Applying engineering principles to the design and construction of transcriptional devices

by

Reshma P. Shetty

Submitted to the Department of Biological Engineering
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

May 2008

Author .....
Department of Biological Engineering
May 23, 2008

Certified by .....
Thomas F. Knight, Jr.
Senior Research Scientist, Computer Science & Artificial Intelligence Laboratory
Thesis Supervisor

Certified by .....
Drew Endy
Cabot Assistant Professor of Biological Engineering
Thesis Supervisor

Accepted by .....
Alan Grodzinsky
Chairman, Department Committee on Graduate Students

Thesis Committee

Accepted by .....

Bruce Tidor

Professor of Biological Engineering and Computer Science

Chairman of Thesis Committee

Accepted by .....

Thomas F. Knight, Jr.

Senior Research Scientist, Computer Science & Artificial Intelligence Laboratory

Thesis Supervisor

Accepted by .....

Drew Endy

Cabot Assistant Professor of Biological Engineering

Thesis Supervisor

Accepted by .....

Robert T. Sauer

Salvador E. Luria Professor of Biology

Thesis Committee Member

Applying engineering principles to the design and construction of transcriptional devices

by

Reshma P. Shetty

Submitted to the Department of Biological Engineering
on May 23, 2008, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Abstract

The aim of this thesis is to consider how fundamental engineering principles might best be applied to the design and construction of engineered biological systems. I begin by applying these principles to a key application area of synthetic biology: metabolic engineering. Abstraction is used to compile a desired system function, reprogramming bacterial odor, to devices with human-defined function, then to biological parts, and finally to genetic sequences. Standardization is used to make the process of engineering a multi-component system easier. I then focus on devices that implement digital information processing through transcriptional regulation in *Escherichia coli*. For simplicity, I limit the discussion to a particular type of device, a trancriptional inverter, although much of the work applies to other devices as well. First, I discuss basic issues in transcriptional inverter design. Identification of key metrics for evaluating the quality of a static device behavior allows informed device design that optimizes digital performance. Second, I address the issue of ensuring that transcriptional devices work in combination by presenting a framework for developing standards for functional composition. The framework relies on additional measures of device performance, such as error rate and the operational demand the device places on the cellular chassis, in order to proscribe standard device signal thresholds. Third, I develop an experimental, proof-of-principle implementation of a transcriptional inverter based on a synthetic transcription factor derived from a zinc finger DNA binding domain and a leucine zipper dimerization domain. Zinc fingers and leucine zippers offer a potential scalable solution to the challenge of building libraries of transcription-based logic devices for arbitrary information processing in cells. Finally, I extend the principle of physical composition standards from parts and devices to the vectors that propagate those parts and devices. The new vectors support the assembly of biological systems. Taken together, the work helps to advance the transformation of biological system design from an *ad hoc*, artisanal craft to a more predictable, engineering discipline.

Thesis Supervisor: Thomas F. Knight, Jr.
Title: Senior Research Scientist, Computer Science & Artificial Intelligence Laboratory

Thesis Supervisor: Drew Endy
Title: Cabot Assistant Professor of Biological Engineering

Acknowledgments

I am forever indebted to my advisors Tom Knight and Drew Endy for bringing me to synthetic biology. Their willingness to tackle hard problems and to do so when everyone thinks you're crazy has given me a more valuable graduate experience than I could have possibly imagined. They've inspired me by their ability to see the world for what it should be rather than what it is and their tireless work to make it so. Tom, in particular, taught me what engineering really means. And Drew showed me how to look beyond my own research and examine the larger context and community in which I work. Their mentorship has made me a better researcher and a better person. Five years on, I definitely made the right choice.

To Austin, thank you for discussing every failed experiment and wild idea any time, day or night. I'm in awe of your rational logic in the face of any problem. To the Knight and Endy labs and the Registry/iGEM team, I've been honored to work with such a talented and passionate group of people. It's been a fantastic ride. To the OpenWetWare crew, especially Sri, Ilya, Jason, Barry, and Austin, thanks for not accepting the status quo and daring to change the way research is done. To the 2006 MIT iGEM team, thanks for pulling off our very own genetically-encoded "hack".

To Bruce Tidor, chair of my thesis committee, for encouraging me to use modeling to inform the design of transcription-based logic devices. And to Bob Sauer, another member of my thesis committee, for his practicality and sound experimental advice. I am also grateful to Scot Wolfe, J. Keith Joung, Amy Keating, Bambang Adiwijaya, Carl Pabo, and Kathleen McGinness for insightful advice and help at critical junctures in my work.

To Doug Lauffenburger, Linda Griffith, and the Biological Engineering department, thank you for creating a very special environment that welcomes contributions and input from every member of the community.

To the Olivera lab, especially Pradip Bandyopadhyay, Craig Walker, and Toto Olivera, thank you for sharing with me your love of science and for teaching me most of the biology I know.

To my 105 Chestnut roommates, Diana, Nicholas/Marc/Antonio, and Tyberius, thank you for West Wing nights, summer bbq's, and ridiculing me out of my daily morning grumpiness. And to my 52 Simpson roommates, S and K, thank you for Sunday night dinners and late night conversations.

To beh-one, be-2002, be-2003 and girls_night, especially the Aldridges, Ale, Amy, Brian, Dave, Diana, "cruel" Eric, the Frooks, Jason, Jonnie B., Kathryn, Lara and husband, Maxx, Nate, Nick, Samantha, Sri, Teather, and Victor, thank you for being an amazing group of friends and endless source of laughter and fun. You got me through classes, quals, and the ups and downs of research. To Michelle, thank you for your friendship through the years ... you've taught me a lot about what it means to be a friend. And to Jing, thank you for convincing me to major in computer science and for fatigue-induced, delirious silliness during late night coding sessions in the computer lab.

To Barry, you've been my rock and my best friend during the years. Thanks for

always seeing the best in me.

To my sister Rekha, thank you for your love, wrestling fights, fort-building, and most recently, for going shopping for me so I don't have to. And to the rest of my family spread out all over the world, your love and support means a lot.

Finally, I dedicate this thesis to my mom and dad for their belief in me, their unconditional love, and their sacrifices. Everything I am I owe to them.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Formally, engineering is defined as the application of scientific knowledge to meet human needs. In practice, a key goal of engineering is to make the process of engineering many component systems faster, cheaper, and more predictable. To do so, engineers employ a variety of approaches that are broadly applicable to all branches of engineering. Mathematical models inform the design of complex systems. Standards ensure that components can be combined to build composite components that work as predicted. Finally, rigorous measurement tools and techniques characterize and quantify component behavior. I begin by discussing two examples from mature engineering disciplines that highlight how fundamental engineering principles make the engineering of many component systems easier.

Engineers have long known that many component engineered systems should be built from standard parts. For instance, after the War of 1812, the United States was left with thousands of broken guns that were beyond repair [1]. These guns could not be fixed because, at the time, guns were hand-crafted and lacked uniformity. Each gun part only fitted together with the other parts from that same gun. As a result, when individual gun parts broke, they could not be easily replaced and the gun was rendered useless. In response to prohibitive gun maintenance and repair costs, a handful of armory authorities began to advocate for the standardization of gun part design and manufacturing across national armories. The result was a massive engineering effort, spanning many years, to standardize gun part design and manufacture, so

that gun parts were interchangeable. In the long term, the armories found that standardization led to less development effort, higher quality and cheaper production costs. Advancements in engineering of standard, interchangeable parts also eventually propagated into other industries like manufacturing of sewing machines, agricultural machinery, and railroad equipment.

Arguably the best example of how functional standards can enable engineering of many component systems is the Internet. In the early 1970s, the Defense Advanced Research Projects Agency (DARPA) was interested in enabling satellite networks and ground-based networks to communicate with each other [2]. In response, DARPA-sponsored researchers developed common standards that govern how computers send data back and forth across networks [3]. The eventual result, the TCP/IP protocol suite, is still used today. The set of standards underlying the Internet allows desktop computers, laptop computers, server farms, cell phones, and more to connect and share information. Moreover, these devices can share information via a variety of high-level services, such as email, video streaming, file sharing, and static webpages, all using a common set of underlying standards specifying how data packets should be sent back and forth. Stated more generally, functional standards that allow components to connect and work reliably and predictably in combination enable the engineering of complex, multi-component systems, such as the Internet.

In 1978, Waclaw Szybalski and Ann Skalka wrote, in an editorial congratulating Werner Arber, Hamilton Smith, and Daniel Nathans on winning the Nobel Prize, that, "The work on restriction nucleases not only permits us easily to construct recombinant DNA molecules and to analyze individual genes but also has led us into the new era of 'synthetic biology' where not only existing genes are described and analyzed but also new gene arrangements can be constructed and evaluated" [4]. Over the course of the past thirty years, researchers have made extensive use of recombinant DNA technology both to understand how cells work and to build cells that have novel functions [5, 6]. Yet biological research remains largely a craft. The field is primarily driven by a small number of highly-skilled experts. To be successful in the field, newcomers must usually learn the trade through years of apprenticeship. To

realize the extraordinary capabilities of biology to produce materials and to sense and respond to information, design and construction of biological systems must transition from a craft to an industrialized process, just as other technologies have. The goal of synthetic biology is to bring about this transition, by systematizing and standardizing the process of engineering biology.

Many argue that biology can't be standardized in the same way that gun manufacturing and computer networks were. Biology is simply too complicated, and we understand too little of it, to engineer multi-component systems of the scale that we can in mature engineering disciplines. There is little doubt that the challenge is tremendous: in addition to the challenges of physical and functional composition in the face of imperfect understanding that is common to other engineering disciplines, biological systems also have the added dimension of being self-reproducing in the presence of error (mutation). Yet it is worth noting that standardization did not come easily in mature engineering disciplines either. When the United States Ordnance Department decided to standardize gun part design and manufacturing in 1815, the idea of uniformity and interchangeability in gun parts had already been promulgated as early as the 1780s by several French armories [1]. Indeed, it took many years of sustained effort as well as several machine innovations post-1815 to achieve a significant standardization across just two national armories. Interestingly, a significant hurdle in the process was the resistance and skepticism of armory personnel that standardization was possible or even desirable. Even over a century later when engineers were trying to establish functional standards for computer networks, it took several years of iterative refinements to develop the TCP/IP protocol suite used today. In biological engineering, an even larger scale of foundational engineering investment is needed, since there is an interest not in any one problem but in opening a diverse array of problems to biological solutions [7].

In this thesis, I consider how fundamental engineering principles might be best applied to the design and construction of engineered biological systems. In Chapter 2, I apply these principles to a key application area of synthetic biology: metabolic engineering. Abstraction is used to compile a desired system function, reprogramming

bacterial odor, to devices with human-defined function, then to biological parts, and finally to genetic sequence. Standardization is used to make the process of engineering a multi-component system easier. In the following three chapters, I focus on devices that implement digital information processing through transcriptional regulation in *Escherichia coli*. For simplicity, I limit the discussion to a particular type of device, a trancriptional inverter, although much of the work applies more generally as well. In Chapter 3, I discuss basic issues in transcriptional inverter design. A transfer curve describes static device behavior. Identification of key metrics for evaluating the quality of a transfer curve allows informed device design that optimizes performance. In Chapter 4, I address the issue of ensuring that transcriptional devices work in combination by presenting a framework for developing standards for functional composition. The framework relies on additional measures of device performance, such as error rate and the operational demand that the device places on the cellular chassis, to proscribe device signal thresholds. Digital devices must have compatible signal thresholds to work in combination. In Chapter 5, I describe a proof-of-principle implementation of a transcriptional inverter based on a synthetic transcription factor derived from a zinc finger DNA binding domain and a leucine zipper dimerization domain. Zinc fingers and leucine zippers offer a potential scalable solution to the challenge of building libraries of transcription-based logic devices for arbitrary information processing in cells. Finally in Chapter 6, I extend the principle of physical composition standards from parts and devices to the vectors that propagate those parts and devices. The new vectors support the assembly of biological systems. Taken together, the work advances the transformation of biological system design from an *ad hoc*, artisanal craft to a more predictable engineering discipline.

Chapter 2

Eau d'E coli: A synthetic biology approach to reprogramming bacterial odor

This chapter is based on a manuscript that I co-wrote with Stephen T. Payne and Drew Endy (Payne *et al.*, submitted). The initial project conception was mine. The research work was done primarily by a very talented undergraduate Stephen T. Payne, in collaboration with Veena Venkatachalam, Kate Broadbent, Delbert Green II, and Boyuan Zhu, and supervised by Barry Canton, Austin J. Che, Jason R. Kelly, Samantha C. Sutton, Thomas F. Knight, Jr., Drew Endy, and myself.

2.1 Summary

The underlying goal of synthetic biology is to make the design, construction, and characterization of engineered biological systems easier. Here, we evaluate whether synthetic biology approaches can support the process of metabolic engineering. As a model problem, we chose to reprogram the odor of *Escherichia coli*. We first designed and produced a bacterial chassis with reduced fecal odor. Then, by applying fundamental engineering principles such as abstraction and standardization, we (1) implemented wintergreen and banana odorant generators that use exogenously sup-

plied precursors, (2) developed transcriptional control devices for exponential and stationary phase protein production, and (3) combined a stationary phase transcriptional control device with a banana odorant generator to produce banana odor in a growth phase-dependent manner. Our results also confirm that the enzymes that produce odorants can serve as reporters of gene expression, complementing existing genetically-encoded reporters such as β-galactosidase, fluorescent proteins, and luciferases.

2.2 Introduction

Metabolic engineers have demonstrated successful construction of novel biosynthetic pathways in industrial microorganisms for the purpose of producing commercially useful compounds [8, 9, 10, 11, 12, 13]. However, such engineering feats require huge investments of labor, time, and capital by world-renowned genetic engineers [14, 15, 16]. Typically, such large resource investments are only justified when the product is overwhelmingly compelling from an industrial or medical perspective.

The underlying goal of synthetic biology is to make the design, construction, and characterization of engineered biological systems easier. A proposed approach advocates applying fundamental engineering principles such as abstraction, standardization, and decoupling to the substrate of biology and the process of biological engineering [7]. However, the relevance of principles from classical engineering disciplines to biological engineering has not yet been fully explored.

We examined how the ideas of abstraction and standardization can be applied to metabolic engineering problems. Abstraction is an approach for managing complexity by hiding unnecessary detail [7, 17]. An abstraction hierarchy is organized around a set of functional layers (Figure 2-1). For synthetic biology, the lowest abstraction layer is currently defined at the level of primary nucleic acid sequences. Moving up one level in the abstraction hierarchy, parts are defined as nucleic acid sequences that encode basic biological functions such as a transcriptional promoter or enzyme coding sequence [18]; parts are specified by their innate biological function(s), while details

regarding nucleic acid sequence remain hidden. At the next level in the abstraction hierarchy, devices are defined as human-defined functions that can be realized via a combination of one or more parts. Devices are specified in terms of their inputs and outputs, with details regarding all underlying parts hidden. Different classes of devices have different types of inputs and outputs. For example, transcriptional devices either receive transcriptional input(s), produce transcriptional output(s), or both. As a second example, biosynthetic devices convert one or more chemical precursor inputs to one or more chemical product outputs. At the topmost layer of the abstraction hierarchy, engineered biological systems are defined as combinations of devices that produce more powerful behaviors, such as a tumor-killing microbe [19]. Systems are specified by their overarching behavior, while details regarding the input/output relationships of the component devices remain hidden. By using an abstraction hierarchy, the behavior of an engineered biological system can be implemented as a combination of devices, each device can be defined in terms of its component parts, and each part can be specified by its primary nucleic acid sequence.

The utility of the abstraction hierarchy depends on standards that define how components in each layer of the hierarchy are combined and shared across layers. For example, standards for physical composition specify how parts physically connect; Knight developed the BioBrick standard for physical composition of genetic parts [20]. Using the BioBrick standard, the synthetic biology community has developed a collection of genetic parts in the Registry of Standard Biological Parts (Registry, `http://partsregistry.org`). All parts in the Registry can be readily assembled using the BioBrick assembly standard. As a second example, standards for functional composition specify how device inputs and outputs are functionally connected. A proposed signal standard for transcriptional devices is the rate at which RNA polymerases move past a particular point on a strand of nucleic acid, Polymerases Per Second (PoPS) [21, 22]. By using the PoPS signal standard, transcriptional devices can have one or more PoPS inputs, one or more PoPS outputs, or both. Thus, PoPS defines a standard, common signal carrier for transcriptional devices. In practice, while the goal of functional composition standards is to ensure that devices work reli-

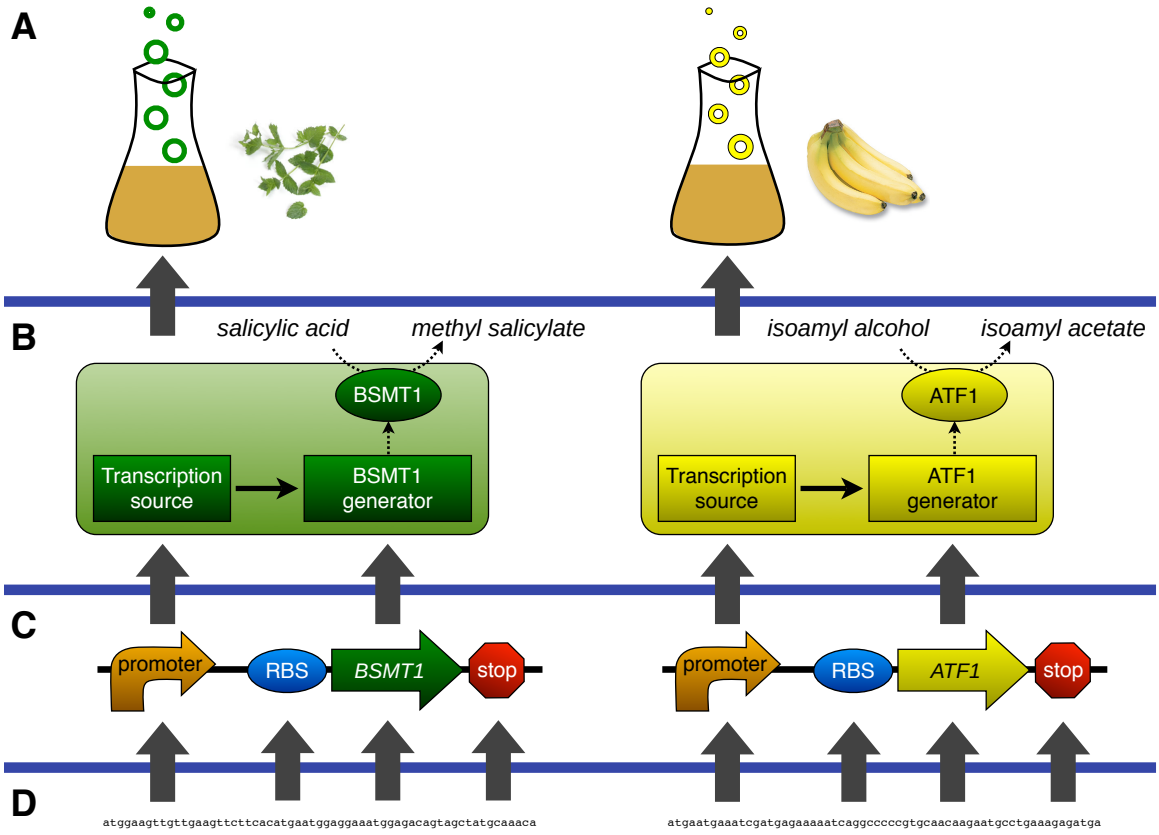


Figure 2-1: (A) Systems are defined by their overarching behavior, such as reprogramming bacterial odor. (B) A biosynthetic device, such as a wintergreen or banana odorant generator, takes one or more chemical inputs and produces one or more chemical outputs. Biosynthetic devices are made up of two transcriptional devices: a transcription source and a biosynthetic enzyme generator. A transcription source produces a transcriptional signal output. A biosynthetic enzyme generator takes as input a transcriptional signal and produces as output an enzyme, such as BSMT1 or ATF1, that catalyzes the conversion of a precursor to a product. (C) Biological parts are nucleic acid sequences that encode basic biological functions. A transcription source can be made up of a single part such as a promoter. An odorant enzyme generator is made up of three parts: a ribosome binding site (RBS), an enzyme coding sequence, and a transcriptional terminator (stop). (D) In synthetic biology, the lowest layer of the abstraction hierarchy is nucleic acid sequence.

ably and predictably when used in combination, only a few parts in the Registry and elsewhere have yet been shown to function reliably in combination [23, 24, 25]. Nevertheless, the modularity afforded by abstraction and standardization already enables parts and devices to be used as "off-the-shelf" components that can be independently tested, optimized for function as necessary, and improved over time.

As a model problem for exploring the relevance of synthetic biology to metabolic engineering, we chose to reprogram the odor of *Escherichia coli*. Odorants are volatile chemicals that have an odor or smell detectable by the human olfactory system [26]. For example, most flowers produce a complex array of odorants to generate a unique smell [27]. As a second example, the common laboratory chemical dithiothreitol has a distinctive "wet dog" odor [28]. As a final example, cultures of most laboratory strains of *E. coli* have a fecal odor. To reprogram bacterial odor, we had the option of eliminating natural odorants from *E. coli*, adding novel odorants to *E. coli*, or both. Eliminating natural odorants from *E. coli* requires modifying existing cellular biosynthetic pathways. Adding novel odorants to *E. coli* involves engineering new biosynthetic pathways for chemical production.

2.3 Results

2.3.1 Identifying an odor-free chassis

We named our project to reprogram bacterial odor, Eau d'E coli. The first challenge in the project was to ensure that the natural, fecal odor of *E. coli* did not overpower our engineered odors. Indole was suggested to be the primary contributor to the fecal odor of *E. coli* (E Pichersky, personal communication, 2006). We confirmed that indole is the primary odorant produced by *E. coli* by smelling LB Lennox medium supplemented with indole at a concentration comparable to that produced in LB cultures of *E. coli* strain MG1655 ($\sim$300 μM) [29]. LB medium supplemented with indole smelled similar to typical *E. coli* laboratory cultures. In nature, *E. coli* uses indole for intercellular signaling in biofilm formation [30, 31, 32]. There are also reports of

indole playing a role in multidrug exporter gene regulation and ColE1 plasmid maintenance [33, 34]. Regardless, since indole is not essential for cell viability, we could begin to reprogram bacterial odor by attempting to modify cellular metabolism for decreased indole production.

Indole is a product of the tryptophanase enzyme encoded by the *tnaA* gene of the *tna* operon in *E. coli* [35]. Mutations to the *tna* operon can reduce indole levels [36]. We tested four *E. coli* strains as potential odor-free chassis for the Eau d'E coli project: YYC912, JC12337, MEB61, and MB408 (CGSC 7602, CGSC 6373, CGSC 6836, and CGSC 7152, respectively, and provided by Mary Berlyn at The Coli Genetic Stock Center) [37, 38, 39]. The four strains all carry mutations in the *tnaA* gene. By smelling overnight liquid LB cultures of each strain, we determined that *E. coli* strain YYC912 likely did not produce indole. We confirmed via gas chromatography analysis that *E. coli* strain YYC912 produced no measurable indole in comparison to *E. coli* strain TOP10 (Figure 2-2) [40]. Thus, we selected *E. coli* strain YYC912 as an odor-free chassis for Eau d'E coli.

2.3.2 Engineering *E. coli* to smell like wintergreen and banana

Next, we started our work to add new odors to *E. coli* by considering what was already known about bacterial odorant production [41]. Several groups have demonstrated odorant production in *E. coli* during studies of enzymes involved in floral odorant emission [42, 43, 44, 45]. However, published reports do not state whether the *E. coli* cultures produced sufficient levels of odorant to actually smell. When contacted, the authors confirmed that expression of some enzymes resulted in *E. coli* cultures that had the expected floral odor (N Dudareva and E Pichersky, personal communication, 2006). Similarly, Horton *et al.* demonstrated production of the odorant isoamyl acetate in *E. coli* but did not report whether the isoamyl acetate-producing cultures had the expected banana odor [46]. From our initial survey of the literature, we considered five candidate odorants that might be added to *E. coli* (Table 2.1).

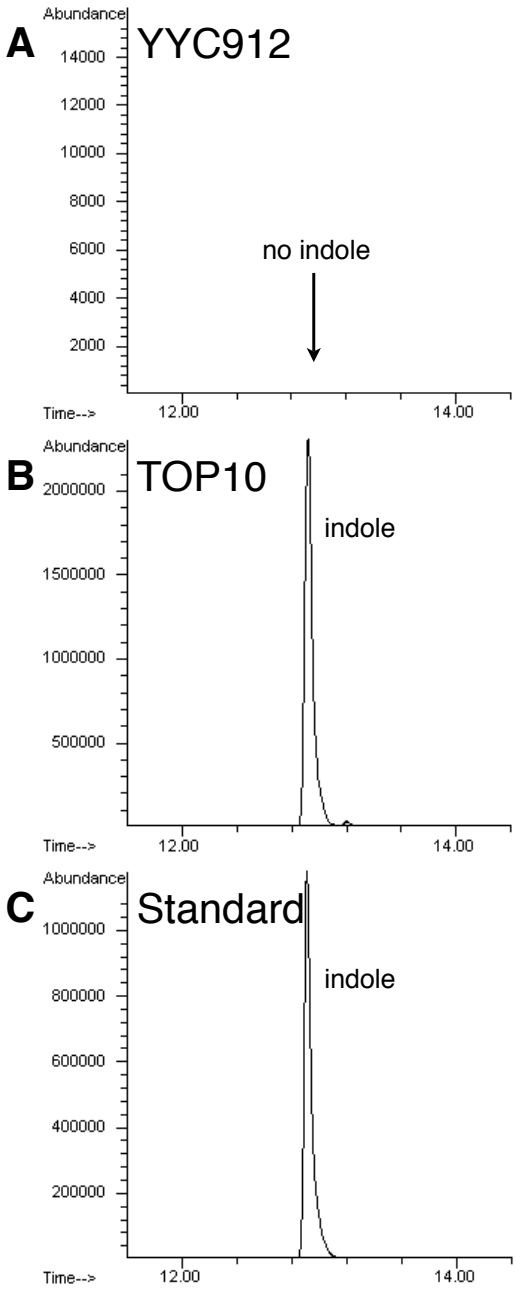


Figure 2-2: To verify the selection of *E. coli* strain YYC912 as an odor-free chassis, cultures of each strain for indole production were analyzed by gas chromatography. A) *E. coli* strain YYC912 had no detectable indole. B) *E. coli* strain TOP10 had high levels of indole. C) The retention time of the indole peak from *E. coli* strain TOP10 is identical to that of the pure indole standard.

Odorant	*Part no.*	*Enzyme*	*Organism*	*Evaluation*	*Reference*
Cinnamon	n/a	*CCMT*	*Ocimum basilicum*	No sequence available	[47]
Jasmine	BBa_J45003	*JMT*	*Arabidopsis thaliana*	Failed construction	[48]
Floral	BBa_J45002	*BAMT*	*Antirrhinum majus*	No odor	[49] [50]
Wintergreen	BBa_J45005	*SAMT*	*Clarkia breweri*	No source DNA	[44]
	BBa_J45001	*SAMT*	*Antirrhinum majus*	Works	[45]
	BBa_J45004	*BSMT1*	*Petunia* × *hybrida*	Works; selected	[51] [52]
Banana	BBa_J45014	*ATF1*	*Saccharomyces cerevisiae*	Works; selected	[53] [46]

Table 2.1: Candidate odorants and enzymes for reprogramming bacterial odor

We selected wintergreen and banana odorant production for the initial Eau d'E coli system.

We constructed two biosynthetic devices for odorant production in *E. coli* (Figure 2-1). Biosynthetic devices catalyze the conversion of one or more chemical precursors to one or more chemical products. For example, an odorant generator is a biosynthetic device that catalyzes the conversion of a precursor to an odorant. The wintergreen odorant generator (BBa_J45120) is based on the *S*-adenosyl-L-methionine:benzoic acid/salicylic acid carboxyl methyltransferase I (*BSMT1*) gene from *Petunia* × *hybrida* (plasmid encoding *BSMT1* provided by Natalia Dudareva, Department of Horticulture and Landscape Architecture, Purdue University) [51, 52]. BSMT1 catalyzes the conversion of the precursor salicylic acid to the odorant methyl salicylate, which has a wintergreen odor. The banana odorant generator (BBa_J45200) is based on the *Saccharomyces cerevisiae* alcohol acetyltransferase I (*ATF1*) gene [53, 46, 54]. ATF1 catalyzes the conversion of the precursor isoamyl alcohol to the odorant isoamyl acetate, which has a banana odor. Each biosynthetic device is made up of two transcriptional devices: a transcription source and an odorant enzyme generator. Transcription sources, such as promoters, produce a tran-

Description	*Salicylic acid*	*Retention time*	*Abundance*
Wintergreen odorant generator (BBa_J45120)	+	11.471	1.0×10^8
	−	11.473	3.3×10^4
Methyl salicylate standard	n/a	11.461	2.1×10^8
E. coli strain TOP10	+	11.347	2.2×10^4

Table 2.2: Wintergreen odorant generator produces methyl salicylate.

Description	*Isoamyl alcohol*	*Retention time*	*Abundance*
Banana odorant generator (BBa_J45200)	+	4.456	4.7×10^8
	−	4.400	3.6×10^6
Isoamyl acetate standard	n/a	4.388	9.6×10^7
E. coli strain TOP10	+	4.432	8.0×10^5

Table 2.3: Banana odorant generator produces isoamyl acetate.

scriptional signal output. Odorant enzyme generators take as input a transcriptional signal and produce as output an enzyme that catalyzes production of an odorant from a chemical precursor. All transcriptional devices in this work use PoPS as a common signal carrier.

To confirm that our biosynthetic devices produced the correct odorants we analyzed the *E. coli* cultures for odorant production by gas chromatography. The wintergreen odorant generator (BBa_J45120) produced high levels of methyl salicylate when the precursor salicylic acid was added to the culture medium (Table 2.2, Figure 2-3). The cellular chassis, *E. coli* strain TOP10, did not produce methyl salicylate in the presence of exogenous salicylic acid, demonstrating that our biosynthetic device was indeed responsible for methyl salicylate production. Similarly, the banana odorant generator (BBa_J45200) produced high levels of isoamyl acetate when the precursor isoamyl alcohol was added to the culture medium, whereas the cellular chassis did not (Table 2.3, Figure 2-4).

A blind smell test demonstrated that we had successfully reprogrammed the odor of bacteria. Smell test participants distinguished between cultures producing wintergreen odorant, banana odorant, or the natural fecal odorant of *E. coli*. Of the 116 respondents, 64% were able to correctly identify the culture producing methyl sali-

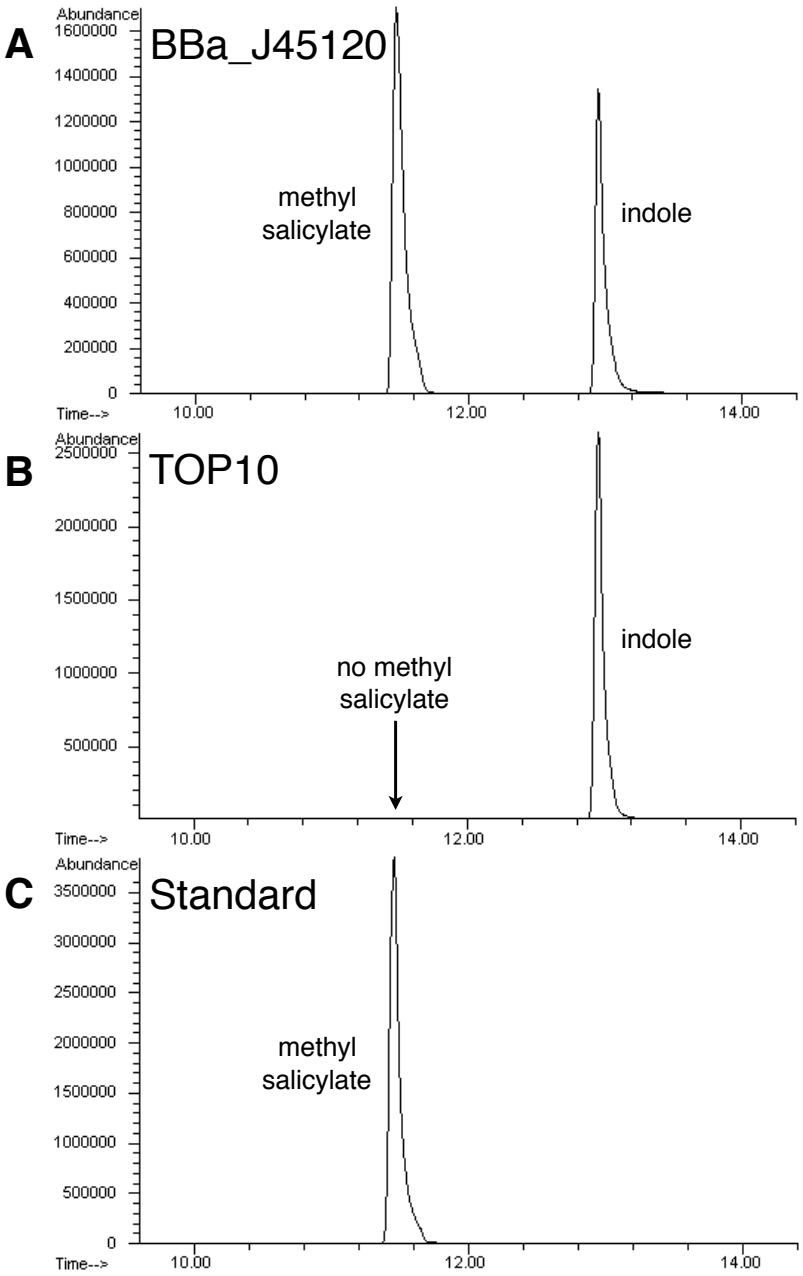


Figure 2-3: To confirm that the wintergreen odorant generator produced methyl salicylate, cultures with the device supplemented with 2 mM salicylic acid were analyzed by gas chromatography. (A) The wintergreen odorant generator (BBa_J45120) produced high levels of methyl salicylate when the precursor salicylic acid was added to the culture medium. (B) The cellular chassis alone (*E. coli* strain TOP10) did not produce methyl salicylate, although salicylic acid was added to the culture medium. (C) The retention time of the methyl salicylate peak from the wintergreen odorant generator (BBa_J45120) is identical to that of the pure methyl salicylate standard. Most *E. coli* strains produce indole.

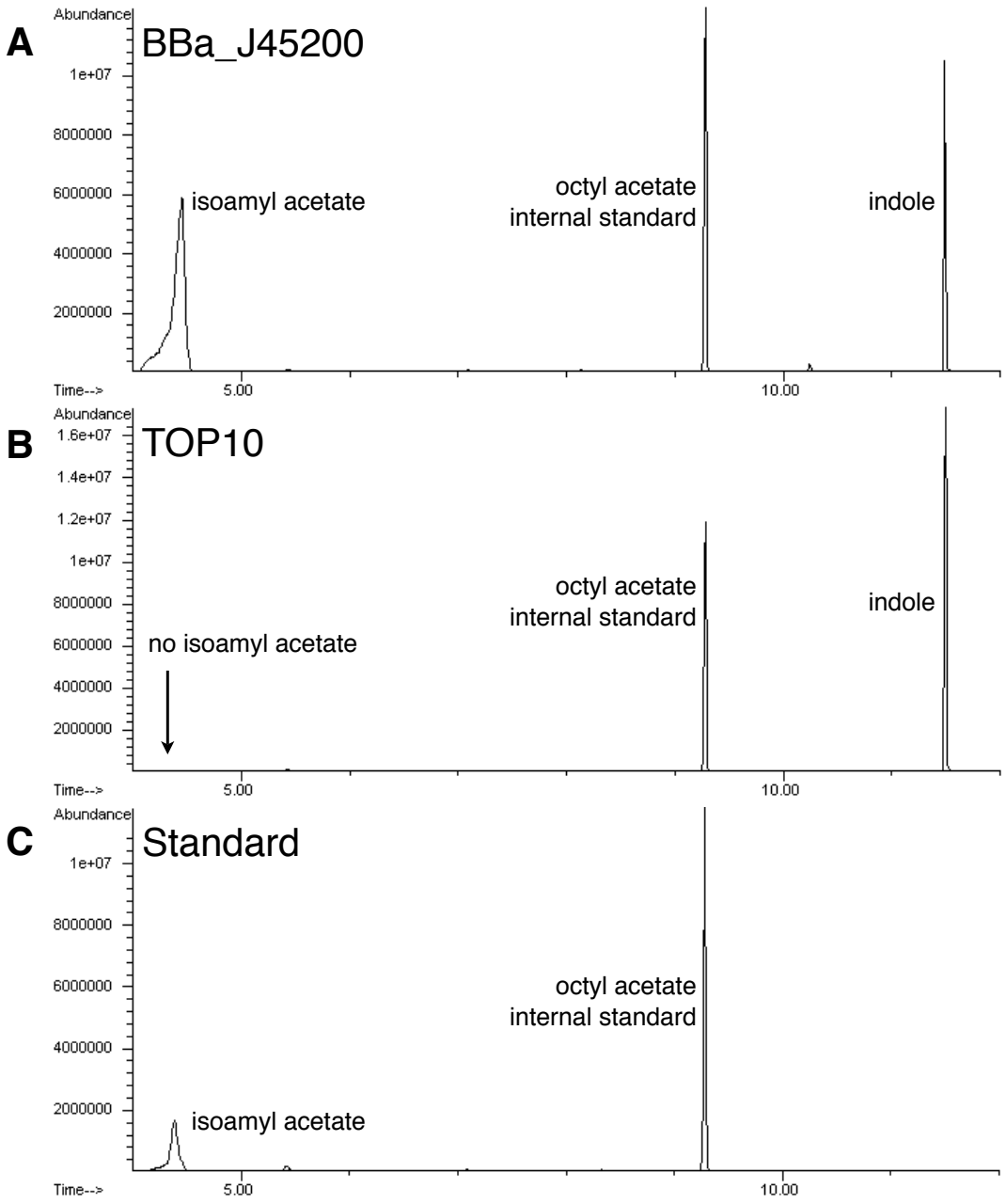


Figure 2-4: To confirm that the banana odorant generator produced isoamyl acetate, cultures with the device supplemented with 5mM isoamyl alcohol were analyzed by gas chromatography. (A) The banana odorant generator (BBa_J45200) produced high levels of isoamyl acetate when the precursor isoamyl alcohol was added to the culture medium. (B) The cellular chassis alone (*E. coli* strain TOP10) did not produce isoamyl acetate, although isoamyl alcohol was added to the culture medium. (C) The retention time of the isoamyl acetate peak from the banana odorant generator (BBa_J45200) is identical to that of the pure isoamyl acetate standard. Most *E. coli* strains produce indole. Octyl acetate was used as an internal standard for all samples containing isoamyl acetate.

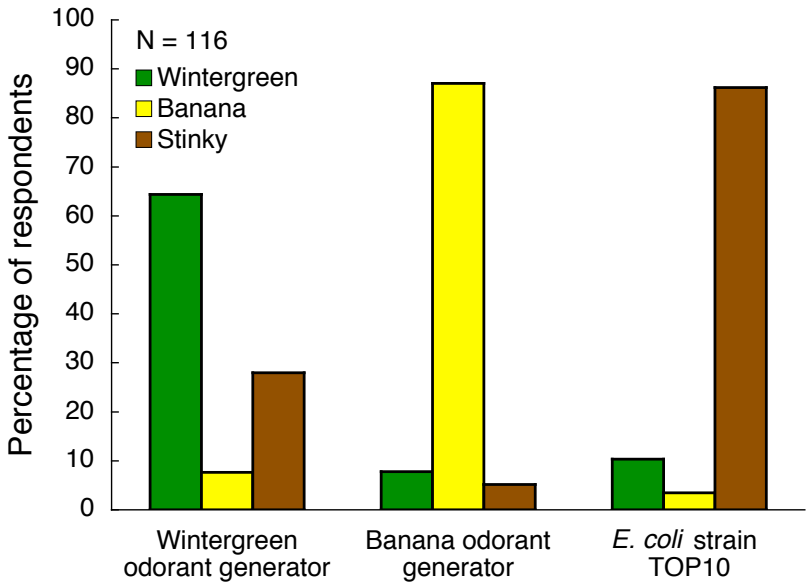


Figure 2-5: During a blind smell test at the 2006 iGEM Jamboree (`http://www.igem2006.com`), participants smelled cultures of the wintergreen odorant generator, the banana odorant generator, and *E. coli* strain TOP10. Participants were asked to characterize each culture as smelling like wintergreen (green bars), banana (yellow bars), or the natural fecal odor of *E. coli* (brown bars). Based on the survey results, people can smell the odorant from both odorant generators (Pearson's chi-square test yields p ¡ 0.01). For the smell test, the odor-free chassis (*E. coli* strain YYC912) was used for the wintergreen and banana odorant generators.

cylate through its wintergreen odor, 87% were able to correctly identify the culture producing isoamyl acetate through its banana odor, and 86% were able to correctly identify the laboratory *E. coli* strain TOP10 through its fecal odor (Figure 2-5). Both the wintergreen and banana odorant generators were propagated in the odor-free chassis for the smell test. Based on the survey results, humans can smell the odorant produced by both odorant generators (Pearson's chi-square test yields $p < 0.01$).

2.3.3 Engineering growth-dependent transcriptional control devices

We sought to extend the initial Eau d'E coli system by developing and demonstrating that odorant production could be regulated and, in turn, used as a genetically

encoded reporter of cell state. Specifically, we sought to engineer *E. coli* to produce one odorant during exponential growth and a different odorant during stationary phase. We use the terms exponential growth and stationary phase practically; we define exponential growth as the period of culture during which cells are growing and dividing and stationary phase as the subsequent period during which cells undergo little or no growth. In our smell tests of the constitutive odorant generators, the culture producing banana odorant had a stronger odor than the culture producing wintergreen odorant. Thus, we opted to design a system to produce wintergreen odorant only during exponential growth and, as the culture transitions to stationary phase, wintergreen odorant production should plateau or decrease, while banana odorant production begins. We predicted that batch cultures of such cells would initially smell like wintergreen, and then the banana odor would overpower any residual wintergreen odor.

We considered different designs for exponential and stationary phase regulation of odorant production. All designs focused on using transcriptional control devices to regulate the odorant enzyme generators and thus odorant production. To start, we noted that several *E. coli* promoters that are primarily active in stationary phase have been previously characterized [55, 56]. We evaluated two stationary phase promoters as potential transcriptional control devices. First, we tested the promoter that controls transcription of *rpoS* in *E. coli* (plasmid pBS-rrnBTrpoSpUV provided by Masayasu Mie and Masuo Aizawa, Department of Biological Information, Tokyo Institute of Technology) [57]. The *rpoS* gene encodes σ^S factor, a transcription factor known to be present at increased levels during late exponential phase and early stationary phase [58, 59]. Funabashi *et al.* previously demonstrated that cells with an *rpoS::GFP* fusion only showed fluorescence in stationary phase [60]. Second, we tested the promoter that controls transcription of *osmY* [61, 62]. Expression of *osmY* is dependent on σ^S *in vivo* [63]. Schellhorn *et al.* previously demonstrated that an *osmY::lacZ* fusion generated the highest transcriptional signal in stationary phase as compared to nine other σ^S-dependent promoter-*lacZ* fusions [55, 56]. In addition, the *osmY::lacZ* fusion generated only a small transcriptional signal during exponential

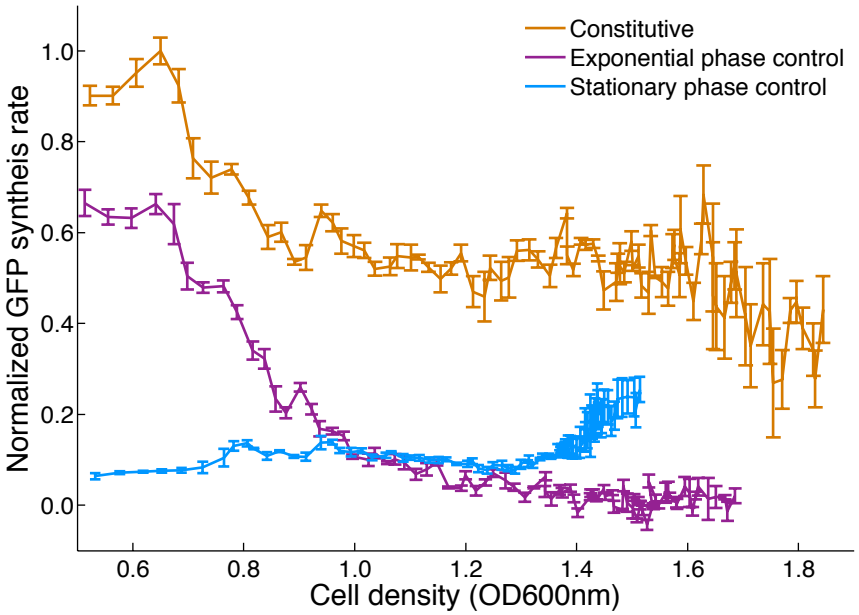


Figure 2-6: To test and verify function of the constitutive, stationary phase and exponential phase transcriptional control devices, each control device was assembled with the GFP generator (BBa_E0840), and the fluorescence of *E. coli* cultures with each device was monitored over time. A plot of the change in fluorescence per unit time (normalized GFP synthesis rate) versus the cell density (OD600nm) for each device is shown. The constitutive transcriptional control device produced a high GFP synthesis rate irrespective of cell density. The stationary phase transcriptional control device produced a low initial GFP synthesis rate which increased with culture cell density. The exponential phase transcriptional control device produced an initially high GFP synthesis rate which dropped off as cell density increased. Data shown are averages of triplicate measurements of cultures grown from three individual colonies of each device. Error bars are the 95% confidence interval of the mean of the three independent cultures.

growth. We constructed both a short and long version of the *osmY* promoter. The short *osmY* promoter (BBa_J45993) consisted of only 57 base pairs encompassing the -35 and -10 promoter regions, while the long *osmY* promoter (BBa_J45992) included 199 base pairs [62]. Preliminary tests demonstrated that only the long *osmY* promoter met our requirements for control device function: it produced a low GFP synthesis rate during exponential growth and a high GFP synthesis rate in stationary phase (Figure 2-6). Thus, we selected the long *osmY* promoter as our stationary phase transcriptional control device.

