metabolite rise-time (the time needed to reach half of the steady-state concentration). The findings of this study offer a quantitative and systematic guideline on choosing the genetic circuit architectures and biochemical parameters to control metabolite dynamics.

## 4.2 Results

### 4.2.1 Long Rise-Times of Metabolite Levels in Unregulated Pathways

To avoid natural regulation and minimize crosstalk between metabolic pathways (e.g., competition for metabolic precursors), we used an engineered FFA-producing pathway in *Escherichia coli* as a model to study metabolite dynamics. The engineered pathway branches from the high-flux fatty acid biosynthetic pathway by expressing a cytosolic thioesterase (encoded by *tesA*) under the control of an inducible promoter, $P_{Lac}$. The thioesterase uses acyl-ACPs as metabolic precursors to produce FFAs, which cannot be metabolized in our engineered cell due to the deletion of a β-oxidation gene (*fadE*). If needed, FFAs can be converted to other metabolic products by introducing heterologous enzymes with controlled conversion rates. Thus, the FFA-producing pathway satisfies all the requirements of this study: 1) a sufficient supply of metabolic precursors (acyl-ACPs from high-flux fatty acid biosynthetic pathway)(Zhang *et al*, 2011), 2) defined transcriptional regulation, 3) minimal interaction with other metabolic pathways, and 4) controllable product consumption.

To test the dynamics of a product in an unregulated pathway, the enzyme TesA was expressed from an IPTG-inducible promoter $P_{Lac}$ to create an open loop pathway (OL, Figure 4.1A). A red fluorescent protein (RFP, encoded by *rfp*) was placed 3' of TesA in the same operon to monitor the expression dynamics of TesA. Cells were cultivated in the exponential growth phase. After

induction by IPTG, the changes in RFP fluorescence and FFA concentration over time were measured. Cell-density-normalized RFP fluorescence was used to indicate enzyme dynamics, and the time course of cell-density-normalized FFA was measured as metabolite dynamics (Fig 4.1B&4.1C). Surprisingly, FFA takes as many as seven cell cycles to reach its steady state concentration after tuning on the expression of TesA.

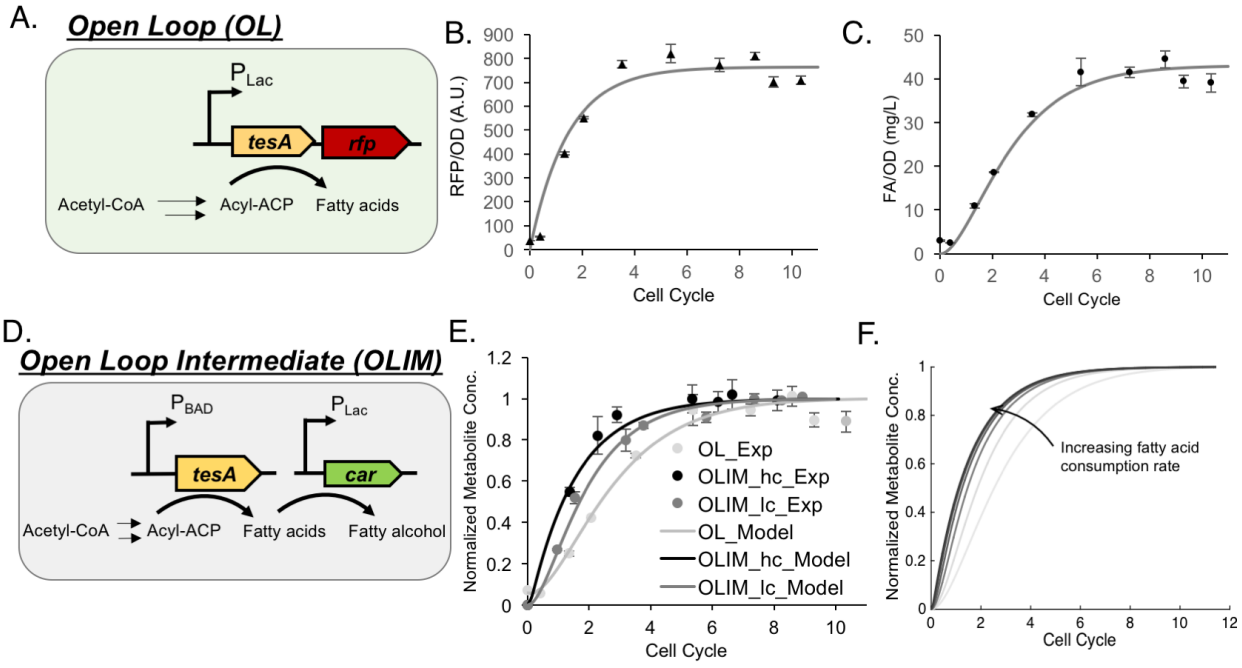



**Figure 4.1 Protein and metabolite production dynamics in unregulated metabolic pathways.**

A. An engineered free fatty acid (FFA) pathway was chosen as the model pathway to study the dynamics of a pathway product. FFAs were produced by expressing a thioesterase (encoded by *tesA*) under the control of a $P_{Lac}$ promoter. RFP was expressed in the same operon to monitor the

expression dynamics of the thioesterase.

B, C. Expression dynamics of the (B) thioesterase and (C) FFA. TesA and FFA dynamics (dots) were fitted to equations (1) and (2), respectively (solid line).

D. A fatty alcohol pathway was used as a model pathway to study the dynamics of a pathway intermediate. The fatty alcohol pathway was constructed by expressing a thioesterase under the control of a $P_{BAD}$ promoter and a carboxylic acid reductase (encoded by *car*) under the control of a $P_{Lac}$ promoter.

E. FFA production dynamics with high (labelled as OLIM_hc) and low (labelled as OLIM_lc). FFA consumption rates were measured and fitted to a model. Data are plotted together with the production dynamics of a pathway product for comparison.

F. Simulation results of the intermediate dynamics with different intermediate consumption rates. As the consumption rate of the intermediate increases (from light gray to dark gray lines), the intermediate dynamics further speed up and hit an upper limit.

To quantitatively understand the slow FFA accumulation, we constructed a general mathematical model to describe the enzyme expression and metabolite dynamics in unregulated metabolic pathways. In the case of the single-step FFA-producing pathway, the enzyme is thioesterase and the metabolite is FFA. In the exponential growth phase and with a constant supply of precursors, enzyme expression and metabolite concentration can be described by differential equations (1) and (2) in the Supplemental Information M1. Solving these equations, we obtain

$$[Enzyme] = [Enzyme]_{ss} \cdot (1 - \exp(-ln2 * N)) \qquad (1)$$

$$[Metabolite] = [Metabolite]_{ss} \cdot (1 - (ln2 * N + 1) \cdot \exp(-ln2 * N)) \quad (2)$$

where N represents the number of cell cycles. $[Enzyme]_{ss}$ and $[Metabolite]_{ss}$ represent the steady state concentrations of the enzyme and the metabolite, respectively. Our experimental results matched well with the model prediction ($R^2$=0.98): After a biosynthetic pathway is turned on, the settling time (defined as the time required for a response curve to reach and stay within 5% of the steady state level, Figure 4S1) for the protein and metabolite are five and seven cell cycles, respectively. Moreover, since equation (2) is independent of the catalytic parameters, the settling times are functions of cell cycles only, and not pathway specific. Further analysis revealed that the long metabolite settling time was mostly caused by the slow rise of the enzyme concentration (a settling time of five cell cycles) and the slow metabolite consumption/dilution time. The slowly rising protein concentration has been observed previously and is caused by the long protein lifetime. Similarly, as an end product, a metabolite is not consumed and can only be diluted by cell division, thus explaining why the settling time of a metabolite product depends only on cell cycles. Thus, based on our analysis we hypothesized that the metabolite dynamics can be accelerated by 1) a faster metabolite consumption/dilution rate and 2) faster protein production dynamics.

Next, to speed up the metabolite dynamics by increasing its consumption rate, we studied the

dynamics of an intermediate metabolite. A simple two-step metabolic pathway was simulated using ODE equations (See Supplemental Information M2). Mathematical analysis revealed that an intermediate metabolite could reach to its steady state faster than that of a product, depending on its enzymatic consumption rate. The theoretical prediction was verified experimentally using an engineered two-step metabolic pathway, where a carboxylic acid reductase (CAR, encoded by *car*) was constantly expressed in the FFA-producing pathway (See Methods)(Akhtar *et al*, 2013), converting FFA to fatty alcohols and making FFA a pathway intermediate (Fig 4.1D) (open loop intermediate, OLIM). As we increased the FFA consumption rate by tuning the expression level of CAR, FFA reached its steady state faster (Fig 4.1E). Further analysis also showed that as the intermediate's consumption rate increased, its rise-time decreased until reaching a limit, where consumption of the intermediate was so rapid that metabolite dynamics overlapped with the dynamics of enzyme expression (Fig 4.1F).

Overall, our results indicated that the dynamics of a metabolite from an open loop, regardless of whether the metabolite is a pathway intermediate or an end product, are slow and are ultimately limited by the dynamics of enzyme expression. Following transcriptional activation, a metabolite takes at least five (for an intermediate) or seven (for a product) cell cycles to reach steady state when a pathway is unregulated. Such slow responses greatly delay cell growth and adaptation if the metabolite is essential to cell growth, thus necessitating regulatory strategies to expedite metabolite dynamics.

### 4.2.2 Genetic Negative Feedback Circuits Accelerate Metabolic Dynamics

One of the most common transcriptional regulations of metabolism is negative feedback, where the product of a metabolic pathway inhibits the transcription of pathway enzymes. We hypothesized that such metabolic negative feedback circuits can speed up metabolite rise-time, just as negative gene circuits speed up the rise-time of under-controlled proteins(Rosenfeld *et al*, 2002). To test this hypothesis, three different metabolic feedback systems were designed and constructed, representing three commonly found distinct regulation architectures in nature and engineered systems. The first feedback system contains a negative auto-regulated gene circuit, called a negative gene loop (NGL), where *tesA* is co-transcribed with a repressor, *tetR*, which feedback inhibits the expression of *tesA-tetR* via a hybrid promoter, $P_{TL}$ (Fig 4.2A). $P_{TL}$ was engineered by placing a TetR-binding site (TetO) between the -35 and -10 region of a strong promoter and a LacI-binding site (LacO) 3' of the -10 region, thus allowing $P_{TL}$ to be repressed by both TetR and LacI. While the FFA-producing pathway can be transcriptionally activated by addition of IPTG, expression of TesA is negatively auto-regulated, thus affecting FFA dynamics (Fig 4.3A). This architecture is prevalent in nature for the regulation of ligand-independent transcription factors. The second feedback system, named a negative metabolic loop (NML), contains a FFA-repressed transcription factor, FadR, which activates the expression of *tesA* through a hybrid promoter, $P_{FL}$ (Fig 4.3B). $P_{FL}$ was engineered based on a natural FA-activating promoter, $P_{fabA}$. Specifically, $P_{fabA}$ contains a FadR-binding site (FadRO) 5' of the -35 region, allowing the promoter to be activated by FadR by recruiting RNA polymerase. A LacO was placed

3’ of the -10 region of $P_{fabA}$, allowing the resulting promoter $P_{FL}$ to be repressed by LacI, as experimentally verified (Fig 4.2B). Thus, upon transcriptional activation of FFA production via IPTG induction, FFA can be quickly synthesized. A proportion of the produced FFA is reversibly converted to acyl-CoA by an endogenous enzyme, acyl-CoA synthase. Acyl-CoA then binds to FadR and antagonizes FadR’s DNA-binding activity, decreasing the transcription rate from $P_{FL}$, thus slowing FFA production. Overall, the NML forms a negative feedback loop that involves dynamic sensing of the metabolic product, FFA. NML mimics the prevalent natural regulation via ligand-dependent TFs, such as transcriptional inhibition of genes by the pathway end-product in a few *E. coli* amino acid biosynthetic pathways. The third feedback system, named a layered negative metabolic loop (LNML), contains three inhibition layers of regulation and requires two gene expression steps to feedback control the production of FFA. Specifically, an FA-activated promoter, $P_{AR2}$, was engineered by inserting a FadRO between the -35 and -10 region of a phage T7 promoter, so that $P_{AR2}$ can be repressed by FadR (Fig 4.2C). $P_{AR2}$ was then used to control the expression of TetR, which represses the expression of the *tesA-rfp* from $P_{TL}$. Thus, transcriptional activation of the pathway by IPTG induces *tesA* expression from $P_{TL}$. The produced FFA antagonizes FadR’s DNA-binding activity via acyl-CoA, leading to activation of TetR expression from $P_{AR2}$. TetR then decreases FFA production by down-regulating *tesA* expression (Fig 4.3C). To monitor enzyme expression, an *rfp* was cloned 3’ of TesA in all three feedback loops.

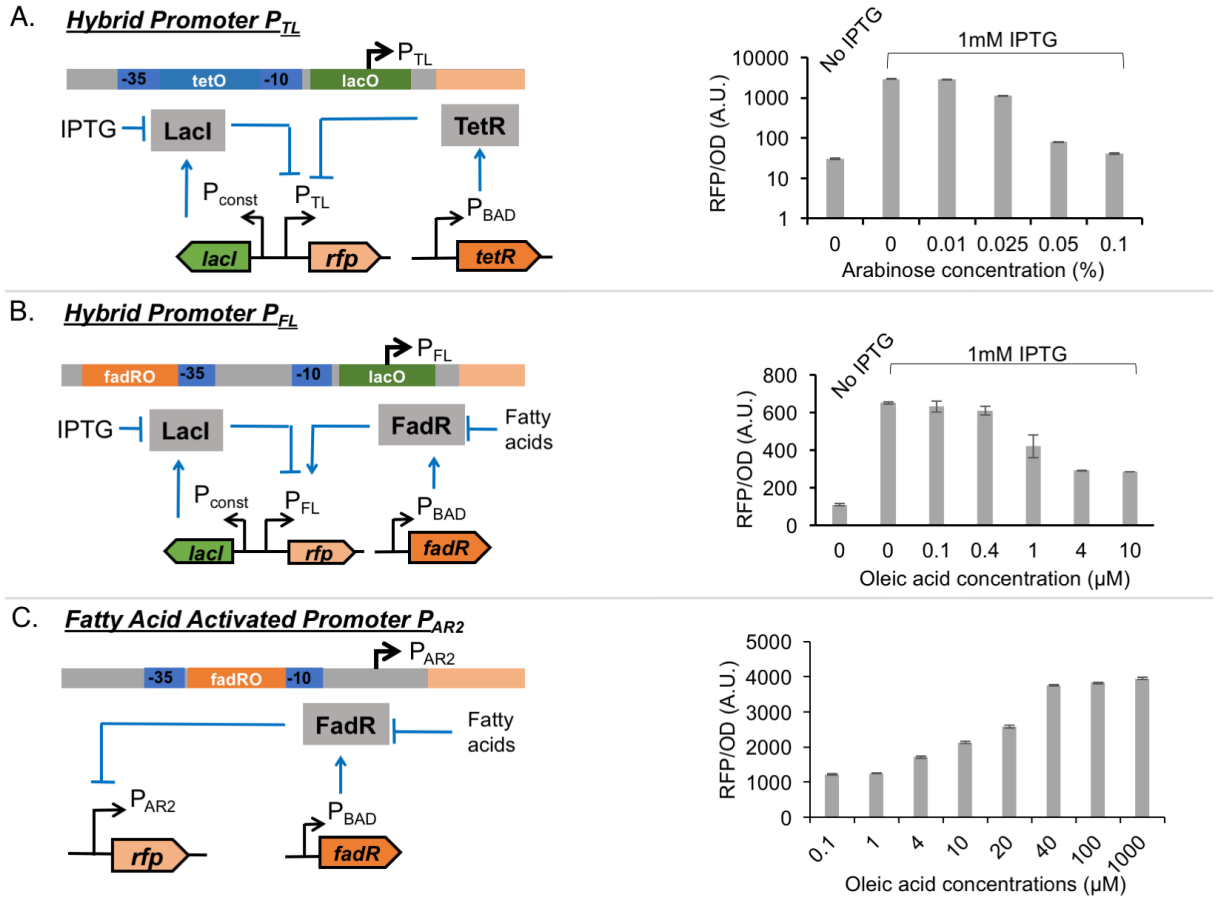



**Figure 4.2 Design of synthetic promoters to create negative feedback circuits.**

A. Design and characterization of the $P_{TL}$ promoter. A hybrid promoter, $P_{TL}$, was created by incorporating TetO and LacO into the promoter. The promoter was activated by IPTG and repressed by TetR in a dose-dependent manner.

