Biosensing for Multiplexed Genome Engineering: Applications in Renewable Chemical Production

Biosensing for Multiplexed Genome Engineering: Applications in Renewable Chemical Production

A dissertation presented

by

Jameson Kerr Rogers

to the

School of Engineering and Applied Sciences

in partial fulfillment of the requirements
for the degree of

Doctor of Philosophy

in the subject of

Engineering Sciences

Harvard University
Cambridge, Massachusetts

April 2015

Dissertation Advisor: Professor George M. Church Jameson Kerr Rogers

Biosensing for Multiplexed Genome Engineering: Applications in Renewable Chemical Production

Abstract

Engineered biological systems are increasingly used to produce fuels, pharmaceuticals and industrial chemicals. While transforming cells into renewable chemical factories presents an enormous opportunity, development timelines are long, costly and often uncertain. Engineering microbes for chemical production is accomplished through the biological design-build-test cycle: many designs are formulated, the corresponding organisms are constructed, and their ability to produce the desired chemical is evaluated. Designs that perform well become the starting point for the next round of the cycle. Faster design cycles result in shorter and less costly product development timelines.

Advances in DNA sequencing, synthesis and genome engineering technologies have sped up the design and build steps of the design cycle by enabling billions of organism variants to be designed and constructed simultaneously. However, evaluation of the resulting designs continues to rely on low-throughput technologies with evaluation rates on the order of thousands per day. Because the engineering process is a cycle, it can only proceed at the rate of the slowest step. A high-throughput method for design evaluation would increase the throughput of the design cycle by up to a million-fold.

This thesis describes an engineering framework that makes high-throughput design evaluation a reality. By programming cells to keep track of their own success in making a desired product, I enable screens and selections to be used for the optimization of metabolic pathways. I develop biosensors that maintain gene expression at a rate proportional to the concentration of several different chemical products and show that higher product concentration results in a higher fluorescent output. I then construct metabolic pathways for the production of the renewable plastic precursors 3-hydroxypropionate, acrylate, glucarate and muconate. I combine each pathway with the appropriate biosensor and use fluorescence to observe product formation in real-time. Next, I replace the fluorescent protein with an antibiotic resistance gene and link the level of product formation to the cell's ability to survive an antibiotic challenge. I deploy the selection to optimize production of both glucarate and naringenin from glucose.

I further develop the characterization of these new biosensors to promote their use as genetic switches for synthetic biological circuits. Finally, I develop a device called the fluorimostat that makes long-term closed-loop programmable control of gene expression a reality.

Table of contents

Acknowledgements

I would like to thank George Church for taking me on as a student and providing such a valuable learning experience. George has crafted a unique environment that is ideally suited to maximize creativity and foster world-changing innovation. There are two mantras that I will always associate with my time in the Church lab. First, nothing is impossible. Something might be cost prohibitive or take a long time, but *impossible* is not part of the Church lab vocabulary. This philosophy is one of the reasons why George's students are able to accomplish so much. The second mantra is related to scope. If you are going to do an experiment, why not do it on such a scale that the implications will resonate around the world? Why do one experiment when you can do all the experiments simultaneously? While George is an exemplar for scientific thought and innovation, his drive to improve the world makes him an excellent role model for how to live life outside the lab as well.

I thank my thesis committee for listening to my ideas and guiding their development over the last several years. Neel Joshi welcomed me to Harvard five years ago and has led me through the PhD process since then. Thank you, Don Ingber for advising me as a committee member and making the Wyss Institute a reality. When I decided to come to Harvard for my PhD, I was really deciding to join the Wyss Institute for Biologically Inspired Engineering. There is nowhere else I could have pursued immediately translatable research while surrounded by entrepreneurs and engineers alike. I want to thank Brian Baynes for making time to serve on my committee despite his busy schedule as a full-time entrepreneur. Brian contributes a commercial perspective on my research and has provided consistently valuable advice on how to maximize my impact in the real world.

While my mentors within Harvard have helped me develop the core of my scientific thinking, my mentors outside Harvard have helped me hone the other skills required for a successful career as an entrepreneur-scientist. As such, I want to thank my colleagues and mentors at Pronutria and Flagship Ventures. At Pronutria, I would like to acknowledge the mentorship of Rajeev Chillakuru, Subhayu Basu, and Greg Sieczkiewicz. At Flagship Ventures, I want to thank Geoff von Maltzahn and Avak Kahvejian for teaching me the VentureLabs entrepreneurial process and guiding me through the early stages of company creation. I also want to thank Noubar Afeyan for welcoming me to VentureLabs and providing such a unique environment for experimentation in biotech entrepreneurship.

I would like to acknowledge my wonderful colleagues who I have worked with every day. I thank my closest collaborators Vatsan Raman, Noah Taylor and Ben Pruitt. Much of this thesis has been a team effort and the work would not have been the same without these collaborations. I also thank the rest of my colleagues at the Wyss and NRB, especially Dan Goodman, Dima Ter-Ovanesyan, Marc Lajoie, Marc Guell and Wade Hicks for being great colleagues and scientific sounding boards. I especially thank Kevin Esvelt for his early mentorship in both scientific philosophy and molecular biology. I thank Jason Lohmueller and Tom Armel for being the first to teach me the techniques of synthetic biology.

The most important people to thank are my family and friends. To my family, thank you for tolerating me living three thousand miles away and always dropping everything to spend time with me when I visit. Thanks for the constant support and for being excited about all my small discoveries. And thanks for being awesome – I wouldn't be writing a thesis acknowledgement right now without the support of such a great family. And, to my friends in Cambridge who have made my time here so memorable, thank you. Thanks to

Lesley Sykes for being my companion throughout the PhD. You got me through all the hard parts and made all the good parts great by sharing them with me. Thanks to Daniel Dadon for bringing some of San Diego to Cambridge and to Kyle Eriss for being a great Cantabrigian friend and neighbor.

Chapter 1: Introduction

Biosensing for Multiplexed Genome Engineering

Imagine a future where chemicals and materials are produced from CO_2 and sunlight rather than oil dug from the earth – where microbes are harnessed to create products unconstrained by what can be produced from petrochemical building blocks. Advances in biotechnology bring this future closer every day. Even in its nascent state, the bioeconomy is a large and growing segment of the global economy[1]. Annual U.S. revenues derived from genetically modified systems are more than $350 billion, equivalent in size to the entire U.S. semiconductor industry[2]. Industrial biotechnology accounts for $125 billion of the bioeconomy and is composed of revenue from the sale of fuels, materials, chemicals and enzymes. Biologically derived chemicals account for $66 billion of the total, while biofuels make up $30 billion[3]. Demand for bio-based products is driven not only by their renewable nature, but by potentially lower production costs, independence from foreign oil, and novel product chemistries[4]. While the promise of industrial biotechnology is huge, development timelines are long, costly and often uncertain.

It is no surprise that engineering biological systems is a challenging process. Product development requires the creation of a life form that behaves in a predictable and reproducible way, a process in which traditional engineering paradigms tend to be inadequate. Each component of a biological system interacts with thousands of other components within the cell. This is in stark contrast to electrical or mechanical engineering, where a given component interacts with just a handful of adjacent components. The

astounding complexity that results from the high connectivity of biological systems is further exacerbated by poor characterization of components at the individual level. Consequently, finding a solution that maintains cellular viability, while meeting desired design goals, requires many iterations of the engineering process. Each iteration can be broken into three steps: biological design, genetic construction, and phenotype evaluation. This is the biological design-build-test cycle, and the primary technological driver for an advanced bioeconomy.

The rates of biotechnological innovation and product development increase with higher design cycle throughput. Because design-build-test throughput is the product of cycle speed and bandwidth, the throughput of a given step increases with greater step speed or wider step bandwidth. The speed of the build and test steps are limited by the rate at which cells are grown and manipulated. In contrast to speed, the physical limits on bandwidth are vast. Billions of cells or trillions of DNA molecules fit in a single droplet. If each cell or molecule contains a unique design, the designs are multiplexed in space. In contrast to multiplexing, parallel experimentation requires spatial separation of designs. This places a physical limit on the number of designs that can be processed due to constraints of space and the logistics of design manipulation.

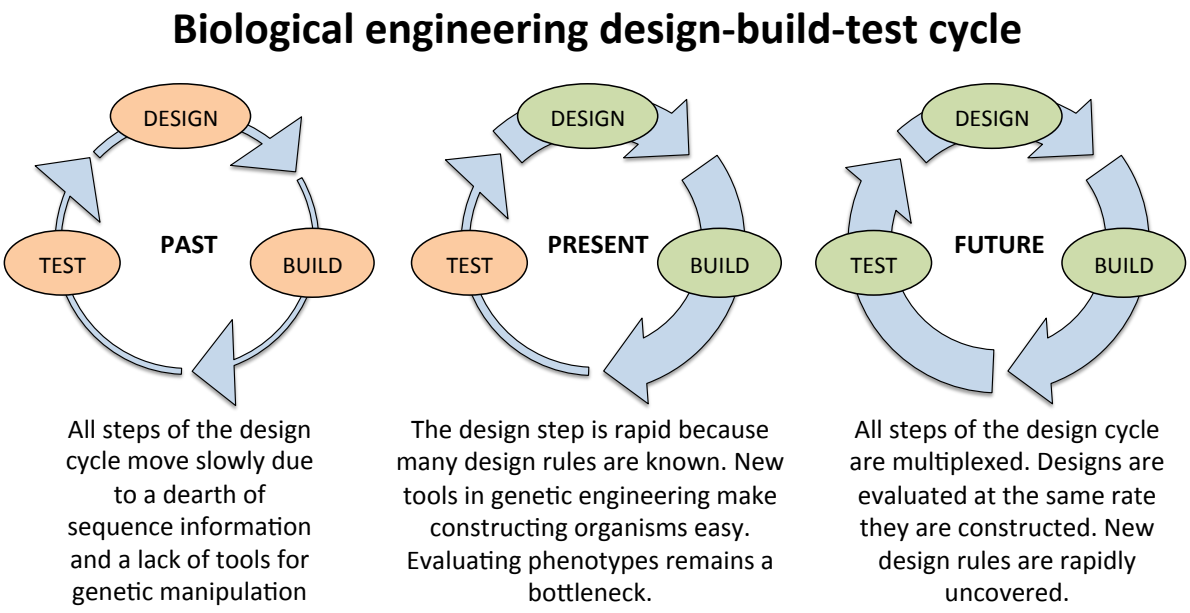


Figure 1.1: The evolution of the biological engineering design-build-test cycle.

Using the evaluation step to learn about the design space and inform subsequent design steps enables product development rates to scale non-linearly with the throughput of the design cycle. The simplest feedback between the evaluation and design steps occurs as it does in nature – the most successful designs survive to become the templates for the next iteration of the cycle and the design space converges to a local maximum. More advanced feedback from the test step to the design step provides an explicit understanding of which designs function best. This enables the formulation of design rules that more quickly define the design space and inform subsequent engineering endeavors. This is a form of directed evolution in which an engineer monitors the flow of designs and intervenes as necessary. However, such a process requires the technological capability to quickly rank and identify the best designs, and to build new designs with enough speed and precision to actuate the knowledge gained.

Because total cycle throughput is limited by the throughput of the slowest step, every step of the cycle must be multiplexed to achieve a fully multiplexed design cycle. While innovation in a single step results in higher step throughput, it also changes how adjacent steps are approached. Widely available gene synthesis is an example of how an innovation in the build-step freed design-step engineers from the constraints of using existing DNA sequences. A more dramatic leap in how engineers approach biological design will occur once the cycle is multiplexed from start to finish.

A perfect design step would obviate the other steps in the design cycle. Some engineering disciplines have approached this scenario – it is rare to see a bridge constructed repeatedly until it works as desired. However, it seems safe to assume our knowledge of biological design principles will lag behind those of structural and civil engineering for the foreseeable future, necessitating continued innovation in biological design.

The effectiveness of the design step has risen meteorically with DNA sequencing technology. In the burgeoning era of biotechnology, locating or amplifying a gene was hindered by a lack of sequence knowledge. This made even simple design endeavors, such as recombinant protein expression, an arduous task. Even once a gene's sequence was known, exploration of novel biological designs was constrained to random walks in adjacent sequence space because hypotheses about functional regions had yet to be formulated. Later, as the amount of sequenced DNA exploded, so did the potential for designing novel biological systems. The vast repertoire of sequenced genomes has enabled the formulation of design rules; in turn enabling engineers to hone in on the active sites of

enzymes, borrow homologous sequences from distant species and to locate regulatory elements for transcription and translation.

Modern-day design tools expand on these design principles and enable forward engineering of biological systems. Precise prediction of ribosomal binding site strength and promoter activity is now possible[5-7]. New proteins can be constructed *in silico* before being implemented *in vivo*[8]. The metabolism of entire organisms can be modeled mathematically, revealing which genes should be turned up or down to accomplish a given metabolic goal[9,10]. But even these sophisticated design tools exist within the ambiguity of biology – they provide guides for design spaces that we would otherwise have no hope of constraining. For example, a small enzyme consisting of 300 amino acids exists in a design space of $1x10^{390}$ possible protein sequences. If our design tools allow us to identify an active site of seven amino acids, that active site exists in a design space of one billion potential active sites. If we know that those amino acids should be positively charged, then our design space converges to two thousand possible proteins. How do we construct so many designs?

The build step of the design cycle is the process in which potential designs are constructed in DNA and integrated into a cell. The cost of gene synthesis has fallen to the point where ordering several genes is trivial for most laboratories [11,12]. Combined with technologies enabling seamless plasmid construction[13,14] and simple methods for modifying the genome[15,16], parallel construction is a robust process. However, achieving a meaningful increase in build bandwidth by multiplexing with synthesized genes will remain cost prohibitive as long as the construction of those genes is a parallel process itself. As such, engineers are left with either random or site directed mutagenesis to

achieve multiplexed construction of genetic elements. This has been a serious constraint on what designs can actually be constructed in a multiplexed way. Recently, chip-based oligonucleotide synthesis has enabled multiplexed construction of precisely designed sequences of up to 200 base pairs, allowing for the evaluation of hundreds of thousands of complex hypotheses at a time[6,17,18]. The capability to multiplex the build step is transforming how engineers approach the design cycle: experiments that were previously infeasible and now within reach.

Simultaneous advances in genome engineering have made the construction of billions of genome variants a routine process[16,19-22]. Multiplexed genome engineering allows specified or degenerate mutations to be targeted anywhere in the genome. Such facile genome engineering enables new classes of experiments. As an example, the entire set of metabolic modifications outlined by flux balance analysis (FBA) can now be explored simultaneously. When optimizing the production of a target compound, FBA identifies genes that are important to modulate but does not accurately specify what their level of expression should be[9]. Multiplexed genome engineering enables the combinatorial exploration of gene expression levels for each of the genes of interest[16,19]. If 10 genes are targeted with mutations corresponding to 10 levels of expression, the resulting design space is composed of 10 billion genomes.

Despite the success in multiplexing the design and build steps, evaluating 10 billion designs with current technology would take decades because the test step of the engineering cycle remains a parallel process. Evaluation of the multiplexed designs requires demultiplexing, which negates the value of multiplexing at the outset. One reason phenotype evaluation lacks an adequate multiplexed solution is that analytical methods are

dramatically different for different phenotypes. When cells are engineered to produce fuels or chemicals, design success is often determined by the amount of compound produced. This requires cells to be separated into individual designs (the demultiplexing step) and cultured independently in small volumes, such as in 96-well plates. Next, either the supernatant or cell lysates are prepared such that the concentration of the molecule of interest can be determined using chromatography or mass spectrometry[23]. Cells that produce high concentrations of the metabolite are retained and used as the starting point for the next round of design. Alternatively, promising designs are sequenced to reveal what aspects of the design worked well, informing subsequent rounds of the cycle. In typical labs, throughput is limited to hundreds of design evaluations per day, while specialized labs are able to evaluate thousands of designs per day[23].

Enabling cells to report their own progress in making a specific chemical provides a multiplexed solution to the test step. Rather than assaying individual designs, engineers should be able to define a design goal and immediately separate cells that meet the specified level of performance from cells that do not. If cells keep track of their own progress, then the time required for separation of productive cells from unproductive cells no longer scales with the number of cells evaluated. Selections are an example of such a multiplexed evaluation method. In a selection, only cells that have a certain phenotype survive. This decouples the time required for evaluation from the number of cells evaluated. However, selections are typically based upon an ad-hoc link between a phenotype of interest and a necessary cell function. For example, selecting for increased utilization of a new sugar is possible if all other energy sources are withheld. But selecting for the increased production of a novel chemical is not so simple. A general method for

multiplexed phenotype evaluation is the last step required for a fully multiplexed design-build-test cycle.

One such method is based upon genetically encoded biosensors. Biosensors provide a general framework for linking intracellular chemical concentration to transcription and provide a generalizable method for multiplexing design evaluation in cases of metabolic engineering. Genetically encoded biosensors are based upon allosteric transcription factors that allow expression of a target gene when bound by a specific small molecule. Transcription factors cluster into more than 20 major families[24]. Currently, the *lacI* family contains 29 thousand sequenced members, while the *gntR* family contains 49 thousand members[25]. There are over 200 thousand sequences available for members of the *tetR* family of transcriptional repressors[24]. These naturally occurring transcription factors bind to an incredible range of compounds. If a microbe has an incentive to consume or avoid a compound, there is likely a transcription factor that has evolved to bind it. Recent advances in protein engineering and directed evolution have produced designer transcription factors that bind compounds for which natural transcription factors have yet to be discovered[26,27].

Biosensors enable screens and selections for a vast repertoire of compounds by providing a transcriptional readout to intracellular metabolite concentration. When the transcriptional output of the biosensor is an antibiotic resistance gene, biosensor activation confers antibiotic resistance. Treating a population of cells with the appropriate antibiotic allows cells to survive only if they produce the required amount of product. Alternatively, if the transcriptional readout of the biosensor is a fluorescent protein, cells with more effective designs will fluoresce more brightly. Fluorescent biosensors enable millions of

cells to be screened per minute with high-throughput methods such as fluorescent activated cell sorting (FACS).

Combining multiplexed phenotype evaluation with next-generation sequencing enables a deep understanding of the design space being explored. Sorting cells into bins based upon their fluorescence is a multiplexed method for assigning each cell a rank based upon the quality of the design it contains. Deep sequencing of the bin contents provides a list of designs for each bin. Ranks are assigned to each design based on which bin they were found in. Each iteration of the design cycle provides millions of design-rank pairs. This wealth of information allows design rules to be developed much more rapidly than would otherwise be possible.

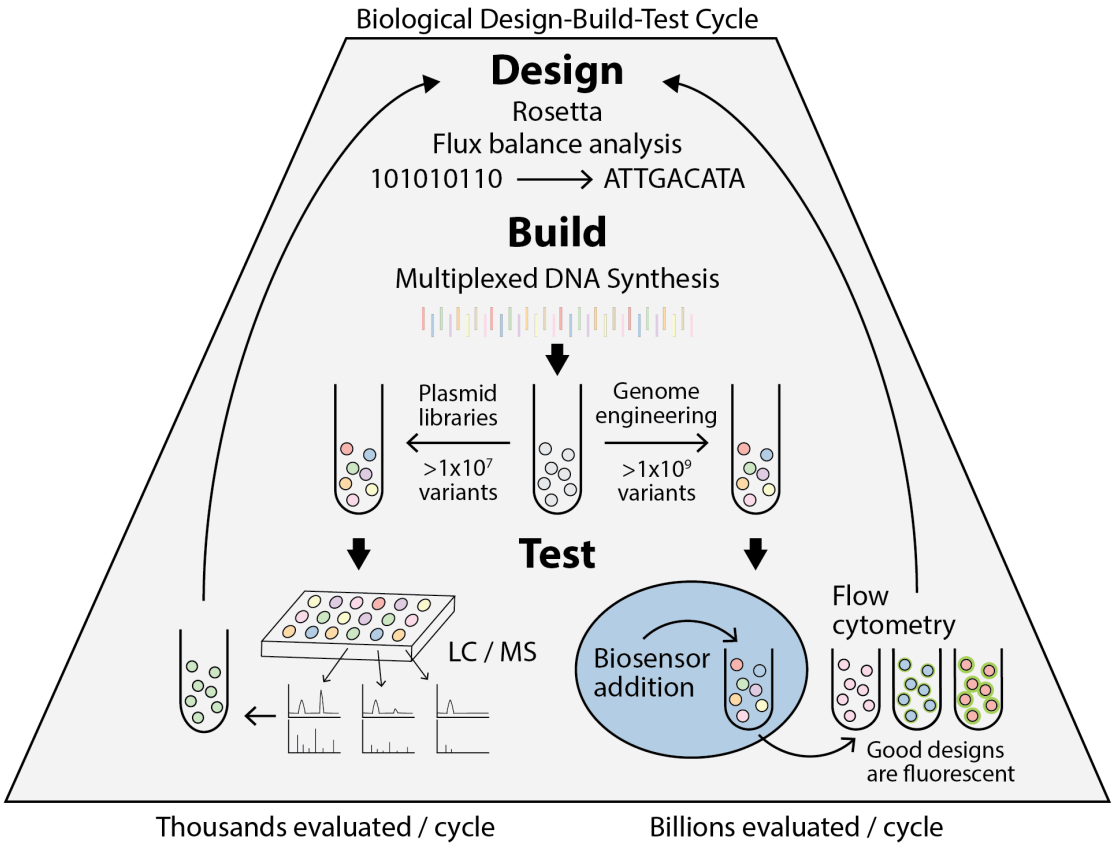


Figure 1.2: Biological design-build-test cycle for metabolic engineering. A reliable link between product titer and cellular fluorescence is the last piece of the puzzle to enable a fully multiplexed design cycle (right side of the triangle). Traditional metabolite analysis (left side of the triangle) requires separation of designs into individual cultures before analysis by liquid chromatography and/or mass spectrometry (LC/MS). Plasmid libraries, genome engineering and other building techniques can be used with either test methodology.

Simple design rules and extensive characterization of metabolite-responsive biosensors are necessary for multiplexed phenotype evaluation to be broadly adopted. Using biosensors to optimize several different metabolic pathways allows for the formulation of design rules that transform screens and selections from ad-hoc solutions to well-characterized tools. In the case of selections, three main areas of characterization are needed. First, the dynamic range of the sensor must be evaluated. This is the range of

chemical concentrations over which an increase in the production of the target compound results in an increase in fitness. Second, the failure rate of the sensor must be defined. The failure rate is the likelihood that a cell erroneously survives the selection without producing the required amount of product. The inverse of the failure rate is the maximum bandwidth for the selection. Third, there must be a set of methods for modulating the dynamic range and failure rate, such that the selection is appropriate for the evaluation task at hand.

An analogous set of characterizations are necessary for screens built upon fluorescent biosensors. The transfer function between product concentration and fluorescence response should be evaluated. The transfer function provides the dynamic range of the system and allows product concentration to be mapped onto the observed fluorescence. Determining the kinetics of biosensor activation over a range of product concentrations reveals the optimal timing for fluorescent sorting. Highly productive cells may be missed if sorting is done too soon, while poorly producing cells may be retained if sorting is carried out too late.

In this thesis I use biosensors to transduce small-molecule concentration into a transcriptional readout and develop a framework for multiplexed phenotype evaluation. I apply biosensors to monitor the production of glucarate, muconate, acrylate and 3-hydroxypropionate in real time and ultimately optimize the production of naringenin and glucarate. Glucarate, muconate and acrylate are plastic precursors. Enabling the renewable production of these compounds has the potential to lower costs and decouple pricing from oil. Glucarate polymers are entering the market as next-generation time-release fertilizers and ultra-absorbent fibers. Both Rivertop Renewables and Kalion have been actively

developing the glucarate market. Muconate is a precursor to nylon with a global market of $20B[28]. Muconate is also used to produce polyethylene terephthalate with a global market of $31B[29]. Amyris is pioneering the commercialization of these and other muconate products. The global acrylate market is $14 billion[30], with BASF, Cargill and Dow each pursuing bio-based drop-in replacements. 3-hydroxypropionate (3HP) is used to produce 1,3-propandiol, malonate and is a precursor for renewable acrylate. Genomatica, Metabolix, Myriant, Novozymes and OPX are all competing to achieve commercial scale fermentation of 3HP from sugar and other types of biomass. Naringenin is a flavonoid isolated from grapefruit with several emerging health applications.

Altogether, I have applied biosensor technology towards the optimization of five pathways, developing ten sensors for sixteen compounds selected from diverse chemical classes such as macrolide antibiotics, alkanes, vitamins, flavonoids, diacids, and phenol.

The field of synthetic biology aims to make the process of engineering biology more akin to electrical engineering. However, the complexity of synthetic circuits is limited by the number of orthogonal input channels available to control gene expression in real-time[31-33]. Biosensors can be used as novel communication channels into (and out of) engineered cells, providing a way to increase the complexity of human-cell interaction. To this end, I characterized biosensors as genetic rheostats such that they are tunable input channels for synthetic circuits. Building further on the principle of programmable biology, I have constructed a device, termed the fluorimostat, that is able to autonomously characterize the properties of inducible systems. The fluorimostat is controlled through a Python module known as cellscript, enabling biological engineers a direct programming interface for closed- and open-loop control of gene expression.

Throughout this thesis I have aimed to maximize my impact on the biological design cycle by producing tools that are effective, yet simple to implement. My hope is that by enabling a fully multiplexed design-built-test cycle for bio-based chemical production, I can help accelerate the rate of product development in industrial biotechnology. Previous efforts to multiplex the evaluation of experimental results have become foundations of modern biology. These technologies include immune sequencing[34], RNA-Seq[35], CHIP-Seq[36] and Flow-Seq[6,18]. The common element between all of these transformative technologies is the transduction of a signal to a nucleotide sequence. Multiplexed phenotype evaluation for metabolic engineering follows in the same tradition by linking intracellular small-molecule concentration to a transcriptional readout. Screens and selections are convenient methods to monitor the presence of these transcripts. Direct sequencing of the transcriptional output will enable even greater multiplexing, since dependence on translation and protein function would no be longer necessary.

References

1 Carlson, R. Biodesic 2011 Bioeconomy Update. (Biodesic, 2011).

2 Carlson, R. *synthesis*, <http://www.synthesis.cc/cgi-bin/mt/mt-t.cgi/392> (2014).

3 Solomon, D. in *ACS Science & the Congress.*

4 (ed The White House) (2012).

5 Salis, H. M., Mirsky, E. A. & Voigt, C. A. Automated design of synthetic ribosome binding sites to control protein expression. *Nature biotechnology* **27**, 946-950, doi:10.1038/nbt.1568 (2009).

6 Kosuri, S. *et al.* Composability of regulatory sequences controlling transcription and translation in Escherichia coli. *Proceedings of the National Academy of Sciences of the United States of America* **110**, 14024-14029, doi:10.1073/pnas.1301301110 (2013).

7 Mutalik, V. K. *et al.* Precise and reliable gene expression via standard transcription and translation initiation elements. *Nat Methods* **10**, 354-360, doi:10.1038/nmeth.2404 (2013).

8 Rohl, C. A., Strauss, C. E., Misura, K. M. & Baker, D. Protein structure prediction using Rosetta. *Methods in enzymology* **383**, 66-93, doi:10.1016/S0076-6879(04)83004-0 (2004).

9 Orth, J. D., Thiele, I. & Palsson, B. O. What is flux balance analysis? *Nature biotechnology* **28**, 245-248, doi:10.1038/nbt.1614 (2010).

10 Varma, A. & Palsson, B. O. Stoichiometric flux balance models quantitatively predict growth and metabolic by-product secretion in wild-type Escherichia coli W3110. *Applied and environmental microbiology* **60**, 3724-3731 (1994).

11 Carr, P. A. & Church, G. M. Genome engineering. *Nature biotechnology* **27**, 1151-1162, doi:10.1038/nbt.1590 (2009).

12 Carlson, R. Vol. 2015 (2014).

13 Gibson, D. G. *et al.* Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat Methods* **6**, 343-345, doi:10.1038/nmeth.1318 (2009).

14 Engler, C. & Marillonnet, S. Combinatorial DNA assembly using Golden Gate cloning. *Methods in molecular biology* **1073**, 141-156, doi:10.1007/978-1-62703-625-2_12 (2013).

15 Datsenko, K. A. & Wanner, B. L. One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. *Proceedings of the National Academy of Sciences of the United States of America* **97**, 6640-6645, doi:10.1073/pnas.120163297 (2000).

16 Wang, H. H. *et al.* Programming cells by multiplex genome engineering and accelerated evolution. *Nature* **460**, 894-U133, doi:Doi 10.1038/Nature08187 (2009).

17 Quan, J. *et al.* Parallel on-chip gene synthesis and application to optimization of protein expression. *Nature biotechnology* **29**, 449-452, doi:10.1038/nbt.1847 (2011).

18 Goodman, D. B., Church, G. M. & Kosuri, S. Causes and effects of N-terminal codon bias in bacterial genes. *Science* **342**, 475-479, doi:10.1126/science.1241934 (2013).

19 Raman, S., Rogers, J. K., Taylor, N. D. & Church, G. M. Evolution-guided optimization of biosynthetic pathways. *Proceedings of the National Academy of Sciences of the United States of America* **111**, 17803-17808, doi:10.1073/pnas.1409523111 (2014).

20 Wang, H. H. *et al.* Genome-scale promoter engineering by coselection MAGE. *Nat Methods* **9**, 591-+, doi:Doi 10.1038/Nmeth.1971 (2012).

21 Isaacs, F. J. *et al.* Precise manipulation of chromosomes in vivo enables genome-wide codon replacement. *Science* **333**, 348-353, doi:10.1126/science.1205822 (2011).

22 Lajoie, M. J. *et al.* Genomically recoded organisms expand biological functions. *Science* **342**, 357-360, doi:10.1126/science.1241459 (2013).

23 Dietrich, J. A., McKee, A. E. & Keasling, J. D. High-Throughput Metabolic Engineering: Advances in Small-Molecule Screening and Selection. *Annu Rev Biochem* **79**, 563-590, doi:Doi 10.1146/Annurev-Biochem-062608-095938 (2010).

24 Cuthbertson, L. & Nodwell, J. R. The TetR family of regulators. *Microbiology and molecular biology reviews : MMBR* **77**, 440-475, doi:10.1128/MMBR.00018-13 (2013).

25 Finn, R. D. *et al.* Pfam: the protein families database. *Nucleic acids research* **42**, D222-230, doi:10.1093/nar/gkt1223 (2014).

26 Tang, S. Y. & Cirino, P. C. Design and application of a mevalonate-responsive regulatory protein. *Angewandte Chemie* **50**, 1084-1086, doi:10.1002/anie.201006083 (2011).

27 Raman, S., Taylor, N., Genuth, N., Fields, S. & Church, G. M. Engineering allostery. *Trends in genetics : TIG* **30**, 521-528, doi:10.1016/j.tig.2014.09.004 (2014).

28 Liu, X. (2012).

29 Newswire, P. (Yahoo Finance, 2014).

30 Newswire, P. (Bloomberg News, 2013).

31 Moon, T. S., Lou, C., Tamsir, A., Stanton, B. C. & Voigt, C. A. Genetic programs constructed from layered logic gates in single cells. *Nature* **491**, 249-253, doi:10.1038/nature11516 (2012).

32 Stanton, B. C. *et al.* Genomic mining of prokaryotic repressors for orthogonal logic gates. *Nature chemical biology* **10**, 99-105, doi:10.1038/nchembio.1411 (2014).

33 Litcofsky, K. D., Afeyan, R. B., Krom, R. J., Khalil, A. S. & Collins, J. J. Iterative plug-and-play methodology for constructing and modifying synthetic gene networks. *Nat Methods* **9**, 1077-1080, doi:10.1038/nmeth.2205 (2012).

34 Laserson, U. *et al.* High-resolution antibody dynamics of vaccine-induced immune responses. *Proceedings of the National Academy of Sciences of the United States of America* **111**, 4928-4933, doi:10.1073/pnas.1323862111 (2014).

35 Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L. & Wold, B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods* **5**, 621-628, doi:10.1038/nmeth.1226 (2008).

36 Baba, T. *et al.* Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol Syst Biol* **2**, doi:Artn 2006.0008 Doi 10.1038/Msb4100050 (2006).

Chapter 2: Synthetic biosensors for precise gene control and real-time monitoring of metabolites

Abstract

Characterization and standardization of inducible transcriptional regulators has transformed how scientists approach biology by allowing precise and tunable control of gene expression. Despite their utility, only a handful of well-characterized regulators exist, limiting the complexity of engineered biological systems. We apply a characterization pipeline to four genetically encoded sensors that respond to acrylate, glucarate, erythromycin and naringenin. We evaluate how the concentration of the inducing chemical relates to protein expression, how the extent of induction affects protein expression kinetics, and how the activation behavior of single cells relates to ensemble measurements. We show that activation of each sensor is orthogonal to the other sensors, and to other common inducible systems. We demonstrate independent control of three fluorescent proteins in a single cell, chemically defining eight unique transcriptional states. To demonstrate biosensor utility in metabolic engineering, we apply the glucarate biosensor to monitor product formation in a heterologous glucarate biosynthesis pathway and identify superior enzyme variants. Doubling the number of well-characterized inducible systems makes more complex synthetic biological circuits accessible. Characterizing sensors that transduce the intracellular concentration of valuable metabolites into fluorescent readouts enables high-throughput screening of biological catalysts and alleviates the primary bottleneck of the metabolic engineering design-build-test cycle.

Introduction

In-depth biological part characterization forms the foundation for abstraction and complexity in engineered biological systems. Sensors are one of the most important components to characterize as they provide the channels of communication into and out of the cell. Biosensors that respond to external agents such as chemicals or light allow real-time control of gene expression. Furthermore, sensors enable online monitoring of metabolic phenotypes by transducing intracellular chemical concentration into gene expression. Because phenotype evaluation is a major rate-limiting step in metabolic engineering, coupling sensors to reporter gene expression enables rapid and multiplexed phenotype evaluation, facilitating faster design-build-test cycles.

Small molecule inducible systems are genetically encoded biosensors that modulate gene expression in response to the presence of a small molecule inducer. One of the most widely used biosensors is the allosteric DNA binding protein LacI, which natively regulates the lactose catabolism operon in *E. coli* by binding near the transcriptional start site and repressing transcription initiation (1). When an inducing molecule such as isopropyl β-D-1-thiogalactopyranoside (IPTG) is present in the cell, it binds to the LacI protein and the LacI-IPTG complex disassociates from DNA, allowing transcription to proceed. Construction and characterization of engineered LacI-inducible systems (2,3) has resulted in widespread use in applications ranging from protein over-expression (4), to signal processing (5,6), and even chromosomal visualization (7).

Because of their general applicability and extensive characterization, a small set of canonical inducible regulators (LacI, TetR (2), AraC (8), LuxR (9)) are repeatedly used for a diverse range of applications. Other well-characterized inducible systems are available

(PrpR (10), RhaRS (11), CymR (12), XylS (13)), but with the exception of CymR, these suffer from catabolite repression and/or weak induction. Other expression control paradigms include riboswitches (14), which provide ligand-mediated control of translation, and light-regulated optogenetic systems (15), which are a promising complement to chemical induction. However, there is a pressing need for additional inducible systems, as genetically encoded biosensors allow facile control of transcription by merely supplying the inducer in the growth medium.

Robust inducible systems are valuable tools facilitating adjustable and on-the-fly control of specific genes. Tunable expression of one or more genes over the course of an organism's growth provides unique insights into gene function (16) and developmental programs (17). Dynamic regulation is therefore distinct from static methods that disrupt genes altogether (18-20) or that change expression through modification of *cis*-regulatory elements such as promoters (21) and ribosomal binding sites (RBS) (22). Because extensive characterization is often lacking, inducible systems are typically operated as all-or-nothing switches without regard for the speed or extent of induction. This mode of operation is adequate for conditional overexpression of potentially toxic genes, but yields less information than careful titration of gene dosage when probing cellular behavior (23).

When metabolite-responsive biosensors regulate fluorescent reporters, they facilitate real-time observation of internal cell states. Measurement of intracellular metabolites is often desirable, but is typically a slow and destructive process. Standard methods require macroscopic cultures or prepared lysates, which are assayed by chromatography and mass spectrometry, or by absorbance in special cases (24). When intracellular metabolite concentration is transduced to fluorescence, high-throughput single cell measurements

become possible (24). Fluorescent monitoring approaches can leverage allosteric transcriptional regulators that are already known for many common intermediates of metabolism, such as pyruvate (25), phosphoenol pyruvate (26), citrate (27), lactate (28), unsaturated fatty acids (29) and NADH (30).

Genetically encoded biosensors are valuable in metabolic engineering applications as they enable cells to report on their individual progress in producing a target compound from glucose or other low cost starting materials. Each cell expresses a fluorescent protein or antibiotic resistance gene at a rate proportional to its ability to produce the target compound. This link between internal metabolite concentration and reporter expression allows engineers to screen (by fluorescence) or select (with an antibiotic) for the most desirable cells (31). Coupling selections or screens to small molecule sensors have yielded new enzymes (32-35) and genomes enhanced for target metabolite production (36).

In addition to online observation of key metabolites, metabolic engineers benefit from advances in real-time control of biosynthetic gene expression. Independent control of each enzyme in a metabolic pathway facilitates the careful balancing of expression that is often necessary for optimal product production (37). In other cases, carefully timed expression of pathway enzymes has been shown to increase product titer (38). Keasling and coworkers used a genetically encoded biosensor to increase biodiesel titers by tying enzyme expression to the concentration of a pathway intermediate (39). Monitoring unwanted side-products, such as lactate, or desired products, such as fatty acids, with fluorescent reporters allows screening of millions of cells by flow-cytometry (32). If the reporter is an antibiotic resistance marker, selection can then be used in directed evolution applications (40).

Thorough characterization of an inducible system should provide enough information for applications ranging from environmental sensing and signal processing to metabolic engineering. Many characterization approaches provide this information and new methodologies are being developed. The BioFab has created analysis pipelines resulting in the characterization of transcriptional initiators (41) and terminators (42). Furthermore, they have developed methods to indicate how robust these characterized parts are to changing environments (43). While these characterization projects rely primarily on end-point measurements in 96-well plates, other approaches have exclusively used flow-cytometry data to fit parameters to complex models of induction (44).

Biosensor parameters dictate how useful a given biosensor will be for the application at hand. Characterizing these properties requires measuring: (1) the relationship between stimulus strength and circuit activation; (2) the response time of the biosensor to a stimulus; (3) the heterogeneity of biosensor activation between cells in an isogenic population and (4) the cross-reactivity with stimuli of other biosensors. Pioneering work from the Endy lab has introduced the idea of biological part 'datasheets' that capture the information subsequent engineers might need in order to use a new part(9). This engineering approach to biology has subsequently been used to characterize the commonly used inducible systems XylS, LacI and AraC (45). In this study, we characterize four additional allosteric transcription factors that modulate transcription in response to small molecule concentration. We include two commonly used biosensors in our analysis in order to ground our results in a context that many bioengineers are familiar with.

Results

We chose to characterize biosensors based upon two criteria: (1) published experimental validation of the DNA binding protein, its cognate promoter/operator and the inducer chemical; and (2) potential for the inducing chemical to be produced enzymatically through metabolic engineering. AcuR binds acrylate in order to regulate dimethylsulfoniopropionate (DMSP) catabolism in *Rhodobacter sphaeroides* (51). CdaR is a transcriptional activator from *E. coli* that has been shown to regulate transcription in response to several diacids: glucarate, galactarate and glycerate (52). MphR mediates transcription in the presence of erythromycin and other macrolide antibiotics such as josamycin and azithromycin (53). MphR was first identified in a macrolide resistant strain of *E. coli* (53) and has subsequently been used in both mammalian(54,55) and microbial (56) transgene activation. In *Pseudomonas putida*, TtgR regulates expression of a multi-drug efflux pump in response to flavonoids such as naringenin, phloretin and genistein (57), and has also been used for mammalian transgene activation (58). AcuR, MphR and TtgR are members of the TetR transcriptional repressor family. We also included the well-characterized regulators TetR and AraC for comparison.

Sensor Characterization

Biosensors were constructed as a single plasmid encoding both the allosteric transcriptional regulator and a fluorescent reporter. The reporter mRNA is transcribed from a promoter/operator sequence controlled by the allosteric transcriptional regulator. For transcriptional repressors, a medium-strength constitutive promoter (50) was used to drive regulator transcription. For the transcriptional activators, the native promoter

sequence of the activator was used in order to preserve the auto-regulating behavior of the AraC and CdaR regulators (52,59). We constructed the biosensors in commonly used high and low copy plasmids to evaluate their behavior in different contexts. High copy plasmids employed a pUC origin of replication (~100-500 copies), while the low copy plasmids encoded the SC101 replication origin (2-5 copies). All plasmids expressed beta-lactamase, enabling the use of carbenicillin for plasmid maintenance. In the case of the MphR biosensor, an erythromycin resistance gene was also included to protect the cells from the high macrolide concentrations required for induction.

The relationship between inducer concentration and expression of the fluorescent reporter was evaluated for six inducible systems (Fig. 2.1). The resulting biosensor transfer functions encompass the complete range of sensor outputs, allowing determination of each biosensor's dynamic range. Evaluation of the transfer function also reveals the minimum and maximum expression level obtainable in each biosensor implementation. The calculated fold-induction (maximum fluorescence divided by uninduced fluorescence) of high-copy biosensors ranged from 63 to 210 (Table 1). The fold-induction values indicated as being above a certain number are the result of the mean uninduced fluorescence residing within, or very near, the intensity of the cellular auto-fluorescence. A true number for fold-induction is undefined in this scenario, so a minimum fold-induction was determined using the bounds of the 95% confidence intervals. The greatest magnitude of induction among the high-copy biosensors was achieved with AraC, followed by CdaR and MphR (Fig. 2.1a). For the low-copy biosensors, fold induction ranged from 3 to 78. The AcuR biosensor demonstrated the lowest uninduced accumulation of GFP, with no fluorescence above background in the absence of acrylate for both the high and low copy

systems. This is in contrast to the TtgR biosensor, which showed higher uninduced accumulation of GFP in the low-copy configuration. The opposite effect was observed for the TetR biosensor, which demonstrated a lower uninduced accumulation of GFP, such that fluorescence was within that of the background in the low-copy configuration (Fig. 2.1b).

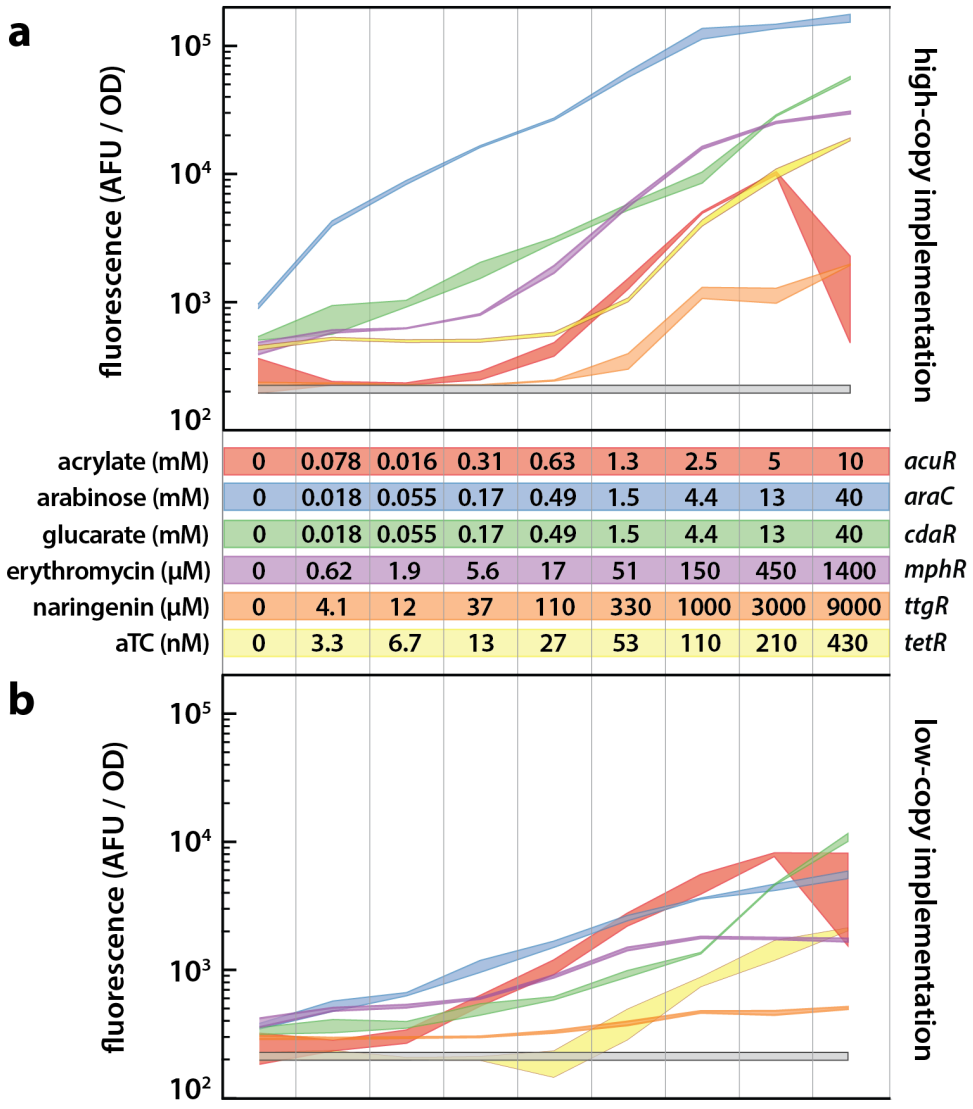


Figure 2.1: Induction dynamics for each of the inducible systems are reported. The relationship between fluorescent response and inducer concentration is represented as a 95% confidence band (n=3) for both the high-copy (a) and low-copy (b) implementations

of the inducible systems. The plots are log scale to capture the wide range of inducer concentrations and biosensor responses. Inducer concentrations are the same for both high and low-copy implementations. Each curve is matched to a color-coded table of inducer concentration ranges. Acrylate and anhydrotetracycline (aTC) increase in 2-fold increments while arabinose, glucarate, erythromycin and naringenin increase in 3-fold increments. The inducing chemical and biosensor name are indicated to the left and right of the table, respectively. The gray band is the fluorescent response of a control strain containing no fluorescent reporter. Fluorescence measurements are performed 15 hours after addition of the inducing chemicals.