Since we had already engineered a stationary phase transcriptional control device (BBa_J45992) and we had access to an "off-the-shelf" working transcriptional inverter (BBa_Q04401) from the Registry [64], we opted to construct an exponential phase transcriptional device by combining the *osmY* promoter and transcriptional inverter (an inverter is a device that converts a HIGH input signal to a LOW output signal and vice versa). The resulting composite exponential phase device (BBa_J45994) worked well when tested: the device only produced a high GFP synthesis rate in exponential phase (Figure 2-6). As expected, the timing of the exponential and stationary phase devices are well-coordinated, with the GFP synthesis rate of the exponential phase device decreasing just as the GFP synthesis rate of the stationary phase device increases. Our reuse of a preexisting transcriptional inverter from the Registry saved us considerable effort in constructing an exponential phase control device. Moreover, functional composition of the stationary phase promoter and transcriptional inverter yielded an exponential phase transcriptional control device that worked as designed.

2.3.4 Growth-dependent regulation of odorant production

To enable growth-dependent regulation of odorant production, we used the engineered exponential and stationary phase control devices to control the wintergreen and banana odorant enzyme generators, respectively. Since reliable functional composition of genetically-encoded devices remains a challenge [23], we could not assume that the transcriptional control devices would properly regulate the odorant enzyme generators simply because they correctly regulated a GFP test device. Thus, to evaluate the function of the constitutive, exponential and stationary phase odorant generators, we quantified the odorant production of cultures at different cell densities using gas chromatography. The exponential phase wintergreen odorant generator (BBa_J45181) produced methyl salicylate, but its methyl salicylate levels were indistinguishable from the constitutive device during stationary phase (Figure 2-7). In contrast, the stationary phase banana odorant generator (BBa_J45250) worked as designed: the composite device produced little isoamyl acetate at low cell densities and more isoamyl acetate in stationary phase (Figure 2-8). As a comparative control,

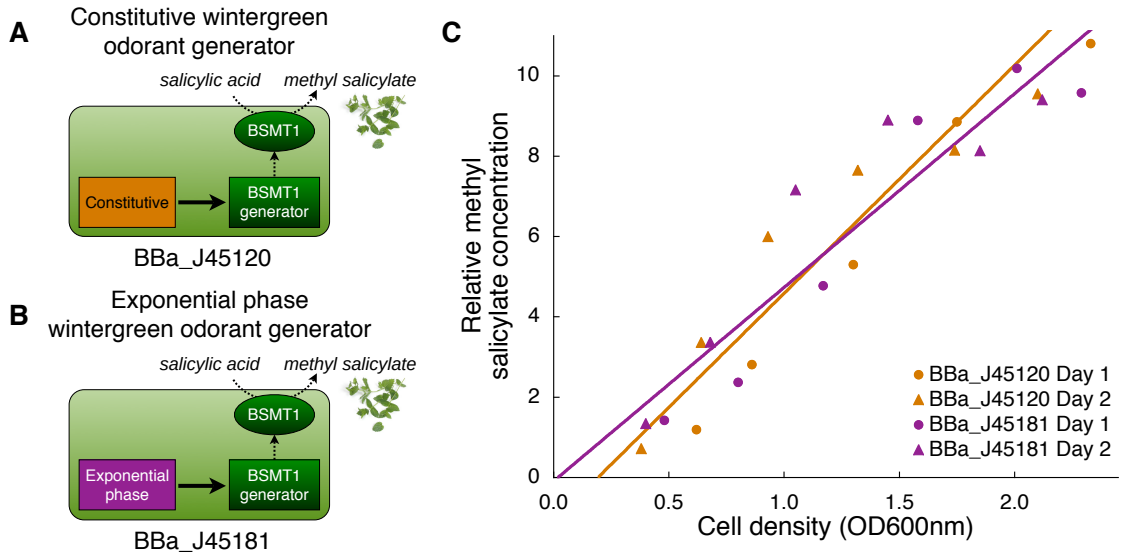


Figure 2-7: (A) The constitutive wintergreen odorant generator (BBa_J45120) is made up of the constitutive transcriptional control devices and the BSMT1 enzyme generator. (B) The exponential phase-dependent wintergreen odorant generator (BBa_J45181) is made up of the exponential phase transcriptional control device and the BSMT1 enzyme generator. (C) To demonstrate growth phase-dependent wintergreen odorant production, relative methyl salicylate concentrations of cultures of the constitutive and exponential phase-dependent wintergreen odorant generators were measured at different cell densities (OD600nm). The constitutive and exponential phase wintergreen odorant generators produced similar levels of methyl salicylate at all cell densities examined. We conducted two independent experiments (days 1-2). To aid visual comparison of the two odorant generators, a linear fit to the data for each device is shown.

the constitutive banana odorant generator (BBa_J45200) produced isoamyl acetate across all cell densities.

2.4 Discussion

2.4.1 Successes in reprogramming bacterial odor

There are five successes from our work that are worth noting. First, we identified a bacterial chassis for odorant production that is free of the natural, fecal odor of most *E. coli* strains. The odor-free chassis is useful for ensuring that the natural odor of *E.*

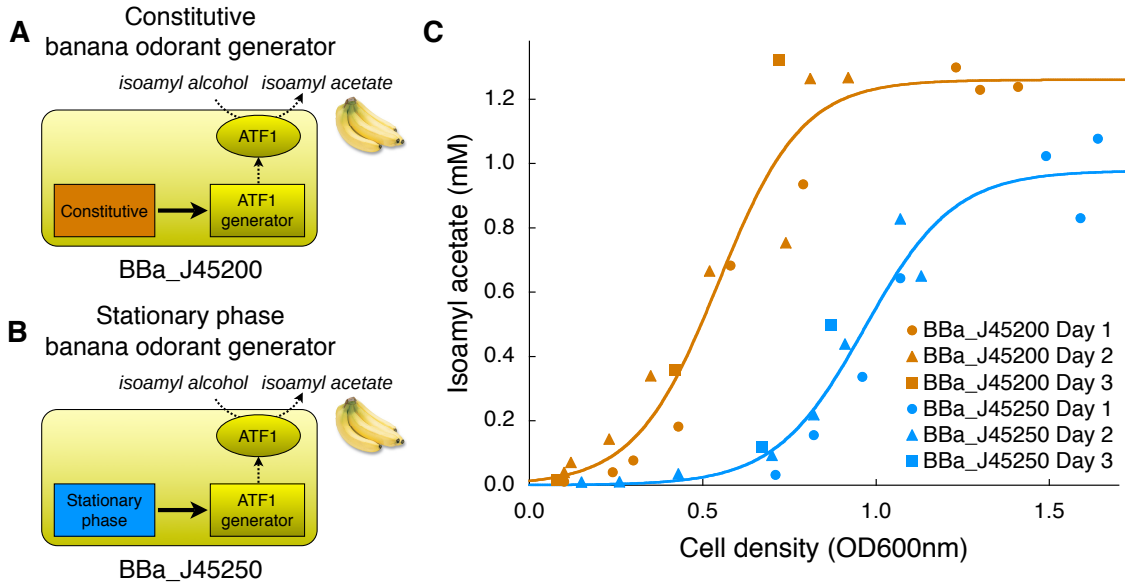


Figure 2-8: (A) The constitutive banana odorant generator (BBa_J45200) is made up of the constitutive transcriptional control device and the ATF1 enzyme generator. (B) The stationary phase-dependent banana odorant generator (BBa_J45250) is made up of the stationary phase transcriptional control device and the ATF1 enzyme generator. (C) To demonstrate growth phase-dependent banana odorant production, isoamyl acetate concentrations of cultures of the constitutive and stationary phase banana odorant generators were measured at different cell densities (OD600nm). As expected, the stationary phase banana odorant generator produced very little isoamyl acetate at low cell densities but its isoamyl acetate production increased with cell density. By comparison, the constitutive banana odorant generator produced more isoamyl acetate at lower cell densities than the stationary phase banana odorant generator. We conducted three independent experiments (days 1-3). To aid visual comparison of the two odorant generators, an empirical fit to the data for each device is shown.

coli does not overpower any engineered odors. Second, we implemented wintergreen and banana odorant generators that use exogenously supplied precursors. Blind smell tests demonstrated that most people can smell the wintergreen and banana odorants produced in culture. Third, we successfully engineered exponential and stationary phase transcriptional control devices. The exponential and stationary phase control devices can be combined with a GFP generator in order to produce regulated, growth-dependent protein production. Fourth, we combined the stationary phase transcriptional control device with the banana odorant enzyme generator to produce regulated banana odorant production. Finally, taken together, our results demonstrate that odorant-producing enzymes can serve as genetically-encoded reporters of gene expression. Odor-based reporters complement existing optically-based reporters such as β-galactosidase, fluorescent proteins, and luciferases. Furthermore, odor-based reporters may prove useful in situations in which direct culture sampling and measurement is difficult, such as industrial fermentation where off-gas analysis by gas chromatography is already common [65].

2.4.2 Failures in reprogramming bacterial odor

In designing a system for producing wintergreen odorant during exponential growth and banana odorant in stationary phase, we were unable to regulate wintergreen odorant production in a growth-dependent fashion. The exponential phase wintergreen odorant generator (BBa_J45181) did produce wintergreen odorant when cultures were supplemented with salicylic acid, but the exponential device produced methyl salicylate levels indistinguishable from the constitutive device during stationary phase (Figure 2-7). Furthermore, methyl salicylate production was roughly linear in cell density. There are two possible explanations for the experimental results. First, assuming that the control devices regulated the BSMT1 generator similarly to the GFP test device, then the experimental results suggest that enzyme concentration was not rate-limiting in methyl saliylate production. Instead, the substrate salicylic acid or the cofactor *S*-adenosyl-L-methionine may be limiting. Although rough estimates of absolute methyl saliclylate levels suggest that at most $\sim 10\%$ of the exogenously

supplied precursor is consumed in the assay, the intracellular concentration of salicylic acid could be limiting. Alternatively, the constitutive and exponential phase transcriptional control devices may regulate the BSMT1 generator differently than the GFP test device. Either both control devices maintained BSMT1 expression in stationary phase, or both devices turned off BSMT1 expression. In other words, functional composition of the transcriptional control devices with the BSMT1 generator failed. We cannot definitively exclude either explanation based on the data.

2.4.3 Application of synthetic biology approaches to metabolic engineering.

In drawing lessons from our experiences in reprogramming bacterial odor, it is worth considering what is unique to the work and what may apply more generally to metabolic engineering. In most classical metabolic engineering projects, the goal is to produce a particular chemical at a target yield as defined by a specific application [66, 67]. In contrast, the goal of Eau d'E coli was to engineer bacterial odorants to change the odor of cultures in a regulated fashion. Thus, we were able to choose which odorants to produce based on a preliminary evaluation of different candidate odorants. Moreover, since the human olfactory system is known to be quite sensitive [68, 69], production of just 0.3 μM methyl salicylate or 20 nM isoamyl acetate can be sufficient to smell and therefore constitute successful implementation of our engineered system (`http://www.leffingwell.com/odorthre.htm`) [70]. However, as a practical aside, increased isoamyl acetate production by industrial microbes does have commercial applications in the food flavoring industry [46].

In metabolic engineering, most engineered biosynthetic systems are built from *ad hoc* collections of genetic components that can be assembled, tested, and used in screens or selections if necessary. In this work, we instead used abstraction to systematically compile overall system function, reprogramming bacterial odor, into two biosynthetic devices. The biosynthetic devices were in turn compiled to transcriptional devices, and then BioBrick standard biological parts, and finally nucleic acid

sequences. Thus, abstraction provides an approach for systematically mapping high-level system behavior, such as reprogramming bacterial odor, to low-level primary sequence data. As a result, abstraction allows biological engineers to cope with the complexity of engineering multi-component synthetic biological systems; as the number of components in engineered biological systems increases from a few dozen parts to hundreds of interacting devices, abstraction will become even more important.

Meanwhile, standards that support the physical composition of genetic parts make construction of many-component, engineered biological systems, including metabolic engineering projects, both easier and faster. For example, in the Eau d'E coli project, we used the BioBrick standard for physical composition of genetic parts. Our use of the BioBrick physical composition standard offered four advantages over classical molecular cloning approaches. First, our use of a uniform part assembly procedure reduced the learning curve associated with system construction. Making construction easier was critical for our team of novice biological engineers to begin construction of Eau d'E coli devices quickly despite limited prior research experience. Second, standardization of the assembly procedure tends to make the device and system construction process more reliable since the same reagents and protocols are used at each stage. Third, since our system was constructed of BioBrick parts, we could readily reuse preexisting parts from the Registry in our system design. For example, we reused a promoter, a ribosome binding site, transcriptional terminator, GFP generator, and transcriptional inverter from the Registry (BBa_R0040, BBa_B0032, BBa_B0015, BBa_E0840, and BBa_Q04401). Each reused part could be readily combined "off-the-shelf" with our newly constructed parts because all parts adhered to the BioBrick assembly standard. Our reuse of parts resulted in significant time and effort savings since we did not have to develop parts *de novo* or redesign reused parts. Fourth, the parts that encode growth-dependent transcriptional regulation and odorant production are now freely available to the community via the Registry of Standard Biological Parts for reuse and improvement (Table 2.4).

Functional composition builds upon physical composition: parts must not only be readily connected but should also function as expected. Just as standards for physical

Part number	*Description*	*Source*	*Works?*
BBa_J45004	*BSMT1*: converts salicylic acid to wintergreen odorant	*Petunia* × *hybrida*	Yes
BBa_J45014	*ATF1*: converts isoamyl alcohol to banana odorant	*S. cerevisiae*	Yes
BBa_R0040	constitutive promoter	Registry	Yes
BBa_J45992	stationary phase-dependent transcriptional control device	*E. coli osmY* promoter	Yes
BBa_Q04401	*tetR* transcriptional inverter	Registry	Yes
BBa_J45994	exponential phase-dependent transcriptional control device	composite	Yes
BBa_J45995	stationary phase-dependent GFP generator	composite	Yes
BBa_J45996	exponential phase-dependent GFP generator	composite	Yes
BBa_I7100	constitutive GFP generator	Registry	Yes
BBa_B0015	part without GFP	Registry	Yes
BBa_J45119	BSMT1 enzyme generator	composite	Yes
BBa_J45199	ATF1 enzyme generator	composite	Yes
BBa_E0840	GFP generator	Registry	Yes
BBa_J45120	constitutive wintergreen odorant generator	composite	Yes
BBa_J45181	exponential phase-dependent wintergreen odorant generator	composite	No
BBa_J45200	constitutive banana odorant generator	composite	Yes
BBa_J45250	stationary phase-dependent banana odorant generator	composite	Yes

Table 2.4: BioBrick standard biological parts for reprogramming bacterial odor.

composition ensure that any two parts that adhere to a physical composition standard can be readily combined, standards for functional composition ensure that any two parts or devices function as expected when combined. Today, we lack adequate standards to ensure reliable functional composition; the only proposed standard for functional composition is the use of PoPS as a common signal carrier for transcriptional devices. The PoPS standard ensures that the output(s) of one transcriptional device can be connected to the input(s) of another PoPS-based device. Additional standards, such as prescribed PoPS signal ranges that ensure transcriptional device signal levels are well-matched, are needed so that devices can be developed to meet proscribed functional specifications, and evaluated for the reliability of their use in combination.

In the absence of sufficient standards for functional composition, we relied on trial-and-error in building the Eau d'E coli system. Such *ad hoc* approaches can yield success. For example, to construct an exponential phase transcriptional control device, we combined a stationary phase promoter with an "off-the-shelf" transcriptional inverter. Both devices were independently characterized, and we successfully combined them to demonstrate growth-phase dependent GFP production (Figure 2-6). Similarly, to construct a stationary phase banana odorant generator, we combined a stationary phase PoPS source with the banana odorant enzyme generator. Again, both devices had been independently characterized and behaved as expected when combined (Figure 2-8). Nevertheless, functional composition of transcriptional devices remains challenging. For example, when we combined the exponential phase control device with the wintergreen odorant enzyme generator, the resulting composite device produced methyl salicylate during exponential phase, but we did not measure lower methyl salicylate levels in stationary phase as compared to the constitutive wintergreen odorant generator (Figure 2-7).

Metabolic engineering depends on functional composition of not only transcriptional devices but also biosynthetic devices. For instance, the Eau d'E coli system could be further extended to produce odorants from endogenous cellular metabolites rather than supplied exogenous precursors. Such an extension would require addi-

tional biosynthetic devices that convert natural cellular metabolites to the odorant precursors salicylic acid and isoamyl alcohol. In fact, salicylate production from the cellular metabolite chorismate has already been shown in *E. coli* [71]. (Salicylate is the anion of salicylic acid; at intracellular pH, salicylic acid is primarily in its anion form.) Thus, we could construct biosynthetic device(s) that catalyze the conversion of chorismate to salicylate. By combining the salicylate generator with the wintergreen odorant generator, we could construct a complete wintergreen odorant biosynthetic system based on methyl salicylate production from cellular metabolites (Figure 2-9). Similarly, in the case of banana odor, isoamyl alcohol production from the cellular metabolite α-ketoisocaproate has recently been reported in *E. coli* [72]. Thus, we might similarly engineer biosynthetic device(s) that catalyze the conversion of α-ketoisocaproate to isoamyl alcohol and again combine the device(s) with the banana odorant generator to make a complete banana odorant biosynthetic system (Figure 2-9). The challenge in functional composition of biosynthetic devices is therefore the classical metabolic engineering challenge of matching or maximizing flux through the set of devices [73, 74, 75, 76]. It is interesting to speculate whether it would be useful or even possible to define a common signal carrier, such as flux, for biosynthetic devices. Could we then set standards for minimum and maximum fluxes through biosynthetic devices? Similarly, could biological engineers construct libraries of catabolic devices that degrade any number of feedstocks to a small set of core metabolites as well as libraries of anabolic devices that can convert those core metabolities to any number of useful chemicals? Although such steps pose great technical challenges, they also hold promise for further reducing the work needed to develop engineered biosynthetic systems.

2.5 Conclusions

Synthetic biology approaches are clearly not now sufficiently advanced to replace classical metabolic engineering techniques. Nevertheless, fundamental engineering principles, such as abstraction and standardization, can already enable rapid proto-

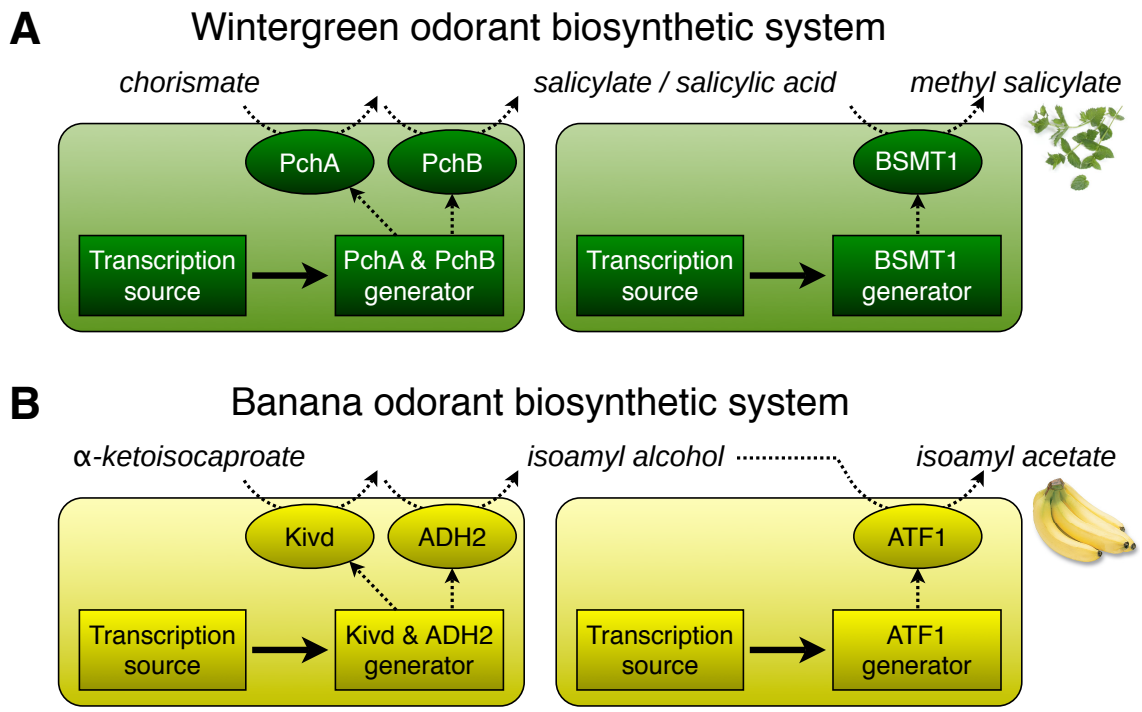


Figure 2-9: (A) The wintergreen odorant biosynthetic system is made up of two biosynthetic devices: a salicylate generator and a wintergreen odorant generator (BBa_J45120). The salicylate generator catalyzes the conversion of the cellular metabolite chorismate to salicylate via enzymes PchA and PchB. (Salicylate is the anion of salicylic acid; at intracellular pH, salicylic acid is primarily in its anion form.) The wintergreen odorant generator catalyzes the conversion of salicylic acid to methyl salicylate via the enzyme BSMT1. Methyl salicylate has a wintergreen odor. (B) The banana odorant biosynthetic system is made up of two biosynthetic devices: an isoamyl alcohol generator and a banana odorant generator (BBa_J45200). The isoamyl alcohol generator catalyzes the conversion of the cellular metabolite α-ketoisocaproate to isoamyl alcohol via enzymes Kivd and ADH2. The banana odorant generator catalyzes the conversion of isoamyl alcohol to isoamyl acetate via the enzyme ATF1. Isoamyl acetate has a banana odor.

typing of many-component, biosynthetic systems. One benchmark for the power of synthetic biology approaches is the annual international Genetically Engineered Machines (iGEM) competition (`http://igem.org`). In iGEM, teams of undergraduate students engineer synthetic biological systems of their own design. In fact, we reprogrammed bacterial odor as a part of the 2006 competition. The iGEM students are largely novice biological engineers, many of whom have little or no prior biological research experience. Using the BioBrick physical composition standard, the teams reuse standard genetic parts from the Registry and design any new parts needed for their engineered biological system. As new technical standards are developed, iGEM provides a convenient framework to further examine whether these standards make the process of engineering biology easier. We propose that synthetic biology approaches, when combined with classical metabolic engineering techniques, have the potential to dramatically reduce the resources needed to engineer biosynthetic systems. In particular, synthetic biology may expand access to metabolic engineering from a small number of expert labs and companies to a broader base of novice biological engineers.

Chapter 3

Design of transcription-based logic devices

3.1 Summary

To aid in the design of transcription-based logic devices, I constructed a physicochemical model of the simplest transcription-based logic device: an inverter composed of a homodimeric repressor and an operator to which it binds. The model serves as a tool to explore which parameters most strongly impact device performance and to identify target values for those biochemical parameters over which I have some control, such as repressor-DNA affinity, translation rate and copy number.

My modeling results lead to a few key observations:

1. Quantitative measures of device performance, as characterized by the transfer curve, include the trip point, swing, gain, and noise margins.

2. The lumped parameter α_i, defined as the product of the ratio of mRNA and protein synthesis to their decay rates and copy number, determines the device input protein swing, the range of input protein concentration over which the device operates, as well as the device fan out, the maximum number of outputs the device can drive.

3. The dissociation constants governing the binding equilibrium between monomers

and dimers and between unbound and bound operator DNA are the primary determinants of the shape of the device transfer curve. I obtain some estimates for target values of the dissociation constants.

4. The effects of explicit inclusion of nonspecific DNA binding on device performance must be compensated for by increased cooperativity of protein binding to operator DNA.

5. An alternate device design in which several nonfunctional, high-affinity protein binding sites are present yields a substantially improved transfer curve, as measured by the noise margin.

3.2 Introduction

Synthetic biology distinguishes itself from previous efforts in genetic engineering in its emphasis on genetically-encoded control of system behavior, among other things. Thus, in Chapter 2, the system specification was not merely to produce odorants but to do so in a regulated fashion. To achieve the desired system behavior, I used transcriptional regulation and in particular, transcription-based logic. Here, I explore the design of transcription-based logic devices in more detail.

Digital devices are devices that represent signals as being in one of two possible states: logical zero or logical one [77]. Digital devices receive one or more digital signals as input and produce one or more digital signals as output. Digital logic devices perform simple boolean logic operations, such as logical NOT, logical AND, and logical OR [78]. For example, an inverter implements a logical NOT operation by converting its input to the opposite output: logical zero input is converted to logical one output, and vice versa. An inverter is one of the simplest digital logic devices, because it receives only one input and produces only one output. Thus, inverters offer a useful model logic device through which to explore issues in device design.

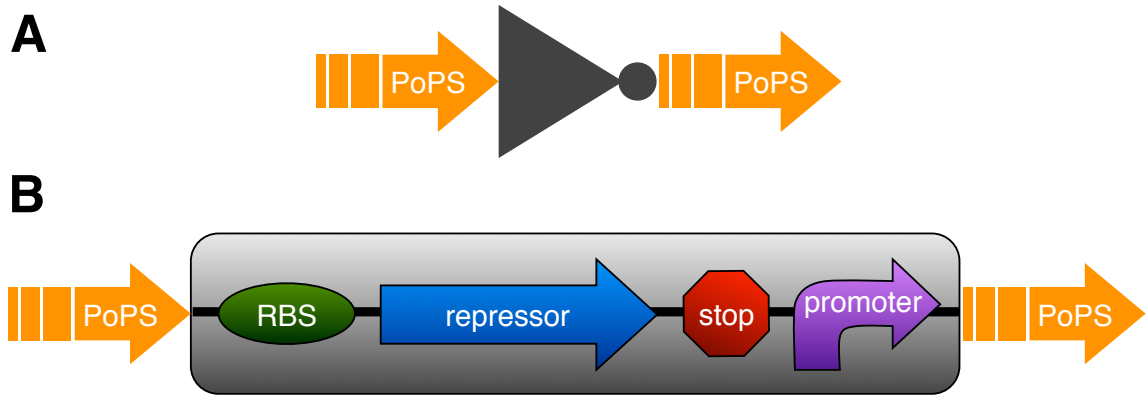


Figure 3-1: (A) An inverter receives an input, performs a logical NOT operation, and produces a corresponding output. In transcription-based inverters, both input and output signals are encoded as a transcription rate in units of PoPS. (B) An inverter is composed of four parts: a ribosome binding site (RBS), a coding sequence for a homodimeric repressor, a transcriptional terminator (stop), and a cognate promoter.

3.2.1 Representing digital signals in transcription-based devices

Transcription-based logic devices encode device signals as a transcription rate (Figure 3-1). Logical zero is represented as a low transcription rate, and logical one is represented as a high transcription rate. Transcriptional devices must be able to unambiguously represent the two signal states to function as digital devices. In transcription-based logic, an inverter is composed of four parts: a ribosome binding site (RBS), a coding sequence for a homodimeric repressor, a transcriptional terminator (stop), and a cognate promoter. A high input drives transcription of the repressor. Upon translation and folding, the repressor dimerizes and binds to a cognate operator DNA sequence within the promoter to turn off transcription (Figure 3-2). A low input, meaning no transcription of the repressor, results in the promoter producing a transcription output signal.

Since transcription rates are the basis for logical signals in transcription-based devices, it is important to specify the units of transcription rate. Here, the unit of measure for transcription rate is the rate at which RNA polymerase transcribes past a certain point on the DNA molecule called polymerases per second and abbreviated

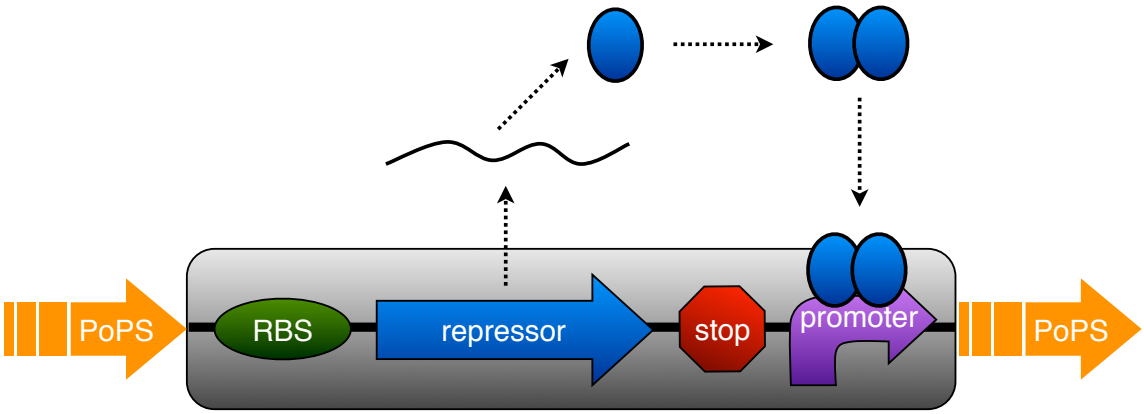


Figure 3-2: The input signal to a transcription-based inverter drives transcription of a gene encoding a repressor. The mRNA is translated to yield a repressor monomer. The repressor dimerizes and binds to its cognate promoter thereby regulating the output signal.

PoPS [21, 22]. I use to the symbol Π to denote a signal in units of PoPS. Π_i denotes the input signal to a device or the PoPS just before the ribosome binding site (at the device input). Π_o denotes the output signal from a device or the PoPS just after the promoter (at the device output).

A related but distinct parameter is the mRNA synthesis rate for a coding sequence or how many full length transcripts are synthesized from a coding sequence in a given period of time. This parameter is useful since it governs how many transcripts are produced from a DNA molecule per unit time. The value of the mRNA synthesis rate depends on the upstream promoter and the coding sequence itself. Since in general only the full length transcripts give rise to functional proteins, mRNA synthesis rate is often the parameter of interest in a system. Usually, the transcription rate and mRNA synthesis rate are treated as equivalent. However, in this work I draw a precise distinction between the two rates. Transcription rate, measured in PoPS, is unique to a particular position on the DNA molecule whereas the mRNA synthesis rate, measured in transcripts per second, is unique to a particular coding sequence on the DNA molecule. So the transcription rate at the end of a coding sequence should be equivalent to the mRNA synthesis rate, but the transcription rate at the beginning of a coding sequence may not be. (It is likely, however, that the transcription rate

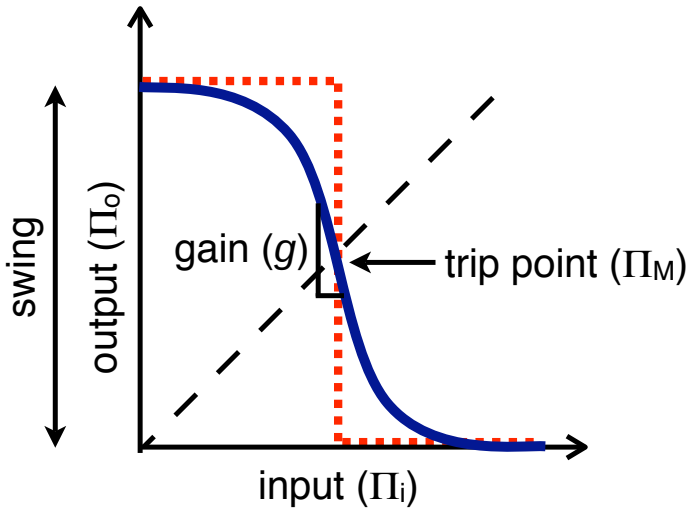


Figure 3-3: A transfer curve for an inverter device is a plot of the device output as a function of the device input. An ideal transfer curve has high output for all low inputs and low output for all high inputs (red dashed line). Real transfer curves approximate idea digital transfer curves (blue solid line). The swing, trip point, and gain are all features of the device transfer curve.

at the end of a coding sequence is proportional to the the transcription rate at the beginning.) In order for devices to be universally composable, the boundaries of the device must be those shown (Figure 3-1), and therefore the relevant units for input and output signals is PoPS.

3.2.2 Performance metrics for the device transfer curve

A transfer curve describes digital device behavior: it is a plot of device output(s) as a function of input(s) (Figure 3-3) [78]. In electrical engineering, various metrics have been developed to evaluate the quality of a device transfer curve [79, 80]. These metrics facilitate comparison of different device designs. Here, I focus on four measures of the device transfer curve: the swing, trip point, gain, and noise margin.

The swing is the range of PoPS values over which the device operates. For the device to operate properly, the swing should be much greater than the stochastic fluctuation in the signal or else the device will change states sporadically. For devices to behave well in series, the input and output swing should be comparable. Hence,

the swing is a quantitative parameter of the device transfer curve.

The trip point (also known as switch point or switching threshold) Π_M is the point at which the $\Pi_i = \Pi_o$. Ideally, Π_M should be approximately half the swing of the device, so that both noise margins (discussed below) have approximately equal size. Thus, the deviation of Π_M from this value is one metric for assessing the quality of a transfer curve.

The gain of a device is simply the slope of the transfer curve, denoted g here. (Note that the gain is defined as the absolute value of the slope so that the sign can be ignored.)

$$g = \left| \frac{d\Pi_o}{d\Pi_i} \right| \tag{3.1}$$

As mentioned before, for inputs in the valid low or valid high ranges, the device should suppress noise meaning that the absolute value of the slope of the transfer curve should be less than one ($g < 1$) and for the transition region, the gain is then necessarily greater than one ($g > 1$) to ensure that the device never produces an ambiguous signal. Unless otherwise specified, I use the term gain to refer to the slope of the transfer curve around the trip point.

Taken together, the swing, trip point, and gain offer useful measures of device performance. Each characteristic quantifies a particular aspect of the device transfer curve. Additionally, I can experimentally measure each characteristic either from the device transfer curve itself or independently from the device transfer curve. The drawback to these three metrics is that their relative importance to device performance is not clear. For example, if device A has a high gain but a very low trip point, how does it compare to device B which has lower gain but a trip point close to half the swing? Thus, in terms of both modeling device behavior and comparing different device designs, it is useful to have a single performance metric to assess device performance. Such a metric would enable direct comparison between different designs. The noise margin, discussed below, offers a single quantitative measure of device performance.

In logic devices, an important issue is that of noise. A key question is how much

noise can a device tolerate and still behave correctly. To cope with noise, a digital device must improve the quality of the signals it propagates. To do so, devices must suppress noise for signal values representing logical zero or one and amplify noise over the transition range. Such a transfer curve ensures that the device output will always reside in the logical zero or one state. The noise margin is "the maximum spurious signal that can be accepted by the device when used in a system whilst still giving correct operation" [81]. Noise which exceeds the noise margin can cause erroneous switching of state. As the device engineer, I want to maximize the noise margin so that the device is as robust as possible to signal fluctuation from either intrinsic (stochastic signal fluctuation) or extrinsic noise (variation in chassis components) [82]. Thus, an important quantitative measure of device performance is the size of the noise margin. Hill demonstrates a simple technique for calculating the noise margin [81, 83]. The method, called maximum square, defines the noise margin as the length of a side of the largest square that can be fit entirely within the loops created by two superimposed transfer curves (Figure 3-4). The noise margin, under this definition, specifies exactly how much noise the device can tolerate in the worst-case scenario of an infinite series of devices that experience noise at every input. The noise margin is a convenient measure of device performance, because it encapsulates the information in the trip point, gain or swing. A device with a poor swing, trip point, or gain will have a small noise margin. The primary drawback to the noise margin metric is that it cannot be directly measured experimentally but rather must be inferred from the transfer curve.

3.2.3 Modeling a transcriptional inverter

In order to gain a better understanding of the behavior of transcription-based logic devices, I developed a mathematical model that describes the behavior of these devices. The purpose of the model is to identify those parameters to which device behavior is most sensitive and to elucidate the relationship between these biochemical parameters and device performance, as captured by the transfer curve. The model should aid device design in two ways. First, it should provide target ranges for those parameters

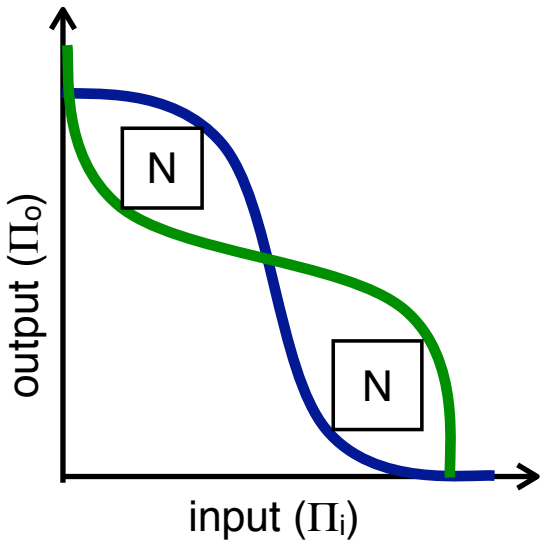


Figure 3-4: To compute the noise margin of a transcription-based inverter, superimpose two transfer curves for an inverter device. The blue curve is a plot of output (y axis) versus input (x axis), whereas the green curve is a plot of input (y axis) versus output (x axis). Then, draw the largest possible square inside each loop to obtain the maximum noise margin (N) of the device.

whose values I can control as the device engineer. Second, it should facilitate the study of alternate device designs that lead to improved device behavior.

There have been several previous efforts to model the behavior of simple logic devices [84, 85, 86, 25]. My work distinguishes itself from previous efforts in two key respects. First, as discussed already, I use transcription rate (PoPS) rather than protein concentration as the common signal carrier. The advantage of using transcription rate as the signal carrier is that devices are composable: any device may be connected to any other device. Devices whose output is protein concentration can only be connected to devices which take the same protein as input. The focus on composable devices is critical to developing general guidelines for device design. Second, most previous modeling efforts ask the question, given typical biological parameter values, what kind of device performance is expected? In this work, I instead ask, given that I as the device engineer have some measure of control over device design, how should I design the device in order to achieve the best possible device

performance? By approaching the model from a purely design perspective, I obtain somewhat different results.

In this chapter, I reserve the terms *input* and *output* to refer to the input signal and output signal respectively. I refer to the DNA encoding the repressor as the *device input* and the DNA encoding the promoter as the *device output*. Similarly, I will use the term *device input mRNA* to refer to the mRNA encoding the repressor and *device input protein* to refer to the repressor itself.

3.3 Model formulation

3.3.1 Assumptions

As in any model of a biological system, I make several simplifying (and possibly invalid) assumptions in deriving the model. These assumptions significantly impact the construction of the model and the device behavior that the model represents. I list here some of the more sweeping assumptions behind my model so that the model development is as transparent as possible.

1. I use reaction rate equations to model all reaction events meaning that the system is assumed to be continuous and deterministic in nature. In reality, within an *Escherichia coli* cell, molecular species are often present in small numbers and thought to be subject to stochastic fluctuation.

2. All binding events are at equilibrium. It is relatively common to treat the binding of two proteins as a rapid process that is in equilibrium because of the relatively short timescale of binding events relative to other reactions in the system like synthesis and decay. Additionally, I sometimes assume that synthesis/decay processes are at steady-state. The purpose of this model is to gain a general understanding of which parameters have a significant impact on inverter transfer curves. Therefore, by making these assumptions, I may concentrate on the statics of inverter behavior rather than the dynamics. However, it is not

clear that synthesis and decay processes will be at steady state in the cell, and therefore this assumption may not hold.

3. The device does not place a significant demand upon the cell's resources. My model implicitly assumes that the critical machinery, materials, and energy of the cell necessary for device operation is not a limiting factor in device behavior. Therefore, the effect of the available resource pool is captured in the parameter values themselves. Again, this assumption is relatively common in these types of models, but its validity has not been conclusively established.

4. I do not explicitly model the effects of cell growth and DNA replication upon device behavior. As the cell grows and divides, mRNA and protein undergoes dilution. Typically, biological models describe this process as a constant growth rate which simply increases the degradation rate of molecular species. I omit this from my model and assume that my degradation rate constants encompass these effects as well. Additionally, the amount of DNA in the cell is not fixed. In fact, just before division the DNA content of the cell should be twice its value immediately after division. I neglect these effects and assume a constant DNA copy number.

5. I ignore many of the details of the biological processes that I model. For example, I describe transcription and translation each with a single rate constant corresponding to the number of transcripts per gene copy per second and the number of proteins per transcript per second respectively. These two lumped rate constants omit the fact that both transcription and translation are really a series of binding and reaction steps that include both initiation and elongation. This assumption is again relatively common in these types of models but may be inappropriate in this context.

6. The device input protein binds as a dimer to the device output regulating the output signal. Transcription from the naked device output occurs at some maximum rate. The bound device output is in a completely repressed state

such that the output signal is zero. The fraction of device outputs bound to a single device input protein is negligible and considered zero.

3.3.2 Specification of biochemical reactions

The first step in modeling any biological system is to define the system: enumerate the relevant chemical reactions that will be included in the model. In this step, I utilize several of the assumptions previously discussed. Since my models describe transcription-based logic devices, I define device behavior as output transcription rate as a function of input transcription rate. In the case of the transcription-based inverter, the input signal leads to mRNA synthesis from DNA encoding a repressor (the device input). Translation of the resulting mRNA transcripts leads to repressor synthesis. The repressor binds to the operator region (device output) regulating the output transcription rate and correspondingly the mRNA synthesis rate. The model must capture these biological reactions.

Defining a device in this way allows me to draw a distinction between the device input (the repressor) and the device output (the promoter). The device input and output need not exist on the same piece of DNA and thus may in fact have different copy number. The complete list of variables, species and parameters used are listed (Tables 3.1 and 3.2). In the species names, DNA, mRNA and protein species are denoted with d, m and p respectively. In the parameter names, k denotes a kinetic rate and K represents a equilibrium dissociation constant. For the subscripts, i denotes input, o denotes output, s denotes synthesis, and d denotes decay.

I list the biological reactions of the model below. Since I prefer to begin with the simplest possible model that captures the relevant level at which I can design devices, the only reactions that I include are transcription, translation, mRNA and protein degradation, repressor dimerization and repressor-operator binding.

Variable	*Definition*	*Units*
d_i	device input	*copies/cell*
d_o	device output	*copies/cell*
m_i	device input mRNA	*transcripts/cell*
m_o	device output mRNA	*transcripts/cell*
p_i	device input protein	*proteins/cell*
$p_i \cdot p_i$	dimerized device input protein	*complexes/cell*
$p_i \cdot p_i \cdot d_o$	dimerized device input protein bound to device output	*complexes/cell*
d_n	nonspecific DNA	*copies/cell*
$p_i \cdot p_i \cdot d_n$	dimerized device input protein bound to nonspecific DNA	*complexes/cell*
d_a	device alternate operators	*copies/cell*
$p_i \cdot p_i \cdot d_a$	dimerized device input protein bound to device alternate operators	*complexes/cell*

Table 3.1: Model molecular species

$$
\begin{aligned}
d_i &\xrightarrow{k_{sm_i}} d_i + m_i && \text{(mRNA synthesis)} \\
m_i &\xrightarrow{k_{dm_i}} \phi && \text{(mRNA degradation)} \\
m_i &\xrightarrow{k_{sp_i}} m_i + p_i && \text{(protein synthesis)} \\
p_i &\xrightarrow{k_{dp_i}} \phi && \text{(protein degradation)} \\
p_i \cdot p_i &\xrightarrow{k_{dp_i}} \phi && \text{(protein degradation)} \\
p_i \cdot p_i \cdot d_o &\xrightarrow{k_{dp_i}} d_o + \phi && \text{(protein degradation)} \\
p_i + p_i &\xleftrightarrow{K_1} p_i \cdot p_i && \text{(protein dimerization)} \\
p_i \cdot p_i + d_o &\xleftrightarrow{K_2} p_i \cdot p_i \cdot d_o && \text{(protein-operator binding)} \\
d_o &\xrightarrow{k_{sm_o}} d_o + m_o && \text{(mRNA synthesis)}
\end{aligned}
$$

Note that k_{sm_i} and k_{sm_o} are not rate *constants* but rather rate *variables* because both will vary according to the input transcription signal to the device. This aspect of the model is discussed further in section 3.4.1. The rates governing mRNA degradation, protein synthesis and protein degradation are constants in this model.

Parameter	*Definition*	*Typical value*	*Units*	*Reference*
$\max(k_{sm_i})$	maximum input mRNA synthesis rate	1	$\frac{transcripts/genecopy}{second}$	Drew Endy
$\max(k_{sm_o})$	maximum output mRNA synthesis rate	1	$\frac{transcripts/genecopy}{second}$	Drew Endy
k_{sp_i}	input protein synthesis rate	0.17	$\frac{proteins/transcript}{second}$	[87]
$t_{1/2m}$	mRNA halflife	2	$minutes$	[87]
$t_{1/2p}$	protein halflife	40	$minutes$	[88]
$[d_i]^T$	device input sites	1	$copies/cell$	none
$[d_o]^T$	device output sites	1	$copies/cell$	none
$[d_n]^T$	nonspecific binding sites	$4.2 * 10^6$	$copies/cell$	[89]
$[d_a]^T$	device alternate sites		$copies/cell$	
K_1	device input protein dimerization dissociation constant	8.4	$proteins/cell$	[89, 90]
K_2	device input protein dimer-device input dissociation constant	1.3	$molecules/cell$	[89, 90]
K_5	input protein-nonspecific DNA dissociation constant	$4.2 * 10^5$	$molecules/cell$	[89]
K_6	input protein-singly bound nonfunctional site dissociation constant	$K_4/10$	$molecules/cell$	none

Table 3.2: Model parameters

3.4 Model derivation

I begin with a physicochemical model of my sample inverter system. Such a model is simply a translation of the biological reactions into kinetic equations.

$$\begin{aligned}
\frac{d[m_i]}{dt} &= k_{sm_i}[d_i] - k_{dm_i}[m_i] \\
\frac{d[p_i]^T}{dt} &= k_{sp_i}[m_i] - k_{dp_i}[p_i]^T \\
\frac{d[m_o]}{dt} &= k_{sm_o}[d_o]
\end{aligned} \tag{3.2}$$

I derive the degradation rate constants k_{dm_i}, k_{dp_i}, k_{dm_o} and k_{dp_o} by assuming that molecules decay at a rate proportional to their concentration. Therefore, the concentration of a molecule X as a function of time X(t) will adhere to the following equation.

$$X(t) = X(0)e^{-k_{dx}t} \tag{3.3}$$

X(0) denotes the initial concentration of X, k_{dx} is the degradation rate constant for X and t is time. The halflife of X is the time it takes for the concentration of X to decrease to half of its initial value X(0). If I know the the halflife of X then I can calculate the degradation rate constant for X using equation (3.4). Estimates for the halflives of molecular species in the model are listed (Table 3.2).

$$k_{dx} = \frac{ln(2)}{t_{1/2x}} \tag{3.4}$$

It is easy to overlook some of the subtleties of the model in the mapping of the biological reactions into kinetic equations. In this section, I discuss the nuances of the model, since the decisions I make in developing the model critically impact the device behavior that the model predicts.