B. Design and characterization of the $P_{FL}$ promoter. A hybrid promoter, $P_{FL}$, was created by introducing LacO into the promoter region of the FabA promoter. The promoter was activated by IPTG and repressed by fatty acids in a dose-dependent manner.

C. Design and characterization of the $P_{AR2}$ promoter. A fatty acid responsive promoter, $P_{AR2}$, was engineered by introducing FadRO into the promoter region of a phage lambda promoter. The

promoter was activated by oleic acids.

Responses of the three engineered hybrid promoters to their corresponding effectors were first individually validated in expression tests. As expected, both $P_{TL}$ and $P_{FL}$ could be activated by 1 mM IPTG (Fig 4.2). $P_{TL}$ was then subjected to down-regulation by TetR, as tested by increasing the TetR expression level from a $P_{BAD}$ promoter using arabinose. $P_{FL}$ was subjected to downregulation by the addition of oleic acids in a dose-dependent manner (Fig 4.2A & 4.2B). $P_{AR2}$ could be activated by increasing the oleic acid concentration, per our design (Fig 4.2C). Next, three *E. coli* strains carrying engineered metabolic feedback systems were cultivated. Enzyme expression and metabolite dynamics after pathway activation were monitored by measuring cell-density-normalized fluorescence and FFA concentration. Both time course curves were then normalized by their steady state values.

FFA displayed distinctly different dynamics in all three types of regulatory architectures. While NGL increased the speed of FFA dynamics only slightly compared to that of an unregulated pathway (Fig 4.3A), NML allowed FFA to reach steady state more rapidly, in three cell cycles (settling time), as compared to seven cell cycles for an unregulated pathway (Fig 4.3B). LNML, on the other hand, raised the FFA concentration quickly to a high level and gradually lowered it to its steady state, causing an overshoot of FFA concentration (Fig 4.3C). Compared to the rise-time for an unregulated pathway (2.48 cell cycles), the rise-times of FFA in three regulated pathways

were 1.75, 1.33 and 0.21 cell cycles, decreased by 1.4-, 1.9-, and 11.8-fold, respectively. The dynamics of TesA also changed, with patterns similar to that of FFA in each strain. The rise-times of TesA decreased from one cell cycle in the unregulated pathway to 0.41, 0.38, and 0.03 cell cycles in NGL, NML, and LNML, respectively, indicating that the rapid metabolite dynamics were results of transcriptional regulation of the enzyme expression level, consistent with our design (Fig 4.3A, 4.3B, &4.3C). Overall, our results demonstrated that the metabolic negative feedback loops could accelerate metabolite dynamics, decreasing metabolite rise-time by up to 11.8-fold.

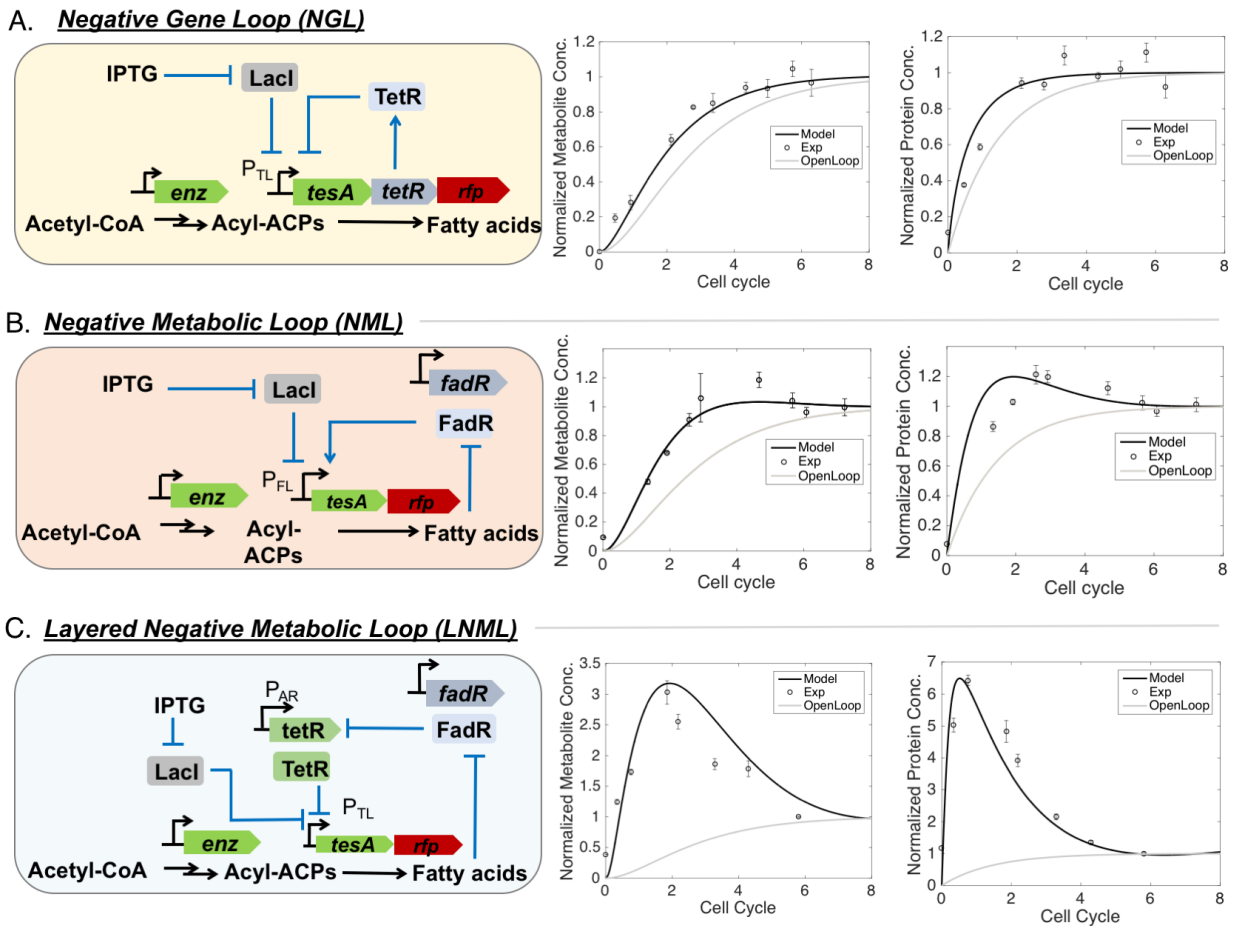



**Figure 4.3 Negative feedback circuits with three different architectures and their metabolite**

**and protein dynamics.**

A. Design and dynamics of metabolite and protein in a negative gene loop (NGL). In the NGL, $P_{TL}$ promoter is used to control the expression of the *tesA-tetR-rfp* operon. Induction of the operon initiates the synthesis of TetR, which feeds back to inhibit the operon.

B. Design and dynamics of metabolite and protein in a negative metabolic loop (NML). In the NML, $P_{FL}$ is used to control the expression of the *tesA-rfp* operon. Upon induction of $P_{FL}$, FFA synthesis is initiated, and its accumulation turns down the expression of the operon.

C. Design and dynamics of metabolite and protein in a layered negative metabolic loop (LNML). In the LNML, $P_{TL}$ is used to control the expression of the *tesA-rfp* operon. The inverter TetR is placed under the control of $P_{AR2}$. Production of free fatty acids is turned on by inducing TesA synthesis, and the accumulation of free fatty acids induces TetR expression. The accumulation of TetR turns down TesA synthesis by inhibiting $P_{TL}$.

Normalized free fatty acid and protein accumulation dynamics were all fitted with kinetic models (solid black lines) and are plotted with those of a non-regulated pathway (solid gray lines) for comparison.

### 4.2.3 Model Analysis Elucidates Metabolite Dynamics Behavior for the Three Feedback Architectures

To obtain a quantitative understanding of the regulated metabolite dynamics and to explore the limits of each regulatory architecture, general mathematical models were developed and fitted

to our experimental data in each system (see Supplemental Information, Fig 4.3). Circuit parameters that could be experimentally tuned using cost-effective methods were varied in each model to explore the boundaries of metabolic control. In the case of the NGL, we varied the dissociation constant ($K_d$) between TetR and $P_{TL}$ and simulated FFA dynamics. As seen in Fig 4.4A and 4.4B, as $K_d$ decreases (tighter repression of $P_{TL}$ by TetR), FFA reaches its steady state faster and approaches a theoretical upper limit, the fastest metabolite dynamics under NGL control. Our experimental results lie close to this upper limit, consistent with the tight interaction between TetR and its operator site TetO in $P_{TL}$ ($K_d$ = 30 nM). Under this theoretical upper limit condition, the protein's rise-time was deceased by 5-fold (from one cell cycle to 0.21 cell cycle, Fig 4.4B), and the metabolite's rise-time was decreased by only 1.57-fold (from 2.48 cell cycle in OL to 1.58 cell cycle in the fastest NGL). Thus, an NGL speeds up the response of protein expression upon transcriptional activation, but has mild effects on metabolite dynamics. Consistent with this conclusion, natural NGLs are mostly found to regulate transcription factors but rarely to regulate enzymes, possibly due to their inefficiency in controlling metabolite dynamics.

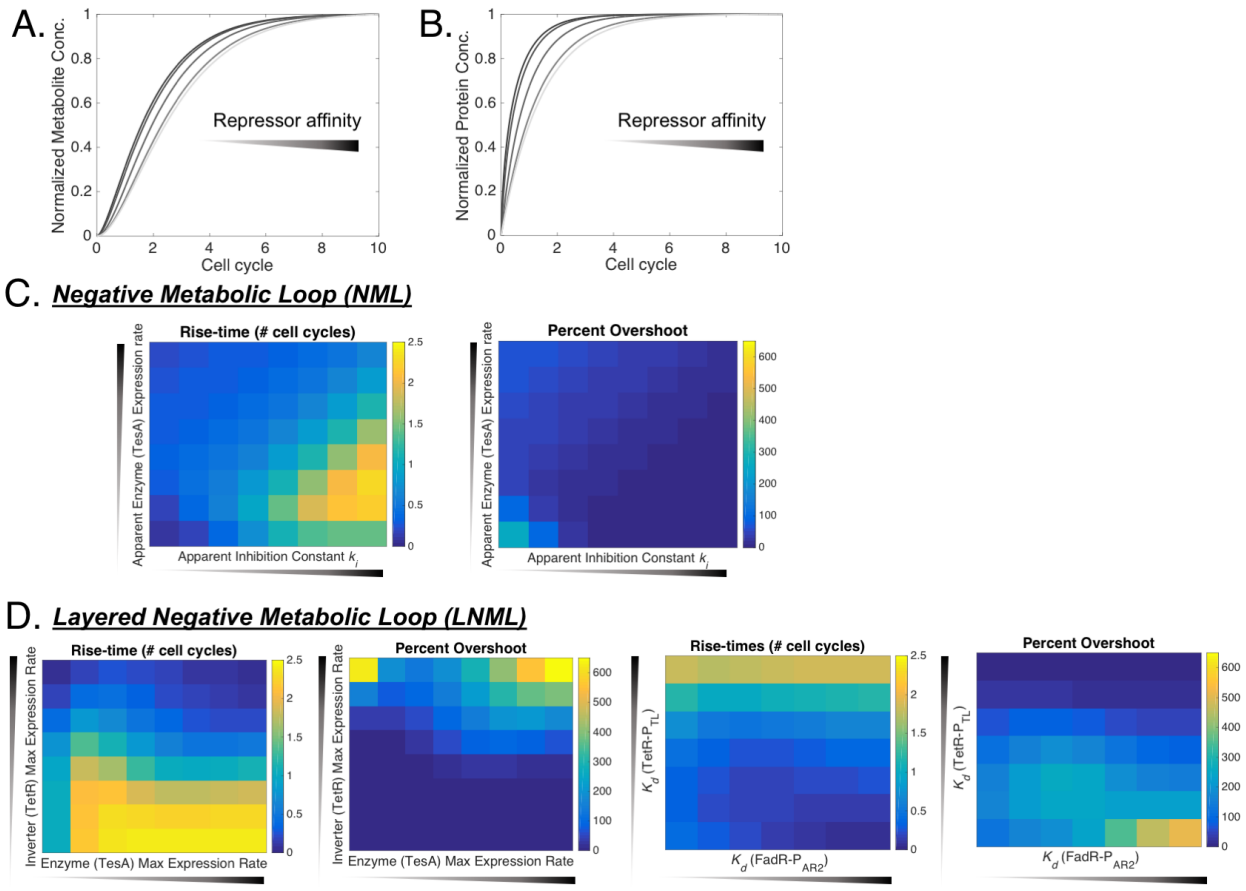



**Figure 4.4 Tuning of circuit parameters to control protein and metabolite dynamics.**

A, B. Dynamics of (A) metabolite and (B) protein in NGL with varying dissociation constants between TetR and $P_{TL}$. $K_d$ decreases from light gray to dark gray lines. Upper limits of achievable metabolite and protein dynamics were observed.

C. Rise-time and percent overshoot of NML under varied tunable circuit parameters.

D. Rise-time and percent overshoot of LNML under varied tunable circuit parameters.

Compared to NGL, NML exhibited a more dramatic acceleration in the metabolite dynamics, with a decrease in both the rise-time and the settling time. To further explore the effects of circuit

parameters on the metabolite dynamics, we varied the apparent enzyme expression rate ($r'_{FL}$, see Supplemental Information M4) from $10^{-13}$ M/s to $10^{-8}$ M/s, and varied the apparent inhibition constant ($k_i$, see Supplemental Information M4) from 10μM to 1 M to cover large biologically relevant ranges (Fig 4.4C). As the apparent inhibition constant increases, the repression from the negative feedback loop decreases, which increases the rise-time. On the other hand, the metabolite rise-time correlates non-monotonically with the enzyme expression rate. Because the steady state enzyme expression level is correlated with the enzyme expression rate, a very low TesA expression rate leads to a low TesA steady state level (close to its initial concentration), therefore resulting in a short rise-time. As the TesA expression rate increases, FFA accumulates to higher levels, which gradually cross the inhibition threshold ($k_i$). When the FFA level is much lower than the threshold, TesA expression is barely inhibited, which causes an increase in rise-time. As the TesA expression rate further increases, FFA reaches a concentration much higher than the threshold, which leads to a strong repression that decreases the rise-time. In the parameter space we explored, the metabolite rise-time was decreased by more than two-fold in 80% of the parameter space with very little overshoot (percent overshoot < 50 in 86% of the parameter space). This suggests that NML is particularly effective in accelerating the metabolite dynamics without causing large overshoots.