The time required for induction was evaluated for each biosensor (Figs. 2.2, A.1, A.2). Reporter expression was monitored for eight hours with a wide range of inducer concentrations. All high-copy biosensors began producing fluorescence above background within 30 minutes, and achieved maximum levels of fluorescence within five hours under the highest induction conditions. Low-copy biosensors began producing measurable fluorescence within 50 minutes, but could require more than eight hours to achieve maximum fluorescence at the highest levels of induction. While the onset of expression began rapidly and without much variability between biosensors, the maximum fluorescence was sensor-dependent. For example, the CdaR biosensor achieved maximum fluorescence from moderate induction in nearly one hour, while the highest glucarate induction condition required six hours to reach maximal fluorescence. This is in contrast to the high-copy MphR biosensor, which approached maximal fluorescence around three hours regardless of the intensity of induction. The low copy variant of the MphR biosensor showed a similar trend, but required additional time to achieve maximum fluorescence (Fig. A.1). Variability in the kinetics of induction may be related to the intrinsic strength of the regulated promoter, sensor-DNA equilibrium, or the relationship between biosensor activity and growth-phase. Each repressor besides AcuR ceased to accumulate additional fluorescence at the onset of stationary phase. This is in contrast to the activators, which

achieved maximal fluorescence well before stationary phase. Contrary to the behavior of the other repressors, strong induction of AcuR by acrylate is enabled by entry into stationary phase.

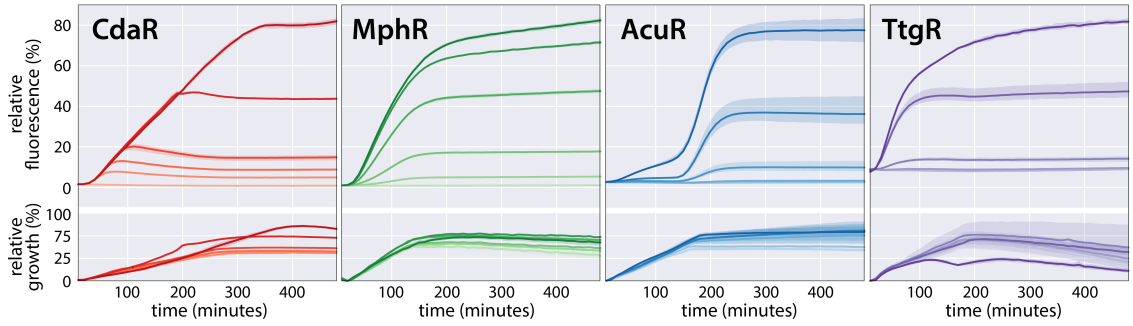


Figure 2.2: Induction and growth kinetics for the low-copy glucarate (CdaR), erythromycin (MphR), acrylate (AcuR) and naringenin (TtgR) biosensors. Chemical inducers are added at time zero and fluorescence is observed for eight hours. Lower panels show the optical density of the induced cultures over time. Induction levels are indicated by shade, with darker colors indicating higher inducer concentrations. Glucarate induction levels are 40mM, 13mM, 4.4mM, 1.5mM, 0.49mM and no inducer addition. Erythromycin induction levels are 1400μM, 450μM, 150μM, 51μM, 17μM and no inducer addition. Acrylate induction levels are 5mM, 2.5mM, 1.3mM, 0.63mM, 0.31mM and no inducer addition. Naringenin induction levels are 9mM, 3mM, 0.33mM, 0.11mM, 0.037mM and no inducer addition. Fluorescence and optical density are normalized as described in the Methods. The standard error of the mean is represented with a 95% confidence interval (n=3).

Complex synthetic circuits can be mathematically modeled to aid in component selection and system design. In order to make our biosensors compatible with such forward engineering efforts, we applied a simple model of gene activation to relate promoter activity to inducer concentration. We defined promoter activity as the time derivative of fluorescence corrected for cell growth. The time required for fluorophore maturation was considered and found to be less than two minutes (49). Likewise, degradation of GFP was ignored because the half-life in *E. coli* is greater than 24 hours (60).

We fit gene expression rates to a Hill function adapted to account for both the maximum velocity of gene expression and the basal expression of the uninduced cells. We found that both activators, AraC and CdaR, had Hill coefficients indicating low cooperatively. The repressors TetR, AcuR and TtgR all exhibited high cooperatively. The exception is the repressor MphR, which has a lower Hill coefficient (Table 1). Examination of the activity-induction curves of the high-copy sensors reveals that the induction behavior of MphR is indeed more similar to that of the activators AraC and CdaR, rather than the other repressors TetR, AcuR and TtgR (Fig. 2.3). The same trend holds for the activity-induction curves of the low-copy sensors except that AcuR demonstrates less cooperatively in this implementation, potentially due to its dependence on growth phase for activation (Fig. A.3). Due to the toxicity of acrylate at 10mM, we omitted this induction condition from the data used for modeling. Likewise, the highest concentrations of erythromycin were omitted from the low-copy MphR biosensor model as they showed substantial toxicity, likely due to lower expression of the erythromycin resistance gene. The maximum velocity of the high-copy sensors was always greater than the low-copy versions, however the magnitude of the change was greater in the repressors than the activators. The activator-based sensors control their own expression and this feedback may provide some expression stability in the face of copy number variation. Basal promoter activity was less than 3% of the maximum promoter activity for each high-copy biosensor. Low-copy biosensors had higher and more variable basal promoter activity due to lower maximum activities, and in some cases, less effective transcriptional repression.

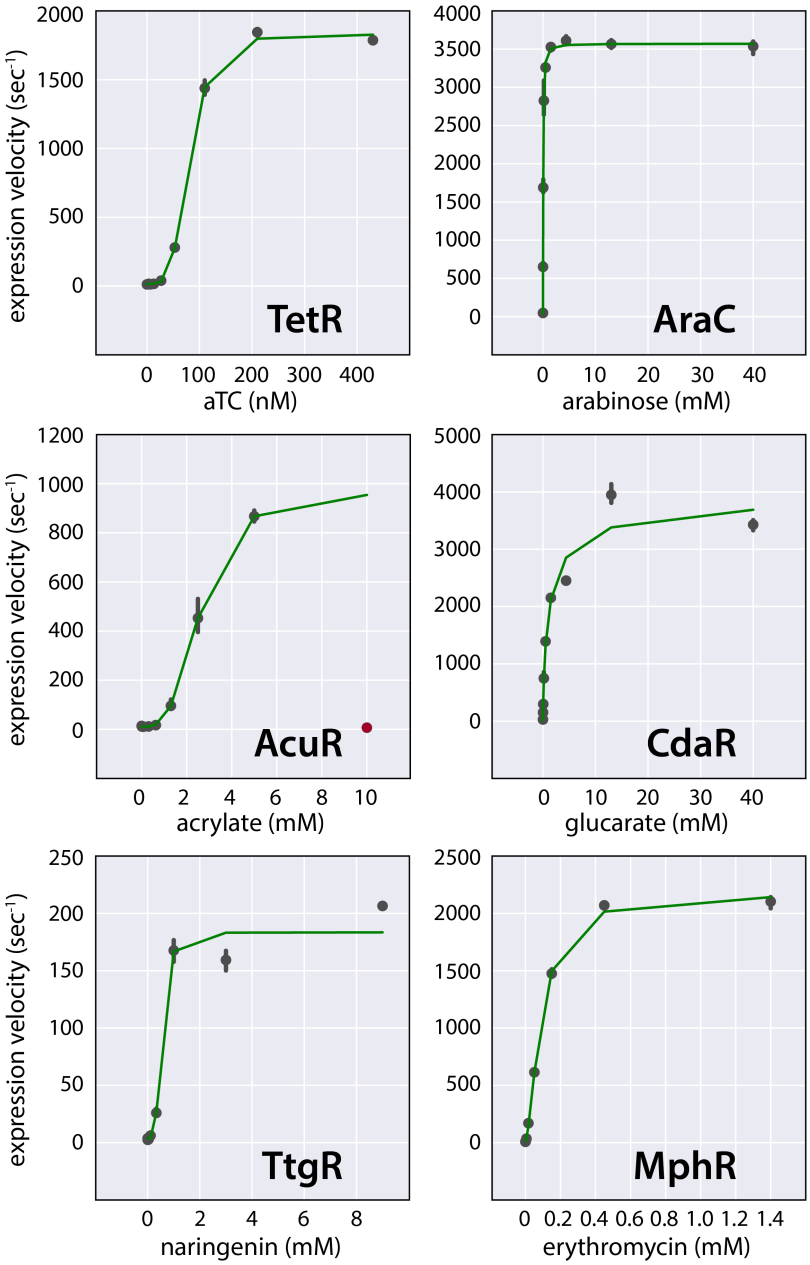


Figure 2.3: High-copy promoter activity was fit to a model of inducible gene expression. The maximum expression velocity of each inducible promoter was determined at various levels of induction (points). The data was fit to a Hill function modified to account for basal and maximal promoter activity (green lines). The anhydrotetracycline (TetR), acrylate (AcuR) and naringenin (TtgR) biosensors all show high induction cooperativity. The arabinose (AraC), glucarate (CdaR) and erythromycin (MphR) biosensors show low or moderate levels of cooperativity. The 10mM acrylate induction condition was omitted from the modeling data due to high toxicity (red point). Error bars reflect the 95% confidence interval for the measured expression velocity.

Table 2.1: Induction characteristics of the small-molecule inducible systems.

	Copy Number	Fold Induction	Hill Coefficient	Half Maximal Parameter		Max Expression Velocity (s^{-1})	Basal Expression Velocity (s^{-1})
AcuR	High	>90	3.2 ± 0.3	2.6 ± 0.1	mM	910 ± 40	5 ± 7
	Low	>50	1.3 ± 0.1	1.0 ± 0.1	mM	150 ± 10	15 ± 2
AraC	High	210 ± 9	1.3 ± 0.1	59 ± 3	μM	3150 ± 60	20 ± 50
	Low	29 ± 3	1.3 ± 0.2	250 ± 30	μM	1260 ± 50	40 ± 30
CdaR	High	168 ± 6	1 ± 0.1	490 ± 60	μM	2600 ± 100	0 ± 60
	Low	78 ± 8	1 ± 0.2	8 ± 2	mM	1000 ± 100	30 ± 20
MphR	High	108 ± 9	1.6 ± 0	97 ± 2	μM	2070 ± 20	9 ± 7
	Low	8 ± 1	1.6 ± 0.1	22 ± 1	μM	66 ± 1	5 ± 0.4
TetR	High	63 ± 3	4.2 ± 0.1	81 ± 1	nM	1760 ± 10	8 ± 6
	Low	>50	3.1 ± 0.3	54 ± 2	nM	116 ± 2	2 ± 1
TtgR	High	70 ± 20	3.8 ± 0.6	550 ± 50	μM	180 ± 6	4 ± 3
	Low	3 ± 0	2.3 ± 0.4	190 ± 20	μM	25 ± 1	6 ± 1

Individual cells were evaluated by flow cytometry to determine whether the ensemble induction dynamics were indicative of single cell behavior, or were instead an averaged result of stochastic, all-or-nothing responses in individual cells (Fig. 2.4, Fig. A.4). This type of characterization is important, as some inducible systems have been observed to produce bimodal or otherwise heterogeneous induction patterns due to positive feedback or inducer transport properties (61,62). Basal, low and high induction levels were measured after overnight induction. For each biosensor, it was shown that the majority of individual cells adjust their fluorescence in response to inducer concentration. In cases where a small group of cells do not fluoresce, the population represents less than 2% of the total cell population and may consist of dead cells, or cells containing dysfunctional plasmids. Nonetheless, the individual cell responses reflect the population-averaged behavior observed in ensemble measurements. High-copy AraC and CdaR biosensors both have high basal levels of reporter expression when evaluated in bulk (Fig. 2.1a).

Unsurprisingly, these sensors demonstrated the widest uninduced fluorescence distributions when evaluated at the single cell level (Fig. 2.4). The TetR biosensor has a broad fluorescence distribution when partially induced, possibly precluding its use in sensitive induction applications. When partially induced, both low copy MphR and high copy TtgR fluorescence distributions are compressed against the limit of detection (Fig. A.4). This could indicate that the left tail of the distribution is below the limit of detection, or that some cells are not activating in response to the inducer. As observed in the ensemble measurements, TtgR induction is weak. In the case of the low copy TtgR plasmid, the induced and uninduced populations almost entirely overlap when observed by flow cytometry (Fig. A.4).

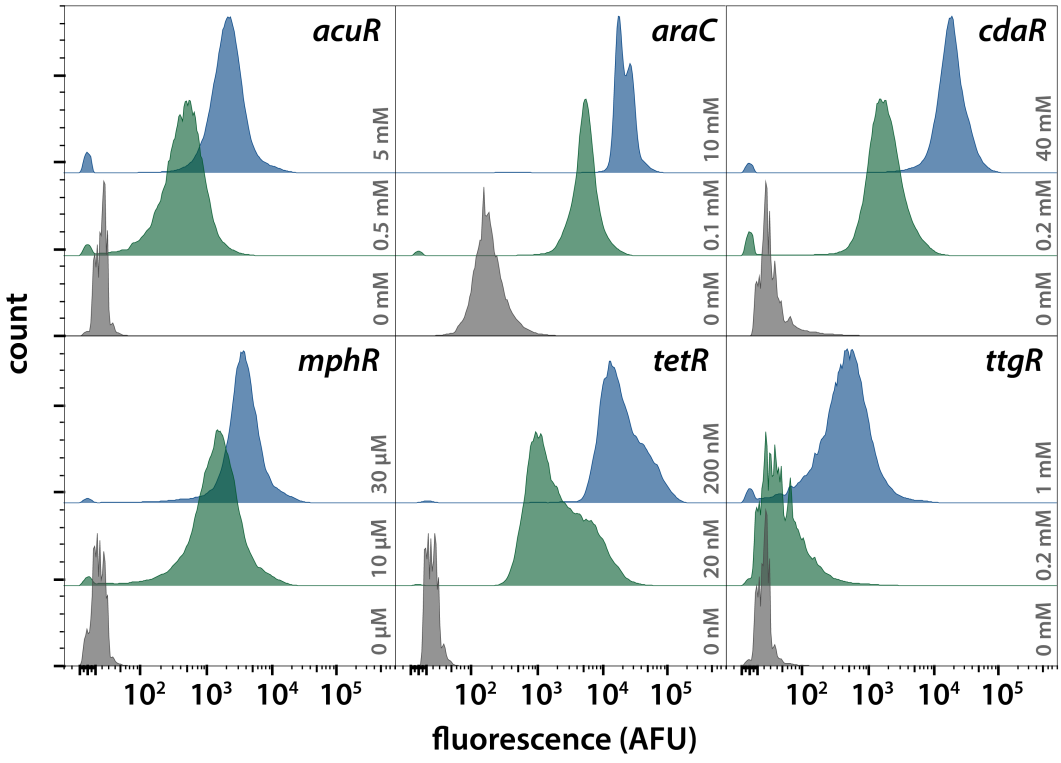


Figure 2.4: The behavior of single cells in response to chemical induction was evaluated by flow cytometry. 100,000 cells from uninduced (grey), partially induced (green) and fully induced (blue) populations were observed for each high copy biosensor. The specific

concentration of inducer is indicated in the plot. Histograms are plotted with a biexponential scale to render the wide range of biosensor activation. The absence of large, well-separated bimodal distributions indicates that bulk fluorescent measurements do indeed reflect the induction behavior of individual cells.

Toxicity of the inducer chemicals was measured to help guide the choice of inducer concentration for future biosensor applications (Fig. 2.5). In applications where maximum protein production is the goal, toxicity will be less of a consideration. In contrast, sub-toxic induction is important for complex circuits that require the cell to maintain a healthy cell state. As expected, erythromycin was toxic to *E. coli* at concentrations as low as 50 μM. However, with expression of the erythromycin resistance gene (*eryR*), only slight toxicity was observed at erythromycin concentrations up to 1.4 mM. A similar growth defect was observed with 430 nM anhydrotetracycline (aTC). Both growth defects are likely due to the solvent, in this case ethanol. Naringenin showed significant toxicity at concentrations of 330 μM and above. This toxicity is likely due to the flavonoid itself, rather than the solvent dimethylsulfoxide (DMSO). Acrylate showed substantial toxicity at 5 mM and 10 mM, corroborating previous observations (63). High concentrations of arabinose resulted in higher growth rates due to *E. coli's* ability to use the sugar as a carbon source. Similar but more modest growth benefits were observed at the highest concentration of glucarate and at low levels of ethanol supplementation.

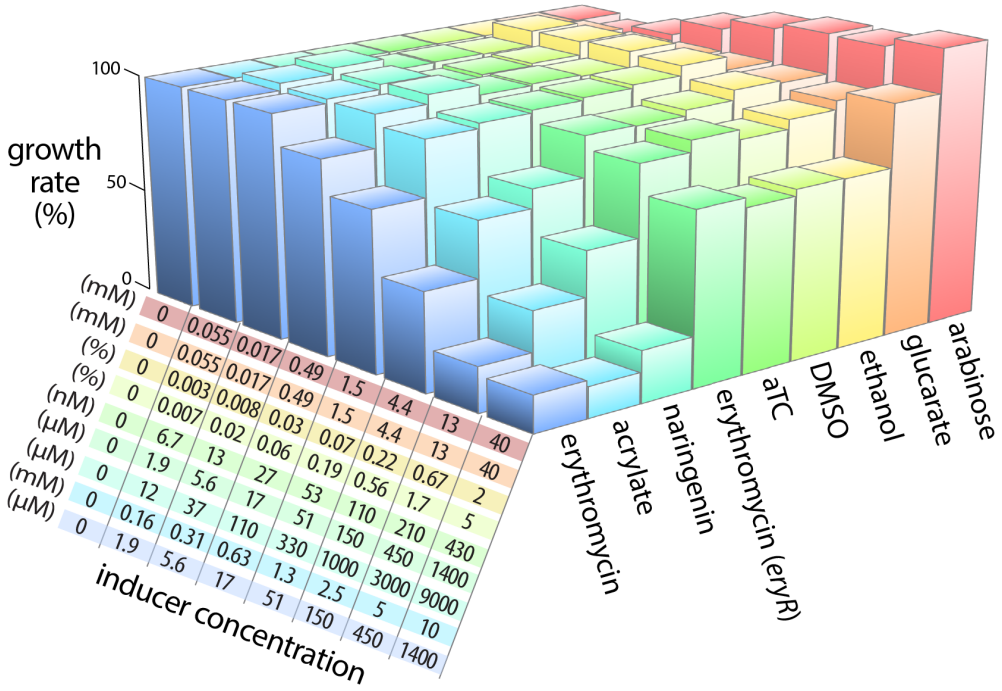


Figure 2.5: The toxicity of each inducer chemical was evaluated over a wide range of concentrations. Growth rate was measured for each combination of chemical and concentration during the exponential phase of growth. Rates were normalized to the growth rate of cells without any added chemical and plotted as bar height. Concentration of each inducer is indicated in the table, corresponding to the bar chart by order and color. Ethanol and DMSO were included as they are the solvents for aTC and naringenin, respectively. Erythromycin was evaluated twice: with and without the erythromycin resistance gene, *eryR*. Inducer concentrations mirror the concentrations used in the induction experiments.

Sensor Orthogonality

The cross-reactivity of each biosensor was evaluated with a panel of inducing compounds: acrylate, arabinose, glucarate, erythromycin, aTC, naringenin, IPTG, rhamnose, cumate and the solvents, DMSO and ethanol. The inducing compounds not otherwise evaluated in this work were included to provide forward compatibility for future biosensor

implementations. No sensor was observed to respond to any of the evaluated compounds except for its cognate inducer (Fig. 2.6). Cumate, glucarate and acrylate all feature a carboxylate, yet are discriminated by their respective sensors. TtgR is known to be promiscuous (57,64), and it is surprising that it is not activated by cumate since it binds many similar molecules, one of which is chloramphenicol (57). TtgR activation by chloramphenicol precludes engineered systems containing TtgR alongside a plasmid maintained by chloramphenicol acetyl transferase.

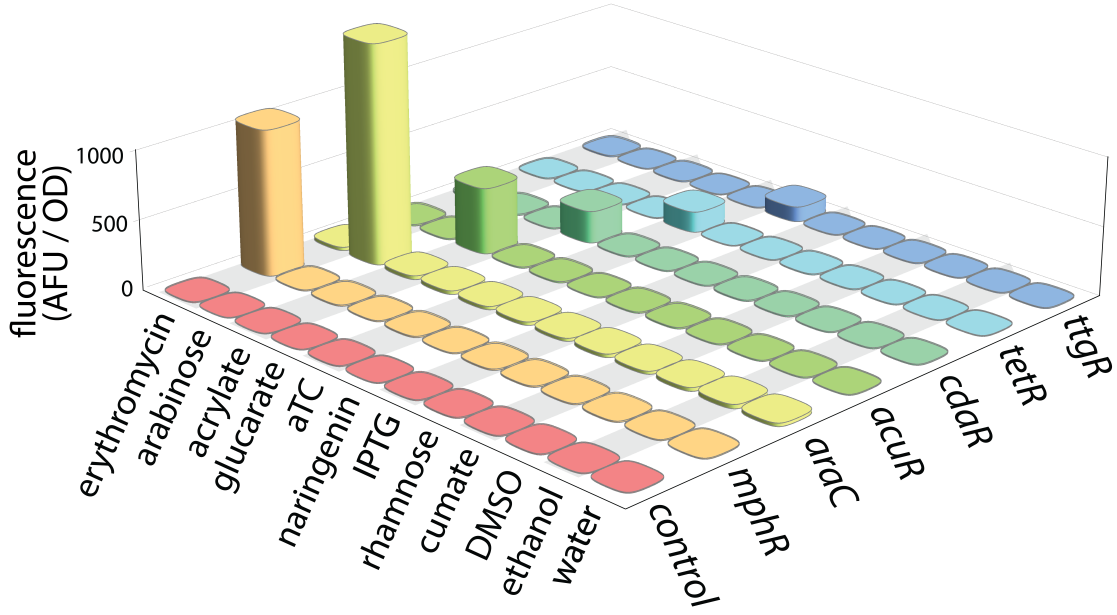


Figure 2.6: The potential for the chemical inducers to activate non-target sensors was evaluated. The cross-reactivity of the new inducers, along with a selection of other commonly used inducers and inducer solvents, was evaluated against each of our six inducible systems. The growth-normalized fluorescent response of each biosensor-inducer pair is plotted as height (n=3). No cross-reactivity was observed.

While the cross-reactivities of the biosensors were evaluated with a single sensor per cell, the real utility of orthogonal sensors comes from controlling a single cell with

multiple sensors. To this end, we redeployed several sensors to allow stable maintenance and non-overlapping fluorescent readouts in the same cell. The MphR biosensor was reconstructed such that erythromycin controlled the expression of mCherry in a vector backbone encoding the p15a replication origin and spectinomycin resistance. Similarly, the AcuR biosensor was reconstructed in a vector backbone encoding the colA origin and kanamycin resistance to facilitate acrylate-mediated expression of CFP. These plasmids were co-transformed with the pJKR-H-CdaR plasmid (encoding GFP) and stably maintained in DH5α cells. Overnight induction of this strain with every combination of glucarate, erythromycin and acrylate induction resulted in eight distinct cell states as measured by fluorescence in the three channels (Fig. 2.7). High, but non-toxic, levels of inducer were chosen for each orthogonal induction channel.

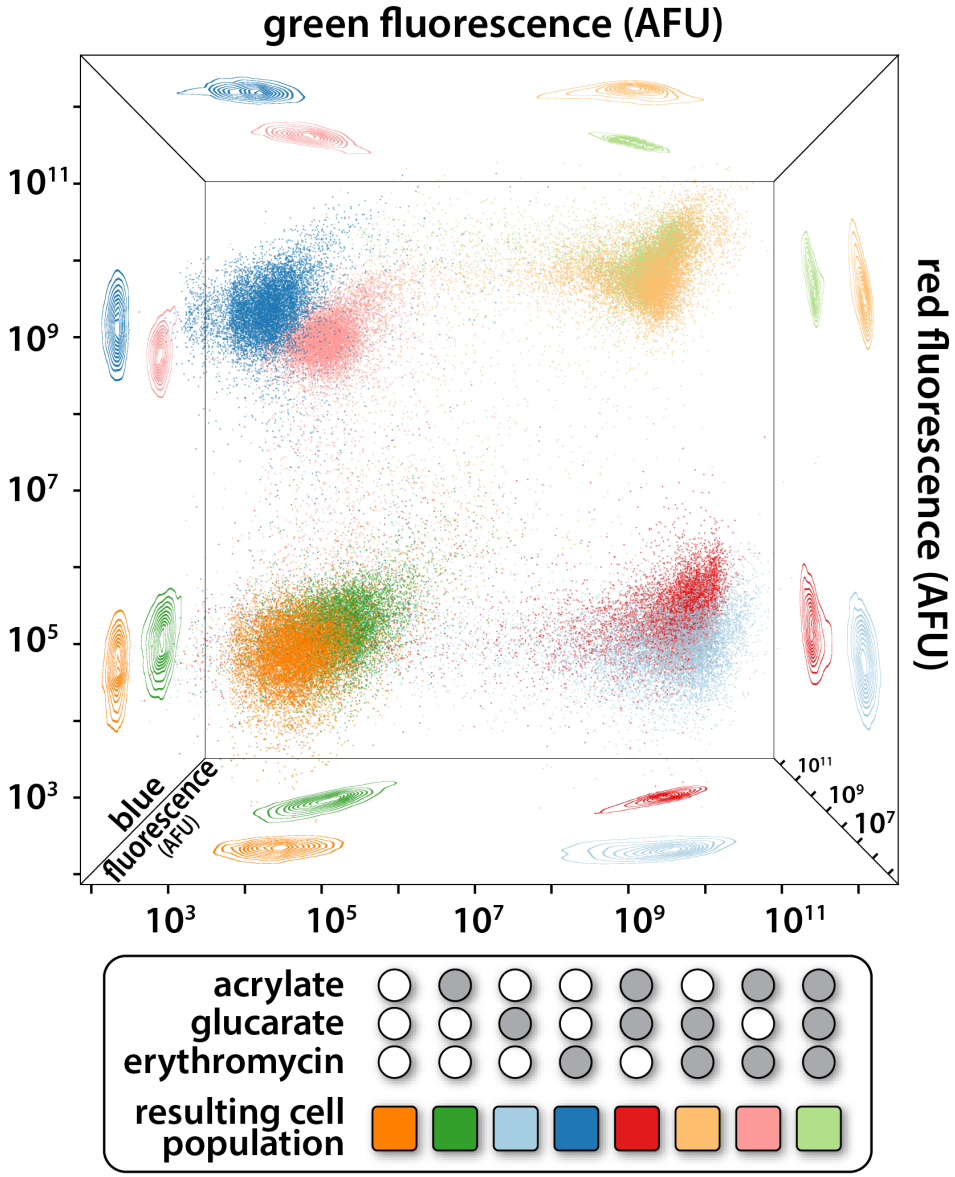


Figure 2.7: Compatible CdaR-GFP, AcuR-CFP and MphR-mCherry biosensors were transformed into the same cell. The potential for these biosensors to be controlled independently was evaluated by flow cytometry. The isogenic cell population was exposed to no inducer (orange), glucarate (light blue), acrylate (dark green), erythromycin (dark blue), glucarate and acrylate (red), glucarate and erythromycin (tan), erythromycin and acrylate (pink) or glucarate, acrylate and erythromycin (light green). The eight combinations of binary induction resulted in eight distinct cell populations when characterized in the three fluorescent channels. The point clouds, each point representing 1 of 10,000 cells, are projected onto the faces of the cube in order to aid in visualization of the 3D space. All axes are log scale to capture the wide range of fluorescent responses.

Flow cytometry was used to evaluate individual cell behavior. Without induction, there is low fluorescence in all channels, as represented by the orange population in Figure 2.7. Induction with only glucarate results in individual cells changing their cell state by producing GFP with no CFP or RFP expression (light blue population in Figure 2.7). The trend continues with induction by erythromycin producing the dark blue cell population exhibiting high fluorescence in the red channel, but low fluorescence in the blue and green channels. Similarly, the dark green points represent acrylate-induced cells with high fluorescence in the blue channel. Induction by all three ligands results in the light green cell population demonstrating high fluorescence in all three channels.

In principle, 16 distinct cell states can be defined with four orthogonal inducible systems, and 32 distinct cell states can be defined with five inducible systems. In these cases, output channels become limiting, as there are limited distinct fluorescent proteins. If three levels of induction are considered (none, intermediate and high), rather than the binary case examined here, the number of cell states increases from 8 to 27 for the system of three orthogonal biosensors, and from 16 to 81 for the theoretical system of four orthogonal biosensors.

Sensors for Metabolic Flux Monitoring

The CdaR biosensor was used to monitor production of glucarate from myo-inositol. Glucarate can be produced from biomass as a renewable replacement for nylon and other plastics (65), however high titers are currently limited by the activity of myo-inositol oxygenase (MIOX) (66), which converts myo-inositol into glucuronate. Glucuronate is in turn oxidized to glucarate by the fast-acting enzyme glucuronate dehydrogenase (Udh). By

co-transforming plasmids containing the CdaR biosensor, a constitutively expressed Udh gene and a library of four constitutively expressed MIOX orthologs, we were able to rapidly identify enzymes producing higher glucarate titers in *E. coli*. The four MIOX variants produced a 20-fold range in fluorescence after 16 hours (Fig. 2.8). Mass spectrometry was used to determine actual glucarate titers in order to determine if biosensor readout was predictive of an enzyme's potential for glucarate production. Glucarate titers were well correlated with fluorescence (Fig. 2.8), encouraging future work in which biosensors enable high throughput discovery and optimization of enzymatic activity. The previously characterized *Mus musculus* MIOX ortholog (66,67) produced the highest fluorescence and titer. Interestingly, a very similar glucarate titer and biosensor response was obtained from the *Flavobacterium johnsoniae* MIOX ortholog that shares only 45% identity with the *Mus musculus* variant.

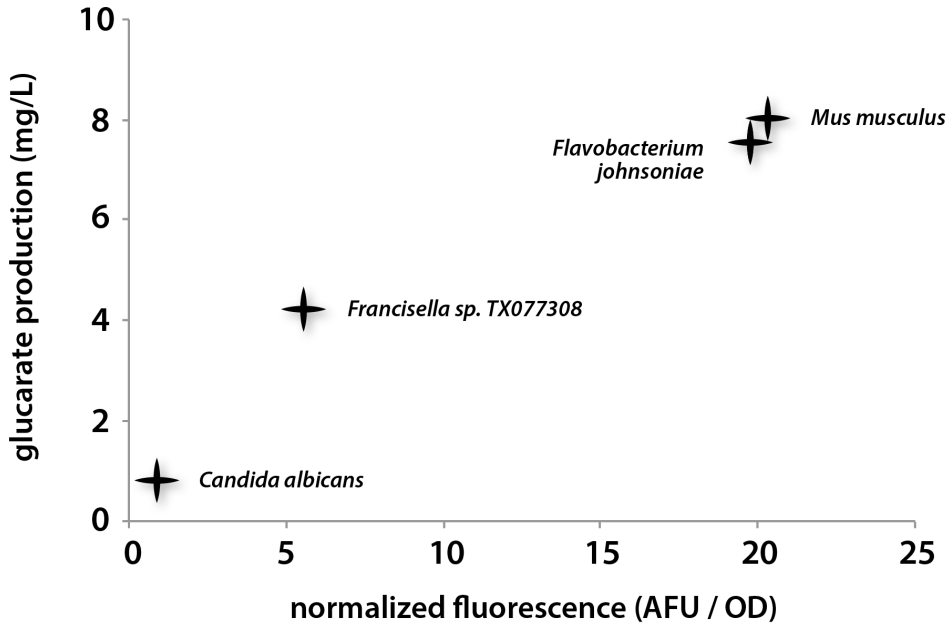


Figure 2.8: Activation of the CdaR biosensor is well correlated with glucarate titers. Glucarate can be produced from myo-inositol by the enzymes MIOX and Udh. MIOX orthologs were transformed into cells containing Udh and the CdaR biosensor. Fluorescence was observed 48 hours after addition of myo-inositol. Glucarate titers were measured after the same period of time in identical strains without the glucarate biosensor. All coefficients of variation are less than 10% (n=3).

Phenol is an important commodity chemical for which novel production routes would provide economic and environmental benefits. Some success has been shown in enzymatic conversion of benzene to phenol (68,69), however these enzymes function with low k_{cat}, and in some cases continue oxidizing phenol to catechol, an undesirable side product (69). We hypothesized that TtgR would be able to respond to phenol as it responds to many other compounds with an activated benzene. Separate populations of the TtgR biosensor strain were incubated with 0.1% phenol, 0.1% catechol and up to 0.4% benzene. Surprisingly, only phenol activated the sensor (Fig. 2.9). The selective response of TtgR to phenol, but not byproducts of phenol production, may enable the directed evolution of

phenol-producing enzymes by coupling phenol production to expression of fluorescent proteins or antibiotic selection markers in individual cells.

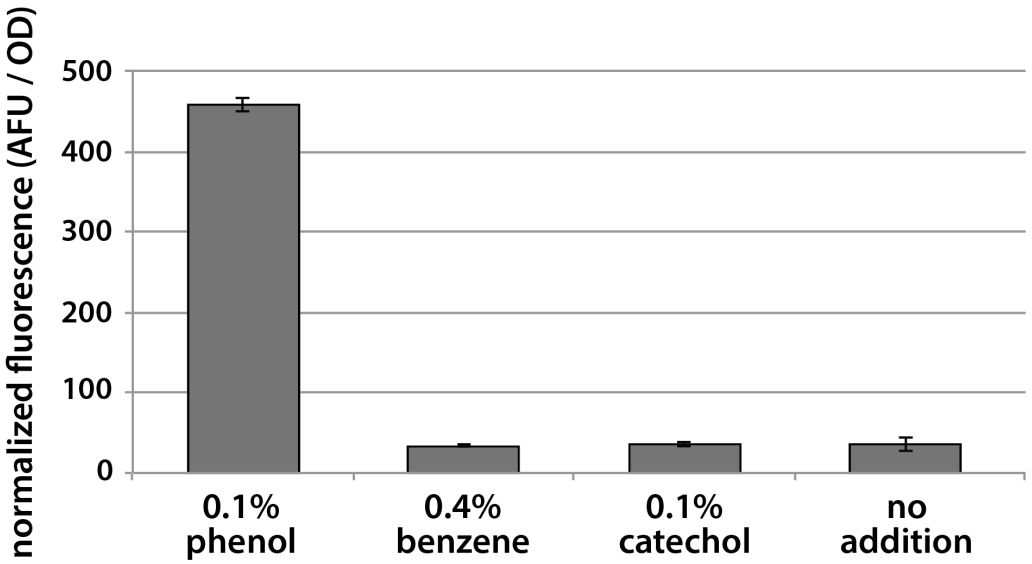


Figure 2.9: The TtgR biosensor was evaluated for its ability to aid in the directed evolution of enzymatic phenol production. A fluorescent response was observed in the presence of 0.1% phenol, while background levels of fluorescence were observed when the sensor was induced with the precursor molecule benzene. Catechol is a side-product of phenol production and did not activate the sensor. All experiments were carried out in triplicate.

Discussion

Enabling the simultaneous control of multiple reporter-coupled genes expands our ability to rapidly probe cellular behavior. Characterization of new and interoperable genetically encoded sensors provides additional input and output nodes for engineered biological systems. Demonstrating the utility of these sensors in metabolic engineering applications provides a basis for future work in directed evolution and enzyme discovery applications.

A balance in the basal strength of the regulated promoter, the copy number of that promoter, and the expression level of the regulator protein is key in achieving tight repression with high dynamic range upon induction. Such behavior has been exploited in previously well-characterized inducible systems to achieve tuned responses to induction (70-72). The TtgR biosensor is an example of a promoter/repressor system that would benefit from further tuning of these parameters. The low dynamic range of the TtgR biosensor and the observation that shifting from a high- to low-copy sensor implementation results in elevated basal reporter expression may be characteristics of a poorly tuned system. Future work will involve multiplexed evaluation of these design specifications in order to understand how system characteristics such as wide dynamic range, low off-state expression and ultra-high expression can be achieved.

By applying a modeling framework to our biosensors, we are able to capture their behavior with a small number of key parameters. Basal and maximal promoter activities have clear biological meaning while the Hill coefficient and lumped half-maximal parameter are more complicated. For inducible systems regulated by a transcription factor, the lumped half-maximal parameter of the Hill equation represents $K \cdot K_d \cdot K_p \cdot T$, where K is an equilibrium constant reflecting the binding event(s) between the transcription factor and the promoter, K_d is the disassociation constant for inducer bound to transcription factor, K_p is the partition coefficient for intracellular to extracellular inducer, and T is the concentration of transcription factor in the cell (73). The K_d is known for several of the transcription factors evaluated here, however the partitioning coefficient and equilibrium of the DNA-protein binding events are not. We did not determine the absolute concentration of transcription factor in the cells, however we note that the lumped

parameter is almost always smaller in the low-copy biosensor variants, reflecting the lower relative concentration of transcription factor. We fit the half-maximal parameter as a lumped value, as we are unable to deconvolute the individual parameters that make it up.

The magnitude of the Hill coefficient reflects the cooperatively of the system, with higher values resulting in more sigmoidal induction curves. The Hill coefficient determined for both low- and high-copy AraC is 1.3, which is in close agreement with previously measured values that range between 1 and 2 (74). The Hill coefficients for low- and high-copy TetR are 3.1 and 4.2, which are similar to previously reported values (75,76). The TtgR biosensor has low- and high-copy Hill coefficients of 2.3 and 3.8, respectively. The CdaR biosensors demonstrated no cooperatively with a Hill coefficient of 1. The lack of cooperatively may be the result of CdaR acting as a monomer, or because there is feedback between biosensor activation and CdaR expression. The moderate Hill coefficient of 1.6 observed for the MphR biosensors reflects a cooperativity substantially lower than the other high-copy repressors. As noted above, the AcuR biosensor shows substantially different induction responses in the high- and low–copy implementations, possibly due to its dependence on growth-phase for induction.

Demonstration of biosensor interoperability is critical for complicated synthetic biological applications in which multiple genes need to be controlled simultaneously (77). We approached interoperability characterization in two ways: by demonstrating that only cognate inducer compounds activated a given sensor (Fig. 2.6) and by combining three biosensors, each controlling an orthogonal output, in a single cell (Fig. 2.7). Ideally, we would have constructed a cell containing every biosensor; however, spectral overlap between fluorescent reporter proteins constrained the system. Reporter protein

expression was similar in strength for singly, doubly and triply induced populations for each of the three biosensors evaluated: AcuR, MphR and CdaR. This result confirms our expectation that cellular behavior prototyped with a single sensor should maintain a similar behavior when operated alongside at least two other inducible systems. While the cell states defined in Figure 2.7 are discussed as a digital system, in reality, induction level is a continuous variable and analog descriptions of system dynamics are more precise (78).

Finally, we demonstrated the use of CdaR and TtgR as biosensors in the observation of intracellular metabolite concentrations. In the case of CdaR, we were able to demonstrate that more effective variants of the MIOX enzyme provided a higher fluorescent signature when the production phenotype was observed through biosensor response. Using fluorescence as a proxy for metabolite production will enable screening millions of enzyme variants per day, rather than the thousands of metabolites that can be screened by relying solely on HPLC or mass spectrometry (24). The TtgR biosensor was found to respond to phenol, a valuable compound that was not previously known to bind and activate the TtgR transcriptional repressor. Crucial for the directed evolution of phenol producing enzymes is the observation that TtgR does not respond to the precursor molecule, benzene, or the unwanted byproduct, catechol.

Innovation in the characterization of biological parts is enabling new options for biological design. Fluorescent activated cell sorting (FACS) combined with multiplexed DNA synthesis and sequencing has been used to characterize biological parts at a throughput of more than 10,000 parts per experiment (79). While these parts were constitutive promoter and RBS variants, advances in microfluidics have enabled new strategies for the characterization of inducible systems. These strategies allow culture

conditions to be changed over time while maintaining cells in specific growth phases (80). Combined with automated microscopy and image processing, these microfluidic platforms open the door for more comprehensive characterization pipelines (81).

Thorough characterization of biological parts is the bedrock for abstraction and complexity in engineered biological systems. Of primary importance are sensors, as they provide channels of communication into and out of the cell. *A priori* design of complex biological systems is challenging, and greater success may be achieved by first engineering a simple reporter that monitors the desired phenotype (*e.g.* metabolite production). Once the desired phenotype produces an output such as fluorescence, millions if not billions of designs can be evaluated rapidly, allowing biological engineering to more closely mimic the process of evolution for which the biological medium was optimized.

Materials and Methods

Chemicals and Reagents

All reagents were obtained from Sigma (St. Louis, MO) unless otherwise noted. Antibiotics and IPTG were obtained from Gold Biotechnology (St. Louis, MO). Anhydrotetracycline (aTC) was obtained from Clontech (Mountain View, CA). Polymerase chain reaction (PCR) mix was purchased from Kapa Biosystems (Wilmington, MA). Erythromycin and aTC were dissolved in ethanol while naringenin was dissolved in dimethyl sulfoxide. All other inducers were dissolved in deionized water.

Plasmid Construction

Plasmids were constructed using Gibson isothermal assembly methods (46) and transformed into DH5α electrocompetent cells (New England Biolabs, Ipswich MA). All standard induction plasmids contained the *rrnB* strong terminator (47) followed by the inducible promoter and the strong g10 RBS (48) 'tttaactttaagaaggagatatacat,' driving the expression of sfGFP (49) (except in the case of CdaR which used the native RBS). GFP was followed by a transcriptional terminator prefixing the proB promoter (50) and strong RBS 'gaaataaggaggtaatacaa,' which facilitated expression of the transcriptional regulator. Each inducible system was implemented on high and low copy plasmids. High copy pJKR-H plasmids were constructed with the pUC origin and beta lactamase antibiotic resistance marker derived from pUC19 (New England Biolabs, Ipswich MA). Low copy pJKR-L plasmids were constructed in the same way, except that the pUC origin was replaced by the SC101 origin (including repA) from pSC101 obtained from American Type Culture Collection (ATCC #37032). In the case of the MphR inducible system, the *eryR* erythromycin resistance cassette was included as well. The sequences of the transcriptional regulators and their cognate promoters are supplied in Table S1. The plasmids MphR-p15a-SPEC-mCherry and AcuR-colA-KAN-CFP were designed for compatible maintenance with pJKR-H-CdaR. In both these plasmids the antibiotic resistance gene and origin of replication were replaced with p15a-*aadA* and colA-*kanR*. Sequence and organism names for each of the MIOX enzymes can be found in Table S2. Each enzymes was cloned downstream of the constitutive promoter P2(41) and g10 RBS to create pJKR-MIOX variants. These expression plasmids used the colA origin of replication

and a kanamycin resistance gene for maintenance. The Udh enzyme was expressed constitutively on a p15a origin of replication providing spectinomycin resistance. Sequences and plasmids are available on Addgene (plasmid numbers 62557-62570).