3.4.1 Obtaining the device transfer curve from the model

A transfer curve specifies the device output as a function of its input. For transcription-based devices, transcription rates measured in units of PoPs are the physical implementation of logical signals. Ideally, I should be able to obtain from the model a device transfer curve of the following form.

$$\Pi_o = f\left(\Pi_i\right) \tag{3.5}$$

On the basis of equation (3.5), it is clearly important that I be able to easily specify the strength of the input to the device. In the model's current form in equation set (3.2), there is no way to specify input transcription rate to the device except indirectly by changing the value of the mRNA synthesis rate k_{sm_i}. Instead, I introduce what is called a drive term in electrical engineering. Denoted S_i, the drive term can vary between zero and one and represents the fraction of maximal strength at which the input is being driven. Both the input transcription rate Π_i and the mRNA synthesis rate k_{sm_i} should be proportional to the input signal strength S_i. Since S_i is dimensionless, the following equation holds.

$$S_i = \frac{\Pi_i}{\max(\Pi_i)} = \frac{k_{sm_i}}{\max(k_{sm_i})} \tag{3.6}$$

Thus, although the value of $\max(\Pi_i)$ and the exact relationship between k_{sm_i} and Π_i may be unknown, I can still explicitly specify S_i in the model. Effectively, S_i is a non-dimensional form of the device input signal.

In the model, I assume that all device outputs exist in either the unbound (d_o) or doubly bound ($p_i \cdot p_i \cdot d_o$) state. All unbound device outputs lead to the same mRNA synthesis rate $\max(k_{sm_o})$ (in units of transcripts per DNA copy per second) of the downstream coding sequence. All doubly bound device outputs are not transcribed at all (or $\min(k_{sm_o}) = 0$). The maximum output transcription rate occurs when there is no device input protein present so all device outputs are in the unbound state d_o. Therefore, the output transcription rate is proportional to the fraction of device

outputs in the unbound (d_o) state. I can therefore define a variable S_o as the fraction of maximum output transcription rate and the following equations should hold.

$$\begin{aligned} S_o &= \frac{\Pi_o}{\max(\Pi_o)} = \frac{k_{sm_o}}{\max(k_{sm_o})} \\ &= \frac{[d_o]}{[d_o] + [p_i \cdot p_i \cdot d_o]} \end{aligned} \tag{3.7}$$

Rewriting of the model to include S_i and S_o leads to the following set of equations. Note that the differential equation for $[m_o]$ is no longer necessary since the output variable of interest is S_o, and it is specified by a separate equation. Again, S_o is effectively a non-dimensional form of the device output signal.

$$\begin{aligned} \frac{d[m_i]}{dt} &= S_i \max(k_{sm_i})[d_i] - k_{dm_i}[m_i] \\ \frac{d[p_i]^T}{dt} &= k_{sp_i}[m_i] - k_{dp_i}[p_i]^T \\ S_o &= \frac{[d_o]}{[d_o] + [p_i \cdot p_i \cdot d_o]} \end{aligned} \tag{3.8}$$

Although the model does not explicitly include the absolute input and output signals Π_i and Π_o, it does include the fractional signal strengths S_i and S_o. Thus, this model should yield the transfer curve as S_o as a function of S_i. I make use of the additional assumptions enumerated in section 3.3.1 to simplify the model further and ultimately use the model to generate the device transfer curve.

3.4.2 Use of the Boltzmann distribution to compute the fraction of molecules in each binding state

I have some control over the copy number of the inverter. I can even place the device input and device output on different plasmids such that they have different copy number (denoted by $[d_i]^T$ and $[d_o]^T$, respectively). The model should make use of the constant number of DNA molecules in the cell. Since I assume that all binding reactions are at equilibrium, the DNA molecules are partitioned between the two binding states ($[d_o]$ and $[p_i \cdot p_i \cdot d_o]$) according to a Boltzmann distribution with the

total number of DNA molecules held fixed [91, 92, 93].

$$[d_i]^T = [d_i] \tag{3.9}$$

$$[d_o]^T = [d_o] + [p_i \cdot p_i \cdot d_o] \tag{3.10}$$

$$[d_o] = [d_o]^T \frac{[d_o]}{[d_o] + K_1^{-1} K_2^{-1} [d_o][p_i]^2} \tag{3.11}$$

$$[p_i \cdot p_i \cdot d_o] = [d_o]^T \frac{K_1^{-1} K_2^{-1} [d_o][p_i]^2}{[d_o] + K_1^{-1} K_2^{-1} [d_o][p_i]^2} \tag{3.12}$$

Combining the above equations permits specification of S_o in terms of the dissociation constants K_1 and K_2 and the free input protein concentration $[p_i]$.

$$\begin{aligned} S_o &= \frac{[d_o]}{[d_o]^T} \\ &= \frac{[d_o]}{[d_o] + K_1^{-1} K_2^{-1} [d_o][p_i]^2} \\ &= \frac{K_1 K_2}{K_1 K_2 + [p_i]^2} \end{aligned} \tag{3.13}$$

Since I treat the DNA molecules as distributed between different thermodynamic states and since all binding reactions are at equilibrium relative to synthesis and decay reactions, it follows that the regulatory protein p_i should also be distributed among binding states according to the Boltzmann distribution. In the expression for S_o above, $[p_i]$ refers to the unbound input protein concentration in the cell. It is important to distinguish between the total concentration of input protein ($[p_i] + 2 * [p_i \cdot p_i] + 2 * [p_i \cdot p_i \cdot d_o]$) in the cell and the amount of unbound protein in the cell ($[p_i]$). The total amount of input protein, which I denote $[p_i]^T$, depends on the synthesis and decay rates. The amount of free input protein in the cell, $[p_i]$, depends on the dissociation constants and the number of binding sites within the cell. I can use mass conservation to obtain an implicit expression for $[p_i]$ (equation (3.14)) and solve this expression numerically for $[p_i]$.

$$[p_i]^T = [p_i] + 2K_1^{-1}[p_i]^2 + \frac{2[d_o]^T [p_i]^2}{K_1 K_2 + [p_i]^2} \tag{3.14}$$

I now rewrite the entire model once more, being careful to distinguish between $[p_i]$ and $[p_i]^T$.

$$\begin{aligned}
\frac{d[m_i]}{dt} &= S_i \max(k_{sm_i})[d_i]^T - k_{dm_i}[m_i] \\
\frac{d[p_i]^T}{dt} &= k_{sp_i}[m_i] - k_{dp_i}[p_i]^T \\
[p_i]^T &= [p_i] + 2K_1^{-1}[p_i]^2 + \frac{2[d_o]^T[p_i]^2}{K_1K_2 + [p_i]^2} \\
S_o &= \frac{K_1K_2}{K_1K_2 + [p_i]^2}
\end{aligned} \tag{3.15}$$

3.4.3 mRNA levels are at steady-state relative to protein levels

Generally, mRNA halflives are much shorter than protein halflives. Estimates of mRNA lifetimes are approximately two minutes [87]. Others have extended mRNA halflives to eight minutes by adding secondary structural elements to the transcript [94, 95, 96]. Protein halflives, however, are much longer with some of the shortest halflives being between twenty and thirty minutes [97, 98, 88]. Therefore, on average proteins will outlast mRNA transcripts a minimum of two- to three-fold and usually at least ten-fold if not more. So I can make an additional simplifying assumption that the mRNA species is at its steady-state levels relative to protein concentration.

$$[m_i]_{ss} = S_i \left(\frac{\max(k_{sm_i})}{k_{dm_i}} [d_i]^T \right) \tag{3.16}$$

The mRNA steady-state assumption leads to the following simplified model.

$$\begin{aligned}
\frac{d[p_i]^T}{dt} &= k_{sp_i} S_i \left(\frac{\max(k_{sm_i})}{k_{dm_i}} [d_i]^T \right) - k_{dp_i}[p_i]^T \\
[p_i]^T &= [p_i] + 2K_1^{-1}[p_i]^2 + \frac{2[d_o]^T[p_i]^2}{K_1K_2 + [p_i]^2} \\
S_o &= \frac{K_1K_2}{K_1K_2 + [p_i]^2}
\end{aligned} \tag{3.17}$$

3.4.4 Use of the steady-state assumption to generate the device transfer curve

The model in equation set 3.17 describes the input and output protein concentration as a function of time for a given input signal (provided as a signal strength between zero and one). To obtain the transfer curve for the device, I compute the output signal strength S_o as a function of the input signal strength S_i at steady-state. I obtain the transfer curve from the following equations.

$$\begin{aligned}
[p_i]_{ss}^T &= S_i \left(\frac{k_{sp_i} \max(k_{sm_i})}{k_{dp_i} k_{dm_i}} [d_i]^T \right) \\
[p_i]_{ss}^T &= [p_i] + 2K_1^{-1}[p_i]^2 + \frac{2[d_o]^T [p_i]^2}{K_1 K_2 + [p_i]^2} \\
S_o &= \frac{K_1 K_2}{K_1 K_2 + [p_i]^2}
\end{aligned} \tag{3.18}$$

The key parameters in the model are the dissociation constants, the output part copy number and the lumped parameter defined in equation (3.19) which I term α_i. The lumped parameter is the maximum attainable steady-state concentrations of the input protein: when the input signal is at a maximum ($S_i = 1$), $[p_i]_{ss}^T = \alpha_i$.

$$\alpha_i \equiv \frac{k_{sp_i} \max(k_{sm_i})}{k_{dp_i} k_{dm_i}} [d_i]^T \tag{3.19}$$

Equating the first two expressions in the model (equation set (3.18)) and substituting α_i leads to the following simplified model describing the transfer curve of the inverter.

$$\begin{aligned}
0 &= [p_i] + 2K_1^{-1}[p_i]^2 + \frac{2[d_o]^T [p_i]^2}{K_1 K_2 + [p_i]^2} - S_i \alpha_i \\
S_o &= \frac{K_1 K_2}{K_1 K_2 + [p_i]^2}
\end{aligned} \tag{3.20}$$

Although I do not have an explicit equation for S_o as a function of S_i, I do have an equation specifying S_o as a function of the free monomeric input protein concentration $[p_i]$ as well as an equation relating $[p_i]$ to S_i. Therefore, the model will still generate a device transfer curve for S_o versus S_i for the transcription-based inverter.

3.5 Model analysis

The device behavior that the model predicts depends most critically on the assumptions made during construction of the model but also on the parameter values chosen. The parameters which determine device behavior are α_i, $[d_o]^T$, K_1 and K_2. Each of these parameters will have a different impact on device behavior, and experimentally, I have differing degrees of control over them. I discuss each model parameter in turn.

3.5.1 α_i determines the input protein swing and is responsive to tuning

The lumped parameter α_i represents the maximum steady-state device input protein concentration and thus determines the input protein swing of the device. Although the device signal is encoded as a transcription rate, the protein concentration range over which the device operates is also important. The value of α_i should be sufficiently high such that stochastic noise in protein number is unlikely to switch signal state. Yet if α_i is too high, then just a handful of devices will place an unacceptably high operational demand upon the cell which may cause altered cell behavior. Ideally, the target value for α_i should be as low as possible without rendering the device vulnerable to stochastic fluctuation, in order to minimize device demand on the cell.

As the device designer, I have some measure of control over α_i by tuning the various parameters that constitute the lumped parameter. I can alter the value of k_{sp_i} by using different ribosome binding sites. However, translation initiation appears to be a somewhat complex variable to tune, since changes in the Shine-Dalgarno sequence may also affect transcription rates [99]. Researchers have shown 200-fold range in gene expression by changing the translational initiation signal, but most of this change is attributable to altered mRNA levels possibly due to protection of the mRNA by bound ribosomes [99]. Yet if the desired goal is to simply change α_i then the fact that $\max(k_{sm_i})$ and k_{sp_i} are coupled may not be problematic.

In addition to controlling synthesis rates of transcripts and proteins, I can also adjust degradation rates. I have very coarse control of the protein degradation rate

k_{dp_i}, since I can place different degradation tags on the protein to speed up protein turnover. For example, green fluorescent protein (GFP) had an estimated halflife of 40, 60, or 110 minutes, depending on the sequence of a carboxy-terminal degradation tag [88]. In the absence of a tag, GFP is indefinitely stable over the course of most experiments. Enhancing protein stability to decrease degradation rate is significantly more difficult and therefore not a promising means of control. In the case of the mRNA degradation rate k_{dm_i}, *in vivo* mRNA functional halflives range between 40 seconds and 20 minutes which again provides an approximate upper bound on my range of control [100]. In fact, some engineering work has been done on stabilizing mRNA transcripts. Transcript halflives between two and eight minutes have been observed depending on the presence of stem loops at the 5' and 3' end of the transcripts [94, 95, 96]. In summary, I currently have relatively coarse and weak control over protein and mRNA degradation rates.

Finally, I can tune copy number in two ways to influence α_i. I may place my device on plasmids with different replication origins or include tandem duplicate copies of my device. Thus, I should be able to set $[d]^T$ to either single copy, medium copy (15-20) or high copy (100-200) or integer multiples thereof. Of course, in reality, the copy number of my device will vary two-fold *in vivo* over the cell cycle.

Based on typical parameter values (Table 3.2), $\alpha_i = 50,000$ proteins/cell. Since typical parameter values are usually measured or estimated from genetic parts designed to optimize protein expression, they tend to yield very high protein concentrations. The value of α_i is clearly high enough such that random fluctuations in protein concentration should not be a factor. In fact, the expected value of α_i may be too high since a typical *E. coli* cell only has 2.6 million proteins in the cell which means that a single inverter device consumes nearly 2% of the cell's protein content [101]. Regardless, it is clear that α_i primarily determines the input protein swing and is likely regulatable over a few orders of magnitude using the techniques described above.

Transfer curves for different values of α_i with all other parameter values set to their typical values are shown (Figure 3-5, Table 3.2). A couple of issues in the transfer

curve are immediately apparent. First, the trip point is near the half-maximal input only when $\alpha_i = 50$. For most of the plots, the trip point is lower than the ideal value. Thus, either α_i should have a value lower than typical parameter estimates, or other parameter values need to change to compensate for high α_i. A second issue is the small noise margin of the transfer curves. The ideal device transfer curve should have a low gain region for low as well as high inputs to ensure adequate noise margins. The low gain regions are necessary to dampen noise effects reducing the likelihood of erroneous signal propagation, especially when devices are in series. Devices that have low gain regions for both low and high inputs are desirable, because they restore the signal at every stage in a series of devices. In the plotted transfer curves, a low gain region only exists for high inputs. Devices with this kind of skewed transfer curve are functional, since signal restoration still occurs at every other stage in a device cascade. Nevertheless, a device with a larger low gain region for low inputs is preferable.

3.5.2 Transcription-based devices have high fan out

The parameter $[d_o]^T$ represents the copy number of the device output. Every copy of the device output is capable of driving the transcription of a different output protein. Therefore, $[d_o]^T$ specifies the number of outputs the device controls. The maximum number of outputs that a device can reliably drive is known as the "fan out" in electrical engineering. If $[d_o]^T$ exceeds the device fan out, then the device will be unable to completely repress transcription from each of its outputs leading to a poor device transfer curve. The number of outputs a device can successfully drive will depend on the maximum level of input protein and the repressor-operator binding affinity. Thus, α_i is an important factor in the fan out. Based on the estimated value of α_i (Table 3.2), the device fan out is very high, because α_i will likely exceed $[d_o]^T$ by multiple orders of magnitude, and repressor-operator binding affinity tends to be quite high.

In considering device behavior for different values of $[d_o]^T$, it is important to distinguish between the scenario in which each copy of the device output drives transcription of a different protein and the scenario in which each copy of the device

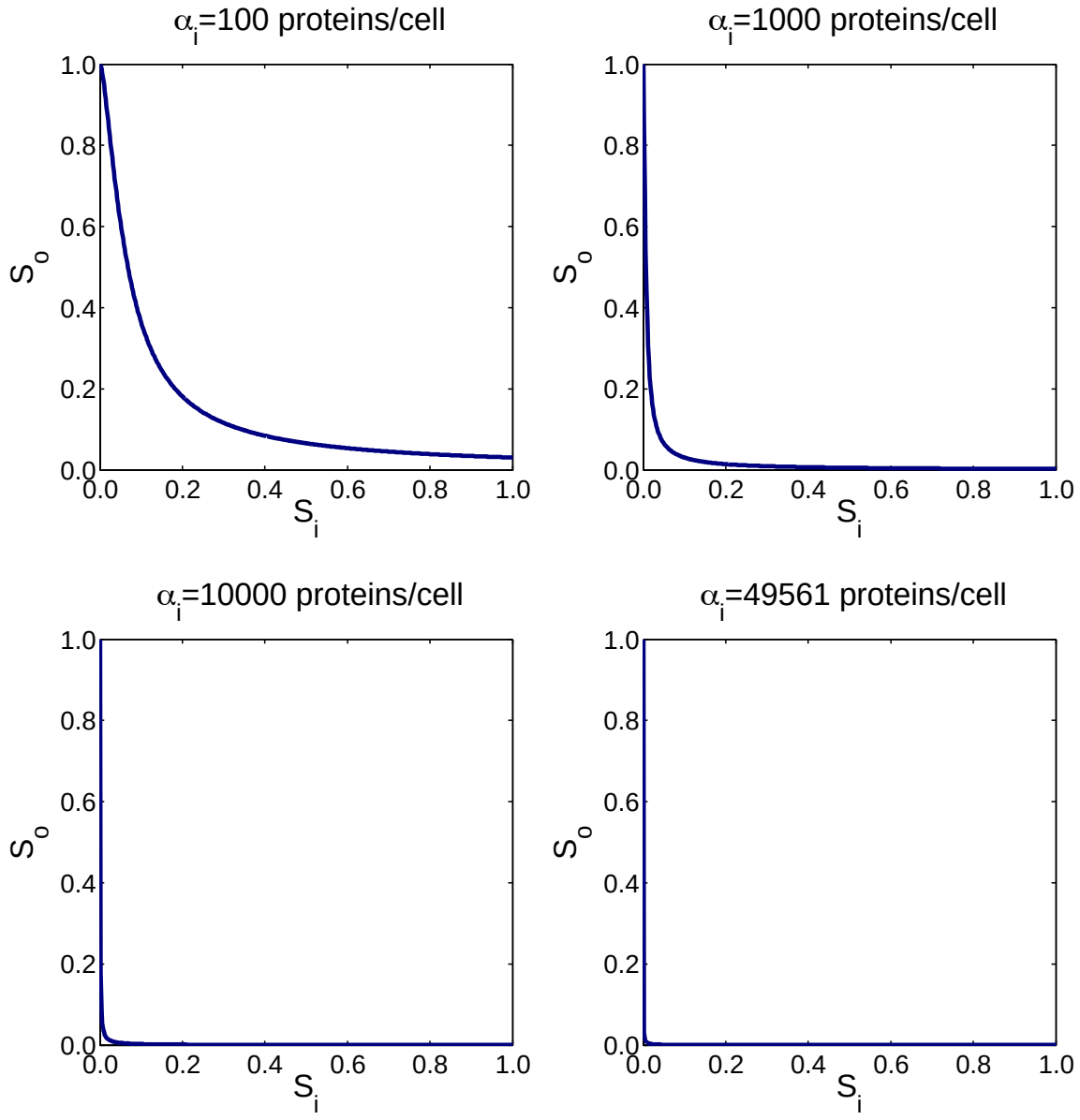


Figure 3-5: Transfer curve of transcription-based inverter using equation set (3.20) for different values of the lumped parameter α_i. The last plot shows the transfer curve for the typical value of α_i (Table 3.2). α_i determines the input protein swing of the device: the range of input protein concentrations over which the device operates.

output drives transcription of the same protein. Transfer curves for different values of $[d_o]^T$ assuming that every copy of the device output regulates transcription of a different device output protein are shown (Figure 3-6). As $[d_o]^T$ increases, the output signal for high input remains very low suggesting that transcription-based devices are likely capable of driving a large number of different outputs. In fact, the transfer curve actually looks better as $[d_o]^T$ increases, since the trip point moves closer to the ideal value of half the swing.

In the alternate scenario in which every copy of the device output drives transcription of the same protein, the shape of the transfer curve will depend on the copy number of the device output relative to the device input. If the device output copy number exceeds the device input copy number by a sufficient margin, even in the repressed state the output protein may experience a relatively high rate of transcription. If the opposite situation occurs, then only a small input signal will be sufficient to repress synthesis of the output protein.

3.5.3 Dissociation constants determine transfer curve shape

The dissociation constants K_1 and K_2 determine the concentration of protein at which half-maximal binding occurs. K_1 characterizes the monomer-dimer equilibrium, while K_2 determines the binding equilibrium between the device input protein dimer and the device output. The transfer curve should have a high gain region for those input values over which the device output signal transitions from high to low. In biological terms, such behavior is generally characterized as "switch-like" behavior [102]. One of the most common origins for switch-like behavior is cooperative interactions in which the binding of one protein to the operator greatly enhances the affinity of the operator for a second protein. In the transcription-based device I describe here, the input proteins indeed form dimers with the operator. Since I assume that the dimer binds to the operator and the monomer does not, I essentially assume cooperative binding of the device input protein to the DNA.

Deconvolving the influence of K_1 and K_2 on the device transfer curve is difficult since the two dissociation constants appear together in the model, yet K_1 also appears

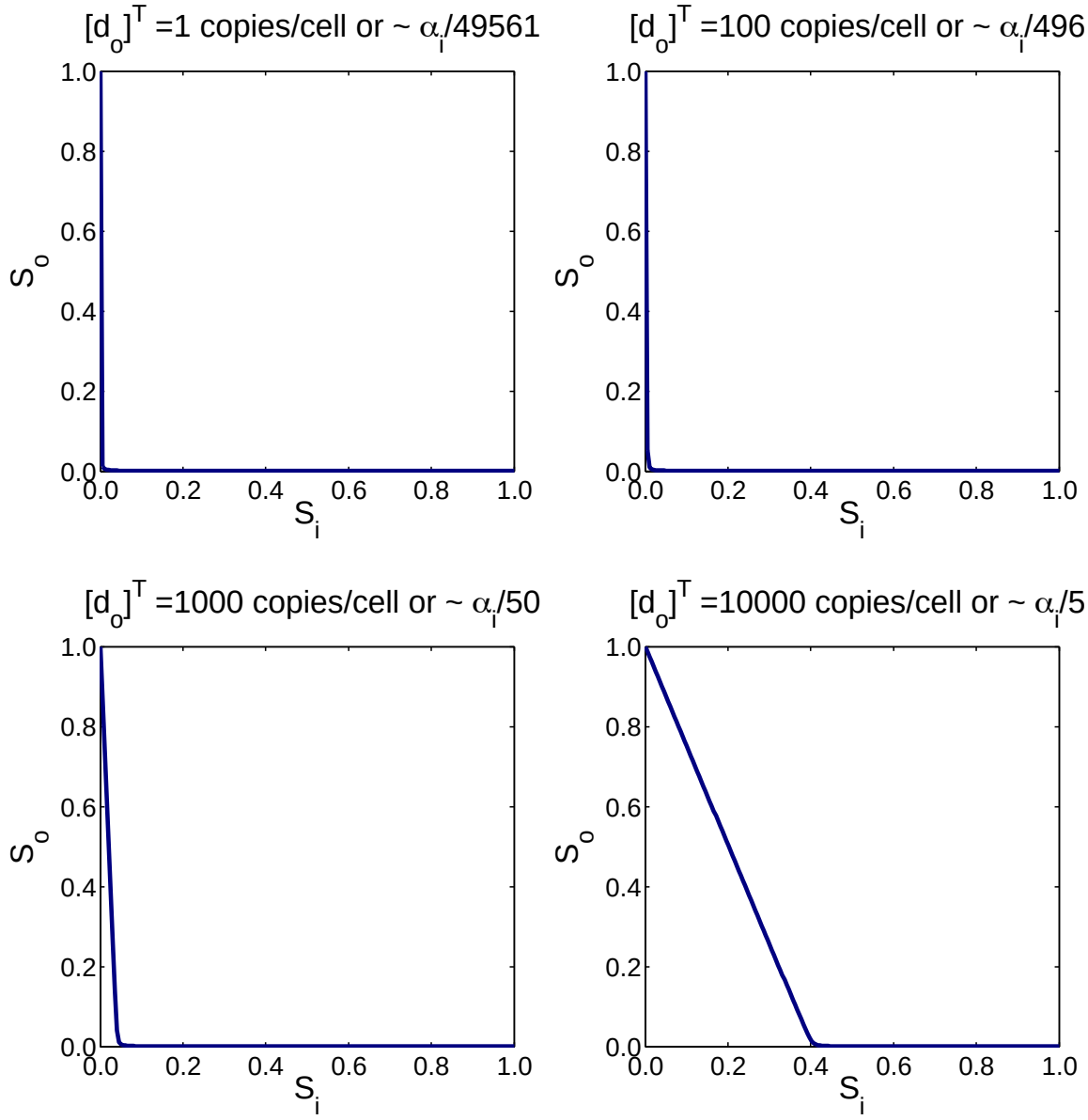


Figure 3-6: Transfer curve of transcription-based inverter using equation set (3.20) for different values of the parameter $[d_o]^T$, the device output copy number, assuming that each output drives transcription of a different protein. The first plot shows the transfer curve for the typical value of $[d_o]^T$ (Table 3.2). Although the values shown for $[d_o]^T$ are not all realistic, they serve to illustrate the effect of $[d_o]^T$ on the trip point of the transfer curve.

alone in the equation governing the free monomeric protein concentration $[p_i]$ (see equation set (3.20)). Experimentation with different values for K_1 and K_2 shows that a reasonable transfer curve is obtained using the following approximations.

$$K_1 \approx \alpha_i * 10 \tag{3.21}$$

$$K_2 \approx \alpha_i/500 \tag{3.22}$$

Transfer curves for three sets of parameter values are shown (Figure 3-7): the typical dissociation constant values (Table 3.2), dissociation constant values that adhere to expressions (3.21) and (3.22), and an intermediate set of values. The curve adhering to expressions (3.21) and (3.22) more closely resembles the ideal transfer curve since its trip point is closer to half the swing, and it has a small but non-zero noise margin. The model produces transfer curves of similar shape irrespective of the value of α_i (not shown) indicating that it is not the absolute value of the dissociation constants but rather the value of the dissociation constants relative to α_i that is important.

The observed approximations for K_1 and K_2 can be rationalized intuitively. A good device transfer curve must have two low gain regions separated by a high gain region. K_2, which defines the binding affinity between the repressor dimer and the operator, primarily influences the slope in the high gain region, since it is repressor dimer binding that is responsible for the transition from high output to low output with increasing input. K_2 must be low relative to α_i, so that the transition region of the transfer curve is high gain. As one might expect, if K_2 is too low then the device trip point is very low, because small quantities of protein are sufficient for repression. K_1, on the other hand, defines the binding affinity between repressor monomers. A relatively high value for K_1 (on the order of α_i or slightly higher) ensures a low gain region for low input values. For low inputs, there are low numbers of proteins around and because the dimerization affinity is low, very few repressors are in the dimeric state and thus capable of repression. Hence, for low input values, the input remains

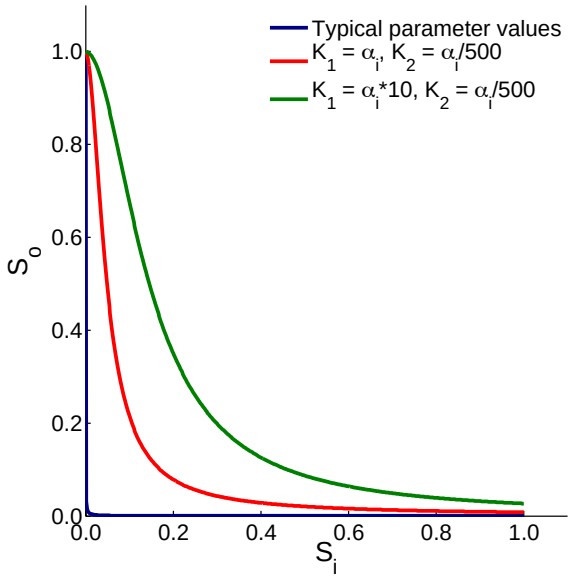


Figure 3-7: Transfer curve of transcription-based inverter using equation set (3.20) for different sets of dissociation constant values. Typical values obtained from relevant literature for all model parameters are listed (Table 3.2). In the blue transfer curve, the following parameter values change: $K_1 = \alpha_i$ and $K_2 = \alpha_i/500$. In the green transfer curve, the following parameter values change: $K_1 = \alpha_i * 10$ and $K_2 = \alpha_i/500$.

high. If K_1 is too high, then dimers do not form and device output is high irrespective of input.

Since the dissociation constants have such a crucial impact upon the shape and quality of the device transfer curve, a natural question is how much control does the device engineer have over these parameter values. In general, rational design of protein-protein interactions and protein-DNA interactions is tractable but not routine. To facilitate control over the dissociation constants, I use synthetic transcription factors to implement a transcriptional inverter (Chapter 5). Leucine zippers constitute the dimerization domains, and zinc fingers make up the DNA binding domains. Much work has been done on both of these protein domain families, and there are many possible candidate domains from which to choose. Leucine zipper pairs dimerize with dissociation constants as low as 50 nM (∼20 proteins/cell) and as high as 3 uM (∼1264 proteins/cell) [103]. Zinc finger proteins have been engineered to bind DNA sequences with very high affinity (pM range) [104]. Based on the model, such

binding affinities are more than sufficient for transcription-based logic.

3.5.4 Enhanced cooperativity compensates for explicit consideration of nonspecific DNA

The model as presented in equation set (3.20) neglects nonspecific binding: the binding of transcription factors to noncognate DNA sequences. In the case of bacterial repressors, nonspecific DNA binding can involve a significant fraction of the total protein concentration [105, 106, 107]. The amount of noncognate DNA exceeds the amount of operator DNA in the cell by several orders of magnitude since every base in the *E. coli* genome represents a potential nonspecific binding site for the repressor. Therefore, nonspecific DNA binding competes with cognate DNA binding for the free protein concentration in the cell, thereby affecting the device transfer curve. Inclusion of nonspecific DNA binding requires the addition of a single reaction to the list of biochemical reactions (see Table 3.1 and 3.2 for species and parameter definitions).

$$p_i \cdot p_i + d_n \overset{K_5}{\longleftrightarrow} p_i \cdot p_i \cdot d_n$$

Note that I assume that only repressor dimers are capable of binding to DNA. Such an assumption is essentially equivalent to assuming that repressor monomers make insufficient contacts with the DNA to constitute a significant fraction of the protein in the cell. The addition of nonspecific binding only results in a minor modification to the model.

$$\begin{aligned} 0 &= [p_i] + 2K_1^{-1}[p_i]^2 + \frac{2[d_o]^T[p_i]^2}{K_1K_2 + [p_i]^2} + \frac{2[d_n]^T[p_i]^2}{K_1K_5 + [p_i]^2} - S_i\alpha_i \\ S_o &= \frac{K_1K_2}{K_1K_2 + [p_i]^2} \end{aligned} \tag{3.23}$$

The presence of a large number of nonspecific DNA binding sites creates competition for the pool of free device input protein. Thus, for typical parameter values, a higher input signal is necessary to achieve the same output signal, when nonspecific DNA binding is included in the model (Figure 3-8). Thus, the trip point is shifted

to slightly higher signal values, and the transfer curve is slightly improved. Parameter values that yield reasonable transfer curves when nonspecific DNA binding is not included in the model do not yield reasonable transfer curves when nonspecific DNA binding is explicitly included in the model: in order for the device to achieve significant repression of the output for logic one input, the affinity of the operator for the protein dimer must be correspondingly higher when nonspecific DNA binding is explicitly included. Otherwise, most of the device input proteins are bound nonspecifically and little repression occurs. However, at very high operator affinities, even small quantities of protein are sufficient for repression, because the specific operator sites are assumed to have higher affinity than the nonspecific sites, resulting in a small low gain region for low input. Thus, when nonspecific DNA binding is included in the model, the device exhibits either a low gain region for low input or complete repression at high input, but not both. Thus, in order to compensate for the increase in operator affinity needed for complete repression, the repressor dimerization affinity must be decreased, so that a low gain region for low input values exists. In short, inclusion of nonspecific DNA binding in the model exacerbates the tension between the dual goals of complete repression at high input values and little repression at low input values. Regardless, nonspecific DNA binding is an important consideration in transcriptional device design that cannot be neglected.

3.5.5 Transcriptional devices perform sufficiently well to implement combinational digital logic

The transfer curve for the transcription-based inverter device derived from equation set (3.23) is shown (Figure 3-9). Typical parameter values were used to generate the curve with the following exceptions: $K_1 = \alpha_i * 100$ and $K_2 = \alpha_i/10,000$ (Table 3.2). A plot of not only the output signal strength as a function of input signal strength but also the input as a function of output is shown. Graphing the superimposed transfer curves permits calculation of the device noise margin. The noise margin is about 0.05. Although the noise margin is not very high, note that it represents the noise margin

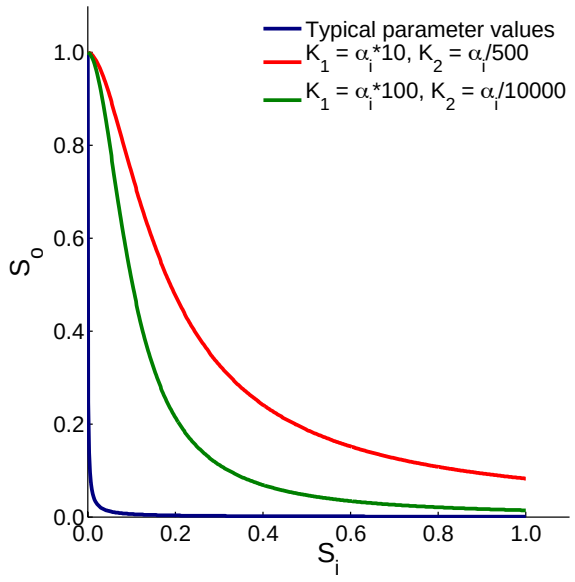


Figure 3-8: Transfer curve of transcription-based inverter using equation set (3.23) for different sets of dissociation constant values. The model used to generate these transfer curves explicitly includes nonspecific binding. Typical values obtained from relevant literature for all model parameters are listed (Table 3.2). To retain the same basic shape of the transfer curve as obtained in the absence of nonspecific DNA binding, K_1 must be increased ten-fold and K_2 decreased ten- to twenty-fold (Figure 3-8). The increased affinity between the input protein dimer and operator compensates for the competition from nonspecific DNA binding to achieve sufficiently low output for logic one input. The decreased dimerization affinity maintains the existence of a low gain region for low input.

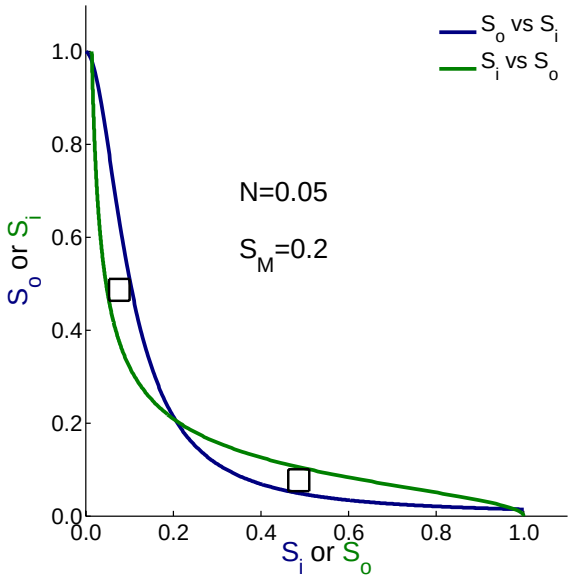


Figure 3-9: Superimposed transfer curves of a transcription-based inverter using equation set (3.23). Typical parameter values were used to generate the curve with the following exceptions:: $K_1 = \alpha_i * 100$ and $K_2 = \alpha_i/10000$ (Table 3.2). The noise margin N and trip point S_M are shown. The transfer curve is of sufficient quality that transcription-based devices can be used in combinational digital logic.

in the worst-case scenario of an infinite number of inverters with noise at every input. The trip point S_M, at 0.2, is relatively close to the ideal value of 0.5. Therefore, using these parameter values, the model yields a transfer curve that suggests that transcription-based devices may behave sufficiently well to implement combinational digital logic. Nevertheless, there remains considerable room for improvement.

3.6 Parameter sensitivity analysis

A prime consideration in the rational design of synthetic transcription factors is how sensitive device behavior is to the biochemical parameters describing the parts. For instance, if the device only performs well if the protein binds to the DNA within a narrow target range of affinities, then it is unlikely that the device will work when built. Sensitivity analysis can give insight into which biochemical parameters are most important to device behavior. Since the noise margin is a single quantitative

measure of device performance (encompassing the quality of the swing, trip point, and gain), I use the noise margin to quantify device behavior in the sensitivity analyses.

There are two possible ways to perform parameter sensitivity analysis. In the first method, all parameters are held constant except one which is varied over an appropriate range of values. This approach reflects the fact that it may not be possible to achieve the target parameter value when designing the device. Therefore, it is useful to know how close to the target value is sufficient for acceptable device performance. In the second method, a particular parameter is varied over a range of parameter values and the other parameters are reoptimized appropriately to give the best possible device behavior.[1] The latter approach reflects the fact that the influence of different parameters on device performance is often co-dependent. Thus, it may be that other parameters can compensate if one parameter is not at its optimal value.

Equation set (3.23) was used in all parameter sensitivity analyses. It is reproduced here for reference. As before, S_i is the independent variable describing device input and S_o is the dependent variable describing device output.

$$0 = [p_i] + 2K_1^{-1}[p_i]^2 + \frac{2[d_o]^T[p_i]^2}{K_1K_2 + [p_i]^2} + \frac{2[d_n]^T[p_i]^2}{K_1K_5 + [p_i]^2} - S_i\alpha_i$$
$$S_o = \frac{K_1K_2}{K_1K_2 + [p_i]^2}$$

3.6.1 Device behavior is insensitive to α_i

If the values of the dissociation constants are held fixed as α_i is varied, device behavior is very dependent on α_i (Figure 3-5). If instead, the dissociation constants are defined in terms of α_i, and α_i is varied over an appropriate range of values, device performance is insensitive to α_i (Figure 3-10). The sensitivity analysis confirms that it is the values of the dissociation constants relative to α_i, not the absolute parameter values, that affect device performance. Thus, one method of tuning device behavior is simply

[1]The optimization function used in this sensitivity analysis minimized the difference between the transfer curve and a square transfer function. Computational difficulties prevented use of noise margin maximization as the optimization function.

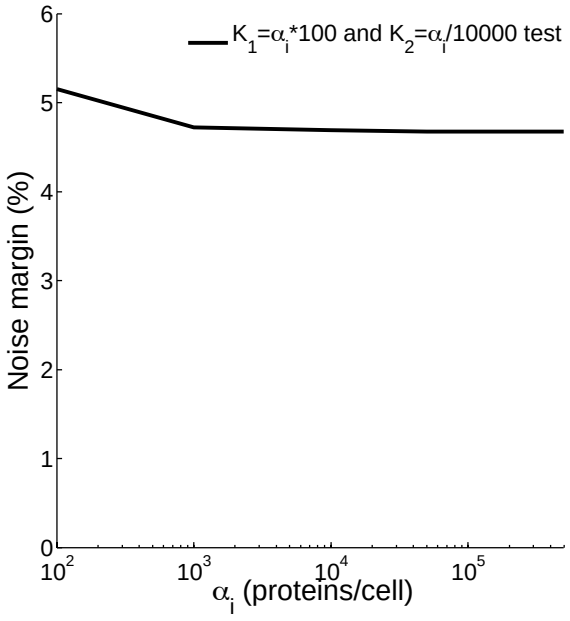


Figure 3-10: Parameter sensitivity analysis of α_i: the maximum steady-state protein concentration in the cell (or the input protein swing). Device performance is insensitive to α_i as long as the dissociation constants are kept in constant proportion to α_i. Thus, it is the value of the dissociation constants relative to α_i not the absolute values of the parameters that affects device performance.

to change the protein concentration range. Clearly, however, the device designer must use some discretion in tuning the device swing, since stochastic fluctuations can impact performance for low α_i, and excessive device demand may be a factor for high α_i.

3.6.2 Device noise margin is sensitive to K_1 and K_2

K_1 and K_2 are the binding affinities for input protein dimerization and input protein dimer-DNA binding, respectively. Device performance is critically dependent on the value of K_1 and K_2 relative to the input protein swing α_i. The device noise margin (in percent of signal strength) versus K_1 and K_2 using equation set (3.23) is shown (Figure 3-11). The noise margin is nonzero for high values of K_1 (greater than α_i) and very low values of K_2 (Figure 3-11). Again, the sensitivity analysis confirms that low affinity for protein dimerization ensures that there is a low gain region for low

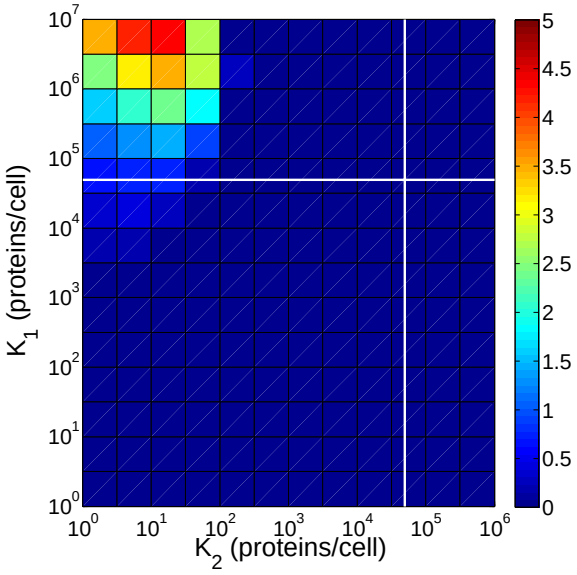


Figure 3-11: Parameter sensitivity analysis of K_1 (the dissociation constant for input protein dimerization) and K_2 (the dissociation constant for input protein dimer-DNA binding). Device performance is very sensitive to the value of K_1 and K_2 relative to α_i. In the plot, the noise margin (in percent of signal strength) is indicated via color as a function of K_1 (y axis) and K_2 (x axis). See the colorbar to the right of the plot for the correspondence between color and noise margin. The white lines on the plot indicate the value of α_i used to construct the plot and serve as a reference for the value of the dissociation constants relative to α_i. Thus, K_1 should be higher than α_i and K_2 should be very low in order to achieve nonzero noise margins.

inputs and high affinity for DNA binding by the dimer ensures that there is a steep transition region from high to low output.

3.7 Alternate device designs

The model presented in section 3.4 focused on the simplest possible design of a transcription-based inverter: an input signal measured in PoPS which drives the transcription of a homodimeric device input protein which regulates the signal from the device output. By tuning the various biochemical parameters appropriately, the model yields reasonable device transfer curves. Yet in naturally occurring systems, biology generally relies on sophisticated layers of regulation to precisely regulate system behavior. Here, I consider alternate device designs that lead to improved static

device performance.

3.7.1 Including nonfunctional high affinity protein binding sites improves the noise margins of the transfer curve

The primary problem in device behavior is that for low inputs, the size of the low gain region is very small (Figure 3-9). The result being that the device has rather small noise margins. If the size of this low gain region could be increased, the device noise margin would similarly increase. Thus, the device should ideally be altered such that a higher level of input is needed for repression of the output signal. One possible solution is to introduce additional, high affinity protein binding sites into the device. The sites may effectively be able to sequester the repressor for low inputs. Thus, such a modification may increase the size of the low gain region (for low input) resulting in enhanced noise margins. The additional protein binding sites should be nonfunctional, meaning that they have no direct effect on the output signal.

Inclusion of nonfunctional DNA binding sites requires the addition of another reaction to the list of biochemical reactions. The dimer binds to this nonfunctional DNA binding site with a dissociation constant of K_6.

$$p_i \cdot p_i + d_a \overset{K_6}{\longleftrightarrow} p_i \cdot p_i \cdot d_a$$

By again assuming that there are a fixed number of nonfunctional binding sites ($[d_a]^T$) and that these DNA molecules are partitioned between the different binding states (d_a and $p_i \cdot p_i \cdot d_a$) according to a Boltzmann distribution, I obtain the following equations.

$$[d_a]^T = [d_a] + [p_i \cdot p_i \cdot d_a] \tag{3.24}$$

$$[d_a] = [d_a]^T \frac{[d_a]}{[d_a] + K_1^{-1} K_6^{-1} [d_a][p_i]^2} \tag{3.25}$$

$$[p_i \cdot p_i \cdot d_a] = [d_a]^T \frac{K_1^{-1} K_6^{-1} [d_a][p_i]^2}{[d_a] + K_1^{-1} K_6^{-1} [d_a][p_i]^2} \tag{3.26}$$

The addition of these reactions requires modification of the final kinetic model describing the transfer curve.

$$0 = [p_i] + 2K_1^{-1}[p_i]^2 + \frac{2[d_o]^T[p_i]^2}{K_1K_2 + [p_i]^2} + \frac{2[d_n]^T[p_i]^2}{K_1K_5 + [p_i]^2} + \frac{2[d_a]^T[p_i]^2}{K_1K_6 + [p_i]^2} - S_i\alpha_i$$
$$S_o = \frac{K_1K_2}{K_1K_2 + [p_i]^2} \tag{3.27}$$

The purpose of including these nonfunctional binding sites is to increase the size of the low gain region in the transfer curve in order to enlarge the noise margins. Therefore, at low input signal, the input protein should preferentially bind to the nonfunctional sites. In other words, the input protein should have greater affinity for the nonfunctional binding sites than for the operator DNA. To ensure this result, the following inequality must hold.

$$K_6 < K_2 \tag{3.28}$$

The number of nonfunctional binding sites $[d_a]^T$ determines how much the noise margins are increased in the transfer curve. When $[d_a]^T = 0.1 * \alpha_i$ and $K_6 = K_2/10$, the noise margin N increases to 0.1 (Figure 3-12). Thus, the inclusion of nonfunctional binding sites leads to a much improved transfer curve as long as these sites have higher affinity than the operator sites and there are a sufficient number of them to actually make an impact on the transfer curve. Interestingly, by including these alternate binding sites, the dimerization affinity need not be as weak (Figure 3-9), because the alternate binding sites are used to sequester low levels of protein dimer. Hence, the new design permits dimerization affinities that are more characteristic of dimerization domains.