Comparing the three architectures, LNML had the shortest metabolite rise-time, but at the same time caused a large metabolic overshoot. Metabolite overshoot is an interesting phenomenon that can be potentially exploited for engineering purposes, including metabolite-triggered bistable

switches(Kotte *et al*, 2015), transient metabolic signals(Shin *et al*, 2006; Ray *et al*, 2011), and metabolic oscillators(Elowitz & Leibler, 2000; Novák & Tyson, 2008). On the other hand, too large an overshoot may also adversely affect cell growth if the metabolite concentration is above the cellular tolerance level. Thus, it is essential to fine tune circuit parameters to balance the metabolite rise-time and the size of metabolic overshoot.

Because LNML consists of multiple layers of interactions, there is a large set of experimental tunable parameters. These include the protein production rate (such as TesA and TetR), and the dissociation constants between TFs and their cognate promoters (such as FadR-$P_{AR2}$ and TetR-$P_{TL}$). We systematically varied the experimentally-tunable parameters in our model and studied how they affected the metabolite rise-time and the size of the overshoot. The maximum protein production rates of TesA and TetR were varied from $10^{-13}$ M/s to $10^{-8}$ M/s, and the dissociation constants of FadR-$P_{AR2}$ and TetR-$P_{TL}$ were scanned from 0.1 nM to 10 μM, both covering broad ranges of biologically relevant values. As the maximum TetR production rate increases, FFA exhibits a shorter rise-time and a higher percent overshoot (defined as the ratio of the amount of overshoot to the steady-state) (Fig 4.4D). Interestingly, the rise-time and the percent overshoot correlates non-monotonically with the maximum TesA expression rate (Fig 4.4D). A slow TesA production rate leads to minimal TetR expression, causing a long delay in TetR accumulation to repress the circuit, which generates a large percent overshoot and results in a short rise-time. Contrariwise, a high enzyme expression rate leads to strong activation, and a delay of repression

follows because high levels of repressors are accumulated to bring down the enzyme expression rate, which generates overshoot and a short rise-time (Figure 4S2). In addition, as we increase the dissociation constant between TetR and $P_{TL}$, the repression strength of the circuit becomes weaker and thus approaches the dynamic behavior of an OL (Fig 4.4D). However, as the binding affinity between FadR and $P_{AR2}$ decreases, the FFA rise-time decreases and reaches a valley before it increases. At low $K_d$ (FadR-$P_{AR2}$), high FFA concentrations are required to turn on $P_{AR2}$, which causes a delay in repression and a long rise-time. As $K_d$ (FadR-$P_{AR2}$) increases, less FFA is needed to turn on $P_{AR2}$, and thus rise-time decreases. When $K_d$ (FadR-$P_{AR2}$) further increases beyond a certain level, TetR expression becomes insensitive to FFA concentration, leading to a longer rise-time (Figure 4S2). Interestingly, the $K_d$ (FadR-$P_{AR2}$) of the rise-time valley shifts right as the $K_d$ (TetR-$P_{TL}$) decreases. This was because as $K_d$ (TetR-$P_{TL}$) decreases, TesA expression becomes more sensitive to TetR levels, thus increasing the threshold of $K_d$ (FadR-$P_{AR2}$) that can produce varying $P_{TL}$-sensitive TetR concentrations at different FFA levels. Overall, our model results suggest that a decrease in rise-time is always coupled with an increase in the size of overshoot. To shorten the rise-time, a high inverter (TetR) maximum expression rate and low $K_d$ (TetR-$P_{TL}$) should be used. In addition, fine tuning of the maximum enzyme (TesA) expression rate and $K_d$ (FadR-$P_{AR2}$) are also crucial to obtain a short rise-time for the LNML circuit.

Comparing different negative feedback topologies, metabolic overshoot is often seen in LNML, indicating that metabolic overshoot is more tightly associated with the regulatory

architecture rather than the circuit parameters. The addition of an inverter in LNML (the TetR layer) creates an additional step of gene expression, increasing the delay time from metabolite sensing to its repression. Indeed, when the TetR degradation rate is increased (we denote this architecture as LNML-Deg), the overshoot region shrinks and the overshoot size decreases dramatically, while rise-times remain short (See Supplemental Information, Figure 4S3). This finding suggests that using an inverter with a short life time, such as fusing a degradation tag to the repressor or using RNA-based repression mechanisms, can potentially shorten the settling time.

Overall, these results suggest that NML is the most effective architecture to accelerate metabolite dynamics, with dramatically decreased rise-time and little overshoot. In NML, since low and high $k_i$ may lead to large metabolic overshoots and long rise-times, respectively, these parameter values should be avoided when constructing metabolic circuits. By comparison, LNML is also effective in shortening the rise-time but mostly at the cost of generating large overshoots. Since a high enzyme expression rate is often preferred in engineering applications to increase metabolite concentrations, a medium expression level of the inverter is preferred to speed up the response while maintaining a relatively small percent overshoot. In addition, because the rise-time and percent overshoot correlate non-monotonically with $K_d$ (FadR-$P_{AR2}$), fining tuning of $K_d$ (FadR-$P_{AR2}$) allows further precise control of the metabolite dynamics. NGL is inefficient in accelerating metabolite dynamics but may be considered when a metabolite biosensor is not available. These design guidelines can be used to choose a proper regulatory architecture and parameters to achieve

desirable metabolite dynamics.

## 4.3 Conclusions and Discussions

While previous efforts mostly focused on the effects of regulatory networks on the steady state metabolite level, this work provided a systematic and quantitative view of how transcriptional feedback circuits affect metabolite dynamics. Three types of commonly observed architectures of metabolic feedback loops were constructed and analyzed both experimentally and mathematically. Under these feedback controls, metabolite rise-time can be dramatically increased, by up to 11.8-fold over that of unregulated metabolic pathways. The effects of several regulatory parameters on metabolite dynamics were also systematically studied, allowing a deeper understanding of metabolic regulation in both natural and engineered systems.

It is suggested that natural enzymes are usually expressed at abundant levels at steady state, and thus serve as a buffer to maintain a steady cell metabolism under small environmental perturbations(Kochanowski *et al*, 2013; Fendt *et al*, 2010). However, under drastic environmental changes, such as the depletion of amino acids, transcriptional regulation turns on the expression of an entire biosynthetic pathway, causing large metabolic changes. During such a transition, the ability to rapidly adjust enzyme expression and catalysis to produce a desirable level of the target metabolite increases the rate of adaptation. Thus, negative metabolic feedback loops provide rapid

response to large environmental changes by quickly optimizing both enzyme and metabolite levels. On the other hand, the vast majority of engineered pathways in the field of metabolic engineering and synthetic biology are unregulated and could be slow to respond to large metabolic changes. Such slow response may cause delays in signal detection by biosensors that sense a metabolic product of an environmental signal(Xiao *et al*, 2017). Therefore, this study also provides guidelines to design synthetic regulatory circuits to speed up metabolite dynamics in engineered systems.

The benefit of a faster metabolite rise-time is, however, also accompanied with a possible metabolic overshoot. Metabolic overshoot may be undesired in most natural systems, due to the overproduction of unnecessary proteins and metabolites. Although overshoot can be mitigated to some degree by fine-tuning the regulatory parameters of the metabolic feedback circuits, our study found that it is more tightly associated with regulatory architectures. Compared to LNML, NML has a greater parameter space where a shorter rise-time is achieved without overshoot, strongly suggesting why NML is the most commonly found transcription metabolic regulation in prokaryotes. For synthetic biology applications where metabolite overshoot is needed, LNML provides a large parameter space to generate overshoot with tunable size.

Gene regulation lies at the center of systems biology, and its functional role has inspired researchers to build synthetic regulation for various engineering applications. In this work, we

exploited genetic circuits to speed up metabolite dynamics. Our study filled the gap that links gene regulation with metabolite dynamics and indicated that, beyond optimizing cellular resources for enzyme production, transcriptional regulation also plays a critical role in quickly adapting metabolite levels to large environmental shifts. Our work also provides a systematic design principle that illustrates how regulation architectures and parameter fine-tuning affect metabolite productivity, which can be used to engineer synthetic feedback circuits to better control metabolite dynamics.

# 4.4 Methods

## 4.4.1 Materials

Phusion DNA polymerase, restriction enzymes, and T4 ligase were purchased from Thermo Fisher Scientific (Waltham, Massachusetts, U.S.A.). Gel purification and plasmid miniprep kits were purchased from iNtRON Biotechnology (Lynnwood, WA, U.S.A.). All primers were synthesized by Integrated DNA Technologies (Coralville, IA, U.S.A). All reagents were purchased from Sigma Aldrich (St. Louis, MO, U.S.A.). *E. coli* DH10B was used for cloning purposes, and *E. coli* DH1 (Δ*fadE*) was used to construct negative feedback circuits.

### 4.4.2 Plasmids and strains

Plasmid pB5k-tesA-RFP was constructed by cloning a cytosolic thioesterase gene *tesA* (*'tesA*: leader sequence deleted) to the 5' of RFP in a Biobrick vector pBbB5k-RFP. To create pB8k-tesA, *tesA* was amplified and cloned 3' of the $P_{BAD}$ promoter in a Biobrick plasmid pBbB8k-RFP. Plasmid pA5c-CAR was constructed by cloning a carboxylic reductase 3' of a $P_{LacUV5}$ promoter in a Biobrick plasmid pBbA5c-RFP. Plasmid pETLa-RFP was created by inserting one TetO site in between the -35 and -10 region of the $P_{lacUV5}$ promoter in a Biobrick plasmid pBbE5a-RFP, using a one-step Golden-Gate DNA assembly method. To create pETLa-tesA-RFP, *tesA* was cloned 5' of *rfp* in the pETLa-RFP plasmid. The gene *tetR* was amplified from a BioBrick plasmid, pBbA2c-RFP, and cloned 3' of the $P_{BAD}$ promoter in a Biobrick plasmid pB8k-RFP, yielding pB8k-tetR. Plasmid pBAR2k-RFP was constructed from phage $P_{A1}$ promoter by placing one FadRO site between the -35 and -10 regions of the promoter. Plasmid pETLa-tesA-tetR-RFP was constructed by cloning *tetR* 3' of *tesA* and 5' of *rfp* in the pETLa-tesA-RFP plasmid. Plasmid pBAR2k-tetR and pA8c-fadR were constructed by replacing *rfp* in pBAR2k-RFP and pBbA8c-RFP with *tetR* and *fadR*, respectively. Strains were created by transforming the corresponding plasmids into DH1 (Δ*fadE*) competent cells. Plasmids (Table 4S1), strains (Table 4S2), and promoter sequences (Table 4S3) are summarized in the Supplemental Information.

### 4.4.3 Cell growth and fluorescence assay

Cell growth curves and cell culture fluorescence were recorded on an Infinite F200PRO (TECAN)

plate reader. Strains were first cultivated overnight in Luria−Bertani (LB) medium (220 rpm, 37 °C) supplemented with appropriate antibiotics (50 mg/L ampicillin, 50 mg/L kanamycin, and 30 mg/L chloramphenicol). To test $P_{TL}$ promoter behavior, the overnight LB cultures were inoculated 2% v/v into fresh minimal medium(Liu *et al*, 2013) with 2% glucose, and the overnight minimal medium culture was then inoculated into fresh minimal medium. To test $P_{FL}$ promoter behavior, the overnight LB cultures were inoculated into fresh minimal medium with 1% glycerol as a carbon source (supplemented with 0.5% tergitol NP-40), and the overnight minimal medium culture was then inoculated into fresh minimal medium. The $P_{AR2}$ promoter was tested using a method modified from previous publication(Zhang *et al*, 2012). Specifically, cells were induced with the corresponding inducer concentrations at OD=0.6, and then were incubated in a 96-well plate inside the plate reader with shaking (218.3 rpm, 37 °C). The cell density ($OD_{600}$) and red fluorescence (excitation, 535 ± 9 nm; emission, 620 ± 20 nm) were recorded every 1000 s until the cell culture reached the stationary phase. Fluorescence from a wild-type *E. coli* DH1 (Δ*fadE*) cell culture was used as the background, and was subtracted from all fluorescence measurements. The background-corrected fluorescence was later normalized by cell density as measured at $OD_{600}$.

### 4.4.4 Protein and metabolite production dynamics

The OL, OLIM, NGL, NML, and LNML strains were inoculated into LB medium with appropriate antibiotics. For adaptation, overnight LB cultures of OL, NGL, and LNML strains were inoculated

2% v/v into M9 medium supplemented with 1% glycerol and amino acids, the same composition as EZ-rich medium. The overnight culture of OLIM was inoculated 2% v/v into M9 medium supplemented with 2% glucose and 0.5% yeast extract. This medium was used to maintain a constant growth rate throughout the two subsequent induction processes. For adaptation, the overnight LB culture of NML was inoculated 2% v/v into minimal medium, supplemented with 1% glycerol. The overnight cultures in minimal medium were used to inoculate 25 mL of the corresponding fresh minimal medium with an initial $OD_{600}$ of 0.08 and then induced when the $OD_{600}$ reached 0.6. Strains NGL and LNML were supplemented with 50 nM of aTc to remove leaky expressed TetR. Strain OLIM was first induced with 4 μM or 40 μM IPTG and maintained under steady state for five cell cycles before 0.4% arabinose was added to induce CAR. Cell cultures were diluted under the same induction conditions periodically to maintain the cells in the exponential growth phase. $OD_{600}$ and fluorescence were measured every 1-2 cell cycles, and cell cultures were collected at each sampling point and stored at -20 °C for FFA quantification. FFA were quantified using a previously published method(Liu *et al*, 2013). Fluorescence and FFA concentration were normalized by cell density as measured by $OD_{600}$, and the data were then normalized by the steady state to obtain the protein and metabolite dynamic curves.