Induction and Toxicity

DH5α cells transformed with pJKR plasmids and maintained with carbenicillin were used in the induction assays. For each induction evaluation experiment, the cells were grown overnight to saturation before being diluted 1:100 into fresh LB media and incubated at 200RPM and 37°C. After four hours, 150µl of the log-phase cells were transferred to 96-well plates and stock inducer was added to achieve the desired range of induction concentrations. Three separate wells were inoculated and independently supplemented with the appropriate amount of inducing chemical for each level of induction. Measurements were made on the same Biotek (Winooksi, VT) HT plate reader using the same settings: excitation 485/20, emission 528/20, 37°C and fast shaking. Fluorescence and absorbance were measured every 10 minutes for 15 hours. Fluorescence was measured in arbitrary units (AFU) while optical density was determined by absorbance (OD). Normalized fluorescence was determined by dividing fluorescence by optical density for a given measurement. Five independent wells containing control strains, transformed with pUC19, were included to provide a measurement of background autofluorescence. The same protocol modified to observe fluorescence after 90 minutes was used to evaluate induction of the TtgR biosensor with phenol and related compounds.

The ratio of fluorescence to absorbance at 600nm was used in order to compensate for changes in cell density over time and between experiments (AFU / OD). Normalized fluorescence at the 15th hour was used to determine the relationship between inducer concentration and fluorescent response. This transfer function is plotted on a log-log scale in Figure 2.1 to capture the wide range of inducer concentrations and resulting fluorescence values. The Mathematica Hypothesis Testing Package function MeanCI was used to calculate the 95% confidence interval of the estimated mean based upon a Student's t-distribution derived from the three induction replicates. Time-courses of cell growth and biosensor activation were normalized and plotted with the Python module Seaborn using bootstrapping to produce 95% confidence intervals for the standard error of the sample mean (assuming a normal error distribution) for the three independently induced replicates (Figs. 2.2, A.1, A.2). For visualization purposes, normalization was performed on the fluorescence time-course data by dividing all data in a graph by 110% of the highest value such that the trends in each graph can be observed on a common axis. For visualization purposes, growth time-course data was normalized such that each growth curve was divided by its mid-point value and offset to zero at time zero. The raw data produced by all kinetic induction experiments is provided in Supplementary Material.

Induction ratios were determined after 15 hours of induction. Standard error of the derived fold-induction value was determined from the standard error of the mean (assuming a normal error distribution) for the induced, uninduced and control sample means such that the standard error of fold-induction is:

$$\sigma_{\bar{F}} = \bar{F}\sqrt{\frac{\sigma_{\bar{I}}^2 + \sigma_{\bar{C}}^2}{\bar{I}_B^2} + \frac{\sigma_{\bar{U}}^2 + \sigma_{\bar{C}}^2}{\bar{U}_B^2}}$$

$\bar{F}$ is the mean magnitude of the fold induction, $\sigma_{\bar{I}}$, $\sigma_{\bar{U}}$ and $\sigma_{\bar{C}}$ are the SEM for the fluorescence of induced, uninduced and control cells, while $\bar{I}_B$ and $\bar{U}_B$ are the mean fluorescence of the induced and uninduced cells with mean background fluorescence subtracted. For cases where the mean fluorescence of the uninduced cells was within the background fluorescence of the strain, a lower bound on the fold-induction was determined by dividing the 95% confidence interval lower bound of $\bar{I}_B$ by the 95% confidence interval upper bound of $\bar{U}_B$. The range of the 95% confidence interval was approximated by doubling the standard error of the background-subtracted fluorescence.

Toxicity of the inducer chemicals and their solvents were determined at each of the concentrations evaluated for induction response. In these experiments, DH5α cells were diluted 1:100 from overnight growth into fresh selective LB and grown for 2 hours at 200RPM and 37°C. pUC19 was used as a control plasmid in each case except for the erythromycin evaluation in which pJKR-H-MphR was used to provide *eryR* expression. 150µl of cells were transferred into 96-well plates and the assayed chemical was immediately added in triplicate before further incubation at 600RPM and 37°C. After 15 hours the absorbance at 600nm was measured and normalized to the absorbance observed in the control wells in which no chemical was added the cells.

The cross-reactivity matrix was determined by inducing cells that contained target and off-target biosensors. The cells were prepared and evaluated in the same way as described in the induction evaluation experiments. The following inducer concentrations were used: acrylate (5mM), arabinose (165µM), glucarate (4.4mM), erythromycin (51µM), naringenin (9mM), IPTG (1mM), rhamnose (10mM), cumate (20µM), DMSO (1%), ethanol (1%).

Mathematical Modeling

The GFP expression rate was calculated at each time-point with the formula $\Delta GFP/OD$. Scipy was used to perform a non-linear least-squares fit of the maximum GFP expression rate to the corresponding inducer concentration using the Hill function

$$\left(\frac{\Delta GFP}{\Delta t}\right)_{max} = V_{max} \cdot \frac{I^h}{I^h + K_L^h} + V_{min}$$

V_{max} is the maximum rate of GFP expression, V_{min} is basal rate of GFP expression, I is the concentration of inducer, h is the hill coefficient and K_L is the lumped half-maximal parameter described in Results. The variance of each parameter was determined from the least-squares covariance matrix. The square of the variance is the parameter error reported in Table 1. Points in Figures 2.3 and A.3 reflect the mean of three independently induced replicates with error bars corresponding to the 95% confidence interval determined for the standard error of the mean by bootstrapping. Lines reflect the model fitted to the data.

Flow Cytometry

D5Hα cells containing the plasmid to be evaluated were grown to saturation overnight and diluted 1:100 in 1mL of selective LB media and incubated in 96-well deep well blocks at 900RPM and 37°C. After 4 hours, inducers were added to the desired final concentration and incubation was resumed for 15 hours. Induced cultures were diluted 1:100 in cold phosphate buffered saline (PBS) and kept on ice until evaluated on the LSRFortessa flow cytometer (BD Biosciences, San Jose, CA). At least 100,000 events were

captured for each sample. Gating was performed on forward and side scatter to avoid debris and clumped cells. Data was exported to FloJo for visualization and Mathematica for subsequent analysis.

Cells that were transformed with pJKR-H-CdaR, MphR-p15a-SPEC-mCherry (designated pJKR-O-MphR) and AcuR-colA-KAN-CFP (designated pJKR-O-AcuR) were maintained in LB with all three antibiotics. The cells were induced in the same manner as above with induction concentrations of 5mM acrylate, 40mM glucarate and 37μg/mL erythromycin. Collection and gating was performed as above. 10,000 events were plotted in Figure 2.7 for clarity.

Glucarate Production

For observation of glucarate production via fluorescence, BL21 DE3 (New England Biolabs, Ipswich MA) cells that were triply transformed with pJKR-H-CdaR, pJKR-UDH and pJKR-MIOX were diluted 1:100 from saturated culture into carbenicillin, spectinomycin and kanamycin selective LB. After 4 hours the cells were transferred to 96-well plates in triplicate and 50mM myo-inositol was added to the media. Fluorescence and absorbance (600nm) were measured with a Biotek HT plate reader in 15-minute intervals for 48 hours with fast shaking at 37°C.

In order to directly measure glucarate titers, BL21 DE3 cells transformed with pJKR-MIOX variants were prepared as above, except for production took place in 1mL cultures within 96-well deep well blocks incubating at 900RPM and 37°C for 48 hours. Supernatants were collected via centrifugation and filtration and glucarate was determined by mass spectrometry.

References

1. Gilbert, W. and Muller-Hill, B. (1966) Isolation of the lac repressor. *Proceedings of the National Academy of Sciences of the United States of America*, **56**, 1891-1898.

2. 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.

3. Ceroni, F., Furini, S., Giordano, E. and Cavalcanti, S. (2010) Rational design of modular circuits for gene transcription: A test of the bottom-up approach. *Journal of biological engineering*, **4**, 14.

4. Baneyx, F. (1999) Recombinant protein expression in Escherichia coli. *Current opinion in biotechnology*, **10**, 411-421.

5. 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.

6. Litcofsky, K.D., Afeyan, R.B., Krom, R.J., Khalil, A.S. and Collins, J.J. (2012) Iterative plug-and-play methodology for constructing and modifying synthetic gene networks. *Nat Methods*, **9**, 1077-1080.

7. Belmont, A.S. and Straight, A.F. (1998) In vivo visualization of chromosomes using lac operator-repressor binding. *Trends in cell biology*, **8**, 121-124.

8. Guzman, L.M., Belin, D., Carson, M.J. and Beckwith, J. (1995) Tight Regulation, Modulation, and High-Level Expression by Vectors Containing the Arabinose P-Bad Promoter. *J Bacteriol*, **177**, 4121-4130.

9. Canton, B., Labno, A. and Endy, D. (2008) Refinement and standardization of synthetic biological parts and devices. *Nature biotechnology*, **26**, 787-793.

10. Lee, S.K. and Keasling, J.D. (2005) A propionate-inducible expression system for enteric bacteria. *Applied and environmental microbiology*, **71**, 6856-6862.

11. Wegerer, A., Sun, T. and Altenbuchner, J. (2008) Optimization of an E. coli L-rhamnose-inducible expression vector: test of various genetic module combinations. *BMC biotechnology*, **8**, 2.

12. Choi, Y.J., Morel, L., Le Francois, T., Bourque, D., Bourget, L., Groleau, D., Massie, B. and Miguez, C.B. (2010) Novel, versatile, and tightly regulated expression system for Escherichia coli strains. *Applied and environmental microbiology*, **76**, 5058-5066.

13. Mermod, N., Ramos, J.L., Lehrbach, P.R. and Timmis, K.N. (1986) Vector for regulated expression of cloned genes in a wide range of gram-negative bacteria. *J Bacteriol*, **167**, 447-454.

14. Wittmann, A. and Suess, B. (2012) Engineered riboswitches: Expanding researchers' toolbox with synthetic RNA regulators. *FEBS letters*, **586**, 2076-2083.

15. Drepper, T., Krauss, U., Meyer zu Berstenhorst, S., Pietruszka, J. and Jaeger, K.E. (2011) Lights on and action! Controlling microbial gene expression by light. *Applied microbiology and biotechnology*, **90**, 23-40.

16. Manasherob, R., Zaritsky, A., Metzler, Y., Ben-Dov, E., Itsko, M. and Fishov, I. (2003) Compaction of the Escherichia coli nucleoid caused by Cyt1Aa. *Microbiology*, **149**, 3553-3564.

17. Gueiros, F.J. and Losick, R. (2002) A widely conserved bacterial cell division protein that promotes assembly of the tubulin-like protein FtsZ. *Gene Dev*, **16**, 2544-2556.

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

19. Giaever, G., Chu, A.M., Ni, L., Connelly, C., Riles, L., Veronneau, S., Dow, S., Lucau-Danila, A., Anderson, K., Andre, B. *et al.* (2002) Functional profiling of the Saccharomyces cerevisiae genome. *Nature*, **418**, 387-391.

20. Kobayashi, K., Ehrlich, S.D., Albertini, A., Amati, G., Andersen, K.K., Arnaud, M., Asai, K., Ashikaga, S., Aymerich, S., Bessieres, P. *et al.* (2003) Essential Bacillus subtilis genes. *Proceedings of the National Academy of Sciences of the United States of America*, **100**, 4678-4683.

21. Wang, H.H., Kim, H., Cong, L., Jeong, J., Bang, D. and Church, G.M. (2012) Genome-scale promoter engineering by coselection MAGE. *Nat Methods*, **9**, 591-+.

22. Wang, H.H., Isaacs, F.J., Carr, P.A., Sun, Z.Z., Xu, G., Forest, C.R. and Church, G.M. (2009) Programming cells by multiplex genome engineering and accelerated evolution. *Nature*, **460**, 894-U133.

23. Balagadde, F.K., Song, H., Ozaki, J., Collins, C.H., Barnet, M., Arnold, F.H., Quake, S.R. and You, L.C. (2008) A synthetic Escherichia coli predator-prey ecosystem. *Mol Syst Biol*, **4**.

24. Dietrich, J.A., McKee, A.E. and Keasling, J.D. (2010) High-Throughput Metabolic Engineering: Advances in Small-Molecule Screening and Selection. *Annu Rev Biochem*, **79**, 563-590.

25. Ogasawara, H., Ishida, Y., Yamada, K., Yamamoto, K. and Ishihama, A. (2007) PdhR (pyruvate dehydrogenase complex regulator) controls the respiratory electron transport system in Escherichia coli. *J Bacteriol*, **189**, 5534-5541.

26. Cortay, J.C., Negre, D., Galinier, A., Duclos, B., Perriere, G. and Cozzone, A.J. (1991) Regulation of the Acetate Operon in Escherichia-Coli - Purification and Functional-Characterization of the Icir Repressor. *Embo J*, **10**, 675-679.

27. Martin, M.G., Magni, C., de Mendoza, D. and Lopez, P. (2005) CitI, a transcription factor involved in regulation of citrate metabolism in lactic acid bacteria. *J Bacteriol*, **187**, 5146-5155.

28. Gao, C., Hu, C.H., Zheng, Z.J., Ma, C.Q., Jiang, T.Y., Dou, P.P., Zhang, W., Che, B., Wang, Y.J., Lv, M. *et al.* (2012) Lactate Utilization Is Regulated by the FadR-Type Regulator LldR in Pseudomonas aeruginosa. *J Bacteriol*, **194**, 2687-2692.

29. Henry, M.F. and Cronan, J.E. (1991) Escherichia-Coli Transcription Factor That Both Activates Fatty-Acid Synthesis and Represses Fatty-Acid Degradation. *J Mol Biol*, **222**, 843-849.

30. Wang, E., Bauer, M.C., Rogstam, A., Linse, S., Logan, D.T. and von Wachenfeldt, C. (2008) Structure and functional properties of the Bacillus subtilis transcriptional repressor Rex. *Mol Microbiol*, **69**, 466-478.

31. van Sint Fiet, S., van Beilen, J.B. and Witholt, B. (2006) Selection of biocatalysts for chemical synthesis. *Proceedings of the National Academy of Sciences of the United States of America*, **103**, 1693-1698.

32. Dietrich, J.A., Shis, D.L., Alikhani, A. and Keasling, J.D. (2013) Transcription factor-based screens and synthetic selections for microbial small-molecule biosynthesis. *ACS synthetic biology*, **2**, 47-58.

33. Tang, S.Y. and Cirino, P.C. (2011) Design and application of a mevalonate-responsive regulatory protein. *Angewandte Chemie*, **50**, 1084-1086.

34. Tang, S.Y., Qian, S., Akinterinwa, O., Frei, C.S., Gredell, J.A. and Cirino, P.C. (2013) Screening for enhanced triacetic acid lactone production by recombinant Escherichia coli expressing a designed triacetic acid lactone reporter. *Journal of the American Chemical Society*, **135**, 10099-10103.

35. Yang, J., Seo, S.W., Jang, S., Shin, S.I., Lim, C.H., Roh, T.Y. and Jung, G.Y. (2013) Synthetic RNA devices to expedite the evolution of metabolite-producing microbes. *Nature communications*, **4**, 1413.

36. Binder, S., Schendzielorz, G., Stabler, N., Krumbach, K., Hoffmann, K., Bott, M. and Eggeling, L. (2012) A high-throughput approach to identify genomic variants of bacterial metabolite producers at the single-cell level. *Genome biology*, **13**, R40.

37. Pitera, D.J., Paddon, C.J., Newman, J.D. and Keasling, J.D. (2007) Balancing a heterologous mevalonate pathway for improved isoprenoid production in Escherichia coli. *Metab Eng*, **9**, 193-207.

38. Shiue, E. (2014) Improvement of D-glucaric acid production in Escherichia coli. MIT.

39. 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. *Nature biotechnology*, **30**, 354-U166.

40. Raman, S., Rogers, J.K., Taylor, N.D. and Church, G.M. (2014) Evolution-guided optimization of biosynthetic pathways. *Proceedings of the National Academy of Sciences of the United States of America*, **111**, 17803-17808.

41. Mutalik, V.K., Guimaraes, J.C., Cambray, G., Lam, C., Christoffersen, M.J., Mai, Q.A., Tran, A.B., Paull, M., Keasling, J.D., Arkin, A.P. *et al.* (2013) Precise and reliable gene expression via standard transcription and translation initiation elements. *Nat Methods*, **10**, 354-360.

42. Cambray, G., Guimaraes, J.C., Mutalik, V.K., Lam, C., Mai, Q.A., Thimmaiah, T., Carothers, J.M., Arkin, A.P. and Endy, D. (2013) Measurement and modeling of intrinsic transcription terminators. *Nucleic acids research*, **41**, 5139-5148.

43. Mutalik, V.K., Guimaraes, J.C., Cambray, G., Mai, Q.A., Christoffersen, M.J., Martin, L., Yu, A., Lam, C., Rodriguez, C., Bennett, G. *et al.* (2013) Quantitative estimation of activity and quality for collections of functional genetic elements. *Nat Methods*, **10**, 347-353.

44. Lillacci, G. and Khammash, M. (2013) The signal within the noise: efficient inference of stochastic gene regulation models using fluorescence histograms and stochastic simulations. *Bioinformatics*, **29**, 2311-2319.

45. Balzer, S., Kucharova, V., Megerle, J., Lale, R., Brautaset, T. and Valla, S. (2013) A comparative analysis of the properties of regulated promoter systems commonly used for recombinant gene expression in Escherichia coli. *Microbial cell factories*, **12**, 26.

46. Gibson, D.G., Young, L., Chuang, R.Y., Venter, J.C., Hutchison, C.A., 3rd and Smith, H.O. (2009) Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat Methods*, **6**, 343-345.

47. Orosz, A., Boros, I. and Venetianer, P. (1991) Analysis of the complex transcription termination region of the Escherichia coli rrnB gene. *European journal of biochemistry / FEBS*, **201**, 653-659.

48. Olins, P.O., Devine, C.S., Rangwala, S.H. and Kavka, K.S. (1988) The T7 phage gene 10 leader RNA, a ribosome-binding site that dramatically enhances the expression of foreign genes in Escherichia coli. *Gene*, **73**, 227-235.

49. Pedelacq, J.D., Cabantous, S., Tran, T., Terwilliger, T.C. and Waldo, G.S. (2006) Engineering and characterization of a superfolder green fluorescent protein. *Nature biotechnology*, **24**, 79-88.

50. Davis, J.H., Rubin, A.J. and Sauer, R.T. (2011) Design, construction and characterization of a set of insulated bacterial promoters. *Nucleic acids research*, **39**, 1131-1141.

51. Sullivan, M.J., Curson, A.R., Shearer, N., Todd, J.D., Green, R.T. and Johnston, A.W. (2011) Unusual regulation of a leaderless operon involved in the catabolism of dimethylsulfoniopropionate in Rhodobacter sphaeroides. *PloS one*, **6**, e15972.

52. Monterrubio, R., Baldoma, L., Obradors, N., Aguilar, J. and Badia, J. (2000) A common regulator for the operons encoding the enzymes involved in D-galactarate, D-glucarate, and D-glycerate utilization in Escherichia coli. *J Bacteriol*, **182**, 2672-2674.

53. Noguchi, N., Takada, K., Katayama, J., Emura, A. and Sasatsu, M. (2000) Regulation of transcription of the mph(A) gene for macrolide 2'-phosphotransferase I in Escherichia coli: characterization of the regulatory gene mphR(A). *J Bacteriol*, **182**, 5052-5058.

54. Kramer, B.P., Viretta, A.U., Daoud-El-Baba, M., Aubel, D., Weber, W. and Fussenegger, M. (2004) An engineered epigenetic transgene switch in mammalian cells. *Nature biotechnology*, **22**, 867-870.

55. Weber, W., Fux, C., Daoud-el Baba, M., Keller, B., Weber, C.C., Kramer, B.P., Heinzen, C., Aubel, D., Bailey, J.E. and Fussenegger, M. (2002) Macrolide-based transgene control in mammalian cells and mice. *Nature biotechnology*, **20**, 901-907.

56. Mohrle, V., Stadler, M. and Eberz, G. (2007) Biosensor-guided screening for macrolides. *Analytical and bioanalytical chemistry*, **388**, 1117-1125.

57. Teran, W., Felipe, A., Segura, A., Rojas, A., Ramos, J.L. and Gallegos, M.T. (2003) Antibiotic-dependent induction of Pseudomonas putida DOT-T1E TtgABC efflux pump is mediated by the drug binding repressor TtgR. *Antimicrobial agents and chemotherapy*, **47**, 3067-3072.

58. Rossger, K., Charpin-El-Hamri, G. and Fussenegger, M. (2013) A closed-loop synthetic gene circuit for the treatment of diet-induced obesity in mice. *Nature communications*, **4**, 2825.

59. Hahn, S. and Schleif, R. (1983) In vivo regulation of the Escherichia coli araC promoter. *J Bacteriol*, **155**, 593-600.

60. Andersen, J.B., Sternberg, C., Poulsen, L.K., Bjorn, S.P., Givskov, M. and Molin, S. (1998) New unstable variants of green fluorescent protein for studies of transient gene expression in bacteria. *Applied and environmental microbiology*, **64**, 2240-2246.

61. Hansen, L.H., Knudsen, S. and Sorensen, S.J. (1998) The effect of the lacY gene on the induction of IPTG inducible promoters, studied in Escherichia coli and Pseudomonas fluorescens. *Current microbiology*, **36**, 341-347.

62. Khlebnikov, A., Risa, O., Skaug, T., Carrier, T.A. and Keasling, J.D. (2000) Regulatable arabinose-inducible gene expression system with consistent control in all cells of a culture. *J Bacteriol*, **182**, 7029-7034.

63. Todd, J.D., Curson, A.R., Sullivan, M.J., Kirkwood, M. and Johnston, A.W. (2012) The Ruegeria pomeroyi acuI gene has a role in DMSP catabolism and resembles yhdH of E. coli and other bacteria in conferring resistance to acrylate. *PloS one*, **7**, e35947.

64. Teran, W., Krell, T., Ramos, J.L. and Gallegos, M.T. (2006) Effector-repressor interactions, binding of a single effector molecule to the operator-bound TtgR homodimer mediates derepression. *The Journal of biological chemistry*, **281**, 7102-7109.

65. Werpy, T.a.P., G. (2004) In Energy, U. S. D. o. (ed.).

66. Moon, T.S., Yoon, S.H., Lanza, A.M., Roy-Mayhew, J.D. and Prather, K.L. (2009) Production of glucaric acid from a synthetic pathway in recombinant Escherichia coli. *Applied and environmental microbiology*, **75**, 589-595.

67. Shiue, E. and Prather, K.L. (2014) Improving D-glucaric acid production from myo-inositol in E. coli by increasing MIOX stability and myo-inositol transport. *Metab Eng*, **22**, 22-31.

68. Farinas, E.T., Alcalde, M. and Arnold, F. (2004) Alkene epoxidation catalyzed by cytochrome P450 BM-3 139-3. *Tetrahedron*, **60**, 525-528.

69. Karich, A., Kluge, M., Ullrich, R. and Hofrichter, M. (2013) Benzene oxygenation and oxidation by the peroxygenase of Agrocybe aegerita. *AMB Express*, **3**, 5.

70. Moser, F., Horwitz, A., Chen, J., Lim, W. and Voigt, C.A. (2013) Genetic sensor for strong methylating compounds. *ACS synthetic biology*, **2**, 614-624.

71. Wang, B., Barahona, M. and Buck, M. (2013) A modular cell-based biosensor using engineered genetic logic circuits to detect and integrate multiple environmental signals. *Biosensors & bioelectronics*, **40**, 368-376.

72. Zucca, S., Pasotti, L., Mazzini, G., De Angelis, M.G. and Magni, P. (2012) Characterization of an inducible promoter in different DNA copy number conditions. *BMC bioinformatics*, **13 Suppl 4**, S11.

73. Sauro, H.M. (2012) Enzyme Kinetics for Systems Biology. Ambrosius Publishing.

74. 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 systems biology*, **5**, 111.

75. Hooshangi, S., Thiberge, S. and Weiss, R. (2005) Ultrasensitivity and noise propagation in a synthetic transcriptional cascade. *Proceedings of the National Academy of Sciences of the United States of America*, **102**, 3581-3586.

76. Nevozhay, D., Adams, R.M., Murphy, K.F., Josic, K. and Balazsi, G. (2009) Negative autoregulation linearizes the dose-response and suppresses the heterogeneity of gene expression. *Proceedings of the National Academy of Sciences of the United States of America*, **106**, 5123-5128.

77. Lee, S.K., Chou, H.H., Pfleger, B.F., Newman, J.D., Yoshikuni, Y. and Keasling, J.D. (2007) Directed evolution of AraC for improved compatibility of arabinose- and lactose-inducible promoters. *Applied and environmental microbiology*, **73**, 5711-5715.

78. Sarpeshkar, R. (2014) Analog synthetic biology. *Philosophical transactions. Series A, Mathematical, physical, and engineering sciences*, **372**, 20130110.

79. Kosuri, S., Goodman, D.B., Cambray, G., Mutalik, V.K., Gao, Y., Arkin, A.P., Endy, D. and Church, G.M. (2013) Composability of regulatory sequences controlling transcription and translation in Escherichia coli. *Proceedings of the National Academy of Sciences of the United States of America*, **110**, 14024-14029.

80. Ball, D.A., Lux, M.W., Adames, N.R. and Peccoud, J. (2014) Adaptive imaging cytometry to estimate parameters of gene networks models in systems and synthetic biology. *PloS one*, **9**, e107087.

81. Charvin, G., Cross, F.R. and Siggia, E.D. (2008) A microfluidic device for temporally controlled gene expression and long-term fluorescent imaging in unperturbed dividing yeast cells. *PloS one*, **3**, e1468.

Chapter 3: Genetically encoded sensors enable real-time observation of metabolite production

Abstract

Engineering cells to produce valuable metabolic products is hindered by the low-throughput methods available for evaluating candidate designs. Even as our capacity to design and build genetic variants surpasses billions of cells per day, the throughput of direct metabolite measurement remains limited to hundreds of samples per day. Consequently, the vast majority of designs go unevaluated. In this paper we develop a framework for observing product formation in real-time without the need for sample preparation or laborious analytical methods, in turn laying the foundation for evaluation rates of up to a billion cells per day. We use genetically encoded biosensors to provide a fluorescent readout that is proportional to the intracellular concentration of a target metabolite. Combining an appropriate biosensor with cells designed to produce a metabolic product allows us to track product formation by observing fluorescence. With individual cells exhibiting fluorescent intensities proportional to the amount of metabolite they produce, high-throughput methods such as flow-cytometry can be used to rank the quality of billions of genetic variants per day. We observe production of several renewable plastic precursors with fluorescent readouts and demonstrate that higher fluorescence is indeed an indicator of higher product titer. Using fluorescence as a guide, we identify culture conditions that produce 3-hydroxypropionate at 4.2 g/L, 23-fold higher than previously reported. We also report the first direct biological production of acrylate, a plastic precursor with global sales of $14 billion. Finally, we monitor the production of

glucarate, a replacement for environmentally damaging detergents, and muconate, a renewable precursor to PET and nylon with combined markets of $51 billion, in real-time, demonstrating that our method is applicable to a wide range of molecules. Fluorescent monitoring of product formation enables high-throughput phenotype evaluation and alleviates the final bottleneck in the metabolic engineering design cycle.

Introduction

Biological production of valuable products such as pharmaceuticals or renewable chemicals holds the potential to transform the global economy. However, the rate at which bioengineers are able to engineer new living catalysts is hampered by an exceedingly slow design-build-test cycle. We describe a method to accelerate the design-build-test cycle for metabolic engineering by several orders of magnitude, enabling rapid gains for the bioeconomy.

Biological production of a desired product is accomplished by guiding a low-cost starting material such as glucose through a series of intracellular enzymatic reactions, ultimately yielding a molecule of economic interest. The choice of the catalytic strategy, the creation of enzyme variants, and the tuning of endogenous cellular metabolism create a vast universe of potential designs. Because of the complexity of biology, appropriate designs are not known *a priori*. Even sophisticated modeling paradigms can result in design spaces that are on the orders of millions and billions (1,2). Until recently, generation of this kind of targeted genetic diversity was impossible, but with the cost of DNA synthesis rapidly dropping and the advent of multiplexed genome editing (2), the creation of massive libraries of defined mutations is now possible. Evaluating each cellular design for the

desired phenotype is therefore the last major bottleneck in the bioengineering design-build-test cycle.

Current methods for evaluating biological production of chemicals rely on slow and laborious techniques such as high-pressure liquid chromatography (HPLC) and mass spectrometry (3). Each potential design must be separated from the other cellular designs and cultured independently before the compound of interest is isolated and run through the desired analytical system. Generous estimates of the throughput of these methods are around thousands of samples per day in highly specialized labs (3). These rates of evaluation are exceedingly small compared to typical enzyme library sizes, or the $1x10^9$ unique genomes that can be built in a day using multiplexed genome engineering (2).

Engineers have begun developing strategies for multiplexed evaluation of metabolite production phenotypes in order to enable a fully multiplexed design-build-test cycle (1,4-7). Such a design cycle would more closely resemble biological evolution, rather than the design approaches inspired by electrical engineering that currently dominate the fields of synthetic biology and metabolic engineering. Strategies for multiplexed design evaluation include selections and screens. Selections, which only allow cells exhibiting a desired phenotype to live, have the potential to evaluate billions of designs simultaneously (3). Sensor-selectors are standardized selections aimed at enhancing metabolite production for a wide range of compounds (1). Selections are limited by their false-positive rate and can be challenging to troubleshoot, especially if high production of the metabolite of interest provides a negative growth phenotype. Genetically encoded biosensors link intracellular metabolite levels to fluorescent protein expression and enable fluorescence-based screens. Combined with fluorescent activated cell sorting (FACS), biosensor-based

screens provide evaluation rates of up to $1x10^9$ designs per day (3). Despite the power of FACS for design evaluation, few attempts have been made at optimizing metabolic pathways due to a lack of biosynthesis pathways coupled to appropriately characterized biosensors (8).

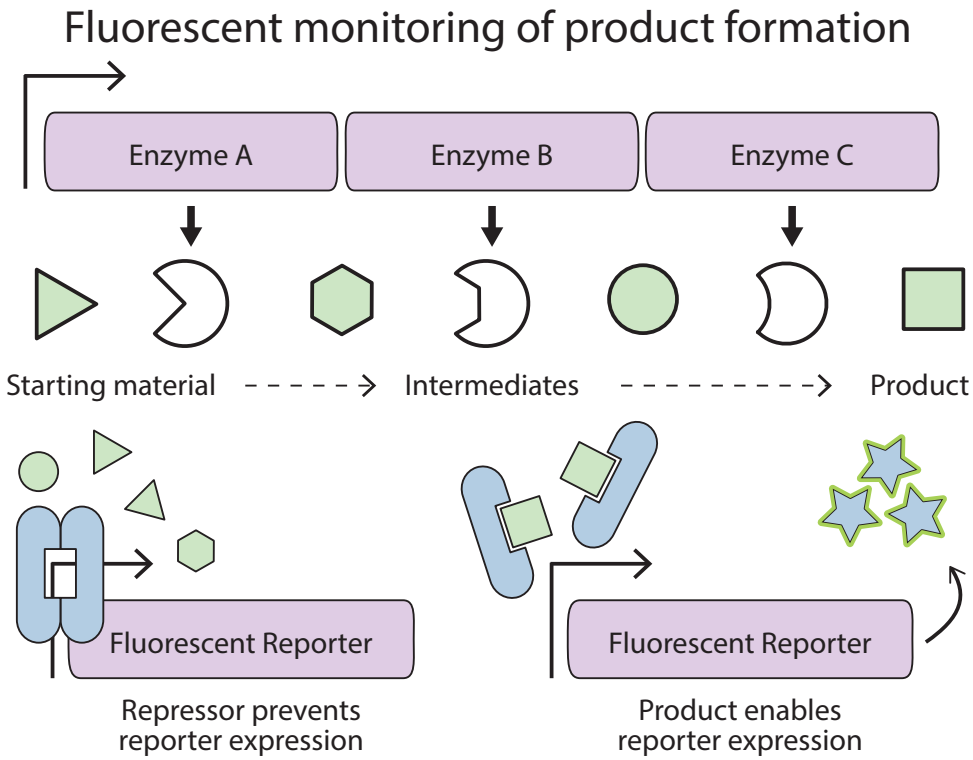


Figure 3.1: Biosensors enable real-time monitoring of biological chemical production. A metabolic pathway converts a low-cost starting material such as glucose (green triangle) to a desired product (green square) through a series of enzymatic reactions. A biosensor (blue ellipse) regulates the production of a fluorescent reporter by preventing transcription in the absence of the correct inducing molecule. The fluorescent response of the biosensor is proportional to the amount of product produced by the cell because the starting material and intermediates do not activate the biosensor. Each cell reports its rate of chemical production through its fluorescence intensity. Cells with high intensity, and consequently highly productive metabolic pathways, are easily identified.

Genetically encoded biosensors are the most versatile method for coupling cellular fluorescence to the quality of a metabolic engineering design. Genetically encoded biosensors link the expression of a fluorescent protein to the intracellular concentration of a target metabolite through the use of an allosteric transcription factor. In addition to the classic small-molecule inducible systems such as *lacI*-IPTG and *tetR*-aTC (9), many new biosensors have been characterized that respond to valuable products as diverse as macrolide antibiotics, flavonoids and plastic precursors. A biosensor designed to respond to a product, while ignoring intermediates, allows the fluorescent readout to act as a real-time proxy for product formation from glucose or other precursors (Fig. 3.1). This allows simple observation of performance characteristics such as the rate of product formation or the titer, all without the need for low-throughput analytical pipelines.

In this work we develop the first strategy for real-time monitoring of metabolic product formation and demonstrate its utility in observing the production of 3-hydroxypropionate (a renewable plastic precursor) (10), acrylate (the monomer for several common plastics), glucarate (a renewable building block for superabsorbent polymers and a replacement for phosphates in detergents) (11) and muconate (a building block for renewable nylon) (12). We develop two unique biosensors for 3-hydroxypropionate (3HP) and compare their ability to observe 3HP production. We use the real-time observation of 3HP formation to select culture conditions that result in a 23-fold increase in 3HP production over previously reported titers. We achieve the first direct biological route to acrylate by converting 3HP to acrylate *in vivo*. We go on to demonstrate the generalizable nature of real-time metabolite observation by deploying glucarate and muconate biosensors with their respective heterologous metabolic pathways.

We aim to alleviate the primary bottleneck in the bioengineering design-build-test cycle and enable next-generation metabolic engineering by creating a phenotype evaluation framework that can be applied to a wide range of products and pathways.

Results and Discussion

The framework for real-time observation of metabolite production consists of two components: the biosensor and the pathway (Fig. 3.1). The biosensor is a small-molecule inducible system that produces a fluorescent readout proportional to the amount of product inside of the cell. The pathway consists of all the genes necessary to produce the product molecule from a desired starting point, typically a low-cost feedstock such as glucose or biomass. Rates of product formation vary depending on the amount of intermediate supplied, the number of reactions leading from that intermediate to the final product and how fast those reactions take place. Final titers depend on these factors as well as the amount of starting material that is shunted into side reactions or used for energy by the cell. We selected pathways for 3HP, glucarate and muconate production from the literature. The acrylate biosynthesis pathway was developed in this study and represents the first direct biological rate to acrylate. Previously characterized muconate and glucarate biosensors were combined with their respective production pathways in order to monitor product formation. Because there are no existing biosensors for 3HP, novel sensors were developed and evaluated for their ability to sense 3HP production in real-time.

The first 3HP biosensor was developed from the *Escherichia coli* 2-methylcitrate responsive transcriptional regulator *prpR* (13). Since no 3HP-responsive allosteric transcriptional regulator is known, it was necessary to use a transcriptional regulator that

binds to a molecule that 3HP can be converted to intracellularly. The principle of relying on a downstream molecule to affect a response from a non-binding target compound was pioneered in the Keasling lab when *prpR* and two endogenous enzymes were used to construct a propionate biosensor (14). Here, we use the endogenous enzyme, 2-methylcitrate synthase (*prpC),* and the heterologous multifunctional enzyme propionyl-CoA synthase (*pcs*) from the carbon fixation pathway of *Chloroflexus aurantiacus* (15) to produce 2-methylcitrate from 3HP. Together the system of three genes (*pcs*, *prpC*, *prpR*) comprises the *prpR*-based 3HP biosensor (Fig. 3.2A). We evaluated the fluorescent response of the *prpR*-based biosensor to concentrations of 3HP up to 25mM with, and without, the presence of *pcs* (Fig. 3.2B). Indeed, *pcs* is necessary for a fluorescent response to 3HP. The response increases with increasing concentrations of 3HP, indicating that we have successfully linked the intracellular concentrations of 3HP and 2-methylcitrate. As a result of this linkage, 3HP concentration controls expression of green fluorescent protein (GFP). When induced with 12mM 3HP, the *prpR*-based 3HP biosensor produces a fluorescent response 2.4-fold greater than uninduced fluorescence levels. For comparison, the *prpR*-based propionate biosensor developed in the Keasling lab shows 4.5-fold induction under similar conditions (14). The fluorescent response of the *prpR*-based biosensor to 3HP induction is half-maximal at 5 hours and requires approximately 8 hours to reach 90% induction (Fig. 3.2C).

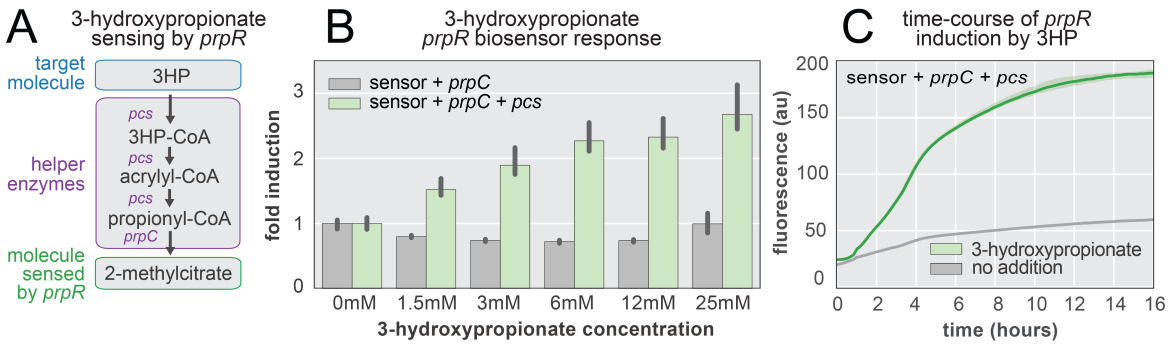


Figure 3.2: Development of the *prpR*-based 3-hydroxypropionate (3HP) biosensor. (A) Two helper enzymes, *pcs* from *Chloroflexus aurantiacus* and the endogenous *prpC*, convert 3HP into the *prpR*-binding compound 2-methylcitrate. (B) Exogenously supplied 3HP triggers a fluorescent response in cells containing the biosensor (green bars). Increasing the concentration of 3HP results in a higher fluorescent output. No biosensor activation is observed without the helper enzyme *pcs* (grey bars). (C) The fluorescent response of the biosensor begins after one hour and achieves 90% saturation after ten hours (green line). Basal induction increases over time but remains low (grey line). Error bars and confidence bands represent the 95% confidence interval (n=3).

The second 3HP biosensor was developed from *acuR*, an acrylate responsive transcriptional regulator found in the aquatic bacterium *Rhodobacter sphaeroides*(16). A novel pathway was constructed that converts 3HP to acrylate (Fig. 3.3A), allowing the acrylate biosensor to report intracellular 3HP concentration. In this case, a truncated version of the multifunctional enzyme *pcs* is used to convert 3HP into acrylyl-CoA, which is subsequently hydrolyzed to acrylate by the acrylyl-CoA hydrolase (*ach*) from *Acinetobacter baylyi* (17). In *Chloroflexus aurantiacus*, *pcs* catalyzes three subsequent reactions: 3HP to 3HP-CoA to acrylyl-CoA to propionyl-CoA (15). We made use of all three reactions in the *prpR*-based biosensor, but for the *acuR*-based biosensor accumulation of acrylyl-CoA rather than propionyl-CoA is necessary. Separation of *pcs* into its functional domains has been shown to increase the rates of the individual reactions (18). Because of this, we reasoned that we could remove the domain responsible for conversion of acrylyl-CoA to propionyl-

CoA while preserving the activity of the other two domains. We refer to the truncated enzyme as *pcs*$^{\Delta 3}$, and its co-expression with *ach* and *acuR* constitute the *acuR*-based 3HP biosensor (Fig. 3.3A). Increasing concentrations of 3HP in the media resulted in increasing levels of fluorescence when *pcs*$^{\Delta 3}$ and *ach* were present, but resulted in no biosensor activation in their absence (Fig. 3.3B). A 90-fold increase in fluorescence was obtained when the *acuR*-based biosensor was induced with 10mM 3HP. This is a much more dramatic activation than that achieved with the *prpR*-based biosensor. The induction kinetics of 3HP and the authentic activator acrylate were compared by monitoring biosensor activation in real-time. 3HP-mediated induction only slightly lagged the time-course of acrylate induction (Fig. 3.3C). Fluorescence remained at background levels for greater than 16 hours in the absence of *pcs*$^{\Delta 3}$ and *ach* (Fig. B.1).

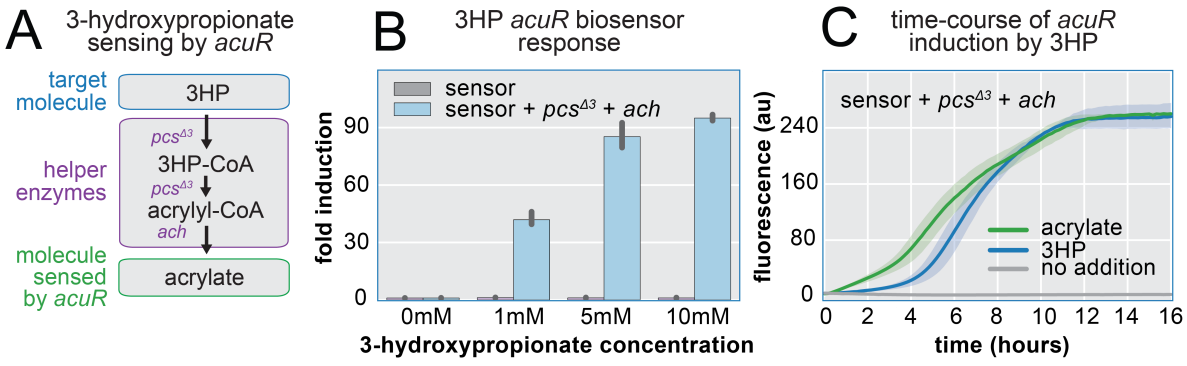


Figure 3.3: Development of the *acuR*-based 3-hydroxypropionate (3HP) biosensor. (A) Two heterologous helper enzymes, a truncated form of *pcs* (*pcs*$^{\Delta 3}$) and the acrylyl-CoA hydrolase *ach*, convert 3HP into the *acuR*-binding compound acrylate. (B) Exogenously supplied 3HP triggers a fluorescent response in cells containing the biosensor (blue bars). Higher concentrations of 3HP result in higher fluorescent outputs. No biosensor activation is observed without the helper enzymes *pcs*$^{\Delta 3}$ and *ach* (grey bars). (C) The fluorescent response of the biosensor to 3HP begins immediately and achieves 90% saturation after eight hours (blue line). Induction by acrylate is initially more rapid but achieves the same final fluorescence (green line). Basal induction is low over the duration of the experiment (grey line). Error bars and confidence bands represent the 95% confidence interval (n=3).