The explicit inclusion of nonfunctional binding sites leads to a better transfer curve than nonspecific DNA binding alone. I can design nonfunctional binding sites to have a higher DNA binding affinity than the functional operator, whereas I assume that the input protein dimer generally has lower affinity for nonspecific DNA than its cognate operator. Thus, nonfunctional binding sites create a larger low gain region for low inputs than nonspecific DNA binding does, thereby leading to a better transfer curve.

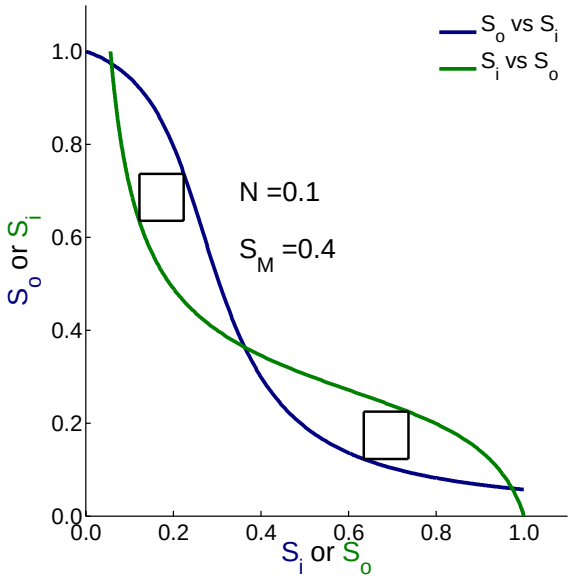


Figure 3-12: Superimposed transfer curves of a transcription-based inverter using equation set (3.27). Typical parameter values were used to generate the curve with the following exceptions: $K_1 = \alpha_i * 10$, $K_2 = \alpha_i/1000$, $K_6 = K_2/10$ and $[d_a]^T = 0.1 * \alpha_i$ (Table 3.2. The noise margin N and trip point S_M are improved as a result of the inclusion of nonfunctional alternate high affinity protein binding sites.

Additionally, since I can more easily control the number of nonfunctional binding sites than the number of nonspecific DNA binding sites, I can also ensure that the device is capable of complete repression. Natural repressors, such as lac repressor and λ cI repressor are well know for having multiple operator sites as well [108, 89]. However, the multiple operator sites are generally believed to contribute to improved repression through local concentration and DNA looping effects [108, 109, 110, 89].

3.8 Conclusions

The central challenge in designing digital logic devices using transcription factors and cognate promoters is achieving good digital behavior, as measured by device performance characteristics such as the noise margin. Many of the potential device designs explored in this chapter yield devices that either do not repress transcription completely or have a small low gain region for low input signals; devices with either

issue generally have poor noise margins. Device performance is most sensitive to the binding affinities for input protein dimerization and input protein dimer-DNA binding, respectively. Careful tuning of the binding affinities relative to the steady-state concentration of input protein can yield improved device performance. Finally, I demonstrate that device performance can be further improved via alternative device designs that include high affinity, nonfunctional protein binding sites that ensure a low gain region for low inputs without sacrificing complete transcriptional repression at high inputs.

Chapter 4

A framework for developing device family specifications

4.1 Summary

A primary goal of this thesis is to lay the groundwork for implementation of combinational digital logic using transcription-based devices. To implement combinational digital logic, devices must work in combination: the output of one device must be able to drive the input of a second device. A critical issue in combinational logic is ensuring that device signals are well-matched so that devices operate over the same range of signal values. In electrical engineering, the challenge of signal matching is addressed through device family specifications that prescribe thresholds for valid signal ranges. Devices that adhere to the device family specification should work in combination. Here, I present a framework for developing device family specifications based on device performance metrics, such as device error rate and the amount of resources that a device draws from the cellular chassis. The framework should aid device engineers in setting device family specifications and building logic devices that meet those specifications.

4.2 Introduction

Engineering many-component biological systems relies not only on designing devices that work but also designing devices that work in combination. Therefore, an overarching goal of this thesis is to lay the groundwork for implementation of combinational digital logic using transcription-based devices. Pioneering work by several groups has demonstrated the implementation of genetically-encoded digital logic devices and simple systems built from those devices [111, 112, 113]. However, eight years later, the engineering of systems from digital logic devices based on gene expression remains largely ad hoc. In the Eau d'E coli project, I showed that even if two transcription-based devices work independently, they will not necessarily work in combination (Chapter 2). Thus, reliable functional composition of devices remains a key challenge [23]. A common problem in functional composition of logic devices is that of signal level matching: the output from one device must be able to drive the input of another. If device signal levels are mismatched, then devices will not work in combination (Figure 4-1).

To address the problem of signal level matching, devices can be engineered to meet a device family specification [78]. A device family specification consists of a set of functional standards to which a device must adhere. In particular, device family specifications prescribe thresholds for device signals (Figure 4-2). Devices that adhere to a device family specification must produce signals that fall within the valid range defined by the thresholds. Such signal thresholds are set so that the output of one device can successfully drive the input of a second device, assuming both devices meet the specification. The signal thresholds of digital devices are further constrained by the requirement that the output signal must have equal or better quality than the input signal [79, 80]. In the absence of such a constraint, the signal will degrade as it is transmitted through a device and information will be lost. Therefore, the thresholds must be set so that the output signal representing logical zero should be lower than the highest acceptable input signal for logical zero, and the output signal representing logical one should be higher than the lowest acceptable input signal for

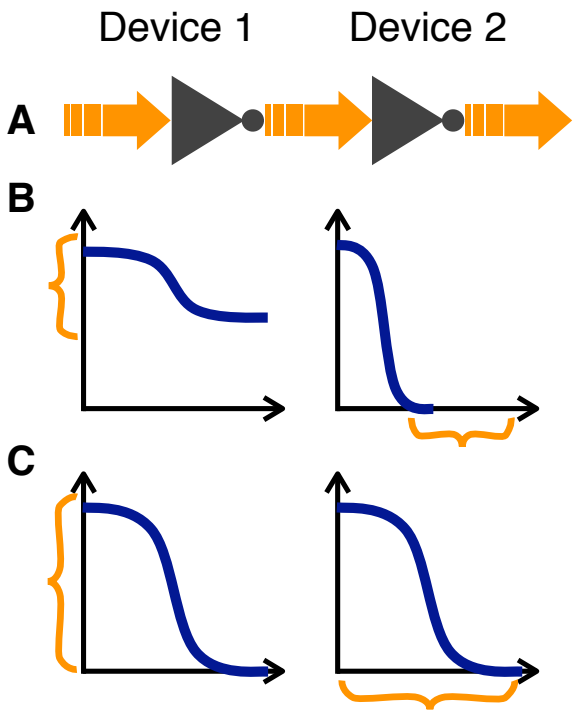


Figure 4-1: For devices to work in combination, device signal levels must matched. A) Two digital inverters are combined in series so that the output of device 1 drives the input of device 2. B) Example transfer curves of two devices whose signal levels are not matched. Device 1 switches from logical one to logical zero over a high output signal range. All device 1 output signal values are interpreted by device 2 as a logical one input signal. Thus, the two devices do not work in combination. B) Example transfer curves of two devices whose signal levels are well-matched. The output signal range of device 1 is similar to the input signal range of device 2.

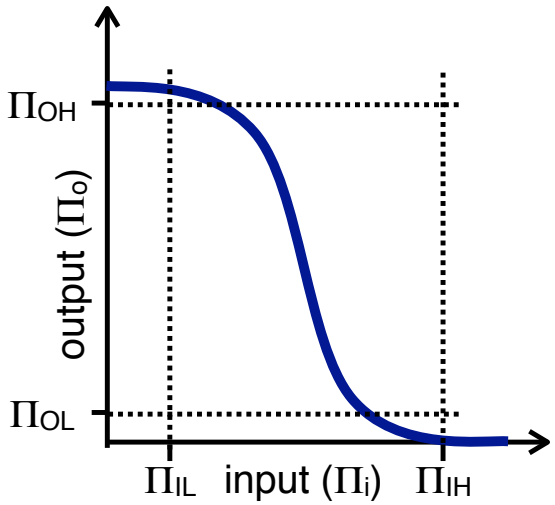


Figure 4-2: Signal thresholds for a transcriptional inverter. Π_{IL} denotes the maximum PoPS that the device will recognize as a logic zero input. Similarly, Π_{IH} denotes the minimum PoPS that the device will recognize as a logic one input. Likewise, Π_{OL} and Π_{OH} represent the threshold PoPS for the output.

logical one. These constraints on digital device signal thresholds ensure that digital devices operate reliably and accurately even in the presence of noise.

Currently, a device family specification for transcription-based devices has not been proposed. In this work, I seek to provide a framework for developing novel device family specifications. I focus on developing guidelines for establishing device signal thresholds based on two criteria: the target device demand and target device error rate. Device demand is the chassis resources that the device uses during operation [114, 115]. If the device demand is too high, a chassis will not be able to provide sufficient resources for the device to operate reliably. To build many-component systems, biological engineers must ensure that each device has a sufficiently small demand, such that the chassis can support the operation of multiple devices. The device error rate is the frequency with which a device produces an incorrect output for a valid input. Errors occur because physical device implementations are vulnerable to noise [116, 117, 118, 119, 120]. Different applications will require devices with different error rates. For example, engineered metabolic systems might tolerate reasonably high error rates as long as they achieve a target product yield. In contrast, systems used

in medical applications, such as tumor killing microbes, might require significantly lower device error rates [19]. There is an inherent tradeoff between device demand and device error rate: devices that use more resources can have lower error rates and vice versa. To establish suitable device signal thresholds, the device engineer must balance the device demand with an acceptable error rate.

4.3 Results and discussion

4.3.1 Device demand and latency increases with swing

The demand of a device is the resources that the device consumes from the cellular chassis during operation. Demand can take the form of the device occupying cellular machinery such as RNA polymerases or ribosomes, using materials such as charged tRNAs, or consuming energy such as ATP (Barry Canton, unpublished work). One convenient measure of demand is the number of amino acids synthesized per unit time [22]. Since, in general, the number of nucleotides in a transcript is approximately proportional to the number of amino acids, the nucleotide demand per unit time should scale with the amino acid demand. Similarly, assuming that the device is not saturating the chassis, the machinery and energy demands of the cell should also scale with the amino acid demand. Thus, I use translational demand as a surrogate measure for the total demand a device places on the cellular chassis.

The translational demand of a device depends on the translation rate, the steady-state mRNA level, and the protein length. The steady-state mRNA level in turn depends on the transcription rate, the mRNA degradation rate, and the DNA copy number. The device engineer can crudely control several of these gene expression parameters by altering the genetic sequence that encodes device function (Chapter 3). However, for the purpose of a framework for developing a device family specification, I assume that the device engineer can only tune the mean high signal value μ_H representing logical one. To meet a proscribed device family specification, the mean high signal value must be compatible with the signal thresholds set by a device family

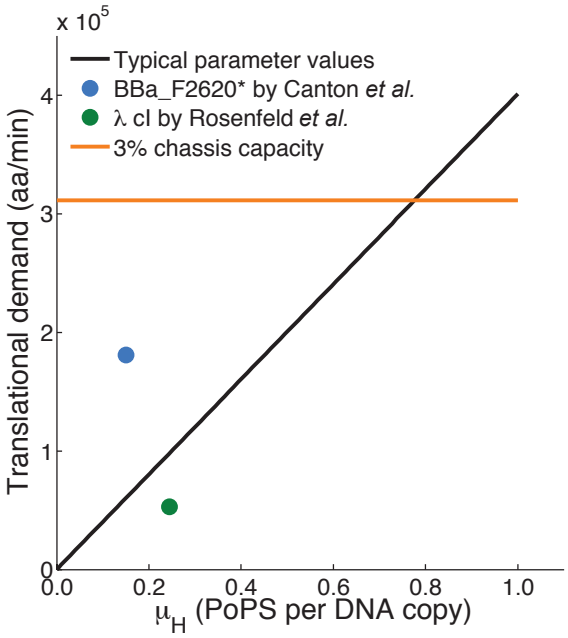


Figure 4-3: The translational demand, measured in amino acids synthesized per unit time, scales linearly with the mean high signal based on simple models of gene expression. The relationship between translational demand and mean high signal derived for a device with typical biochemical parameters is shown. Some devices, such as the receiver (BBa_F2620), place a greater demand on the cell for a given mean high signal [22]. Other devices, such as a transcriptional inverter based on λ cI repressor, place a smaller demand on the cell [124]. For reference, a line corresponding to 3% of the total chassis capacity for an *E. coli* cell with 60 minute doubling time is indicated [125].

specification. Similarly, to develop suitable device family specifications, the signal thresholds must be selected according to physically attainable mean high signal values. According to a simple model of gene expression (Chapter 3), the translational demand should scale linearly with the mean high signal (Figure 4-3). The simple model is obviously insufficient to account for the sophisticated regulation and feedback that determines a cell's capacity to provide resources to a device [121, 122, 123]. Nevertheless, it does provide an crude indicator of how much of the cell's resources a device consumes, and therefore how many devices I might expect to operate simultaneously in a cell. Thus, the percentage of the cell's capacity consumed by the device provides an upper limit to the device's mean high signal and by extension to the signal thresholds defined by a device family specification.

Although I do not explicitly consider dynamic device behavior in this work, note that the device propagation delay also increases with swing. The device propagation delay is the time it takes for a change in input signal to result in a change in output signal [78]. The device propagation delay, when switching from a low to high input, depends on the time needed for transcription, translation, and protein folding. The device propagation delay, when switching from a high to low input, depends primarily on the protein degradation rate: how long it takes for the repressor to degrade so that the output switches from low to high. For most proteins, the timescale of synthesis is of order tens of minutes [126, 22], while the timescale of degradation is even longer [97, 98, 88]. As a result, the device propagation delay is generally determined by the protein degradation rate. Higher mean signal values generally mean higher numbers of proteins and thus slower device propagation delays.

4.3.2 PoPS signals are noisy

As with all physical instantiations of digital signals, transcriptional device signals are noisy [82, 127]. Thus, I expect signals to be a distribution of values with corresponding mean μ and standard deviation σ. Noise in device signals may be intrinsic arising from stochastic fluctuations in biochemical reactions or extrinsic arising from fluctuations in cellular resources [82, 128]. However, for the purposes of developing a framework for a device family specification, I focus on noise in transcriptional device signals that can impact device performance, by causing a device to produce erroneous signals. Typically, noise in device signals is characterized by its amplitude and frequency. Noise with an autocorrelation time shorter that the device propagation delay does not impact device performance, because the noise is averaged out by the device. Noise with an autocorrelation time longer than the device propagation delay can lead to erroneous device signals. Rosenfeld *et al.* determined that intrinsic noise in gene expression has an autocorrelation time of less than ten minutes, whereas extrinsic noise has an autocorrelation time of forty minutes [124]. Thus, much of the noise relevant to device performance likely arises from extrinsic sources. Regardless, for the purposes of the this work, I do not assume anything about the source of signal

noise, only that it exists.

To our knowledge, there is no available experimental characterization of PoPS signal value distributions. Few single cell *E. coli* mRNA measurements have been performed to date. Le *et al.* measured RNA concentration in single cells after addition of an inducer [129]. They measured an initial peak of RNA concentration ∼20 minutes after induction, followed by a drop in RNA concentration, followed by another peak in RNA concentration after cell division. Similarly, Golding *et al.* measured the number of transcripts per cell as a function of time after addition of an inducer [130]. Golding *et al.* reported that transcription is bursty with exponentially-distributed periods of transcriptional inactivity punctuated by Poissonian transcriptional events. Neither Le *et al.* nor Golding *et al.* directly reported the distribution of transcription rates observed in their experiments. However, based on the supplementary data provided by Le *et al.*, I estimate that their observed PoPS signal distribution has a mean of 0.2 μM/minute and a standard deviation of 0.1 μM/minute (N=18). It is difficult to draw definitive conclusions regarding the PoPS signal value distribution given the small sample size.

Alternatively, single cell *E. coli* protein measurements are more common, but most measurements are done using fluorescence and do not report results in actual molecule number. As an exception, Rosenfeld *et al.* recently measured the protein production rate of a transcriptional inverter as a function of repressor concentration in single cells [124]. They characterized a transcriptional inverter based on λ cI repressor and the cognate P_R promoter. One version of the promoter had a wild-type O_R2 operator, and a second promoter had a mutated O_R2^* operator. In the supporting online materials, they report that the measured protein production rates, when divided by the mean protein production rate at a given repressor concentration, were log-normally distributed with mean close to 1 and standard deviation of 0.35 for the O_R2^* operator. They report obtaining a similar distribution of protein production rates at a given repressor concentration. The observed log-normal distribution is unsurprising, since log-normal distributions tend to arise for variables that are the multiplicative product of many small independent factors [131, 132, 133]. Gene expression is dependent on

a sequence of several biochemical steps that give rise to an overall protein production rate. By a similar argument, I suggest that PoPS signal distributions are also log-normal. However, the framework I present for developing device family specifications is not strictly dependent on PoPS signal values having a log-normal distribution and may be straightforwardly re-derived for other distributions.

4.3.3 Computing the device error rate from the device signal cumulative distribution function

Errors in digital logic devices occur when the device produces an incorrect output signal for a valid input signal. Errors arise because device signals are noisy. The device error rate for a transcriptional device can be considered in one of two ways. First, given a population of cells, each containing an identical genetically-encoded device that should be producing a particular output either in the high or low state, what fraction of cells are producing an incorrect output? Second, for any particular cell, what fraction of time does the cell spend producing an incorrect signal? The error rate calculated from either approach should be the same assuming that in the former approach, a sufficiently large population is measured, and in the latter approach, the signal is measured a sufficient number of times over a period longer than the autocorrelation time of the noise.

To estimate the device error rate, I use the expected signal distributions (Figure 4-4). Only signal values that adhere to the prescribed signal thresholds are considered correct signals. Initially, assume that all thresholds are set to the trip point or $\Pi_{IL} = \Pi_{IH} = \Pi_{OL} = \Pi_{OH} = \Pi_M$. Therefore, signal values that are lower than the device trip point are interpreted as logical zero, and signal values that are higher than the trip point are interpreted as logical one. Logical zero signals are usually low but occasionally assume higher values that are erroneously interpreted as logical one. Similarly, logical one signals are usually high but occasionally assume lower values that are erroneously interpreted as logical zero. Such situations yield errors in device behavior. The device error rate is computed by summing the probability that a logical

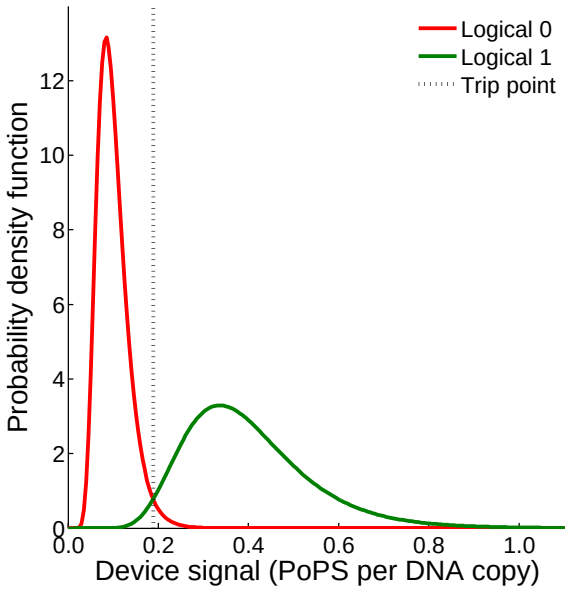


Figure 4-4: Device signals are noisy, and thus device signal values have a characteristic probability density function. Sample probability density functions for signals representing logical zero (low) and logical one (high) are shown. Both distributions are log-normal with a coefficient of variation of 0.35. A sample trip point for the device is indicated. Logical zero signal values that are greater than the trip point and logical one signal values that are less than the trip point give rise to errors.

zero signal (Π_0) is greater than the trip point and the probability that a logical one signal (Π_1) is less than the trip point. Stated differently, the device error rate is the sum of the logical zero signal error and the logical one signal error.

$$Error = P(\Pi_0 > \Pi_M) + P(\Pi_1 < \Pi_M) \tag{4.1}$$

$$Error = (1 - P(\Pi_0 < \Pi_M)) + P(\Pi_1 < \Pi_M) \tag{4.2}$$

The errors in the logical zero and logical one signals can be computed using the cumulative distribution function of the respective signal distributions. Using our assumption that PoPS-based signals are log-normally distributed, I can calculate the device error rate as a function of the mean low signal μ_L, the standard deviation in the low signal σ_L, the mean high signal μ_H, the standard deviation in the high signal σ_H, and the trip point Π_M.

$$Error = \left(1 - \left(\frac{1}{2} + \frac{1}{2}\text{erf}\left[\frac{\ln(\Pi_M) - m_L}{s_L\sqrt{2}}\right]\right)\right) + \left(\frac{1}{2} + \frac{1}{2}\text{erf}\left[\frac{\ln(\Pi_M) - m_H}{s_H\sqrt{2}}\right]\right) \tag{4.3}$$

$$Error = \frac{1}{2}\text{erf}\left[\frac{\ln(\Pi_M) - m_L}{s_L\sqrt{2}}\right] + \frac{1}{2}\text{erf}\left[\frac{\ln(\Pi_M) - m_H}{s_H\sqrt{2}}\right] \tag{4.4}$$

$$m_L = \ln\left(\frac{\mu_L^2}{\sqrt{\sigma_L^2 + \mu_L^2}}\right) \tag{4.5}$$

$$s_L = \sqrt{\ln\left(\left(\frac{\sigma_L}{\mu_L}\right)^2 + 1\right)} \tag{4.6}$$

$$m_H = \ln\left(\frac{\mu_H^2}{\sqrt{\sigma_H^2 + \mu_H^2}}\right) \tag{4.7}$$

$$s_H = \sqrt{\ln\left(\left(\frac{\sigma_H}{\mu_H}\right)^2 + 1\right)} \tag{4.8}$$

Since there is no closed form solution to the log-normal cumulative distribution function [134], I compute device error rates numerically as needed.

4.3.4 Device error rate decreases as swing increases

Although the goal of this work is ultimately to enable device engineers to develop a transcriptional device specification, I begin by assuming that the signals thresholds are already defined. The challenge then becomes designing a device that both adheres to the device specification and has an error rate appropriate for the end-user application. In designing transcriptional logic devices, the device engineer has some discretion over device design (Chapter 3). The device engineer can alter the device transfer curve by changing the translation rate, repressor-DNA binding affinity, and more [135, 136, 124]. To change the device error rate, the device engineer has the option of varying the mean or standard deviation of the signals representing logical zero and one. As discussed, the standard deviation of the PoPS signal distributions is largely dependent on noise extrinsic to the device. Extrinsic noise tends to vary based on the chassis and environmental conditions rendering it largely beyond the control of the device engineer [137]. The mean low signal tends to be set by the interaction of the repressor with the promoter and RNA polymerase. Efforts to reduce mean low signals have yielded mixed success (Chapter 5). The mean high signal, in contrast, largely depends on the promoter strength, and the relationship between promoter sequence and strength has been studied extensively [138, 139, 140, 141, 142, 143, 144, 145, 146, 147, 148]. Therefore, I can change the device error rate by changing the device swing via the mean high signal μ_H. The device error rate drops off with increasing swing (Figure 4-5). In designing a device to meet a particular specification, device engineers must select a swing that is high enough to yield a suitable device error rate but low enough to avoid placing too much demand on the chassis resources.

4.3.5 Choosing appropriate signal thresholds

I now explore the dependence of error rate on signal thresholds in order to inform the selection of appropriate signal thresholds for transcription-based logic devices. Thus far, we've calculated the device error rate by assuming that logical zero and logical one signal thresholds are equal to the trip point. Generally however, signal thresholds

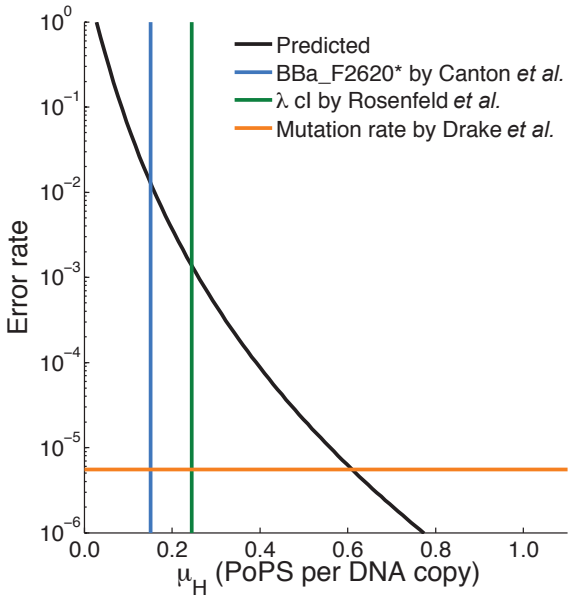


Figure 4-5: The device error rate drops off as the mean high signal μ_H increases. The mean high signal determines the device swing if the mean low signal is held constant. The relationship between error rate and mean high signal is shown for $\mu_L = 0.028$ PoPS per DNA copy, coefficient of variation for logical zero and one signal distributions of 0.35, and with a trip point Π_M selected to minimize device error rate. For reference, the estimated mean high signal for the receiver (BBa_F2620) and the λ cI-based inverter is indicated [124, 22]. The estimated mutation rate for a device with length $\sim$1 kilobase operating for 10 generations is also shown [149].

are not set to the trip point, but rather specified so that signal values representing logical zero and logical one are readily distinguished. Or more formally, the following inequalities hold.[1]

$$\Pi_{IL} < \Pi_{IH} \tag{4.9}$$

$$\Pi_{OL} < \Pi_{OH} \tag{4.10}$$

Thus, a logical zero signal is a valid low output signal only if it is less than Π_{OL}. Similarly, a logical one signal is a valid high output signal only if it is greater than Π_{OH}. (Input signals have similar valid signal ranges set by their corresponding thresholds.) The device error rate is again calculated using the cumulative distribution function of the logical zero and logical one signals, but using the relevant signal thresholds in place of the trip point. For the purposes of calculating the device error rate, I focus on the output signals produced by a device. To simplify notation, I drop the subscript O indicating output.

$$Error = P(\Pi_0 > \Pi_L) + P(\Pi_1 < \Pi_H) \tag{4.11}$$

$$Error = (1 - P(\Pi_0 < \Pi_L)) + P(\Pi_1 < \Pi_H) \tag{4.12}$$

Again, the device error rate is the sum of the error in the logical zero and logical one signals. It is a function of the mean low signal μ_L, the standard deviation in the low signal σ_L, the mean high signal μ_H, the standard deviation in the high signal σ_H, and now the signal thresholds Π_L and Π_H. Note that the error rate in the signal is dependent not on the mean and threshold values themselves but rather on their relative values, for a given coefficient of variation in the signal. Such a result can be understood intuitively by recognizing that multiplying both the mean signal value and threshold by the same factor is the equivalent of sliding the signal probability density

[1] In truth, the inequality $\Pi_{OL} < \Pi_{IL} < \Pi_{IH} < \Pi_{OH}$ should hold for digital logic devices, so that the quality of the output signal is guaranteed to be better than the input signal.

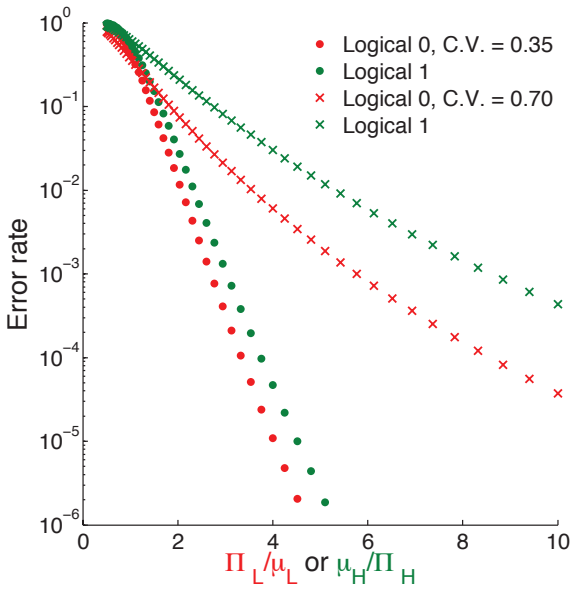


Figure 4-6: The error rate in the logical zero signal decreases roughly exponentially with the ratio of the low signal threshold Π_L to the mean low signal μ_L. Similarly, the error rate in the logical one signal decreases with the ratio of the mean high signal μ_H to the high signal threshold Π_H. The relationship between signal error rate and relative value of the mean signal and threshold are plotted assuming a coefficient of variation of either 0.35 or 0.70 for the signal distribution. The logical zero and one signals have a slightly different dependence on the relative value of the mean signal and threshold, because log-normal distributions are asymmetric.

function to the left or right. The error in the signal, derived from the cumulative distribution function, is the same regardless. The error rate of a logical zero signal decreases approximately exponentially with the ratio of the low signal threshold to the mean low signal (Figure 4-6). Similarly, the error rate of a logical one signal decreases exponentially with the ratio of the mean high signal to the high signal threshold.

Choosing appropriate signal thresholds for a new device family specification is complicated by both the summed contributions of the logical zero and logical one signal errors to the overall device error rate as well as the interdependence of the mean signal value and signal threshold. In the absence of any information on typical mean signal values and desired error rate, selection of signal thresholds is arbitrary. However, if there is any available information regarding expected mean low or high signals and desired device error rates, then selection of signal thresholds is much

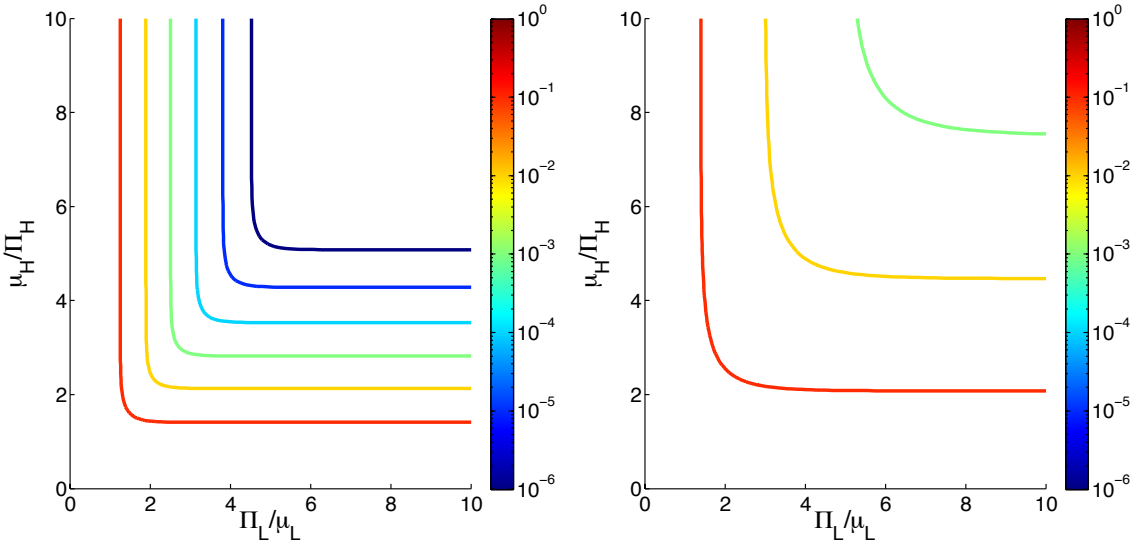


Figure 4-7: The device error rate as a function of the ratio of the mean high signal to high signal threshold and the low signal threshold to mean low signal. Contour lines corresponding to selected error rates are shown. High error rates in the logical zero signal due to small Π_L/μ_L can be compensated by high μ_H/Π_H and vice versa. The plots were obtained assuming that both the logical zero and logical one signals have a coefficient of variation of 0.35 (left) or 0.7 (right).

easier (Figure 4-7). For example, assume that a device engineer wishes to engineer a family of devices with an error rate of 0.1. Further assume that the device family specification should be compatible with Rosenfeld *et al.*'s inverter based on the λ cI repressor and cognate P_R promoter [124, 22]. Rosenfeld *et al.* measured a mean low signal of 25 proteins per cell per minute and a mean high signal of 220 proteins per cell per minute when encoded on the genome. (By way of comparison, the receiver device BBa_F2620 has a mean low signal of 900 proteins per cell per minute and a mean high signal of 30,180 proteins per cell per minute, when encoded on a medium copy plasmid [22].) Protein synthesis rates can be converted to transcription rates with units of PoPS via the following equation where b is the burst size or proteins produced per transcript [128, 22].

$$\text{PoPS} = \frac{\text{Protein synthesis rate}}{60 * b} \tag{4.13}$$

Using equation (4.13), I estimate that Rosenfeld *et al.*'s inverter has $\mu_L = 0.028$ and

$\mu_H = 0.24$. Finally, assume that the device signal is quite noisy and has coefficient of variation of 0.7. Then to achieve an error rate of 0.1, I can set $\Pi_L = 2 * \mu_L$ and $\Pi_H = \mu_H/2.6$ (Figure 4-7). The chosen thresholds constitute a valid device specification, because they satisfy inequality (4.10) as required for digital logic. I estimate the translational demand of Rosenfeld *et al.*'s inverter to be approximately 0.5% of chassis capacity (Figure 4-3), so under this device family specification I can operate 20 devices in parallel and only consume 10% of the chassis capacity. It is worth noting that in some cases, selection of appropriate signal thresholds is impossible. Consider again the example of Rosenfeld *et al.*'s inverter with a signal coefficient of variation of 0.7: it is impossible to choose Π_L and Π_H to achieve an error rate of 0.01 without violating inequality (4.10). Rosenfeld *et al.*'s inverter based on the λ cI repressor and cognate P_R promoter has an estimated error rate of 0.0014, based on the measured coefficient of variation in the protein production rate signal of 0.35.

In practice, the process of developing device family specifications will likely be an iterative process of building multiple transcription-based devices, prescribing a device family specification that is compatible with those devices, reengineering of the devices to better meet the specification, and refining the specification again. Looking forward, assuming that useful and realizable device family specifications for transcription-based logic are developed, the challenge of engineering new devices will likely focus more and more on tuning device behavior to meet a proscribed device family specification. I anticipate that the framework presented here will still prove relevant to device engineering. Rather than choosing signal thresholds, device engineers will instead choose target mean low and high signals that both meet the device family specification and yield acceptable error rates and device demands.

4.4 Conclusions

In the future, there will be many available families of transcription-based logic. Just as electrical engineers wishing to use TTL logic devices can choose from among the 74xx, 74Lxx, 74Hxx, 74Sxx, or other families [78], biological engineers wishing to

use transcription-based logic will choose from several device family specifications. Some will be optimized for low error rates, others for low device demand, and still others will likely be optimized for parameters not considered in detail here, such as device propagation delay or portability across chassis. The work presented here sought to provide an initial framework for developing device family specifications for transcription-based logic. Critical next steps are both to develop the framework further and to engineer libraries of transcription-based devices that meet a common, proscribed device family specification.

Chapter 5

Implementation of transcription-based logic using synthetic transcription factors

5.1 Summary

A grand challenge in synthetic biology is to implement memory and logic in cells. Many potential applications in environmental remediation, therapeutics, and sophisticated materials production rely on the ability of cell to sense cues from the environment, process that information, and actuate an appropriate response. Yet the scale of information processing systems that are realizable is severely limited by the limited number of available devices. For example, many well-known engineered biological systems, such as the repressilator, the "coliroid" bacterial photography system, and the band detector, are constructed from just a handful of devices [112, 150, 151]. To address this problem, I present an initial, proof-of-principle implementation of a transcription-based inverter using a synthetic transcription factor constructed from a zinc finger DNA binding domain and a leucine zipper dimerization domain. Zinc fingers and leucine zippers are both common protein domains that have been well-studied and are readily amenable to redesign via computational and experimental

techniques. Thus, synthetic transcription factors built from zinc fingers and leucine zippers potentially offer a scalable solution to the challenge of engineering memory and logic in cells. Although the resulting device could benefit from further improvements, data on device performance suggests that the performance may be sufficient to implement transcription-based logic.

5.2 Introduction

A grand challenge in synthetic biology is to implement memory and logic in cells. With even modest amounts of information processing, a range of applications becomes realizable. Specifically, any applications that requires cells to sense a set of inputs, process that information and actuate a response depends on implementation information processing [152, 153]. As such, several groups are working to implement information processing in cells at the transcriptional, translational and post-translational level [135, 154, 155, 156, 157, 158]. In electrical engineering, digital logic has proven an attractive framework for information processing due to its reliability [77]. More importantly, the implementation of the digital abstraction using imperfect physical devices is well-understood.

The scale of the information processing systems that we can currently build in biology is limited. Transcription-based circuits are arguably the most common kind of system, yet most transcription-based genetic circuits are constructed from just a handful of components [112, 113, 159, 151]. Most available logic gates are based on bacterial repressors/activators and their cognate promoters, such as *lacI*, *tetR*, λ cI, *luxR*, and *araC* [64, 135]. Transcription-based logic devices, like most other biological devices, rely on diffusion and intermolecular binding to operate. Since the cellular chassis can often be thought of as a self-mixing system [160], logic devices cannot be reused in biological engineering as they can in electrical engineering. Therefore, a critical hurdle in implementing multi-component systems using transcription-based logic is the relatively small number of logic gates available. A central contribution of my thesis work is to demonstrate an alternative implementation of transcription-based

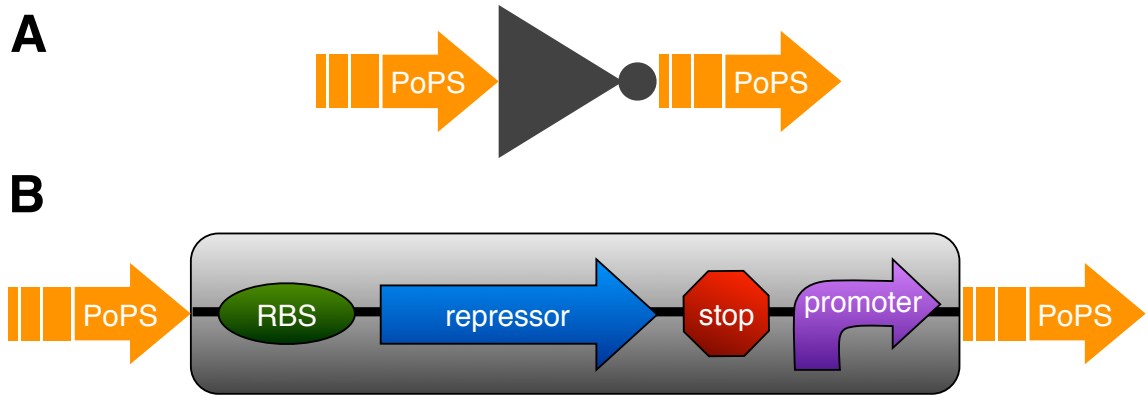


Figure 5-1: (A) An inverter receives an input, performs a logical NOT operation, and produces a corresponding output. In transcription-based inverters, both input and output signals are encoded as a transcription rate in units of PoPS. (B) An inverter is composed of four parts: a ribosome binding site (RBS), a coding sequence for a homodimeric repressor, a transcriptional terminator (stop), and a cognate promoter. (Reproduced here from Figure 3-1 for convenience.)

logic using synthetic transcription factors.

As a first step towards a general implementation of transcription-based logic, I focus on engineering a transcriptional inverter. A transcriptional inverter is composed of four genetic parts: a ribosome binding site (RBS), a repressor coding sequence, a transcriptional terminator (stop), and a transcriptional promoter with operator site(s) for repressor binding (Figure 3-1, reproduced here for convenience). Transcriptional inverters both receive an input and produce an output using the common transcriptional signal carrier PoPS [21, 22]. A high input signal to the device drives transcription of the repressor. The mRNA is subsequently translated, and the repressor folds and dimerizes. The dimerized repressor can bind to the operator site(s) of the promoter to repress transcription producing a low output signal. A low input signal to the device means that the repressor is not expressed, so the transcriptional promoter produces a high output signal.

5.2.1 Zinc fingers DNA binding domains and and leucine zippers dimerization domains are promising components for synthetic transcription factors

To implement transcription-based digital logic, synthetic transcriptions factors must fulfill two distinct roles: DNA binding and dimerization. Specific, high-affinity binding to cognate DNA ensures that we can eventually operate multiple logic devices in cells, without interference between devices or with the chassis itself. Dimerization is necessary because it provides some nonlinearity in the device transfer curve, which is critical for implementing digital logic (Chapter 3). Thus, to develop a scalable implementation of transcription-based logic, a large set of DNA binding domains and dimerization domains are needed. In my thesis work, I focused on using zinc fingers domains for DNA binding and leucine zipper domains for dimerization.

The Cys_2His_2 zinc finger DNA binding domains are an abundant family of DNA binding proteins in eukaryotes [161]. The canonical Cys_2His_2 zinc finger DNA binding domain, Zif268, consists of 3 fingers, each approximately 30 amino acids in length [162, 163, 164]. Each finger binds a DNA subsite 3-4 base pairs in length [165, 166]. Zinc fingers bind DNA with high affinities and are capable of discriminating between very similar DNA sequences [167, 168]. Moreover, zinc fingers are quite amenable to redesign via directed evolution or rational approaches to bind specifically to arbitrary DNA sequences [169, 170, 171, 172]. Individual fingers can be combined in tandem to bind longer DNA sequences [173, 174, 175].

Significant resources have been invested in developing artificial DNA binding domains in eukaryotes for industrial and medical applications [176, 177, 178, 179]. Several groups have developed libraries of zinc fingers that can bind to most of the sixty-four possible codons [180, 181, 182, 183, 184]. More recently, groups have invested effort into standardizing the design and construction of zinc finger arrays for use in targeting specific genome loci in eukaryotic cells, including accompanying software tools [185, 186, 187].

In contrast, less work has been done with respect to Cys_2His_2 zinc finger DNA

binding domains in prokaryotes. Zinc finger DNA binding domains are naturally present in some prokaryotes but are quite rare relative to eukaryotes [188]. In *E. coli*, most work to date involving zinc-finger proteins focuses on using bacterial one- and two-hybrid methods for studying DNA binding specificities and selecting new specificities [189, 104, 190, 185, 191, 192]. Such work demonstrates that three finger domains can successfully be expressed, bind DNA, and activate transcription in *E. coli*. My work seeks to build on these efforts by demonstrating that zinc finger proteins can also be used as a component in synthetic repressors for transcription-based logic in *E. coli*.

Though not as prevalent as zinc fingers, leucine zippers are also a common domain in eukaryotes [193, 161, 194]. Leucine zippers mediate protein-protein interactions and are distinguished by a characteristic heptad repeat of leucine residues [193, 195]. Each leucine zipper is an α-helix that, when bound to its cognate partner, forms a coiled-coil structure [196, 197]. Leucine zippers are well-known for their ability to homo- and heterodimerize with other leucine zippers [194]. As with zinc fingers, the interaction specificities of leucine zippers have been extensively studied [198, 199, 200]. In particular, Newman and Keating exhaustively characterized each of the pairwise interactions between the leucine zipper regions of bZIP transcription factors from humans and yeast [103]. Newman and Keating made use of protein arrays to assess binding affinity between protein pairs and validated interaction results with solution studies and previous small-scale studies. From the Newman and Keating data set, biological engineers can select leucine zipper pairs with high specificity for use in synthetic transcription factors. Although the bZIP proteins in which leucine zippers are found are a transcription factor family themselves, their DNA binding domains have not been subject to as many systematic design efforts as zinc finger proteins and therefore are less suitable for scalable synthetic transcription factor design.

5.2.2 Previous designs of dimeric zinc fingers

Carl Pabo's lab has previously designed DNA binding proteins from dimeric zinc fingers. Pomerantz *et al.* fused fingers 1 and 2 of Zif268 to the dimerization domain

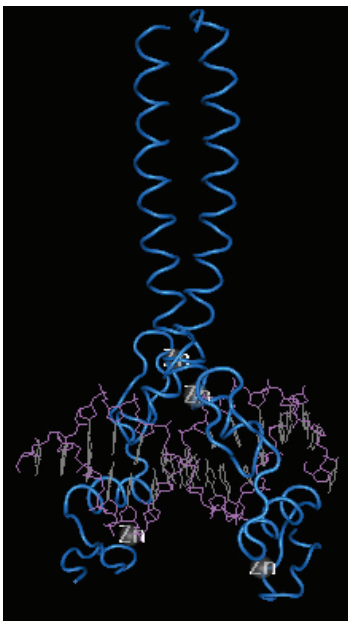


Figure 5-2: The structure of a dimeric zinc finger protein, Zif23-GCN4(-2) bound to DNA has been solved (PDB 1LLM) [203]. The ribbon diagram of the protein backbone is shown in blue, the DNA is shown in purple and gray, and the zinc ions are labeled. The two zinc fingers of each monomer bind to the DNA. The α-helix of finger 3 forms a continuous α-helix with the leucine zipper.

of GAL4 to form the fusion protein ZFGD1 [201]. The ZFGD1 homodimer bound to two six-nucleotide subsites separated by a thirteen nucleotide spacer region. They found that half-maximal binding occurred at a monomer concentration of 1 nM. Later, Wolfe *et al.* designed a different dimeric zinc finger by directly fusing the α-helix of finger 3 to the α-helix of GCN4 leucine zipper [202]. The resulting protein, Zif23-GCN4, has similar affinities as ZFGD1. Wolfe *et al.* optimized the design with phage display to achieve dissociation constants of up to $\sim$100-fold lower depending on the operator site (protein Zif23-GCN4(-2)). Wolfe *et al.* also measured the Zif23-GCN4(-2) protein to be more specific than Zif268. Finally, Wolfe *et al.* demonstrated that the Zif23-GCN4(-2) protein, when fused to an activation domain, could activate gene expression in human embryonic kidney cells. The structure of the dimeric zinc finger was also solved (Figure 5-2) [203]. Thus, dimeric zinc fingers are capable of binding DNA with high affinity and specificity and serve as a prototype for synthetic transcription factors for a transcriptional inverter in *E. coli*.