### 4.4.5 Mathematical model and fitting to experimental data

Fitting the model to the experimental data was performed by implementing least-squares non-linear optimization, using the MATLAB Global Optimization Toolbox function MultiStart. All model parameters were left free to be fitted, and the optimization was initiated from thousands of random guesses of the set of parameter values to ensure convergence onto a global solution. Kinetic models, coefficient of determination ($R^2$), and the fitted parameters of OL, OLIM, NGL, NML, and LNML are in the Supplemental Information. The fitted parameters and $R^2$ are summarized in the Supplemental Information Table 4S4. In Fig 4.4C, the parameter value, $k_{tesA}$=77.75 $s^{-1}$, is used for tuning the apparent expression rate of TesA and apparent inhibition constant. Parameters used in Fig 4.4D for tuning the maximum expression rates of TesA and TetR are $k_{tesA}$=230.90 $s^{-1}$, $k_{d,tetR}$=3.85e-8 M, $k_{fadR,\ FFA}$=0.001 M, and $k_2$=138.50. Parameters used in Fig 4.4D for tuning $K_d$ (TetR-$P_{TL}$) and $K_d$ (FadR-$P_{AR2}$) are $r_{TL}$=1e-10 M/s, $r_{AR2}$=1e-10 M/s, $k_{tesA}$=230.90 $s^{-1}$, $k_{fadR,\ FFA}$=0.001 M, and $k_{deg}$=0.0004 $s^{-1}$. All data points are logarithmically spaced in Fig 4.4C & 4.4D. Protein degradation was considered as first-order kinetics and the degradation rate constant was $k_{deg}$=0.0004 $s^{-1}$.

# Acknowledgements

The authors would like to thank Ahmad Mannan for his help in data fitting. This work was supported by Washington University in St. Louis, the National Science Foundation

(MCB1453147), the Human Frontier Science Program (RGY0076/2015), and the International Center for Advanced Renewable Energy and Sustainability (I-CARES).

# References


Akhtar MK, Turner NJ & Jones PR (2013) Carboxylic acid reductase is a versatile enzyme for the conversion of fatty acids into fuels and chemical commodities. *Proc. Natl. Acad. Sci.* **110:** 87–92 Available at: http://www.pnas.org/cgi/doi/10.1073/pnas.1216516110

Bennett MR (2015) Emergent genetic oscillations in a synthetic microbial consortium. **349:**

Berthoumieux S, Jong H de, Baptist G, Pinel C, Ranquet C, Ropers D & Geiselmann J (2013) Shared control of gene expression in bacteria by transcription factors and global physiology of the cell. *Mol. Syst. Biol.* **9:** 634 Available at: http://msb.embopress.org/content/9/1/634%5Cnhttp://msb.embopress.org/content/9/1/634.long%5Cnhttp://msb.embopress.org/content/msb/9/1/634.full.pdf%5Cnhttp://www.ncbi.nlm.nih.gov/pubmed/23340840

Chin CS, Chubukov V, Jolly ER, DeRisi J & Li H (2008) Dynamics and design principles of a basic regulatory architecture controlling metabolic pathways. *PLoS Biol.* **6:** 1343–1356

Chubukov V, Zuleta IA & Li H (2012) Regulatory architecture determines optimal regulation of gene expression in metabolic pathways. *Proc. Natl. Acad. Sci. U. S. A.* **109:** 5127–5132 Available at: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3324031/

Dahl RH, Zhang F, Alonso-Gutierrez J, Baidoo E, Batth TS, Redding-Johanson AM, Petzold CJ, Mukhopadhyay A, Lee TS, Adams PD & Keasling JD (2013) Engineering dynamic pathway regulation using stress-response promoters. *Nat. Biotechnol.* **31:** 1039–46 Available at: http://dx.doi.org/10.1038/nbt.2689

Elowitz MB & Leibler S (2000) A synthetic oscillatory network of transcriptional regulators. *Nature* **403:** 335–338

Fendt S-M, Buescher JM, Rudroff F, Picotti P, Zamboni N & Sauer U (2010) Tradeoff between enzyme and metabolite efficiency maintains metabolic homeostasis upon perturbations in enzyme capacity. *Mol. Syst. Biol.* **6:** 356 Available at: http://eutils.ncbi.nlm.nih.gov/entrez/eutils/elink.fcgi?dbfrom=pubmed&id=20393576&retmode=ref&cmd=prlinks

Gupta A, Reizman IMB, Reisch CR & Prather KLJ (2017) Dynamic regulation of metabolic flux in engineered bacteria using a pathway-independent quorum-sensing circuit. *Nat. Biotechnol.* **3:** Available at: http://www.nature.com/doifinder/10.1038/nbt.3796

Kochanowski K, Sauer U & Chubukov V (2013) Somewhat in control - the role of transcription in regulating microbial metabolic fluxes. *Curr. Opin. Biotechnol.***:** 1–6 Available at: http://www.ncbi.nlm.nih.gov/pubmed/23571096%5Cnhttp://linkinghub.elsevier.com/retrieve/pii/S0958166913000712

Kotte O, Volkmer B, Radzikowski JL & Heinemann M (2015) Phenotypic bistability in Escherichia coli ' s central carbon metabolism Phenotypic bistability in Escherichia coli ' s central carbon metabolism. *Mol. Syst. Biol.* **10:** 1–11 Available at: http://www.ncbi.nlm.nih.gov/pubmed/24987115

Kotte O, Zaugg JB & Heinemann M (2010) Bacterial adaptation through distributed sensing of metabolic fluxes. *Mol. Syst. Biol.* **6:** 355 Available at: http://www.ncbi.nlm.nih.gov/pubmed/20212527

Link H, Kochanowski K & Sauer U (2013) Systematic identification of allosteric protein-metabolite interactions that control enzyme activity in vivo. *Nature.Com* **31:** 357–361 Available at: http://www.nature.com/nbt/journal/vaop/ncurrent/pdf/nbt.2489.pdf%5Cnfile:///Users/niko/Dropbox/Papers2/Files/Untitled-8401.pdf%5Cnhttp://eutils.ncbi.nlm.nih.gov/entrez/eutils/elink.fcgi?dbfrom=pubmed&id=23455438&retmode=ref&cmd=prlinks

Liu D, Xiao Y, Evans BS & Zhang F (2013) Negative Feedback Regulation of Fatty Acid Production Based on a Malonyl-CoA Sensor–Actuator. *ACS Synth. Biol.* **4:** 132–140 Available at: http://dx.doi.org/10.1021/sb400158w

Mukherji S & van Oudenaarden A (2009) Synthetic biology: understanding biological design from synthetic circuits. TL - 10. *Nat. Rev. Genet.* **10 VN-r:** 859–871 Available at: /Users/yurikoharigaya/Documents/ReadCube Media/nrg2697.pdf%5Cnhttp://dx.doi.org/10.1038/nrg2697

Novák B & Tyson JJ (2008) Design principles of biochemical oscillators. *Nat. Rev. Mol. Cell Biol.* **9:** 981–991 Available at: http://www.nature.com/doifinder/10.1038/nrm2530

Paddon CJ & Keasling JD (2014) Semi-synthetic artemisinin: a model for the use of synthetic biology in pharmaceutical development. *Nat. Rev. Microbiol.* **12:** 355–367 Available at: http://dx.doi.org/10.1038/nrmicro3240%5Cn10.1038/nrmicro3240%5Cnhttp://www.nature.com/doifinder/10.1038/nrmicro3240

Peralta-Yahya PP, Ouellet M, Chan R, Mukhopadhyay A, Keasling JD & Lee TS (2011) Identification and microbial production of a terpene-based advanced biofuel. *Nat. Commun.* **2:** 483 Available at: http://dx.doi.org/10.1038/ncomms1494

Peralta-Yahya PP, Zhang F, del Cardayre SB & Keasling JD (2012) Microbial engineering for the production of advanced biofuels. *Nature* **488:** 320–328

Pisithkul T, Patel NM & Amador-Noguez D (2015) Post-translational modifications as key regulators of bacterial metabolic fluxes. *Curr. Opin. Microbiol.* **24:** 29–37 Available at: http://dx.doi.org/10.1016/j.mib.2014.12.006

Ray JCJ, Tabor JJ & Igoshin OA (2011) Non-transcriptional regulatory processes shape transcriptional network dynamics. *Nat. Rev. Microbiol.* **9:** 817–828 Available at: http://www.nature.com/doifinder/10.1038/nrmicro2667

Rosenfeld N, Elowitz MB & Alon U (2002) Negative autoregulation speeds the response times of transcription networks. *J. Mol. Biol.* **323:** 785–793

Shen-Orr SS, Milo R, Mangan S & Alon U (2002) Network motifs in the transcriptional regulation network of Escherichia coli. *Nat. Genet.* **31:** 64–8 Available at: http://www.nature.com/ng/journal/v31/n1/full/ng881.html

Shin D, Lee E-J, Huang H & Groisman EA (2006) A Positive Feedback Loop Promotes Transcription Surge That Jump-Starts Salmonella Virulence Circuit. *Science (80-. ).* **314:** 1607–1609 Available at: http://www.sciencemag.org/cgi/doi/10.1126/science.1134930

Xiao Y, Bowen CH, Liu D & Zhang F (2016) Exploiting nongenetic cell-to-cell variation for enhanced biosynthesis. *Nat. Chem. Biol.* **12:** 339–344 Available at: http://www.nature.com/doifinder/10.1038/nchembio.2046

Xiao Y, Jiang W & Zhang F (2017) Developing a Genetically Encoded, Cross-Species Biosensor for Detecting Ammonium and Regulating Biosynthesis of Cyanophycin.

Xu P, Li L, Zhang F, Stephanopoulos G & Koffas M (2014) Improving fatty acids production by engineering dynamic pathway regulation and metabolic control. *Proc. Natl. Acad. Sci.* **111:** 11299–11304 Available at: http://www.pnas.org/cgi/doi/10.1073/pnas.1406401111

Zaslaver A, Mayo A, Rosenberg R, Bashkin P, Sberro H, Tsalyuk M, Surette M & Alon U (2004) Just-in-time transcription program in metabolic pathways. *Nat. Genet.* **36:** 486–491

Zhang F, Carothers JM & Keasling JD (2012) Design of a dynamic sensor-regulator system for production of chemicals and fuels derived from fatty acids. *Nat. Biotechnol.* **30:** 354–359 Available at: http://www.nature.com/doifinder/10.1038/nbt.2149%5Cnhttp://www.nature.com/nbt/journal/v30/n4/full/nbt.2149.html?WT.ec_id=NBT-201204%5Cnhttp://www.ncbi.nlm.nih.gov/pubmed/22446695

Zhang F, Rodriguez S & Keasling JD (2011) Metabolic engineering of microbial pathways for advanced biofuels production. *Curr. Opin. Biotechnol.* **22:** 775–783 Available at: http://dx.doi.org/10.1016/j.copbio.2011.04.024

# Appendix: Supplementary Information for Chapter 4

**Supplementary Table 4S1** Plasmids used in this study.

| Plasmids | Replication Origin | Overexpressed Operon | Resistance |
|---|---|---|---|
| pB5k-tesA-RFP | BBR1 | $P_{lacUV5}$-*tesA-rfp* | $Kan^R$ |
| pB5k-tesA | BBR1 | $P_{lacUV5}$-*tesA* | $Kan^R$ |
| pB8k-tesA | BBR1 | $P_{BAD}$-*tesA* | $Kan^R$ |
| pA5c-CAR | p15A | $P_{LacUV5}$-*car* | $Cm^R$ |
| pETLa-RFP | ColE1 | $P_{TL}$-*rfp* | $Amp^R$ |
| pETLa-tesA-RFP | ColE1 | $P_{TL}$-*tesA-rfp* | $Amp^R$ |
| pB8k-tetR | BBR1 | $P_{BAD}$-*tetR* | $Kan^R$ |
| pE8a-fadR | ColE1 | $P_{BAD}$-*fadR* | $Amp^R$ |
| pBAR2k-RFP | BBR1 | $P_{AR2}$-*rfp* | $Kan^R$ |
| pETLa-tesA-tetR-RFP | ColE1 | $P_{TL}$-*tesA-tetR-rfp* | $Amp^R$ |
| pEFLa-tesA-RFP | ColE1 | $P_{FL}$-*tesA-rfp* | $Amp^R$ |
| pBAR2k-tetR | BBR1 | $P_{AR2}$-*tetR* | $Kan^R$ |
| pA8c-fadR | p15A | $P_{BAD}$-*fadR* | $Cm^R$ |

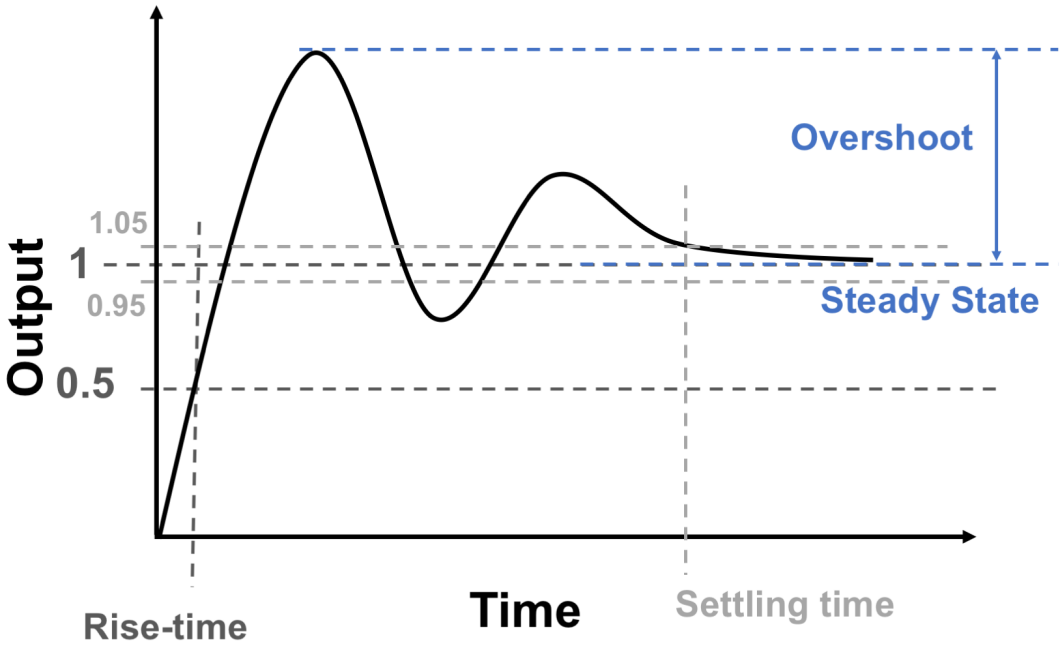
**Supplementary Table 4S2.** Strains used in this study.

| Strains | Relevant genotype |
|---|---|
| DH1 | F– λ– endA1 recA1 relA1 gyrA96 thi-1 glnV44 hsdR17(rK–mK–) |
| DH1(Δ*fadE*) | *E. coli* DH1: Δ*fadE* |
| PTLC | *E. coli* DH1 (Δ*fadE*): pETLa-RFP, pB8k-tetR |
| PFLC | *E. coli* DH1 (Δ*fadE*): pEFLa-RFP, pA8c-fadR |
| PAR2C | *E. coli* DH1 (Δ*fadE*): pBAR2k-RFP, pE8a-fadR |
| OL | *E. coli* DH1 (Δ*fadE*): pB5k-tesA-RFP |
| OLIM | *E. coli* DH1 (Δ*fadE*): pB8k-tesA, pA5c-CAR |
| NGL | *E. coli* DH1 (Δ*fadE*): pETLa-tesA-tetR-RFP |
| NML | *E. coli* DH1 (Δ*fadE*): pEFLa-tesA-RFP, pA8c-fadR |
| LNML | *E. coli* DH1 (Δ*fadE*): pETLa-tesA-RFP, pBAR2k-tetR, pA8c-fadR |

**Supplementary Table 4S3** Sequences of the engineered promoters. The bold sequences represent the -35 and -10 regions. LacI, TetR, and FadR binding sites are colored in red, blue and green lines.