We co-expressed the 3HP biosensors with the 3HP production pathway in order to monitor 3HP production in real-time (Fig. 3.4A). The production pathway consists of the endogenous biosynthesis of malonyl-CoA and the bi-functional enzyme malonyl-CoA reductase (*mcr*) from the carbon fixation pathway of *Chloroflexus aurantiacus* (15). *Mcr* shunts malonyl-CoA away from fatty acid biosynthesis by catalyzing the conversion of malonyl-CoA, first into malonate semialdehyde, and then into 3HP at the expense of two NADPH+. This route from glucose to 3HP has been published previously, achieving titers of 60 mg/L with expression of *mcr* alone (19). Titers were increased to 180 mg/L with overexpression of the ACC complex and *pntAB*, increasing availability of malonyl-CoA and NADPH+, respectively. For our study, we chose to increase the amount of malonyl-CoA available for 3HP production by use of the fatty acid inhibitor cerulenin, rather than through genetic manipulations. Fatty acid biosynthesis is the primary sink for malonyl-CoA and operates at a much higher velocity than heterologously expressed *mcr* (19). Since cerulenin inhibits the activities of *fabB* and *fabF,* increasing its concentration results in lower fatty acid biosynthesis rates and a higher concentration of available malonyl-CoA (20). In each of the 3HP implementations, the biosensor helper enzymes, *pcs* and $pcs^{\Delta 3}$ / *ach*, were constitutively expressed while *mcr* was expressed conditionally with the addition of IPTG.

Co-expression of *mcr* and the *prpR*-based 3HP biosensor facilitates observation of 3HP production without the need for HPLC or mass spectrometry (Fig. 3.4B). Cells containing both the biosensor and *mcr* resulted in higher fluorescence over time than cells containing just the biosensor. When *mcr* activity was increased by induction with IPTG, the *mcr* containing cells showed a higher rate of GFP accumulation, ultimately achieving higher

levels of fluorescence. The cells without *mcr* were unaffected by IPTG induction. When the *prpR*-based biosensor was used for 3HP observation, it was necessary to produce 3HP using rich LB media as the carbon source. Sensitivity of *prpR* to catabolite repression precluded the use of glucose as the starting material for 3HP production. Even low levels of glucose result in the *prpR* transcriptional regulator becoming non-responsive to 2-methylcitrate (21). The uninduced expression of GFP from the *prpR* biosensor is significant, likely due to basal levels of 2-methylcitrate in the cell. Nonetheless, end point fluorescence measurements revealed that cells with the capacity to produce 3HP (*mcr*+) are 20% more fluorescent than cells without *mcr* (Fig. 3.4C). When induced, the *mcr*+ cells are 50% more fluorescent, approaching induction levels observed with 1.5mM exogenous 3HP.

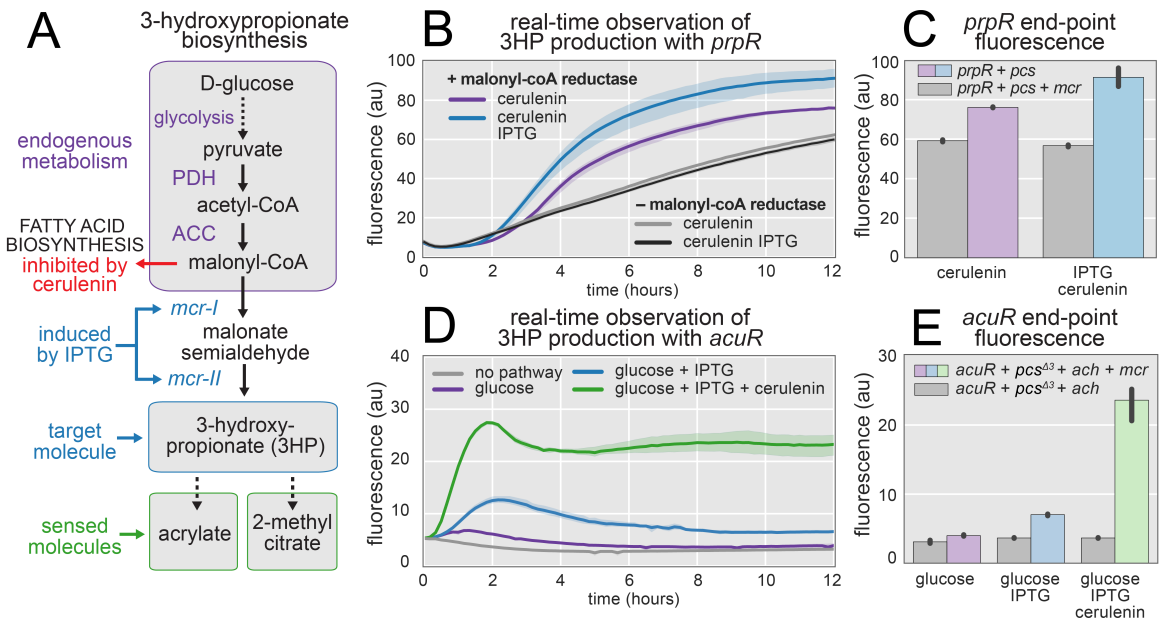


Figure 3.4: Formation of 3-hydroxypropionate is observed in real-time. (A) 3HP is produced from glucose by converting malonyl-CoA into malonate semialdehyde and then on to 3HP. Malonyl-CoA reductase (*mcr*) performs both of these reactions but competes with fatty acid biosynthesis for malonyl-CoA. (B) The *prpR*-based 3HP biosensor reports 3HP production progress in real-time. Addition of cerulenin increases the malonyl-CoA pool, providing a boost in 3HP production (purple line). Addition of IPTG increases *mcr* activity and further increases 3HP production (blue line). Cerulenin and IPTG have no

impact on the fluorescent response of cells without *mcr* (grey and black lines). (C) Fluorescence of the *prpR*-based biosensor after 12 hours. (D) The *acuR*-based 3HP biosensor reports the progress of 3HP production in real-time. The addition of 50mM glucose (purple line) results in a small increase in fluorescence over background (grey line). Addition of IPTG increases the production of *mcr* and the activation of the biosensor (blue line). Providing glucose, IPTG and cerulenin together results in the highest rate of biosensor activation (green line). (E) Fluorescence of the *acuR*-based biosensor after 12 hours reveals an approximately 5-fold increase in fluorescence under optimized culture conditions (glucose, IPTG, cerulenin) when compared to 3HP production with glucose alone. In the absence of *mcr,* culture conditions have no effect on biosensor activation (grey bars). Fluorescence measurements are in arbitrary units and different panels should not be quantitatively compared. Error bars and confidence bands represent the 95% confidence interval (n=3).

The *acuR*-based 3HP biosensor enables real-time observation of 3HP production under culture conditions more representative of what would be experienced at scale (Fig. 3.4D). The *acuR* biosensor is not affected by catabolite repression and can be used to observe 3HP production from glucose. *Mcr* was co-expressed with the *acuR*-based biosensor and fluorescence was observed for twelve hours. Cells that were incubated with 50mM glucose, but without IPTG or cerulenin, produced fluorescence indistinguishable from background levels. Cells incubated with glucose and IPTG showed a significant increase in fluorescence. The most dramatic increase in fluorescence was observed when both IPTG and cerulenin were used. Production of 3HP with glucose, IPTG and cerulenin resulted in higher rates of GFP expression and end-point fluorescence values than either of the other two culture conditions. End-point measurements reveal an eight-fold increase in fluorescence for *mcr*+ cells versus *mcr*- cells when incubated with glucose, IPTG and cerulenin (Fig. 3.4E). Incubation with glucose and IPTG results in a two-fold increase. Incubation with glucose alone results in only a 20% increase in fluorescence compared to cells lacking *mcr*.

The culture conditions optimized by use of the 3HP biosensors resulted in a 3HP titer of 4.2 g/L (Fig. 3.5A). To our knowledge, this is the highest titer reported in the literature for 3HP production from glucose in *E. coli*. These titers were obtained with rich media in 96-well plates supplemented once with 50mM glucose at the beginning of fermentation. Production was carried out in BL21 transformed with the *mcr* plasmid under the same culture conditions used for the biosensor experiments. Glucose alone produced no detectable 3HP, while the addition of 1mM IPTG resulted in 1.5 ± 0.2 mg/L. The addition of cerulenin resulted in 7.1 ± 2.5 mg/L while addition of both cerulenin and IPTG to the media resulted in 4.2 ± 1.2 g/L. The trend in 3HP titer among the various culture conditions is reflected in the fluorescent output of the two 3HP biosensors. We were able to produce 23-fold more 3HP than previously reported for the malonyl-CoA route to 3HP. More importantly, the titers achieved here are high enough to warrant further commercial exploration. A major caveat is that these levels of 3HP production rely on the addition of cerulenin, which is not feasible for scale-up due to cost. Our group has previously shown that multiplexed genome engineering combined with static biosensors can obviate the need for cerulenin by finding mutants with increased malonyl-CoA availability (1). Future work will aim to find a similar mutant for 3HP production using the fluorescent biosensors developed here.

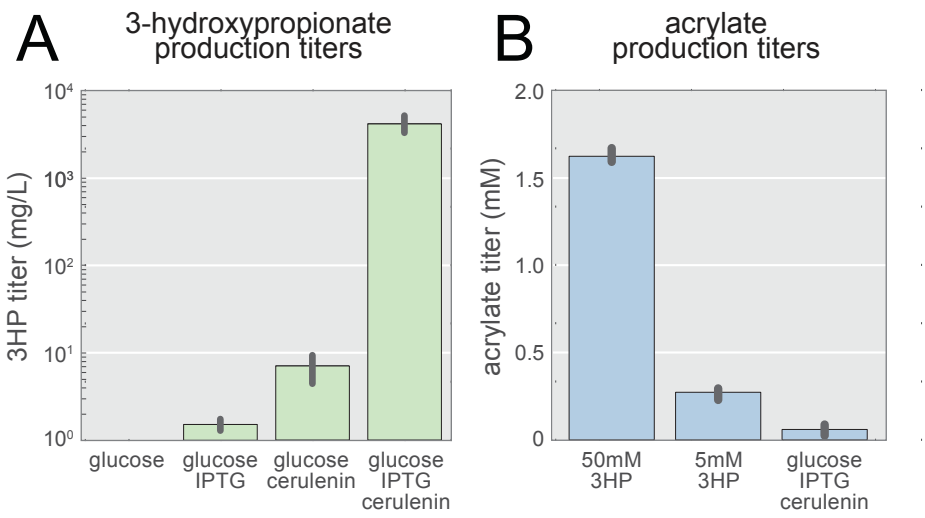


Figure 3.5: Production titers for 3-hydroxypropionate (3HP) and acrylate. (A) The culture conditions evaluated for biosensor activation were also evaluated for 3HP production. Titers were measured by LC/MS and found to correspond to biosensor activation. A record 4.2 g/L 3HP was achieved with glucose, IPTG and cerulenin. (B) Acrylate production from 3HP and glucose was determined by LC/MS. Error bars represent the 95% confidence interval derived from the standard error of the mean (n=3).

Co-expression of *pcs*$^{\Delta 3}$ and *ach* enables *in vivo* production of acrylate (Fig. 3.5B, Fig. B.2). While biologically-derived 3HP is used to produce several materials, it is most importantly a precursor to bio-based acrylate. Currently, 3HP is isolated from the cell culture and chemically converted to acrylate. *In vivo* production of acrylate obviates this step by allowing its direct fermentation. We produced 1.62 ± 0.05 mM and 0.27 ± 0.04 mM acrylate from 50mM and 5mM 3HP, respectively. Addition of the *mcr* plasmid resulted in 60 ± 37 μM acrylate produced from glucose. These conversion efficiencies are low for commercial acrylate production, but perfect for tracking 3HP production.

The *acuR*-based 3HP biosensor will be a better choice than the *prpR*-based biosensor for most 3HP production enhancement applications. The lack of catabolite repression and dramatically higher dynamic range are key considerations. Substantial

work has been done to remove the catabolite repression from *prpR,* but with little success (21). The fact that *acuR* senses acrylate, which is not naturally present in cells, while *prpR* senses 2-methlycitrate, is another important consideration. The background levels of 2-methlycitrate in the cells may account for the elevated basal fluorescence and lower dynamic range of the *prpR* biosensor.

We apply the real-time sensing paradigm to two additional metabolic pathways to demonstrate that real-time observation of product formation is generalizable to many biosensor and pathway combinations. The first additional pathway produces glucarate from glucose through the action of three heterologous enzymes. Glucarate is a DoE 'top value added chemical' to produce from biomass, with applications as a renewable replacement for petrochemicals and as a building block for new ultra-hydrophilic polymers (11). Several papers describe construction and optimization of the glucarate biosynthesis pathway (22-24). However, attaining high titers remains challenging due to the low stability and activity of myo-inositol oxygenase, making this pathway a prime target for optimization by directed evolution. The ability to monitor glucarate production in individual cells will enable studies combining flow-cytometry and directed mutagenesis.

We combined the previously characterized glucarate biosensor with the glucarate biosynthesis pathway comprised of IPTG inducible myo-inositol-1-phosphate synthase (Ino1, *Saccharomyces cerevisiae*), myo-inositol oxygenase (MIOX, *Mus musculus*) and uronate dehydrogenase (Udh, *Agrobacterium tumefaciens*) (Fig. 3.6A). We anticipated that the glucarate biosensor would produce a fluorescent response proportional to the amount of glucarate produced within the cell. To evaluate whether or not this was the case, we maintained identical production conditions (e.g., genetics, media composition) but varied

the exogenously supplemented precursor molecules with the hypothesis that compounds further along in the biosynthesis pathway (i.e., separated from glucarate by fewer reactions) would result in a faster rate of glucarate formation. Addition of glucarate itself resulted in the fastest rate of GFP production and ultimately the highest amount fluorescence (Fig. 3.6B). As expected, we observed a similar fluorescent response to glucarate in biosensor strains with, and without, the biosynthesis pathway. In contrast, none of the other exogenously supplied molecules resulted in a fluorescent response in the biosensor strain lacking the glucarate biosynthesis pathway (Fig. 3.6C). This indicates that glucarate and not the precursor molecules trigger the fluorescent response. In the strain containing both the biosensor and the biosynthesis pathway, addition of glucuronate to the media resulted in a fluorescent response lagging glucarate by about 90 minutes, ultimately achieving an end-point fluorescence that was 80% that of glucarate. This is in contrast to the addition of myo-inositol, which resulted in a fluorescence response lagging that of glucuronate by 60 minutes. Notably, the end-point fluorescence achieved by the addition of myo-inositol is just 20% that of glucuronate addition. Media supplemented with 50mM glucose resulted in no fluorescent response within the duration of the experiment (Fig. 3.5B).

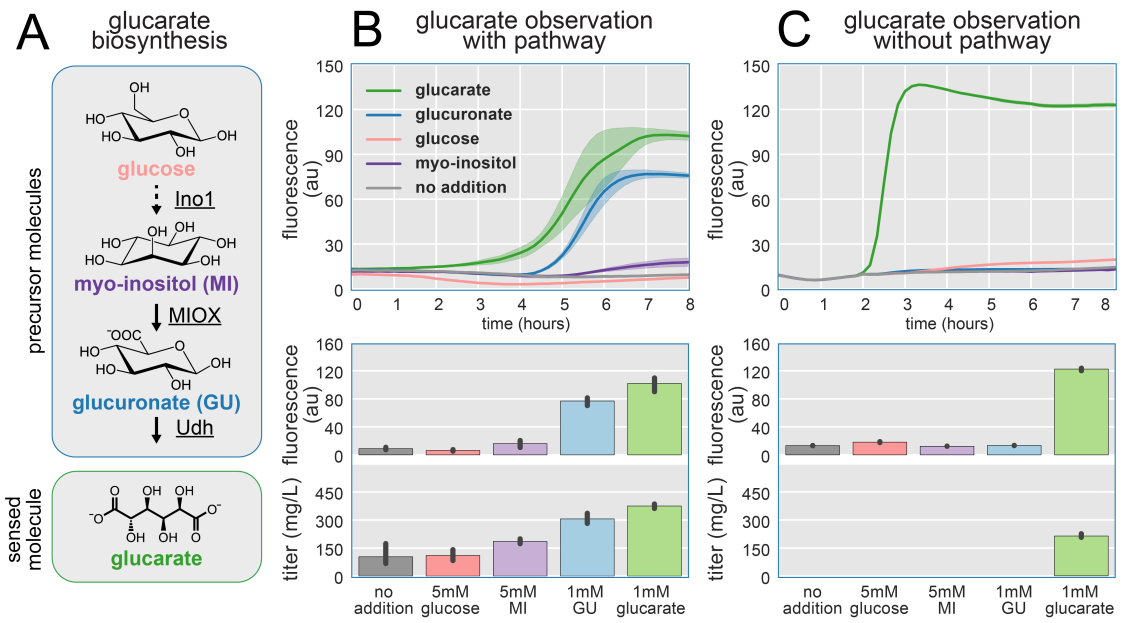


Figure 3.6: Real-time observation of glucarate production. (A) Glucarate can be produced from glucose with the expression of three heterologous enzymes of various activities. Udh has high activity, MIOX has low activity and Ino1 competes with glycolysis for glucose-6-phosphate. The presence of glucarate activates the biosensor. (B) Intermediates of glucarate biosynthesis are added to the media. Fluorescence is observed over time as the intermediates are converted to glucarate. Biosensor activation by glucuronate (blue line) lags behind activation by glucarate (green line). Both glucarate and glucuronate activation are faster than activation by myo-inositol (purple line) or glucose (tan line), reflecting the dynamics of the biosynthesis pathway. End-point fluorescence trends well with LC/MS determined glucarate titers. (C) In the absence of the glucarate biosynthesis pathway there is no biosensor activation or glucarate production (as determined by LC/MS) from any of the pathway intermediates. Addition of glucarate resulted in biosensor activation (green line). Error bars and confidence bands represent the 95% confidence interval (n=3).

The fluorescent output of the glucarate biosensor reflects the properties of the glucarate biosynthesis pathway. The conversion of glucuronate to glucarate is known to be the fastest heterologous reaction in the biosynthesis of glucarate (23). Correspondingly, we see robust biosensor activation when glucuronate is the starting material. The fluorescent response to myo-inositol addition is slow and corroborates the difficulty of using *M. musculus* MIOX in the catalysis of myo-inositol to glucuronate in *E. coli* (24). The lack of

biosensor response to additional glucose may reflect the fact that glucarate biosynthesis is competing with glycolysis for glucose-6-phosphate. It may take substantial Ino1 activity to create meaningful quantities of myo-inositol. Low myo-inositol production would be further compounded by weak MIOX activity, ultimately yielding low glucarate titers and biosensor mediated fluorescence. Tuning the endogenous metabolism to balance glycolysis with glucarate production, while screening for fluorescence resulting from glucose supplementation, would be a powerful strategy for finding strains that produce high glucarate titers. A similar approach can be taken in searching for a more effective variant of MIOX from a library of targeted or untargeted mutations.

In addition to revealing the relative rates of product formation, the fluorescent response is a good proxy for product titer. The fluorescence observed eight hours after addition of the precursor molecules was measured and compared to the glucarate titers achieved under similar conditions. Glucarate production was observed for every condition tested when the cells contained the biosynthesis pathway. No production was observed without the pathway. Glucarate formation was observed in rich LB media even without additional substrate added. Addition of 5mM glucose did not result in a significant increase in glucarate titer within eight hours, consistent with the observed lack of a fluorescent response. However, addition of myo-inositol resulted in an elevated glucarate titer as reflected in the fluorescent response. Addition of glucuronate resulted in glucarate production at a 97% yield (background production of glucarate from LB was subtracted). This yield reflects the high fluorescence achieved by glucarate addition. Plotting titer as a function of fluorescence reaffirms that fluorescence is a good predictor of titer across the four culture conditions evaluated here ($R^2 = 0.96$, Fig. B.3).

For the last example of real-time observation of metabolite production we use a muconate biosensor to monitor product formation in a muconate biosynthesis pathway. Muconate is a diacid that can be hydrogenated to adipate, a major platform chemical with global sales of two million tons per year (25). Production of muconate from renewable sources such as glucose enables renewable production of adipate products such as nylon and polyurethane. Real-time observation of muconate production is achieved by linking intracellular muconate concentration to GFP expression through a muconate biosensor constructed from *benM*, a LysR-type transcriptional regulator derived from *Acinetobacter baylyi* (26). In order to achieve various muconate production rates for real-time observation, we produced muconate from a range of precursor molecules by implementing the biosynthetic pathway developed by Draths and Frost (12,27). This pathway uses three heterologous enzymes to convert 3-dehydroshikimate (DHS), an intermediate of aromatic amino acid biosynthesis, to cis,cis-muconate (Fig. 3.7A). The branch-point from endogenous metabolism is catalyzed by DHS dehydratase (*Acinetobacter baylyi*, *quiC*) resulting in protocatechuate, which in turn is decarboxylated to catechol by protocatechuate decarboxylase (*Klebsiella pneumonia*, *aroY*) before oxygenation to muconate by catechol 1,2-dioxygenase (*Acinetobacter baylyi*, *catA*). Combining the non-optimized muconate biosynthesis pathway with the newly constructed muconate fluorescent biosensor provides an ideal test case for real-time observation of muconate production.

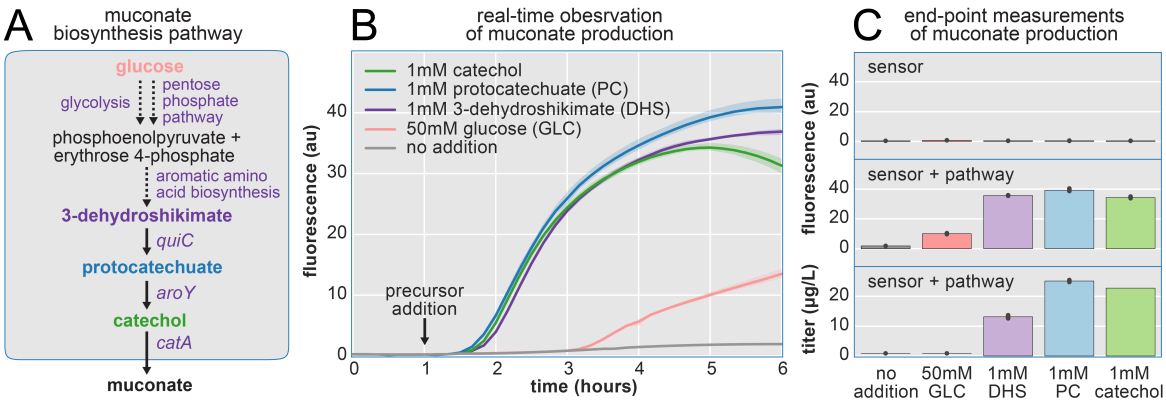


Figure 3.7: Real-time observation of muconate production. (A) Muconate is produced from glucose by expression of three heterologous enzymes. The committed step of muconate biosynthesis is the conversion of 3-dehydroshikimate (DHS) to protocatechuate (PC). (B) Muconate pathway intermediates were added after one hour and fluorescence was monitored over time. Each late pathway intermediate activated the biosensor within an hour. Conversion of glucose to muconate was much slower (pink line). (C) End-point measurements of fluorescence reveal that the muconate biosynthesis pathway is necessary for biosensor activation by pathway intermediates (top panel). The presence of the pathway enables precursors to trigger the biosensor. Consistent with its place far upstream in the metabolic pathway, glucose is the only substrate that achieves lower fluorescence (pink bar – middle panel). Background fluorescence is observed without glucose (grey bar – middle panel). Muconate supernatant titers were determined by HPLC. Less muconate was produced by DHS (purple bar – bottom panel) than the subsequent intermediates (blue, green bars – bottom panel). Muconate production from glucose (pink bar – bottom panel) was below the limit of quantification by HPLC. Error bars and confidence bands represent the 95% confidence interval (n=3).

A rapid fluorescence response was observed for the pathway intermediates DHS, protocatechuate and catechol, while a slow response was observed when muconate production was started from glucose (Fig. 3.7B). No response to any of the intermediates was observed when the muconate biosynthesis pathway was absent. The observation that DHS produces a response as fast as the other pathway intermediates suggests that competition with aromatic amino acid biosynthesis is not limiting the rate of muconate production under these conditions. In contrast, the slow rate of GFP expression resulting from the addition of glucose provides evidence that achieving a sufficient supply of

endogenous DHS may be limiting in the genetic background used here. This is consistent with the negative feedback that exists at both the transcriptional and allosteric levels in aromatic amino acid biosynthesis (12). The negative feedback is designed to throttle the production of DHAP (the precursor to DHS and consequently muconate) in the presence of aromatic amino acids. Strains optimized for muconate biosynthesis overexpress a feedback resistant mutant of *aroF* that is defective in product inhibition. Other genetic modifications aimed at increasing the DHS pool have included knockout of *aroE* and overexpression of *aroB* and *tktA* (12).

Examining end-point fluorescence reveals that the final fluorescence achieved is consistent with the muconate titers measured in the supernatant at that time-point (Fig. 3.7C). While the end-point fluorescence of the late intermediates is similar, it is dramatically lower for glucose. This is reflected in the titers, as muconate production from glucose was lower than our limit of detection at this early time-point. In this case, the biosensor is more sensitive than the traditional methods of quantification. This may be attributed to intracellular versus extracellular sensing; unless muconate is actively transported out of the cell, intracellular concentrations would be expected to rise before supernatant concentrations. A biosensor for protocatechuate was recently developed and used to measure end-point fluorescence resulting from the production of protocatechuate from DHS(28). Using this intermediate biosensor with our muconate biosensor to control GFP and RFP would enable screens aimed at minimizing the concentration of potentially toxic pathway intermediates while maximizing end-product formation.

Through this work we have developed a framework for tracking the formation of metabolic products in real-time using fluorescent biosensors. We demonstrated that the

fluorescence achieved by a cell is indicative of its productivity: higher fluorescence is indicative of higher product titers. Fluorescence as a proxy for product titer not only allows for real-time observation of product formation, but also enables cells to report their own individual progress in producing a chemical. This allows millions of cells to be evaluated in minutes by combining biosensor technology with fluorescent activated cell sorting (FACS). FACS has increasingly been used to characterize libraries of mutant cells (28,29). The throughput of phenotype evaluation in metabolic engineering can be increased by more than three orders of magnitude by relying on FACS as a primary screen and only using HPLC or mass spectrometry to evaluate the most promising individuals (3). By reliably linking product formation to fluorescence, we enable the last low-throughput step of the metabolic engineering design-build-test cycle to be multiplexed, alleviating the primary bottleneck in the development of biologically-based products.

Materials and Methods

Chemicals and Reagents

All reagents were obtained from Sigma unless otherwise noted. Antibiotics and IPTG were obtained from Gold Biotechnology. PCR mix was purchased from Kapa Biosystems. 3-hydroxypropionate was purchased from Toronto Research Chemicals. Cerulenin was purchased from Cayman Chemical and dissolved in ethanol. Acrylic acid was stored at room temperature with 200ppm MEHQ as an inhibitor and diluted immediately prior to use. All cell culture additives were dissolved in deionized water to achieve appropriate working concentrations.

Strains and Plasmids

Plasmids were constructed using Gibson isothermal assembly methods (30) and cloned into DH5α electrocompetent cells purchased from New England Biolabs. Biosynthesis of product molecules was carried out in either BL21 (DE3) or DH5α, as noted in the text. The *prpR*-based 3HP biosensor was implemented as a two-plasmid system. The first plasmid is pPro24-GFP (Addgene plasmid #18880), which expresses GFPuv under the control of the methyl-citrate responsive transcription factor, *prpR*, on a pBR322 origin of replication providing β-lactam resistance. The second plasmid (pJKR-PCS) was constructed such that the enzyme propionyl-CoA synthase was under the control of the constitutive promoter prod (31) on a ColA origin of replication providing kanamycin resistance. The *acuR*-based 3HP biosensor is composed of two plasmids. The first is the previously characterized high-copy acrylate biosensor pJKR-H-*acuR* (Addgene plasmid #62567), which expresses sfGFP under the control of the acrylate responsive transcription factor, *acuR*, on a pUC origin of replication providing β-lactam resistance. The second is derived from pJKR-PCS such that PCS is truncated between amino acids 1400 and 1401. The enzyme acrylyl-CoA hydrolase from *Acinetobacter baylyi* was subsequently cloned into the plasmid under the control of the P2 (32) constitutive promoter. The resulting plasmid is designated pJKR-PCSfrag-ACH. The 3HP biosynthesis plasmid, designated pJKR-MCR, was constructed such that malonyl-CoA reductase from *Chloroflexus aurantiacus* was expressed by the pLlacO (9) promoter under the control of LacI on a p15a origin of replication with spectinomycin resistance. The glucarate biosensor is the previously characterized plasmid pJKR-H-*cdaR* (Addgene plasmid #62557), which expresses sfGFP under the control of the

glucarate responsive transcription factor, *cdaR*, on a pUC origin of replication providing β-lactam resistance. The glucarate biosynthesis pathway was implemented on a single plasmid, pJKR-GA-EXP, which expresses the genes MIOX, Ino1 and Udh co-cistronically from an IPTG-regulated T7 promoter on a p15a origin of replication providing kanamycin resistance. MIOX from *Mus musculus* and Ino1 from *Saccharomyces cerevisiae* were synthesized with codons optimized for *E. coli* expression. Udh was obtained from *Agrobacterium tumefaciens* genomic DNA (ATCC #33970D-5). The muconate biosensor was constructed with sfGFP under the control of the muconate-responsive transcription factor *benM* (*Acinetobacter baylyi*) on a pUC origin of replication providing spectinomycin resistance. *BenM* was constitutively expressed with the proB (31) promoter. To provide an option for expression normalization, the RFP mKate2 was constitutively expressed from the P11 (32) promoter. The resulting plasmid is designated pJKR-H-*benM*. The muconate biosynthesis pathway was constructed as a single plasmid with codon optimized variants of the biosynthesis genes expressed co-cistronically from an IPTG inducible T7 promoter on a p15a origin of replication providing β-lactam resistance.

3-Hydroxypropionate Biosensor Characterization

DH5α cells doubly transformed with plasmids pPro24-GFP and pJKR-PCS, or plasmids pJKR-H-*acuR* and pJKR-PCSfrag-ACH, were exposed to increasing concentrations of 3HP and monitored for GFP expression. Cells were grown overnight to saturation before being diluted 1:100 into fresh LB media and incubated at 200RPM and 37°C. After four hours, 150µl of the log-phase cells were transferred to 96-well plates and 3HP was added to the appropriate final concentration. Each inoculation and induction was performed in

triplicate. Strains lacking the biosensor helper plasmids were included to reveal that the biosensor activation was indeed dependent on the presence of the helper plasmids. In the case of the end-point measurements, fluorescence was measured 16 hours after 3HP addition with a Biotek HT plate reader (excitation 485/20, emission 528/20). Time course data was collected over a 16 hour period on the same plate reader at 37°C with fast shaking and 10 minute measurement intervals. Fluorescence was normalized by optical density. Fold induction was determined by dividing the fluorescence obtained at the current induction level by the fluorescence obtained without induction. Error bars represent the 95% confidence interval derived from the standard error of the mean.

3-Hydroxypropionate / Acrylate Production and Monitoring

BL21 cells containing the plasmids for the *prpR*- and *acuR*-based 3HP biosensors were transformed with the plasmid pJKR-MCR. These production strains were grown up overnight and back-diluted 1:100 into fresh LB media and incubated at 200RPM and 37°C. After four hours, 150µl of the log-phase cells were transferred to 96-well plates and exposed to 3HP production conditions. The *prpR*-based biosensor production strain was incubated with and without 1mM IPTG and 20µg/ml cerulenin in LB. The *acuR*-based biosensor production strain was incubated with 50mM glucose and different combinations of 1mM IPTG and 20µg/ml cerulenin. Growth-normalized fluorescence was observed in the Biotek HT plate reader as described above. End-point measurements were taken after 12 hours. 3HP production titers were determined by liquid-chromatography and mass spectrometry (LC/MS). The strain used for 3HP titer measurements only contained the

production plasmid pJKR-MCR. Overnight cultures were inoculated 1:100 into 1mL of LB supplemented with 1mM IPTG, 20μg/ml cerulenin and 50mM glucose in 96-well blocks. Production took place at 900RPM and 37C for 16 hours before supernatants were isolated and filtered at 0.2μm for LC/MS. Samples were prepared for acrylate analysis in an identical manner with DH5α as the strain background. Production of acrylate from glucose was carried out with 50mM glucose, 1mM IPTG and 20μg/ml cerulenin. Acrylate was measured on a Thermo q-Exactive mass spectrometer equipped with a Thermo 3000 Ultimate μHPLC. All production runs were setup in triplicate. Error bars represent the 95% confidence interval derived from the standard error of the mean.

Glucarate Production and Monitoring

Glucarate production monitoring was carried out in BL21 cells doubly transformed with pJKR-H-*cdaR* and pJKR-GA-EXP. BL21 transformed with pJKR-H-*cdaR* alone was used as a control. Overnight cultures back-diluted 1:100 into Davis media supplemented with 5g/L glucose. After four hours incubating at 200RPM and 37°C, 150μl of the log-phase cells were transferred to 96-well plates and exposed to 1mM IPTG with the pathway intermediates noted in the text. Normalized fluorescence was observed for eight hours after addition of the pathway intermediates as described above. End-point measurements were taken after eight hours. Glucarate production for titer measurement was carried out in BL21 transformed with pJKR-GA-EXP. An overnight culture was inoculated 1:100 into 1mL of LB supplemented with 1mM IPTG and the specified concentration of pathway intermediate. Production took place in 96-well blocks at 900RPM and 37°C for eight hours before supernatants were isolated and 0.2μm filtered for LC/MS. All production runs were

setup in triplicate. Error bars represent the 95% confidence interval derived from the standard error of the mean.

Muconate Production and Monitoring

Muconate production monitoring was carried out in BL21 cells doubly transformed with pJKR-H-*benM* and the muconate production plasmid pMuc1. Overnight cultures were grown overnight before being diluted 1:100 into LB and incubated at 200RPM and 37°C. After four hours, 150µl of the log-phase cells were transferred to 96-well plates and monitored in the plate reader. After one hour, the specified concentration of pathway intermediate was added in triplicate and fluorescence monitoring was resumed. End-point measurements were made five hours after the addition of intermediates. Strains transformed with the muconate production plasmid alone were used to determine product titers. An overnight culture was inoculated 1:100 into 1mL of LB supplemented with the specified concentration of pathway intermediate. Production took place in 96-well blocks at 900RPM and 37°C for five hours. The quantity of muconate in the supernatant was determined by absorbance at 210nm on an Agilent HPLC equipped with an Aminex HPX-87H column (Bio-Rad) operated in isocratic mode with 0.1% phosphoric acid. All production runs were setup in triplicate. Error bars represent the 95% confidence interval derived from the standard error of the mean.

References

1. Raman, S., Rogers, J.K., Taylor, N.D. and Church, G.M. (2014) Evolution-guided optimization of biosynthetic pathways. *Proceedings of the National Academy of Sciences of the United States of America*, **111**, 17803-17808.

2. Wang, H.H., Isaacs, F.J., Carr, P.A., Sun, Z.Z., Xu, G., Forest, C.R. and Church, G.M. (2009) Programming cells by multiplex genome engineering and accelerated evolution. *Nature*, **460**, 894-U133.

3. Dietrich, J.A., McKee, A.E. and Keasling, J.D. (2010) High-Throughput Metabolic Engineering: Advances in Small-Molecule Screening and Selection. *Annu Rev Biochem*, **79**, 563-590.

4. Yang, J., Seo, S.W., Jang, S., Shin, S.I., Lim, C.H., Roh, T.Y. and Jung, G.Y. (2013) Synthetic RNA devices to expedite the evolution of metabolite-producing microbes. *Nature communications*, **4**, 1413.

5. Binder, S., Schendzielorz, G., Stabler, N., Krumbach, K., Hoffmann, K., Bott, M. and Eggeling, L. (2012) A high-throughput approach to identify genomic variants of bacterial metabolite producers at the single-cell level. *Genome biology*, **13**, R40.

6. Dietrich, J.A., Shis, D.L., Alikhani, A. and Keasling, J.D. (2013) Transcription factor-based screens and synthetic selections for microbial small-molecule biosynthesis. *ACS synthetic biology*, **2**, 47-58.

7. Tang, S.Y., Qian, S., Akinterinwa, O., Frei, C.S., Gredell, J.A. and Cirino, P.C. (2013) Screening for enhanced triacetic acid lactone production by recombinant Escherichia coli expressing a designed triacetic acid lactone reporter. *Journal of the American Chemical Society*, **135**, 10099-10103.

8. Tracy, B.P., Gaida, S.M. and Papoutsakis, E.T. (2010) Flow cytometry for bacteria: enabling metabolic engineering, synthetic biology and the elucidation of complex phenotypes. *Current opinion in biotechnology*, **21**, 85-99.

9. 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.

10. Kumar, V., Ashok, S. and Park, S. (2013) Recent advances in biological production of 3-hydroxypropionic acid. *Biotechnology advances*, **31**, 945-961.

11. Werpy, T.a.P., G. (2004) In Energy, U. S. D. o. (ed.).

12. Niu, W., Draths, K.M. and Frost, J.W. (2002) Benzene-free synthesis of adipic acid. *Biotechnology progress*, **18**, 201-211.

13. Palacios, S. and Escalante-Semerena, J.C. (2004) 2-Methylcitrate-dependent activation of the propionate catabolic operon (prpBCDE) of Salmonella enterica by the PrpR protein. *Microbiology*, **150**, 3877-3887.

14. Lee, S.K. and Keasling, J.D. (2005) A propionate-inducible expression system for enteric bacteria. *Applied and environmental microbiology*, **71**, 6856-6862.

15. Mattozzi, M., Ziesack, M., Voges, M.J., Silver, P.A. and Way, J.C. (2013) Expression of the sub-pathways of the Chloroflexus aurantiacus 3-hydroxypropionate carbon fixation bicycle in E. coli: Toward horizontal transfer of autotrophic growth. *Metab Eng*, **16**, 130-139.

16. Sullivan, M.J., Curson, A.R., Shearer, N., Todd, J.D., Green, R.T. and Johnston, A.W. (2011) Unusual regulation of a leaderless operon involved in the catabolism of dimethylsulfoniopropionate in Rhodobacter sphaeroides. *PloS one*, **6**, e15972.

17. Valle F., A., N.J., Noriega, C. (2013).

18. Alber, B.E. and Fuchs, G. (2002) Propionyl-coenzyme A synthase from Chloroflexus aurantiacus, a key enzyme of the 3-hydroxypropionate cycle for autotrophic CO2 fixation. *The Journal of biological chemistry*, **277**, 12137-12143.

19. Rathnasingh, C., Raj, S.M., Lee, Y., Catherine, C., Ashok, S. and Park, S. (2012) Production of 3-hydroxypropionic acid via malonyl-CoA pathway using recombinant Escherichia coli strains. *Journal of biotechnology*, **157**, 633-640.

20. Heath, R.J. and Rock, C.O. (1995) Regulation of malonyl-CoA metabolism by acyl-acyl carrier protein and beta-ketoacyl-acyl carrier protein synthases in Escherichia coli. *The Journal of biological chemistry*, **270**, 15531-15538.

21. Lee, S.K., Newman, J.D. and Keasling, J.D. (2005) Catabolite repression of the propionate catabolic genes in Escherichia coli and Salmonella enterica: evidence for involvement of the cyclic AMP receptor protein. *J Bacteriol*, **187**, 2793-2800.

22. Moon, T.S., Dueber, J.E., Shiue, E. and Prather, K.L. (2010) Use of modular, synthetic scaffolds for improved production of glucaric acid in engineered E. coli. *Metab Eng*, **12**, 298-305.

23. Moon, T.S., Yoon, S.H., Lanza, A.M., Roy-Mayhew, J.D. and Prather, K.L. (2009) Production of glucaric acid from a synthetic pathway in recombinant Escherichia coli. *Applied and environmental microbiology*, **75**, 589-595.

24. Shiue, E. and Prather, K.L. (2014) Improving D-glucaric acid production from myo-inositol in E. coli by increasing MIOX stability and myo-inositol transport. *Metab Eng*, **22**, 22-31.

25. Weber, C., Bruckner, C., Weinreb, S., Lehr, C., Essl, C. and Boles, E. (2012) Biosynthesis of cis,cis-muconic acid and its aromatic precursors, catechol and protocatechuic acid, from renewable feedstocks by Saccharomyces cerevisiae. *Applied and environmental microbiology*, **78**, 8421-8430.

26. Craven, S.H., Ezezika, O.C., Haddad, S., Hall, R.A., Momany, C. and Neidle, E.L. (2009) Inducer responses of BenM, a LysR-type transcriptional regulator from Acinetobacter baylyi ADP1. *Mol Microbiol*, **72**, 881-894.

27. Draths, K.M. and Frost, J.W. (1994) Environmentally Compatible Synthesis of Adipic Acid from D-Glucose. *Journal of the American Chemical Society*, **116**, 399-400.

28. Jha, R.K., Kern, T.L., Fox, D.T. and CE, M.S. (2014) Engineering an Acinetobacter regulon for biosensing and high-throughput enzyme screening in E. coli via flow cytometry. *Nucleic acids research*, **42**, 8150-8160.

29. Kosuri, S., Goodman, D.B., Cambray, G., Mutalik, V.K., Gao, Y., Arkin, A.P., Endy, D. and Church, G.M. (2013) Composability of regulatory sequences controlling transcription and translation in Escherichia coli. *Proceedings of the National Academy of Sciences of the United States of America*, **110**, 14024-14029.

30. Gibson, D.G., Young, L., Chuang, R.Y., Venter, J.C., Hutchison, C.A., 3rd and Smith, H.O. (2009) Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat Methods*, **6**, 343-345.

31. Davis, J.H., Rubin, A.J. and Sauer, R.T. (2011) Design, construction and characterization of a set of insulated bacterial promoters. *Nucleic acids research*, **39**, 1131-1141.

32. Mutalik, V.K., Guimaraes, J.C., Cambray, G., Lam, C., Christoffersen, M.J., Mai, Q.A., Tran, A.B., Paull, M., Keasling, J.D., Arkin, A.P. *et al.* (2013) Precise and reliable gene expression via standard transcription and translation initiation elements. *Nat Methods*, **10**, 354-360.

Chapter 4: Evolution-guided optimization of biosynthetic pathways

Abstract

Engineering biosynthetic pathways for chemical production requires extensive optimization of the host cellular metabolic machinery. Because it is challenging to specify *a priori* an optimal design, metabolic engineers often need to construct and evaluate a large number of variants of the pathway. We report a general strategy that combines targeted genome-wide mutagenesis to generate pathway variants with evolution to enrich for rare high producers. We convert the intracellular presence of the target chemical into a fitness advantage for the cell by using a sensor domain responsive to the chemical to control a reporter gene necessary for survival under selective conditions. Because artificial selection tends to amplify unproductive cheaters, we devised a negative selection scheme to eliminate cheaters while preserving library diversity. This scheme allows us to perform multiple rounds of evolution (addressing ~10^9 cells/round) with minimal carryover of cheaters after each round. Based on candidate genes identified by flux balance analysis, we used targeted genome-wide mutagenesis to vary the expression of pathway genes involved in the production of naringenin and glucaric acid. Through up to four rounds of evolution, we increased production of naringenin and glucaric acid by 36- and 22-fold, respectively. Naringenin production (61 mg/L) from glucose was more than double the previous highest titer reported. Whole genome sequencing of evolved strains revealed additional untargeted mutations that likely benefit production, suggesting new routes for optimization.

Introduction

Microbial production of chemicals presents an alternative to ubiquitous chemical synthesis methods. Biosynthetic production is attractive because it can utilize a broad assortment of organic feedstocks, proceed under benign physiological conditions, and reduce environmentally deleterious byproducts. Biosynthetic alternatives are being pursued for wide range of chemicals, from bulk commodity building blocks to specialty chemicals.

Natural cells are seldom optimized to produce a desired molecule. To achieve economically viable production, extensive modifications to host cell metabolism are often required to improve metabolite titer, production rate and yield. The optimizations of biosynthetic pathways for 1,3-propanediol(1), flavonoids(2, 3), L-tyrosine(4), and 1,4-butanediol(5) illustrate this complexity. Fortunately, computational models of cellular metabolism, such as flux-balance analysis (FBA), aid in predicting metabolic changes likely to improve the production of a target molecule. Powerful methods including oligonucleotide-directed genome engineering(6) (MAGE) and Cas9-mediated editing can specifically mutate the genomic targets predicted by FBA. But the combinatorial space of these genomic mutations quickly outstrips the throughput of current analytical methods for evaluating chemical production in individual clones ($<10^3$ samples machine^{-1} day^{-1}).

Biosensors that report on the concentration of a chemical within each individual cell can alleviate this screening bottleneck. Such sensor-reporters transduce the binding of a target small molecule by a sensory protein or RNA into a gene expression readout(7). The resulting expression of a fluorescent reporter gene or antibiotic resistance gene allows facile identification of mutant cells with increased production of the target chemical.