5.3 Results and discussion

5.3.1 Initial inverter design repressed transcription *in vitro* but not *in vivo*

Although Carl Pabo's lab demonstrated *in vitro* DNA binding of the designed dimeric zinc fingers, there was an open question as to whether the proteins could function as repressors in *E. coli*. I based my initial designs for a repressor upon the Zif23-GCN4 (BBa_C2002) and Zif23-GCN4(-2) (BBa_C2003) proteins designed by Wolfe *et al.*. Interestingly, I was unable to propagate a plasmid expressing Zif23-GCN4(-2) stably in *E. coli* but could propagate Zif23-GCN4. I refer to the synthetic repressor based on Zif23-GCN4 by the generic terms repressor or synthetic repressor. I used a cognate promoter (BBa_R2000) designed by the 2004 MIT IAP class in synthetic biology. To test the effectiveness of the repressor/promoter pair as a transcriptional inverter, I assembled the promoter with a GFP generator on one plasmid (BBa_I2000) and a constitutive promoter with a repressor generator on second plasmid (BBa_P20020). (A protein generator is a device that receives a transcriptional input and produces a protein output.) Cultures containing BBa_I2000, BBa_P20020, and plasmids encoding both parts together were assayed for fluorescence as a function of cell density. The fluorescence of cultures with just the promoter driving GFP production was only slightly higher than cultures with the promoter driving GFP production and the cognate repressor, suggesting that the repressor only repressed transcription slightly (Figure 5-3).

There were at least five potential reasons why the synthetic repressor failed to repress transcription significantly. First, the repressor may not bind to the cognate promoter. Second, the repressor may bind to the cognate promoter but not sterically interfere with transcription initiation by *E. coli* RNA polymerase. Third, the repressor may not be expressed in *E. coli*. Fourth, the protein may not be soluble in the cell. Fifth, the repressor may bind nonspecifically to the cell's genome. *In vitro* experiments enabled testing of four of the five potential issues in transcriptional in-

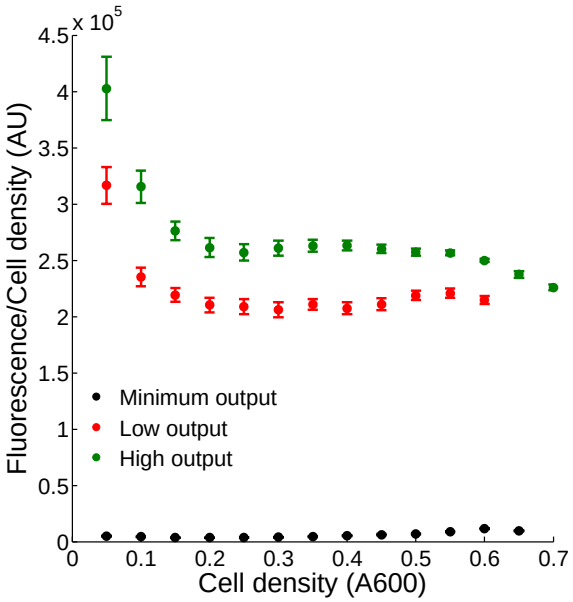


Figure 5-3: To test function of the inverter, the cognate promoter was assembled with the GFP generator. Culture fluorescence was measured over time in the presence and absence of the synthetic repressor (low and high output, respectively). The repressor was encoded on a high copy number plasmid under the control of a constitutive promoter. A plot of the fluorescence scaled to the culture cell density versus the cell density is shown. The low and high output of the inverter are similar indicating that the promoter is only slight repressed by the synthetic repressor. As a reference for the minimum possible output from the inverter, cells not expressing any GFP were also measured (minimum output).

verter function. Results from these experiments helped to focus later inverter design efforts.

To elucidate whether the protein could bind to the cognate promoter, I performed an *in vitro* electrophoretic mobility shift assay [204]. As expected from previously published work [202], the synthetic repressor binds to the cognate promoter based on the observed electrophoretic shift of the promoter DNA in the presence of repressor (Figure 5-4). A shift in the noncognate promoter DNA is only observed at very high protein concentrations, indicating that nonspecific DNA binding may be a factor. Note that the experimental setup of the electrophoretic mobility assay only provides a qualitative assessment of *in vitro* DNA binding, not a quantitative one.

To elucidate whether the repressor could sterically hinder transcription initiation, I tested the inverter *in vitro*. *In vitro* transcription reactions were electrophoresed on a native agarose gel to visualize nucleic acids. The template DNA encoding the cognate promoter controlling the GFP generator (BBa_I2000) is transcribed efficiently *in vitro* by commercially-obtained *E. coli* RNA polymerase holoenzyme (Figure 5-5, lane 3). However, if the template DNA is incubated with purified repressor prior to addition of *E. coli* RNA polymerase, then transcription is repressed (Figure 5-5, lane 4). Transcriptional repression is dependent on the addition of the repressor, since incubation with the eluant from cells not expressing the repressor was insufficient to repress transcription (Figure 5-5, lane 5). When RNA polymerase and purified repressor were added simultaneously to the template DNA, little repression occurred (Figure 5-5, lane 7). Given that I purified the 6xHis-tagged repressor under denaturing conditions with a Ni-NTA spin columns (QIAGEN), the data suggest that pre-incubation of the template DNA was necessary for the binding reaction to equilibrate to a state where most of the template DNA were bound by the correctly-folded dimeric zinc finger. Thus, although the synthetic repressor was physically capable of interfering with transcription from the cognate promoter, the reaction kinetics seemed to favor transcription by RNA polymerase over repressor binding *in vitro*. It is unclear to what extent such kinetic issues are also a factor *in vivo*.

A common issue with expression of heterologous proteins, including some zinc

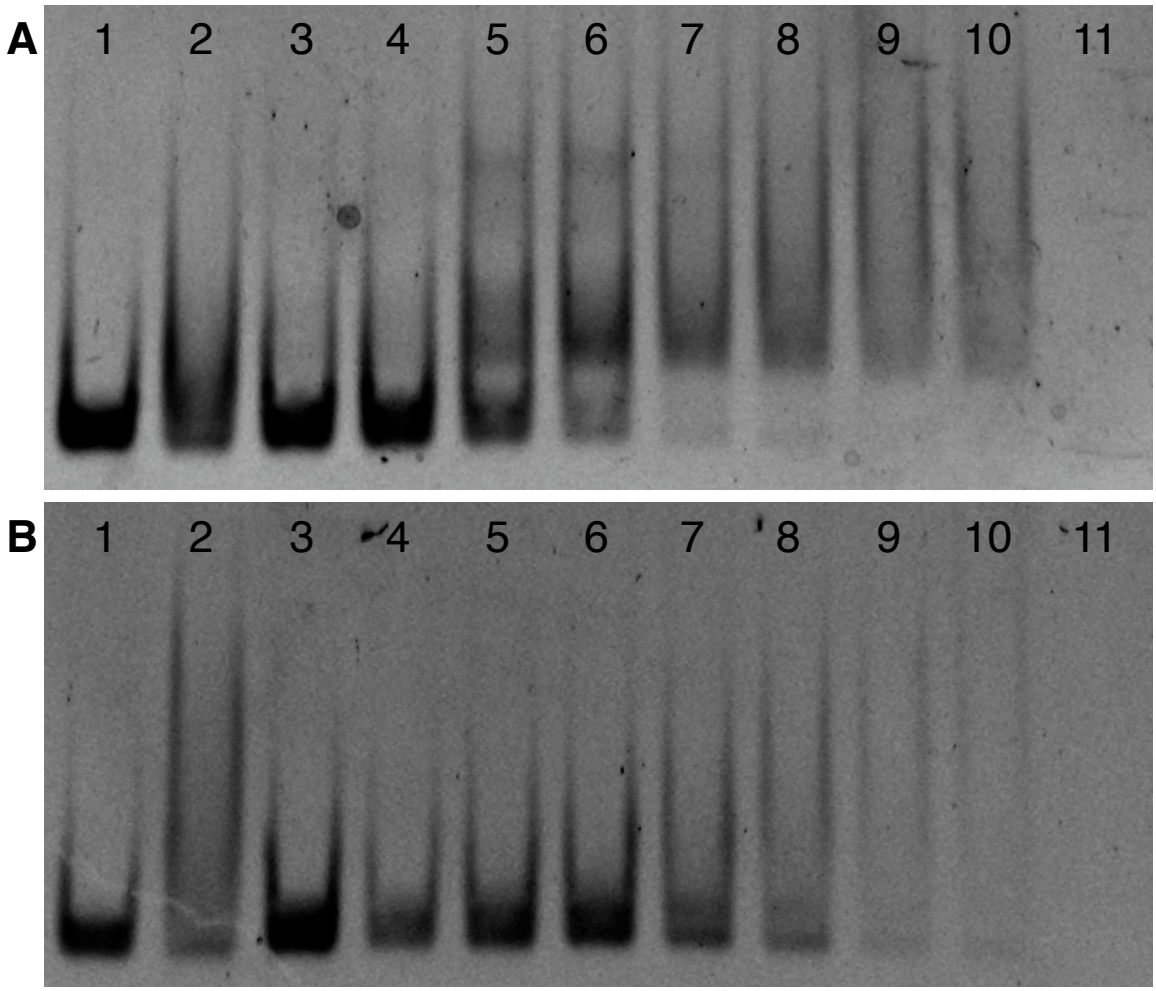


Figure 5-4: The synthetic repressor binds to the cognate promoter (BBa_R2000) but not to a noncognate promoter (BBa_R0040). A) Electrophoretic mobility shift assay results for the synthetic repressor and the cognate promoter. Lane 1 is 50 ng of linear DNA encoding the promoter. Lane 2 is DNA plus 9 μL eluant from a protein purification of cells not expressing the synthetic transcription factor. Lanes 3-10 is DNA plus increasing amounts of purified synthetic repressor (0.3 μL, 0.5 μL, 1 μL, 2 μL, 4 μL, 6 μL, 8 μL, and 9 μL). Lane 12 is purified synthetic repressor only. The promoter DNA undergoes a mobility shift in the presence of increasing amounts of synthetic repressor, indicating repressor-DNA binding. B) Electrophoretic mobility shift assay results for the synthetic repressor and a noncognate promoter. Lanes are identical to those described in (A). At the highest protein concentrations, some nonspecific retardation of DNA is apparent.

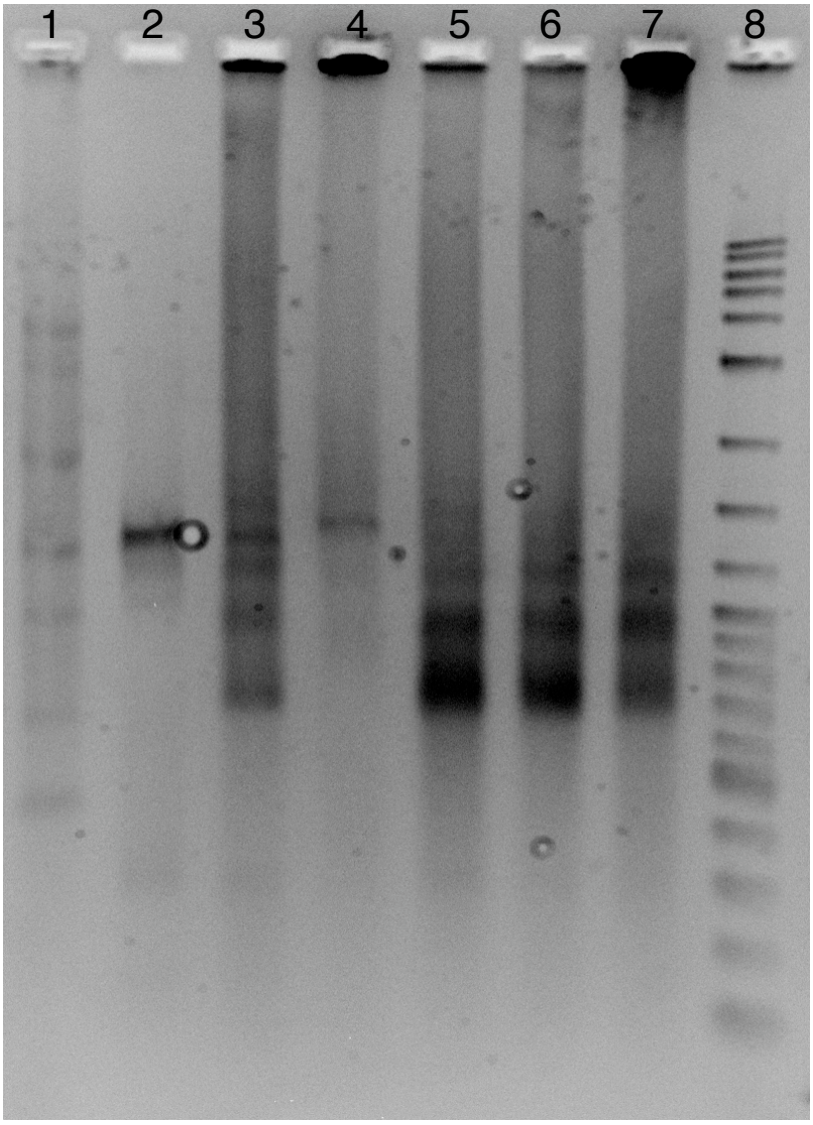


Figure 5-5: The initial design for the transcriptional inverter worked *in vitro*: transcription from the cognate promoter is repressed by the synthetic repressor (BBa_C2002) when the purified repressor is incubated with the DNA prior to addition of RNA polymerase. Lane 1 is 1 μg ssRNA ladder (New England Biolabs, Inc.). Lane 2 is linear DNA encoding the cognate promoter regulating transcription of a GFP generator (BBa_R2000 and BBa_E0840 assembled to make BBa_I2000). Lane 3 is the DNA transcribed by 2.5 units *E. coli* RNA polymerase holoenzyme (RNAP, EPICENTRE® Biotechnologies). Lane 4 is the DNA pre-incubated with purified synthetic repressor and subsequently transcribed by RNAP. Lane 5 is the DNA pre-incubated with eluant from a protein purification of cells not expressing the synthetic transcription factor and subsequently transcribed by RNAP. Lane 6 is the DNA pre-incubated with the elution buffer used during protein purification and subsequently transcribed by RNAP. Lane 7 is the DNA transcribed by RNAP in the presence of purified synthetic transcription factor. Lane 8 is 1 μg of 2-log DNA ladder (New England Biolabs, Inc.).

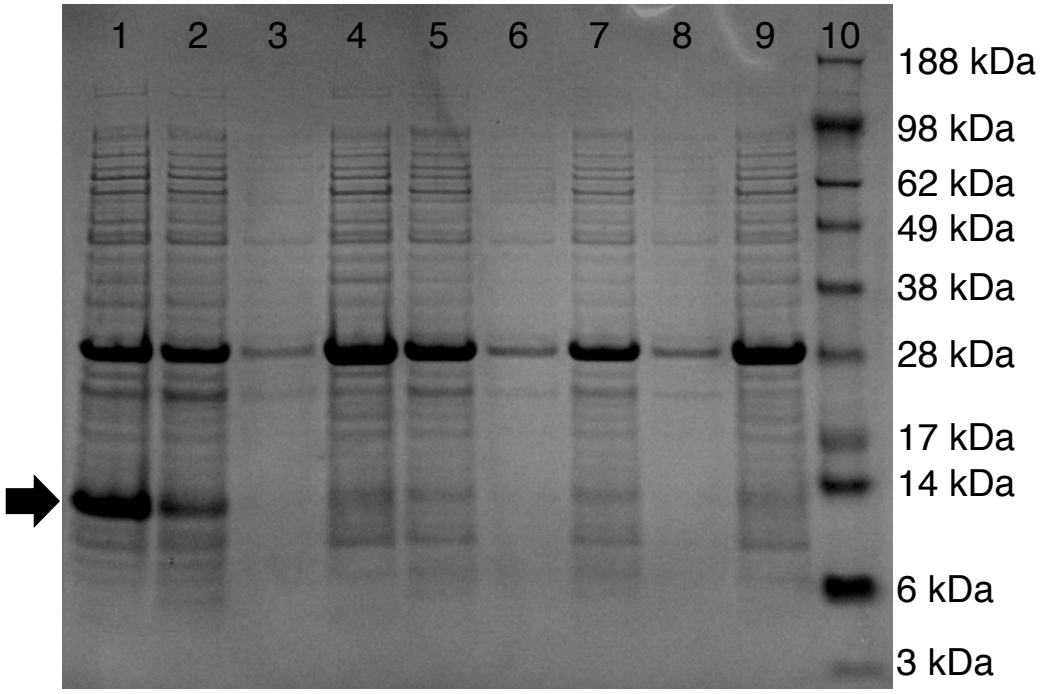


Figure 5-6: The purified synthetic repressor does not appear in the soluble fraction of crude *E. coli* cell lysates. Lanes 1-3 are the insoluble, total, and soluble crude lysate fractions, respectively, from cells expressing the synthetic repressor (BBa_C2002). The repressor is 11.5 kDa, and the corresponding band is indicated by the arrow. Lanes 4-6 are the insoluble, total, and soluble crude lysate fractions, respectively, from cells expressing a different zinc finger-leucine zipper fusion (BBa_C2100). Lanes 7-9 are the total, soluble, and insoluble fractions from cells not expressing any zinc finger protein. Lane 10 is 20 μL SeeBlue®Plus2 Pre-Stained Standard (Invitrogen). The other bright protein band in the gel corresponds to the molecular weight of GFP (27 kDa).

finger proteins, in *E. coli* is protein solubility [205, 206]. To elucidate whether either repressor expression or solubility was a factor in inverter function *in vivo*, I assayed the solubility of the synthetic repressor in *E. coli* in cultures containing the transcriptional inverter. The repressor was easily detectable in the crude cell lysate by polyacrylamide gel electrophoresis (Figure 5-6). However, although some GFP is found in the soluble fraction, little repressor was apparent in the soluble fraction of crude cell lysate, indicating that protein solubility could be a factor in transcriptional inverter function.

Taken together, the experimental data suggest that the synthetic repressor was physically capable of binding the cognate promoter and repressing transcription *in*

vitro but not *in vivo*. The protein solubility assay suggested that the solubility and folding of the protein was a factor. Finally, it may be that much of the repressor is bound nonspecifically to the genome, since the number of potential nonspecific binding sites greatly outnumbers the handful of cognate operator sites.

5.3.2 Redesigned transcriptional inverter demonstrates improved transcriptional repression

In redesigning the synthetic repressor, I made two key modifications. First, I removed the C-terminal 6xHis tag that I had added to facilitate protein purification. The 6xHis tag could possibly interfere with proper zinc coordination by the two cysteine and two histidine residues in each finger, thereby preventing proper protein folding. Second, I fused maltose binding protein domain to the N-terminus of the repressor. A N-terminal maltose binding protein fusion to the repressor offered three potential advantages to synthetic transcription factor function. First, N-terminal fusions of maltose binding protein improve the solubility of a variety of heterologous proteins in *E. coli* [207, 208]. Second, maltose binding protein has a molecular weight of ∼40 kDa [209], so it provided additional bulk to the rather small synthetic transcription factor (∼11.5 kDa). The *in vitro* transcription results suggested that the repressor was capable of repressing transcription; however, since the inverter relied primarily on steric hindrance of RNA polymerase for transcriptional repression, increasing repressor size seemed potentially helpful. Third, maltose binding protein enables easy protein purification in case additional *in vitro* experiments are necessary [210, 211]. The resulting redesigned synthetic repressor was designated BBa_C2006 and the corresponding transcriptional inverter BBa_Q20060.

To evaluate the function of the redesigned inverter, I used a PoPS generator to regulate inverter input and a PoPS reporter to measure inverter output (Figure 5-7). I used a PoPS generator (BBa_F2620) that produces a variable PoPS output dependent on the concentration of exogenous inducer [22]. A PoPS reporter receives a PoPS input and produces a signal that can be easily measured using common laboratory

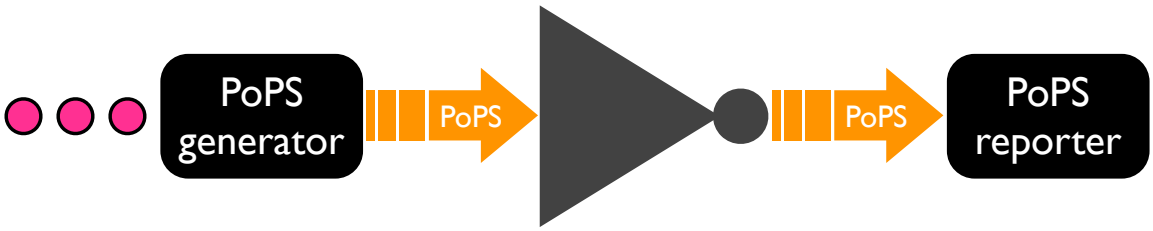


Figure 5-7: To measure transcriptional inverter behavior, the inverter input was regulated using an inducible PoPS generator (BBa_F2620) [22]. Thus, the input to the transcriptional inverter could be regulated via exogenous inducer addition to the culture. The inverter output was measured by a PoPS reporter: either a β-galactosidase activity-based reporter derived from *lacZ*α (BBa_E0435) or a fluorescence-based reporter derived from *gfp* (BBa_E0840).

equipment. I used two different PoPS reporters to measure inverter output: one based on *lacZ*α and one based on *gfp* [212, 213]. Consistency in measured behavior across two different reporters provides confidence that the inverter is working as intended. β-galactosidase assays indicated the inverter has a 2-3-fold difference between its high and low output (Figure 5-8). Measurements using the second PoPS reporter, GFP fluorescence as measured via flow cytometry, confirmed that the redesigned inverter demonstrates 2-3-fold repression (Figure 5-9). To further characterize inverter function, I used the same measurement rig and assayed culture fluorescence at several inducer concentrations. Although the resulting data does not depict a true device transfer curve since the axis units are not defined in the common signal carrier PoPS, the data do indicate that the inverter performs a logical NOT of its input to produce its output as designed (Figure 5-10).

5.3.3 Repression by synthetic transcription factor is specific for cognate promoter

Although the transcriptional inverter worked consistently in different assays, I could not rule out the fact that the measured repression could be due to a nonspecific

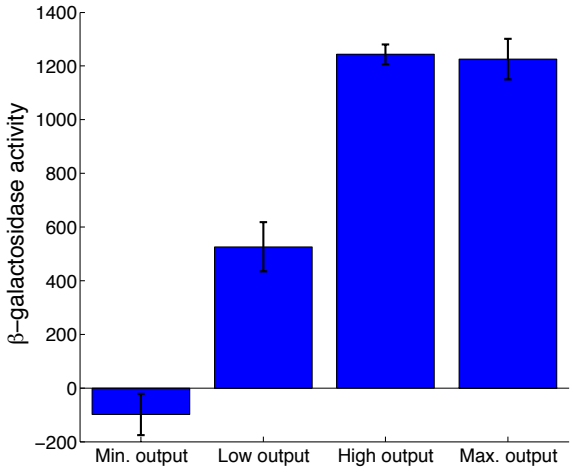


Figure 5-8: β-galactosidase activity was used as a reporter for inverter output. In the absence of lacZα, cell cultures show no β-galactosidase activity (minimal output). Cultures of the inverter, grown in the presence of AHL to induce a high input signal to the device, show some β-galactosidase activity (low output). Cultures of the inverter, grown in the absence of AHL so there is little input to the device, show 2-3 fold higher β-galactosidase activities (high output). Cultures of just the cognate promoter driving transcription of *lacZ*α show similar levels of β-galactosidase activity (maximum output). Data shown are averages of triplicate measurements of cultures grown from three individual colonies with an A600 between 0.16 and 0.22. Error bars are 95% confidence intervals of the mean.

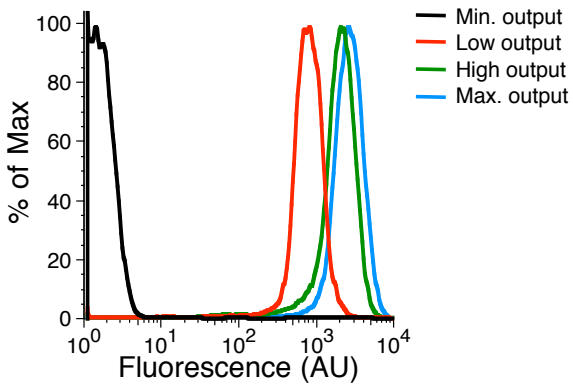


Figure 5-9: GFP fluorescence, as measured by flow cytometry, was used as a second reporter for inverter output. In the absence of GFP, cells produce little fluorescence (minimal output). Cultures of the inverter, grown in the presence of AHL to induce a high input signal to the device, yield cells with some fluorescence (low output). Cultures of the inverter, grown in the absence of AHL so there is little input to the device, yield cells with 2-3 fold higher fluorescence (high output). Cultures of just the cognate promoter driving transcription of *gfp* yield cells with similar levels of fluorescence (maximum output).

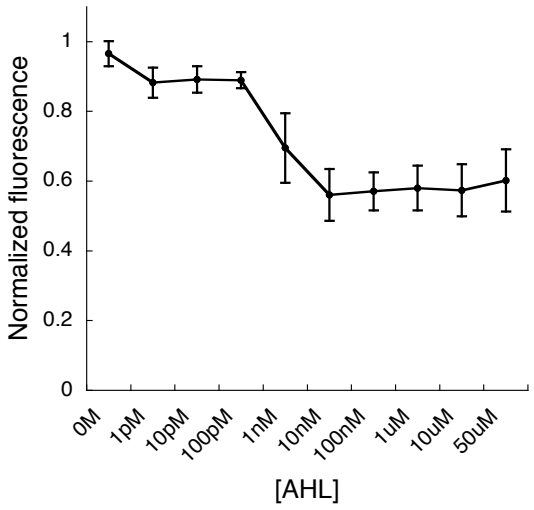


Figure 5-10: To characterize inverter output as a function of input, cultures of the inverter were grown in the presence of various inducer concentrations and their population fluorescence measured. The culture fluorescence, scaled to the cell density and normalized, is plotted as a function of inducer concentration ([AHL]). The inverter performs a logical not of its input to produce its output as designed. Data shown are averages of triplicate measurements of cultures grown from three individual colonies. Error bars are 95% confidence intervals of the mean.

impact of repressor expression on the PoPS reporter. In some cases, expression of one protein may cause a drop in expression of another due to, for example, excessive demands on cellular resources. To confirm that the measured drop in inverter output was not due to nonspecific effects, I co-transformed a plasmid encoding the inverter measurement rig containing the β-galactosidase activity-based reporter with a second plasmid encoding a noncognate promoter containing a fluorescence-based reporter (BBa_I7101). Expression of the redesigned synthetic repressor caused a clear drop in β-galactosidase activity but no measurable drop in fluorescence (Figure 5-11). Thus, the redesigned synthetic repressor appears to be specifically repressing transcription from its cognate promoter. As a second test for nonspecific repression, I replaced the cognate promoter in the inverter measurement rig using the fluorescence-based reporter with a non-cognate promoter. Again, there was no measurable drop in fluorescence upon expression of the redesigned synthetic repressor, as measured by flow cytometry (Figure 5-12). Taken together, the data demonstrate that the inverter

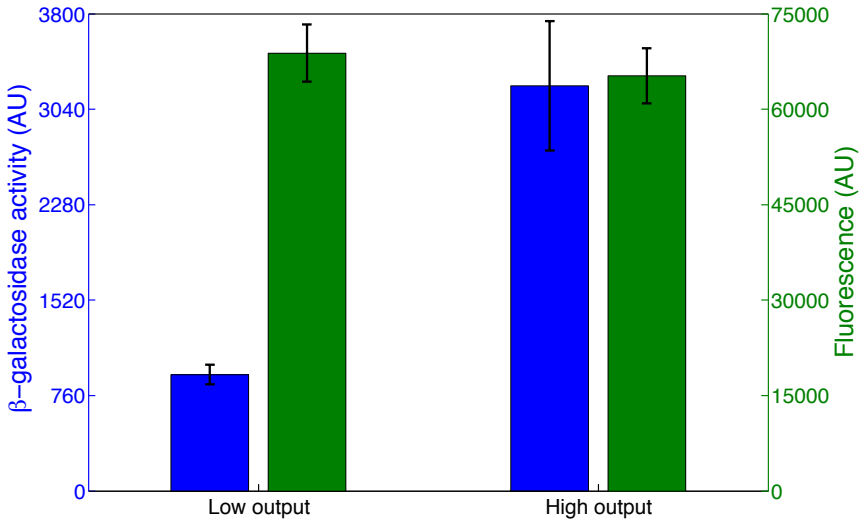


Figure 5-11: To establish whether the measured repression was due to nonspecific effects, a plasmid encoding the inverter measurement rig using the β-galactosidase activity-based reporter was co-transformed with a second plasmid encoding a noncognate promoter driving transcription of *gfp*. Cultures grown in the absence of inducer so there is little input to the device have high levels of both β-galactosidase activity and fluorescence (high output). Cultures grown in the presence of inducer to induce a high device input have a lower level of β-galactosidase activity but the same level of fluorescence. Thus, redesigned synthetic repressor expression does not impact expression from a noncognate promoter. Data shown are averages of triplicate measurements of cultures grown from three individual colonies with an A600 close to 0.1. Error bars are 95% confidence intervals of the mean.

encoded by BBa_Q20060 represents the first functional inverter based on zinc finger and leucine zipper technology. It serves as a proof-of-principle that transcription-based logic based on synthetic transcription factors is tractable, and thus that a scalable implementation of transcription-based logic may be a realizable goal.

5.3.4 Biochemical mechanism of transcriptional inverter is unclear

An underlying premise behind the use of zinc finger and leucine zipper domains is that the resulting synthetic transcription factors may be more amenable to redesign and reengineering than naturally-occurring repressors. The implication is that we, as biological engineers, "understand" the function of zinc fingers and leucine zippers to

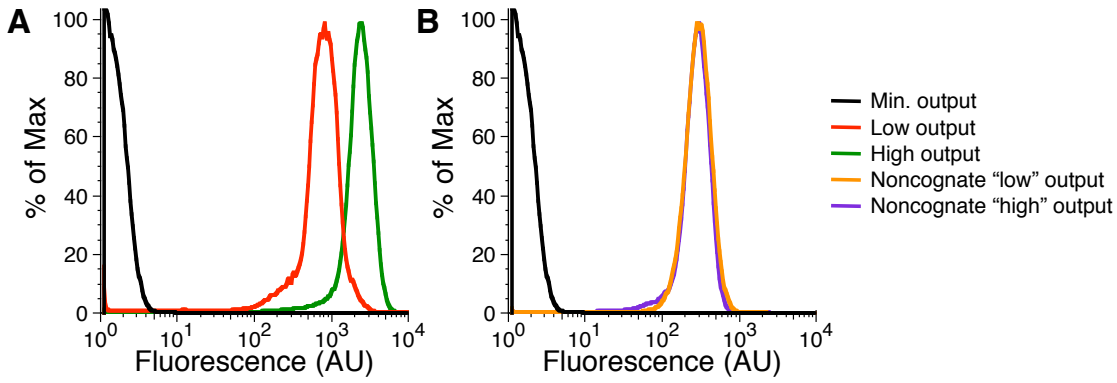


Figure 5-12: A) Cultures with the fluorescence-based inverter measurement rig demonstrate a drop in fluorescence when grown in the absence versus presence of inducer (high versus low output). B) When the cognate promoter is replaced by a noncognate promoter (BBa_R0040), there is no observed drop in fluorescence between cultures grown in the absence versus presence of inducer (high versus low output). Thus, redesigned synthetic repressor expression does not impact expression from a noncognate promoter, even in the absence of an operator site.

a greater extent than natural repressors. To test this hypothesis, I made three mutants of the redesigned synthetic repressor (BBa_C2006); I designed each mutant to impede function of the repressor. The first mutant was a cysteine to serine mutation at position 396 (BBa_C2009). The cysteine is one of the four residues responsible for zinc coordination in finger 2 of the repressor [214, 215, 206]. Mutation of zinc coordination residues often, but not always, disrupts zinc finger folding and function [216, 217, 218]. The second mutant was a deletion of the zinc finger domain from the redesigned synthetic repressor, resulting in a direct fusion of maltose binding protein to the leucine zipper (BBa_C20101). Such a deletion should eliminate any DNA binding. The third mutant was a frameshift mutation that resulted in a complete deletion of both the zinc finger and leucine zipper domains, leaving only the maltose binding protein in place (BBa_C20091). Each mutant was characterized in the measurement test rig using the β-galactosidase activity-based PoPS reporter. The function of the device with a C396S mutation in the repressor was indistinguishable from the inverter encoded by BBa_Q20060 (Figure 5-13). Thus, the cysteine to serine mutation was insufficient to impact the structure and DNA binding of the redesigned synthetic

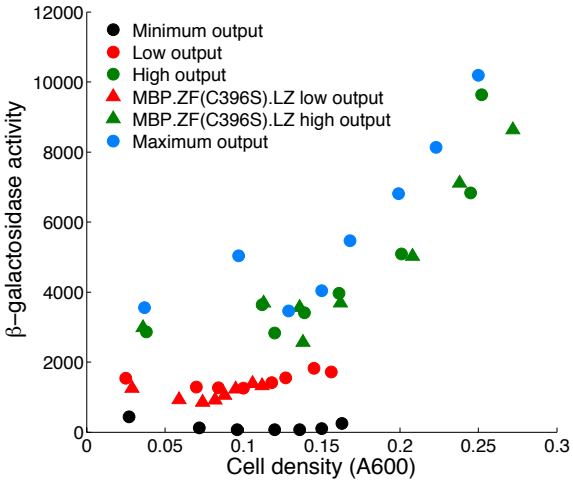


Figure 5-13: Mutation of a cysteine involved in zinc coordination in finger 2 to a serine does not impact repression. A plot of culture β-galactosidase activity as a function of cell density, as measured by the absorbance at 600nm (A600), is shown. In the absence of lacZα, a culture shows no β-galactosidase activity (minimal output). Cultures of the inverter and the C396S mutant, grown in the presence of AHL to induce a high input signal to the device, show similar levels of β-galactosidase activity (low output and MBP.ZF(C396S).LZ low output, respectively). Cultures of the inverter and the C396S mutant, grown in the absence of AHL so there is little input to the device, show similar higher β-galactosidase activities (high output and MBP.ZF(C396S).LZ high output, respectively), as does a culture of just the cognate promoter driving transcription of *lacZ*α (maximum output).

repressor. The device based on a repressor with a deleted zinc finger DNA binding domain demonstrated approximately half the repression of the inverter (Figure 5-14). The result was unexpected, since deletion of the zinc fingers should completely eliminate repression of β-galactosidase activity if the zinc finger domain is responsible for DNA binding. Deletion of both the zinc finger and leucine zipper domains eliminated any observed repression (Figure 5-15).

The experimental result that deletion of the zinc finger domain does not completely eliminate repression is at odds with my previous result that the redesigned synthetic repressor does not repress a noncognate promoter at all. If the zinc finger domain were entirely responsible for the specific binding of the redesigned synthetic repressor to the operator DNA and therefore specificity of repression of the cognate promoter, then deletion of the zinc finger domain should completely eliminate transcriptional

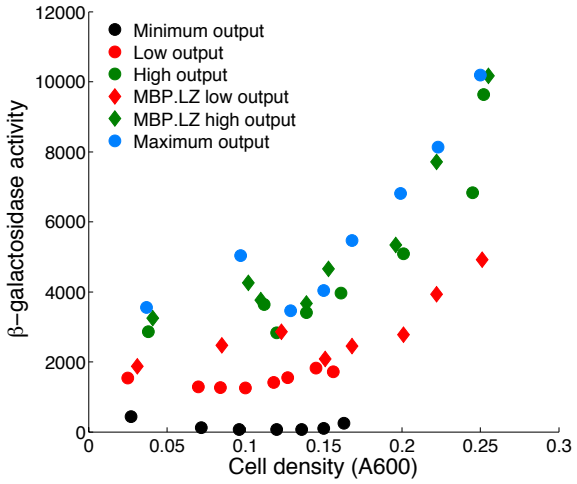


Figure 5-14: Deletion of the entire zinc finger DNA binding domain from the redesigned synthetic repressor reduces repression. A plot of culture β-galactosidase activity as a function of cell density, as measured by the absorbance at 600nm (A600), is shown. In the absence of lacZα, a culture shows no β-galactosidase activity (minimal output). A culture of the inverter, grown in the presence of AHL to induce a high input signal to the device, shows some β-galactosidase activity (low output). A culture of the device with a zinc finger domain deletion has slightly higher β-galactosidase activity when provided with a high input (MBP.LZ low output). Cultures of the inverter and the deletion mutant, grown in the absence of AHL so there is little input to the device, show higher β-galactosidase activities (high output and MBP.LZ high output, respectively), as does a culture of just the cognate promoter driving transcription of *lacZ*α (maximum output).

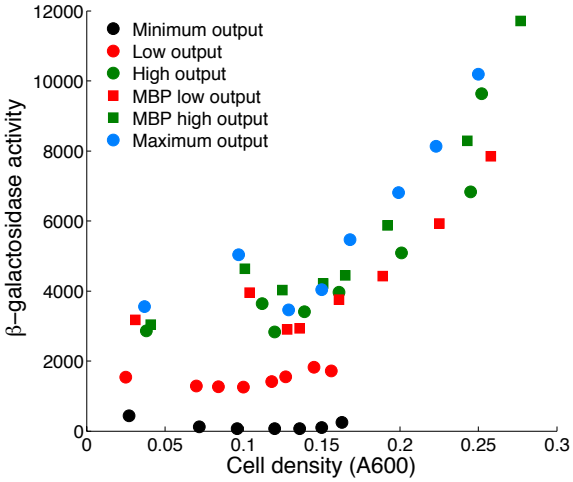


Figure 5-15: Deletion of both the zinc finger DNA binding domain and the leucine zipper dimerization domain from the redesigned synthetic repressor eliminates repression. A plot of culture β-galactosidase activity as a function of cell density, as measured by the absorbance at 600nm (A600), is shown. In the absence of lacZα, a culture shows no β-galactosidase activity (minimal output). A culture of the inverter, grown in the presence of AHL to induce a high input signal to the device, shows some β-galactosidase activity (low output). A culture of the inverter, grown in the absence of AHL so there is little input to the device, shows higher β-galactosidase activity (high output), as does a culture of just the cognate promoter driving transcription of *lacZ*α (maximum output). Similarly, cultures of the deletion mutant show the same high level of β-galactosidase activity, irrespective of the input provided to the device.

repression. In the absence of any unforeseen experimental errors, the data suggest that both the zinc finger and leucine zipper domains contribute to DNA binding specificity and thus transcriptional repression; however, I cannot definitively exclude other, alternative explanations. Additional structure-function studies are necessary to elucidate the details of how the transcription factor binds to the cognate promoter to repress transcription. Thus, although the synthetic transcription factor yields a functional, transcriptional inverter, we do not fully understand its mode of action.

5.3.5 Transcriptional inverter is sufficient for combinational digital logic

The overarching goal of this work is to lay the groundwork for a scalable implementation of transcription-based logic, so that we can eventually engineer families of hundreds of devices that work well in combination. Therefore, it is appropriate to consider, at this point, whether the measured device performance is sufficient to implement combinational digital logic. From the data, the transcriptional inverter clearly meets the basic requirement of the digital abstraction: it is capable of representing two distinguishable states given a suitable input. Estimates of the device demand and the device error rate provide additional guidance as to whether multiple transcription-based logic gates, based on the presented design, could be used in combination. Finally, I explore whether inverter performance may be tuned to meet a particular device specification.

In Chapter 4, a surrogate measure for the total device demand is the number of amino acids synthesized per unit time. Although I presented my device characterization data in arbitrary units such as β-galactosidase activity or fluorescence, absolute measurements of translational demand are straightforward, albeit labor-intensive [22]. In the case of DNA binding proteins, however, there is an additional component to the stress that a device places upon the chassis that I have thus far neglected: nonspecific DNA binding. If a DNA binding protein binds nonspecifically in the genome, particularly in intergenic regions, the cell physiology may be affected. In the *E. coli*

strain MG1655 genome sequence, there no exact matches for the 14 base pair operator site used in the inverter BBa_Q20060, nor any 1-base pair mismatches. There are 7 2-base pair mismatch operator sites in the genome, but none occur in an intergenic region. Finally, there are 148 3-base pair mismatch sites and 1292 4-base pair mismatch sites. Thus, based on sequence analysis alone, it is difficult to conclusively establish whether nonspecific DNA binding is an issue. An alternative approach to assessing the stress or demand that a device places on the chassis is measuring the device impact on growth. Cultures expressing the synthetic repressor exhibit slower growth than cultures not expressing the synthetic repressor (Figure 5-16); therefore, it appears that either device demand is exceeding cell chassis capacity, or the synthetic repressor is binding nonspecifically in the cell thereby interfering with chassis operation. Given that expression of maltose binding protein alone does not result in slowed culture growth (Figure 5-15), nonspecific DNA binding is the more likely explanation for slowed growth. Regardless of the explanation, it is likely that only a handful of transcription-based devices based on the redesigned inverter could operate simultaneously in the chassis. Future design improvements should aim to reduce the impact of the device on the chassis.

The device error rate can be estimated based on the ratio of the mean high signal to the mean low signal and the expected signal coefficient of variation (Chapter 4). According to the experimental data, the transcriptional inverter demonstrated a 2-3-fold change between the high and low output. To estimate the device error rate, I make two assumptions. First, I assume that the measured 2-3-fold repression is a reflection of the change in PoPS between the high and low state. Second, I assume that the PoPS signal of my device will have a coefficient of variation of 0.35, similar to that measured for λ cI and its cognate promoter [124]. Given the two assumptions, the device has an estimated error rate between 10-30%. Such an error rate may in fact be suitable for particular classes of applications, such as metabolic engineering, but is likely unacceptable for others, such as therapeutics. Thus, my inverter can implement digital logic but is only suitable for applications tolerant of relatively high error rates. Improvements in device performance may render this implementation of

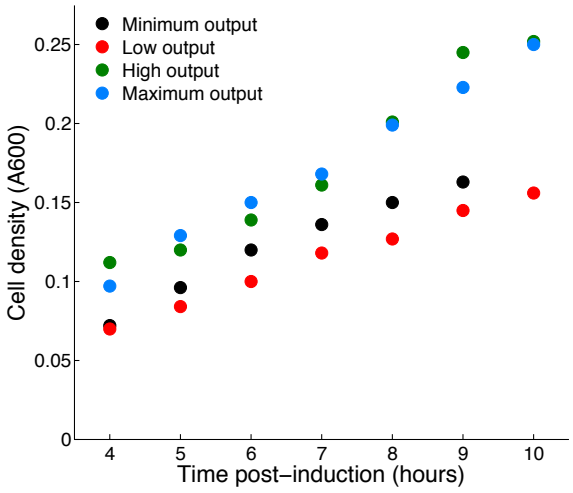


Figure 5-16: Expression of the synthetic repressor slow cell culture growth. A plot of culture cell density, as measured by the absorbance at 600nm (A600), versus time post-induction (hours) is shown. Only growth data corresponding to relevant device characterization conditions are shown. Cultures expressing the synthetic repressor (minimum output and low output) show slower growth than cultures not expressing the synthetic repressor (high output and maximum output).

transcription-based logic suitable for a broader range of applications. Note that more rigorous characterization of the device behavior would yield PoPS signal values and therefore more accurate estimates of the device error rate [22].

A final consideration in evaluating the utility of synthetic transcription factors in implementing digital logic is whether device behavior might be tuned to meet a particular device specification. Device family specifications help to ensure that devices function reliably in combination by setting thresholds for what constitutes valid input and output signals (Chapter 4). To meet a particular device family specification, the mean low and high signal values must be adjusted. One attractive mechanism for tuning device signals is through promoter redesign. Promoters are generally short and the sequence-function relationship is reasonably well-understood [138, 139, 140, 141, 142, 143, 144, 145, 146, 147, 148]. Hence, *de novo* promoter construction is easily realizable. However, a handful of failed attempts to reduce the mean low signal via redesign of the cognate promoter, by varying operator location, promoter strength, and number of operator sites, suggests that varying the mean low signal

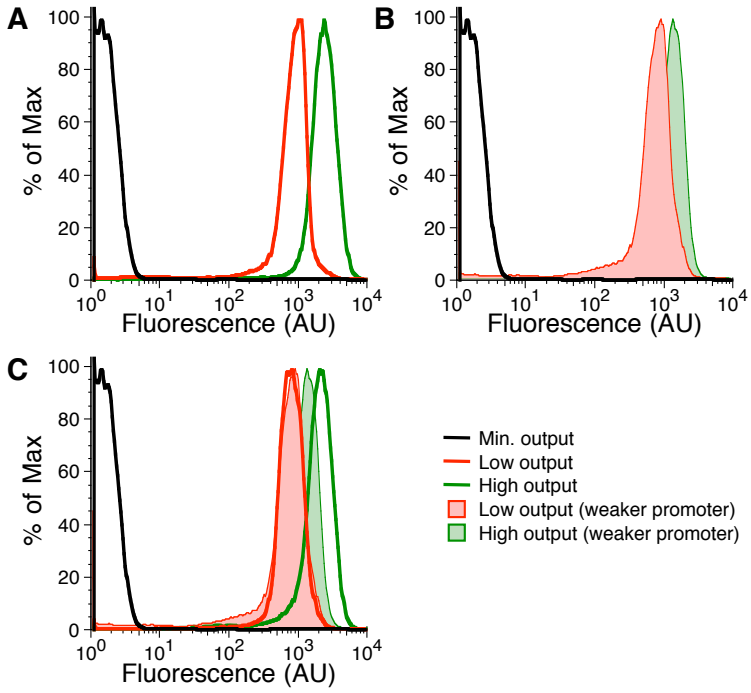


Figure 5-17: The mean high signal of the transcriptional inverter may be more readily tunable than the mean low signal. A) The minimum, low, and high output state of the transcriptional inverter BBa_Q20060. B) The minimum, low, and high output of the transcriptional inverter BBa_Q20061, with a modified promoter BBa_R2201. C) Overlay of the minimum, low, and high output of both inverters. Redesign of the cognate promoter to make it weaker lowered the high output but left the low output the same, suggesting that the mean high signal may be independently tunable from the mean low signal.

may be challenging. However, I was able to independently lower the mean high signal without affecting the mean low signal by redesigning the cognate promoter (Figure 5-17). Thus, to tune inverter function to meet a particular device specification, a combination of rational and library-based approaches may be necessary [219].