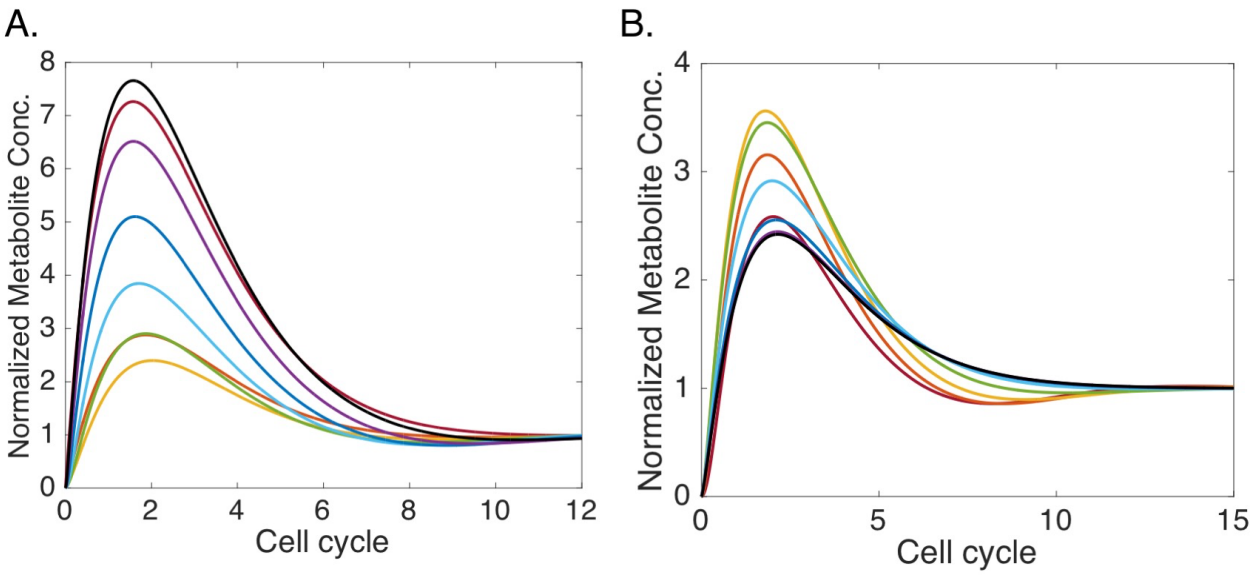
| Promoter | Sequence |
|---|---|
| $P_{TL}$ | CGAAAT**TTGACT**TCCCTATCAGTGATAGAGA**TACT**GTGTGGAATTGTGAGCGGATAACAATT |
| $P_{FL}$ | ATTCCGAACTGATCGGAC**TTGTT**CAGCGTACACGTGTTAGC**TATCCT**GCGTGCAATTGTGAGCGGATAACAATTTTC |
| $P_{AR2}$ | TCAAAAAGAGTG**TTGACT**ATCTGGTACGACCAGAT**GATACT**TAGATTCATTTATGCTTCCGGC |

**Supplementary Figure 4S1. Parameters used in this study to characterize circuit dynamics.** Rise-time is defined as the time needed to reach half of the steady state concentration. Settling time is the time required for a response curve to reach and stay within 5% of the steady state level. Percent overshoot is the ratio of the amount of overshoot to the steady state.

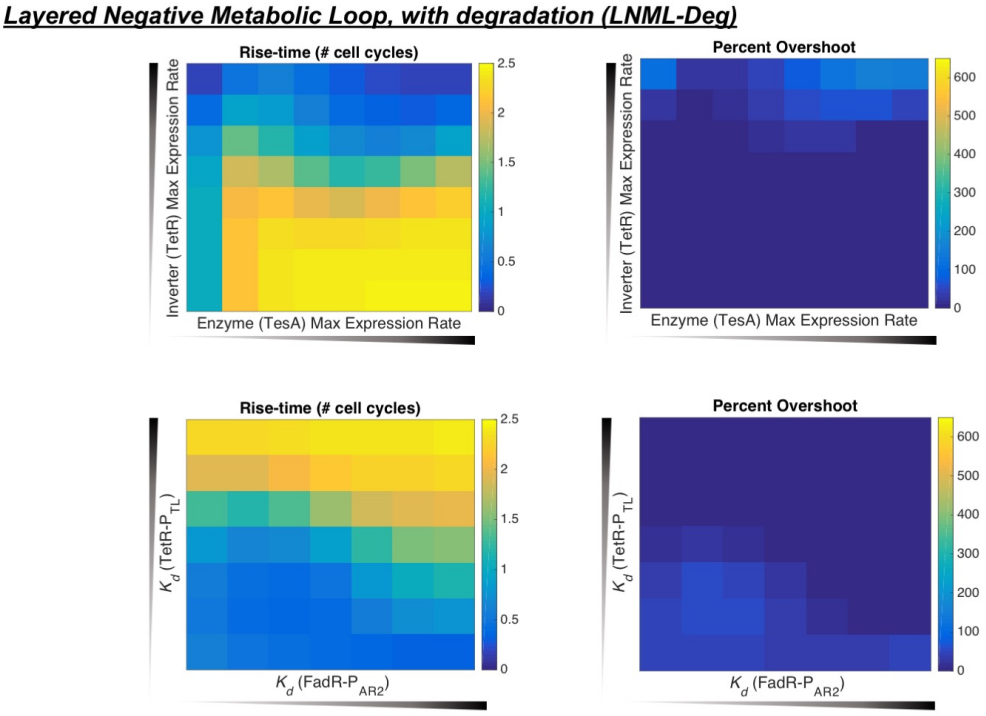



$$\text{Percent Overshoot} = \frac{\text{Overshoot}}{\text{Steady State}} \cdot 100\%$$

**Supplementary Figure 4S2. Metabolic dynamics with A) varying enzyme (TesA) maximum expression rate, and B) varying $K_d$ (FadR-$P_{AR2}$) values.** The value of the parameters increases following the rainbow color sequence, with black for the highest value. The inverter (TetR) max expression rate was fixed at 7.8e-9 M/s in A, and the $K_d$ (TetR-$P_{TL}$) was fixed at 25 nM in B. All the other parameters were the same as the fitted values in Table 4S4.



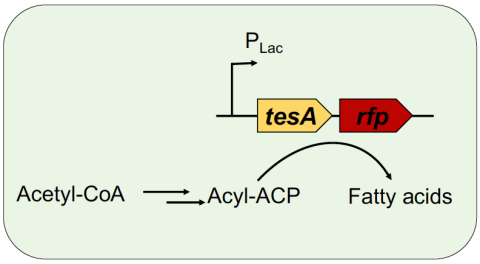
**Supplementary Figure 4S3. Response time and percent overshoot of LNML-Deg under varying tunable circuit parameters.** TetR degradation is considered as first-order kinetics (Supplemental Information M5) and the modeling parameters can be found in the Methods section. The fast degradation of TetR dramatically decreases the percent overshoot.



***Mathematical models of negative feedback circuits***

## M1. Mathematical model of the Open Loop (OL)

The open loop topology is shown below:




Transcription of genes from $P_{Lac}$ promoter at a fixed IPTG concentation can be regarded at a constant rate of $r_{Lac}$. Thus, the expression kinetics of TesA are described by

$$\frac{d[TesA]}{dt} = r_{Lac} - \mu[TesA] \qquad (1)$$

where μ is the specific growth rate of cells in the exponential growth phase.

The kinetics of the free fatty acid (FFA) are described by

$$\frac{d[FFA]}{dt} = \frac{k_{cat} \cdot [TesA] \cdot [S]}{K_m + [S]} - \mu[FFA] \qquad (2)$$

where [S] is the concentration of the precursor acyl-ACPs, and $k_{cat}$ and $K_m$ are the turnover rate and Michaelis constant of TesA. The acyl-ACP concentration is assumed constant in the exponential growth phase. Thus, the FFA kinetics can be simplified to

$$\frac{d[FFA]}{dt} = k_{tesA}[TesA] - \mu[FFA] \qquad (3)$$

where

$$k_{tesA} = \frac{k_{cat} \cdot [S]}{K_m + [S]} \qquad (4)$$

Solving equations (1) and (2), the kinetics of TesA and FFA can be described by

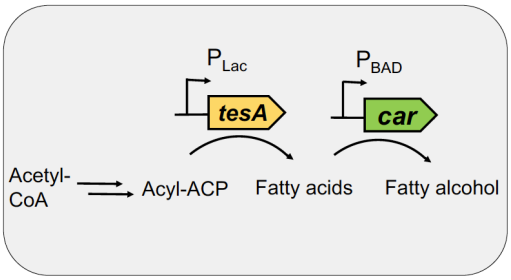
$$[TesA] = [TesA]_{ss} \cdot (1 - \exp(-ln2 * N)) \qquad (5)$$

$$[FFA] = [FFA]_{ss} \cdot (1 - (ln2 * N + 1) \cdot \exp(-ln2 * N)) \qquad (6)$$

Thus, the kinetics of TesA and FFA are independent of the circuit parameters.

## M2. Mathematical Model of the Open Loop Intermediate (OLIM)

The topology of OLIM is shown below:




The kinetics of the FFA concentration are described by

$$\frac{d[FFA]}{dt} = \frac{k_{cat} \cdot [TesA] \cdot [S]}{K_m + [S]} - \mu[FFA] - \frac{k_{cat,CAR} \cdot [CAR] \cdot [FFA]}{K_{m,CAR} + [FFA]} \qquad (7)$$

where [S], [TesA], and [CAR] are the concentrations of acyl-ACPs, TesA, and CAR, respectively. $k_{cat}$ and $K_m$ are the turnover rate and Michaelis constant of TesA, respectively. $k_{cat,CAR}$ and $K_{m,CAR}$ are the turnover rate and Michaelis constant of CAR. The FFA consumption was

simplified into first-order kinetics by assuming $K_{m,CAR} \gg [FFA]$ to minimize the number of model parameters. Thus, the kinetics of FFA concentration can be described as

$$\frac{d[FFA]}{dt} = k_{tesA}[TesA] - \mu[FFA] - k_{CAR}[FFA] \qquad (8)$$

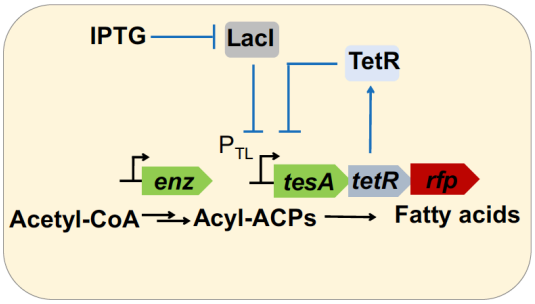
where

$$k_{tesA} = \frac{k_{cat} \cdot [S]}{K_m + [S]} \qquad (4)$$

$$k_{CAR} = \frac{k_{cat,CAR} \cdot [CAR]}{K_{m,CAR}} \qquad (10)$$

## M3. Mathematical model of the Negative Gene Loop (NGL)

The regulatory topology of the NGL is shown below:




With maximum IPTG induction, the activity of the $P_{TL}$ promoter ($A_{TL}$) is regulated by TetR concentration and can be described by

$$A_{TL} = \frac{r_{TL}}{1 + \frac{[TetR]}{K_{d,tetR}}} \qquad (11)$$

where $r_{TL}$ represents the fully activated promoter activity of $P_{TL}$. $K_{d,tetR}$ is the dissociation constant between TetR and $P_{TL}$.

Thus, the expression kinetics of TesA can be described by

$$\frac{d[TesA]}{dt} = \frac{r_{TL}}{1 + \frac{[TetR]}{k_{d,tetR}}} - \mu[TesA] \qquad (12)$$

where $k$ is the specific growth rate of cells in the exponential growth phase.

As in the OL, the kinetics of FFA are described by

$$\frac{d[FFA]}{dt} = k_{tesA}[TesA] - \mu[FFA] \qquad (3)$$

where

$$k_{tesA} = \frac{k_{cat} \cdot [S]}{K_m + [S]} \qquad (4)$$

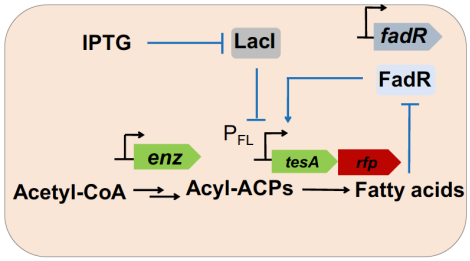
The level of TetR concentration can be described by

$$\frac{d[TetR]}{dt} = \frac{r_{TL,tetR}}{1 + \frac{[TetR]}{k_{d,tetR}}} - \mu[TetR] \qquad (13)$$

where $r_{TL,tetR}$ represents the fully activated synthesis rate of TetR.

## M4. Mathematical model of the Negative Metabolic Loop (NML)

The regulatory topology of the NML is shown below:




With maximum IPTG induction, the activity of the $P_{FL}$ promoter ($A_{FL}$) in response to FadR concentration [FadR] is described by

$$A_{FL} = \frac{r_{FL}[FadR]_f}{[FadR]_f + k_{d,fadR}} \qquad (14)$$

where $r_{FL}$ represents the fully activated promoter activity of $P_{FL}$ and $k_{d,fadR}$ is the dissociation constant between FadR and $P_{FL}$. At equilibrium, the free FadR concentration $[FadR]_f$ can be described by

$$[FadR]_f = \frac{k_{fadR,FFA}}{k_{fadR,FFA} + [FFA]}[FadR]_t \qquad (15)$$

where $k_{fadR,FFA}$ represents the apparent dissociation constant between FFA and FadR. $[FadR]_t$ is the total concentration of FadR.