Sensor-reporters have been employed to screen for increased microbial production of several chemicals, including the isoprenoid precursor mevalonate(8), L-lysine (9, 10), 1-butanol(11) and triacetic acid lactone(12). These studies evaluated a set of variants that altered the expression or coding sequences of one or two key enzyme genes encoded on a plasmid(8, 10-12). Similarly, a lysine-responsive sensor-reporter was used to uncover new endogenous enzyme mutants in *Corynebacterium glutamicum* implicated in higher L-lysine production(9).

We sought to expand the scope of sensor-directed metabolic engineering to the directed evolution of whole endogenous pathways. Using FBA as a guide, we simultaneously targeted up to 18 *E. coli* genomic loci to induce mutations in regulatory or coding sequence of genes implicated in biosynthesis of a target molecule. We established a robust selection, utilizing a sensor protein responsive to the target chemical to regulate the expression of an antibiotic resistance gene. Nearly a billion pathway variants could be evaluated simultaneously, enriching for the best producers when selection pressure was applied.

A major challenge faced by this selection approach (and a difficulty for most genetic selections) is the incidence of cheater cells that survive without producing the target molecule. These cheaters evolve to survive selection by mutating the sensor or selection machinery, rather than through higher target molecule synthesis. Lacking a metabolic burden, these 'evolutionary escapees' outcompete the top producers during a selection. Multiple selection cycles compound escape, obscuring productive cells and making further pathway evolution infeasible. We therefore devised a selection scheme that, by toggling between negative and positive selection, allows us to remove escapees from the population

when they arise. This strategy maintained high selection fidelity, permitting multiple rounds of evolution to progressively enrich for higher producing cells.

For sensor-reporter metabolic engineering to be generalizable, sensor domains specific to many different target molecules must be available. Fortunately, natural sensors exist for a wide array of industrially-relevant chemicals, including aliphatic hydrocarbons, short-chain alcohols, sugars, amino acids, polymer building blocks, and vitamins. Many more sensor domains are likely to be present among the thousands of additional bacterial regulators known from sequence(13-15) that remain to be characterized. We adapted 10 regulators to our selection system, creating synthetic dependence on their cognate inducer molecules, and demonstrated the utility of two of these for genome-wide metabolic engineering.

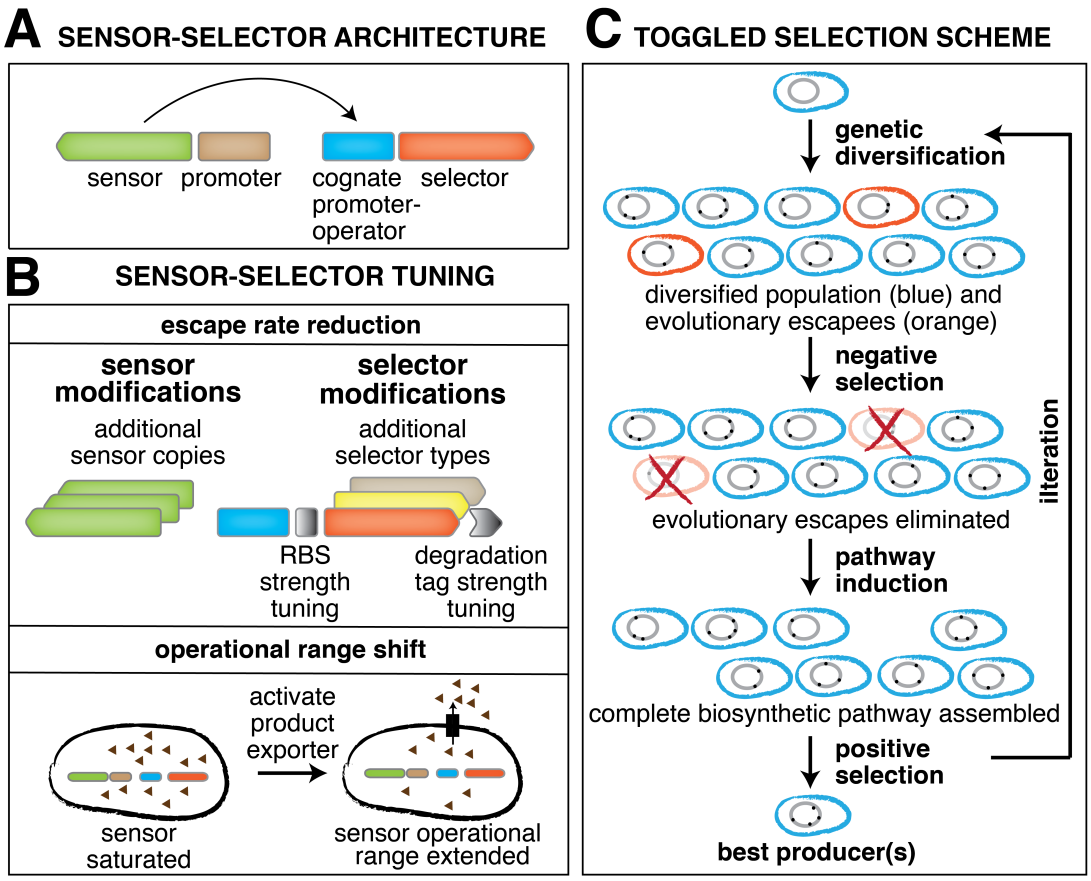


Figure 4.1: Sensor-selector design and pathway optimization through toggled selection (**A)** Sensor-selector genetic architecture. (**B)** Methods for tuning sensor-selectors to reduce escape rate and shift operational range. Escape rate is reduced by (1) adding a degradation tag, (2) mutating the RBS of the selector, (3) including multiple orthogonal selectors or (4) including an additional copy of the sensor. Activating an exporter shifts the sensor-selector operational range. (**C)** Toggled selection protocol for biosynthetic pathway optimization through multiple rounds of evolution. Negative selection eliminates cheaters; subsequent positive selection identifies higher producing clones from a diverse library.

Results

Sensor-selectors are a specific example of the sensor-reporter paradigm that use a gene whose product confers a fitness advantage (e.g. antibiotic resistance) as the reporter. Our sensor-selector architecture encodes a chemical-responsive sensor domain together with its cognate promoter, which controls a selectable reporter (Fig. 4.1A). We show that

this general implementation is suitable for transcriptional regulators (both activators and repressors) and riboswitches, that collectively respond to a wide variety of chemicals (Fig. 4.2A and Table C.1).

Each sensor-selector exhibits unique behavior, dependent on sensor affinity for the chemical, sensor type and induction response; for example, the escape rate and operational range can vary over orders of magnitude for different sensors (Fig. 4.2A). For each sensor, the operational range is defined as the chemical concentration range over which cells continue to experience a marginal fitness advantage with increasing concentration. The lower bound of the range reflects the lowest concentration of exogenously supplied inducer that provides a selective advantage. The upper bound of the range indicates that higher inducer concentration provides no additional fitness advantage. This range informs the utility of a sensor for optimizing a pathway. We measured the operational range of ten sensor-selectors; the MphR(21), TtgR(22) and TetR (24) operational ranges were measured for multiple inducers (Fig. 4.2A).

Under selection pressure, most cells in a sensor-selector strain population survive only when the target chemical is detected. But a small fraction of the cells survive absent the chemical. 'Evolutionary' escape results from mutations that permanently reduce selection sensitivity, and additionally, natural sensors may not have evolved to completely repress the basal expression level of the genes they regulate. In our selections, the resultant constitutive or leaky selector expression generates false positives, making it difficult to identify rare winners. Promoter engineering to optimize the placement of operator sites can yield very tight repression(16), but this approach requires specific development for each sensor. Instead, our standardized construction allows us to reduce the effect of leaky

selector expression through common *cis*-regulatory modifications that are sensor-independent. These modifications include appending a degradation tag to the selector to accelerate its proteolysis and mutating the ribosome binding site (RBS) of the selector gene to attenuate translation (Fig. 4.1B).

We implemented several modifications in the TtgR-TolC sensor-selector strain for comparison. Appending *ssrA* degradation tag variants to TolC reduced escape, in correlation to the strength of the degradation tag(17), by as much as six orders of magnitude (Fig. 4.2B). However, reduced escape also decreased the operational range. We adjusted the spacing between the RBS and translation start site of TolC to achieve fine-grained translation control(18). Five of ten spacing mutations reduced escape rate while maintaining a measurable operational range (Figs. 4.2B, C.6). For a dual selector strain, in which TtgR regulates both *tolC* and a kanamycin resistance gene, observed escape rates support the hypothesis of escape through leaky reporter expression: with both SDS and kanamycin present, the escape rate was much lower ($5.2\pm0.21 \times 10^{-8}$ cell/cell) than with either SDS alone ($1.7\pm0.092 \times 10^{-5}$ cell/cell) or kanamycin alone ($4.4\pm0.44 \times 10^{-4}$ cell/cell). Finally, we observed substantial escape rate reduction using two copies of the *ttgR* sensor gene and a single TolC selector (Fig. 4.2B). Because TtgR acts as a transcriptional repressor, evolutionary escape requires inactivating mutations to both gene copies, and higher sensor expression may reduce escape through tighter basal repression of the selector.

Sensors are useful for pathway optimization only when the intracellular concentration of the target chemical is within the operational range of the sensor. We hypothesized that expressing an exporter of the target chemical should decrease the intracellular concentration, shifting the operational range (Fig. 4.1B). We studied this

export effect by expressing a tetracycline exporter, TetA, in cells that place the tetracycline-responsive sensor, TetR in control of chloramphenicol acyltransferase (CAT) expression. When this strain expressed *tetA*, the entire operational range for tetracycline, including both the lower detection threshold and upper saturation point, shifted about ten-fold higher (Fig. 4.2D). This effect was tunable by controlling *tetA* expression from the arabinose-inducible pBAD promoter (Fig. 4.2E). The CAT selector was used here due to improved titration of drug sensitivity.

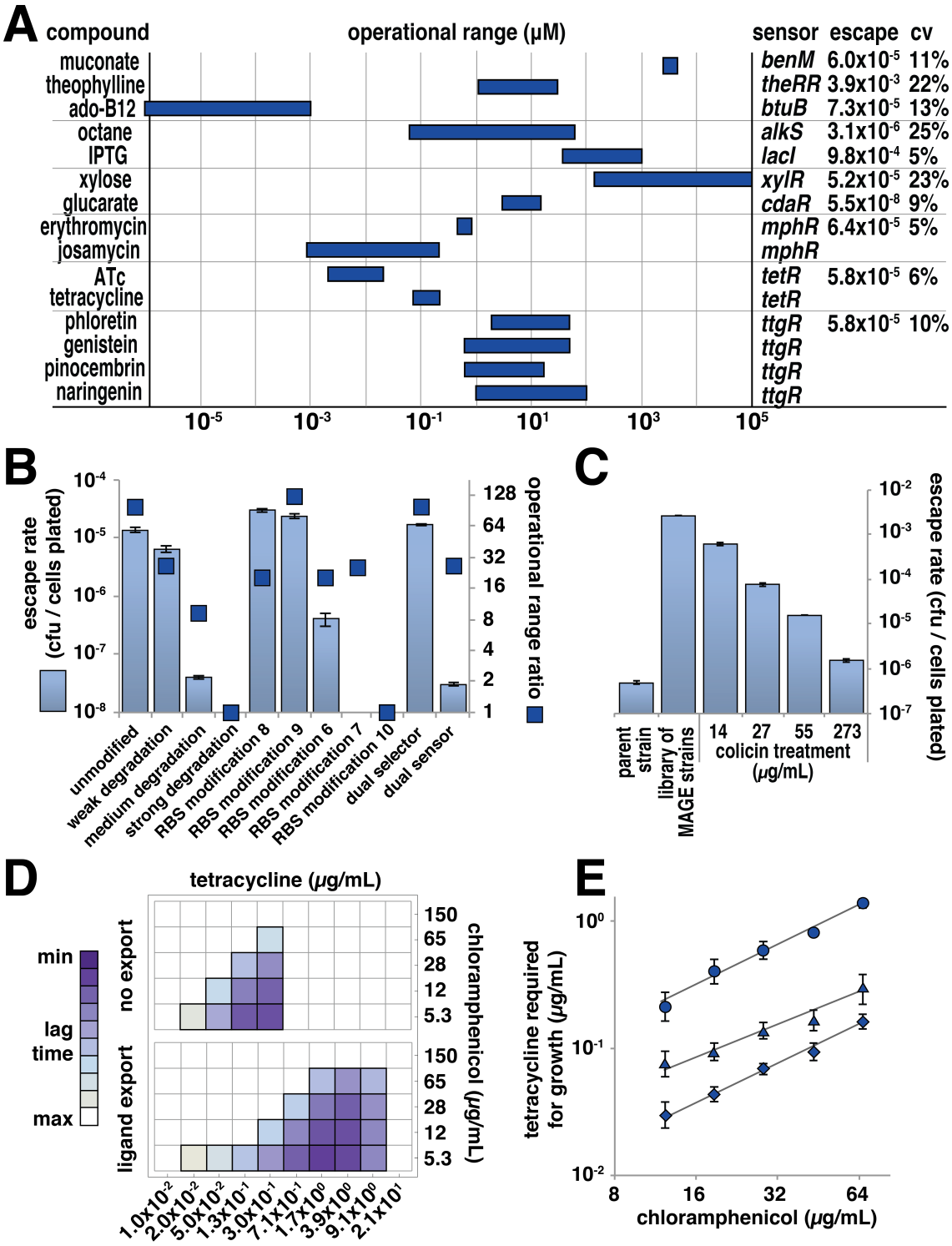


Figure 4.2: Characterization of sensor-selector modifications. **(A)** Escape rate and operational range of ten sensors with cognate inducer chemicals and TolC as a selector. Horizontal bars depict the operational range. The lower bound of the range reflects the lowest concentration of exogenously supplied inducer that provides a selective advantage.

The upper bound of the range indicates that higher inducer concentration does not increase fitness advantage. (**B)** Effect of genetic modifications on the TtgR-TolC sensor-selector escape rate and operational range. Escape rate (light blue bars, left axis) is the proportion of cells that evade selection (cfu/cells plated). Escape rate not shown if below the limit of detection (10^{-10} cfu/cells plated). Escape rate operational range ratio (blue boxes, right axis) is the ratio of the high concentration of the operational range to the low concentration of the operational range. (**C)** MAGE mutagenesis increases the escape rate (CFU/cells plated) in the CdaR-TolC strain. Treatment with colicin E1 removes escapees in a dose-dependent manner. (**D)** Tetracycline exporter (*tetA*) expression shifts the operational range of the TetR-CAT (chloramphenicol acetyltransferase) sensor-selector. Growth lag times reported for orthogonal concentration gradients of tetracycline vs. chloramphenicol in the absence of *tetA* (top panel) compared to *tetA* expression (bottom panel). (**E)** The shift in TetR-CAT operational range is tunable by titration of *tetA* expression. The minimum tetracycline concentration required for growth (y-axis) at a given selection pressure (x-axis) for three *tetA* expression levels: none (diamonds), intermediate (triangles), high (circles). Error bars represent S.E.M. of production from 3 biological replicates.

Pathway Evolution by Toggled Selection

To maximize the likelihood of identifying rare cells with a higher production phenotype, we developed a toggled selection scheme (Fig. 4.1C) that preserves library complexity while eliminating evolutionary escapees. Evolutionary escapees are cells that acquire mutations to survive selection without producing the target chemical. This escape prevents the identification of rare winners in a selection, and confounds multiple rounds of evolution as these escapees outcompete the productive cells. Through toggled selection we can selectively kill the escapees at each round, and carry over the productive cells for further improvements in subsequent rounds. We chose TolC(19) as a selector because of its utility for both positive selection (using sodium dodecyl-sulfate; SDS) and negative selection (using colicin E1). MAGE is highly mutagenic, increasing the escape rate from below 10^{-7} to above 10^{-3} after five cycles in the CdaR-TolC sensor selector strain. This increase could be reversed by incubation with colicin E1 (Fig. 4.2C), because evolutionary

escapees evade SDS toxicity through mutations that constitutively express *tolC*, making them highly susceptible to colicin E1. Crucially, we ensure that productive cells are not also killed during negative selection by maintaining a pathway gene under tight transcriptional control, which prevents prematurely triggering the sensor (Fig. 4.1C). After negative selection, we induce the regulated enzyme, allowing cells to produce the target chemical, and the sensor expresses *tolC* in proportion to chemical production. By toggling to positive TolC selection with SDS, we enrich for higher producers, and these can be characterized for their production phenotypes or subjected to further pathway evolution (Fig. 4.1C).

Naringenin Pathway

We implemented the toggled selection scheme to evolve *E. coli* toward higher production of two chemicals: naringenin and glucaric acid. Naringenin, a pharmacologically useful plant flavonoid molecule, was chosen because previous efforts serve to benchmark our optimization(2, 3, 20). *E. coli* requires four heterologous enzymes to synthesize naringenin from glucose: tyrosine ammonia lyase (TAL), 4-coumaroyl ligase (4CL), chalcone synthase (CHS) and chalcone isomerase (CHI)(3) (Fig. 4.3A). Because this pathway consumes tyrosine and malonyl-CoA, our strain engineering strategy targeted endogenous *E. coli* gene regulatory and coding loci to increase the availability of these precursors (SI Appendix, Table S4). The scope of this work was genomic mutagenesis, so the heterologous genes were left untargeted.

We performed flux balance analysis (FBA) toward increased malonyl-CoA, because its availability limits naringenin production (SI Appendix, Table S6)(2, 20). FBA identified three key pathways: glycolysis, fatty acid biosynthesis and the tricarboxylic acid (TCA) cycle (Fig. 4.3A and Table C.3). Greater flux through glycolysis by up-regulation of *gapA*,

pgk and *pdh* should increase pools of acetyl-CoA, which is converted to malonyl-CoA by acetyl-CoA carboxylase enzymes *accABCD*. Because acetyl-CoA is oxidized in the TCA cycle, we targeted for down-regulation TCA enzymes *mdh*, *fumBC* and *acnAB*. To throttle acetyl- and malonyl-CoA consumption in fatty acid biosynthesis, we targeted *fabBDFH* for down-regulation. Availability of tyrosine, the other precursor for naringenin production, is limited by activity of two enzymes in aromatic biosynthesis, *aroG*(21) and *tyrA*(4) that are inhibited by 3-deoxy-D-arabinoheptulosonate 7-phosphate (DAHP) and chorismate, respectively. We targeted *aroG* and *tyrA* for coding sequence changes shown to alleviate product inhibition. These predictions (Fig. 4.3A) corroborate interventions experimentally shown to increase production of malonyl-CoA(20), tyrosine(22) and naringenin(2, 3).

Previous efforts to engineer the naringenin pathway have relied on plasmid-based over-expression or complete knockouts(20); for tightly-regulated or essential central metabolism genes, such drastic modifications can have deleterious growth defects. For finer control of gene expression states, which can more closely balance biosynthetic and survival objectives, we used multiplex automated genome engineering (MAGE)(6). Oligonucleotides for MAGE mutagenesis were targeted to Shine-Dalgarno sequences to finely increase or decrease translation efficiency, to alternative start codons (CTG, GTG or TTG) to yield larger translational attenuation, or to introduce premature stop codons or coding frameshifts for complete inactivation (SI Appendix, Table S4). Seven genes were identified by FBA for overexpression to increase flux through glycolysis and to convert acetyl-CoA to malonyl-CoA.

Four rounds of evolution by toggled selection were performed on the strain containing two copies of the *ttgR* sensor gene controlling TolC, due to its favorable

combination of escape rate and operational range (Fig. 4.2B). We verified that TtgR responds only to naringenin but not the pathway intermediate *p*-coumaric acid (Fig. C.3). After four rounds, each consisting of about 15 cycles of targeted mutagenesis followed by toggled selection, the best strain identified produced 36 times more naringenin than the parent strain (Fig. 4.3B). We screened approximately 20 colonies to identify the highest producer at each round. With a supernatant concentration of 39 mg/L, the production titer of this strain surpasses the highest published production of naringenin (29 mg/L) directly from glucose(3) (Fig 4.3B). We further enhanced the production titer to 61 mg/L by overexpressing *E. coli* acetyl-CoA carboxylase genes (*accABCD*), which have been shown to increase endogenous malonyl-CoA levels (Figs. 4.3B, C.1). Through genetic changes alone, we were able to nearly recapitulate the high naringenin titer (84 mg/L) previously achieved by addition of cerulenin, an inhibitor of fatty acid biosynthesis, which is prohibitively expensive for industrial scale production (3).

We sequenced the genomes of the starting strain and seven high-producing strains isolated after evolution round four. All seven strains incorporated RBS or start codon changes at several targeted loci (Figure 4.3C). We found a number of mutations associated with malonyl-CoA production (Fig. 4.3C and Table C.7). In the TCA cycle, fumarase was down-regulated by a *fumC* start codon mutation in all seven strains (likely due to its selection in an early round). Several fatty acid genes were also down-regulated. Fatty acid biosynthesis genes whose products initiate synthesis from acetyl-CoA (*fabH*) or malonyl-CoA (*fabD*) were down-regulated by start codon or RBS mutations in seven and four strains, respectively. The fatty acid elongation gene *fabF* had start codon attenuation (GTG to TTG) in four strains and a purine to pyrimidine mutation in the RBS predicted to lower

translation rate(23) in a fifth strain (Fig. 4.3C and Table C.7). None of the seven strains had a down-regulation target knocked out, and none of the strains had mutations affecting *fabB*, an essential gene, reflecting a balance between production and growth objectives. Computational prediction of translation rate shows that selected clones enrich for RBS and start codon mutations that attenuate translation of genes, consistent with FBA predictions (Fig. C.4).

Three strains exhibited targeted mutations in tyrosine biosynthetic genes shown to alleviate product inhibition. All three produced substantially more coumaric acid, including two strains with the *tyrA* mutation A354V, which produced at least an order of magnitude more coumaric acid (Fig. 4.3C). This large coumaric acid buildup suggests that malonyl-CoA may be limiting for naringenin production in these strains. In support of this idea, overexpression of the enzymes *accABCD,* which convert acetyl-CoA to malonyl-CoA, increased naringenin production almost 1.5-fold in the evolved strain (Fig. 4.3B).

While the MAGE process concentrates diversity generation on targeted loci and increases the probability of sampling specific mutations hypothesized to confer beneficial phenotypes, it also has unintended mutagenic effects. Whole genome sequencing revealed many non-targeted mutations in the producer strains (Fig. 4.3C and Table C.8), including several mutations likely involved in higher naringenin production. Frameshifts inactivated *mhpD*, which catabolizes aromatic compounds similar to coumaric acid(24), and *hcaT,* a putative transporter of phenylpropionates like coumaric acid(25). Similarly, a frameshift in *entB*, which diverts chorismate from aromatic biosynthesis, may increase tyrosine production(26). We speculate that knocking out all three enzymes facilitates production of naringenin by increasing the concentration of the precursor, p-coumaric acid. Attributing

function to non-coding regulatory mutations is more tenuous. However, we observed a mutation in the Shine-Dalgarno sequence of *rpoD*, mutation of which increases tyrosine production(22).

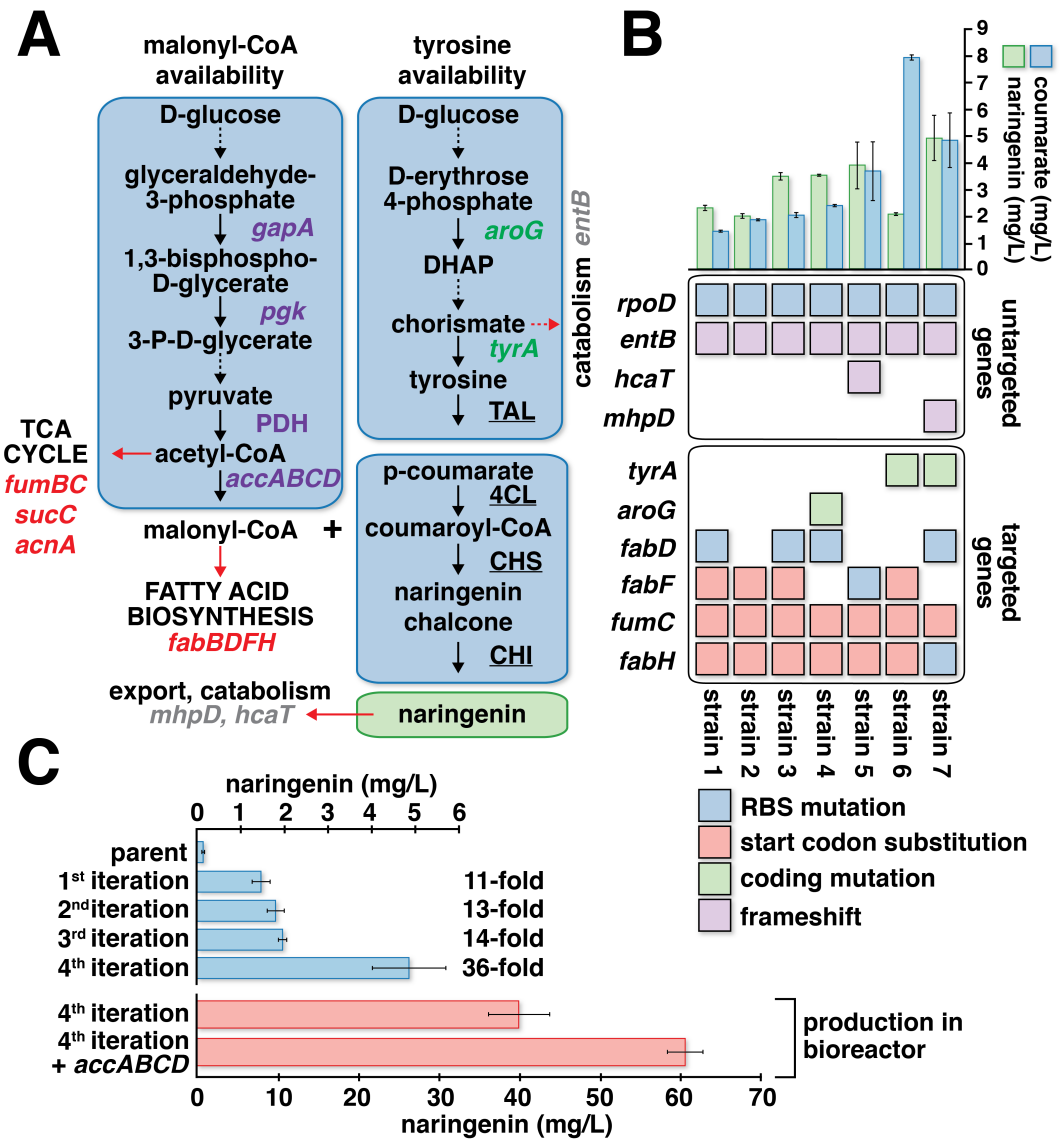


Figure 4.3: Optimization of the naringenin biosynthetic pathway. (**A**) Endogenous *E. coli* genes targeted by MAGE to increase malonyl-CoA and tyrosine availability for naringenin production; targeted genes are colored: blue: up-regulation, red: down-regulation, green: coding changes, gray: untargeted knocked out genes. (**B**) Average naringenin production titers (mg/L) for parent and highest producer after each round of evolution (blue bars). Production titer (mg/L) from fed batch bioreactor fermentation of the highest producer

and highest producer with *accABCD* overexpressed (red bars). **(C)** Genotype and production phenotype of the top seven producers (in no particular order) from the fourth round of toggled selection. Colored boxes denote the type of genetic modification. Mutations founds at targeted genes are shown in the left box, those at untargeted genes, right box. Naringenin (light bars) and coumaric acid (dark bars) concentrations for single production measurements are shown alongside the corresponding genotype. Error bars represent S.E.M. of production from 3 biological replicates.

Glucaric Acid Pathway

In order to validate directed evolution by sensor-selectors as a generalizable method, we optimized the production of glucaric acid in *E. coli*. Glucaric acid was chosen for two reasons. First, unlike naringenin production, previous work to modulate endogenous pathways was absent. Second, glucaric acid was identified as a key renewable chemical for the replacement of petroleum-based polymer production. Glucaric acid can be synthesized in *E. coli* by expression of three exogenous enzymes: myo-inositol-1-phosphate synthase (Ino1), myo-inositol oxygenase (MIOX) and uronate dehydrogenase (Udh)(27) (Fig. 4.4A).

To ensure that the heterologous enzymes were functional and provided a growth advantage under selective conditions, we measured growth lag times in the cdaR-tolC sensor-selector strain after exogenously providing pathway intermediates (glucose, myo-inositol and glucuronic acid). Furthermore, we verified cdaR is specifically activated by glucaric acid, and does not respond to pathway intermediates myo-inositol and glucuronic acid (Fig. C.7).

Increasing concentrations of glucaric acid result in lower lag times for cells grown in the presence of SDS. Under selective conditions, decreasing growth lag times reflect the decreasing number of enzymatic reactions required to produce glucaric acid for cdaR-tolC activation (Fig. 4.4B). Higher concentrations of myo-inositol and glucuronic acid resulted in

shorter lag times under selective conditions, but increasing glucose or glucaric acid concentrations in the media did not result in a growth advantage. In the case of glucaric acid this is expected as both 1 mM and 10 mM are above the operational range. With glucose, one possible explanation is that an increase in glucose in the media results in additional flux through glycolysis and central metabolism rather than increased flux through the glucaric acid pathway, which likely operates slower than glycolysis. The lag time observed at the high glucuronic acid concentration is comparable to the lag time observed with glucaric acid supporting the previous finding that the Udh enzyme acts on a fast time scale when compared to the selection (27, 28). A long lag time even at a high concentration of myo-inositol indicates that the MIOX enzyme is less efficient as reported in previous work(29).

Efforts to increase glucaric acid production in *E. coli* have focused on co-localization of pathway enzymes(30) and improving MIOX solubility(29). To date, modifying endogenous *E. coli* pathways has not been explored. We hypothesized that glycolysis and the pentose phosphate pathway were competing with Ino1 for glucose-6-phosphate (g6p), the branch-point for glucaric acid production. We used MAGE to introduce degeneracy in the RBS of genes involved in catabolism of g6p (SI Appendix, Table S5). We similarly targeted the RBS sequences of *mdh* and *suhB*, the endogenous phosphatase responsible for dephosphorylating myo-inositol-1-phosphate (31) (Fig. 4.4A). Degeneracy in the RBS sequences allowed the selection to sample both up- and down-regulation of the genes. We hypothesized that tuning the rate of glycolysis would allow the glucaric acid pathway to compete for glucose more effectively while still facilitating robust cell growth. The product of the *pgi* gene shuttles g6p into glycolysis and its disruption has been shown to increase

the intracellular pool of g6p(32), the substrate of Ino1. The growth defect of a *pgi* mutant can be rescued by overexpression of *sthA*(31) and thus *pgi* and *sthA* were chosen for simultaneous expression modulation. The other major pathway for g6p catabolism is the pentose phosphate pathway and is initiated by the product of *zwf,* which was also targeted for expression modification. To prevent flux diversion of the intermediate molecule glucuronic acid into the Entner-Doudoroff pathway, we targeted uronate isomerase (*uxaC*) for a knockout. To avoid catabolism of glucaric acid, we also targeted glycerate kinase (*garK*) for a knockout.

We performed 5 cycles of MAGE on seven genomic targets (SI Appendix, Table S5) to achieve a predicted prevalence of approximately $1x10^{-6}$ for strains incorporating mutations at all seven loci. The statistically most common strain contained a single mutation and was predicted to account for 40% of the cell population. After MAGE followed by toggled selection, the enriched non-clonal culture produced 7-fold more glucaric acid than the parent. The best clone isolated from this population produced 22-fold more than the parent (Fig. 4.4C and Table C.7). This highest producing strain contained a targeted nonsense mutation in *garK*, a gene not previously shown to enhance glucaric acid production. None of the other targeted genes were mutated, but an untargeted nonsense mutation in the L-glyceraldehyde 3-phosphate reductase gene (*yghZ*) was found. As an aldo-keto reductase, *yghZ* has fairly broad substrate specificity(33) and could be diverting carbon flux away from glucaric acid by reducing glucuronate to gluconate.

Glucaric acid titers were improved 22-fold over the parent strain; however, absolute production of glucaric acid remained substantially lower (1.2 mg/L, Fig. 4.4C) than previously reported titers(27). Moon *et. al.* carried out glucaric acid production in an *E. coli*

B-strain (BL21), while we optimized the pathway in the MAGE-competent E. coli K-strain. To investigate the possible role of strain background (B vs K strains) in glucaric acid production, we measured glucaric acid titer in our parent K-strain and BL21. We found that glucaric acid titer was 300 times higher in BL21 with the same glucaric acid enzymes and culture conditions (Fig. C.5).

There are substantial differences between B and K strains of *E. coli* that are difficult to bridge through naïve mutagenesis. Notably, B strains have altered carbohydrate metabolism when compared to K strains as well as an enhanced capacity for recombinant protein production(28). Previous work to produce glucaric acid in *E. coli* has revealed MIOX to be a highly unstable enzyme(27) and the primary limit on production may lie in protein folding and stability, rather than host-cell glucose metabolism. Our evolved K strain grew just slightly worse than the parent strain, ruling out gross metabolic deficiency as the cause of low production (Fig. C.8). In light of these considerations, subsequent rounds of diversification and selection were not pursued in the K-strain background. Currently, work is underway to enable MAGE in BL21 for optimization of production pathways better suited for *E. coli* B-strains. These results highlight that directed evolution is not a replacement for the careful choice of a host strain, but should complement thoughtful strain selection.

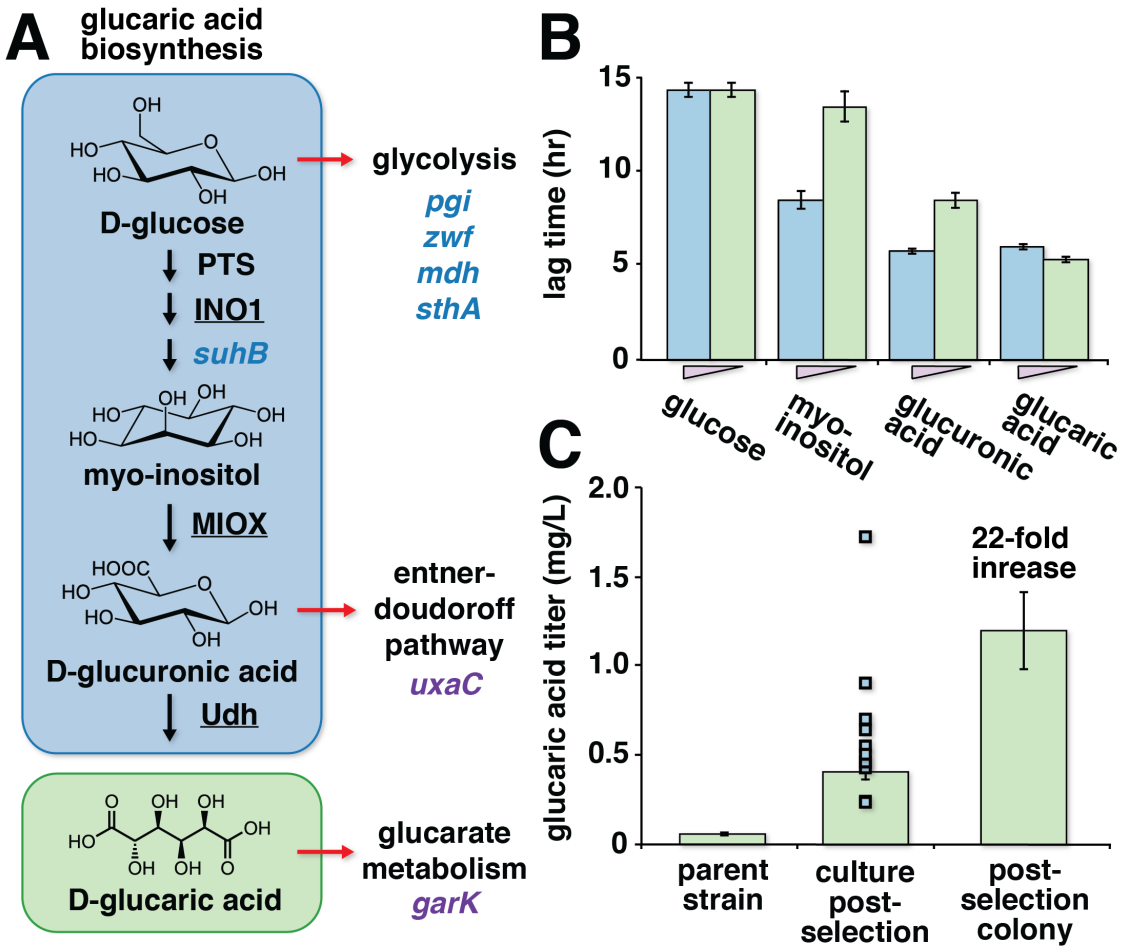


Figure 4.4: Optimization of the glucaric acid biosynthetic pathway. (**A)** Glucaric acid biosynthetic pathway showing key intermediate metabolites and enzymes. Heterologous gene names are underlined. Endogenous *E. coli* genes targeted by MAGE for expression modification: blue: RBS modification, red: knockout. (**B)** Lag time in growth reflects time required for the pathway enzymes to produce activating levels of glucaric acid in the sensor-selector strain CdaR-TolC. Pathway intermediates are supplied exogenously (dark blue 10 mM; light blue 1 mM). Error bars represent S.E.M. of production from 3 biological replicates (**C)** Glucaric acid titers in parent strain, post-selection mixed population and highest producing clone (bars). Squares indicate titers produced by clones isolated from the post-selection population. Error bars represent S.E.M. of production from 3 biological replicates.

Discussion

Rapid advances in DNA sequencing and DNA synthesis technologies(34, 35) have not been accompanied by similar advances to enable the high throughput evaluation of phenotypes. Our implementation of small-molecule sensors coupled to selection advances a versatile platform that can transform biosynthetic phenotypes into fitness differences. These differences empower us to employ evolution followed by sequencing to reveal clues to potential metabolic pathway inefficiencies and to identify targets for subsequent rounds of evolution. The multiplex mutations facilitated by MAGE enable us to target all candidate genes predicted by FBA without prior assumption of the relative importance of each target. Because selection amplifies faster dividing cells, we indirectly enrich for variants that suitably balance biomass and biosynthetic objectives. We show that toggled selection refreshes the pool of productive cells by removing evolutionary escapees. Toggled selection enables multiple rounds of evolution to progressively enrich for higher producing variants. Combining beneficial mutations from independently evolved strains could lead to even higher metabolite production due to epistatic synergies. The incidence of evolutionary escapees and off-target mutations is likely to be significantly reduced by transiently repressing mismatch repair(36). While this may decrease untargeted beneficial mutations (*mhpD*, *entB* and *hcaT* in naringenin biosynthesis) in a single round of evolution, mutations that provide significant selective advantage will ultimately be enriched over multiple rounds.

Besides pathway optimization, we can use sensor-selectors to screen libraries of synthetic or metagenomic sequences for novel biosynthetic operons, new enzyme

functions, and transporters. The vast reservoir of natural chemicals found in microbial species remain largely inaccessible because the enzymatic pathways for their synthesis are not known. With sensor-selectors, large libraries encoding natural or synthetic operons can be interrogated to identify the putative pathway for a target chemical.

Natural sensor domains exist for many classes of molecules that are of economic interest; however, some metabolite targets have no known sensor to detect them. We expect this challenge to be addressed by advances in protein design and by efforts to characterize new transcription factors encoded in metagenomes. Clever use of existing sensors will also allow the optimization of multiple pathways that use common intermediates. For biosynthetic pathways diverging only in late 'decoration' steps, we can leverage class-specific sensors to optimize the production of many related molecules by simply exchanging terminal enzymes. For example, our best naringenin production strain likely has an elevated intracellular concentration of malonyl-CoA, which could be used immediately for the improved production of fatty acid-derived targets or polyketides.

Evolution is a powerful tool for resolving the complexity of biology. Using evolution to guide rational design should ultimately lead to a better understanding of the genotypic basis of biological function.

Materials and Methods

Sensor-Selector Strain Construction

All sensor-selector strains were constructed from *E. coli* MG1655 derivative EcNR2 (Δ*bioAB::*Red-λ prophage-*bla* Δ*mutS::Cm*) to facilitate recombineering and MAGE(6).

Sensor-selector constructs were genomically integrated using a standard architecture (Fig. 4.1A).

Glucaric Acid Pathway Construction and Optimization

A plasmid (pT7GAEXP) enabling glucaric acid biosynthesis in *E. coli* was constructed, encoding: the *Mus musculus* myo-inositol-oxygenase (MIOX) gene; the *Saccharomyces cerevisiae* inositol-1-phosphate synthase (INO1) gene; and the *Agrobacterium tumefaciens* uronate dehydrogenase (UDH) gene. MAGE (8) mutagenesis was used to target seven genes (SI Appendix, Table S5) for expression change in sensor selector strain CdaR-TolC expressing pT7GAEXP. One cycle of toggled negative and positive selection was used to enrich for mutations benefitting glucaric acid production, as assayed by clonal production and mass spectrometry.

Naringenin Pathway Construction and Optimization

Four heterologous genes enabling naringenin production were cloned into two plasmids for expression in the dual TtgR-TolC sensor-selector strain: tyrosine-ammonia lyase (TAL), 4-coumarate:CoA ligase (4CL), chalcone synthase (CHS) and chalcone isomerase (CHI) (3) MAGE (8) mutagenesis targeted 20 endogenous genes (SI Appendix, Table S3) for expression and coding changes in this strain. Four iterations of mutation and toggled negative and positive selection enriched for mutations benefitting naringenin production, as assayed by clonal production and mass spectrometry.

References

1. Nakamura CE, Whited GM (2003) Metabolic engineering for the microbial production of 1,3-propanediol. *Curr Opin Biotechnol* 14:454–459.

2. Leonard E, Lim K-H, Saw P-N, Koffas MAG (2007) Engineering central metabolic pathways for high-level flavonoid production in Escherichia coli. *Appl Environ Microbiol* 73:3877–3886.

3. Santos CNS, Koffas M, Stephanopoulos G (2011) Optimization of a heterologous pathway for the production of flavonoids from glucose. *Metab Eng* 13:392–400.

4. Lütke-Eversloh T, Stephanopoulos G (2005) Feedback inhibition of chorismate mutase/prephenate dehydrogenase (TyrA) of Escherichia coli: generation and characterization of tyrosine-insensitive mutants. *Appl Environ Microbiol* 71:7224–7228.

5. Yim H et al. (2011) Metabolic engineering of Escherichia coli for direct production of 1,4-butanediol. *Nat Chem Biol* 7:445–452.

6. Wang HH et al. (2009) Programming cells by multiplex genome engineering and accelerated evolution. *Nature* 460:894–898.

7. van Sint Fiet S, van Beilen JB, Witholt B (2006) Selection of biocatalysts for chemical synthesis. *Proc Natl Acad Sci USA* 103:1693–1698.

8. Tang S-Y, Cirino PC (2011) Design and application of a mevalonate-responsive regulatory protein. *Angew Chem Int Ed Engl* 50:1084–1086.

9. Binder S et al. (2012) A high-throughput approach to identify genomic variants of bacterial metabolite producers at the single-cell level. *Genome Biol* 13:R40.

10. Yang J et al. (2013) Synthetic RNA devices to expedite the evolution of metabolite-producing microbes. *Nat Commun* 4:1413.

11. Dietrich JA, Shis DL, Alikhani A, Keasling JD (2013) Transcription factor-based screens and synthetic selections for microbial small-molecule biosynthesis. *ACS Synth Biol* 2:47–58.

12. Tang S-Y et al. (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.

13. Gallegos MT, Schleif R, Bairoch A, Hofmann K, Ramos JL (1997) Arac/XylS family of transcriptional regulators. *Microbiol Mol Biol Rev* 61:393–410.

14. Ramos JL et al. (2005) The TetR family of transcriptional repressors. *Microbiol Mol Biol Rev* 69:326–356.

15. Tropel D, van der Meer JR (2004) Bacterial transcriptional regulators for degradation pathways of aromatic compounds. *Microbiol Mol Biol Rev* 68:474–500– table of contents.

16. Lutz R, 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 Res* 25:1203–1210.

17. Andersen JB et al. (1998) New unstable variants of green fluorescent protein for studies of transient gene expression in bacteria. *Appl Environ Microbiol* 64:2240–2246.