5.3.6 Synthetic versus natural transcription factors

The driving motivation behind my work on synthetic transcription factors was to overcome the limitations of the current small number of natural bacterial transcription factors available to implement transcriptional logic. Yet natural bacterial transcription factors have a distinct advantage over synthetic transcription factors in that they

have undergone stringent evolutionary selection. Evolution has yielded transcription factors that repress transcription significantly, bind to their cognate operator(s) specifically, and do not interfere with cellular growth and operation. Despite extensive studies of bacterial repressors over several decades, our understanding of how bacterial repressors work remains imperfect. Hence, our ability to design functional synthetic transcription factors is correspondingly impaired. For example, designing a synthetic transcription factor capable of significant levels of transcriptional repression was a core obstacle in my thesis work, yet natural repressors, such as *lacI*, are readily capable of over 100-fold repression [64]. As a second example, although the dimeric zinc finger protein used here was found to be quite specific *in vitro* [202], expression of the protein still adversely impacted cell growth, likely due to nonspecific DNA binding. Nevertheless, the goal of implementing memory and logic in cells is sufficiently compelling to warrant the investment needed to overcome the challenge of building multiple, orthogonal logic devices in the cell.

5.4 Conclusions

The primary contribution of this work is a proof-of-principle demonstration that a scalable implementation of transcription-based logic using zinc fingers and leucine zippers is possible. The transcriptional inverter BBa_Q20060 is capable of representing a digital signal and is potentially sufficient to implement combinational digital logic, albeit with nontrivial error rates and depressed chassis growth rate. Taken together, my work highlights twin challenges in synthetic transcription factor design: minimizing nonspecific DNA binding and decreasing mean low output signals. With respect to the former challenge, despite the work that has been invested in designing specific zinc fingers [220, 168, 104, 192], nonspecific DNA binding continues to be an issue in synthetic transcription factor design. Although the dimeric zinc finger upon which the transcriptional inverter was based has high measured specificity *in vitro* [202], it still appears to bind nonspecifically *in vivo*. With respect to the latter challenge of decreasing mean low signals, there remains a general lack of understanding of how

proteins repress transcription by RNA polymerase at the molecular-level, hindering design of synthetic transcription factors for transcription-based logic.

Chapter 6

Engineering BioBrick vectors from BioBrick parts

This chapter is based on a manuscript that I co-wrote with Drew Endy and Thomas F. Knight, Jr. [221].

6.1 Summary

The underlying goal of synthetic biology is to make the process of engineering biological systems easier. Recent work has focused on defining and developing standard biological parts. The technical standard that has gained the most traction in the synthetic biology community is the BioBrick standard for physical composition of genetic parts. Parts that conform to the BioBrick assembly standard are BioBrick standard biological parts. To date, over 2,000 BioBrick parts have been contributed to, and are available from, the Registry of Standard Biological Parts. Here we extended the same advantages of BioBrick standard biological parts to the plasmid-based vectors that are used to provide and propagate BioBrick parts. We developed a process for engineering BioBrick vectors from BioBrick parts. We designed a new set of BioBrick parts that encode many useful vector functions. We combined the new parts to make a BioBrick base vector that facilitates BioBrick vector construction. We demonstrated the utility of the process by constructing seven new BioBrick vectors. We also success-

fully used the resulting vectors to assemble and propagate other BioBrick standard biological parts. We extended the principles of part reuse and standardization to BioBrick vectors. As a result, myriad new BioBrick vectors can be readily produced from all existing and newly designed BioBrick parts. We invite the synthetic biology community to (1) use the process to make and share new BioBrick vectors; (2) expand the current collection of BioBrick vector parts; and (3) characterize and improve the available collection of BioBrick vector parts.

6.2 Introduction

The fundamental goal of synthetic biology is to make the process of engineering biology easier. Drawing upon lessons from the invention and development of other fields of engineering, we have been working to produce methods and tools that support the design and construction of genetic systems from standardized biological parts. As developed, collections of standard biological parts will allow biological engineers to assemble many engineered organisms rapidly [18]. For example, individual parts or combinations of parts that encode defined functions (devices) can be independently tested and characterized in order to improve the likelihood that higher-order systems constructed from such devices work as intended [124, 158, 22]. As a second example, parts or devices that do not function as expected can be identified, repaired, or replaced readily [222, 136].

We define a biological part to be a natural nucleic acid sequence that encodes a definable biological function, and a standard biological part to be a biological part that has been refined in order to conform to one or more defined technical standards. Very little work has been done to standardize the components or processes underlying genetic engineering [7]. For example, in 1996, Rebatchouk *et al.* developed and implemented a general cloning strategy for assembly of nucleic acid fragments [223]. However, the Rebatchouk *et al.* standard for physical composition of biological parts failed to gain widespread acceptance by the biological research community. As a second example, in 1999, Arkin and Endy proposed an initial list of useful standard

biological parts but such a collection has not yet been fully realized [224]. In 2003, Knight proposed the BioBrick standard for physical composition of biological parts [20]. Parts that conform to the BioBrick assembly standard are BioBrick standard biological parts. In contrast to the previous two examples, the BioBrick physical composition standard has been used by multiple groups [225, 226, 150, 22], and adoption of the standard is growing. For example, each summer, hundreds of students develop and use BioBrick standard biological parts to engineer biological systems of their own design as a part of the International Genetically Engineered Machines competition (`http://igem.org`). Additional technical standards defining BioBrick parts are set via an open standards setting process led by The BioBricks Foundation (`http://biobricks.org`).

The key innovation of the BioBrick assembly standard is that a biological engineer can assemble any two BioBrick parts, and the resulting composite object is itself a BioBrick part that can be combined with any other BioBrick parts. The idempotent physical composition standard underlying BioBrick parts has two fundamental advantages. First, the BioBrick assembly standard enables the distributed production of a collection of compatible biological parts. Two engineers in different parts of the world who have never interacted can each design a part that conforms to the BioBrick assembly standard, and those two parts will be physically composable via the standard. Second, since engineers carry out the exact same operation every time that they want to combine two BioBrick parts, the assembly process is amenable to optimization and automation, in contrast to more traditional *ad hoc* molecular cloning approaches.

The Registry of Standard Biological Parts (Registry) exemplifies the advantage offered by a physical composition standard such as the BioBrick assembly standard (`http://partsregistry.org`). The Registry currently maintains a collection of over 2,000 BioBrick standard biological parts. Every part in the Registry has a BioBrick part number that serves as the unique identifier of the part (for example, BBa_I51020). The Registry maintains information about each part including its sequence, function, and, if available, user experiences. DNA encoding each BioBrick standard biological

part is stored and propagated in *Escherichia coli* plasmid-based vectors [227, 228, 229, 230]. Biological engineers can obtain parts from the Registry and assemble them using the BioBrick assembly standard in order to construct many-component synthetic biological systems.

All BioBrick parts are currently maintained on a set of plasmids that includes pSB1A3-P1010, pSB3K3-P1010, pSB4A3-P1010 (section C). However, these BioBrick vectors are *ad hoc* designs that were cobbled together from common cloning plasmids such as pUC19 [231, 232, 233]. As a result, whenever a new vector is needed for use with BioBrick parts, a biological engineer must design and assemble the new BioBrick vector from scratch.

Several plasmid-based cloning systems that support the manipulation, propagation, and expression of DNA fragments have been developed [234, 235, 236, 237, 231, 232, 233, 238, 239, 64]. The Gateway® recombinational cloning system and associated vectors are arguably the closest example of a vector standard in biological research [240, 241]. For example, several genome-wide collections of open reading frames (ORFeomes) have been compiled using the Gateway® cloning system [242, 243, 244]. The Gateway® system has even been extended to allow assembly of multiple DNA fragments [245, 246]. However, the Gateway® system generally requires customized assembly strategies for each new system and therefore does not provide the advantages afforded by the BioBrick standard (above).

Thus, we sought to extend the advantages of BioBrick standard biological parts to the vectors that propagate BioBrick parts. To do this, we developed a new process for engineering BioBrick vectors. The process leverages existing and newly designed BioBrick parts for the ready construction of many BioBrick vectors. To demonstrate the utility of the new process, we constructed seven new BioBrick vectors from the base vector. We also successfully used the new vectors to assemble BioBrick standard biological parts.

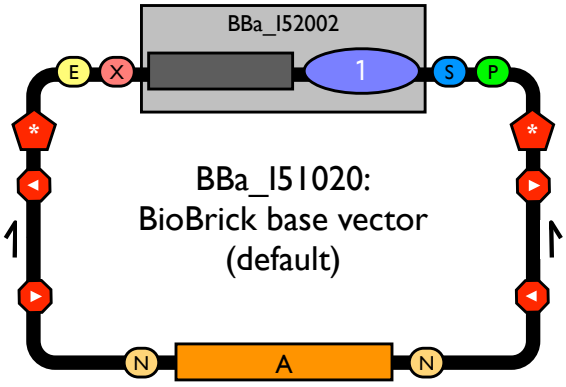


Figure 6-1: Schematic diagram of BBa_I51020: a BioBrick base vector designed to facilitate construction of new BioBrick vectors. Parts from the collection listed in Figure 6-5 were used to construct BBa_I51020.

6.3 Results

6.3.1 The BioBrick base vector (BBa_I51020)

The process for engineering BioBrick vectors from BioBrick parts is primarily based upon a newly designed BioBrick part: BBa_I51020 [Genbank:EU496089]. The new part is a BioBrick base vector that serves as a scaffold for construction of new BioBrick vectors (Figure 6-1). Starting from the base vector, new vectors can be built using plasmid replication origins and antibiotic resistance markers that conform to the BioBrick standard for physical composition. Thus, the base vector enables the ready reuse of vector parts available from the Registry of Standard Biological Parts. Use of the base vector to construct BioBrick vectors ensures standardization and uniformity in any resulting BioBrick vectors. For convenience, the base vector includes both a high copy replication origin and ampicillin resistance marker, so the base vector itself is capable of autonomous plasmid replication for easy DNA propagation and purification [247].

All BioBrick vectors derived from the BioBrick base vector have five key features. First, BioBrick vectors include a complete BioBrick cloning site to support the propagation and assembly of BioBrick standard biological parts [20]. Second, BioBrick vectors contain a positive selection marker in the cloning site to ameliorate one of

the most common problems during assembly of BioBrick parts: contamination of the ligation reaction with uncut plasmid DNA [248]. Any cells transformed with the BioBrick vector produce the toxic protein CcdB and do not grow [249, 250, 251]. Cloning a BioBrick part into the cloning site of the vector removes the toxic *ccdB* gene. Third, BioBrick vectors contain a high copy origin in the cloning site to facilitate increased yields from plasmid DNA purification [252, 253]. Again, cloning a BioBrick part into the cloning site removes the high copy origin in the cloning site thereby restoring replication control to the vector origin. Fourth, BioBrick vectors include transcriptional terminators and translational stop codons flanking the cloning site to insulate the proper maintenance and propagation of the vector from any possibly disruptive function encoded by inserted BioBrick parts [254, 255, 256, 257]. Fifth, BioBrick vectors include verification primer annealing sites sufficiently distant from the cloning site to check the length and sequence of the cloned BioBrick part. The primer annealing sites are identical to those found in commonly used BioBrick vectors, such as pSB1A3-P1010, to support backwards compatibility.

6.3.2 Constructing new BioBrick vectors using the BioBrick base vector

Constructing new BioBrick vectors starting from the BioBrick base vector requires just two assembly steps (Figure 6-2). The replication origin and antibiotic resistance marker should each be BioBrick standard parts. To construct a BioBrick vector, assemble the origin and antibiotic resistance marker via BioBrick standard assembly (first assembly step). Then, digest the resulting composite part with restriction enzymes XbaI and SpeI, and digest the BioBrick base vector with NheI to excise the ampicillin resistance marker. Next, ligate the composite origin and resistance marker to the linearized base vector (second assembly step). XbaI, SpeI, and NheI all generate compatible DNA ends that, when ligated with a DNA end from one of the other enzymes, produce a non-palindromic sequence that cannot be cut by any of the three enzymes. Thus, proper assembly of the vector eliminates any BioBrick

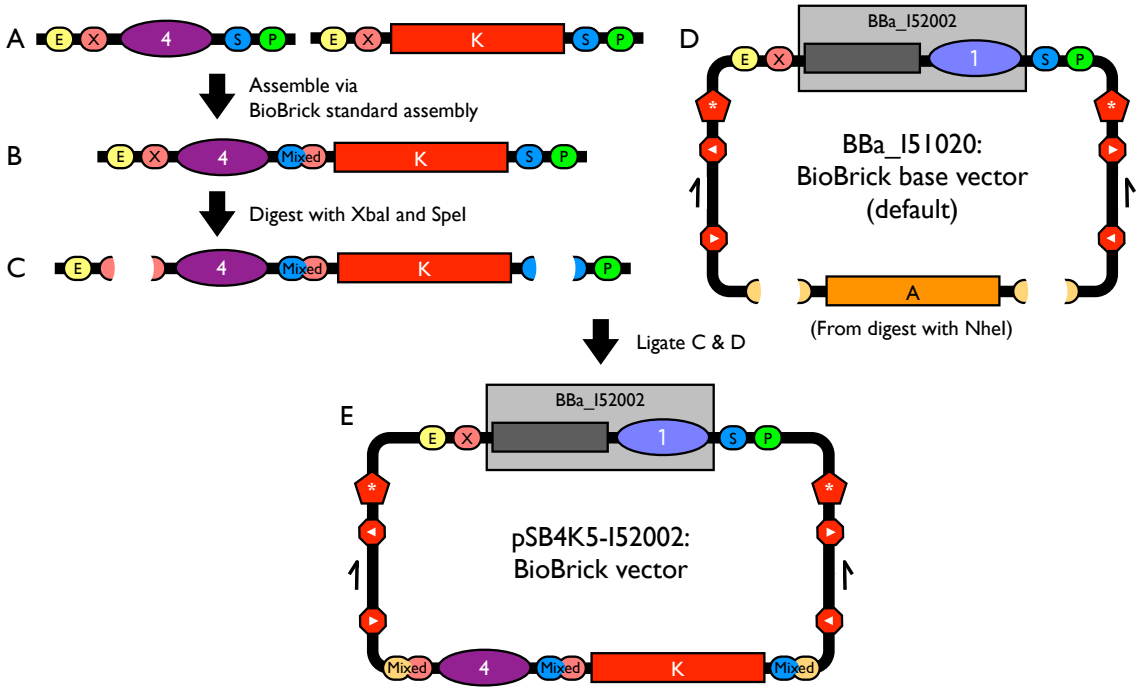


Figure 6-2: Assembly strategy for a new BioBrick vector using the BioBrick base vector BBa_I51020. (A) The replication origin and antibiotic resistance cassette should each be BioBrick standard biological parts. (B) Assemble the desired replication origin and antibiotic resistance cassette via BioBrick standard assembly to construct a composite origin and antibiotic resistance cassette. (C) Digest the resulting BioBrick composite part with XbaI and SpeI. (D) To excise the ampicillin resistance marker, digest the base vector with NheI. XbaI, SpeI, and NheI all generate compatible cohesive DNA ends that, when ligated with a DNA end from a one of the other enzymes, produce a non-palindromic sequence that cannot be cut by any of the three enzymes. Finally, ligate the digested composite origin and resistance marker to the digested base vector. (E) The result is the new BioBrick vector pSB4K5-I52002.

enzyme sites and ensures that the resulting vector adheres to the BioBrick physical composition standard. Finally, transform the ligation product into a strain tolerant of *ccdB* expression, such as *E. coli* strain DB3.1 [258, 259].

To support the construction of new BioBrick vectors, we built four new antibiotic resistance markers and two replication origins all as BioBrick standard biological parts. The four antibiotic resistance markers express proteins that confer resistance to ampicillin (BBa_P1002 [Genbank:EU496092]), kanamycin (BBa_P1003 [Genbank:EU496093]), chloramphenicol (BBa_P1004 [Genbank:EU496094]), and tetracy-

cline (BBa_P1005 [Genbank:EU496095]), respectively [260, 261, 262, 263]. The two replication origins were derived from the pSC101 (BBa_I50042 [Genbank:EU496096]) and p15A (BBa_I50032 [Genbank:EU496097]) replicons, respectively [264, 265]. We used the described procedure, base vector, and new vector parts to construct seven new BioBrick vectors: pSB4A5-I52002, pSB4K5-I52002, pSB4C5-I52002, pSB4T5-I52001, pSB3K5-I52002, pSB3C5-I52001, and pSB3T5-I52001 [Genbank:EU496098-EU496104] (section 6.5.4).

6.3.3 Assembling BioBrick parts using a new BioBrick vector

BioBrick vectors support assembly of new BioBrick standard parts. The new vectors are compatible with prefix or postfix insertions of BioBrick parts as originally described [20]. Alternatively, the new vectors also support three antibiotic based assembly (3A assembly; Figure 6-3; Shetty, Rettberg, and Knight, in preparation) [266]. 3A assembly is a method for assembling one part (the prefix part) upstream or 5' to a second part (the suffix part) in the BioBrick cloning site of a BioBrick vector (the destination vector). 3A assembly favors correct assembly of the prefix and suffix BioBrick parts in the destination vector through a combination of positive and negative selection. Briefly, 3A assembly works as follows: Digest the prefix part with EcoRI and SpeI, the suffix part with XbaI and PstI, and the destination vector with EcoRI and PstI. Then, ligate the two parts and destination vector and transform into competent *E. coli*. Plate the tranformed cells on LB agar plates supplemented with antibiotic corresponding to the destination vector resistance marker. Most of the resulting colonies should contain the composite BioBrick part cloned into the destination vector.

To confirm that our new BioBrick vectors function as expected, we assembled new BioBrick standard biological parts using four of the vectors that we constructed. To demonstrate that the composite BioBrick parts were correctly assembled using our new vectors, we performed a colony PCR amplification of the assembled parts and determined that the PCR product length was correct (Figure 4). Each part was also verified to be correct via sequencing with primers that anneal to the verification

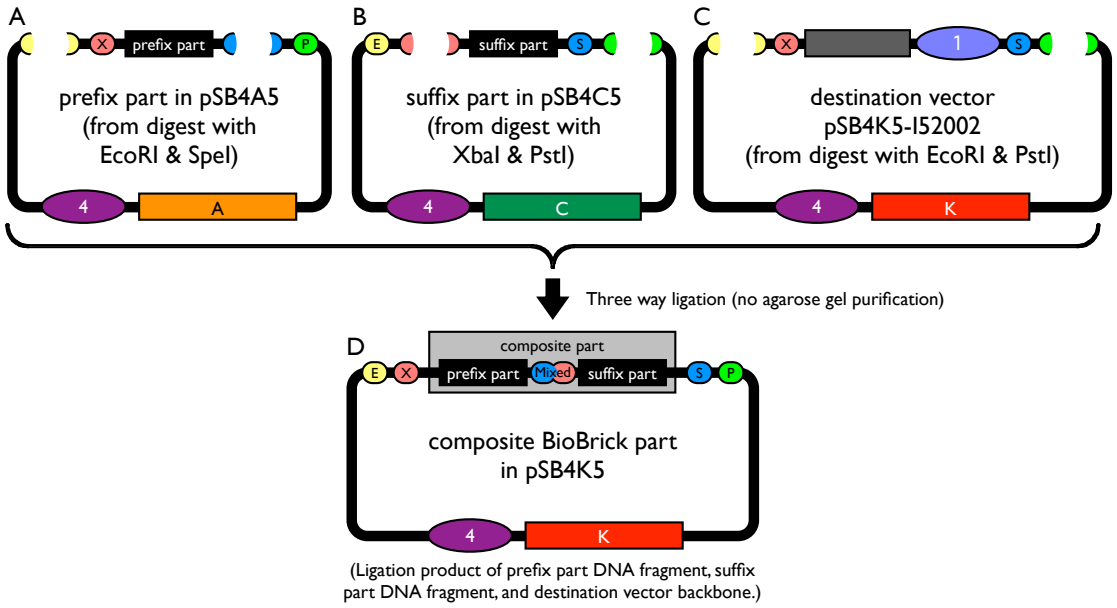


Figure 6-3: Assembly strategy for two BioBrick standard biological parts using a new BioBrick vector. (A) Digest the prefix part with enzymes EcoRI and SpeI. (B) Digest the suffix part with restriction enzymes XbaI and PstI. (C) Digest the destination vector (pSB4K5-I52002) into which the two parts will be assembled with restriction enzymes EcoRI and PstI. Without agarose gel purification of the linearized DNA, ligate the three fragments, transform into *E. coli* and plate on LB agar plates supplemented with the antibiotic corresponding to the destination vector resistance marker. (D) Most of the resulting colonies contain the composite BioBrick part cloned into the destination vector.

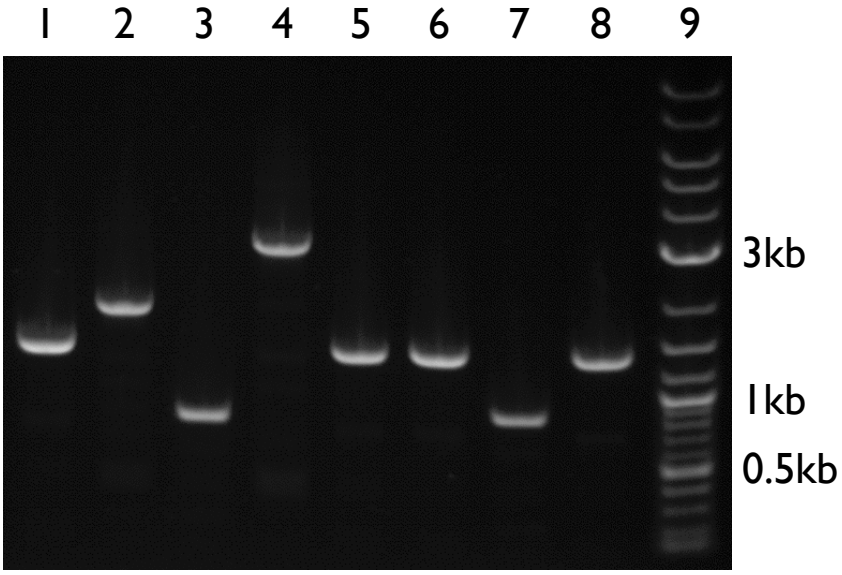


Figure 6-4: To verify the function of the new BioBrick vectors, we performed a colony PCR using primers that anneal to the verification primer binding sites. To check the length of the resulting PCR products, we electrophoresed the reactions through an 0.8% agarose gel. Lanes 1-8 are the PCR products resulting from the amplification of the following BioBrick parts cloned into new BioBrick vectors. The desired PCR product lengths are in parentheses. Lane 1 is pSB4A5-I52001 (1370 bp), lane 2 is pSB4K5-T9003 (1883 bp), lane 3 is pSB4C5-E0435 (814 bp), lane 4 is pSB4T5-P20061 (2988 bp), lane 5 is pSB3K5-I52002 (1370 bp), lane 6 is pSB3C5-I52001 (1370 bp), lane 7 is pSB3T5-I6413 (867 bp), and lane 8 is BBa_I51020 (1370 bp). Lane 9 is 1 μg of 2-log DNA ladder (New England Biolabs, Inc.). The 0.5 kb, 1 kb, and 3 kb DNA fragments in the DNA ladder are annotated.

primer binding sites (BBa_G00100 and BBa_G00102).

6.4 Discussion

We developed a new process for engineering BioBrick vectors from BioBrick parts. The process now makes possible the ready construction of many, new BioBrick vectors using the growing collection of BioBrick parts available from the Registry of Standard Biological Parts. Moreover, new BioBrick vectors can be constructed from the BioBrick base vector in just two assembly steps. Finally, any BioBrick vectors derived from the BioBrick base vector have five key features designed to facilitate the cloning,

assembly, and propagation of BioBrick parts. We used the process to construct seven new BioBrick vectors and used the vectors to assemble new BioBrick parts.

6.4.1 Design of new BioBrick vectors parts

To adhere to the BioBrick standard for physical composition, BioBrick vector parts need only be free of the BioBrick restriction enzyme sites. However, we chose to design anew all BioBrick vector parts (Figure 6-5), so that we could completely specify their DNA sequences. We compiled a list of potentially useful endonuclease sites for removal from all new BioBrick vector parts (Table 6.1). We targeted each group of endonuclease sites for removal for a different reason. We targeted recognition sites of enzymes that produce compatible cohesive ends to the BioBrick enzymes because such enzymes often prove useful in constructing new variants of BioBrick vectors. We targeted offset cutter sites because they may be useful in alternative restriction enzyme-based assembly methods [267]. We targeted homing endonuclease sites because they are commonly used in genome engineering [268]. We targeted some nicking endonuclease sites because they can be useful for specialized cloning applications [269]. Finally, we targeted several additional restriction endonuclease sites to keep them available for use by new standards for physical composition. Our list of endonuclease sites constitutes a set of target sequences that should be considered for removal from any newly synthesized BioBrick part, if possible. The target sequence set will change as the synthetic biology community develops new standards for physical composition of BioBrick parts. Some of the targeted endonuclease sites were naturally absent from the DNA sequences encoding our new vector parts. For any remaining sites, we removed the recognition sequences from the BioBrick vector parts by introducing point mutations. However, the functions of the pSC101 and pUC19-derived plasmid replication origins were sensitive to introduced mutations, so the replication origins used in this work are not free of all targeted endonuclease sites (sections 6.5.2 and 6.5.3). Similarly, issues during synthesis led to an unnecessary redesign of the *ccdB* positive selection marker, so it too is not free of all targeted endonuclease sites.

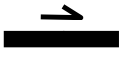
Part Number	*Function*	*Notation*
BBa_G00000	BioBrick cloning site prefix	
BBa_G00001	BioBrick cloning site suffix	
BBa_P1016	*ccdB* positive selection marker	
BBa_I50022	pUC19-derived high copy replication origin	
BBa_B0042	translational stop sequence	
BBa_B0053 & BBa_B0054	forward transcriptional terminator	
BBa_B0055 & BBa_B0062	reverse transcriptional terminator	
BBa_G00100	forward verification primer annealing site (VF2)	
BBa_G00102	reverse verification primer annealing site (VR)	
BBa_B0045	NheI restriction site	
BBa_P1006	ampicillin resistance marker (reverse orientation)	
BBa_P1002	ampicillin resistance marker	
BBa_P1003	kanamycin resistance marker	
BBa_P1004	chloramphenicol resistance marker	
BBa_P1005	tetracycline resistance marker	
BBa_I50042	pSC101 replication origin	
BBa_I50032	p15A replication origin	

Figure 6-5: The Registry part number, function, and graphical notation of each constructed BioBrick vector part are listed. The part collection includes (1) BBa_G00000: BioBrick cloning site prefix including the EcoRI (E) and XbaI (X) restriction enzyme sites, (2) BBa_G00001: BioBrick cloning site suffix including the SpeI (S) and PstI (P) restriction enzyme sites which, together with the BioBrick prefix, forms a BioBrick cloning site for compatibility with all BioBrick standard biological parts, (3) BBa_P1016: positive selection marker *ccdB* to improve yield of insert-containing clones during part assemblies, (4) BBa_I50022: pUC19-derived high copy replication origin within the BioBrick cloning site that allows for easy plasmid DNA purification of the base vector and any derived vectors, (5) BBa_B0042: a short DNA sequence that has translational stop codons in all six reading frames to prevent translation into or out of the BioBrick cloning site, (6) BBa_B0053-B0055 and BBa_B0062: forward and reverse transcriptional terminators flanking the BioBrick cloning site to prevent transcription into or out of the BioBrick cloning site, (7) BBa_G00100 and BBa_G00102: sequence verification primer annealing sites for primers VF2 and VR, (8) BBa_B0045: NheI (N) restriction site for insertion of desired replication origin and resistance marker to construct vector of interest, (9) BBa_P1006: ampicillin resistance selection marker to facilitate propagation of the base vector, (10) BBa_P1002-P1005: four antibiotic resistance markers, and (11) BBa_I50042 and BBa_I50032: pSC101 and p15A replication origins. Each part is used either as a component of the BioBrick base vector BBa_I51020 (1-9) or to construct new BioBrick vectors (10-11).

Endonuclease	*Description*
EcoRI, XbaI, SpeI, PstI	BioBrick restriction site
ApoI, MfeI AvrII, NheI NsiI SbfI	Produces compatible ends to EcoRI Produces compatible ends to XbaI and SpeI Produces compatible ends to PstI
AarI, AcuI, BbsI, BciVI, BfuAI, BmrI, BsaI, BsgI, BsmBI, BsmI, BspMI, BsrDI, BtgZI, EarI, EcoP15I, FokI, SapI, TspRI	Offset cutter
I-CeuI, I-SceI, PI-PspI, PI-SceI, I-PpoI	Homing endonuclease
Nt.BbvCI, Nt.BstNBI, Nt.AlwI	Nicking endonuclease
AgeI, AscI, BamHI, BbvCI, FseI, HindIII, KasI, NcoI, NdeI, NgoMIV, PacI, PmeI (MssI), RsrII, SacI, SalI, SfiI, SgfI, SgrAI, SrfI, SwaI (SmiI), XcmI, XhoI, XmaI, XmnI, ZraI	Restriction endonuclease

Table 6.1: Endonuclease sites targeted for removal from BioBrick vector parts.

6.4.2 Construction of BioBrick base vector

To realize our designs for new BioBrick vectors, we contracted for DNA synthesis of the four antibiotic resistance markers, pSC101 replication origin and the entire BioBrick base vector. However, synthesis of the BioBrick base vector was problematic (section 6.5.3). The issues that arose during synthesis are briefly discussed here, because they are relevant to anyone interested in synthesizing new BioBrick parts. Difficulties during synthesis stemmed from the inclusion of both a *ccdB* positive selection marker that is toxic to most *E. coli* strains and a synthetic replication origin that proved incapable of supporting replication of the BioBrick base vector. Commercial DNA synthesis processes currently rely on cloning, assembly, and propagation of synthesized DNA in *E. coli*. In general, for parts whose function are incompatible with growth and replication of *E. coli*, the processes of DNA design and DNA synthesis cannot be easily decoupled. Improvements in commercial DNA synthesis are needed that free the process from dependence on *in vivo* DNA propagation and replication.

6.5 Materials and methods

6.5.1 Design of BioBrick vector parts and the BioBrick base vector

We designed all BioBrick vector parts and the BioBrick base vector using Vector NTI®Suite 7 for Mac OS X by Invitrogen Life Science Software in Carlsbad, CA. We removed endonuclease recognition sites from the designed parts either manually or using GeneDesign vβ2.1 Rev 5/26/06 [270].

6.5.2 Construction of BioBrick vector parts

We contracted for DNA synthesis of the four antibiotic resistance markers and the pSC101 replication origin to the DNA synthesis company Codon Devices, Inc. in Cambridge, MA. The four antibiotic resistance markers (BBa_P1002-P1005) were

easily synthesized as designed. Testing confirmed that the four markers conferred resistance to the corresponding antibiotics. Synthesis of the pSC101 origin was also straightforward. However, testing revealed that our design for the pSC101 origin (BBa_I50040) was nonfunctional as a replication origin. We successfully reconstructed a functional pSC101 replication origin (BBa_I50042) via PCR of an existing plasmid. Thus, we presume that one or more of the introduced point mutations to eliminate endonuclease sites were deleterious to the plasmid replication function of the designed origin. We did not attempt to synthesize the p15A replication origin (BBa_I50032). Instead, like the pSC101 origin, we constructed p15A origin by PCR of an existing plasmid.

We constructed the functional pSC101 replication origin by PCR using pSB4A3-P1010 as a template and amplification primers I50042-f (`5'-GTT TCT TCG AAT TCG CGG CCG CTT CTA GAG CTG TCA GAC CAA GTT TAC GAG-3'`) and I50042-r (`5'-GTT TCT TCC TGC AGC GGC CGC TAC TAG TAG TTA CAT TGT CGA TCT GTT C-3'`). We constructed the p15A replication origin by PCR using pSB3K3-P1010 as a template and amplification primers I50032-f (`5'-GTT TCT TCG AAT TCG CGG CCG CTT CTA GAG ATG GAA TAG ACT GGA TGG AG-3'`) and I50032-r (`5'-GTT TCT TCC TGC AGC GGC CGC TAC TAG TAA ACA CCC CTT GTA TTA CTG-3'`). Each reaction was a mix of 45 μL PCR SuperMix High Fidelity, 31.25 pmoles each of forward and reverse primer, and 1 ng template DNA in a 50 μL total volume. The PCR conditions were an initial denaturation step of 95°C for 15 mins followed by 40 cycles of 94°C for 30 seconds, 56°C for 30 seconds, and 68°C for 2.5 minutes. Finally, the reactions were incubated at 68°C for 20 minutes. We then added 20 units DpnI restriction enzyme to each reaction to digest the template DNA. The reactions were incubated for 2 hours at 37°C and then heat-inactivated for 20 minutes at 80°C. We purified both reactions using a MinElute PCR Purification kit according to the manufacturer's directions (QIAGEN, Germany). The pSC101 and p15A origin PCR products were used directly for assembly of the BioBrick vectors.

6.5.3 Construction of BioBrick base vector

We also contracted for synthesis of the entire BioBrick base vector. However, we encountered two issues during synthesis of the base vector. First, troubleshooting efforts during synthesis compromised the design of the base vector: failed attempts to clone the base vector into an *E. coli* strain intolerant of expression of the toxic protein CcdB led to an unnecessary redesign of the *ccdB* positive selection marker in the BioBrick base vector (from BBa_P1011 to BBa_P1016 [Genbank:EU496090]). Second, faulty part design adversely impacted the synthesis process: our pUC19-based replication origin design was similarly nonfunctional, so the base vector could not be propagated as specified. Yet, synthesized DNA for the BioBrick base vector was nevertheless provided. We eventually determined that the provided DNA was actually a fusion of two slightly different copies of the base vector: one with the designed, nonfunctional version of the pUC19 origin (BBa_I50020) and one with a functional version of the pUC19 origin (BBa_I50022 [Genbank:EU496091]). To obtain a single, corrected version of the BioBrick base vector, we performed a restriction digest of the provided base vector DNA with EcoRI. We then re-ligated 1 μL of a ten-fold dilution of the linearized base vector DNA. For detailed reaction conditions, see section 6.5.5. We transformed the religated BioBrick base vector into *E. coli* strain DB3.1 via electroporation and plated the transformed cells on LB agar plates supplemented with 100 μg/mL ampicillin to obtain the corrected BioBrick base vector BBa_I51020 [258, 271, 272]. Correct construction of the BioBrick base vector was verified by DNA sequencing by the MIT Biopolymers Laboratory.

6.5.4 Assembly of BioBrick vectors

We assembled the new BioBrick vectors as described (Figure 6-2). For detailed reaction conditions, see section 6.5.5. However, we used the synthesized BioBrick base vector BBa_I51019 instead of the corrected BioBrick base vector BBa_I51020, since, at the time, we had not yet identified the issue with the provided synthesized DNA. As a result, we obtained a mixture of new vectors. Four of the constructed vectors have a

functional version of the pUC19 origin (BBa_I50022) in the BioBrick cloning site and propagate at high copy (vectors with BBa_I52002: pSB4A5, pSB4K5, pSB4C5, and pSB3K5). The other three vectors have a nonfunctional version of the pUC19 origin (BBa_I50020) in the BioBrick cloning site and propagate at low copy (vectors with BBa_I52001: pSB4T5, pSB3C5, and pSB3T5). We chose to describe all seven vectors here for two reasons. First, all seven new BioBrick vectors can be used for the propagation and assembly of BioBrick parts; the vectors pSB4T5, pSB3C5, and pSB3T5 are just slightly less convenient for plasmid DNA purification. Second, the difficulties that we encountered during construction of the BioBrick base vector are illustrative of the current interdependence of DNA design and DNA synthesis (section 6.4.2).

6.5.5 Assembly of BioBrick parts using the new BioBrick vectors

We assembled BioBrick composite parts as described (Figure 6-3). We performed all restriction digests by mixing 0.5-1 μg DNA, 1X NEBuffer 2, 100 μg/ml Bovine Serum Albumin, and 1 μL each needed restriction enzyme in a 50 μL total volume. Restriction digest reactions were incubated for at least 2 hours at 37°C and then heat-inactivated for 20 minutes at 80°C. We then dephosphorylated the destination vector into which the parts were assembled. (When assembling BioBrick vectors, we dephosphorylated the composite origin and resistance marker to prevent circularization of this DNA fragment.) We performed dephosphorylation reactions by adding 5 units Antarctic Phosphatase and 1X Antarctic Phosphatase Reaction Buffer in a total volume of 60 μL to the heat-inactivated restriction digest reaction. We incubated dephosphorylation reactions for 1 hour at 37°C and inactivated the phosphatase by heating to 65°C for 5 minutes. We purified all reactions using a MinElute PCR Purification kit according to the manufacturer's directions (QIAGEN). We performed all ligation steps by mixing 2-4 μL of each purified, linearized DNA, 1X T4 DNA Ligase Reaction Buffer, and 200 units T4 DNA Ligase in a 10μL total volume. We incubated the ligation reactions for 20 minutes at room temperature. We transformed

all assembled BioBrick parts into *E. coli* strain TOP10 via chemical transformation [273, 274, 275]. (We transformed the assembled BioBrick vectors into *E. coli* strain DB3.1 via electroporation [258, 271, 272].) Transformed cells were plated on LB agar plates supplemented with 100 μg/mL ampicillin, 50 μg/mL kanamycin, 35 μg/mL chloramphenicol, or 15 μg/mL tetracycline as appropriate. We identified clones with correct construction of BioBrick parts by growth on the plates supplemented with the correct antibiotic, lack of growth on plates supplemented with other antibiotics, length verification by colony PCR (section 6.5.6), and DNA sequencing by the MIT Biopolymers Laboratory.

6.5.6 Verification of correct BioBrick part assembly via colony PCR

To demonstrate the correct assembly of BioBrick parts using the new BioBrick vectors, we performed a colony PCR using primers that anneal to the verification primer binding sites. We picked one colony and diluted it into 100 μL water. Then we mixed 9 μL PCR SuperMix High Fidelity, 6.25 pmoles VF2 primer (`5'-TGC CAC CTG ACG TCT AAG AA-3'`), 6.25 pmoles VR primer (`5'-ATT ACC GCC TTT GAG TGA GC-3'`), and 1 μL colony suspension. The PCR conditions were as described previously but using an annealing temperature of 62°C and an elongation time of 3.5 minutes. We diluted the reactions four-fold with water and then performed an agarose gel electrophoresis of 20 μL of each diluted reaction using a 0.8% E-Gel®. We also electrophoresed 1 μg of 2-log DNA ladder (New England Biolabs, Inc., Ipswich, MA) to verify the length of each PCR product. The gel was imaged with 302 nm transilluminating ultraviolet light using an ethidium bromide emission filter and an exposure time of 614 milliseconds.

Materials for all PCR and agarose gel electrophoresis steps in this work were purchased from the Invitrogen Corporation in Carlsbad, CA unless otherwise specified. Reagents for all restriction digest, dephophorylation, and ligation reactions were purchased from New England Biolabs, Inc., Ipswich, MA. All PCR and temperature-

controlled incubation steps were done in a DNA Engine Peltier Thermal Cycler (PTC-200) or DNA Engine OPTICON^{TM}from MJ Research, Inc. (now Bio-Rad Laboratories, Inc., Hercules, CA).

6.6 Conclusions

The goal of synthetic biology is to make the process of design and construction of many-component, engineered biological systems easier. In support of this goal, a technical standard for the physical composition of biological parts was developed [20]. Here, we extended the same principles of part reusability and standardization of physical composition to the vectors that are used to assemble and propagate BioBrick parts. Using the process described here, new BioBrick vectors can be produced from existing and newly designed BioBrick parts. As a result, myriad new vectors with diverse functions can be built readily to support the engineering of many-component systems. We invite the community to build on this work in several ways. First, we invite the community to use the process described here to construct more BioBrick vectors and share them via the Registry of Standard Biological Parts. Second, we invite the community to expand the collection of parts for making BioBrick vectors. For example, shuttle vector parts, compatible replication origins, and additional antibiotic resistance markers would all be useful contributions to the Registry. Third, we invite the community to further characterize and improve the BioBrick parts that make up BioBrick vectors. For example, important parameters to measure include plasmid copy number, and transcriptional and translational read-through into and out of the BioBrick cloning site.

Chapter 7

Future work

Synthetic biology seeks to make the process of design, construction, and testing of multi-component, engineered biological systems easier. My thesis focused on the application of fundamental engineering principles, such as abstraction, standards for physical composition, standards for functional composition, and design scalability to the engineering of biology. I drew upon a combination of mathematical modeling, experimental implementation, and measurement tools. Yet this thesis merely takes the first few steps towards the larger goal of making biological engineering cheap, fast, and predictable. Here, I list several areas where additional work is needed.

7.1 Relevance of lessons from electrical engineering to biological engineering

A central theme in my thesis is the application of principles from electrical engineering disciplines to biological engineering. In particular, I make extensive use of ideas regarding how to assess digital device performance and how to design digital devices so that they work in combination. Ideas from electrical engineering are relevant to biological engineering, because electrical engineering makes effective use of abstraction. For example, electrical engineers evaluate digital device performance using metrics, such as noise margin, propagation delay, and power consumption, that are indepen-

dent of the underlying implementation. In fact, digital device performance metrics are often used to compare one device implementation to another. Hence, approaches for evaluating digital device performance in electrical engineering readily apply to transcription-based implementations. It should be noted, however, that biological implementations of combinational digital logic do face additional issues that are mostly unique to biology. For example, by relying on molecular specificity to isolate one device's operation from another, crosstalk between transcription-based devices is a key issue. Implementations in electrical engineering largely bypass the crosstalk issue through spatial separation of signals. As a second example, genetically-encoded devices and systems self-replicate in the presence of error. To date, mature engineering disciplines have not needed a rigorous design framework for self-replicating systems that are vulnerable to error. Developing such a framework represents a key challenge in biological engineering that likely cannot readily draw upon lessons from other fields.

7.2 Models of device operation do not capture *in vivo* device behavior

There is an significant disconnect between the design principles obtained from simple gene expression models of transcriptional inverters (Chapter 3) and the issues faced during experimental implementation of transcriptional inverters (Chaper 5). For instance, a general observation from the presented model of a transcriptional inverter is that low input signals are sufficient to turn off the device output. The observed behavior is due in large part to the typically high repressor-operator DNA affinities measured for most naturally occurring bacterial repressors, such as lacI, tetR, and λ cI. Even if the model explicitly accounts for nonspecific binding, it still predicts that relatively low input signals will turn off device output. Such a prediction is at odds with the experimental characterization of the transcriptional inverter based on a synthetic transcription factor (Chapter 5). Even at maximum input, the inverter failed

to turn off the device output completely. The disconnect between the model and the experiments arises because our models generally assume that repressor binding is sufficient to sterically hinder transcriptional initiation. In practice, complete repression of transcriptional initiation is hard to achieve. Thus, either the repressor occupancy time on the promoter DNA is far lower than simple models predict or transcriptional initiation can still occur even in the presence of the repressor. Regardless, there is an apparent lack of understanding of transcriptional repression at the molecular level. Good models for understanding the interplay between repressor binding and RNAP polymerase binding to the promoter do not exist. The postulated mechanisms by which natural repressors achieve high levels of repression, such as local concentration effects and DNA looping, are not so well-understood that we can straightforwardly map the ideas to design of transcriptional inverters.

A second area where our models completely fail to capture device operation at the relevant level of detail is nonspecific DNA binding. Although the presented model takes into account the impact of nonspecific DNA binding on free repressor levels, it clearly has no means of capturing second-order effects of nonspecific DNA binding, such as reduced chassis growth rate. It would be useful to have even a heuristic understanding for how much nonspecific DNA binding a chassis can tolerate without adversely effecting its ability to support system operation.

One source of confusion in mapping model predictions to experimental characterizations is how transfer curves are plotted. I plotted transfer curves obtained from the model on linear axes. However, for expediency, most experimental characterizations of transcriptional devices are measured using indirect reporters of gene expression, such as fluorescence or enzymatic activity. Such measurements usually yield data in arbitrary units rather than absolute numbers of molecules and are often plotted on logarithmic axes rather than linear axes. Logarithmic plots tend to highlight low gain regions for low inputs, whereas linear plots do not. Hence, depending on the shape of the transfer curve, the device behavior may appear to be more "switch-like" on a logarithmic plot than on a linear plot. Absolute measurements of device transfer curves in PoPS are needed for direct comparison of models to experimental data.

In a similar vein, proposal of suitable device family specifications is severely hampered by the lack of data on device signals and variability in device signals. The crude estimates of PoPS signal values and their distributions based on published data that I have presented are not sufficient to prescribe device signal thresholds in good faith. More extensive device characterization data is needed to inform models for device operation and performance. In particular, measurement of not only transcriptional signals but also device material, energy, and machinery demands at the single cell level is needed. Such data is key to being able to develop suitable device family specifications to which engineers can build their devices.

7.3 Engineering combinational, transcription-based logic

The long-term aim of much of the work in this thesis is to develop libraries of transcription-based logic devices. Ideally, devices in the library would meet at least four criteria. First, devices must be well-characterized. For example, the transfer curve, propagation delay, demand, error rate for each transcription-based devices should be measured. Devices should be sufficiently well-characterized that end-users can simply consult a device datasheet to determine whether the device will work in their system or not [22]. Second, families of devices in the collection should work in combination: a string of devices should be able to faithfully propagate a signal. Thus, devices must adhere to proscribed device family specifications. Third, the devices should be orthogonal. Two devices which are not connected should not interfere with each other's operations. In practice, orthogonality means that proteins from two different devices should not dimerize and that proteins should not bind to noncognate operators. Fourth, the devices should implement any logical operations not just logical NOT. By building logic devices that can implement the universal logic operations of NAND or NOR, any logical operation is realizable.

As demonstrated in this thesis, achieving combinational, transcription-based dig-

ital logic in *E. coli* is a challenging goal. My thesis work has largely relied on rational design approaches to build a prototype transcriptional inverter. Although the transcriptional inverter presented here may be sufficient to implement combinational digital logic, many classes of applications will require devices with improved performance. Some applications will require lower error rates, other lower device demand so that more devices can operate in parallel in the cell, and some will require both. Going forward, a hybrid approach of rational design and combinatorial libraries may prove fruitful. Rational design involves diagnostic experiments to identify likely failure modes combined with intentional device redesign to address the observed failure. For example, I saw a significant improvement in inverter function through a rational redesign of the repressor. By combinatorial libraries, I am not referring to the widely used practice of generating sequence diversity through mutagenic PCR or random oligos. Instead, I mean construction of libraries of different part combinations. For example, libraries of promoters with different natural operator sites in different combinations have been constructed and screened for function [219]. Such an approach may also prove useful in designing promoters repressible by synthetic transcription factors for use in transcriptional inverters. Key parameters to vary in the promoter library are operator number, operator position, and promoter strength.