Combining equations (14) and (15), we obtain

$$A_{FL} = \frac{r'_{FL}}{1 + \frac{[FFA]}{ki}} \qquad (16)$$

where

$$r'_{FL} = r_{FL}\frac{[FadR]_t}{[FadR]_t + k_{d,fadR}} \quad (17)$$

$$k_i = \frac{k_{d,fadR} + [FadR]t}{k_{d,fadR}} k_{fadR,FFA} \quad (18)$$

Thus, the expression kinetics of TesA are described by

$$\frac{d[TesA]}{dt} = \frac{r'_{FL}}{1 + \frac{[FFA]}{ki}} - \mu[TesA] \quad (19)$$

As in the OL, the FFA kinetics are described by

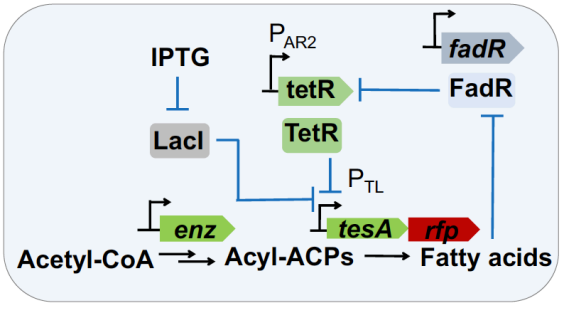
$$\frac{d[FFA]}{dt} = k_{tesA}[TesA] - \mu[FFA] \quad (3)$$

where

$$k_{tesA} = \frac{k_{cat} \cdot [S]}{K_m + [S]} \quad (4)$$

## M5. Mathematical Model of the Layered Negative Metabolic Loop (LNML).

The regulatory topology of the LNML is shown below:




As in the NGL, the activity of $P_{TL}$ promoter ($A_{TL}$) in response to TetR concentration [TetR] can be described as

$$A_{TL} = \frac{r_{TL}}{1 + \frac{[TetR]}{k_{d,tetR}}} \qquad (11)$$

The expression kinetics of TesA can be described by

$$\frac{d[TesA]}{dt} = \frac{r_{TL}}{1 + \frac{[TetR]}{k_{d,tetR}}} - \mu[TesA] \qquad (12)$$

The activity of $P_{AR2}$ ($A_{AR2}$) can be described by

$$A_{AR2} = \frac{r_{AR2}}{1 + \frac{[FadR]_f}{k_{d,fadR}}} \qquad (20)$$

where $r_{AR2}$ represents the fully induced promoter activity of the $P_{AR2}$ promoter. At equilibrium, $[FadR]_f$ can be described by

$$[FadR]_f = \frac{[FadR]_t}{1 + [FFA]/k_{fadR,FFA}} \qquad (21)$$

Combining equations (11) and (12), the activity of $P_{AR2}$ can be described by

$$A_{AR2} = \frac{r_{AR2}}{1 + \frac{[FadR]_t}{k_{d,fadR} \cdot (1 + \frac{FFA}{k_{fadR,FFA}})}} \qquad (22)$$

Thus, the expression kinetics of TetR are described by

$$\frac{d[TetR]}{dt} = \frac{r_{AR2}}{1 + \frac{[FadR]_t}{k_{d,fadR} \cdot (1 + \frac{FFA}{k_{fadR,FFA}})}} - \mu[TetR] \qquad (23)$$

Denote

$$k_2 = \frac{[FadR]_t}{k_{d,fadR}} \qquad (24)$$

Then, combining equations (23) and (24), the expression kinetics of TetR are described by

$$\frac{d[TetR]}{dt} = \frac{r_{AR2}}{1 + \frac{k_2}{(1 + \frac{FFA}{k_{fadR,FFA}})}} - \mu[TetR] \qquad (25)$$

As in OL, FFA kinetics can be described by

$$\frac{d[FFA]}{dt} = k_{tesA}[TesA] - \mu[FFA] \qquad (3)$$

where

$$k_{tesA} = \frac{k_{cat} \cdot [S]}{K_m + [S]} \qquad (4)$$

To model the LNML-Deg, the degradation rate of the repressor was considered as first order kinetics with a rate constant of $k_{deg}$. Thus, the expression kinetics of TetR are described by

$$\frac{d[TetR]}{dt} = \frac{r_{AR2}}{1 + \frac{[FadR]_t}{k_{d,fadR} \cdot (1 + \frac{FFA}{k_{fadR,FFA}})}} - \mu[TetR] - k_{deg}[TetR] \qquad (26)$$

***Model Parameterization***

All model parameters were obtained by either direct experimental measurements or by fitting to protein and metabolite kinetics data. Data fitting was performed using the MultiStart algorithm in MATLAB R2015b. The lower and the upper bounds of the protein synthesis rate were set between $10^{-11}$ M/s and $10^{-8}$ M/s to cover the typical range of overexpressed protein synthesis rates. $k_{tesA}$ was set between 1 $s^{-1}$ and 1000 $s^{-1}$ to cover a broad range of protein turnover rates(Zhang, 2011). The dissociation constant between TetR and $P_{TL}$, $k_{d,tetR}$, was fitted within a range between 3e-8 M and 4e-8 M, considering that the concentration of $P_{TL}$ varies between 6e-8 M and 8e-8

M(Lee *et al*, 2011). The apparent dissociation constant between FFA to $P_{FL}$ was fitted with upper and lower bounds set to be 0 M to 0.12 M, with 0.12 M being the calculated maximum intracellular FFA concentration produced in the OLIM. The leaky expressed FadR was estimated to be between 20 nM to 70 nM(TANIGUCHI, 2011), and $k_{d,\ fadR}$ was reported to be 0.2 nM(Van Aalten, D.M., DiRusso, C.C. and Knudsen, 2001). Thus the bounds of $k_2$ were set between 40 to 140. A table of parameters used in this work is listed below.

**Supplementary Table 4S4** Circuit parameters determined by fitting experimental data to the model and the $R^2$ of the fitting.

| | Parameter | Value | Unit | $R^2$ (Metabolite) | $R^2$ (Protein) |
|---|---|---|---|---|---|
| **OL** | $[Enzyme]_{ss}$ | 764 | A.U. | 0.98 | 0.94 |
| | $[Metabolite]_{ss}$ | 43.13 | mg/L | | |
| **OLIM** | $k_{CAR}$ | 4.78e-4 (OLIM_hc) | $s^{-1}$ | 0.99 | N.A. |
| | | 2.83e-4 (OLIM_lc) | | 0.99 | N.A. |
| **NGL** | $r_{TL}$ | 2.09e-11 | M/s | 0.97 | 0.91 |
| | $k_{tesA}$ | 105.25 | $s^{-1}$ | | |
| | $r_{TL,tetR}$ | 2.18e-11 | M/s | | |
| | $k_{d,tetR}$ | 3.0e-8 | M | | |
| **NML** | $r'_{FL}$ | 1.00e-11 | M/s | 0.93 | 0.92 |
| | $k_{tesA}$ | 77.75 | $s^{-1}$ | | |
| | $k_i$ | 0.079 | M | | |
| **LNML** | $r_{TL}$ | 1.66e-11 | M/s | 0.62 | 0.82 |
| | $k_{tesA}$ | 230.90 | $s^{-1}$ | | |
| | $r_{AR2}$ | 2.21e-9 | M/s | | |
| | $k_{d,tetR}$ | 3.85e-8 | M | | |

| | | | | | |
|---|---|---|---|---|---|
| | $k_{fadR,FFA}$ | 0.001 | M | | |
| | $k_2$ | 138.50 | 1 | | |

# References

Van Aalten, D.M., DiRusso, C.C. and Knudsen J (2001) The structural basis of acyl coenzyme A-dependentregulation of the transcription factor FadR. *EMBO J.* **20:** 2041–2050

Lee T, Krupa RA, Zhang F, Hajimorad M, Holtz WJ, Prasad N, Lee S & Keasling JD (2011) BglBrick vectors and datasheets: A synthetic biology platform for gene expression. *J. Biol. Eng.* **5:** 12

TANIGUCHI Y (2011) Quantifying E. coli Proteome and Transcriptome with Single-molecule Sensitivity in Single Cells. *Seibutsu Butsuri* **51:** 136–137

Zhang YHP (2011) Substrate channeling and enzyme complexes for biotechnological applications. *Biotechnol. Adv.* **29:** 715–725

# Chapter 5 Conclusions and Future Directions

## 5.1 Conclusions

This dissertation introduced metabolite biosensors, with a focus on transcription factor-based metabolite biosensors, genetic circuits, and their applications in metabolic engineering. The following chapters centered on our efforts to develop design principles for metabolite biosensors (Chapter 2), to engineer genetic circuits to regulate metabolic pathways (Chapter 3), and to study the effects of regulation architectures and the associated biochemical parameters on metabolite dynamics (Chapter 4). All together, these contribute to rational design and tuning of biosensors and genetic circuits for metabolic engineering applications.

Based on our model analysis and experimental results, we revealed that the metabolite response threshold and dynamic range are inherently coupled and further identified a design constraint whereby a maximal achievable dynamic range is expected when tuning the binding affinities between the transcription factors and their promoters. We then developed rational design strategies for orthogonal control of the biosensor response threshold and dynamic range, by tuning the binding affinity between the transcription factor and the metabolite and the dynamic range of the promoter activity. This work directly provides quantitative guidelines for biosensor design, which

facilitate diverse applications in strain screening, dynamic pathway control, and real-time metabolism monitoring.

In Chapter 3, we developed a malonyl-CoA sensor-actuator and applied it to build a negative feedback circuit to regulate an engineered free fatty acid producing pathway. The circuit dynamically controls the expression of acetyl-CoA carboxylase (ACC) in response to intracellular malonyl-CoA levels, therefore alleviating toxicity from ACC overexpression and also recovering cell growth. As a result, fatty acid production was improved. This work directly contributes to the toolbox of synthetic biology by developing a biosensor for malonyl-CoA, which is an important precursor for many valuable products. In addition, this study documents the first example where metabolite-mediated dynamic down-regulation of enzyme expression is applied to enhance pathway productivities.

The understanding of metabolic dynamics is not only important to identify the optimal regulation architecture to engineer synthetic regulatory circuits, but also crucial to understand their benefits to cell fitness from a systems biology perspective. In Chapter 4, we found that upon induction of a metabolic pathway, the metabolite dynamics was ultimately limited by enzyme expression dynamics. We further constructed negative feedback circuits with three regulation architectures and studied the effects of the circuit architectures and parameters on metabolic dynamics. The study revealed that metabolic dynamics can be dramatically accelerated by gene

regulation, which potentially can be explored for applications in designing synthetic circuits to improve the productivities of engineered metabolic pathways.

## 5.2 Recommendations for Future Directions

Based on this dissertation, three research directions are recommended for future work. These include 1) Developing engineering strategies to efficiently expand the variety of metabolites that can be sensed. 2) Exploring how negative feedback circuits respond to environmental perturbations and regulate cellular heterogeneities. 3) Developing general design principles to engineer synthetic circuits to regulate metabolic pathways.

### 5.2.1 Developing engineering strategies to efficiently expand the variety of metabolites that can be sensed

Biosensors have demonstrated valuable applications in metabolic engineering, including strain screening and dynamic pathway regulation. However, their applications are strongly limited by the small number of available biosensors. The design of metabolite biosensors has largely relied on naturally available genetic parts (e.g. transcription factors, RNAs). While some studies have expanded the specificities of transcription factors to detect metabolites of interest where no natural sensors exist, it remains a challenge to develop efficient and general engineering strategies to sense any metabolite of interest(Liu et al., 2015). Therefore, engineering strategies that can link the existing interactions between metabolites of interest and proteins to tunable outputs will be particularly useful in further expanding the biosensor toolbox(Feng et al., 2015).

### 5.2.2 Exploring how negative feedback circuits respond to environmental perturbations and regulate cellular heterogeneities

This dissertation has explored the effects of negative feedback circuits on metabolic dynamics. Our initial modelling results suggest that under environmental perturbations, negative feedback circuits can accelerate cellular response by bringing the metabolite concentration quickly back to its steady state level. This aspect can be further explored to engineer more metabolically stable systems with steady outputs under even large environmental stimuli.

In addition, it has been found that protein expression noise is decreased by negative autoregulation(Beckskei and Serrano, 2000). However, the effects of negative feedback circuits with different architectures (as in Chapter 4) on protein noise remain largely unknown. In addition, the regulation of metabolite noise has been barely explored. It is possible that a decrease in protein noise (by negative feedback circuits) may lead to a more homogenous metabolite distribution. Thus, it will be interesting to explore whether and how the negative feedback circuits considered in Chapter 4 affect protein and metabolite noise.

### 5.2.3 Developing general design principles to engineer synthetic circuits to regulate metabolic pathways

Synthetic regulatory circuits have proved useful to improve production in a variety of metabolic pathways(Dahl et al., 2013; Gupta et al., 2017; Liu et al., 2013; Zhang et al., 2012). Given a metabolic pathway of interest, the number of possible control topologies is tremendous, making it

impractical to test production under all possible control topologies and circuit parameters. Unfortunately, there is little knowledge on the effect of each control topology on pathway activity. As a result, the current construction of synthetic circuits relies mostly on intuition and empirical guesses to choose control topologies, and researchers have to test a variety of parameters (e.g., promoter and RBS strengths) to identify the optimal condition, which is labor-intensive and time-consuming. Thus, it will be essential to combine computational simulations with experimental approaches to predict the optimal control topologies and parameters to minimize the engineering efforts(Eckert and Trinh, 2016). To do this, mathematical models that consider the effects of overexpression of metabolic pathways on cell growth and protein expression need to be incorporated in the kinetic models. In addition, the benefits of dynamic regulation in saving carbon and energy may also need to be considered. Mathematical analysis can analyze productivities for pathways with different bottleneck steps and various regulation architectures, which will provide guidelines for rational design of genetic circuits to regulate a variety of metabolic pathways.

# References


Beckskei, A., Serrano, L., 2000. Engineering stability in gene networks by autoregulation. Nature 405, 590–593. doi:10.1038/35014651

Dahl, R.H., Zhang, F., Alonso-Gutierrez, J., Baidoo, E., Batth, T.S., Redding-Johanson, A.M., Petzold, C.J., Mukhopadhyay, A., Lee, T.S., Adams, P.D., Keasling, J.D., 2013. Engineering dynamic pathway regulation using stress-response promoters. Nat. Biotechnol. 31, 1039–46. doi:10.1038/nbt.2689

Eckert, C.A., Trinh, C.T., 2016. Biotechnology for Biofuel Production and Optimization. Elsevier.

Feng, J., Jester, B.W., Tinberg, C.E., Mandell, D.J., Antunes, M.S., Chari, R., Morey, K.J., Rios, X., Medford, J.I., Church, G.M., Fields, S., Baker, D., 2015. A general strategy to construct small molecule biosensors in eukaryotes 1–23. doi:10.7554/eLife.10606

Gupta, A., Reizman, I.M.B., Reisch, C.R., Prather, K.L.J., 2017. Dynamic regulation of metabolic flux in engineered bacteria using a pathway-independent quorum-sensing circuit. Nat. Biotechnol. 3. doi:10.1038/nbt.3796

Liu, D., Evans, T., Zhang, F., 2015. Applications and advances of metabolite biosensors for metabolic engineering. Metab. Eng. 31, 35–43. doi:10.1016/j.ymben.2015.06.008

Liu, D., Xiao, Y., Evans, B.S., Zhang, F., 2013. Negative Feedback Regulation of Fatty Acid Production Based on a Malonyl-CoA Sensor–Actuator. ACS Synth. Biol. 4, 132–140. doi:10.1021/sb400158w

Zhang, F., Carothers, J.M., Keasling, J.D., 2012. Design of a dynamic sensor-regulator system for production of chemicals and fuels derived from fatty acids. Nat. Biotechnol. 30, 354–359. doi:10.1038/nbt.2149

# Appendix Chapter 1: Enhancing Fatty Acid Production in Escherichia coli by *Vitreoscilla* Hemoglobin Overexpression

Note: This chapter contains text and figures from the published paper (Liu, Di, et al. "Enhancing fatty acid production in Escherichia coli by *Vitreoscilla* hemoglobin overexpression." *Biotechnology and bioengineering*114.2 (2017): 463-467).