18. Chen H, Bjerknes M, Kumar R, Jay E (1994) Determination of the optimal aligned spacing between the Shine-Dalgarno sequence and the translation initiation codon of Escherichia coli mRNAs. *Nucleic Acids Res* 22:4953–4957.

19. DeVito JA (2008) Recombineering with tolC as a selectable/counter-selectable marker: remodeling the rRNA operons of Escherichia coli. *Nucleic Acids Res* 36:e4.

20. Xu P, Ranganathan S, Fowler ZL, Maranas CD, Koffas MAG (2011) Genome-scale metabolic network modeling results in minimal interventions that cooperatively force carbon flux towards malonyl-CoA. *Metab Eng* 13:578–587.

21. L-tyrosine-producing bacterium and a method for producing L-tyrosine (2005) L-tyrosine-producing bacterium and a method for producing L-tyrosine.

22. Santos CNS, Xiao W, Stephanopoulos G (2012) Rational, combinatorial, and genomic approaches for engineering L-tyrosine production in Escherichia coli. *Proc Natl Acad Sci USA* 109:13538–13543.

23. Salis HM, Mirsky EA, Voigt CA (2009) Automated design of synthetic ribosome binding sites to control protein expression. *Nat Biotechnol* 27:946–950.

24. Burlingame R, Chapman PJ (1983) Catabolism of phenylpropionic acid and its 3-hydroxy derivative by Escherichia coli. *J Bacteriol* 155:113–121.

25. Díaz E, Ferrández A, García JL (1998) Characterization of the hca cluster encoding the dioxygenolytic pathway for initial catabolism of 3-phenylpropionic acid in Escherichia coli K-12. *J Bacteriol* 180:2915–2923.

26. Gehring AM, Bradley KA, Walsh CT (1997) Enterobactin biosynthesis in Escherichia coli: isochorismate lyase (EntB) is a bifunctional enzyme that is phosphopantetheinylated by EntD and then acylated by EntE using ATP and 2,3-dihydroxybenzoate. *Biochemistry* 36:8495–8503.

27. Moon TS, Yoon S-H, Lanza AM, Roy-Mayhew JD, Prather KLJ (2009) Production of glucaric acid from a synthetic pathway in recombinant Escherichia coli. *Appl Environ Microbiol* 75:589–595.

28. Yoon S-H, Moon TS, Iranpour P, Lanza AM, Prather KJ (2009) Cloning and characterization of uronate dehydrogenases from two pseudomonads and Agrobacterium tumefaciens strain C58. *J Bacteriol* 191:1565–1573.

29. Shiue E, Prather KLJ (2013) Improving d-glucaric acid production from myo-inositol in E. coli by increasing MIOX stability and myo-inositol transport. *Metab Eng* 22C:22–31.

30. Moon TS, Dueber JE, Shiue E, Prather KLJ (2010) Use of modular, synthetic scaffolds for improved production of glucaric acid in engineered E. coli. *Metab Eng* 12:298–305.

31. Canonaco F et al. (2001) Metabolic flux response to phosphoglucose isomerase knock-out in Escherichia coli and impact of overexpression of the soluble transhydrogenase UdhA. *FEMS Microbiol Lett* 204:247–252.

32. Fraenkel DG (1968) The accumulation of glucose 6-phosphate from glucose and its effect in an Escherichia coli mutant lacking phosphoglucose isomerase and glucose 6-phosphate dehydrogenase. *J Biol Chem* 243:6451–6457.

33. Grant AW, Steel G, Waugh H, Ellis EM (2003) A novel aldo-keto reductase from Escherichia coli can increase resistance to methylglyoxal toxicity. *FEMS Microbiol Lett* 218:93–99.

34. Carr PA, Church GM (2009) Genome engineering. *Nat Biotechnol* 27:1151–1162.

35. Kosuri S et al. (2013) Composability of regulatory sequences controlling transcription and translation in Escherichia coli. *Proc Natl Acad Sci USA*.

36. Nyerges Á et al. (2014) Conditional DNA repair mutants enable highly precise genome engineering. *Nucleic Acids Res* 42:e62.

Chapter 5: A fluorimostat for programmable control of gene expression

Abstract

A simple-to-construct device providing programmatic and closed-loop control of gene expression would enable new experimental capabilities in biology labs everywhere. We have designed the fluorimostat to suit this need. The fluorimostat allows bioengineers to control gene expression *via* cellscript, a new biological scripting module for Python. We implemented a cellscript protocol to autonomously determine the cooperatively of the *mphR* macrolide biosensor over a wide range of *gfp* expression levels. The fluorimostat also enables closed-loop control of expression, allowing complex genetic programs to continuously correct for drift and other perturbations over periods of minutes to weeks. Designs and scripts are freely available on GitHub.

Introduction

Precise control of gene expression is required to study the dynamics of natural and synthetic biological systems(1,2). Achieving such exquisite control remains challenging, as cells studied in batch culture are constantly transitioning from one growth phase to another(2,3). Turbidostats maintain cells at a constant density and are used to achieve long-term steady-state growth, enabling experiments that would otherwise be confounded by the effects of growth phase transition. An even greater challenge is the maintenance of steady-state gene expression, which is difficult to achieve in batch culture, but would

enable much more robust experimentation. We developed the fluorimostat to address this challenge. The fluorimostat is an inexpensive and easy-to-construct device that enables a series of precise gene expression levels to be defined programmatically and maintained in the presence of perturbations.

We designed the fluorimostat to enable scriptable and autonomous control of biological systems for engineers and biologists everywhere. Previous closed-loop expression systems have relied on immobilized cells and epifluorescence microscopy for the evaluation of expression states(4,5). Other demonstrations have relied on batch cultures of cells, or evaluation with flow-cytometry(6,7). Most recently, light-controlled gene expression has been exploited for the convenient property that it can be instantaneously added or removed(7-9). However, relying on light constrains the range of inducible expression systems that can be studied to a limited set of regulators with low dynamic range(10). Furthermore, the above systems rely on expensive equipment not typically dedicated to cell culture, limiting the potential ubiquity of these systems. By relying on inexpensive components and providing detailed instructions for setup and operation, we make the fluorimostat available to labs everywhere.

The fluorimostat is designed as a bridge between biology and software. As such, we have followed open-source principles in making our designs accessible both intellectually and practically. In order to maximize backwards compatibility with existing inducible systems, we have designed the fluorimostat to work with any chemical inducer. To ensure forward compatibility, we monitor fluorescence with an inexpensive inline spectrophotometer, allowing multiplexed observation of any fluorophore of interest. We have developed a stand-alone GUI for observation and manipulation of the system. For

more complex biological programs, the behavior of the system can be scripted through the cellscript Python library. In this paper we develop a cellscript function that enables the fluorimostat to autonomously characterize the induction properties of the *mphR* macrolide antibiotic biosensor. Finally we implement a PID controller to demonstrate closed-loop control of gene expression.

Results and Discussion

The fluorimostat system is built from three discrete modules: the growth chamber, the induction chamber and the spectrophotometric flow cell. The growth chamber, operating as a turbidostat, maintains cells in a logarithmic growth phase through simple hysteretic control. When the optical density of the growth chamber reaches the specified upper limit, fresh media is added while spent media is removed until the optical density reaches the specified lower limit. Cells from the growth chamber are continuously pumped into the induction chamber. The concentration of inducer in the induction chamber is maintained at a defined level by software-actuated pumping of a concentrated inducer solution. The residence time of cells in the induction chamber is controlled by the ratio of flow rate to working volume.

Cells exiting the induction chamber pass through a spectrophotometric flow cell on their way to waste. The flow cell is constructed with 530nm (green) and 470nm (blue) LEDs for excitation of RFP and GFP, respectively. The LEDs and desired filters are held in place with a 3D-printed assembly and can be easily reconfigured. The spectrophotometer monitors the cells *via* a fiber-optic cable coupled to the flow cell at an angle orthogonal to the direction of illumination. The software parses the spectrum, informing the actuation of

inducer addition in the context of closed-loop control. The entire system is operated within a 37°C warm-room and constructed such that a simple clean-in-place (CIP) procedure can be performed in less than 20 minutes.

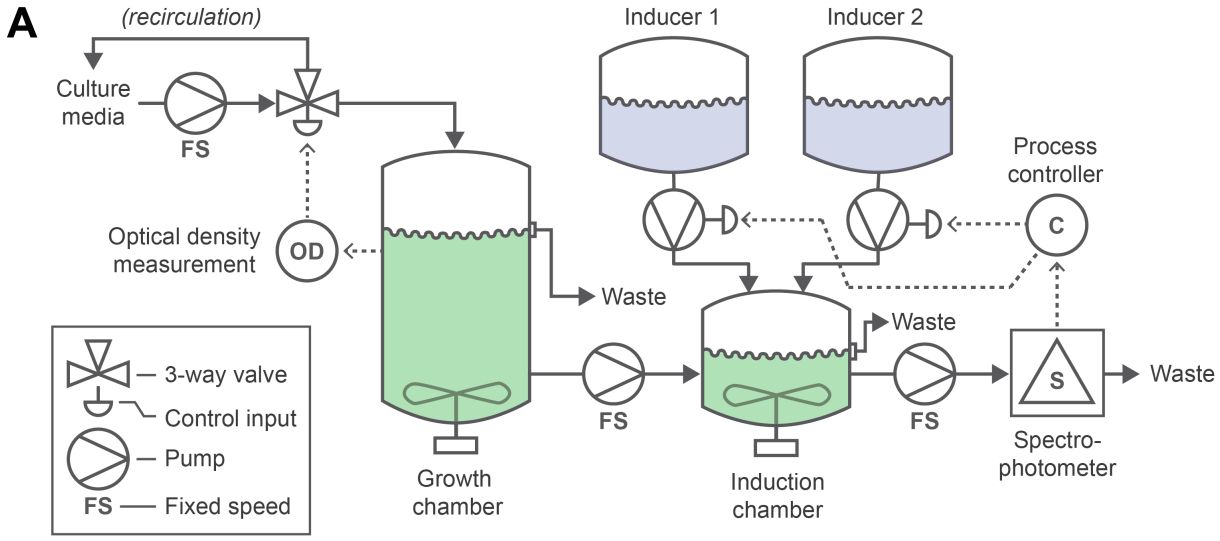


Figure 5.1: Fluorimostat process flow diagram. The growth chamber is maintained at a constant cell density through closed-loop hysteresis. The induction chamber is operated in either closed- or open-loop mode. Effluent from the induction chamber is passed through an LED illuminated spectrophotometric flow cell. The fluorimostat software calculates fluorescence intensity (a.u.) from the spectrum. The process controller actuates the inducer pump to achieve either a specified concentration of inducer (in open-loop mode) or a specified fluorescence level (in closed-loop mode).

All aspects of the fluorimostat are monitored and controlled through a GUI built in Labview and distributed in both binary and source-file formats. Configuration of the fluorimostat and scripting of biological programs can be performed through the cellscript Python library. The fluorimostat is instantiated as an object with user-configurable parameters such as flow rate, cell density and residence time. Biological programs are compiled from commands as simple as setting the induction level to a certain value for a

period of time, or as complicated as a sinusoid or other waveform with a given period and amplitude.

We used the fluorimostat to automate the characterization of a small-molecule inducible system. Characterization of inducible system parameters such as the hill coefficient or the maximal rate of expression is challenging in batch culture because large dilutions are needed to avoid approaching stationary phase before collecting of sufficient fluorescence data(2). Furthermore, in batch culture, the gene expression rate is the derivative of the fluorescence corrected for fluorophore maturation, degradation and dilution(2). In contrast, the fluorimostat operates at steady state, making fluorescence a direct measurement of the dilution-normalized gene expression rate. As a result, induction parameters are determined by simply modeling the transient behavior of the system as it adjusts to changes in inducer concentration. Arbitrary changes in the fluorimostat set point are used to estimate induction parameters using following system of equations:

$$C = C_0 e^{-Dt} + \frac{s \cdot r}{f + r}(1 - e^{-Dt})$$

$$F = F_0 e^{-Dt} + \left(V_{max} \frac{C^h}{C^h + K^h} + V_{min}\right) \frac{(1 - e^{-Dt})}{D}$$

The concentration of inducer and level of fluorescence in the induction chamber are denoted *C* and *F* with the subscript zero indicating initial conditions. *D* is the dilution rate of the induction chamber while *f* and *r* are the flow rates from the growth chamber to the induction chamber and from the inducer reservoir to the induction chamber, respectively. The concentration of inducer in the reservoir is denoted *s*. V_{max}, V_{min}, *h* and *K* are the hill function parameters that describe the induction response of the inducible system.

The concentration of inducer in the induction chamber is modeled precisely from the physical properties of the system. The fluorescence is in turn modeled from inducer concentration and the parameters of the hill function. Evaluating these equations each time the flow rate of inducer is modified results in a prediction of fluorescence over the course of the experiment. Minimizing the error between the predicted fluorescence and the observed fluorescence produces estimates for V_{max}, V_{min}, h and K. Bootstrapping the measured signal generates a distribution for each of the parameters and allows an estimate of the induction transfer function to be plotted. If the range of concentrations explored by the fluorimostat covers the majority of the transfer function then sufficient information was captured for a good fit.

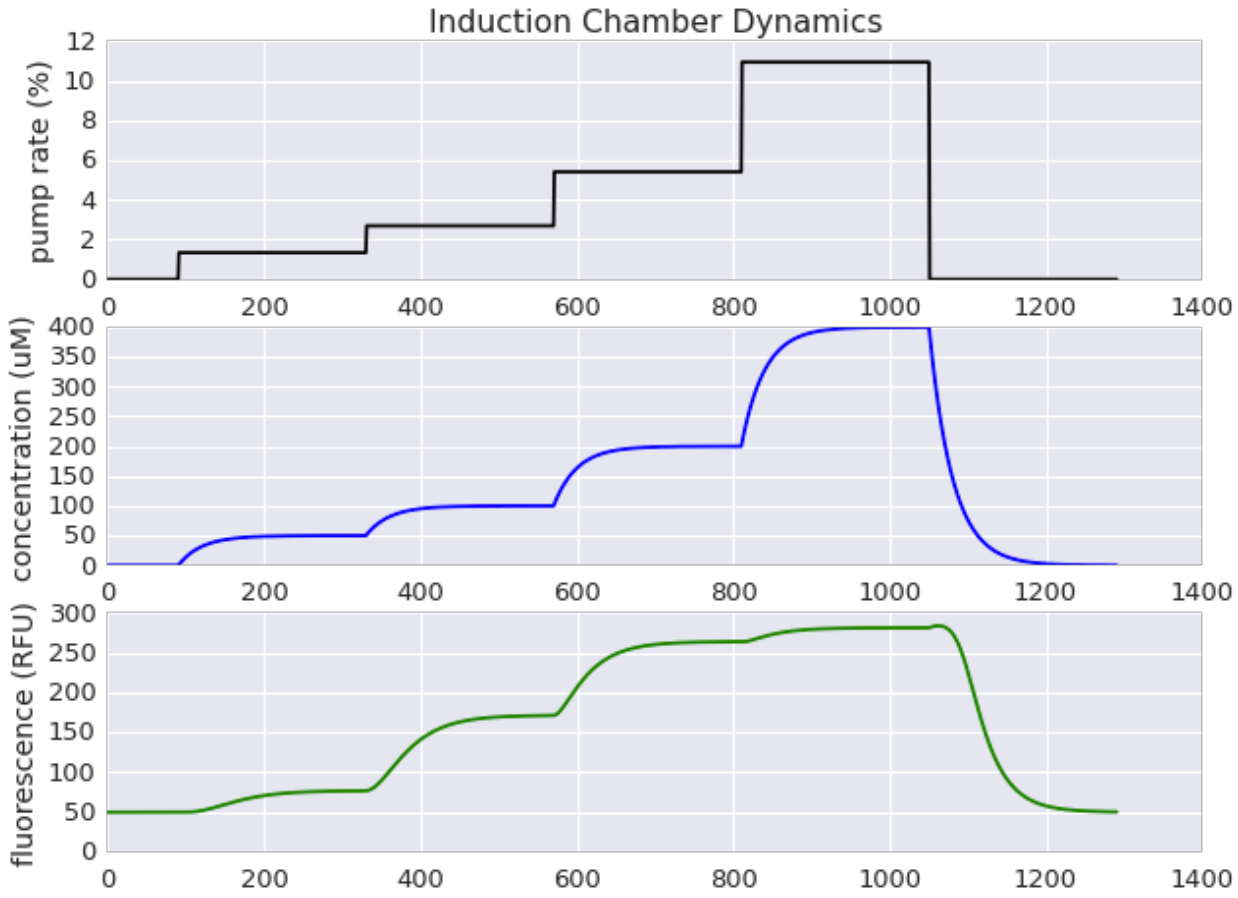


Figure 5.2: Computational design of an open-loop fluorimostat experiment. The user specifies a range of steady-state concentrations and their respective durations. The cellscript software encodes the desired concentrations as a series of inducer pump flow rates (top panel). The physical parameters of the system (*e.g.* pump rate, volume) are used to calculate the transient concentrations of inducer (middle panel). Promoter parameters estimated from previous experiments are used to predict the fluorescence that will be observed during the experiment (bottom panel).

We encoded this characterization pipeline in the cellscript Python module and deployed it for the automated characterization of *mphR*. The mphR transcriptional repressor permits gene expression when bound to erythromycin(11). Higher erythromycin concentrations result in greater gene expression rates from the *mphA* promoter, in turn driving expression of the green fluorescent protein (GFP). Characterization of *mphR* was carried out through a series of four logarithmically-spaced induction levels that we

anticipated would cover the majority of the induction transfer function. The resulting fluorescence response varied over a 25-fold range and yielded a mean *Vmax*, *Vmin*, *h* and *K* of 51 RFU/min, 258 RFU/min, 3.3 and 106μM, respectively. These values are similar to previously reported values derived for this promoter-repressor pair(2).

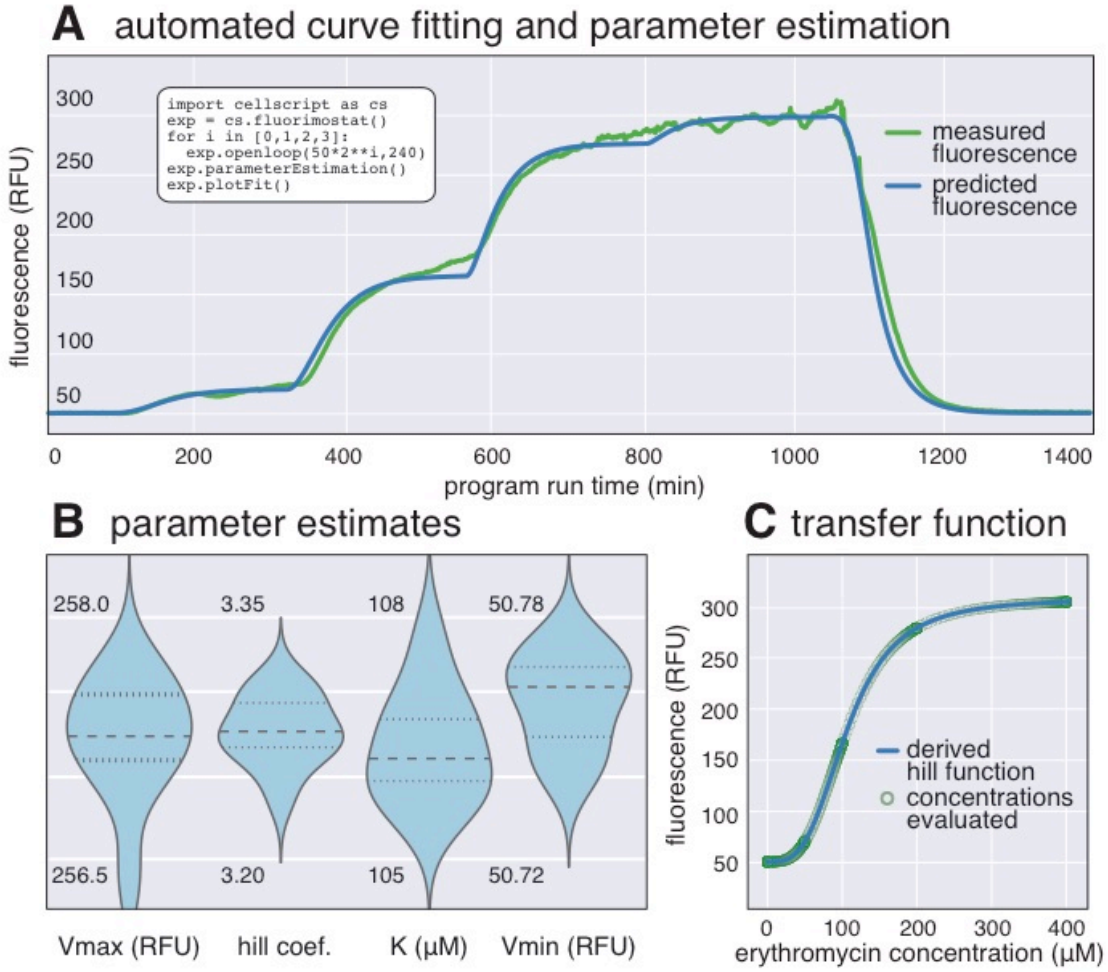


Figure 5.3: The fluorimostat and cellscript Python module were used to characterize genetically encoded biosensors. (A) A simple script encoded a range of erythromycin concentrations each maintained for four hours. A mathematical model of the *mphR* biosensor was fit to the observed fluorescence. (B) The parameters of the Hill function were estimated by fitting the model several times on data bootstrapped from the original measurements. (C) The transfer function between inducer concentration and gene expression rate was determined from the fitted parameters. The actual concentrations evaluated in the experiment are superimposed on the Hill function to visualize the breadth of the fit.

Proportional-integral-derivative (PID) control was implemented to enable long-term induction programs capable of correcting for drift and other perturbations. We demonstrated PID control by allowing the fluorimostat software to autonomously control the inducer pump rate while attempting to achieve three consecutive set points. Set points were achieved with minimal overshoot or oscillation. The controller output is the sum of the proportional, integral and derivative terms. We plotted each of these terms over the course of the experiment to reveal how the controller was operating. Noise is primarily introduced by the derivative term as it amplifies any noise is the input signal. Noisy control did not impact the ability of the system to reach the desired fluorescence levels.

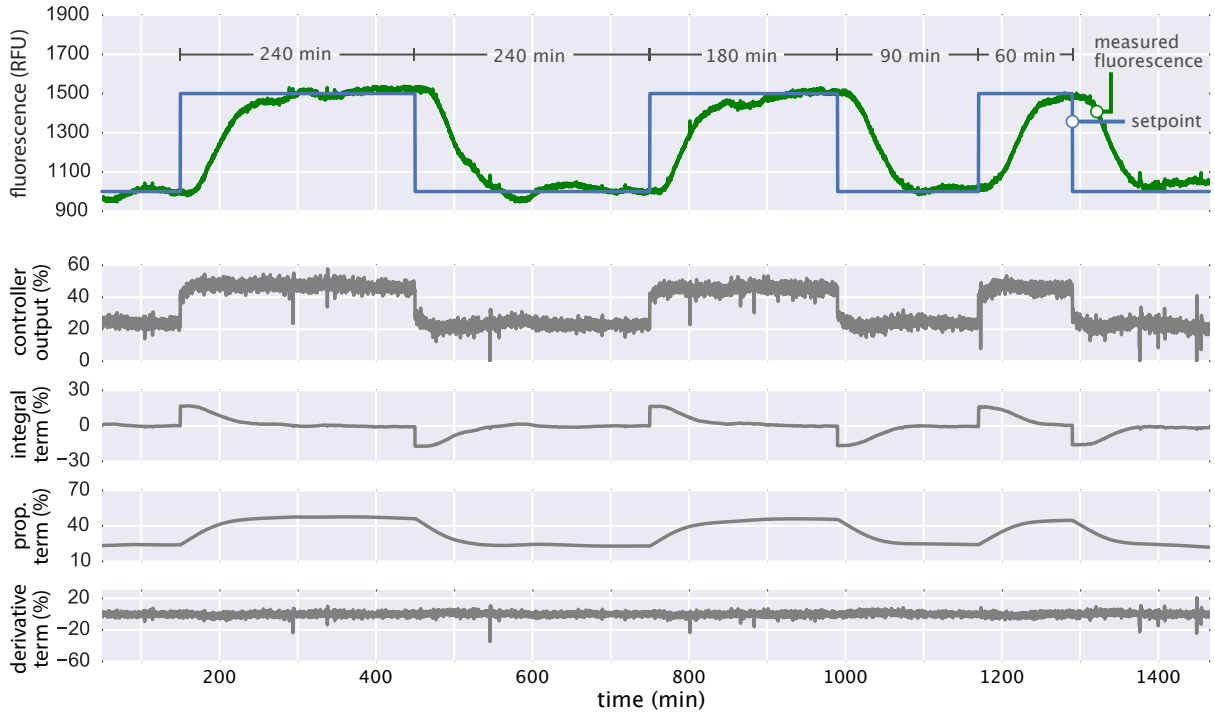


Figure 5.4: Closed-loop control of gene expression. Three consecutive fluorescence set points were evaluated with decreasing periods (first panel). The controller output during the experiment was noisy with the bulk of the values falling within 5% of the mean set point (second panel). Separating the controller output into the individual components

reveals the majority of noise comes from the derivative controller (panels three through five).

Going forward we will continue to support fluorimostat development with the aim of reducing the cost of implementation and increasing utility to researchers with additional functionality. The fluorimostat described here will complement other open source laboratory tools such as the recently released flexistat turbidostat(12) and soon to be released Wyss Evolvulator 2.0.

Materials and Methods

Fluorimostat Design and Operation

The fluorimostat is operated within a humidity and temperature controlled room maintained at 37°C. The growth chamber is fed by a media reservoir that is continuously pumped at a flow rate significantly greater than the growth rate of the cells. When cell density is below the desired set point, fresh media is diverted to the reservoir such that cells can grow without dilution. When cell density is above the desired set point, a valve directs media into the growth chamber such that the cellular density is decreased. A waste port simultaneously removes excess fluid to maintain a constant volume in the growth chamber. The density of cells within the growth chamber is monitored by the transmittance observed by a photodiode located across the growth chamber from a continuously illuminated LED. The growth chamber is located on a stir-plate and mixed by a magnetic stir-bar. Cells are continuously pumped from the growth chamber to the induction chamber. The induction chamber is operated at a smaller volume than the

growth chamber such that the desired induction chamber dilution rate will require a flow rate smaller than the effective dilution rate of the growth chamber, avoiding depletion of growth chamber working volume. The induction chamber receives a constant flow rate from the growth chamber and a variable, but negligible, flow rate from the inducer pump. The control software actuates the inducer pump to achieve either the desired inducer concentration or the desired fluorescence set point, depending on whether the fluorimostat is operating in closed- or open-loop mode. Cells are pumped from the induction chamber to the photometric flow-cell at a low flow rate to avoid generating bubbles. The volume of the induction chamber is maintained at a constant level by a high flow rate waste efflux at a specified fluid level within the chamber. The photometric flow cell consists of a cuvette with three quartz windows. Inlet and outlet fittings allow the induction chamber effluent to flow between the windows. High power LEDs are seated against two of the windows while a fiber-optic cable with collimated lens is seated against the third orthogonal window. The fiber-optic cable directs light to a spectrophotometer, which in turn feeds the observed spectrum back to the control software. Luer-lock fittings were used for all tubing connections. Any connections made within the warm-room were sanitized with 70% ethanol. A simple clean-in-place (CIP) procedure was used to avoid the need for device disassembly between experiments. To perform the CIP procedure, the media reservoir is disconnected and replaced with a 10% bleach solution. The bleach solution is pumped throughout the system and maintained for a residence time of at least 10 minutes. The media reservoir is reconnected and used to rinse remaining bleach out of the system before inoculation with cells. Approximately 1mL of cells grown to saturation are inoculated

through the aseptic waste septum of the growth chamber. Experiments are carried out once cells achieve steady-state growth.

Signal Processing

Induced cell culture was passed into a quartz cuvette flowcell (Starna 583.4-F) at a rate of 1.2 ml/min. Alternating GFP and RFP fluorescence measurements were taken by exciting the culture with a 470nm or 530nm high-powered LED (Thorlabs, Inc M530L3, M470F1), respectively. A BLACK-Comet UV-VIS spectrometer (StellarNet, Inc) was used to capture each emission spectrum with a 4000 ms integration time. The raw GFP and RFP fluorescence values were calculated by summing the amplitudes of the wavelengths from 500-520nm and 574-594nm, respectively. Raw fluorescence values are reported in the log file at 10-second intervals, and a smoothed fluorescence signal was generated as follows for closed-loop control:

$$S = \frac{S_0 \times 3 \ + \ F}{4}$$

S_0 is the previous smoothed fluorescence value, F is the current raw fluorescence value, and S is the calculated smoothed fluorescence value.

Chemicals and Reagents

All chemicals were purchased from Sigma unless otherwise specified. Erythromycin was purchased from Millipore (part #329815) and dissolved in 50% ethanol to 20mM. Davis media consisted of 7g/L dibasic potassium phosphate, 2g/L monobasic potassium phosphate, 1g/L ammonium sulfate, 0.1% tween 20, 4.5g/L glucose, 0.5g/L sodium citrate,

12.5mg/L magnesium sulfate, 5mg/L thiamine and 0.5g/L casamino acids. Appropriate antibiotics were added as necessary to maintain plasmids.

Strains and Plasmids

The previously described *mphR* biosensor was used in this study (pJKR-H-*mphR*). The *E. coli* strain DH5α was used for all experiments. Strains were inoculated from glycerol stocks and grown to saturation overnight in LB media supplemented with appropriate antibiotics. Approximately 1ml of saturated cells were inoculated into the growth chamber and allowed to reach an exponential state of growth prior to experimentation.

Inducible System Characterization

The *mphR* biosensor was characterized with the automated method described below. A strain of DH5α E. coli with the pJKR-H-*mphR* plasmid was maintained at an optical density of 0.18-0.20 within the growth chamber. A flow rate of 1.5ml/min was used to feed the induction chamber maintained at a volume of 45ml. A peristaltic pump with a maximum flow rate of 0.28ml/min dispensed inducer from a 20mM erythromycin reservoir. The cellscript module was used to assemble a protocol for the Labview software to follow. The protocol called for sequential inducer pump duty cycles of 1.3%, 2.7%, 5.4% and 11% with the entire sequence bracketed by long periods without pumping. The different duty cycles were maintained for 4 hours and corresponded to steady-state erythromycin concentrations of 50μM, 100μM, 200μM and 400μM. The Labview software logged the measured fluorescence and current inducer pump duty cycle every minute for

the duration of the experiment. The cellscript module read the log file and determined inducible system parameters using the algorithm described below.

First, concentration for each minute of the experiment was determined using the equation:

$$C = C_0 e^{-Dt} + \frac{s \cdot r}{f + r}(1 - e^{-Dt})$$

where C is the current concentration of inducer in the induction chamber, C_0 is the previous concentration of the inducer in the induction chamber, f is the flow rate from the growth chamber to the induction chamber, r is the current flow rate from the inducer reservoir to the induction chamber and D is the dilution rate of the induction chamber calculated as $D = f + r/\ v$ where v is the induction chamber volume. Next the fluorescence for each minute of the experiment was modeled using the equation:

$$F = F_0 e^{-Dt} + \left(V_{max} \frac{C^h}{C^h + K^h} + V_{min}\right) \frac{(1 - e^{-Dt})d}{D}$$

where F is the current fluorescence, F_0 is the previous fluorescence and V_{min}, V_{max}, K and h are the parameters of the hill function, which relates inducer concentration to gene expression rate. V_{min} is the gene expression rate in the absence of inducer, V_{max} is the maximum gene expression rate achievable with any amount of inducer, K is the half-maximal parameter and reflects the concentration of inducer that results in half-maximal induction and h is the hill coefficient, which indicates the cooperativity of the system. The dilution rate D is calculated as above while d captures the effect that increasing inducer flow rate has on cell density with $d = f/(f + r)$. The inducer flow rate is determined by multiplying the current duty cycle by the maximum inducer pump flow rate.

Finally, the model of fluorescence over time was used to determine the inducible system parameters. The SciPy function *minimize* was used to minimize the quadratic mean of the difference between the modeled fluorescence and the measured fluorescence over the duration of the experiment using sequential least squares programming(13). The entire process outlined above was carried out ten times on samples bootstrapped from the original dataset in order to determine confidence intervals for the fit parameters. The parameters estimates were used to plot the transfer function of inducer concentration to gene expression. The concentrations that were experimentally evaluated are superimposed on this plot such that the breadth of the fit can be visually inspected.

Closed Loop Control

Closed-loop control of induction levels was achieved through the implementation of a conventional parallel proportional-integral-derivative (PID) controller in Labview (National Instruments):

$$u(t) = K_p e(t) + K_i \int_0^t e(\tau)d\tau + K_d \frac{d}{dt} e(t)$$

The Cohen-Coon open-loop tuning method was used to derive initial PID gains from open-loop step function experimental data. Fine-tuning of the gain values was performed using plant simulation of the induction system, with model parameters derived from open-loop experimentation. For the purposes of the PID control, the fluorescence value (a.u.) was the process variable and the set point, on a scale of 0-100, represented the integer percentage of full inducer flow rate. The PID control loop was run one per ten seconds

during an experiment, and was provided with a smoothed input signal as described in "Signal processing."

References

1. Olson, E.J. and Tabor, J.J. (2014) Optogenetic characterization methods overcome key challenges in synthetic and systems biology. *Nature chemical biology*, **10**, 502-511.

2. Rogers, J.K. (2015) Synthetic biosensors for precise gene control and real-time monitoring of metabolites. *Nucleic acids research*.

3. Castillo-Hair, S.M., Igoshin, O.A. and Tabor, J.J. (2015) How to train your microbe: methods for dynamically characterizing gene networks. *Current opinion in microbiology*, **24**, 113-123.

4. Uhlendorf, J., Miermont, A., Delaveau, T., Charvin, G., Fages, F., Bottani, S., Batt, G. and Hersen, P. (2012) Long-term model predictive control of gene expression at the population and single-cell levels. *Proceedings of the National Academy of Sciences of the United States of America*, **109**, 14271-14276.

5. Menolascina, F., Fiore, G., Orabona, E., De Stefano, L., Ferry, M., Hasty, J., di Bernardo, M. and di Bernardo, D. (2014) In-vivo real-time control of protein expression from endogenous and synthetic gene networks. *PLoS computational biology*, **10**, e1003625.

6. Milias-Argeitis, A., Summers, S., Stewart-Ornstein, J., Zuleta, I., Pincus, D., El-Samad, H., Khammash, M. and Lygeros, J. (2011) In silico feedback for in vivo regulation of a gene expression circuit. *Nature biotechnology*, **29**, 1114-1116.

7. Olson, E.J., Hartsough, L.A., Landry, B.P., Shroff, R. and Tabor, J.J. (2014) Characterizing bacterial gene circuit dynamics with optically programmed gene expression signals. *Nat Methods*, **11**, 449-455.

8. Toettcher, J.E., Gong, D., Lim, W.A. and Weiner, O.D. (2011) Light-based feedback for controlling intracellular signaling dynamics. *Nat Methods*, **8**, 837-839.

9. Melendez, J., Patel, M., Oakes, B.L., Xu, P., Morton, P. and McClean, M.N. (2014) Real-time optogenetic control of intracellular protein concentration in microbial cell cultures. *Integrative biology : quantitative biosciences from nano to macro*, **6**, 366-372.

10. Drepper, T., Krauss, U., Meyer zu Berstenhorst, S., Pietruszka, J. and Jaeger, K.E. (2011) Lights on and action! Controlling microbial gene expression by light. *Applied microbiology and biotechnology*, **90**, 23-40.

11. Noguchi, N., Takada, K., Katayama, J., Emura, A. and Sasatsu, M. (2000) Regulation of transcription of the mph(A) gene for macrolide 2'-phosphotransferase I in Escherichia coli: characterization of the regulatory gene mphR(A). *J Bacteriol*, **182**, 5052-5058.

12. Takahashi, C.N., Miller, A.W., Ekness, F., Dunham, M.J. and Klavins, E. (2015) A low cost, customizable turbidostat for use in synthetic circuit characterization. *ACS synthetic biology*, **4**, 32-38.

13. Kraft, D. (1988). DLR German Aerospace Center - Institute for Flight Mechanics, Koln Germany, Vol. DFVLR-FB, pp. 28.

Chapter 6: Conclusion

This thesis advances our capabilities to control and evaluate biological systems. Allosteric transcriptional regulators were repurposed as synthetic biosensors that enable precise control of gene expression. The properties of these biosensors were characterized to enable their use in complex synthetic circuits. The biosensors were then redeployed in selections and screens that enable multiplexed phenotype evaluation for bio-based chemical production. Methods to modulate the parameters important for successfully finding highly productive cells were determined. Real-time observation of product formation was demonstrated for acrylate, muconate, glucarate and 3-hydroxypropionate. Multiplexed phenotype evaluation was used to optimize the production of naringenin and glucarate. Finally, truly programmable biology was enabled through the creation of the fluorimostat.

This thesis describes work completed over the last several years. Many of the ideas are still evolving and follow-up work is being carried out. The biosensors characterized in *E. coli* are being reengineered for use in mammalian cells. Demonstration of biosensor activity in cell-free systems has been completed, and work is underway to create a fully cell-free directed evolution platform that relies on emulsion sorting technology. The cell-free directed evolution system will be especially important for producing toxic products such as phenol or acrylate. The cell-free system also provides an ideal environment to evolve enzymes that are ultimately intended for operation outside of cells.

While we have used the fluorimostat to autonomously characterize inducible systems, it is also valuable in directed evolution and the automated optimization of

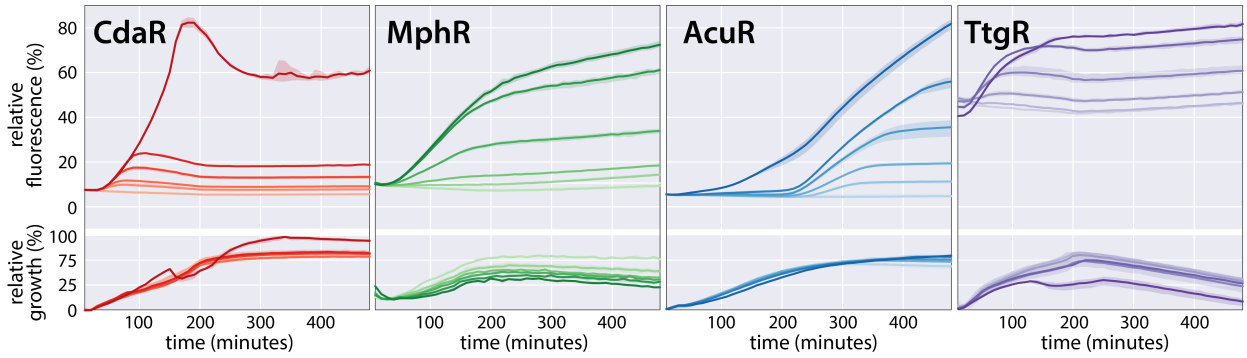
metabolic pathways. We are currently implementing a system where the fluorimostat monitors glucarate production *via* the glucarate biosensor and automatically modulates the timing and magnitude of heterologous enzyme expression such that product formation rates are maximized. We are also initiating collaborations in which the fluorimostat will be used to carry out continuous evolution experiments. To further worldwide adoption of the fluorimostat, we are integrating its development with the Wyss Evolvulator to make an even cheaper and more modular open source hardware platform. In this way, the fluorimostat will continue to evolve and a community will be fostered for further hardware and software development.

Going forward, I will continue promoting the use of multiplexed phenotype evaluation in the design-build-test cycle for metabolic engineering. Companies at the leading edge of metabolic engineering are already beginning to adopt multiplexed genome engineering and phenotype evaluation in their design cycles. As the ease and effectiveness of genome engineering progresses, the efficient evaluation of biological designs will become an increasingly important component of the biological engineering process.

Appendix A: Supplemental information for Chapter 2

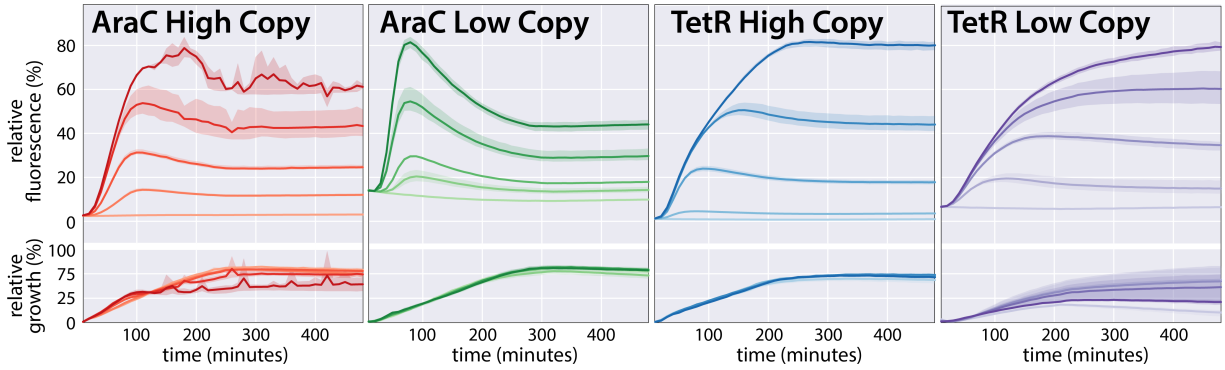
Supplemental Figures

Figure A.1: Fluorescence and growth kinetics for the low-copy implementations of the glucarate, erythromycin, acrylate and and naringenin biosensors.



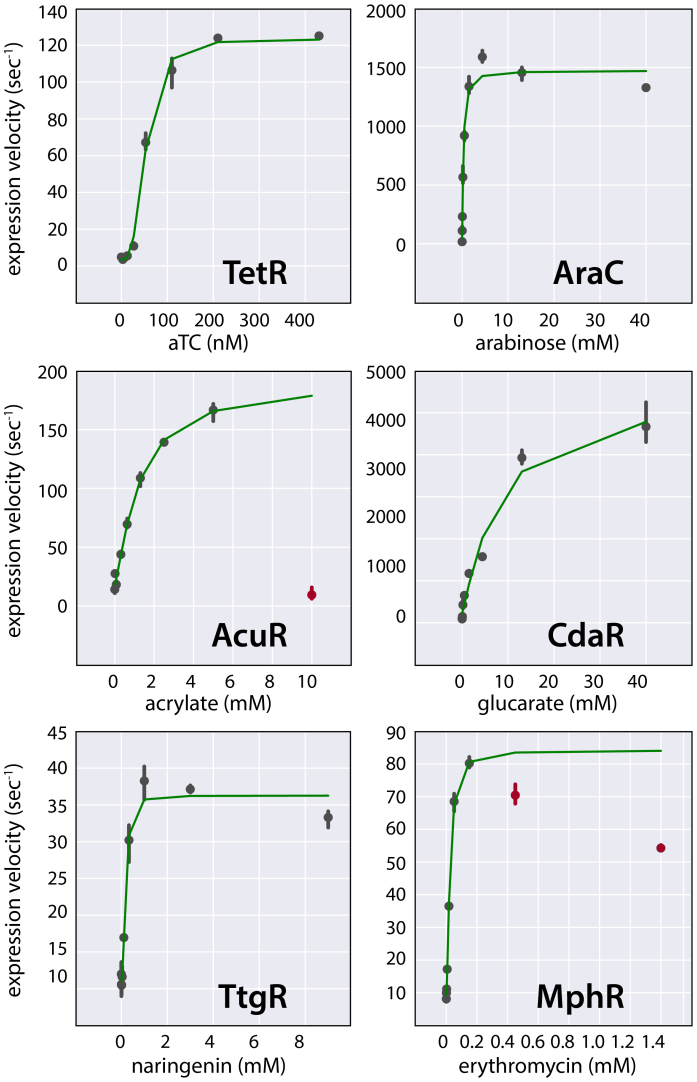
Induction and growth kinetics for the low-copy glucarate (CdaR), erythromycin (MphR), acrylate (AcuR) and naringenin (TtgR) biosensors. Chemical inducers are added at time zero and fluorescence is observed for eight hours. Lower panels show the optical density of the induced cultures over time. Induction levels are indicated by shade, with darker colors indicating higher inducer concentrations. Glucarate induction levels are 13mM, 4.4mM, 1.5mM, 0.49mM, 0.17mM and no inducer addition. Erythromycin induction levels are 150μM, 51μM, 17μM, 5.6μM, 1.9μM and no inducer addition. Acrylate induction levels are 5mM, 2.5mM, 1.3mM, 0.63mM, 0.31mM and no inducer addition. Naringenin induction levels are 9mM, 3mM, 0.33mM, 0.11mM, 0.037mM and no inducer addition. Fluorescence and optical density are normalized as described in the Methods. The standard error of the mean is represented with a 95% confidence interval (n=3).