A more ambitious approach to making devices with improved performance is to construct and screen a library of transcriptional inverters based on different zinc finger domains and cognate operators. Many zinc fingers and cognate binding sites have been characterized based on Zif268 [202, 104, 276, 192]. Similarly, there are several leucine zippers available [103]. Therefore, design of tens of different inverters based on experimentally-validated zinc fingers and leucine zippers is readily realizable. In addition to yielding novel functional, transcriptional inverters, such experiments could help to establish definitively whether zinc fingers and leucine zippers proffer a tractable solution to the challenge of building libraries of transcription-based logic devices from synthetic transcription factors.

A natural question is why a library-based approach was not adopted in this thesis. A library-based approach is greatly facilitated by access to three intersecting tech-

nologies that are only now becoming readily available: ready access to cheap DNA synthesis, high-throughput, automated assembly of genetic parts, and collections of functional, standard biological parts. DNA synthesis allows ready construction of synthetic transcription factors and cognate promoters from sequence information alone. Gene synthesis costs have dropped from around $10 per base in 2002 to $0.70 per base today [277]. High-throughput, automated assembly of genetic parts allows the engineer to easily construct combinatorial libraries of zinc finger domains, leucine zipper domains and cognate promoters. Although few companies offer commercial assembly of genetic parts, the BioBrick idempotent physical composition standard and the improved assembly procedures developed in this thesis renders parallelized assembly of genetic parts far more tractable. Finally, construction of libraries of transcriptional inverters depends on the availability of libraries of ribosome binding sites and transcriptional terminators for use in the inverters, since reuse of the same parts over and over can lead to issues of recombination and loss-of-function [22]. The Registry of Standard Biological Parts has grown from about 10 parts in 2002 to over 2,000 today. Yet, the number of well-characterized parts that implement basic biological functions is still quite small.

The convergence of the technologies of synthesis, assembly, and an available collection of standard biological parts also paves the way for other, related work towards the larger goal of combinational, transcription-based logic. For example, as device family specifications are further flushed out and more transcriptional devices are built, devices will need to be tuned to meet a given device specification. Again, automated assembly of genetic parts will be crucial. Tuning device behavior to meet a particular device specification will likely require screening of device variants with different ribosome binding sites and promoter variants. As a second example, a more diverse set of synthetic transcription factors and cognate promoters could be build to implement a wider array of logic operations. Devices that implement logical NOR can be built directly from the transcriptional inverters discussed here by including multiple copies of the device input: the ribosome binding site, repressor coding sequence, and terminator. Such a layout means that if any one of the inputs receives a high signal, then

repressor is expressed and will turn off the output signal. In fact, devices that implement logical NAND are readily built from heterodimerizing synthetic transcription factors. Thus, successful engineering of transcription factors for transcription-based logic have the potential to make accessible arbitrary *in vivo* information processing using digital devices.

7.4 Extending standardization and design scalability to other classes of devices

More generally, I expect that many of the principles explored in this thesis with respect to abstraction, standards for physical composition, standards for functional composition, device performance measures, and design scalability will also be relevant to other domains of synthetic biology. Just as cells rely on regulation at the transcriptional, translational, degradation, and post-translational level, engineered information systems will rely on digital logic devices that operate at each level. Advancements in genetically-encoded sensors and actuators, such as those for catabolism and anabolism of chemicals, will render information processing devices even more valuable in biological engineering.

Appendix A

Materials and methods for Chapter 2

A.1 Design

All genetic parts used in this work were designed and assembled according to the BioBrick assembly standard unless otherwise specified [20]. We designed all BioBrick standard biological parts using Vector NTI® Suite 7 for Mac OS X by Invitrogen Life Science Software in Carlsbad, CA and the Registry of Standard Biological Parts (`http://partsregistry.org`).

A.2 System construction and assembly

A.2.1 Part construction

We constructed the long *osmY* promoter (BBa_J45992) and all protein coding region parts by PCR. Each reaction was a mix of 15 pmoles each of forward and reverse primer and 1 ng of template with PCR Supermix High Fidelity added to a total volume of 50 μL. The PCR conditions were an initial denaturation step of 95°C for 3 minutes followed by 40 cycles of 94°C for 30 seconds, 55°C for 30 seconds, and 68°C for 2.25 minutes. Finally, the reactions were incubated at 72°C for 10 min-

utes. We performed gel electrophoresis on 5 μL each PCR product using 1% agarose gels supplemented with either 0.5 μg/mL ethidium bromide or 1X SYBR Safe DNA gel stain in 1X TAE running buffer. To verify the length of the PCR products, we also electrophoresed 1 μg of 2-log DNA ladder. If the PCR was successful, we purified the PCR product using the QIAQuick PCR Purification Kit according to the manufacturer's directions (QIAGEN, Germany). We constructed the short *osmY* promoter (BBa_J45993) via primer annealing. We dissolved the primers in 50 mM Tris buffer to yield a concentration of 500-1000 ng/μL. We mixed 8 μL each primer and 1 mM NaCl in total volume of 40 μL. We incubated the reaction mix in a beaker of boiling water and allowed the beaker and reaction to cool to room temperature. Template and primer sequences used for each part are listed (Table A.1). All primers were designed with the aid of IDT SciTools OligoAnalzer 3.1 (`http://scitools.idtdna.com/analyzer/Applications/OligoAnalyzer/`). We cloned each BioBrick part into a BioBrick vector using the same restriction digest, ligation, and transformation procedures described in Assembly of BioBrick standard biological parts below.

A.2.2 Ensuring parts conform to the BioBrick standard for physical composition

To ensure our parts complied with the BioBrick assembly standard, we used site-directed mutagenesis to remove a BioBrick restriction site from the *ATF1* coding sequence (BBa_J45014) with forward mutagenesis primer (`5'-GAA GCA AAT ATT AGA AGA GTT CAA AAA TAG TAA GGG-3'`) and reverse mutagenesis primer (`5'-CCC TTA CTA TTT TTG AAC TCT TCT AAT ATT TGC TTC-3'`). We performed the site-directed mutagenesis reactions using the Quikchange Site-Directed Mutagenesis Kit according to the manufacturer's directions with three modifications (Stratagene, La Jolla, CA). First, we phosphorylated the primers by preparing a reaction mix of 10 units T4 Polynucleotide Kinase, 1 μL T4 Ligase Buffer, and 8 μL primer. We incubated the mix at 37°C for 30 minutes and then heat-inactivated it at 65°C for 20 minutes.

Part number	*Template*	*Forward primer*	*Reverse primer*
BBa_J45004	pET-28a-*BSMT1* [51]	5'-GTT TCT TCG AAT TCG CGG CCG CTT CTA GAT GGA AGT TGT TGA AGT TC-3'	5'-GTT TCT TCC TGC AGC GGC CGC TAC TAG TAT TAT TAA TTT ATT TTG GTC AAG GAG-3'
BBa_J45014	*S. cerevisiae* strain ACLY387 genome	5'-GTT TCT TCG AAT TCG CGG CCG CTT CTA GAT GAA TGA AAT CGA TGA GAA AAA TC-3'	5'-GTT TCT TCC TGC AGC GGC CGC TAC TAG TAT TAT TAA GGG CCT AAA AGG AGA GCT TTG-3'
BBa_J45992	*E. coli* strain MG1655 genome	5'-GTT TCT TCG AAT TCG CGG CCG CTT CTA GCT GGC ACA GGA ACG TTA TC-3'	5'-GTT TCT TCC TGC AGC GGC CGC TAC TAG TAT TGT TAA ATA TAG ATC ACA ATT TTG AAA CCG-3'
BBa_J45993	none	5'-CTA GAG GCT TAT GTT TTC GCT GAT ATC CCG AGC GGT TTC AAA ATT GTG ATC TAT ATT TAA CAA ATA CTA GTA GCG GCC GCT GCA-3'	5'-GCG GCC GCT ACT AGT ATT TGT TAA ATA TAG ATC ACA ATT TTG AAA CCG CTC GGG ATA TCA GCG AAA ACA TAA GCC T-3'

Table A.1: Primers for part construction

Second, the DpnI digestion step was extended to 2-3 hours. Third, we transformed the mutated DNA into *E. coli* strain TOP10. We designed the primers for site-directed mutagenesis with the aid of PrimerX (`http://www.bioinformatics.org/primerx`).

A.2.3 Assembly of BioBrick standard biological parts

We assembled all BioBrick standard biological parts by an iterative process of restriction digest, ligation, transformation, and plasmid DNA purification using three antibiotic assembly [266]. We performed all restriction digests by mixing 0.5-1 μg DNA, 1X NEBuffer 2, 100 μg/mL Bovine Serum Albumin, and each appropriate restriction enzyme (10 units EcoRI, 10 units XbaI, 5 units SpeI, and/or 10 units PstI) in a 50 μL total volume. We incubated all restriction digest reactions for at least 2 hours at 37°C and then heat-inactivated them for 20 minutes at 80°C. We performed all ligation reactions by mixing 1-5 μL of each linearized DNA, 1X T4 DNA Ligase Buffer, and 200 units T4 DNA Ligase in a 10 μL total volume. All ligations were three-way ligations involving (1) BioBrick vector cut with EcoRI and PstI, (2) the upstream or 5' part cut with EcoRI and SpeI, and (3) the downstream or 3' part cut with XbaI and PstI. All ligation reactions were incubated for 20 minutes at room temperature. We transformed 2 μL each ligation reaction into *E. coli* strain TOP10 via chemical transformation [273, 274, 278, 40, 275]. Transformed cells were plated on LB agar plates supplemented with 50 μg/mL ampicillin, 34 μg/mL chloramphenicol, 20 μg/mL kanamycin, and/or 5 μg/mL tetracycline as appropriate.

A.2.4 Verifying successful assemblies

We identified colonies with correctly assembled BioBrick composite parts by growth on the plates supplemented with the correct antibiotic, length verification by colony PCR (optional), and DNA sequencing by the MIT Biopolymers Laboratory. For colony PCR, we picked several colonies and diluted each colony into 20 μL water. Then we mixed 9 μL PCR Supermix High Fidelity, 10 picomoles each forward and reverse primer, and 0.5 μL colony suspension. The PCR conditions were an initial

denaturation step of 98°C for 5 minutes followed by 40 cycles of 98°C for 10 seconds, 58°C for 30 seconds, and 72°C for 1 minute per kilobase of expected product length. Finally, the reactions were incubated at 72°C for 10 minutes. The primers used for both colony PCR and DNA sequencing were VF2 (5'-`TGC CAC CTG ACG TCT AAG AA`-3') and VR (5'-`ATT ACC GCC TTT GAG TGA GC`-3'), respectively. We used mutagenesis primers as additional, internal primers for sequencing longer parts.

A.2.5 Debugging failed assemblies

In the event that the first attempt of an assembly was unsuccessful, we tried one or more problem-solving strategies. First, gel electrophoresis was performed as described above to verify digestion of the vectors and parts to be assembled. If we found that digestion was incomplete, we repeated the digest and extended the incubation at 37°C to several hours or overnight. Second, we purified the digested parts using the QIAQuick PCR Purification Kit as described above. Third, we performed gel extractions of the digested parts using the QIAQuick Gel Extraction Kit according to the manufacturer's directions (QIAGEN). Fourth, we tried two-way BioBrick part assemblies as originally described [20].

The PCR Supermix High Fidelity, SYBR Safe DNA gel stain, and all primers were purchased from the Invitrogen Corporation in Carlsbad, CA. All PCR steps were done in a DNA Engine Peltier Thermal Cycler (PTC-200) from MJ Research, Inc. (now Bio-Rad Laboratories, Inc., Hercules, CA). All reagents for restriction enzyme digestions, phophorylations, and ligations as well as 2-log ladder were purchased from New England Biolabs, Inc., Ipswich, MA.

A.3 Gas chromatography analysis of odorant production

To test function of the constitutive wintergreen odorant generator (BBa_J45120), constitutive banana odorant generator (BBa_J45200), and odor-free chassis (BBa_J45999),

we grew overnight cultures at 37°C from single colonies in LB Lennox medium. Cultures were supplemented with 50 μg/mL ampicillin, 34 μg/mL chloramphenicol, 20 μg/mL kanamycin, 5 μg/mL tetracycline, 2 mM salicylic acid (precursor to wintergreen odorant), and/or 5 mM isoamyl alcohol (precursor to banana odorant) as appropriate. We prepared samples suitable for gas chromatography by extracting 20 mL culture into 2 mL heptane. In each case, we analyzed 200 μL of sample.

We analyzed samples for methyl salicylate or indole using an Agilent 6890N gas chromatograph operated in pulsed spitless mode and equipped with a Restek Rtx-1 (30 m length, 0.25 mm inner diameter, and 1.00 μm film thickness) crossbound, 100% dimethyl polysiloxane column and coupled to an Agilent 5973 mass-selective detector. The carrier gas was helium held at a constant flow rate of 1.0 mL/minute. Samples were injected at 250°C. The GC oven was programmed with an initial 5 minute temperature hold at 100°C, followed by a temperature ramp of 10°C/minute to 180°C, followed by a temperature hold of 2 minutes, followed by a temperature ramp of 70°C/minute, and a final hold of 10 minutes. The Agilent 5973 had an analyzer temperature of 150°C and an ion source temperature of 230°C and was operated at 70 eV in full scan mode with a mass range m/z 50-500. Initial solvent delay was 5 minutes. For reference, we analyzed samples containing 100 ppm pure methyl salicylate and 100 ppm pure indole in the same manner.

We analyzed samples for isoamyl acetate using a HP 6890 gas chromatograph fitted with a PTV injector operated in splitless mode and equipped with a J&W Scientific DB-17 (30 m length, 0.32 mm inner diameter, and 0.25 μm film thickness) fused silica capillary column and coupled to an Agilent 5973 mass-selective detector. The carrier gas was helium held at a constant pressure of 22.7 psi. The samples were injected at 50°C and the GC oven was programmed with an initial 5 minute temperature hold, followed by a temperature ramp of 20°C/minute to 200°C, followed by a final hold of 5 minutes. The Agilent 5973 was operated at 70 eV in full scan mode with a mass range m/z 50-750. Initial solvent delay was 3 minutes. For reference, we also analyzed a sample containing 100 ppm pure isoamyl acetate in the same manner. All isoamyl acetate samples also contained 100 ppm octyl acetate as an internal standard.

A.4 Testing function of transcriptional control devices

To measure the function of the transcriptional control devices, we assembled the constitutive (BBa_R0040), exponential phase (BBa_J45994), and stationary phase (BBa_J45992) control devices with the GFP generator (BBa_E0840). To control for cellular autofluorescence, we also characterized a part which does not synthesize GFP (BBa_B0015). For each of the four constructs, we grew three cultures from single colonies overnight at 37°C. After 20 hours of growth, we diluted each culture 250-fold into 25 mL culture volume in 250 mL flasks. We incubated diluted cultures at 220 rotations per minute (RPM) in an Innova 2300 Large-Capacity Benchtop Shaker (New Brunswick Scientific Co., Inc., Edison, NJ) for 4 hours at 37°C. We then diluted the cultures again to an OD600nm of 0.0448 into a 25 mL final culture volume. The rediluted cultures were grown under the same conditions for 50 minutes. We loaded 200 μL of each culture in triplicate into a 96-well plate. The plate was incubated in a Wallac Victor3 multi-well fluorimeter (Perkin Elmer, Waltham, MA) at 37°C and assayed with an automatically repeating protocol of absorbance measurements (600 nm absorbance filter, 0.1 seconds counting time through 5 mm of fluid) and fluorescence measurements (488 nm excitation filter, 525 nm emission filter, 0.1 sec, CW lamp energy 12901 units). Between each measurement, we included two identical steps of a delay of 250 seconds followed by shaking (1 mm, linear, normal speed, 15 seconds). The time between repeated measurements was 10 minutes and 37 seconds.

A.5 Analysis of growth phase-dependent transcriptional control devices

We analyzed the resulting data to compute the rate of synthesis of GFP per cell as a function of culture density. We subtracted the appropriate backgrounds from the raw data measured by the Wallac Victor3 multi-well fluorimeter. The absorbance

of wells containing LB medium, A_{media}, was subtracted from the sample absorbance data, A_{raw}. The resulting data, $A_{corrected}$, was assumed to be directly proportional to the number of cells in the well.

$$A_{corrected} = A_{raw} - A_{media} \tag{A.1}$$

Similarly, the fluorescence data for the GFP-free cells containing BBa_B0015, G_{cells}, was subtracted from the sample fluorescence data, G_{raw}, and the resulting data $G_{corrected}$ was assumed proportional to the total number of GFP molecules in the well.

$$G_{corrected} = G_{raw} - G_{cells} \tag{A.2}$$

A calibration curve relating absorbance measured by the multi-well fluorimeter to OD600 was produced (data not shown). The calibration curve was used to convert the corrected absorbance data to OD600. The calibration equation used is shown in Equation (A.3).

$$OD = 3.11 * A_{corrected} - 0.016 \tag{A.3}$$

Synthesis rates of GFP per cell in relative units, S_{cell}, were calculated by first assuming the total GFP synthesis rate, S_{total}, to be equal to the time differential of GFP. S_{cell} was then calculated via Equation (A.5). Since we measured the total amount of GFP in the well and since we assume that GFP is not degraded, we calculate the total synthesis rate of GFP and, from that, the per cell synthesis rate of GFP without considering dilution due to cell growth.

$$S_{total} = \frac{dGFP}{dt} \tag{A.4}$$

$$S_{cell} = \frac{S_{total}}{OD} \tag{A.5}$$

We averaged the data for each of the three replicates and three independent cul-

tures. Error bars were calculated as the 95% confidence interval in the mean of the three independent cultures.

A.6 Testing function of growth phase-dependent odorant generators

To test the function of the growth phase-dependent odorant generators, we measured the amount of odorant produced by the constitutive wintergreen odorant generator (BBa_J45120), exponential phase wintergreen odorant generator (BBa_J45181), constitutive banana odorant generator (BBa_J45200), and stationary phase banana odorant generator (BBa_J45250) as a function of the culture density. For each day's experiment, we grew 5 mL cultures of the constitutive and growth phase-dependent device overnight at 37°C in LB Lennox medium supplemented with 50 μg/mL ampicillin and 5 μg/mL tetracycline. In two 2-L flasks, we diluted the overnight cultures 10,000-fold in LB medium supplemented with antibiotic to make master cultures. We incubated the master cultures at 37°C for 20 minutes at 220 RPM in an Innova 2300 Large-Capacity Benchtop Shaker (New Brunswick Scientific Co., Inc., Edison, NJ). Then, we transferred the cultures to be shaken at 110 RPM in a Lab-Line 4645 shaker from Lab-Line Instrument in Melrose Park, IL (time zero). Periodically, we aliquotted 25 mL of master culture into a new 125-mL flask and supplemented the culture with 2 mM pure salicylic acid or 5 mM pure isoamyl alcohol as appropriate. At this time, we measured the OD600nm of the culture aliquot in a 1 cm VWR polystyrene semi-micro cuvette using the CO 8000 Biowave Cell Density Meter from WPA (now Biochrom Ltd, United Kingdom). We returned the culture aliquot to the same growth conditions as the master culture (110 RPM at 37°C). After 3 additional hours of growth, we measured the OD600nm of the culture aliquot again. We prepared and analyzed samples via gas chromatography as previously described. All methyl salicylate samples also contained 10 ppm pentachloronitrobenzene as an internal standard. To quantify banana odorant levels in culture, we generated standard

curves using varying concentrations of pure isoamyl acetate.

A.7 Quantification of isoamyl acetate production by the constitutive and stationary phase banana odorant generators

Isoamyl acetate production by the constitutive and stationary phase banana odorant generators as a function of cell density was analyzed as follows.

1. Reference samples of 5, 25, 100, 200, and 500 ppm isoamyl acetate (38 μM to 3.8 mM) were analyzed by gas chromatography during each day's experiment. Each reference sample also contained 100 ppm octyl acetate.

2. The peak height corresponding to the retention time of isoamyl acetate was divided by the peak height corresponding to the retention time of the octyl acetate internal standard.

3. A linear standard curve was fit to a plot of isoamyl alcohol concentration versus ratio of isoamyl acetate to octyl acetate peak heights.

4. The data obtained for each experimental sample was converted to molar concentration using the standard curve obtained that day.

5. The molar concentration of isoamyl acetate was corrected for compound dilution during sample preparation for gas chromatography and compound concentration during heptane extraction. We did not correct for extraction efficiency. Extraction efficiencies were measured to be $\sim$70%.

6. The OD600nm of each culture was calculated by averaging the OD600nm reading before and after the three hour growth in the presence of exogenous 5 mM isoamyl alcohol.

7. We plotted the corrected isoamyl acetate concentration (mM) versus the average OD600nm.

The data from each of the three days was combined for each device and fit empirically to the logistic equation.

$$[\text{Isoamyl acetate}] = \frac{K}{(1 + C * e^{-r * \text{OD600nm}})} \tag{A.6}$$

The values for K, C, and r were fit by minimizing the sum of the squares of the errors using solver in Microsoft® Excel® 2004 for Mac Version 11.3.7. For the constitutive banana odorant generator (BBa_J45200), we obtained $K = 1.3$, $C = 94$, and $r = 8.3$ with a least squares error of 0.31. For the stationary phase banana odorant generator (BBa_J45250), we obtained $K = 0.98$, $C = 1900$, and $r = 7.8$ with a least squares error of 0.15.

A.8 Quantification of methyl salicylate production by the constitutive and exponential phase wintergreen odorant generators

Methyl salicylate production by the constitutive and exponential phase wintergreen odorant generators as a function of cell density was analyzed using a similar approach to isoamyl acetate production. However, methyl salicylate levels were not converted to absolute concentrations and were instead plotted as relative values to the 10 ppm pentachloronitrobenzene internal standard (Figure 2-7).

The data from both days was combined for each device and fit by linear regression. For the constitutive wintergreen odor generator, we obtained a slope of 5.7 and an intercept of -1.1 ($R^2 = 0.88$). For the exponential phase wintergreen odor generator, we obtained a slope of 4.8 and an intercept of -0.1 ($R^2 = 0.86$). The fitted lines simply serve to emphasize the fact that there is no difference in methyl salicylate levels produced by the constitutive and exponential phase wintergreen odorant generators.

Appendix B

Materials and methods for Chapter 5

B.1 Construction of transcriptional inverters

B.1.1 Fabrication of BioBrick basic parts

A part based on Zif23-GCN4 (BBa_C2001) and cognate promoter (BBa_R2000) were originally designed as a part of the 2004 MIT IAP class in synthetic biology and contracted for DNA synthesis to Blue Heron Biotechnology, Inc. The initial design for the synthetic transcription factor BBa_C2002 used in this work was constructed by PCR using pSB2K3-C2001 as a template and amplification primers BioBrick-f (`5'-GTT TCT TCG AAT TCG CGG CCG CTT CTA G-3'`) and C2002-r (`5'-GTT TCT TCC TGC AGC GGC CGC TAC TAG TAT TAT TAG TGA TGG TGA TGG TGA TGA CGT TCA CCA ACC AGT TTT TTC-3'`). The reaction was a mix of 18 μL PCR SuperMix High Fidelity, 12.5 pmoles each of forward and reverse primer, and 1 ng template DNA in a 20 μL total volume. The PCR conditions were an initial denaturation step of 95°C for 15 mins followed by 40 cycles of 94°C for 30 seconds, 52°C for 30 seconds, and 68°C for 30 seconds. Finally, the reactions were incubated at 68°C for 20 minutes. The PCR product was TOPO TA cloned by mixing 1 μL PCR product with 0.5 μL pCR4-TOPO vector and incubating at room temperature for 5 minutes before plac-

ing on ice. The cloned PCR product was chemically transformed into *E. coli* strain Mach^{TM}(Invitrogen Corporation, Carlsbad, CA). Correct clones were identified as described in section B.1.2.

The redesigned synthetic repressor BBa_C2006 was constructed by PCR. The maltose binding protein domain was constructed by PCR using pMAL-c2x (New England Biolabs, Inc. in Ipswich, MA) as a template and amplification primers C2006-f (`5'-GTT TCT TCG AAT TCG CGG CCG CTT CTA GAT GAA AAT CGA AGA AGG TAA AC-3'`) and C2006link-r (`5'-ACG GCA CTG GAA CGG TTT CAT CCT TCC CTC GAT CCC G-3'`). The Zif23-GCN4 domain was constructed by PCR using pSB1AT3-B0030.C2005 as a template and amplification primers C2006link-f (`5'-CGG GAT CGA GGG AAG GAT GAA ACC GTT CCA GTG CCG T-3'`) and BioBrick-r (`5'-GTT TCT TCC TGC AGC GGC CGC TAC TAG TA-3'`). Both reactions were a mix of 18 μL PCR SuperMix High Fidelity, 12.5 pmoles each of forward and reverse primer, and 100-200 pg template DNA in a 20 μL total volume. The PCR conditions were as before except that an annealing temperature of 56°C and an elongation time of either 90 or 30 seconds for the maltose binding domain or Zif23-GCN4 domain, respectively, was used. 1 μL of each reaction was diluted in a total volume of 40 μL. The composite synthetic transcription factor was constructed by PCR by mixing 36 μL PCR SuperMix High Fidelity, 25 pmoles each of primers C2006-f and C2005-r, and 2 μL diluted PCR product as template. The PCR conditions were the same as before except that the elongation time was extended to 130 seconds. I purified both reactions using a PCR Purification kit according to the manufacturer's directions (QIAGEN, Germany). I performed a restriction digest of the PCR product by mixing the DNA with 1X NEBuffer 3, 100 μg/ml Bovine Serum Albumin, and 0.5 μL each XbaI and PstI restriction enzyme in a 20 μL total volume. The resulting digest was electrophoresed on an agarose gel and purified via a QIAEX II Gel Extraction Kit (QIAGEN). The linear DNA fragment was cloned as described below into BioBrick vector pSB4A3-P1010.

The modified cognate promoter BBa_R2201 was constructed by primer annealing and extension of primers R2201-f (`5'-GTT TCT TCT CTA GAG AGT TTA TTC TTG ACC CAC GCG CGT GGG AAT GTT ATA ATA C-3'`) and R2201-r (`5'-GTT TCT TCC TGC AGC`

GGC CGC TAC TAG TAC TGA CGT ATT ATA ACA TTC CCA CGC G-3'). The reaction was a mix of 9 μL PCR SuperMix High Fidelity and 12.5 pmoles each forward and reverse primer. The reaction conditions were an initial denaturation step of 95°C for 5 mins followed by 3 cycles of 94°C for 30 seconds, 52°C for 30 seconds, and 68°C for 30 seconds. Finally, the reactions were incubated at 68°C for 5 minutes. The product was again cloned by TOPO TA cloning. Correct clones were identified as described in section B.1.2.

B.1.2 Assembly of BioBrick composite parts

I assembled BioBrick composite parts as described (Figure 6-3). I performed all restriction digests by mixing 0.5-1 μg DNA, 1X NEBuffer 2, 100 μg/ml Bovine Serum Albumin, and 1 μL each needed restriction enzyme in a 50 μL total volume. Restriction digest reactions were incubated for at least 2 hours at 37°C and then heat-inactivated for 20 minutes at 80°C. I then dephosphorylated the destination vector into which the parts were assembled. I performed dephosphorylation reactions by adding 5 units Antarctic Phosphatase and 1X Antarctic Phosphatase Reaction Buffer in a total volume of 60 μL to the heat-inactivated restriction digest reaction. I incubated dephosphorylation reactions for 1 hour at 37°C and inactivated the phosphatase by heating to 65°C for 5 minutes. I purified all reactions using a MinElute PCR Purification kit according to the manufacturer's directions (QIAGEN). I performed all ligation steps by mixing 2-4 μL of each purified, linearized DNA, 1X T4 DNA Ligase Reaction Buffer, and 200 units T4 DNA Ligase in a 10μL total volume. I incubated the ligation reactions for 20 minutes at room temperature. I transformed all assembled BioBrick parts into *E. coli* strain TOP10 as described previously. I identified clones with correct construction of BioBrick parts by growth on the plates supplemented with the correct antibiotic, lack of growth on plates supplemented with other antibiotics, length verification by colony PCR, and DNA sequencing by the MIT Biopolymers Laboratory.

B.1.3 Verification of correct BioBrick part assembly via colony PCR

To verify correct assembly of BioBrick parts, I performed a colony PCR using primers that anneal to the verification primer binding sites. I picked one colony and diluted it into 20 μL water. Then I mixed 9 μL PCR SuperMix High Fidelity, 6.25 pmoles VF2 primer (`5'-TGC CAC CTG ACG TCT AAG AA-3'`), 6.25 pmoles VR primer (`5'-ATT ACC GCC TTT GAG TGA GC-3'`), and 1 μL colony suspension. The PCR conditions were as described previously but using an annealing temperature of 56°C and an elongation time of 1 minuter per kb of expected product length. PCR product size was verified by electrophoresis in a 1% or 1.5% agarose gel in 1X TAE running buffer. I also electrophoresed 1 μg of 2-log DNA ladder (New England Biolabs, Inc., Ipswich, MA) to verify the length of each PCR product. The gel was imaged with 302 nm transilluminating ultraviolet light using an ethidium bromide emission filter.

Materials for all PCR and TOPO TA cloning were purchased from the Invitrogen Corporation in Carlsbad, CA. Reagents for all restriction digest, dephophorylation, and ligation reactions were purchased from New England Biolabs, Inc., Ipswich, MA. All PCR and temperature-controlled incubation steps were done in a DNA Engine Peltier Thermal Cycler (PTC-200) or DNA Engine OPTICON^{TM}from MJ Research, Inc. (now Bio-Rad Laboratories, Inc., Hercules, CA).

B.1.4 Construction of transcriptional inverter mutants

I mutated the cysteine residue at position 396 to serine with site-directed mutagenesis using primers C2009-Mf (`5'-GTT CCA GTG CCG TAT CAG CAT GCG TAA CTT CTC-3'`) and C2009-Mr (`5'-GAG AAG TTA CGC ATG CTG ATA CGG CAC TGG AAC-3'`) and pSB4K5-I8500 as a template. The reaction was a mix of 1X PfuTurbo reaction buffer, 0.5 mM each dNTP, 50 ng template, 25 pmoles each forward and reverse primer, and 2.5 units PfuTurbo polymerase (Stratagene, La Jolla, CA) in a 50 μL total volume. The reaction conditions were an initial denaturation step of 95°C for 5 mins followed by 24 cycles of 95°C for 30 seconds, 59°C for 1 minute, and 68°C for 8

minutes. Finally, the reactions were incubated at 68°C for 20 minutes. The reaction was cooled to 37°C and 20 units of restriction enzyme DpnI were added to the tube. The reaction was incubated at 37°C for 2 hours. I transformed 1 μL of the reaction mix into competent *E. coli* strain TOP10 as described previously.

The deletion of the zinc finger domain from the transcriptional inverter was done by PCR. Primers P20101-f (`5'-CGT AAA TTG CAG CAC ATG AAA CAG CTG-3'`) and P20101-r (`5'-AGT CTG CGC GTC TTT CAG GGC-3'`) were first phosphorylated by mixing 1X T4 Polynucleotide Kinase Reaction Buffer, 1 mM magnesium sulfate, 2 mM ATP, 500 pmoles primer, and 10 units T4 Polynucleotide Kinase (New England Biolabs) in a 50 μL total volume. The reaction was incubated at 37°C for 1 hour followed by heat-inactivation at 65°C for 20 minutes. Then a reaction mix identical to the site-directed mutagenesis reaction mix was prepared, except that 10 pmoles each primer was used. After digestion with DpnI, the reaction was heat-inacivated at 80°C for 20 minutes. The PCR product was then purified, ligated, and transformed as described previously.

The complete deletion of the zinc finger domain and leucine zipper domain was constructed inadvertently during a site-directed mutagenesis reaction attempting to introduce the C396S mutation.

B.2 *In vivo* GFP expression assay as a function of cell density

From single colonies, 1 mL cultures were grown overnight at 37°C in EZ Rich media supplemented with either 0.4% glycerol or 0.2% glucose (Teknova, Hollister, CA). For each experimental sample, 3 cultures from different colonies were grown. Cultures were diluted back to OD600nm of 0.01. I loaded 200 μL of each culture in triplicate into a 96-well plate. The plate was incubated in a Wallac Victor3 multi-well fluorimeter (Perkin Elmer, Waltham, MA) at 37°C and assayed with an automatically repeating protocol of absorbance measurements (600 nm absorbance filter, 0.1

seconds counting time through 5 mm of fluid) and fluorescence measurements (488 nm excitation filter, 525 nm emission filter, 0.1 sec, CW lamp energy 12901 units). Between each measurement, we included a shaking step (1 mm, linear, normal speed, 190 seconds), a delay of 150 seconds, and a second, identical shake step. The time between repeated measurements was 12 minutes and 11 seconds.

We analyzed the resulting data to compute the fluorescence per cell density a function of cell density. Raw data from the Wallac Victor3 multi-well fluorimeter was processed by first subtracting the appropriate backgrounds. The absorbance of wells containing EZ medium, A_{media}, was subtracted from the sample absorbance data, A_{raw}. The resulting data, $A_{corrected}$, was assumed to be directly proportional to the number of cells in the well.

$$A_{corrected} = A_{raw} - A_{media} \tag{B.1}$$

Similarly, the fluorescence data for wells containing EZ medium, G_{media}, was subtracted from the sample fluorescence data, G_{raw}, and the resulting data $G_{corrected}$ was assumed proportional to the total number of GFP molecules in the well.

$$G_{corrected} = G_{raw} - G_{media} \tag{B.2}$$

The fluorescence data was then scaled to the corresponding absorbance data.

$$G_{scaled} = \frac{G_{corrected}}{A_{corrected}} \tag{B.3}$$

Measurements for each experimental sample were binned by absorbance into bins of width 0.05 A600 units and averaged. Error bars are 95% confidence intervals for each bin.

B.3 Protein purification

The synthetic repressor encoded by BBa_C2002 was purified using a denaturing purification on Ni-NTA column (QIAGEN). Two LB cultures, one expressing the synthetic

repressor and one not expressing the synthetic repressor, were grown overnight from a single colony. LB cultures (5 mL) were diluted back 50-fold in 50 mL of LB medium supplemented with 10 μM $ZnCl_2$ and the appropriate antibiotic. At around OD600nm of 0.6, the cells were harvested by centrifugation at 400 $\times$ g for 15 minutes at 4°C. The supernatant was decanted and the cell pellet was stored at -80°C. The pellets were thawed on ice, transferred to a 2-mL Eppendorf tube, and resuspended in 1mL lysis buffer with a 1/4 EDTA-free protease inhibitor cocktail tablet (Roche). Lysis mixtures were agitated at room temperature for 1 hour. Lysates were then centrifuged at 10,000 $\times$ g for 30 minutes. Two Ni-NTA columns were equilibrated with 600 μL lysis buffer. The columns were centrifuged at 700 $\times$ g for 2 minutes with an open lid, and the equilibration buffer was discarded. The cleared lysates were loaded onto the columns, and columns were centrifuged at 700 $\times$ g for 5 minutes. This step was repeated until all of the cleared lysates had been loaded. Then the columns were washed twice with 600 μL wash buffer. Wash buffer was removed by centrifugation at 700 $\times$ g for 2 minutes with an open lid. To elute the protein, I then transferred the columns to a clean eppendorf tube and added 200 μL elution buffer. The column was centrifuged at 700 $\times$ g for 2 minutes with an open lid.

The lysis buffer was 8 M urea, 100 mM NaH_2PO_4, 10 mM Tris Cl, 10 mM imidazole at pH 8.0. The wash buffer and elution buffer were the same as the lysis buffer but at pH 6.3 and 4.5, respectively. After pH adjustment, the solutions were degassed for 1 hour under vacuum. Finally, tris(2-carboxyethyl) phosphine or TCEP was added to each to a 1mM final concentration, and the pH was rechecked. A stock urea solution was always freshly prepared and deionized for 1 hour with AG 501-X8(D) resin (Bio-Rad Laboratories, Hercules, CA) prior to use.

B.4 Electrophoretic mobility shift assay

Electrophoretic mobility shift assays were performed as follows. The synthetic repressor and 50 ng linear promoter DNA were incubated in a protein-DNA binding buffer consisting of 15 mM Hepes-sodium hydroxide (pH 7.9), 100 mM potassium

chloride, 5 mM magnesium chloride, 100 μM zinc sulfate, 10% (v/v) glycerol, 100 μg/mL Bovine Serum Albumin, and 10 mM β-mercaptoethanol. The reaction was incubated for 4 hour at room temperature. The reactions were run using a Novex® DNA Retardation Gel in 0.5X TBE buffer and stained in 1X SYBR green stain in 0.5X TBE at room temperature for 20 minutes. The gel was then washed twice with 150 mL water for ∼10 seconds prior to visualization.

B.5 *In vitro* transcription assay

In vitro transcription reactions were performed as follows. Reactions consisted of 1X *E. coli* RNA polymerase transcription buffer, 5 mM TCEP, 0.5 mM each NTP, repressor-DNA mixture, and 2.5 units *E. coli* RNA polymerase holoenzyme (EPICENTRE® Biotechnologies, Madison, WI) in a 50 μL total volume. The repressor-DNA mixture consisted of 20 μL eluant from protein purification and 60 ng linearized PCR amplified DNA. In reactions in which the protein and DNA were pre-incubated prior to addition of RNA polymerase, the incubation was at room temperature for 2 hours. Although the exact repressor concentration is unknown, I estimate that the protein was in a molar excess of the DNA so most of the DNA should be bound. Subsequently, the reactions were run on a 1% agarose gel in 1X TAE buffer at 6.5 V/cm. To verify nucleic acid fragment size, both 1 μg ssRNA ladder and 2-log DNA ladder were run (New England Biolabs). The gel was stained with 1X SYBR gold in 1X TAE buffer for 40 minutes prior to visualization.

B.6 Protein solubility assay

Protein solubility was assayed as follows. Cultures of the synthetic repressor encoded by BBa_C2002, a different synthetic repressor encoded by BBa_C2100, and cells not expressing any DNA binding protein were grown overnight in LB medium supplemented with the appropriate antibiotic from a single colony. I added 2 mL culture aliquots to 2 2-mL Eppendorf tubes. The cells were pelleted by centrifugation at

4,000 $\times$ g for 15 minutes at 4°C. One tube was resuspended in 50 μL denaturing lysis buffer (as above). The second tube was resuspended in 50 μL native lysis buffer (50 mM monobasic sodium phosphate, 300 mM sodium chloridge, 10 mM imidazole at pH 8.0). The lysis mixtures were freeze-thawed for 3 cycles using a dry ice-ethanol bath and slushy ice. The denaturing lysis mixture was agitated for 1 hour at room temperature. Both lysis mixtures were centrifuged at 10,000 $\times$ g at 4°C for 30 minutes. The supernatant of the denaturing lysis mixture was the total protein. The supernatant of the native lysis mixture was the soluble protein fraction. I then resuspended the pellet of the native lysate in 50 μL denaturing lysis buffer and centrifuged at 10,000 $\times$ g at 4°C for 20 minutes to obtain the insoluble protein fraction. All samples were run using a NuPAGE Novex 10% Bis-Tris Gel in 1X MES running buffer. To estimate molecular weight, 20 μL of SeeBlue Plus2 Pre-Stained Standard was run as well. The gel was stained using SimplyBlue^{TM} SafeStain, using a modified version of the manufacturer's instructions [279]. All reagents for SDS-polyacrylamide gel electrophoresis were obtained from Invitrogen.

B.7 β-galactosidase assay

From single colonies, 1 mL cultures were grown overnight in EZ Rich media supplemented with either 0.4% glycerol or 0.2% glucose and 100 μM zinc sulfate (Teknova, Hollister, CA). Cultures were diluted back to OD600nm of 0.01 in a 5 mL volume and grown 37°C for 2 hours. Then, 1 mM IPTG and 1 μM AHL was added to the cultures, as appropriate. I periodically sampled the growing culture and assayed for β-galactosidase activity [280, 281, 282]. Absorbance of 175 μL of culture was measured in a Wallac Victor3 multi-well fluorimeter (Perkin Elmer, Waltham, MA) at 30°C (600 nm absorbance filter, 0.1 seconds counting time through 4.375 mm of fluid). If appropriate, fluorescence measurements were taken as well (488 nm excitation filter, 525 nm emission filter, 0.1 sec, CW lamp energy 12901 units). I then added 20 μL of culture to 80 μL of permeabilization solution (100 mM dibasic sodium phosphate, 20 mM potassium chloride, 2 mM magnesium sulfate, 0.8 mg/mL hex-

adecyltrimethylammonium bromide or CTAB, 0.4 mg/mL sodium deoxycholate, and 2.7 mM TCEP) in a 96-well untreated, polystyrene, flat bottom microplate. When doing multiple time points, wells were parafilmed to prevent evaporation and plate was left at room temperature, since permeabilization renders the sample stable for several hours. When all samples are permeabilized, 25 μL each permeabilized sample was added to 150 μL substrate solution (60 mM dibasic sodium phosphate, 40 mM monobasic sodium phosphate, 1 mg/mL o-nitrophenyl-β-D-Galactoside or ONPG, and 1.35 mM TCEP) in a new microplate. The samples are then assayed using the multi-well fluorimeter at 30°C with an automatically repeating protocol of absorbance measurements (430 nm absorbance filter, 0.1 seconds counting time through 5 mm of fluid). β-galactosidase activity was calculated using the following equation.

$$\beta\text{-galactosidase activity} = 1000 * \frac{\text{Slope}}{v * A600} \tag{B.4}$$

Slope is the slope of a line fit through the linear portion of the A430 versus time data, v is the culture volume used in the assay (0.005 mL), and A600 is the culture absorbance.

B.8 *In vivo* fluorescence measurements using flow cytometry

Cultures were grown under similar conditions to those described for β-galactosidase activity except that 10 μM AHL was used. Cultures were then diluted 5-fold in phosphate buffered saline and the fluorescence measured using a BD FACScan flow cytometer equipped with a 488nm excitation laser and 530/30 filter (Becton Dickinson, MIT Flow Cytometry Facility). Data corresponding to cells were gated by both forward and side scatter. Data were analyzed using FlowJo version 6.3 from Tree Star, Inc. in Ashland, OR.

B.9 Measurement of device transfer curve

From three single colonies, 1 mL cultures were grown overnight at 37°C in EZ Rich media supplemented with either 0.4% glycerol or 0.2% glucose (Teknova, Hollister, CA). Cultures were diluted back to OD600nm of 0.01 and 1 mL aliquots were distributed into 96-deep well culture plates. Varying concentrations of AHL (0 M, 1 pM, 10 pM, 100 pM, 1 nM, 10 nM, 100 nM, 1 μM, 10 μM, and 50 μM) were added to the wells and the cultures grown for several more hours. I loaded 200 μL of each sample into a 96-well plate. The plate was incubated in a Wallac Victor3 multi-well fluorimeter (Perkin Elmer, Waltham, MA) at 37°C and assayed for absorbance (600 nm absorbance filter, 0.1 seconds counting time through 5 mm of fluid) and fluorescence (488 nm excitation filter, 525 nm emission filter, 0.1 sec, CW lamp energy 12901 units).

We analyzed the resulting data to compute the fluorescence per cell density. Raw data from the Wallac Victor3 multi-well fluorimeter was processed by first subtracting the appropriate backgrounds and then scaling the fluorescence data to the corresponding absorbance data as previously described. For each colony, data was normalized to the maximum value across the three replicates. Data across the three colonies were averaged. Error bars are 95% confidence intervals.

Appendix C

Naming of BioBrick vectors

BioBrick vector names take the form pSB#X#. The letters pSB are an acronym for plasmid Synthetic Biology. The first number denotes the origin of replication (Table C.1). The letter X identifies the antibiotic resistance marker(s) present in the vector (Table C.2). Vectors with multiple resistance markers have multiple, successive letters. Finally, the last number in the vector name is a version number to differentiate between the various implementations of the pSB series of vectors (Table C.3). When referring to both a BioBrick standard biological part and the vector in which it is cloned, the convention is to use the form [vector name]-[part number] such as pSB4K5-T9003. To refer to BioBrick vectors to be used for construction of BioBrick parts, use the full vector name and default cloned part. For example, pSB4A3-P1010, pSB1A10-P1010, pSB4K5-I52002, and pSB3T5-I52001 are all available vectors from the Registry of Standard Biological Parts. However, for convenience, vector names are often abbreviated to pSB4A3, pSB1A10, pSB4K5, and pSB3T5, respectively. New plasmid-based vectors constructed from the BioBrick base vector BBa_I51020 should be named pSB#X5-I52002 where the # is determined by the identity of the replication origin and the letter X is determined by the antibiotic resistance marker(s) present. To expand the BioBrick vector nomenclature, submit new vectors or vector parts to the Registry of Standard Biological Parts and then document any new annotation needed at `http://partsregistry.org/wiki/index.php/Help:Plasmids/Nomenclature`. The BioBricks Foundation is leading an open standards

Number	*Replication origin*	*Copy number* [248]	*Purpose*
1	modified pMB1 derived from pUC19	500-700	Easy plasmid DNA purification
2	F and P1 lytic derived from pSCANS-1-BNL [283]	1-2 inducible to high copy	Inducible copy number
3	p15A derived from pMR101	10-12	Multi-plasmid engineered systems
4	rep101, repA derived from pSC101	$\tilde{5}$	Small cell to cell copy number variation
5	derived from F plasmid	1-2	Improved plasmid stability
6	pMB1 derived from pBR322	15-20	Multi-plasmid engineered systems

Table C.1: Numeric abbreviations for plasmid replication origins in BioBrick vector nomenclature.

setting process should any revisions to the BioBrick vector nomenclature beyond addition of new replication origins, antibiotic resistance markers and version numbers be needed.

Code	*Antibiotic*
A	ampicillin
C	chloramphenicol
E	erythromycin
G	gentamycin
K	kanamycin
N	neomycin
Na	nalidixic acid
R	rifampicin
S	spectinomycin
St	streptomycin
T	tetracycline
Tm	trimethoprim
Z	zeocin

Table C.2: Letter abbreviations for antibiotic resistance markers in BioBrick vector nomenclature.

Number	*Key features*	*Purpose*	*Example*	*Designer*
0	absent or incomplete BioBrick cloning site		pSB2K0	Brookhaven National Lab
1	complete BioBrick cloning site (BCS)	assembly of BioBrick parts	pSB4A1	Reshma Shetty
2	5' terminator and BCS	transcriptional insulation of vector upstream of cloned BioBrick part	pSB1A2	Tom Knight
3	5' terminator and BCS and 3' terminator	transcriptional insulation of vector downstream of cloned BioBrick part	pSB1AC3	Reshma Shetty & Tom Knight
4	pSB2K3-derived vector free of many restriction sites	Genome refactoring [284]	pSB2K4	Leon Chan
5	constructed from BioBrick base vector	standardized BioBrick vector design	pSB4K5	Reshma Shetty
6	Reserved	-	-	-
7	BCS flanked by terminators	transcriptional insulation of cloned BioBrick part [285]	pSB1A7	Karmella Haynes *et al.*
8	Unassigned	-	-	-
9	Unassigned	-	-	-
10	Screening plasmid v1.0 [286]	characterization of single input, single output transcriptional devices	pSB1A10	Josh Michener & Jason Kelly

Table C.3: Numeric abbreviations for vector version number in BioBrick vector nomenclature.