## Abstract

Our recent $^{13}$C-metabolic flux analysis ($^{13}$C-MFA) study indicates that energy metabolism becomes a rate-limiting factor for fatty acid overproduction in *E. coli* strains (after "Push-Pull-Block" based genetic modifications). To resolve this bottleneck, *Vitreoscilla* Hemoglobin (VHb, a membrane protein facilitating $O_2$ transport) was introduced into a fatty-acid-producing strain to promote oxygen supply and energy metabolism. The resulting strain, FAV50, achieved 70% percent higher fatty acid titer than the parent strain in shake tube cultures. In high cell-density bioreactor fermentations, FAV50 achieved free fatty acids at a titer of 7.02 g/L (51% of the theoretical yield). In addition to "Push-Pull-Block-Power" strategies, our experiments and flux balance analysis also revealed the fatty acid over-producing strain is sensitive to metabolic burden and oxygen influx, and thus a careful evaluation of the cost-benefit tradeoff with the guidance of fluxome analysis will be fundamental for the rational design of synthetic biology strains.

# 1.1 Introduction

Synthetic biology (SynBio) can produce a broad scope of products, from biofuels to pharmaceutical chemicals, through gene knockouts and heterologous enzyme overexpression. However, extensive pathway modifications may impose considerable burdens on cell metabolism (Glick 1995; Wu *et al*. 2016). With increased steps of genetic manipulation, the metabolic burden from new genetic parts may limit cell production performance and cause metabolic shifts. For example, biosynthesis of fatty acids or related compounds has been a hot field during recent years (Jones *et al*. 2015). SynBio follows Push-Pull-Block strategies to direct carbon flux towards free fatty acids, including introduction of heterogeneous enzymes and knocking out degradation pathways (Lu *et al*. 2008), reversal of a degradation pathway (Dellomonaco *et al*. 2011), engineering regulators to boost pathway activities (Zhang *et al*. 2012b), and creating sensor regulator systems to control biosynthesis fluxes (Xu *et al*. 2014; Zhang *et al*. 2012a).

Nevertheless, the combination of SynBio approaches still cannot achieve production metrics to meet industrial manufacturing needs (Van Dien 2013). To address this problem, we performed $^{13}$C-MFA of an engineered strain and revealed high ATP consumption for cell maintenance during fatty acid over-production (He *et al*. 2014). This intracellular energy crisis may become worse when the $O_2$ supply is insufficient during large-scale fermentations. Hence, we propose to promote cell energy metabolism by introducing *Vitreoscilla* hemoglobin (VHb) into the hosts. VHb is a type of soluble protein that can bind $O_2$ at low concentrations and improve the efficiency of bacterial

aerobic respiration (Dikshit and Webster 1988; Khosla and Bailey 1988a; Khosla and Bailey 1988b). The ability of wild-type VHb (encoded by *vhb*) to facilitate $O_2$ uptake and biomass growth can be further enhanced by introducing point mutations (amino acid substitutions) (Andersson *et al*. 2000). Hemoglobin has demonstrated its effectiveness in improving SynBio microbial fermentations. For example, DuPont has successfully employed different bacterial hemoglobin genes to increase carotenoid production by microbial host cells (Cheng *et al*. 2007).

# 1.2 Results

## 1.2.1 Engineering VHb to enhance oxygen uptake in the fatty acid biosynthetic pathway

In this work, we chose an *E. coli* DH1 strain with *fadE* knockout (*fadE* encodes an enzyme in the fatty acid β-oxidation pathway) as the host (DH1Δ*fadE* strain) (Steen *et al*. 2010). The control strain (denoted as FA0, Table 1) for free fatty acid production carries a plasmid with *tesA* and *fadR* overexpression, and its central metabolism has been investigated by $^{13}$C-MFA in our previous work (He *et al*. 2014). Based on the control strain, we inserted a wild-type *vhb* gene and its two mutants (*vhb20* and *vhb50*) into the same plasmid to generate three strains, FAV1, FAV20, and FAV50, respectively (Table 1). Under M9 minimal medium and shake tube conditions, only FAV50 (expressing VHb50, created by His36→Arg and Gln66→Arg in wild type VHb) showed significantly enhanced cell growth (as shown in Fig. 1a).

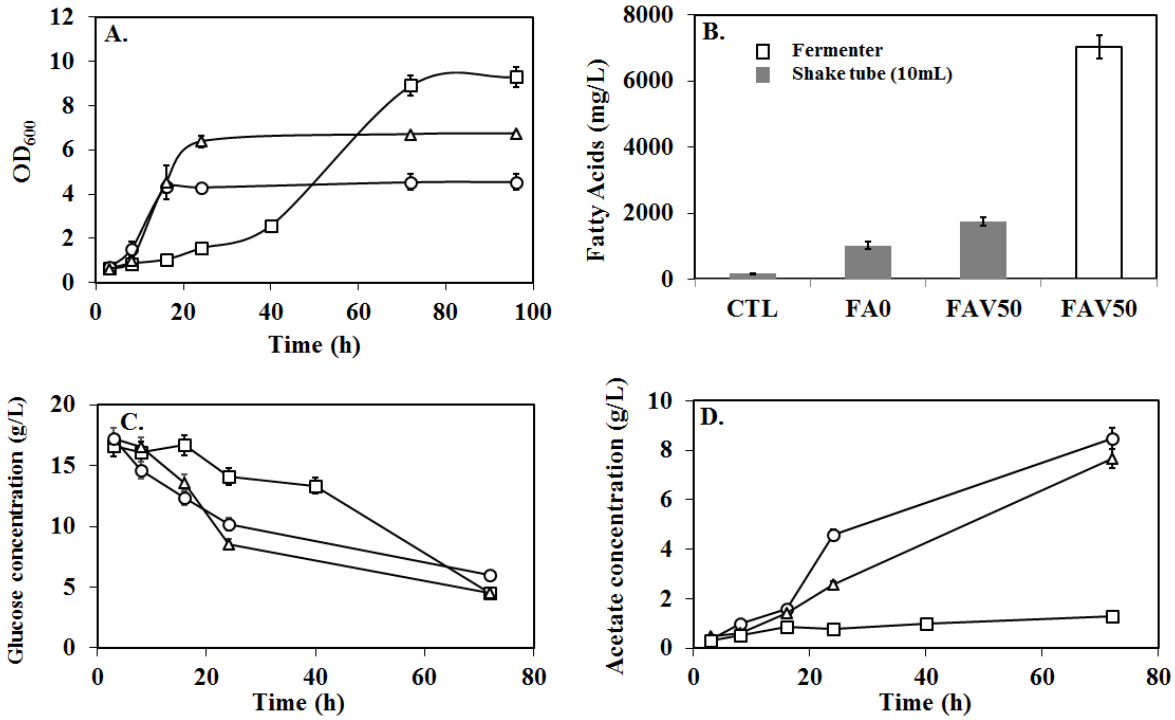



**Figure 1.1 Cell growth and fatty acid production from shake tube cultures (10 mL culture to create oxygen limited conditions).** (a) Growth curve, (b) fatty acid production, (c) glucose consumption, and (d) acetate production of FAV50, FA0, and CTL strains. CTL, the strain carrying an empty plasmid; FA0, the strain carrying a plasmid with tesA and fadR overexpression; FAV50, the strain carrying a plasmid with tesA, fadR and vhb50 overexpression. All strains have a DH1(△ *fadE*) background. Square, FAV50; triangle, FAV0; circle, CTL.

Subsequently, we examined the influence of VHb50 on fatty acid production (Fig. 1b). The strains were cultured in minimal medium in shake tubes with 10 mL culture. The control strain CTL (without *tesA* gene in the plasmid, Table 1) demonstrated a growth rate similar to that of strain

FA0 (overexpressing only *tesA* and *fadR*, Table 1) in the exponential phase. However, CTL culture produced a significant amount of acetate and entered the stationary phase earlier. Strain FAV50 (with *vhb50*, *tesA* and *fadR* gene, Table 1) grew slower at the beginning, but accumulated 30% more biomass and 70% more free fatty acids with a similar consumption of glucose, compared with FA0 (Fig. 1c). Moreover, *vhb50* overexpression significantly reduced acetate secretions (Fig. 1d). These observations indicate the promotion of energy metabolism can reduce waste product synthesis and improve product yields.

**Table 1. Strains and plasmids used in this study**

| Strains | Plasmids | Replication Origin | Overexpressed Operon | Resistance | References |
|---|---|---|---|---|---|
| FAV50 | pA5c-tesA-VHB50-8fadR | p15A | $P_{lacuv5}$-*tesA-vhb50*, $P_{BAD}$-*fadR* | $Cm^R$ | This study |
| FAV20 | pA5c-tesA-VHB20-8fadR | p15A | $P_{lacuv5}$-*tesA-vhb50*, $P_{BAD}$-*fadR* | $Cm^R$ | This study |
| FAV1 | pA5c-tesA-VHB-8fadR | p15A | $P_{lacuv5}$-*tesA-vhb*, $P_{BAD}$-*fadR* | $Cm^R$ | This study |
| FA0 | pA5c-tesA-8fadR | p15A | $P_{lacuv5}$-*tesA*, $P_{BAD}$-*fadR* | $Cm^R$ | Zhang *et al.* 2012b |
| CTL | pA5c-0 | p15A | $P_{lacuv5}$-*none* | $Cm^R$ | This study |
| FAC | pA5c-tesA | p15A | $P_{lacuv5}$-*tesA* | $Cm^R$ | This study |
| FAR1 | pA5c-tesA-RFP-8fadR | p15A | $P_{lacuv5}$-*tesA-rfp*, $P_{BAD}$-*fadR* | $Cm^R$ | This study |
| FAR2 | pA5c-tesA, pB5k-RFP | p15A, BBR1 | $P_{lacuv5}$-*tesA*, $P_{lacuv5}$-*rfp* | $Cm^R$, $Kan^R$ | This study |

During industrial fermentation, oxygen supply always becomes insufficient when cell density

reaches a high level (Garcia-Ochoa and Gomez 2009). To test the performance of VHb under high cell density, we performed semi-batch fermentations on FAV50 in a 1-Liter bioreactor. The culture reached an $OD_{600}$ of over 50, and the final titer of free fatty acids reached 7.04 g/L, with a yield of 0.173 g FA/g glucose (~51% of the theoretical yield) after two days of semi-batch fermentation. In contrast, overexpression of tesA and FadR in DH1(ΔfadE) (without using VHb) produced only 4.8 g/L fatty acids in similar semi-fed batch fermentations (Xiao et al., 2016). The titer and yield of our VHb strain is comparable with recent reports on total fatty acid production via systematic modular optimization (Xu *et al*. 2013) or via a dynamic sensor regulatory system (Xu *et al*. 2014). These experimental results agree with the $^{13}$C-MFA prediction that energy metabolism is one of the key factors limiting fatty acid production.

### 1.2.2 Tradeoff between metabolic burden and benefits in protein overexpression

We also performed other sets of experiments with FAV50 under better aeration conditions (in a shake tube with 5 mL culture or in a baffled shake flask), and observed less significant improvement in biomass growth or fatty acid production than control strain FA0. Moreover, in shake tube cultures the fatty acid titers for FAV1 and FAV20 were lower than in the control strain FA0. This observation indicates that the metabolic burden from VHb expression may offset the benefits from increased $O_2$ transfer, while the function of VHb is significant only under micro-aerobic conditions (Frey and Kallio 2003). To investigate the impact of metabolic burden caused

by protein overexpression on fatty acid production, we replaced the *vhb*50 gene in FAV50 with a gene encoding red fluorescent protein (RFP), thus generating the strain FAR1. Thus, the expression level of VHb is comparable with the RFP level in FAR1. RFP was chosen here for several reasons. First, RFP (25-30 kDa) is relatively small, comparable to VHb. Second, expression of RFP does not directly interfere with cell physiology. Moreover, the expression level of RFP can be readily monitored by fluorescence measurement. Specifically, two strains, FAC (overexpressing the *tesA* gene, Table 1) and FAR1 (overexpressing the *tesA* and *rfp* genes, Table 1), were cultured in M9 medium. Figure 2 shows the final titers of biomass growth and fatty acid production at 72 h post induction. Decreased fatty acid titer for FAR1 indicated that overexpression of even a non-toxic, small-size protein can lead to a significant impact on fatty acid productivity. To further explore the metabolic burden of protein overexpression, we increased the RFP expression level by cloning *rfp* under a higher copy number plasmid, generating the strain FAR2 (Table 1), which has a five times higher RFP expression than FAR1 (Fig. 2c). A dramatic decrease in fatty acid production and cell growth was observed as expected, which further validated the impact of the metabolic burden from heterogeneous gene overexpression. To sum up, our experiments demonstrated various levels of trade-offs between the metabolic burden caused by genetic modifications and the benefits from engineered components (enhanced oxygen flux in this case).

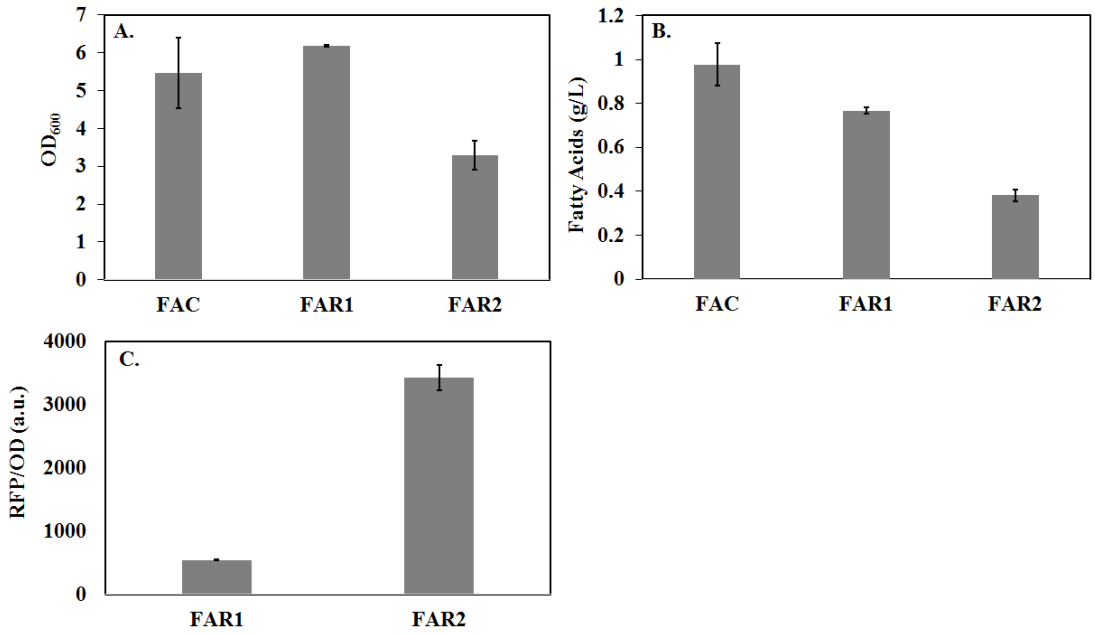



**Figure 2. (a) Biomass growth and (b) fatty acid production of FAC, FAR1, and FAR2 at 72 h post induction**. **c) Relative RFP expression level of FAR1 and FAR2.** No significant biomass growth or fatty acid production was observed in FAV50 under sufficient aeration conditions. FAC, FAR1, and FAR2 are strains expressing *tesA*, with no, medium and high levels of RFP expression, respectively.