Figure A.2: Fluorescence and growth kinetics for the arabinose and anhydrotetracycline (aTC) biosensors.



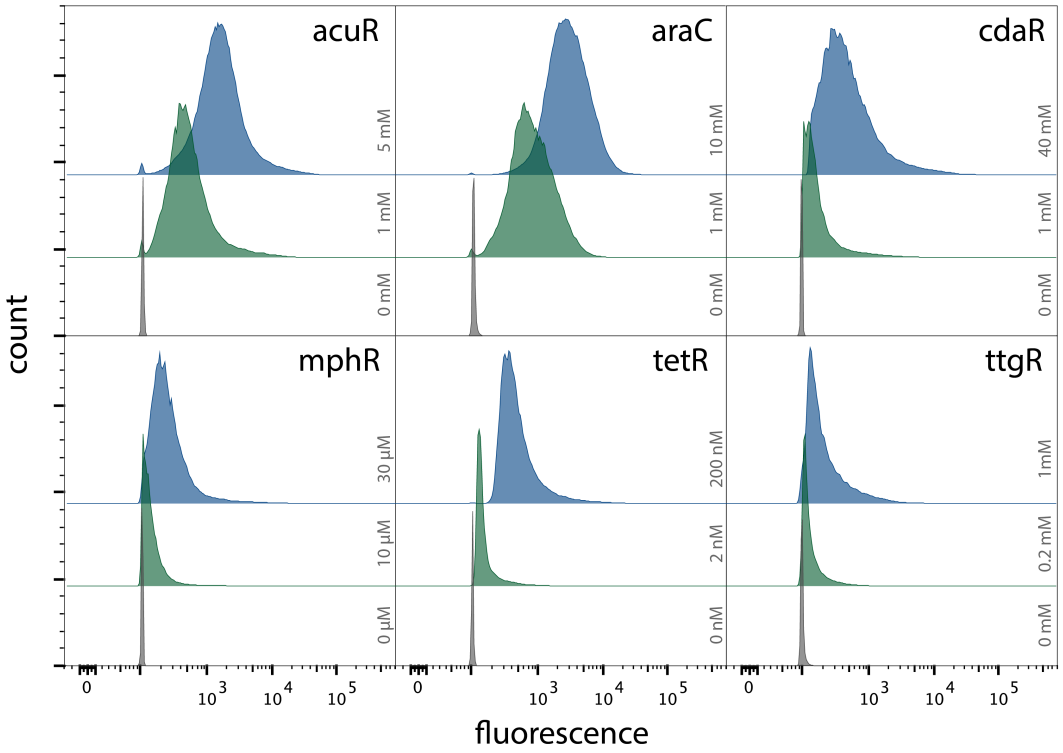
Induction and growth kinetics for the high- and low-copy arabinose (AraC) and anhydrotetracycline (TetR) biosensors. Chemical inducers are added at time zero and fluorescence is observed for eight hours. Lower panels show the optical density of the induced cultures over time. Induction levels are indicated by shade, with darker colors indicating higher inducer concentrations. Arabinose induction levels are 490μM, 170μM, 55μM, 18μM and no inducer addition. Anhydrotetracycline induction levels are 430nM, 210nM, 110nM, 53nM and no inducer addition. Fluorescence and optical density are normalized as described in the Methods. The standard error of the mean is represented with a 95% confidence interval (n=3).

Figure A.3: Promoter activities and model fits for the low-copy biosensors.



Low-copy promoter activity was fit to a model of inducible gene expression. The maximum expression velocity of each inducible promoter was determined at various levels of induction (points). The data was fit to a Hill function modified to account for basal and maximal promoter activity (green lines). The anhydrotetracycline (TetR) and naringenin (TtgR) biosensors show high induction cooperativity. The arabinose (AraC), glucarate (CdaR), acrylate (AcuR) and erythromycin (MphR) biosensors show low or moderate levels of cooperativity. The 10mM acrylate, 1400μM and 450μM erythromycin induction conditions were omitted from the modeling data due to high toxicity (red points). Error bars reflect the 95% confidence interval for the measured expression velocity.

Figure A.4: Single cell analysis of the low-copy biosensors.



The behavior of single cells in response to chemical induction was evaluated by flow cytometry. 100,000 cells from uninduced (grey), partially induced (green) and fully induced (blue) populations were observed for each low-copy biosensor. The specific concentration of inducer is indicated in the plot. Histograms are plotted with a biexponential scale to render the wide range of biosensor activation. The absence of large, well-separated bimodal distributions indicates that bulk fluorescent measurements do indeed reflect the induction behavior of individual cells.

Supplemental Tables

Table A.1: Sequence of regulator proteins and cognate promoter/operators.

Regulator	Promoter / Operator Sequence	Regulator Sequence
acuR	GCTTCACAACCGCACTTGATTTAATAGACCATACCG TCTATTATTTCTGG	ATGCCGCTGACCGACACCCCGCCGTCTGTTCCGCAGAAACCGCGTCGTGGTCGTCCGCGTGGTGCTCCGGACGCTTCTCTGG CTCACCAGTCTCTGATCCGTGCTGGTCTGGAACACCTGACCGAAAAAGGTTACTCTTCTGTTGGTGTTGACGAAATCCTGA AAGCTGCTCGTGTTCCGAAAGGTTCTTTCTACCACTACTTCCGTAACAAAGCTGACTTCGGTCTGGCTCTGATCGAAGCTT ACGACACCTACTTCGCTCGTCTCCTCGACCAGGCGTTCCTGGACGGTTCGCTGGCTCCGCTGGCTCGTCTGCGTCTGTTCAC CCGTATGGCTGAAGAAGGTATGGCTCGTCACGGTTTCCGTCGTGGTTGCCTGGTTGGTAACCTGGGTCAGGAAATGGGTG CTCTGCCGGACGACTTCCGTGCTGCTCTGATCGGTGTTCTGGAAACCTGGCAGCGTCGTACCGCTCAGCTGTTCCGTGAAG CTCAGGCTTGCGGTGAACTGTCTGCTGACCACGACCCGGACGCTCTGGCTGAAGCTTTCTGGATCGGTTGGGAAGGTGCTA TCCTGCGTGCTAAACTGGAACTGCGTCCGGACCCGCTGCACTCTTTCACCCGTACCTTCGGTCGTCACTTCGTTACCCGTAC CCAGGAATAA
araC	AGAAACCAATTGTCCATATTGCATCAGACATTGCC GTCACTGCGTCTTTTACTGGCTCTTCTCGCTAACCC AACCGGTAACCCCGCTTATTAAAAGCATTCTGTAA CAAAGCGGGACCAAAGCCATGACAAAAACGCGTAA CAAAAGTGTCTATAATCACGGCAGAAAAGTCCACA TTGATTATTTGCACGGCGTCACACTTTGCTATGCCA TAGCATTTTTATCCATAAGATTAGCGGATCCTACC TGACGCTTTTTATCGCAACTCTCTACTGTTTCTCCA TACCCGCTTTCATATCTTTCACTTTTTTTGGGCTAA C	ATGGCTGAAGCGCAAAATGATCCCCTGCTGCCGGGATACTCGTTTAACGCCCATCTGGTGGCGGGTTTAACGCCGATTGAG GCCAACGGTTATCTCGATTTTTTTATCGACCGACCGCTGGGAATGAAAGGTTATATTCTCAATCTCACCATTCGCGGTCAG GGGGTGGTGAAAAATCAGGGACGAGAATTTGTCTGCCGACCGGGTGATATTTTGCTGTTCCCGCCAGGAGAGATTCATCA CTACGGTCGTCATCCGGAGGCTCGCGAATGGTATCACCAGTGGGTTTACTTTCGTCCGCGCGCCTACTGGCATGAATGGCT TAACTGGCCGTCAATATTTGCCAATACGGGTTTCTTTCGCCCGGATGAAGCGCACCAGCCGCATTTCAGCGACCTGTTTGG GCAAATCATTAACGCCGGGCAAGGGGAAGGGCGCTATTCGGAGCTGCTGGCGATAAATCTGCTTGAGCAATTGTTACTGC GGCGCATGGAAGCGATTAACGAGTCGCTCCATCCACCGATGGATAATCGGGTACGCGAGGCTTGTCAGTACATCAGCGATC ACCTGGCAGACAGCAATTTTGATATCGCCAGCGTCGCACAGCATGTTTGCTTGTCGCCGTCGCGTCTGTCACATCTTTTCC GCCAGCAGTTAGGGATTAGCGTCTTAAGCTGGCGCGAGGACCAACGCATTAGTCAGGCGAAGCTGCTTTTGAGCACTACCC GGATGCCTATCGCCACCGTCGGTCGCAATGTTGGTTTTGACGATCAACTCTATTTCTCGCGAGTATTTAAAAAATGCACCG GGGCCAGCCCGAGCGAGTTTCGTGCCGGTTGTGAAGAAAAAGTGAATGATGTAGCCGTCAAGTTGTCAtaa
cdaR	ATGCTGTTGATTGACGCCAGTGAGAACCCGGAACC GGAAACGGAATCAAATCCGTGGGTCGAACAGTGGG GCACGCTGTTGTCCTGATATGTTCAGCGAGCGGTA AATGTCGTTTTAGCGGTGCTGAATCGAATCTTTTC AGGCAAATGCCAGTAAAAACTGCTTCATAGCGCGG ATTTTTACTGGCGTTTGCCTGGAGTCAAGCGATCC ATTTCATACTCTTCTTTATTTCTTCGTTTTAACCCT TCCTTTCTTGTTCTTGTTTTCATTTCCGTGAAGTGG ATTCCACCGTCCAGGGCTAATGCCAAAATCGGGCCT CATTGAACGCATTAATGTTGTGTTGTTGCACGGTG AGCCGCTATGGCGCGCTTTTTATACTGCTATTGCCA GATATAAACACGCGCCGTATTCGGCGAACGACCTA TAAAAACGGCAAAAAACACCCTACGTCACCTCTGA TTTCCTGGCGATGTCGCAGTCCAGAGTGAGCGTGG CTAACGCGAATTTTCAGGAGTGCAACA	ATGGCTGGCTGGCATCTTGATACCAAAATGGCGCAGGATATCGTGGCACGTACCATGCGCATCATCGATACCAATATCAAC GTAATGGATGCCCGTGGGCGAATTATCGGCAGCGGCGATCGTGAGCGTATTGGTGAATTGCACGAAGGTGCATTGCTGGT ACTTTCACAGGGACGAGTCGTCGATATCGATGACGCGGTAGCACGTCATCTGCACGGTGTGCGGCAGGGGATTAATCTACC GTTACGGCTGGAAGGTGAAATTGTCGGCGTAATTGGCCTGACAGGTGAACCAGAGAATCTGCGTAAATATGGCGAACTGG TCTGCATGACGGCTGAAATGATGCTGGAACAGTCGCGGTTGATGCACTTGTTGGCGCAGGATAGCCGTTTGCGGGAAGAA CTGGTGATGAACCTGATTCAGGCAGAGGAGAATACTCCCGCACTTACTGAATGGGCGCAACGGCTGGGGATCGATCTCAA TCAACCGCGAGTGGTGGCTATTGTTGAGGTCGACAGCGGTCAGCTTGGCGTGGACAGCGCAATGGCGGAGTTACAACAAC TGCAAAACGCGCTGACTACGCCCGAGCGTAATAATCTGGTGGCGATTGTCTCGCTAACCGAAATGGTGGTGTTGAAACCG GCGTTGAACTCTTTTGGGCGCTGGGATGCAGAAGATCATCGTAAGCGAGTTGAACAACTGATTACCCGCATGAAAGAGTA CGGCCAGCTGCGTTTTCGCGTTTCACTGGGCAACTATTTTACCGGTCCTGGCAGTATTGCCCGATCCTATCGTACGGCGAA AACGACGATGGTGGTGGGTAAACAGCGGATGCCAGAAAGTCGCTGCTATTTTTATCAGGATCTGATGTTACCTGTGTTAC TCGACAGTTTGCGTGGCGACTGGCAGGCCAACGAACTGGCGCGACCGCTGGCGCGGCTGAAAACGATGGACAATAACGGCT TGCTGCGACGAACGCTGGCGGCGTGGTTTCGCCACAATGTGCAACCGCTGGCAACGTCAAAGGCGTTGTTTATTCATCGTA ATACCCTGGAGTATCGGCTTAATCGTATATCGGAACTGACCGGGCTTGATTTGGGCAATTTTGATGACAGGTTGCTGCTG TATGTGGCGTTACAACTGGATGAAGAGCGGtag
mphR	GGATTGAATATAACCGACGTGACTGTTACATTTAG GTGGCTAAACCCGTCAA	ATGCCGCGTCCGAAACTGAAATCTGACGACGAAGTTCTGGAAGCGGCGACCGTTGTTCTGAAACGTTGCGGTCCGATCGAA TTCACCCTGTCTGGTGTTGCGAAAGAAGTTGGTCTGTCTCGTGCGGCGCTGATCCAGCGTTTCACCAACCGTGACACCCTG CTGGTTCGTATGATGGAACGTGGTGTTGAACAGGTTCGTCACTACCTGAACGCGATCCCGATCGGTGCGGGTCCGCAGGGT CTGTGGGAATTCCTGCAGGTTCTGGTTCGTTCTATGAACACCCGTAACGACTTCTCTGTTAACTACCTGATCTCTTGGTAC GAACTGCAGGTTCCGGAACTGCGTACCCTGGCGATCCAGCGTAACCGTGCGGTTGTTGAAGGTATCCGTAAACGTCTGCCG CCGGGTGCGCCGGCGGCGGCGGAACTGCTGCTGCACTCTGTTATCGCGGGTGCGACCATGCAGTGGGCGGTTGACCCGGAC GGTGAACTGGCGGACCACGTTCTGGCGCAGATCGCGGCGATCCTGTGCCTGATGTTCCCGGAACACGACGACTTCCAGCTG CTGCAGGCGCACGCGTAA
tetR	TCGAGTCCCTATCAGTGATAGAGATTGACATCCCT ATCAGTGATAGAGATACTGAGCACATCAGCAGGAC GCACTGACCGAATTCATTAAA	ATGTCTCGTTTAGATAAAAGTAAAGTGATTAACAGCGCATTAGAGCTGCTTAATGAGGTCGGAATCGAAGGTTTAACAAC CCGTAAACTCGCCCAGAAGCTAGGTGTAGAGCAGCCTACATTGTATTGGCATGTAAAAAATAAGCGGGCTTTGCTCGACG CCTTAGCCATTGAGATGTTAGATAGGCACCATACTCACTTTTGCCCTTTAGAAGGGGAAAGCTGGCAAGATTTTTTACGT AATAACGCTAAAAGTTTTAGATGTGCTTTACTAAGTCATCGCGATGGAGCAAAAGTACATTTAGGTACACGGCCTACAGA AAAACAGTATGAAACTCTCGAAAATCAATTAGCCTTTTTATGCCAACAAGGTTTTTCACTAGAGAATGCATTATATGCAC TCAGCGCAGTGGGGCATTTTACTTTAGGTTGCGTATTGGAAGATCAAGAGCATCAAGTCGCTAAAGAAGAAAGGGAAACA CCTACTACTGATAGTATGCCGCCATTATTACGACAAGCTATCGAATTATTTGATCACCAAGGTGCAGAGCCAGCCTTCTTA TTCGGCCTTGAATTGATCATATGCGGATTAGAAAAACAACTTAAATGTGAAAGTGGGTCTTAA
ttgR	CACCCAGCAGTATTTACAAACAACCATGAATGTAA GTATATTCCTTAGCAA	ATGGTGCGTCGCACCAAAGAAGAAGCACAGGAAACGCGTGCGCAGATTATCGAAGCGGCCGAACGCGCGTTTTATAAACG TGGTGTGGCACGTACCACGCTGGCAGATATTGCAGAACTGGCAGGTGTTACCCGCGGTGCAATCTACTGGCATTTCAACAA TAAAGCCGAACTGGTTCAGGCACTGCTGGATTCTCTGCACGAAACGCATGATCACCTGGCCCGTGCAAGCGAATCTGAAGA TGAACTGGACCCGCTGGGCTGCATGCGCAAACTGCTGCTGCAGGTGTTTAACGAACTGGTTCTGGATGCACGTACCCGTCG CATTAATGAAATCCTGCATCACAAATGCGAATTTACGGATGATATGTGTGAAATTCGTCAGCAGCGCCAGAGCGCCGTGC TGGATTGTCATAAAGGTATCACCCTGGCACTGGCAAACGCAGTTCGTCGCGGTCAGCTGCCGGGTGAACTGGATGTGGAA CGCGCAGCGGTTGCGATGTTTGCCTATGTGGATGGCCTGATTGGTCGTTGGCTGCTGCTGCCGGATAGTGTTGATCTGCTG GGCGATGTGGAAAAATGGGTTGATACCGGTCTGGATATGCTGCGTCTGAGCCCGGCGCTGCGCAAATAA

Table A.2: Sequence of MIOX orthologs evaluated in this study.

MIOX Variant	Sequence
Candida albicans	ATGGTAAACAAGGTCGGTAAATCTACTCTCGATAAGAGCACAAACCTAGATAAATCCAAAGGGAATATA TTAGAGAAACTAGATGATGATATACTTCATGTCAATAGAATTCGAGGCTCTTTAACTAACAAAACTCCA ATCACCAAAACCCATTCGATAGATGATGAGCTTAAACTAGAAGAACAATCAGAAACTGCCGCCGATGAA AATTGGCAAATAGCATCGGAATATTATAAAAACATAGACACGAAGGCTTTCCGCCAATATGAATTAGCT TGTGATAGAGTCAAACAGTTTTATGAAGAACAACATGAAAAACAAACCGTGGCGTATAATATTCAAGCA AGAATTAATTTCAAAACTAAAACAAGAGCAAGAATGACAGTTTGGGAAGGACTAGAGAAATTAAACAAA TTGTTAGATGATTCTGATCCCGACACCGAATTGTCACAAATAGATCATGCATTACAGACGGCAGAAGCTA TACGGCGAGATGGGAAACCACGATGGTTTCAATTAGTTGGGTTGATTCATGATTTAGGGAAATTACTAT ATTTTTTTGATTCTCGTGGTCAATGGGATGTAGTGGGTGATACTTTCCCTGTTGGTTGTAAATTCCTGAA ACGGATTATTTTCCCTGATAGTTTTAAAAATAATCCAGATTTCCTAAATCCATTGTATAATACCAAATAT GGCATATATTCAAAACATTGTGGATTAGATAAAGTCATGTTGAGTTGGGGTCATGATGAGTATATGTAT CATGTTGCGAAAAAGAATTCGACATTACCACCGGAAGCATTGGCAATGATAAGGTATCATTCATTTTATC CTTGGCATCAAGAATTGGCATATAGTTATTTAATGGATGAGCATGATAAAGAGATGTTGAAAGCAGTCA AAGCTTTCAATTCCTATGATTTATATTCCAAGATAGATCAACAGTATGATGTTGAAGAGTTGAAACCAT ATTACCTAGAGTTGATTGATGAGTTTTTCCCAAATAAAGTAATTGATTTTTAA
Francisella sp. TX077308	ATGAGTCAGACCGTGGAAAACACGTTTGGCGAATTTCGTAACTACACCGATAGCAAATTCCAGGATCGTG TGGAACGCACGTACAAAGATATGCACATTAACCAGAATCTGGAATACGTTACCCAGATGAAAGATAAAT ACTTCAAACTGGATCTGGGTAAAATGGATGTGTACGAAGTTTTCAAACTGCTGGAAAACGTTCATGATG AAAGCGATCCGGATAATGATCTGCCGCAGATCGAACACGCATATCAGACCGCGGAAGCCTGCCAGAACAA ATTCCTGAAATCTGATACGGAACTGCGCGAAAATGCGCTGATTCGTAGTATCTTTCGCGATCATGAATGG CAGAGCATTCCGAAAATCTGGCAGGATTTCTATACCAAAAAACAGAGTCTGGGCAATCTGTACAGCCATA TTAAAGATTGGTCTTGGTTTCCGCTGGTTGGCTTCGTTCACGATCTGGGTAAAATCATGACCCTGCCGGA ATATGGTCAGCTGCCGCAGTGGAGCACCGTGGGTGATACGTACCCGATTGCCTGCCCGTTTGCAAGCGCG AACGTGTTTTCTCACCGTGAATTTGTTAAAGATTCTAAAGATTACAACAATTACAATACCGAAAGTGAA ACGCATTATGGCAAATACGAGAAAAAATGTGGTTTCGATAACGTGGATATGAGCTTCGGTCACGATGAA TACATCTACAAAGTTTTCGAACAGGGCAGCGATATCCCGTATGAAGGTCTGTACCTGCTGCGCTATCATT CTTTCTACCCGTGGCACACCCCGCAGACGGGCGGTCATGCGTATCAGGAACTGGCCAACGAAAAAGATTG GCTGCTGCTGCCGCTGCTGAAAGCCTTTCAGAAAGCGGATCTGTATTCTAAACTGCCGGAACTGCCGCCG AAAGAAGTGCTGGAGAAAAAATACAAAAGTCTGCTGGATAAATGGGTTCCGAACAAGAAAATTAACTGG TAA
Flavobacterium johnsoniae	ATGAAAAAGCATATAGACACAGACAATCCGTTGAAAAATTTAGATGAGTGGGAAGATGATTTGTTAATG CGATATCCTGACCCTTCTGAAGTAAATGAAAGTTTAAAAGAAAAGCAGAAAGAAGAATTTAGAAATTAT GTCGATTCTGAAAGAGTAGAAACGGTAAAAGAATTTTACAGGATAAACCATACCTACCAAACTTATGAC TTTGTATGCAGTAAAGAACAAGAATTTCTGCAATTTAATAGAAAAGAAATGTCAATCTGGGAAGCTGTC GAGTTTTTAAACACGCTTGTAGACGACAGTGACCCAGATATTGACTTAGACCAGACACAGCACCTTTTAC AGACTTCAGAAGCCATTCGTGCTGATGGTCATCCGGATTGGTTTGTACTGACAGGTTTCATTCACGATTT GGGTAAAGTTTTATGCTTATTTGGAGAACCGCAATGGGCAGTCGTTGGCGATACTTTTCCGGTTGGCTGT GCGTATTCGGATAAAATTGTGTATTCAGAATTTTTTAAAGAAAATCCGGATTATACAGATGAGAGATTC AATACTAAACTAGGAATCTACACTGAAAACTGCGGATTAGATAACGTAAAAATGAGCTGGGGTCATGAC GAATATTTGTATCAGATTATGAAAGATTATTTACCGGATCCTGCTTTATACATGATTCGTTATCACTCTT TTTATTCGCAGCATAAAGAAAATGCGTATGCACATTTAATGAATGAAAAAGACATCGAAATGTTTGACT GGGTTCGAAAATTCAATCCGTACGATTTGTATACAAAGGCTCCTGTAAAACCAGATGTTCAGGCATTACT TCCTTATTATAAAGAATTAGTTGCTAAATATTTGCCTGAAAAATTGAAGTTTTAA
Mus musculus	ATGAAAGTGGATGTTGGCCCGGACCCGAGCCTGGTTTACCGCCCGGATGTGGACCCGGAAATGGCAAAAA GCAAAGATTCGTTTCGTAACTACACCAGTGGCCCGCTGCTGGATCGTGTTTTTACCACGTATAAACTGAT GCATACCCACCAGACGGTTGACTTTGTCAGCCGTAAACGCATTCAATATGGCGGTTTCTCTTACAAGAAA ATGACCATCATGGAAGCGGTGGGCATGCTGGATGACCTGGTTGATGAATCAGATCCGGACGTCGATTTTC CGAATTCGTTTCATGCGTTCCAGACGGCCGAAGGTATTCGCAAAGCCCACCCGGACAAAGATTGGTTCCA TCTGGTCGGCCTGCTGCACGATCTGGGTAAAATCATGGCACTGTGGGGTGAACCGCAGTGGGCTGTGGTT GGTGATACCTTTCCGGTGGGTTGCCGTCCGCAAGCAAGTGTCGTGTTTTGTGACTCCACCTTCCAGGACA ACCCGGATCTGCAAGACCCGCGCTATTCAACGGAACTGGGCATGTACCAGCCGCATTGCGGTCTGGAAAA CGTGCTGATGTCGTGGGGTCACGATGAATACCTGTACCAGATGATGAAATTCAACAAATTCAGCCTGCCG TCTGAAGCCTTCTACATGATCCGTTTCCATAGTTTCTACCCGTGGCACACCGGCGGTGATTATCGCCAGC TGTGCTCCCAGCAAGACCTGGATATGCTGCCGTGGGTGCAAGAATTCAACAAATTCGATCTGTACACGAA ATGTCCGGATCTGCCGGACGTTGAATCTCTGCGTCCGTACTACCAAGGTCTGATTGATAAATACTGTCCG GGCACCCTGTCGTGGTAA

Table A.3: Inducer toxicity.

	Inducer Concentration							
acrylate(μM)	0	156	313	625	1250	2500	5000	10000
arabinose (μM)	0	55	165	494	1481	4444	13333	40000
aTC (nM)	0	6.7	13.0	27	53	110	210	430
DMSO (%)	0	0.0069	0.021	0.062	0.19	0.56	1.7	5
erythromycin (no eryR, μM)	0	1.9	5.6	17	51	150	450	1400
erythromycin (μM)	0	1.9	5.6	17	51	150	450	1400
ethanol (%)	0	0.0027	0.0082	0.025	0.074	0.22	0.7	2
glucarate (μM)	0	55	165	494	1481	4444	13333	40000
naringenin (μM)	0	12	37	111	333	1000	3000	9000
	Growth Rate (hr^{-1})							
acrylate	0.73	0.75	0.74	0.73	0.70	0.50	0.27	0.10
arabinose	0.78	0.76	0.80	0.86	0.90	0.92	0.92	0.95
aTC	0.74	0.75	0.74	0.75	0.75	0.70	0.70	0.54
DMSO	0.73	0.74	0.74	0.74	0.75	0.71	0.66	0.55
erythromycin (no eryR)	0.68	0.68	0.67	0.69	0.68	0.65	0.61	0.52
erythromycin	0.67	0.67	0.67	0.58	0.48	0.29	0.13	0.11
ethanol	0.70	0.75	0.75	0.76	0.75	0.71	0.66	0.52
glucarate	0.74	0.74	0.74	0.74	0.74	0.74	0.72	0.76
naringenin	0.69	0.73	0.72	0.72	0.68	0.53	0.40	0.16

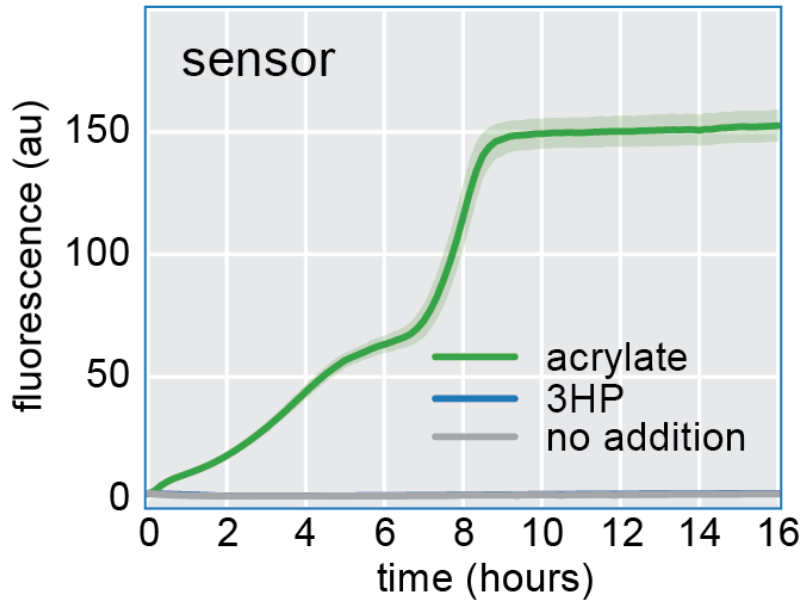
Table A.4: Inducer cross-reactivity (growth-normalized fluorescence)

	TtgR	TetR	CdaR	AcuR	AraC	MphR	control
erythromycin	11	9	14	11	25	1063	8
arabinose	8	8	11	10	1609	8	6
acrylate	9	8	9	485	27	10	6
glucarate	7	7	236	10	27	9	7
aTC	8	152	10	10	24	10	8
naringenin	111	7	9	8	24	8	7
IPTG	8	7	9	10	22	10	8
rhamnose	8	8	10	11	25	10	8
cumate	8	8	9	9	22	9	8
DMSO	8	8	9	9	26	10	8
ethanol	8	6	8	8	30	8	6
water	8	6	8	7	24	7	6

Appendix B: Supplemental Information for Chapter 3

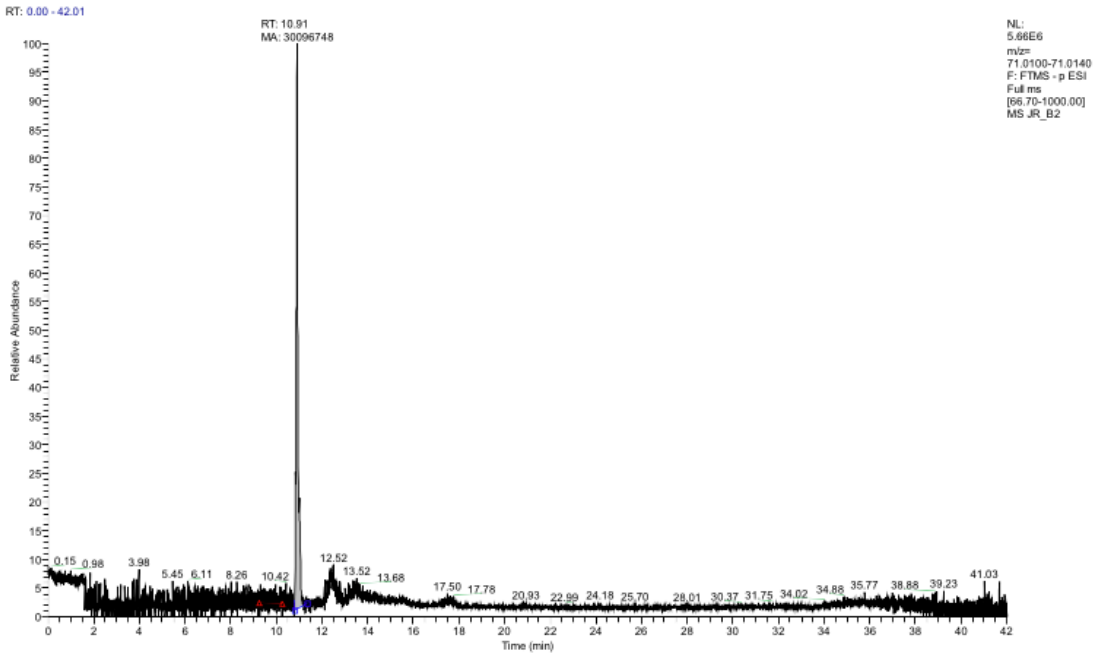
Supplemental Figures

Figure B.1: Cells without the acuR-based 3HP biosensor helper plasmids do not produce fluorescence in the presence of 3HP.



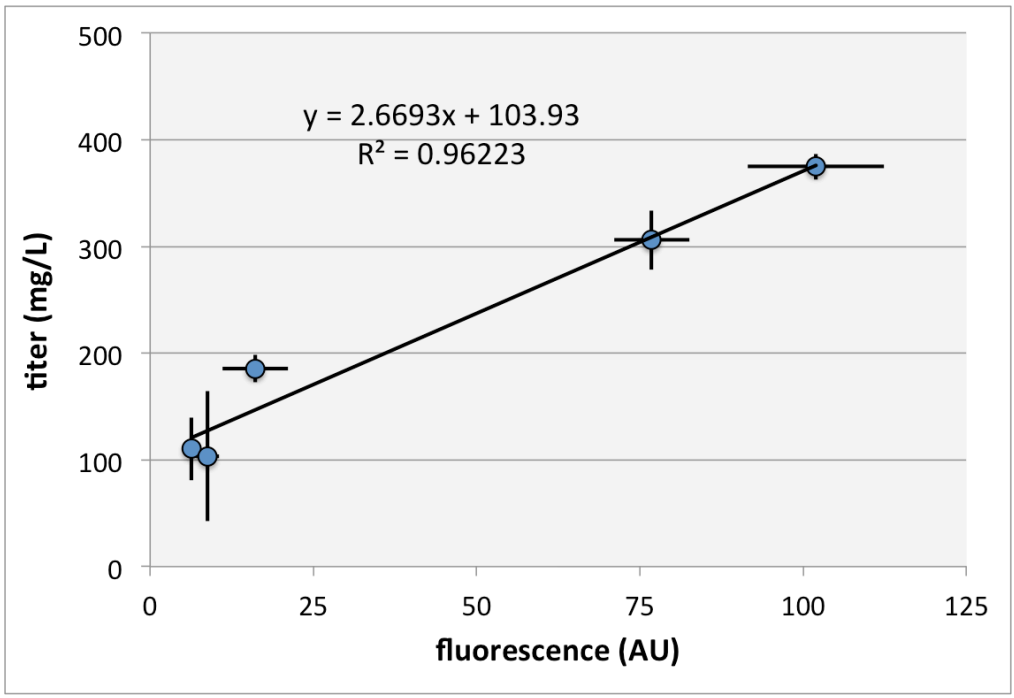
The fluorescent response to the addition of acrylate indicates that the biosensor is functional, but insulated against activation by 3HP without the enzymes necessary to convert 3HP to acrylate. The confidence band represents the 95% confidence interval determined by the standard error of the mean (n=3).

Figure B.2: Representative chromatogram for acrylate measurement by LC-MS.



The samples were analyzed on a Thermo q-Exactive Plus mass spectrometer equipped with a Thermo 3000 Ultimate uHPLC. A resolution of 70,000 was used on the mass spectrometer. A hydrophilic interaction chromatography method was used using a EMD Sequant pHILIC column (150 mm length, 2.1 mm ID, 5um particle size) at a flow rate of 100 uL /minute . Mobile phase A was 20 mM ammonium carbonate, and B was acetonitrile. The gradient started at 100% B and linearly decreased to 40% B over 20 minutes. B was then decreased further to 20% over 10 minutes, and then returned to initial conditions at 100% B over 0.1 minutes and maintained for the next 11.9 minutes to equilibrate the column for the following run.

Figure B.3: Glucarate titer as a function of biosensor fluorescent response.



Glucarate titer as a function of biosensor fluorescent response. Higher fluorescent responses indicate higher glucarate production titers. Error bars represent the 95% confidence interval derived from the standard error of the mean (n=3).

Appendix C: Supplemental Information for Chapter 4

Supplemental Methods

Riboswitch Sensors

Because riboswitches affect expression control through translation of the transcript rather than transcription, a modified sensor architecture was used. The theophylline-responsive theoRR blocks translation unless theophylline is present (23). We included the theoRR as a 5' untranslated region of the *tolC* transcript, allowing for translation only when theophylline was present.

The *btuB* vitamin B12-responsive riboswitch operates in an opposite manner to theoRR, remaining in an open conformation natively, and attenuating translation through RBS occlusion only when B12 or its derivatives are present (17). We included the *btuB* riboswitch and the first 70 codons of *btuB* gene sequence (17) at the beginning of the tetR transcript in a modified tetR-CAT sensor-selector. When the btuB-tetR fusion is transcribed in the absence of vitamin B12, the riboswitch structure within the btuB 5'-UTR is in its open conformation, allowing translation of the btuB-tetR fusion protein, which represses the transcription of CAT, leaving the cells chloramphenicol sensitive. In the presence of vitamin B12, the riboswitch changes conformation to bind B12 and occludes the RBS, preventing btuB-tetR translation and allowing strong transcription of CAT, leading to chloramphenicol resistance. Using this strategy, the translation-attenuating B12-responsive btuB riboswitch was used to control the expression of a positive selection marker.

Sensor-Selector Strain Construction

To construct the TtgR-TolC dual sensor strain, a linear PCR product composed of 5'- zeocin resistance cassette—apFAB101 promoter(48)—RBS B0034 (AAAGAGGAGAAATTA)—*ttgR*-3' was amplified with 50 bp homology and recombineered into the genome at locus 1529620 (numbered relative to the MG1655 sequence) of the TtgR-TolC sensor-selector strain. (For pathway optimization, the pLtetO promoter in the *ttgR* sensor gene construct at the *tolC* locus was also replaced with apFAB101 promoter to avoid homology with pLtetO promoters on naringenin pathway plasmid 2. To construct TtgR strain with dual selectors, a linear PCR product composed of 5'- spectinomycin resistance cassette—*ttgA*p promoter -3' was amplified with 50 bp homology and recombineered 5' to the CAT gene at the Δ*mutS::Cm* locus of the into the TtgR-TolC sensor-selector strain.

Riboswitch-based sensor-selectors necessitated modifications to the standard construction used for all transcription factor-based sensor-selectors. The theophylline-responsive riboswitch (theoRR) sensor used promoter pLtetO to directly transcribe *tolC*, with the theoRR appended 5' to the *tolC* start codon as a 5' untranslated region. The *btuB* riboswitch invertor was a modified version of the TetR-CAT sensor-selector. The *tetR* gene ATG start codon was replaced by the 239

bp of the *E. coli btuB* 5'-untranslated region, and the first 210 nucleotides of the *btuB* coding sequence(17).

Sensor-selector strains CdaR-TolC and dual TtgR-TolC were modified before pathway optimization to include T7 RNA polymerase under control of a pLac promoter. A linear PCR product containing 5'- spectinomycin resistance cassette reverse complement—pLac—T7 RNA polymerase gene - 3' was amplified with 50 bp homology and recombineered into the genome to replace the *bla* gene within the Red-λ prophage.

Escape Rate Measurements

Sensor-selector escape rates were measured by growing cells overnight to saturation in LB medium. Serial dilutions of cells were plated onto non-selective plates to measure the saturation culture density (colony forming units/mL). Serial dilutions of cells were plated onto appropriate selective plates (LB supplemented with SDS or chloramphenicol) to measure the density of cells (CFU/mL) surviving selection in the absence of chemical inducers. The density of cells surviving selection was divided by the total saturation density to calculate the escape rate of each sensor-selector.

TtgR-TolC Sensor-Selector Degradation Tags Modification

We appended three *ssrA* degradation signal variants of varying strengths(28) to the 3' end of the *tolC* selector gene coding sequence in the ttgR-tolC sensor-selector strain. The ssrA variants were inserted in frame to replace the stop codon of the selector by recombineering, using a zeocin resistance cassette as a selection marker for integration. The following degradation tags were appended to the selector gene in frame: strong (RPAANDENYALAA*), medium (RPAANDENYAAAV*), and weak (RPAANDENYALVA Each strain was sequence verified. We determined the escape rate and operational range for each degradation tag using orthogonal gradient growth assays.

TtgR-TolC Sensor-Selector RBS Modification

We generated ten RBS variants by increasing or decreasing the separation between the Shine Dalgarno sequence of promoter *ttgA*p (5'- CCGAGGATCCTC -3') and *tolC* translation start site by 1 to 5 bases (SI Appendix, Fig. S6). The underlined bases remained unmutated. We designed oligonucleotides for each RBS variant and used MAGE to modify the TtgR-TolC sensor-selector strain. We verified the ten variants by sequencing colonies after MAGE, and determined the escape rate and operational range for each of the ten RBS variants using orthogonal gradient growth assays. We measured operational range and escape rates of the five RBS variants that showed ligand induction response (Fig 2B, SI Appendix, Fig S6).

TetA Exporter Plasmid Construction and Assay Conditions

Plasmid pKD46(54) (Genbank accession AY048746) was modified, replacing lambda red *exo*, *beta* and *gam* genes with the *Streptococcus* tetracycline efflux pump (*tetA*) gene, and also replacing beta lactamase (*bla*) gene with spectinomycin resistance gene to create plasmid pBAD-tetA. The *tetA* gene is transcriptionally controlled by the arabinose-inducible pBAD promoter. Plasmid pBAD-*tetA* was transformed into TetR-CAT sensor-selector strain. This strain was characterized using orthogonal gradient growth assays under the following *tetA* expression induction conditions: no arabinose, 0.05 % arabinose (intermediate), or 0.5 % arabinose with 4 hr pre-induction (high).

Orthogonal Gradient Growth Assay

Arrays of growth conditions were evaluated in 96-well plates (BD-Falcon). Serial dilution of selection agent (SDS, chloramphenicol or colicin E1) along one axis of the 96-well plate was followed by serial dilution of an inducer chemical along the second axis to create the arrays in 150 µl of kanamycin supplemented LB media. Overnight cultures picked from fresh colonies were inoculated at a dilution factor of at least 1000 into each well. Plates were incubated at 30 °C with agitation and measurement of optical density at 600 nm (OD600) for at least 16 hours in a Biotek (Winooski, VT) plate reader.

Chemical detection concentration and saturation concentration thresholds were found from these assays by determining the minimum chemical concentration at which cells showed a growth response when compared to no chemical (detection threshold) and the concentration at which higher chemical concentration showed no additional growth benefit (saturation threshold). Data from single measurements are reported in Figure 2A.

Glucaric Acid Production

Overnight cultures of cells were diluted 1:100 into selective LB supplemented with 50mM glucose and 1mM IPTG. 2ml deep 96-well blocks with a working volume of 1ml were incubated humidified at 37C and 900RPM. After 72 hours the plates were spun down for 5 minutes at 4000rpm and the supernatants were filtered prior to analysis.

Naringenin Production

Cell were grown in M9 minimal medium supplemented with 1 % glucose, 1 mM biotin, kanamycin and carbenicillin for 24 hours at 30 °C in deep 96-well plates with fast shaking (900 rpm). After 24 hours, cultures are diluted 1:20 into fresh M9 glucose medium with 1 mM IPTG and incubated for 72 hours in deep 96-well plates with fast shaking.

Glucaric Acid LC-MS Analysis

Detection and quantitation of glucaric acid from culture supernatant was achieved by LC/MS/MS analysis performed with an Agilent (Agilent Technologies, Santa Clara, CA) 6460 triple-quadrupole LC/MS/MS system. The ([M - H+]- / product ion) monitored via electrospray ionization in the negative ion mode with multiple reaction monitoring (MRM) was (209/85.1 amu). Mass spectrometer parameter settings were gas temp (350 °C), gas flow (12 L/min), nebulizer pressure (35 psi), sheath gas heater (400 °C), sheath gas flow (12 L/min), and capillary (4000V). An external standard curve mixture was analyzed at various concentrations 20 pg/uL – 10 ng/uL and utilized for quantitation.

Hydrophilic interaction chromatography (HILIC) conditions with a Phenomenex Luna 5u NH2, 250 x 2.00 mm column (Torrance, CA) were as follows: flow rate: 0.4 mLs/min; solvent A: 20 mM ammonium acetate + 20 mM ammonium hydroxide in 95:5 water:acetonitrile, pH 9.45; solvent B: acetonitrile. The gradient was as follows: t = 0, 85 % B; t = 10 min, 0 % B; t = 11, 0 % B; 3 minute equilibration.

Naringenin LC-MS Analysis

Naringenin was extracted from culture supernatant by mixing with an equal volume of ethyl acetate, mixed vigorously for 30 seconds on a benchtop vortexer, and the mixture briefly centrifuged for phase separation. The organic (upper) phase was transferred to a glass vial, and the ethyl acetate evaporated by gently circulating inert gas into the glass vial. The sample is resuspended in 200 uL of methanol and stored at -20 °C.

Samples were analyzed by LC/MS using a Bruker maXis impact Q-Tof (Billerica, MA) with an Agilent 1290 HPLC (Palo Alto, CA). A 2 x 100 mm Phenomenex Gemini column with 3 µm particle size (Torrance, CA) was used for a gradient separation. Mobile phase A was 0.1 % formic acid in water and mobile phase B was acetonitrile with a flow rate of 0.2 mL/min. The gradient began with 0 % B for 1 min, then increased to 100 % B after 12 min and held for 3 min before returning to 0 % B. Total run time was 20 min. The mass spectrometer was operated in the MRM mode selecting $(M\text{-}H)^-$ for coumaric acid (m/z 163) and naringenin (m/z 271) with a collision energy of 20 eV for each. The total ion chromatograms from each product ion scan were plotted separately and integrated. Naringenin standards at 2 and 10 µM were used as standards.