### 1.2.3 Effect of oxygen uptake flux and maintenance energy on fatty acid yield

SynBio allows the assembly of multiple genetic components in a recombinant host. However, the capability of cell hosts for handling metabolic burden is still hard to be quantified (Wu *et al*. 2016). In particular, cell energy metabolism has a limited ability to generate ATP. ATP is consumed not only for biomass growth and product synthesis, but also for maintenance of SynBio components (such as plasmid synthesis and enzyme overexpression). The shortage of ATP may lead to

undesirable metabolic shifts under sub-optimal cultivations (Wu *et al*. 2015). To illustrate the effects of oxygen uptake flux and metabolic burden from ATP maintenance loss on fatty acid yields, a genome-scale model (Monk *et al*. 2013) was employed to simulate cellular physiologies (Fig. 3). An apparent trend was that with a decrease in oxygen flux and an increase in ATP maintenance cost (representing metabolic burden) to certain levels, the yield of fatty acids dropped sharply off a "cliff", approaching no production (note: the cliff gets steeper as oxygen fluxes decrease). In general, cell metabolism can afford certain metabolic burden without significant decrease of production yield. However, when cells have a higher metabolic burden, the impact of the oxygen supply becomes more significant on fatty acid yields. A small change in oxygen influx may considerably decrease the fatty acid yield (from the blue star to the red star in Fig. 3), which is similar to the case for an increase in maintenance energy (metabolic burden). Moreover, when cell physiological status is located on the "metabolic cliff", its fatty acid production will be highly sensitive to metabolic burdens and the oxygen supply, reducing the reproducibility of strain performance.

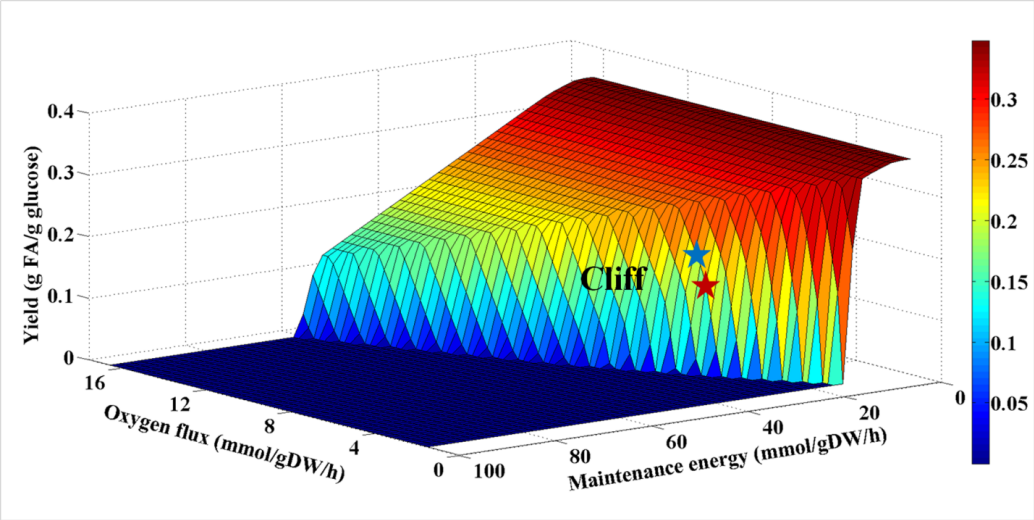



**Figure 3. Effect of oxygen uptake flux and maintenance energy (i.e., ATP consumption for cell maintenance) on fatty acid yield.** Color is used here to visualize the maximal yield of fatty acids (deep blue: 0 → 0.348, deep red) under different conditions (oxygen flux and maintenance energy). To illustrate how a minor difference in oxygen uptake flux abruptly affects the yield of fatty acids at a certain point (cliff), stars represent the status (oxygen flux and maintenance energy) of strains on the cliff slope. Red and blue stars represent two strains with the same maintenance energy and different oxygen uptake rates. The blue star has a better oxygen uptake flux and higher fatty acid yield. The default FBA parameters and constraints are taken from the paper Monk *et al*. 2013). The FBA model assumes a glucose influx, $v_{glucose}$, of 4.5 mmol/gDW/h, a growth rate, μ, of 0.04 $h^{-1}$, and a P/O ratio of 1.5.

# 1.3 Conclusions and Discussions

In summary, this study demonstrates that the introduction of VHb can boost energy metabolism, resulting in enhanced biomass growth and fatty acid titer. The negative effects of metabolic burden on fatty acid production have also been studied by both experiments and FBA

simulation, suggesting that SynBio strategies can achieve expected enhancements only if the benefits outweigh metabolic burden. This work expands SynBio strategies for strain development from Push-Pull-Block to Push-Pull-Block-Power. Moreover, this study illustrates that fluxomics studies can provide valuable guidelines for the SynBio and ME communities. Especially, fluxomes can reveal metabolic burdens and the cell energy metabolism, allowing engineers to properly allocate cell resources during strain development (Wu *et al*. 2016).

# 1.4 Methods

## 1.4.1 Chemicals and strains.

All chemicals were reagent grade and purchased from either Sigma-Aldrich (St. Louis, MO, USA) or Fisher Scientific (Pittsburgh, PA, USA) unless otherwise noted. Restriction enzymes, Phusion DNA polymerase, and T4 ligase were from New England Biolabs (Ipswich, MA, USA). The DNA Purification kit, Gel Recovery kit, and Miniprep kit were from Promega (Madison, WI, USA). The DNA sequences of *vhb* and its mutant *vhb20* and *vhb50* were based on a previous report (Andersson *et al*. 2000). All genes were synthesized by GenScript Inc. (Piscataway, NJ, USA) and cloned into pUC57 vector. *E. coli* DH10B strain was used for plasmid manipulation. The *fadE* knockout *E. coli* DH1 strain and the plasmid pA58c-TR were from Dr. Fuzhong Zhang's lab.

### 1.4.2 Plasmid construction.

Primers used in this study were synthesized by Integrated DNA Technologies (Coralville, IA, USA). Plasmids pA5c-tesA-VHB-8fadR, pA5c-tesA-VHB20-8fadR, and pA5c-tesA-VHB50-8fadR were constructed by inserting *vhb*, *vhb20*, and *vhb50* downstream of *tesA* in plasmid pA58c-TR, respectively. To create the FAC strain, *tesA* was cloned under the control of a pLacUV5 promoter, giving pA5c-tesA. To create the FAR1 strain, *rfp* was cloned downstream of *tesA* in plasmid pA58c-TR. To create the FAR2 strain, *rfp* was cloned into the pB5k vector to construct pB5k-RFP plasmid, which was co-transformed with pA5c-tesA (Lee *et al*. 2011). The DNA sequences of all constructed plasmids were validated by sequencing in the Genome Center at Washington University School of Medicine.

### 1.4.3 Medium and culture conditions.

A modified M9 minimal medium supplied with 2% glucose and appropriate antibiotics was used in this study (Liu *et al*. 2015). In fatty acid production experiments, all strains were first inoculated into LB medium with appropriate antibiotics. The overnight culture was inoculated 2% v/v into minimal medium containing appropriate antibiotics for adaptation. The overnight minimal medium culture was used to inoculate a 10 mL (shake tube)/20 mL (shake flask) fresh minimal medium with an initial OD of 0.08. Cells were induced with 1mM of IPTG when $OD_{600}$ reached 0.6. Cell growth and fatty acids production were monitored at different time points. For semi-batch fermentation, 9 mL of LB culture with the engineered strain was incubated overnight and

inoculated into 450 ml M9 minimal medium (2% glucose, 30 mg/L chloramphenicol with a supply of Vitamin B12) in the bioreactor (New Brunswick BioFlo 110 fermentor). The fermentation was initiated with the following settings: The incubation temperature was controlled between 35°C ~ 37 °C; the pH of the culture medium was adjusted around 7.2 by automatic addition of ammonium hydroxide (6 mol/L); the airflow rate was kept at ~1.5 L/min, and the average stirring rate was 500 rpm. When $OD_{600}$ of culture reaches 7, 0.1 mM of IPTG (final concentration) was added. Four hours after induction, a glucose stock solution (400 g/L glucose and 12 g/L $MgSO_4$) was intermittently pulsed into the bioreactor to re-supply glucose. After 48 h of induction, a total of 40.72 g/L of glucose was consumed, and the final cultures were harvested for measurement of free fatty acids.

### 1.4.4 Fatty acids, glucose, and acetate measurements.

Free fatty acid titer was analyzed following a previous protocol (Liu *et al*. 2015). In brief, 500 μL of cell culture was acidified using 50 μL of concentrated HCl. Fatty acids were extracted twice with ethyl acetate (EtAc) spiked with C19:0 methyl ester (ME) as an internal standard. Fatty acids were then derivatized to fatty acids methyl esters (FAME) using 100 μL of MeOH:HCl (9:1) and 100 μL of TMS-diazomethane (2 M in hexanes). The mixture was incubated for 10 ~ 15 min at room temperature for the reaction to complete. FAME was analyzed using a gas chromatograph-mass spectrometer (GC-MS, Hewlett-Packard model 7890A, Agilent Technologies). Free fatty

acids were quantified based on the standard curve of standard FAME mix. Acetate and D-Glucose measurements followed the protocol of commercial kits (r-biopharm, MO, USA)

### 1.4.5 FBA Simulation.

The genome-scale model iJO1366 (which includes 1366 genes, 2583 reactions, and 1805 metabolites) was employed to simulate fatty acids production in *E.coli* strain (Monk *et al*. 2013). A simplified flux of fatty acid (C16:0) was added as representative of fatty acids production, and the objective function was set to maximize this flux. Default values for the boundary of all fluxes were adopted, except for the following: The upper and lower boundaries of the fatty acid degradation flux were set to zero because Δ*fadE* was knocked out, and the lower boundary of the glucose uptake flux was set based on experimental value. The sensitivity of fatty acid yields to ATP maintenance loss and oxygen influx was tested by FBA. The COBRA toolbox and LibSBML library were employed for genome-scale model manipulation (Bornstein *et al*. 2008; Schellenberger *et al*. 2011), and the Gurobi 5.5 linear solver (Gurobi Optimization Inc.) was utilized for FBA calculation on MATLAB 2012b.

# Acknowledgements

The authors would like to express their gratitude to Dr. Yi Xiao and Wen Jiang for their suggestions and discussions. This work was supported by the National Science Foundation (DBI 1356669 to

YT, MCB1453147 to FZ) and the Human Frontier Science Program (RGY0076/2015 to FZ). Ni Wan was partially supported by the National Science Foundation (CBET 1438125).

# References


Andersson CIJ, Holmberg N, Farrés J, Bailey JE, Bülow L, Kallio PT. 2000. Error-prone PCR of *Vitreoscilla* hemoglobin (VHb) to support the growth of microaerobic *Escherichia coli*. Biotechnology and Bioengineering 70(4):446-455.

Bornstein BJ, Keating SM, Jouraku A, Hucka M. 2008. LibSBML: an API Library for SBML. Bioinformatics 24(6):880-881.

Cheng Q, Perry MP, Tao L. 2006. Bacterial hemoglobin genes and their use to increase carotenoid production. US Patent 7,163,819.

Dellomonaco C, Clomburg JM, Miller EN, Gonzalez R. 2011. Engineered reversal of the β-oxidation cycle for the synthesis of fuels and chemicals. Nature 476(7360):355-359.

Dikshit KL, Webster DA. 1988. Cloning, characterization and expression of the bacterial globin gene from *Vitreoscilla* in *Escherichia coli*. Gene 70(2):377-386.

Frey AD, Kallio PT. 2003. Bacterial hemoglobins and flavohemoglobins: versatile proteins and their impact on microbiology and biotechnology. FEMS Microbiology Reviews 27(4):525-545.

Garcia-Ochoa F, Gomez E. 2009. Bioreactor scale-up and oxygen transfer rate in microbial processes: An overview. Biotechnology Advances 27(2):153-176.

Glick BR. 1995. Metabolic load and heterologous gene expression. Biotechnology Advances 13(2):247-261.

He L, Xiao Y, Gebreselassie N, Zhang F, Antoniewicz MR, Tang YJ, Peng L. 2014. Central metabolic responses to the overproduction of fatty acids in *Escherichia coli* based on $^{13}C$-metabolic flux analysis. Biotechnology and Bioengineering 111(3):575-585.

Jones JA, Toparlak ÖD, Koffas MAG. 2015. Metabolic pathway balancing and its role in the production of biofuels and chemicals. Current Opinion in Biotechnology 33:52-59.

Khosla C, Bailey J. 1988a. The *Vitreoscilla* hemoglobin gene: Molecular cloning, nucleotide sequence and genetic expression in *Escherichia coli*. Molecular and General Genetics MGG 214(1):158-161.

Khosla C, Bailey JE. 1988b. Heterologous expression of a bacterial haemoglobin improves the growth properties of recombinant *Escherichia coli*. Nature 331(6157):633-635.

Lee TS, Krupa RA, Zhang F, Hajimorad M, Holtz WJ, Prasad N, Lee SK, Keasling JD. 2011. BglBrick vectors and datasheets: A synthetic biology platform for gene expression. Journal of Biological Engineering 5(1):1-14.

Liu D, Xiao Y, Evans BS, Zhang F. 2015. Negative Feedback Regulation of Fatty Acid Production Based on a Malonyl-CoA Sensor–Actuator. ACS Synthetic Biology 4(2):132-140.

Lu X, Vora H, Khosla C. 2008. Overproduction of free fatty acids in *E. coli*: Implications for biodiesel production. Metabolic Engineering 10(6):333-339.

Monk JM, Charusanti P, Aziz RK, Lerman JA, Premyodhin N, Orth JD, Feist AM, Palsson BØ. 2013. Genome-scale metabolic reconstructions of multiple *Escherichia coli* strains highlight strain-specific adaptations to nutritional environments. Proceedings of the National Academy of Sciences 110(50):20338-20343.

Schellenberger J, Que R, Fleming R, Thiele I, Orth J, Feist A, Zielinski D, Bordbar A, Lewis N, Rahmanian S. Kang J, Hyduke DR, Palsson BØ. 2011. Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2.0. Nature Protocols 6(9):1290 - 1307.

Van Dien S. 2013. From the first drop to the first truckload: commercialization of microbial processes for renewable chemicals. Current Opinion In Biotechnology 24(6):1061-1068.

Wu G, Yan Q, Jones JA, Tang YJ, Fong SS, Koffas MAG. 2016. Metabolic Burden: Cornerstones in Synthetic Biology and Metabolic Engineering Applications. Trends In Biotechnology 34(7).

Wu S, He L, Wang Q, Tang YJ. 2015. An ancient Chinese wisdom for metabolic engineering: Yin-Yang. Microbial Cell Factories 14(1):39.

Xiao Y, Bowen CH, Liu D, Zhang F. 2016. Exploiting nongenetic cell-to-cell variation for enhanced biosynthesis. Nat Chem Biol. 12(5):339-44.

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*. Nature Communication 4:1409.

Xu P, Li L, Zhang F, Stephanopoulos G, Koffas MAG. 2014. Improving fatty acids production by engineering dynamic pathway regulation and metabolic control. Proceedings of the National Academy of Sciences 111(31):11299-11304.

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):354-359.

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.