Whole Genome Sequencing

Strains for whole genome sequencing were grown overnight to saturation in LB supplemented in appropriate antibiotics. Genomic DNA was extracted from 1 ml culture using the Qiamp DNA Mini kit (Qiagen, Hilden, Germany). 1 μg genomic DNA was sheared using a Covaris S2 Ultrasonicator (Covaris, Woburn, MA). Paired-end libraries were prepared using the TruSeq DNA kit (Illumina, San Diego, CA). After quantitative PCR quantification, libraries were sequenced using paired end 300 bp sequencing on a MiSeq (Illumina) to mean of >100 coverage depth per genome. Reads were aligned to the MG1655 genome sequence, and single nucleotide polymorphisms (SNPs) and structural variations were detected using Lasergene SeqMan Pro 11 software (DNAStar, Madison, WI).

Bioreactor Production of Naringenin

We used an Eppendorf Celligen 310 bioreactor for fed-batch fermentation of the highest producing naringenin strain from the fourth round of evolution. A 1 liter culture of the strain was grown in M9 medium supplemented with 1 % glucose, casamino acids and biotin for 5 days at 32 °C with constant flow of pressurized, filtered air. The pH of the culture was buffered with ammonium hydroxide, and foaming was suppressed with anti-foaming agent. Glucose and casamino acids were provided every 12 hours. Samples were drawn periodically and assayed for coumaric acid and naringenin concentrations.

Supplemental Tables and Figures

Table C.1: Sensor-promoter pairs used in this study.

Sensor	Cognate promoter
BenM	benAp(1)
theoRR	None: riboswitch(2)
btuB	None: riboswitch(3)
AlkS	alkBp(4)
LacI	pLlacO(5)
XylR	xylAp(6)
CdaR	gudPp(7)
MphR	mphAp(8)
TetR	pLtetO(5)
TtgR	ttgAp(9)

Table C.2: List of codon optimized gene sequences used in this study.

Name	Sequence 5' to 3'
ttgR	ATGGTGCGTCGCACCAAAGAAGAAGCACAGGAAACGCGTGCGCAGATTATCGAAGCGGCCGAA CGCGCGTTTTATAAACGTGGTGTGGCACGTACCACGCTGGCAGATATTGCAGAACTGGCAGGTG TTACCCGCGGTGCAATCTACTGGCATTTCAACAATAAAGCCGAACTGGTTCAGGCACTGCTGGA TTCTCTGCACGAAACGCATGATCACCTGGCCCGTGCAAGCGAATCTGAAGATGAACTGGACCCG CTGGGCTGCATGCGCAAACTGCTGCTGCAGGTGTTTAACGAACTGGTTCTGGATGCACGTACCC GTCGCATTAATGAAATCCTGCATCACAAATGCGAATTTACGGATGATATGTGTGAAATTCGTC AGCAGCGCCAGAGCGCCGTGCTGGATTGTCATAAAGGTATCACCCTGGCACTGGCAAACGCAGT TCGTCGCGGTCAGCTGCCGGGTGAACTGGATGTGGAACGCGCAGCGGTTGCGATGTTTGCCTAT GTGGATGGCCTGATTGGTCGTTGGCTGCTGCTGCCGGATAGTGTTGATCTGCTGGGCGATGTGG AAAAATGGGTTGATACCGGTCTGGATATGCTGCGTCTGAGCCCGGCGCTGCGCAAATAA
mphR	ATGCCGCGTCCGAAACTGAAATCTGACGACGAAGTTCTGGAAGCGGCGACCGTTGTTCTGAAAC GTTGCGGTCCGATCGAATTCACCCTGTCTGGTGTTGCGAAAGAAGTTGGTCTGTCTCGTGCGGC GCTGATCCAGCGTTTCACCAACCGTGACACCCTGCTGGTTCGTATGATGGAACGTGGTGTTGAA CAGGTTCGTCACTACCTGAACGCGATCCCGATCGGTGCGGGTCCGCAGGGTCTGTGGGAATTCC TGCAGGTTCTGGTTCGTTCTATGAACACCCGTAACGACTTCTCTGTTAACTACCTGATCTCTTG GTACGAACTGCAGGTTCCGGAACTGCGTACCCTGGCGATCCAGCGTAACCGTGCGGTTGTTGAA GGTATCCGTAAACGTCTGCCGCCGGGTGCGCCGGCGGCGGCGGAACTGCTGCTGCACTCTGTTA TCGCGGGTGCGACCATGCAGTGGGCGGTTGACCCGGACGGTGAACTGGCGGACCACGTTCTGGC GCAGATCGCGGCGATCCTGTGCCTGATGTTCCCGGAACACGACGACTTCCAGCTGCTGCAGGCG CACGCGTAA
MIOX	GTGAAAGTGGATGTTGGCCCGGACCCGAGCCTGGTTTACCGCCCGGATGTGGACCCGGAAATGG CAAAAAGCAAAGATTCGTTTCGTAACTACACCAGTGGCCCGCTGCTGGATCGTGTTTTTACCAC GTATAAACTGATGCATACCCACCAGACGGTTGACTTTGTCAGCCGTAAACGCATTCAATATGGC GGTTTCTCTTACAAGAAAATGACCATCATGGAAGCGGTGGGCATGCTGGATGACCTGGTTGAT GAATCAGATCCGGACGTCGATTTTCCGAATTCGTTTCATGCGTTCCAGACGGCCGAAGGTATTC GCAAAGCCCACCCGGACAAAGATTGGTTCCATCTGGTCGGCCTGCTGCACGATCTGGGTAAAAT CATGGCACTGTGGGGTGAACCGCAGTGGGCTGTGGTTGGTGATACCTTTCCGGTGGGTTGCCGT CCGCAAGCAAGTGTCGTGTTTTGTGACTCCACCTTCCAGGACAACCCGGATCTGCAAGACCCGC GCTATTCAACGGAACTGGGCATGTACCAGCCGCATTGCGGTCTGGAAAACGTGCTGATGTCGTG GGGTCACGATGAATACCTGTACCAGATGATGAAATTCAACAAATTCAGCCTGCCGTCTGAAGCC TTCTACATGATCCGTTTCCATAGTTTCTACCCGTGGCACACCGGCGGTGATTATCGCCAGCTGT GCTCCCAGCAAGACCTGGATATGCTGCCGTGGGTGCAAGAATTCAACAAATTCGATCTGTACAC GAAATGTCCGGATCTGCCGGACGTTGAATCTCTGCGTCCGTACTACCAAGGTCTGATTGATAAA TACTGTCCGGGCACCCTGTCGTGGTAA

Table C.3: Genomic MAGE targets used for naringenin pathway diversification.

Evolution round	Up-regulation targets	Down-regulation targets	Coding targets
1	accABCD aceEF gapA lpdA pgk	acnA fumC mdh sucC	
2	accABCD aceEF gapA lpdA pgk	acnA fumC mdh sucC	
3	accABCD aceEF gapA lpdA pgk	acnA fumC mdh sucC	
4		fabBDFH	tyrA_M53I tyrA_A354V aroG_D146N

Table C.4: Oligonucleotides used for naringenin mutagenesis.

accA T7	accA	C*A*A*G*GAAATTCAGACTcatAGTATTCCTGTATTAtctccctatagtgagtcgtattaGTCAAACTCCAGTTCCACCTGCTCCGAACCAAT
accA_JC23100	accA	A*T*T*C*AGACTcatAGTATTCCTGTATTAgctagcactgtacctaggactgagctagccgtcaaGTCAAACTCCAGTTCCACCTGCTCCGAA
accB_T7	accB	C*G*G*T*GAAACGCCTGTCACAATCACACTAAACAAtaatacgactcactataggggagaAGAGTACGGAACCCACTCatgGATATTCGTAAGA
accB_JC23100	accB	A*A*C*G*CCTGTCACAATCACACTAAACAAttgacggctagctcagtcctaggtacagtgctagcAGAGTACGGAACCCACTCatgGATATTC
accD_T7	accD	G*T*T*C*AATCCAGCTcatTAGGGACCTTTCTGTCTctccctatagtgagtcgtattaGAACCTGGTTCGATGCCAGTTTTATCTTTGGGGA
accD JC23100	accD	T*C*C*A*GCTcatTAGGGACCTTTCTGTCTgctagcactgtacctaggactgagctagccgtcaaGAACCTGGTTCGATGCCAGTTTTATCTT
aceE_T7	aceE	A*A*C*G*TTCTGAcatGGGTTATTCCTTATCTATCTctccctatagtgagtcgtattaAATAACGTTGAGTTTTCTGGAACCTGTCCCATTG
aceE_JC23100	aceE	C*T*G*A*catGGGTTATTCCTTATCTATCTgctagcactgtacctaggactgagctagccgtcaaAATAACGTTGAGTTTTCTGGAACCTGTC
lpdA_T7	lpdA	T*T*T*G*ATTTCAGTACTcatCATGACCTCTATATAtctccctatagtgagtcgtattaTTTATCTCCGGCGGTCATACCCGTCGTCTTTCAG
lpdA_JC23100	lpdA	T*T*C*A*GTACTcatCATGACCTCTATATAgctagcactgtacctaggactgagctagccgtcaaTTTATCTCCGGCGGTCATACCCGTCGTC
gapA_T7	gapA	G*T*A*A*TTTTACAGGCAACCTTTTATTCACTAACAtaatacgactcactataggggagaAATAGCTGGTGGAATATatgACTATCAAAGTAGG
gapA_JC23100	gapA	T*T*A*C*AGGCAACCTTTTATTCACTAACAttgacggctagctcagtcctaggtacagtgctagcAATAGCTGGTGGAATATatgACTATCAA
pgk_T7	pgk	A*T*C*T*TAATTACAGAcatGGTGAATCCTCTCGTTtctccctatagtgagtcgtattaGATTCTAAAAGTTTTGCAGACGCTGCTTGCGTCT
pgk_JC23100	pgk	A*T*T*A*CAGAcatGGTGAATCCTCTCGTTgctagcactgtacctaggactgagctagccgtcaaGATTCTAAAAGTTTTGCAGACGCTGCTT
fumB_BTG	fumB	A*C*G*C*CATTTTCGAATAACAAATACAGAGTTACAGGCTGGAAGCTBtgTCAAACAAACCCTTTATCTACCAGGCACCTTTCCCGATGGGGA
fumC_BTG	fumC	C*A*T*C*AATCGCCCCCATCGAATCTTTTTCGCTGCGTACTGTATTcaVGACCTGCTCCTCACCTGATTAATTTTTTCTTTCTGTTTTGCTTT
mdh_BTG	mdh	C*G*C*C*TGGCCAATACCGCCAGCAGCGCCGAGGACTGCGACTTTcaVCCTAAACTCCTTATTATATTGATAAACTAAGATATGTTGCTCCGC
acnA_BTG	acnA	C*T*G*A*AGAGAATCAGGGCTTCGCAACCCTGTCATTAAGGAGGAGCTBtgTCGTCAACCCTACGAGAAGCCAGTAAGGACACGTTGCAGGCC
tyrA_Met53Ile	tyrA	G*T*A*C*ACCCAGAGCTTCCGCCTCTGCACGACGCGAGGCCAAaATAGATGCCTCGCGCTCCGGAACATAAATAGGCAGTCCAAAGCGGCTTT
tyrA_Ala354Val	tyrA	T*C*A*T*TCGCCTGACGCAATAACACGCGGCTTTCACTCTGAAAACGCTGTaCGTAATCGCCGAACCAGTGCTCCACCTTGCGGAAACTGTCA
aroG Asp146As	aroG	C*C*A*G*CTCATCAGGTCAGCGAGATATTGTGGGGTGATCATATtGAGAAACTCACCTGCCGCTGGCAGACCGCTGTCGTTAATATCAAGCAG
accA_RBS	accA	C*A*A*T*CGGCTGTTCAAAATCAAGGAAATTCAGACTcatAGTHHHHYYYYYHHTTAGTCAAACTCCAGTTCCACCTGCTCCGAACCAATGAG
accB_RBS	accB	C*G*G*T*GAAACGCCTGTCACAATCACACTAAACAAAGAGDDRRRRRDDDDCTCatgGATATTCGTAAGATTAAAAAACTGATCGAGCTGGTT
accC_RBS	accC	C*C*G*G*TAGAATTTGACGAGCCGCTGGTCGTCATCGAGtaaDDRRRRRDDDDAACATGCTGGATAAAATTGTTATTGCCAACCGCGGCGAGA
accD_RBS	accD	G*G*T*G*GGAGTAATGTTGCTTTTAATTCGTTCAATCCAGCTcatTAGHHHHYYYYYHHTCTGAACCTGGTTCGATGCCAGTTTTATCTTTGG
aceE RBS	aceE	C*G*A*T*CGGATCCACGTCATTTGGGAAACGTTCTGAcatGGGHHHHYYYYYHHCTATCTAATAACGTTGAGTTTTCTGGAACCTGTCCCATT
aceF_RBS	aceF	C*C*G*A*TGTCCGGTACTTTGATTTCGATAGCCATTATHHHHYYYYYHHttaCGCCAGACGCGGGTTAACTTTATCTGCATCGATGTTGAATT
lpd_RBS	lpdA	C*C*C*A*AGTACCACGACCTGAGTTTTGATTTCAGTACTcatCATHHHHYYYYYHHATTTATCTCCGGCGGTCATACCCGTCGTCTTTCAGGC
gapA_RBS	gapA	T*T*T*G*TAATTTTACAGGCAACCTTTTATTCACTAACAAATDDRRRRRDDDDTATatgACTATCAAAGTAGGTATCAACGGTTTTGGCCGTA
pgk_RBS	pgk	A*G*C*A*AGATCCAGATCGGTCATCTTAATTACAGAcatGGTHHHHYYYYYHHTTGATTCTAAAAGTTTTGCAGACGCTGCTTGCGTCTTACC
sucC BTG	sucC	A*T*A*G*CGGGCAAAAAGTTGTTTTGCCTGATATTCATGTAAGTTcavGTGTTCTGTCCATCCTTCAGTAATCGTTATCTTTTAAACCGTAGA
fabB_mut1	fabB	A*A*C*A*ATGCCCAGGCCAGTAATCACTGCACGTTTCATTCAATACCTCNGTAAGTCGCACATAGAGTAAGTTTCGAATGCACAATAGCGTAC
fabB_mut2	fabB	A*A*C*A*ATGCCCAGGCCAGTAATCACTGCACGTTTCATTCAATACCTNTGTAAGTCGCACATAGAGTAAGTTTCGAATGCACAATAGCGTAC
fabB_mut3	fabB	A*A*C*A*ATGCCCAGGCCAGTAATCACTGCACGTTTCATTCAATACCNCTGTAAGTCGCACATAGAGTAAGTTTCGAATGCACAATAGCGTAC
fabB_mut4	fabB	A*A*C*A*ATGCCCAGGCCAGTAATCACTGCACGTTTCATTCAATACNTCTGTAAGTCGCACATAGAGTAAGTTTCGAATGCACAATAGCGTAC
fabB mut5	fabB	A*A*C*A*ATGCCCAGGCCAGTAATCACTGCACGTTTCATTCAATANCTCTGTAAGTCGCACATAGAGTAAGTTTCGAATGCACAATAGCGTAC
fabF_mut1	fabF	T*G*C*C*CAGTCCGGTCACAACTACACGACGCTTAGACACGTTTGTCCTCNAGGGAGGGAAAAAATGATTCTAGTGGGACAAAAAGATAAAAC
fabF_mut2	fabF	T*G*C*C*CAGTCCGGTCACAACTACACGACGCTTAGACACGTTTGTCCTNCAGGGAGGGAAAAAATGATTCTAGTGGGACAAAAAGATAAAAC
fabF_mut3	fabF	T*G*C*C*CAGTCCGGTCACAACTACACGACGCTTAGACACGTTTGTCCNCCAGGGAGGGAAAAAATGATTCTAGTGGGACAAAAAGATAAAAC
fabF_mut4	fabF	T*G*C*C*CAGTCCGGTCACAACTACACGACGCTTAGACACGTTTGTCNTCCAGGGAGGGAAAAAATGATTCTAGTGGGACAAAAAGATAAAAC
fabF mut5	fabF	T*G*C*C*CAGTCCGGTCACAACTACACGACGCTTAGACACGTTTGTNCTCCAGGGAGGGAAAAAATGATTCTAGTGGGACAAAAAGATAAAAC
fabF_mut6	fabF	T*G*C*C*CAGTCCGGTCACAACTACACGACGCTTAGACACGTTTGNCCTCCAGGGAGGGAAAAAATGATTCTAGTGGGACAAAAAGATAAAAC
fabD_mut1	fabD	C*C*C*T*GTCCAGGGAACACAAATGCAAATTGCGTCATGTTTTAATCCTNATCCTAGAAACGAACCAGCGCGGAGCCCCAGGTGAATCCACCG
fabD_mut2	fabD	C*C*C*T*GTCCAGGGAACACAAATGCAAATTGCGTCATGTTTTAATCCNTATCCTAGAAACGAACCAGCGCGGAGCCCCAGGTGAATCCACCG
fabD_mut3	fabD	C*C*C*T*GTCCAGGGAACACAAATGCAAATTGCGTCATGTTTTAATCNTTATCCTAGAAACGAACCAGCGCGGAGCCCCAGGTGAATCCACCG
fabD mut4	fabD	C*C*C*T*GTCCAGGGAACACAAATGCAAATTGCGTCATGTTTTAATNCTTATCCTAGAAACGAACCAGCGCGGAGCCCCAGGTGAATCCACCG
fabD_mut5	fabD	C*C*C*T*GTCCAGGGAACACAAATGCAAATTGCGTCATGTTTTAANCCTTATCCTAGAAACGAACCAGCGCGGAGCCCCAGGTGAATCCACCG
fabH_mut1	fabH	C*A*G*A*TAGCTGCCAGTACCAATAATCTTCGTATACATGTACGCTCAGTCANTTTTCGGTTATATACCGTCACTTGCAAACTGCGAGTTCGC
fabH_mut2	fabH	C*A*G*A*TAGCTGCCAGTACCAATAATCTTCGTATACATGTACGCTCAGTNACTTTTCGGTTATATACCGTCACTTGCAAACTGCGAGTTCGC
fabH_mut3	fabH	C*A*G*A*TAGCTGCCAGTACCAATAATCTTCGTATACATGTACGCTCAGNCACTTTTCGGTTATATACCGTCACTTGCAAACTGCGAGTTCGC
fabH mut4	fabH	C*A*G*A*TAGCTGCCAGTACCAATAATCTTCGTATACATGTACGCTNAGTCACTTTTCGGTTATATACCGTCACTTGCAAACTGCGAGTTCGC
fabH_mut5	fabH	C*A*G*A*TAGCTGCCAGTACCAATAATCTTCGTATACATGTACGCNCAGTCACTTTTCGGTTATATACCGTCACTTGCAAACTGCGAGTTCGC
fabH_mut6	fabH	C*A*G*A*TAGCTGCCAGTACCAATAATCTTCGTATACATGTACGNTCAGTCACTTTTCGGTTATATACCGTCACTTGCAAACTGCGAGTTCGC

Table C.5: Oligonucleotides used for glucaric acid mutagenesis.

MAGE Oligo Name	Locus	Function	
garKnull	garK	stop	T*T*C*C*CGAAATCCTTTTTCTATCGCCTGCGCAACCTCGCTGGCAGATTAACTTTCTTTTTAAGAGTCTGGGGCGATTACGATTTTCATACC
uxaCnull	uxaC	stop	C*A*A*A*TGGCAATGGTAATCGAAAATCGGCTGGTCTTTTGCTTAGTCGTGTTACAGACGGCGGGCAAATTCGGTATCTAACAGGAAATCTTC
suhB-degen	suhB	RBS	C*T*C*G*CTGCTATACTCTGCGCCG TTTTCCCGTTCTTTAACATCCDDVVVVVDDDDCCG atgCATCCGATGCTGAACATCGCCGTGCGCGCA
pgi-degen	pgi	RBS	G*G*C*A*GCGGTC TGCGTTGGAT TGATGTTTTT CATTAGHHHH BBBBBHHTGA TTTTGAGAAT TGTGACTTTG GAAGATTGTA GCGCCAGTCA
sthA-degen	sthA	RBS	C*G*C*G*ATAAAATGTTACCATTCT GTTGCTTTTATGTATAAGAACDDVVVVVDDDDACC atgCCACATTCCTACGATTACGATGCCATAGTA
zwf-degen	zwf	RBS	G*T*C*A*CAGGCC TGGGCTGTTT GCGTTACCGC CATGTCHHHH BBBBBHHGTT AACTAACCCG GTACTTAAGC CAGGGTATAC TTGTAATTTT
mdh-degen	mdh	RBS	A*C*C*G*CCAGCA GCGCCGAGGA CTGCGACTTT CATCCTHHHH BBBBBHHTAT ATTGATAAAC TAAGATATGT TGCTCCGCTG CCGCGACCTT

Table C.6: List of key candidate genes for increasing naringenin predicted by flux balance analysis.

Reaction	Direction	Gene(s)
ACCOAC	Up	b0185 and b2316 and b3255 and b3256
G6PDH2r	Down	b1852
GAPD	Up	b1779
CS	Down	b0720
FUM	Down	b1612 or b4122 or b1611
TKT2	Down	b2935
RPE	Down	b3386
PGM	Up	b0755
ENO	Up	b2779
F6PA	Up	b0825
	Up	b2926
DHAPT	Up	b1200 and b1199 and b1198 and b2415 and b2416
TPI	Up	b3919
TALA	Up	b0008
PYK	Up	b1676
PGI	Up	b4025
THD2pp	Up	b1602 and b1603
ACONT	Down	b0118
MDH	Down	b3236
PDH	Up	b0114 and b0115 and b0116
TALA	Up	b0008

Table C.7: Mutations and its corresponding genomic locus found in the seven evolved naringenin pathway strains. Genomic locus is with reference to *E. coli* MG1655 genome.

Strain	Gene name	Type of change	Genomic locus	Mutation
Strain 1	fabD	RBS	1148940	A->G
	fabF	Start codon	1151162	GTG->TTG
	fabH	Start codon	1147982	ATG->GTG
	fumC	Start codon	1684612	ATG->TTG
	entB	Frameshift	627205	G insertion
	rpoD	RBS	3211052	A->G
Strain 2	fabF	Start codon	1151162	GTG->TTG
	fabH	Start codon	1147982	ATG->GTG
	fumC	Start codon	1684612	ATG->TTG
	entB	Frameshift	627205	G insertion
	rpoD	RBS	3211052	A->G
Strain 3	fabD	RBS	1148939	A->T
	fabF	Start codon	1151162	GTG->TTG
	fabH	Start codon	1147982	ATG->GTG
	fumC	Start codon	1684612	ATG->TTG
	entB	Frameshift	627205	G insertion
	rpoD	RBS	3211052	A->G
Strain 4	fabD	RBS	1148940	A->G
	fabH	Start codon	1147982	ATG->GTG
	fumC	Start codon	1684612	ATG->TTG
	aroG	Coding	785291	G->A
	entB	Frameshift	627205	G insertion
	rpoD	RBS	3211052	A->G
Strain 5	fabF	RBS	1151156	A->C
	fabH	Start codon	1147982	ATG->GTG
	fumC	Start codon	1684612	ATG->TTG
	hcaT	Frameshift	2665603	C insertion
	entB	Frameshift	627205	G insertion
	rpoD	RBS	3211052	A->G
Strain 6	fabF	Start codon	1151162	GTG->TTG
	fabH	Start codon	1147982	ATG->GTG
	fumC	Start codon	1684612	ATG->TTG
	tyrA	Coding	2737031	G->A
	entB	Frameshift	627205	G insertion
	rpoD	RBS	3211052	A->G
Strain 7	fabD	RBS	1148942	G->C
	fabH	RBS	1149790	G->C
	fumC	Start codon	1684612	ATG->TTG
	tyrA	Coding	2737031	G->A
	mhpD	Frameshift	372032	G deletion
	entB	Frameshift	627205	G insertion
	rpoD	RBS	3211052	A->G

Mutations and its corresponding genomic locus found in the seven evolved glucaric acid pathway strains

Strain	Gene name	Type of change	Genomic locus	Mutation
Strain 1	garK	Nonsense	3269752	A->T
	garK	Nonsense	3269763	A->T

Table C.8: Number of mutations found in evolved naringenin pathway strains

Strain name	Start codon	Frameshift/Nonsense	Non-synonymous
Strain 1	3	10	65
Strain 2	3	9	67
Strain 3	3	10	67
Strain 4	2	10	70
Strain 5	2	11	65
Strain 6	3	9	70
Strain 7	1	11	67

Number of mutations found in evolved glucaric acid pathway strains

Strain name	Start codon	Frameshift/Nonsense	Non-synonymous
Strain 1	2	8	32

Figure C.1: Time course of naringenin and coumaric acid production titer in bioreactor

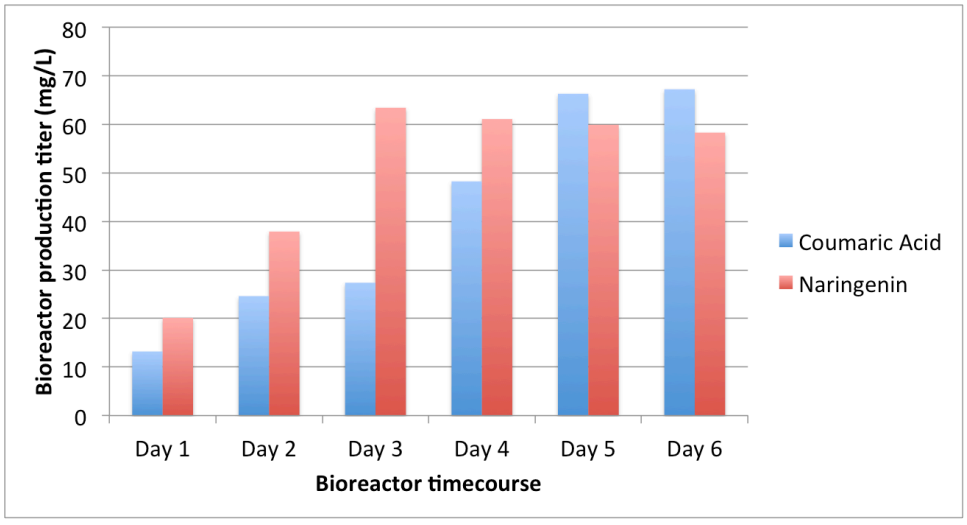


Figure C.2: Growth curve comparison between TtgR-tolC parent strain, evolved strain and evolved strain + accABCD for naringenin production

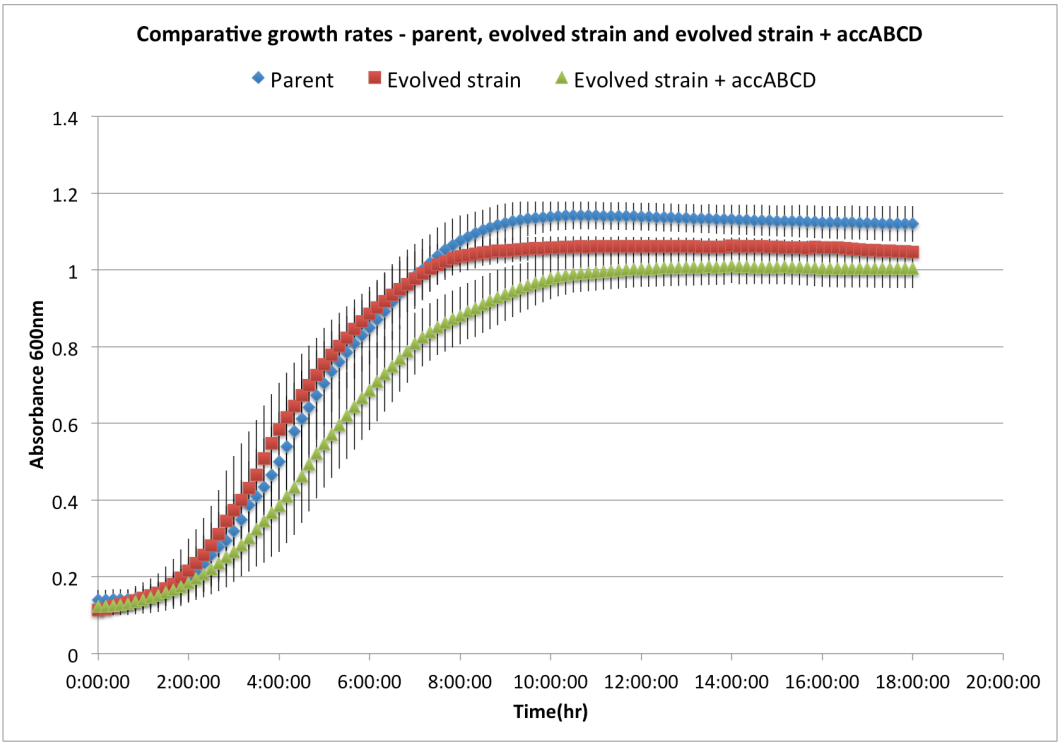


Figure C.3: TtgR-tolC strain growth comparison under SDS selection, exposed to Naringenin or Coumaric acid.

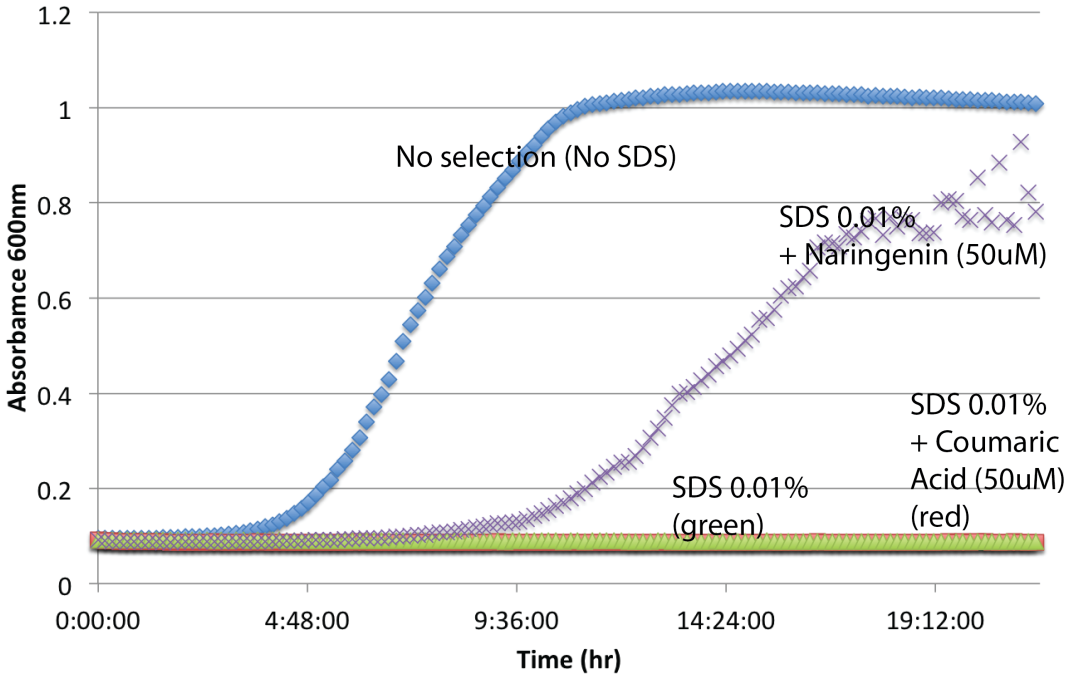


Figure C.4: Salis calculation of predicted RBS change effects

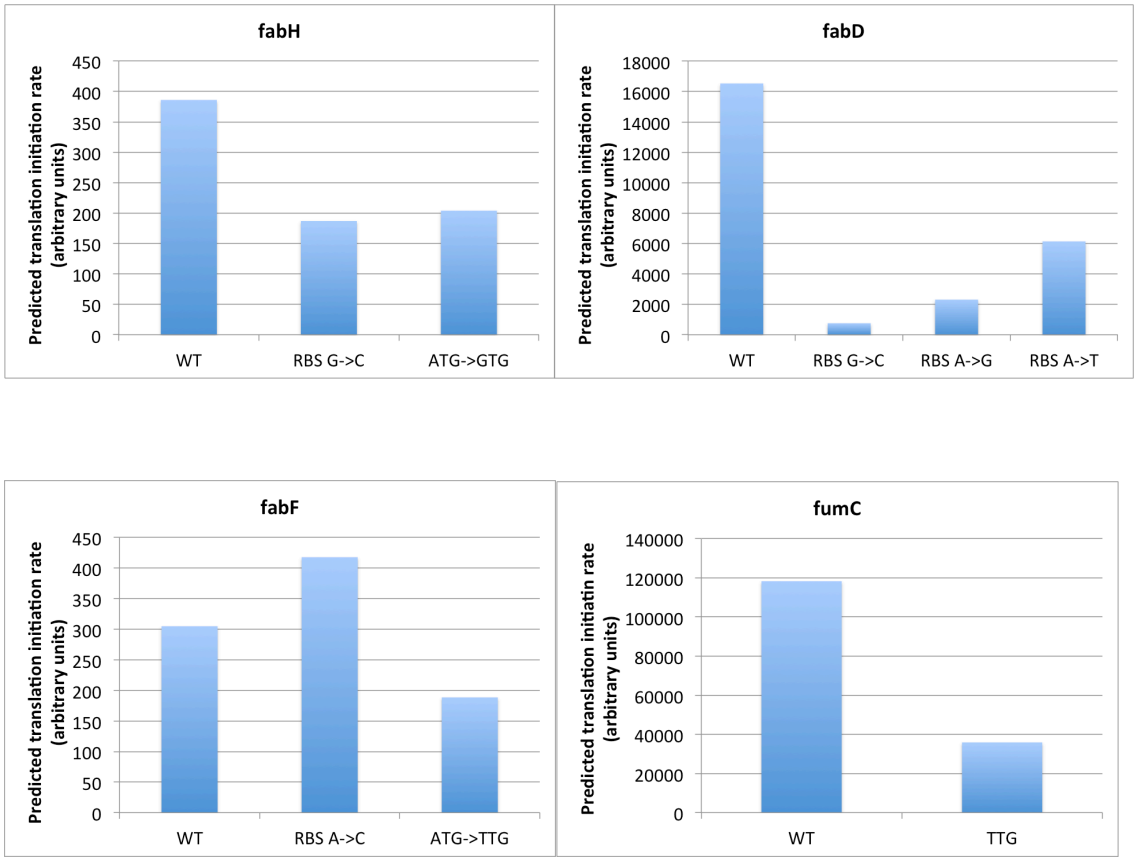


Figure C.5: Glucaric acid production by the pT7GAEXP plasmid in the cdaR-tolC sensor-selector strain (*E. coli* K12 background) and BL21 Star.

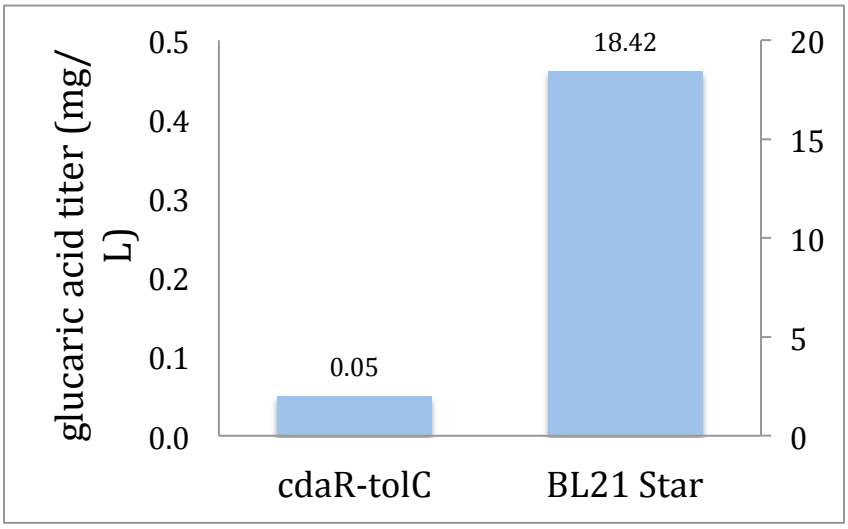


Figure C.6: TtgR operator RBS spacing modifications to reduce escape rate.

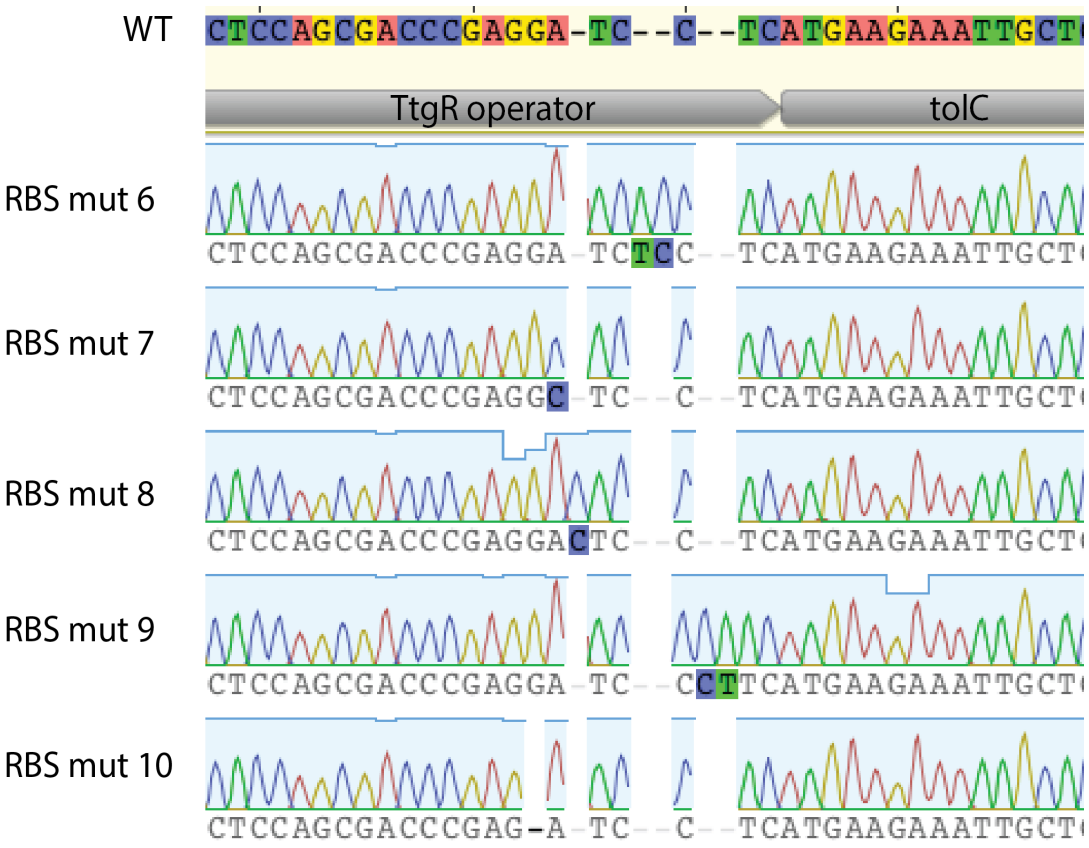


Figure C.7: CdaR-tolC sensor selector strain response to glucaric acid pathway intermediates

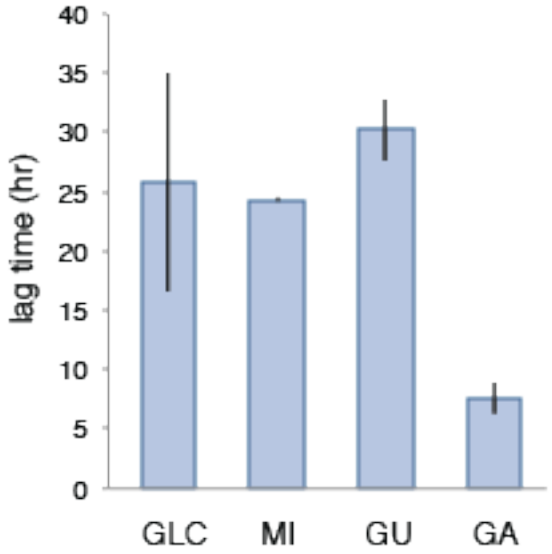
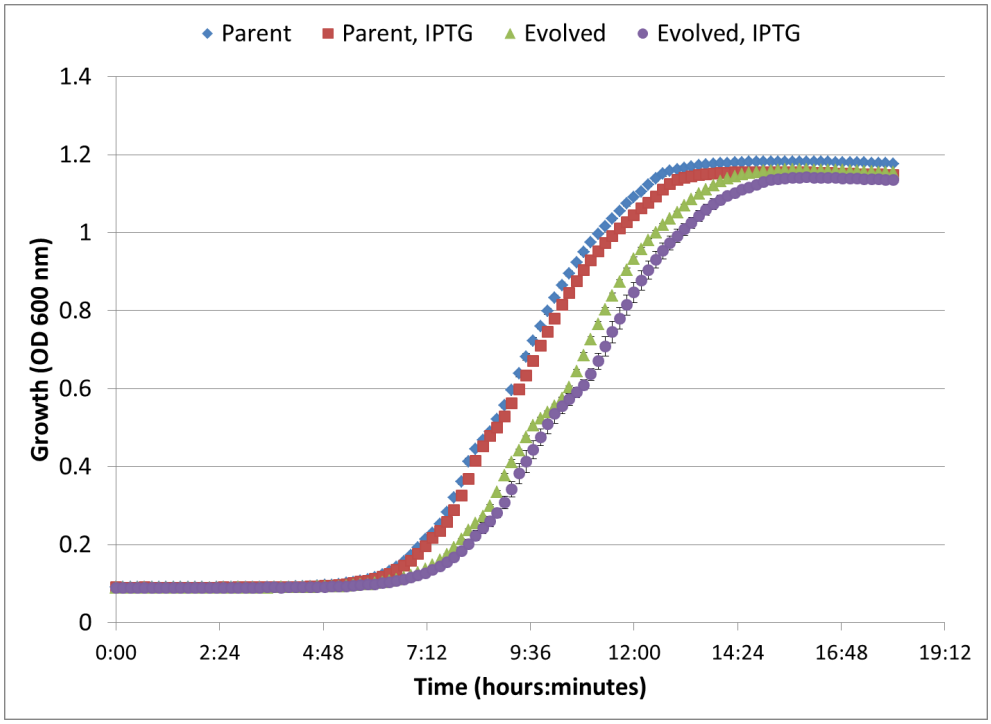


Figure C.8: Growth comparison of CdaR-tolC parent and best evolved glucaric acid production strain. IPTG used to induce pT7GAEXP glucaric acid production plasmid. Mean +/- standard deviation, 5 replicates.



Supplemental References

1. Craven SH et al. (2009) Inducer responses of BenM, a LysR-type transcriptional regulator from Acinetobacter baylyi ADP1. *Mol Microbiol* 72:881–894.

2. Lynch SA, Gallivan JP (2009) A flow cytometry-based screen for synthetic riboswitches. *Nucleic Acids Res* 37:184–192.

3. Fowler CC, Brown ED, Li Y (2010) Using a riboswitch sensor to examine coenzyme B(12) metabolism and transport in E. coli. *Chem Biol* 17:756–765.

4. Canosa I, Yuste L, Rojo F (1999) Role of the alternative sigma factor sigmaS in expression of the AlkS regulator of the Pseudomonas oleovorans alkane degradation pathway. *J Bacteriol* 181:1748–1754.

5. Lutz R, 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 Res* 25:1203–1210.

6. Sumiya M, Davis EO, Packman LC, McDonald TP, Henderson PJ (1995) Molecular genetics of a receptor protein for D-xylose, encoded by the gene xylF, in Escherichia coli. *Recept Channels* 3:117–128.

7. Monterrubio R, Baldoma L, Obradors N, Aguilar J, Badia J (2000) A common regulator for the operons encoding the enzymes involved in D-galactarate, D-glucarate, and D-glycerate utilization in Escherichia coli. *J Bacteriol* 182:2672–2674.

8. Zheng J et al. (2009) Structure and function of the macrolide biosensor protein, MphR(A), with and without erythromycin. *J Mol Biol* 387:1250–1260.

9. Terán W et al. (2003) Antibiotic-dependent induction of Pseudomonas putida DOT-T1E TtgABC efflux pump is mediated by the drug binding repressor TtgR. *Antimicrob Agents Chemother* 47:3067–3072.