Bibliography

[1] Merritt Roe Smith. *Harpers Ferry Armory and the New Technology*. Cornell University Press, Ithaca, New York, 1977.

[2] Barry M Leiner, Vinton G Cerf, David D Clark, Robert E Kahn, Leonard Kleinrock, Daniel C Lynch, Jon Postel, Larry G Roberts, and Stephen Wolff. A brief history of the internet, December 2003. Available from: `http://www.isoc.org/internet/history/brief.shtml`.

[3] Vinton G Cerf and Robert E Kahn. A protocol for packet network intercommunication. *IEEE Transactions on Communications*, 22(5):637–48, 1974.

[4] Waclaw Szybalski and Ann Skalka. Nobel prizes and restriction enzymes. *Gene*, 4(3):181–2, 1978.

[5] Paul Berg. Co-chairman's remarks: reverse genetics: directed modification of DNA for functional analysis. *Gene*, 135(1-2):261–4, 1993.

[6] Irving S Johnson. Human insulin from recombinant DNA technology. *Science*, 219(4585):632–7, 1983.

[7] Drew Endy. Foundations for engineering biology. *Nature*, 438(7067):449–53, 2005.

[8] D A Hopwood, F Malpartida, H M Kieser, H Ikeda, J Duncan, I Fujii, B A Rudd, H G Floss, and S Omura. Production of 'hybrid' antibiotics by genetic engineering. *Nature*, 314(6012):642–4, 1985.

[9] James E Bailey. Toward a science of metabolic engineering. *Science*, 252(5013):1668–75, 1991.

[10] Lara L Madison and Gialt W Huisman. Metabolic engineering of poly(3-hydroxyalkanoates): from DNA to plastic. *Microbiol Mol Biol Rev*, 63(1):21–53, 1999.

[11] Blaine A Pfeifer and Chaitan Khosla. Biosynthesis of polyketides in heterologous hosts. *Microbiol Mol Biol Rev*, 65(1):106–18, 2001.

[12] C E Nakamura and G M Whited. Metabolic engineering for the microbial production of 1,3-propanediol. *Curr Opin Biotechnol*, 14(5):454–9, 2003.

[13] Dae-Kyun Ro, Eric M Paradise, Mario Ouellet, Karl J Fisher, Karyn L Newman, John M Ndungu, Kimberly A Ho, Rachel A Eachus, Timothy S Ham, James Kirby, Michelle C Y Chang, Sydnor T Withers, Yoichiro Shiba, Richmond Sarpong, and Jay D Keasling. Production of the antimalarial drug precursor artemisinic acid in engineered yeast. *Nature*, 440(7086):940–3, 2006.

[14] Chaitan Khosla and Jay D Keasling. Metabolic engineering for drug discovery and development. *Nat Rev Drug Discov*, 2(12):1019–25, 2003.

[15] Jose Manuel Otero, Lisbeth Olsson, and Jens Nielsen. Industrial biotech meets systems biology. *Genetic Engineering & Biotechnology News*, 27(1), 2007.

[16] Gregory Stephanopoulos. Challenges in engineering microbes for biofuels production. *Science*, 315(5813):801–4, 2007.

[17] David Baker, George Church, Jim Collins, Drew Endy, Joseph Jacobson, Jay Keasling, Paul Modrich, Christina Smolke, and Ron Weiss. Engineering life: building a fab for biology. *Sci Am*, 294(6):44–51, 2006.

[18] Christopher A Voigt. Genetic parts to program bacteria. *Curr Opin Biotechnol*, 17(5):548–57, 2006.

[19] J Christopher Anderson, Elizabeth J Clarke, Adam P Arkin, and Chris A Voigt. Environmentally controlled invasion of cancer cells by engineered bacteria. *J Mol Biol*, 355(4):619–27, 2006.

[20] Thomas F Knight. Idempotent vector design for standard assembly of BioBricks. Technical report, MIT Synthetic Biology Working Group Technical Reports, 2003. Available from: `http://hdl.handle.net/1721.1/21168`.

[21] Drew Endy, Isadora Deese, and Chuck Wadey. Adventures in synthetic biology. *Nature*, 438(7067):449–53, 2005.

[22] Barry Canton, Ania Labno, and Drew Endy. BBa_F2620, an engineered cell-cell communication receiver device. *Nat Biotechnol*, (accepted).

[23] Calin C Guet, Michael B Elowitz, Weihong Hsing, and Stanislas Leibler. Combinatorial synthesis of genetic networks. *Science*, 296(5572):1466–70, 2002.

[24] Hal Alper, Curt Fischer, Elke Nevoigt, and Gregory Stephanopoulos. Tuning genetic control through promoter engineering. *Proc Natl Acad Sci USA*, 102(36):12678–83, 2005.

[25] Sara Hooshangi, Stephan Thiberge, and Ron Weiss. Ultrasensitivity and noise propagation in a synthetic transcriptional cascade. *Proc Natl Acad Sci USA*, 102(10):3581–3856, 2005.

[26] Linda Buck and Richard Axel. A novel multigene family may encode odorant receptors: a molecular basis for odor recognition. *Cell*, 65(1):175–87, 1991.

[27] Alexander Vainstein, Efraim Lewinsohn, Eran Pichersky, and David Weiss. Floral fragrance. New inroads into an old commodity. *Plant Physiol*, 127(4):1383–9, 2001.

[28] Joan C Han and Grace Y Han. A procedure for quantitative determination of tris(2-carboxyethyl)phosphine, an odorless reducing agent more stable and effective than dithiothreitol. *Anal Biochem*, 220(1):5–10, 1994.

[29] Dandan Wang, Xuedong Ding, and Philip N Rather. Indole can act as an extracellular signal in *Escherichia coli*. *J Bacteriol*, 183(14):4210–6, 2001.

[30] P Di Martino, R Fursy, L Bret, B Sundararaju, and R S Phillips. Indole can act as an extracellular signal to regulate biofilm formation of *Escherichia coli* and other indole-producing bacteria. *Can J Microbiol*, 49(7):443–9, 2003.

[31] Joanna Domka, Jintae Lee, and Thomas K Wood. YliH (BssR) and YceP (BssS) regulate *Escherichia coli* K-12 biofilm formation by influencing cell signaling. *Appl Environ Microbiol*, 72(4):2449–59, 2006.

[32] Jintae Lee, Arul Jayaraman, and Thomas K Wood. Indole is an inter-species biofilm signal mediated by SdiA. *BMC Microbiol*, 7(1471-2180 (Electronic)):42, 2007.

[33] Hidetada Hirakawa, Yoshihiko Inazumi, Takeshi Masaki, Takahiro Hirata, and Akihito Yamaguchi. Indole induces the expression of multidrug exporter genes in *Escherichia coli*. *Mol Microbiol*, 55(4):1113–26, 2005.

[34] Eleanor L Chant and David K Summers. Indole signalling contributes to the stable maintenance of *Escherichia coli* multicopy plasmids. *Mol Microbiol*, 63(1):35–43, 2007.

[35] E E Snell. Tryptophanase: structure, catalytic activities, and mechanism of action. *Adv Enzymol Relat Areas Mol Biol*, 42(0065-258X (Print)):287–333, 1975.

[36] Ferdousi Rezwan, Ruiting Lan, and Peter R Reeves. Molecular basis of the indole-negative reaction in *Shigella* strains: extensive damages to the *tna* operon by insertion sequences. *J Bacteriol*, 186(21):7460–5, 2004.

[37] Ying-Ying Chang, Ai-Yu Wang, and John E Cronan Jr. Expression of *Escherichia coli* pyruvate oxidase (PoxB) depends on the sigma factor encoded by the *rpoS(katF)* gene. *Mol Microbiol*, 11(6):1019–28, 1994.

[38] Lloyd W Ream, Linda Margossian, Alvin J Clark, Fleming G Hansen, and Kaspar von Meyenburg. Genetic and physical mapping of *recF* in *Escherichia coli* K-12. *Mol Gen Genet*, 180(1):115–21, 1980.

[39] R Mark Edwards and Michael D Yudkin. Location of the gene for the low-affinity tryptophan-specific permease of *Escherichia coli*. *Biochem J*, 204(2):617–9, 1982.

[40] Tim Durfee, Richard Nelson, Schuyler Baldwin, Guy 3rd Plunkett, Valerie Burland, Bob Mau, Joseph F Petrosino, Xiang Qin, Donna M Muzny, Mulu Ayele, Richard A Gibbs, Balint Csorgo, Gyorgy Posfai, George M Weinstock, and Frederick R Blattner. The complete genome sequence of *Escherichia coli* DH10B: insights into the biology of a laboratory workhorse. *J Bacteriol*, 190(7):2597–606, 2008.

[41] Scott Hagedorn and Bryan Kaphammer. Microbial biocatalysis in the generation of flavor and fragrance chemicals. *Annu Rev Microbiol*, 48(0066-4227 (Print)):773–800, 1994.

[42] Jihong Wang, Natalia Dudareva, Shyam Bhakta, Robert A Raguso, and Eran Pichersky. Floral scent production in *Clarkia breweri* (onagraceae). II. localization and developmental modulation of the enzyme S-adenosyl-L-methionine:(iso)eugenol O-methyltransferase and phenylpropanoid emission. *Plant Physiol*, 114(1):213–21, 1997.

[43] Natalia Dudareva, John C D'Auria, Kyoung H N Nam, Robert A Raguso, and Eran Pichersky. Acetyl-coa:benzylalcohol acetyltransferase–an enzyme involved in floral scent production in *Clarkia breweri*. *Plant J*, 14(3):297–304, 1998.

[44] Jeannine R Ross, Kyoung H N Nam, John C D'Auria, and Eran Pichersky. S-adenosyl-L-methionine:salicylic acid carboxyl methyltransferase, an enzyme involved in floral scent production and plant defense, represents a new class of plant methyltransferases. *Arch Biochem Biophys*, 367(1):9–16, 1999.

[45] Florence Negre, Natalia Kolosova, Joseph Knoll, Christine M Kish, and Natalia Dudareva. Novel *S*-adenosyl-L-methionine:salicylic acid carboxyl methyltransferase, an enzyme responsible for biosynthesis of methyl salicylate and methyl benzoate, is not involved in floral scent production in snapdragon flowers. *Arch Biochem Biophys*, 406(2):261–70, 2002.

[46] Catherine Emily Horton, Ke-Xue Huang, George N Bennett, and Frederick B Rudolph. Heterologous expression of the *Saccharomyces cerevisiae* alcohol acetyltransferase genes in *Clostridium acetobutylicum* and *Escherichia coli* for the production of isoamyl acetate. *J Ind Microbiol Biotechnol*, 30(7):427–32, 2003.

[47] Jeremy Kapteyn, Anthony V Qualley, Zhengzhi Xie, Eyal Fridman, Natalia Dudareva, and David R Gang. Evolution of cinnamate/*p*-coumarate carboxyl methyltransferases and their role in the biosynthesis of methylcinnamate. *Plant Cell*, 19(10):3212–29, 2007.

[48] Hak S Seo, Jong T Song, Jong-Joo Cheong, Yong-Hwan Lee, Yin-Won Lee, Ingyu Hwang, Jong S Lee, and Yang D Choi. Jasmonic acid carboxyl methyltransferase: a key enzyme for jasmonate-regulated plant responses. *Proc Natl Acad Sci USA*, 98(8):4788–93, 2001.

[49] Natalia Dudareva, Lisa M Murfitt, Craig J Mann, Nina Gorenstein, Natalia Kolosova, Christine M Kish, Connie Bonham, and Karl Wood. Developmental regulation of methyl benzoate biosynthesis and emission in snapdragon flowers. *Plant Cell*, 12(6):949–61, 2000.

[50] Lisa M Murfitt, Natalia Kolosova, Craig J Mann, and Natalia Dudareva. Purification and characterization of *S*-adenosyl-L-methionine:benzoic acid carboxyl methyltransferase, the enzyme responsible for biosynthesis of the volatile ester methyl benzoate in flowers of *Antirrhinum majus*. *Arch Biochem Biophys*, 382(1):145–51, 2000.

[51] Florence Negre, Christine M Kish, Jennifer Boatright, Beverly Underwood, Kenichi Shibuya, Conrad Wagner, David G Clark, and Natalia Dudareva. Regulation of methylbenzoate emission after pollination in snapdragon and petunia flowers. *Plant Cell*, 15(12):2992–3006, 2003.

[52] Beverly A Underwood, Denise M Tieman, Kenichi Shibuya, Richard J Dexter, Holly M Loucas, Andrew J Simkin, Charles A Sims, Eric A Schmelz, Harry J Klee, and David G Clark. Ethylene-regulated floral volatile synthesis in petunia corollas. *Plant Physiol*, 138(1):255–66, 2005.

[53] Toshio Fujii, Naoshi Nagasawa, Akihiro Iwamatsu, Takayuki Bogaki, Yukio Tamai, and Masaaki Hamachi. Molecular cloning, sequence analysis, and expression of the yeast alcohol acetyltransferase gene. *Appl Environ Microbiol*, 60(8):2786–92, 1994.

[54] R Singh, PV Vadlani, ML Harrison, GN Bennett, and KY San. Aerobic production of isoamyl acetate by overexpression of the yeast alcohol acetyl-transferases AFT1 and AFT2 in *Escherichia coli* and using low-cost fermentation ingredients. *Bioprocess Biosyst Eng*, (1615-7591 (Print)), 2007.

[55] Herb E Schellhorn, Jonathon P Audia, Linda I Wei, and Lily Chang. Identification of conserved, RpoS-dependent stationary-phase genes of *Escherichia coli*. *J Bacteriol*, 180(23):6283–91, 1998.

[56] Somalinga R V Vijayakumar, Mark G Kirchhof, Cheryl L Patten, and Herb E Schellhorn. RpoS-regulated genes of *Escherichia coli* identified by random lacZ fusion mutagenesis. *J Bacteriol*, 186(24):8499–507, 2004.

[57] Regine Hengge-Aronis. Survival of hunger and stress: the role of *rpoS* in early stationary phase gene regulation in e. coli. *Cell*, 72(2):165–8, 1993.

[58] R Lange and Regine Hengge-Aronis. Identification of a central regulator of stationary-phase gene expression in *Escherichia coli*. *Mol Microbiol*, 5(1):49–59, 1991.

[59] Kan Tanaka, Yuko Takayanagi, Nobuyuki Fujita, Akira Ishihama, and Hideo Takahashi. Heterogeneity of the principal σ factor in *Escherichia coli*: the *rpoS* gene product, σ^{38}, is a second principal sigma factor of RNA polymerase in stationary-phase *Escherichia coli*. *Proc Natl Acad Sci USA*, 90(8):3511–5, 1993.

[60] Hisakage Funabashi, Tetsuya Haruyama, Masayasu Mie, Yasuko Yanagida, Eiry Kobatake, and Masuo Aizawa. Non-destructive monitoring of *rpoS* promoter activity as stress marker for evaluating cellular physiological status. *J Biotechnol*, 95(1):85–93, 2002.

[61] Harry H Yim and Merna Villarejo. *osmY*, a new hyperosmotically inducible gene, encodes a periplasmic protein in *Escherichia coli*. *J Bacteriol*, 174(11):3637–44, 1992.

[62] Harry H Yim, Rory L Brems, and Merna Villarejo. Molecular characterization of the promoter of *osmY*, an *rpoS*-dependent gene. *J Bacteriol*, 176(1):100–7, 1994.

[63] Regine Hengge-Aronis, Roland Lange, Nicola Henneberg, and Daniela Fischer. Osmotic regulation of *rpoS*-dependent genes in *Escherichia coli*. *J Bacteriol*, 175(1):259–65, 1993.

[64] Rolf Lutz and Hermann Bujard. Independent and tight regulation of transcriptional units in *Escherichia coli* via the LacR/O, the TetR/O and AraC/I_1-I_2 regulatory elements. *Nucleic Acids Res*, 25:1203–10, 1997.

[65] Govind Rao, Shabbir Bambot, Simon Kwong, Henryk Szmacinski, Jeffrey Sipior, Raja Holavanahali, and Gary Carter. Application of fluorescence sensing to bioreactors. In Joseph R Lakowicz, editor, *Topics in Fluorescence Spectroscopy*, volume 4, chapter 13, pages 417–448. Springer US, 2002.

[66] D C Cameron and I T Tong. Cellular and metabolic engineering. an overview. *Appl Biochem Biotechnol*, 38(1-2):105–40, 1993.

[67] Gregory Stephanopoulos and Joseph J Vallino. Network rigidity and metabolic engineering in metabolite overproduction. *Science*, 252(5013):1675–81, 1991.

[68] Matthias Laska, Alexandra Seibt, and Andreas Weber. 'Microsmatic' primates revisited: olfactory sensitivity in the squirrel monkey. *Chem Senses*, 25(1):47–53, 2000.

[69] Gordon M Shepherd. The human sense of smell: are we better than we think? *PLoS Biol*, 2(5):E146, 2004.

[70] Ron G Buttery, Roy Teranishi, Louisa C Ling, and Jean G Turnbaugh. Quantitative and sensory studies on tomato paste volatiles. *J Agric Food Chem*, 38(1):336–340, 1990.

[71] Laura Serino, Cornelia Reimmann, Heinz Baur, Markus Beyeler, Paolo Visca, and Dieter Haas. Structural genes for salicylate biosynthesis from chorismate in *Pseudomonas aeruginosa*. *Mol Gen Genet*, 249(2):217–28, 1995.

[72] Shota Atsumi, Taizo Hanai, and James C Liao. Non-fermentative pathways for synthesis of branched-chain higher alcohols as biofuels. *Nature*, 451(7174):86–9, 2008.

[73] Vassily Hatzimanikatis, Christodoulos A Floudas, and James E Bailey. Analysis and design of metabolic reaction networks via mixed-integer linear optimization. *AIChE J*, 42(5):1277–92, May 1995.

[74] David A Fell. Metabolic control analysis: a survey of its theoretical and experimental development. *Biochem J*, 286 (Pt 2)(0264-6021 (Print)):313–30, 1992.

[75] Vassily Hatzimanikatis and James E Bailey. MCA has more to say. *J Theor Biol*, 182(3):233–42, 1996.

[76] Gregory Stephanopoulos. Metabolic fluxes and metabolic engineering. *Metab Eng*, 1(1):1–11, 1999.

[77] Richard S Sandige. *The Electrical Engineering Handbook*, chapter Digital Devices. CRC Press, Inc., 1993.

[78] Gregory L Moss, Peter Graham, Richard S Sandige, and H S Hinton. *The Electrical Engineering Handbook*, chapter Logic Elements. CRC Press, Inc., 1993.

[79] Steve Ward and Chris Terman. Digital abstraction. Available from: `http://tinyurl.com/2ppws`.

[80] Kaustav Banerjee. ECE 124A VLSI principles lecture 6, 2004. Available from: `http://www.ece.ucsb.edu/courses/ECE124/124A_F04Banerjee/lecture/Lecture6.pdf`.

[81] C F Hill. Noise margin and noise immunity in logic circuits. *Microelectronics*, 1:16–22, April 1968.

[82] Michael B Elowitz, Arnold J Levine, Eric D Siggia, and Peter S Swain. Stochastic gene expression in a single cell. *Science*, 297(5584):1183–6, 2002.

[83] Jan Lohstroh. Static and dynamic noise margins of logic circuits. *IEEE Journal of Solid-State Circuits*, 14(3):591–598, June 1979.

[84] Thomas F Knight and Gerald Jay Sussman. Cellular gate technology. July 1997. Available from: `http://www.ai.mit.edu/people/tk/ce/cellgates.ps`.

[85] Ron Weiss, Thomas F Knight Jr, and Gerald Sussman. *Cellular Computing*, chapter 4, pages 43–73. Series in Systems Biology. Oxford University Press, 2004.

[86] Christopher Batten, Ronny Krashinksky, and Thomas Knight Jr. A scalable cellular logic technology using zinc-finger proteins. In *3rd Workshop on Non-Silicon Computing*, 2004.

[87] Michael Bernard Elowitz. *Transport, Assembly, and Dynamics in Systems of Interacting Proteins*. PhD thesis, Princeton University, 1999.

[88] J B Andersen, C Sternberg, L K Poulsen, S P Bjorn, M Givskov, and S Molin. New unstable variants of green fluorescent protein for studies of transient gene expression in bacteria. *Appl Environ Microbiol*, 64:2240–6, 1998.

[89] Mark Ptashne. *A Genetic Switch: Phage Lambda Revisited*. Cold Spring Harbor Laboratory Press, 3rd edition, 2004.

[90] Sun Bok Lee and James E Bailey. Genetically structured models for *lac* promoter-operator function in the *Escherichia coli* chromosome and in multi-copy plasmids: *lac* operator function. *Biotechnol Bioeng*, 26:1372–1382, November 1984.

[91] Michael Andrew Gibson and Eric Mjolsness. *Computational Modeling of Genetic and Biochemical Networks*, chapter 1, pages 3–48. The MIT Press, Cambridge, Massachusetts, 2001.

[92] Ken A Dill and Sarina Bromberg. *Molecular Driving Forces: Statistical Thermodynamics in Chemistry and Biology*. Garland Science, New York, New York, 2003.

[93] Lacramioara Bintu, Nicolas E Buchler, Hernan G Garcia, Ulrich Gerland, Terence Hwa, Jane Kondev, and Rob Phillips. Transcriptional regulation by the numbers: models. *Curr Opin Genet Dev*, 15(0959-437X):116–24, 2005.

[94] C D Smolke, A Khlebnikov, and J D Keasling. Effects of transcription induction homogeneity and transcript stability on expression of two genes in a constructed operon. *Appl Microbiol Biotechnol*, 57(5-6):689–96, 2002.

[95] Christina D Smolke and Jay D Keasling. Effect of copy number and mRNA processing and stabilization on transcript and protein levels from an engineered dual-gene operon. *Biotechnol Bioeng*, 78:412–24, 2002.

[96] Christina D Smolke and Jay D Keasling. Effect of gene location, mRNA secondary structures, and RNase sites on expression of two genes in an engineered operon. *Biotechnol Bioeng*, 80:762–76, 2002.

[97] Kenneth C Keiler, Patrick R Waller, and Robert T Sauer. Role of a peptide tagging system in degradation of proteins synthesized from damaged messenger RNA. *Science*, 271(5251):990–3, 1996.

[98] Kenneth C Keiler and Robert T Sauer. Sequence determinants of C-terminal substrate recognition by the Tsp protease. *J Biol Chem*, 271(5):2589–93, 1996.

[99] David E Draper. *Escherichia coli and Salmonella: Cellular and molecular biology*, chapter 59, pages 902–908. ASM Press, Washington, DC, second edition, 1996.

[100] Sidney R Kushner. *Escherichia coli and Salmonella: Cellular and molecular biology*, chapter 56, pages 849–860. ASM Press, Washington, DC, second edition, 1996.

[101] Shan Sundararaj, Anchi Guo, Bahram Habibi-Nazhad, Melania Rouani, Paul Stothard, Michael Ellison, and David S Wishart. The CyberCell Database (CCDB): a comprehensive, self-updating, relational database to coordinate and facilitate *in silico* modeling of *Escherichia coli*. *Nucleic Acids Res*, 32 (Database issue):D293–5, January 2004.

[102] Daniel E Koshland Jr, Albert Goldbeter, and Jeffry B Stock. Amplification and adaptation in regulatory and sensory systems. *Science*, 217(4556):220–5, 1982.

[103] John R S Newman and Amy E Keating. Comprehensive identification of human bZIP interactions with coiled-coil arrays. *Science*, 300(5628):2097–101, 2003.

[104] Jessica A Hurt, Stacey A Thibodeau, Andrew S Hirsh, Carl O Pabo, and J Keith Joung. Highly specific zinc finger proteins obtained by directed domain shuffling and cell-based selection. *Proc Natl Acad Sci USA*, 100(21):12271–6, 2003.

[105] Douglas F Stickle, Karen M Vossen, Daniel A Riley, and Michael G Fried. Free DNA concentration in *E. coli* estimated by an analysis of competition for DNA binding proteins. *J Theor Biol*, 168(1):1–12, 1994.

[106] Audun Bakk and Ralf Metzler. Nonspecific binding of the O_R repressors CI and Cro of bacteriophage λ. *J Theor Biol*, 231(4):525–33, 2004.

[107] Audun Bakk and Ralf Metzler. *In vivo* non-specific binding of λ CI and Cro repressors is significant. *FEBS Lett*, 563(1-3):66–8, 2004.

[108] Benno Müller-Hill. *The lac operon*. Walter de Gruyter, 1996.

[109] Benno Müller-Hill. The function of auxiliary operators. *Mol Microbiol*, 29(1):13–8, 1998.

[110] Bernard Révet, Brigitte von Wilcken-Bergmann, Heike Bessert, Andrew Barker, and Benno Müller-Hill. Four dimers of λ repressor bound to two suitably spaced pairs of λ operators form octamers and DNA loops over large distances. *Curr Biol*, 9(3):151–4, 1999.

[111] Ron Weiss and Thomas F Knight Jr. Engineered communications for microbial robotics. *DNA6: Sixth International Meeting on DNA Based Computers*, June 2000.

[112] Michael B Elowitz and Stanislas Leibler. A synthetic oscillatory network of transcriptional regulators. *Nature*, 403(6767):335–8, 2000.

[113] Timothy S Gardner, Charles R Cantor, and James J Collins. Construction of a genetic toggle switch in *Escherichia coli*. *Nature*, 403(6767):339–42, 2000.

[114] William E Bentley, Noushin Mirjalili, Dana C Andersen, Robert H Davis, and Dhinakar S Kompala. Plasmid-encoded protein: The principal factor in the "metabolic burden" associated with recombinant bacteria. *Biotechnol Bioeng*, 35(7):668–681, 1990.

[115] Bernard R Glick. Metabolic load and heterologous gene expression. *Biotechnol Adv*, 13(2):247–61, 1995.

[116] Otto G Berg. A model for the statistical fluctuations of protein numbers in a microbial population. *J Theor Biol*, 71(4):587–603, 1978.

[117] David R Rigney. Stochastic model of constitutive protein levels in growing and dividing bacterial cells. *J Theor Biol*, 76(4):453–80, 1979.

[118] Harley H McAdams and Adam Arkin. Stochastic mechanisms in gene expression. *Proc Natl Acad Sci USA*, 94(3):814–9, 1997.

[119] Harley H McAdams and Adam Arkin. It's a noisy business! genetic regulation at the nanomolar scale. *Trends Genet*, 15(2):65–9, 1999.

[120] Mukund Thattai and Alexander van Oudenaarden. Intrinsic noise in gene regulatory networks. *Proc Natl Acad Sci USA*, 98(15):8614–9, 2001.

[121] H Bremer, P Dennis, and M Ehrenberg. Free RNA polymerase and modeling global transcription in *Escherichia coli*. *Biochimie*, 85(6):597–609, 2003.

[122] Patrick P Dennis, Mans Ehrenberg, and Hans Bremer. Control of rRNA synthesis in *Escherichia coli*: a systems biology approach. *Microbiol Mol Biol Rev*, 68(4):639–68, 2004.

[123] H Bremer and P Dennis. Feedback control of ribosome function in *Escherichia coli*. *Biochimie*, 90(3):493–9, 2008.

[124] Nitzan Rosenfeld, Jonathan W Young, Uri Alon, Peter S Swain, and Michael B Elowitz. Gene regulation at the single-cell level. *Science*, 307(5717):1962–5, 2005.

[125] Hans Bremer and Patrick P Dennis. *Escherichia coli and Salmonella Typhimurium: Cellular and Molecular Biology*, chapter 97, pages 1553–1569. ASM Publications, 1996.

[126] Drew Endy and Roger Brent. Modelling cellular behaviour. *Nature*, 409(6818):391–5, 2001.

[127] Ertugrul M Ozbudak, Mukund Thattai, Iren Kurtser, Alan D Grossman, and Alexander van Oudenaarden. Regulation of noise in the expression of a single gene. *Nat Genet*, 31(1):69–73, 2002.

[128] Peter S Swain, Michael B Elowitz, and Eric D Siggia. Intrinsic and extrinsic contributions to stochasticity in gene expression. *Proc Natl Acad Sci USA*, 99(20):12795–800, 2002.

[129] Thuc T Le, Sebastien Harlepp, Calin C Guet, Kimberly Dittmar, Thierry Emonet, Tao Pan, and Philippe Cluzel. Real-time RNA profiling within a single bacterium. *Proc Natl Acad Sci USA*, 102(0027-8424):9160–4, 2005.

[130] Ido Golding, Johan Paulsson, Scott M Zawilski, and Edward C Cox. Real-time kinetics of gene activity in individual bacteria. *Cell*, 123(0092-8674 (Print)):1025–36, 2005.

[131] Arthur L Koch. The logarithm in biology: I. mechanisms generating the log-normal distribution exactly. *J Theor Biol*, 12(2):276–90, 1966.

[132] Arthur L Koch. The logarithm in biology: II. distributions simulating the log-normal. *J Theor Biol*, 23(2):251–68, 1969.

[133] Eckhard Limpert, Werner A Stahel, and Markus Abbt. Log-normal distributions across the sciences: Keys and clues. *BioScience*, 51(5):341–52, 2001.

[134] Prabhata K Swamee. Near lognormal distribution. *J of Hydrologic Engineering*, 7(6):441–444, 2002.

[135] Ron Weiss. *Cellular Computation and Communications using Engineered Genetic Regulatory Networks*. PhD thesis, Massachusetts Institute of Technology, 2001. Available from: `http://hdl.handle.net/1721.1/8228`.

[136] Yohei Yokobayashi, Ron Weiss, and Frances H Arnold. Directed evolution of a genetic circuit. *Proc Natl Acad Sci USA*, 99(26):16587–91, 2002.

[137] Johan Paulsson. Summing up the noise in gene networks. *Nature*, 427(6973):415–8, 2004.

[138] Diane K Hawley and William R McClure. Compilation and analysis of *Escherichia coli* promoter DNA sequences. *Nucleic Acids Res*, 11(8):2237–55, 1983.

[139] Takashi Aoyama, Mituru Takanami, Eiko Ohtsuka, Yoshio Taniyama, Ryuji Marumoto, Hiroko Sato, and Morio Ikehara. Essential structure of *E. coli* promoter: effect of spacer length between the two consensus sequences on promoter function. *Nucleic Acids Res*, 11(17):5855–64, 1983.

[140] Takashi Aoyama and Mituru Takanami. Essential structure of *E. coli* promoter II. effect of the sequences around the RNA start point on promoter function. *Nucleic Acids Res*, 13(11):4085–96, 1985.

[141] Ulrich Deuschle, Wolfgang Kammerer, Reiner Gentz, and Hermann Bujard. Promoters of *Escherichia coli*: a hierarchy of *in vivo* strength indicates alternate structures. *EMBO J*, 5(0261-4189 (Print)):2987–94, 1986.

[142] Wolfgang Kammerer, Ulrich Deuschle, Reiner Gentz, and Hermann Bujard. Functional dissection of *Escherichia coli* promoters: information in the transcribed region is involved in late steps of the overall process. *EMBO J*, 5(0261-4189 (Print)):2995–3000, 1986.

[143] Calvin B Harley and Robert P Reynolds. Analysis of *E. coli* promoter sequences. *Nucleic Acids Res*, 15(5):2343–61, 1987.

[144] Michael Brunner and Hermann Bujard. Promoter recognition and promoter strength in the *Escherichia coli* system. *EMBO J*, 6(0261-4189 (Print)):3139–44, 1987.

[145] Gerald Z Hertz and Gary D Stormo. *Escherichia coli* promoter sequences: analysis and prediction. *Methods Enzymol*, 273(0076-6879):30–42, 1996.

[146] Tom Burr, Jennie Mitchell, Annie Kolb, Steve Minchin, and Steve Busby. DNA sequence elements located immediately upstream of the -10 hexamer in *Escherichia coli* promoters: a systematic study. *Nucleic Acids Res*, 28(1362-4962 (Electronic)):1864–70, 2000.

[147] Rutl Hershberg, Gill Bejerano, Alberto Santos-Zavaleta, and Hanah Margalit. PromEC: An updated database of *Escherichia coli* mRNA promoters with experimentally identified transcriptional start sites. *Nucleic Acids Res*, 29(1):277, 2001.

[148] Jennie E Mitchell, Dongling Zheng, Stephen J W Busby, and Stephen D Minchin. Identification and analysis of 'extended -10' promoters in *Escherichia coli*. *Nucleic Acids Res*, 31(1362-4962 (Electronic)):4689–95, 2003.

[149] John W Drake. A constant rate of spontaneous mutation in DNA-based microbes. *Proc Natl Acad Sci USA*, 88(16):7160–4, 1991.

[150] Anselm Levskaya, Aaron A Chevalier, Jeffrey J Tabor, Zachary Booth Simpson, Laura A Lavery, Matthew Levy, Eric A Davidson, Alexander Scouras, Andrew D Ellington, Edward M Marcotte, and Christopher A Voigt. Synthetic biology: engineering *Escherichia coli* to see light. *Nature*, 438(7067):441–2, 2005.

[151] Subhayu Basu, Yoram Gerchman, Cynthia H Collins, Frances H Arnold, and Ron Weiss. A synthetic multicellular system for programmed pattern formation. *Nature*, 434(7037):1130–4, 2005.

[152] Hideki Kobayashi, Mads Kaern, Michihiro Araki, Kristy Chung, Timothy S Gardner, Charles R Cantor, and James J Collins. Programmable cells: Interfacing natural and engineered gene networks. *Proc Natl Acad Sci USA*, 2004.

[153] Michael L Simpson, Chris D Cox, Gregory D Peterson, and Gary S Sayler. Engineering in the biological substrate: Information processing in genetic circuits. *Proceedings of the IEEE*, 92(5):848–863, May 2004.

[154] Subhayu Basu, Rishabh Mehreja, Stephan Thiberge, Ming-Tang Chen, and Ron Weiss. Spatiotemporal control of gene expression with pulse-generating networks. *Proc Natl Acad Sci USA*, 101(17):6355–60, 2004.

[155] Nathan A Sallee, Brian J Yeh, and Wendell A Lim. Engineering modular protein interaction switches by sequence overlap. *J Am Chem Soc*, 129(15):4606–11, 2007.

[156] John E Dueber, Ethan A Mirsky, and Wendell A Lim. Engineering synthetic signaling proteins with ultrasensitive input/output control. *Nat Biotechnol*, 25(6):660–2, 2007.

[157] Maung Nyan Win and Christina D Smolke. A modular and extensible RNA-based gene-regulatory platform for engineering cellular function. *Proc Natl Acad Sci USA*, 104(36):14283–8, 2007.

[158] J Christopher Anderson, Christopher A Voigt, and Adam P Arkin. Environmental signal integration by a modular AND gate. *Mol Syst Biol*, 3(1744-4292 (Electronic)):133, 2007.

[159] Lingchong You, Robert Sidney 3rd Cox, Ron Weiss, and Frances H Arnold. Programmed population control by cell-cell communication and regulated killing. *Nature*, 428(6985):868–71, 2004.

[160] Daniel T Gillespie. Stochastic simulation of chemical kinetics. *Annu Rev Phys Chem*, 58(0066-426X (Print)):35–55, 2007.

[161] G M Rubin, M D Yandell, J R Wortman, G L Gabor Miklos, C R Nelson, I K Hariharan, M E Fortini, P W Li, R Apweiler, W Fleischmann, J M Cherry, S Henikoff, M P Skupski, S Misra, M Ashburner, E Birney, M S Boguski, T Brody, P Brokstein, S E Celniker, S A Chervitz, D Coates, A Cravchik, A Gabrielian, R F Galle, W M Gelbart, R A George, L S Goldstein, F Gong, P Guan, N L Harris, B A Hay, R A Hoskins, J Li, Z Li, R O Hynes, S J Jones, P M Kuehl, B Lemaitre, J T Littleton, D K Morrison, C Mungall, P H O'Farrell, O K Pickeral, C Shue, L B Vosshall, J Zhang, Q Zhao, X H Zheng, and S Lewis. Comparative genomics of the eukaryotes. *Science*, 287(5461):2204–15, 2000.

[162] Raymond S Brown, Christian Sander, and Patrick Argos. The primary structure of transcription factor TFIIIA has 12 consecutive repeats. *FEBS Lett*, 186(2):271–4, 1985.

[163] J Miller, A D McLachlan, and A Klug. Repetitive zinc-binding domains in the protein transcription factor IIIA from *Xenopus* oocytes. *EMBO J*, 4(6):1609–14, 1985.

[164] Barbara A Christy, Lester F Lau, and Daniel Nathans. A gene activated in mouse 3T3 cells by serum growth factors encodes a protein with "zinc finger" sequences. *Proc Natl Acad Sci USA*, 85(21):7857–61, 1988.

[165] Nikola P Pavletich and Carl O Pabo. Zinc finger-DNA recognition: crystal structure of a Zif268-DNA complex at 2.1 Å. *Science*, 252(5007):809–17, 1991.

[166] Mark Isalan, Yen Choo, and Aaron Klug. Synergy between adjacent zinc fingers in sequence-specific DNA recognition. *Proc Natl Acad Sci USA*, 94(11):5617–21, 1997.

[167] Scot A Wolfe, Harvey A Greisman, Elizabeth I Ramm, and Carl O Pabo. Analysis of zinc fingers optimized *via* phage display: evaluating the utility of a recognition code. *J Mol Biol*, 285:1917–34, 1999.

[168] Siyuan Tan, Dmitry Guschin, Albert Davalos, Ya-Li Lee, Andrew W Snowden, Yann Jouvenot, H Steven Zhang, Katherine Howes, Andrew R McNamara, Albert Lai, Chris Ullman, Lindsey Reynolds, Michael Moore, Mark Isalan, Lutz-Peter Berg, Bradley Campos, Hong Qi, S Kaye Spratt, Casey C Case, Carl O Pabo, Judith Campisi, and Philip D Gregory. Zinc-finger protein-targeted gene regulation: genomewide single-gene specificity. *Proc Natl Acad Sci USA*, 100(21):11997–2002, 2003.

[169] Edward J Rebar and Carl O Pabo. Zinc finger phage: affinity selection of fingers with new DNA-binding specificities. *Science*, 263(5147):671–3, 1994.

[170] Yen Choo and Aaron Klug. Toward a code for the interactions of zinc fingers with DNA: selection of randomized fingers displayed on phage. *Proc Natl Acad Sci USA*, 91(23):11163–7, 1994.

[171] Yen Choo and Aaron Klug. Selection of DNA binding sites for zinc fingers using rationally randomized DNA reveals coded interactions. *Proc Natl Acad Sci USA*, 91(23):11168–72, 1994.

[172] Carl O Pabo, Ezra Peisach, and Robert A Grant. Design and selection of novel Cys_2His_2 zinc finger proteins. *Annu Rev Biochem*, 70:313–40, 2001.

[173] Qiang Liu, David J Segal, Jayant B Ghiara, and Carlos F Barbas III. Design of polydactyl zinc-finger proteins for unique addressing within complex genomes. *Proc Natl Acad Sci USA*, 94(11):5525–30, 1997.

[174] Harvey A Greisman and Carl O Pabo. A general strategy for selecting high-affinity zinc finger proteins for diverse DNA target sites. *Science*, 275(5300):657–61, 1997.

[175] Jin-Soo Kim and Carl O Pabo. Getting a handhold on DNA: design of poly-zinc finger proteins with femtomolar dissociation constants. *Proc Natl Acad Sci USA*, 95(6):2812–7, 1998.

[176] Fyodor D Urnov and Edward J Rebar. Designed transcription factors as tools for therapeutics and functional genomics. *Biochem Pharmacol*, 64(5-6):919–23, 2002.

[177] Fyodor D Urnov, Jeffrey C Miller, Ya-Li Lee, Christian M Beausejour, Jeremy M Rock, Sheldon Augustus, Andrew C Jamieson, Matthew H Porteus, Philip D Gregory, and Michael C Holmes. Highly efficient endogenous human gene correction using designed zinc-finger nucleases. *Nature*, 435(7042):646–51, 2005.

[178] Sundar Durai, Mala Mani, Karthikeyan Kandavelou, Joy Wu, Matthew H Porteus, and Srinivasan Chandrasegaran. Zinc finger nucleases: custom-designed molecular scissors for genome engineering of plant and mammalian cells. *Nucleic Acids Res*, 33(18):5978–90, 2005.

[179] Monika Papworth, Paulina Kolasinska, and Michal Minczuk. Designer zinc-finger proteins and their applications. *Gene*, 366(1):27–38, 2006.

[180] David J Segal, Birgit Dreier, Roger R Beerli, and Carlos F Barbas III. Toward controlling gene expression at will: selection and design of zinc finger domains recognizing each of the 5'-GNN-3' DNA target sequences. *Proc Natl Acad Sci USA*, 96:2758–63, 1999.

[181] Birgit Dreier, David J Segal, and Carlos F Barbas III. Insights into the molecular recognition of the 5'-GNN-3' family of DNA sequences by zinc finger domains. *J Mol Biol*, 303:489–502, 2000.

[182] Birgit Dreier, Roger R Beerli, David J Segal, Jessica D Flippin, and Carlos F Barbas III. Development of zinc finger domains for recognition of the 5'-ANN-3' family of DNA sequences and their use in the construction of artificial transcription factors. *J Biol Chem*, 276(31):29466–78, 2001.

[183] Qiang Liu, ZhenQin Xia, Xiaohong Zhong, and Casey C Case. Validated zinc finger protein designs for all 16 GNN DNA triplet targets. *J Biol Chem*, 277:3850–6, 2002.

[184] Birgit Dreier, Roberta P Fuller, David J Segal, Caren Lund, Pilar Blancafort, Adrian Huber, Beate Koksch, and Carlos F Barbas III. Development of zinc finger domains for recognition of the 5'-CNN-3' family DNA sequences and their use in the construction of artificial transcription factors. *J Biol Chem*, 280(42):35588–97, 2005.

[185] David A Wright, Stacey Thibodeau-Beganny, Jeffry D Sander, Ronnie J Winfrey, Andrew S Hirsh, Magdalena Eichtinger, Fengli Fu, Matthew H Porteus,

Drena Dobbs, Daniel F Voytas, and J Keith Joung. Standardized reagents and protocols for engineering zinc finger nucleases by modular assembly. *Nat Protoc*, 1(3):1637–52, 2006.

[186] Jeffrey G Mandell and Carlos F 3rd Barbas. Zinc finger tools: custom DNA-binding domains for transcription factors and nucleases. *Nucleic Acids Res*, 34(Web Server issue):W516–23, 2006.

[187] Jeffry D Sander, Peter Zaback, J Keith Joung, Daniel F Voytas, and Drena Dobbs. Zinc finger targeter (ZiFiT): an engineered zinc finger/target site design tool. *Nucleic Acids Res*, 35(Web Server issue):W599–605, 2007.

[188] Naima Bouhouche, Michael Syvanen, and Clarence I Kado. The origin of prokaryotic C2H2 zinc finger regulators. *Trends Microbiol*, 8(2):77–81, 2000.

[189] J Keith Joung, Elizabeth I Ramm, and Carl O Pabo. A bacterial two-hybrid selection system for studying protein-DNA and protein-protein interactions. *Proc Natl Acad Sci USA*, 97(13):7382–7, 2000.

[190] Xiangdong Meng, Michael H Brodsky, and Scot A Wolfe. A bacterial one-hybrid system for determining the DNA-binding specificity of transcription factors. *Nat Biotechnol*, 23(8):988–94, 2005.

[191] Xiangdong Meng and Scot A Wolfe. Identifying DNA sequences recognized by a transcription factor using a bacterial one-hybrid system. *Nat Protoc*, 1(1):30–45, 2006.

[192] Xiangdong Meng, Stacey Thibodeau-Beganny, Tao Jiang, J Keith Joung, and Scot A Wolfe. Profiling the DNA-binding specificities of engineered Cys2His2 zinc finger domains using a rapid cell-based method. *Nucleic Acids Res*, 35(11):e81, 2007.

[193] William H. Landschulz, Peter F. Johnson, and Steven L. McKnight. The leucine zipper: A hypothetical structure common to a new class of DNA binding proteins. *Science*, 240(4860):1759–64, June 1988.

[194] Charles Vinson, Max Myakishev, Asha Acharya, Alain A Mir, Jonathan R Moll, and Maria Bonovich. Classification of human B-ZIP proteins based on dimerization properties. *Mol Cell Biol*, 22(18):6321–35, 2002.

[195] T Kouzarides and E Ziff. The role of the leucine zipper in the fos-jun interaction. *Nature*, 336(6200):646–51, 1988.

[196] E K O'Shea, R Rutkowski, and P S Kim. Evidence that the leucine zipper is a coiled coil. *Science*, 243(4890):538–42, 1989.

[197] Erin K O'Shea, Juli D Klemm, Peter S Kim, and Tom Alber. X-ray structure of the GCN4 leucine zipper, a two-stranded, parallel coiled coil. *Science*, 254(5031):539–44, 1991.

[198] Erin K O'Shea, Rheba Rutkowski, Walter F Stafford III, and Peter S Kim. Preferential heterodimer formation by isolated leucine zippers from Fos and Jun. *Science*, 245(4918):646–8, 1989.

[199] Christine Pernelle, François F Clerc, Christine Dureuil, Laurent Bracco, and Bruno Tocque. An efficient screening assay for the rapid and precise determination of affinities between leucine zipper domains. *Biochemistry*, 32(43):11682–7, 1993.

[200] Andreas D Baxevanis and Charles R Vinson. Interactions of coiled coils in transcription factors: where is the specificity? *Curr Opin Genet Dev*, 3(2):278–85, 1993.

[201] Joel L Pomerantz, Scot A Wolfe, and Carl O Pabo. Structure-based design of a dimeric zinc finger protein. *Biochemistry*, 37:965–70, 1998.

[202] Scot A Wolfe, Elizabeth I Ramm, and Carl O Pabo. Combining structure-based design with phage display to create new Cys_2His_2 zinc finger dimers. *Structure*, 8(7):739–50, 2000.

[203] Scot A Wolfe, Robert A Grant, and Carl O Pabo. Structure of a designed dimeric zinc finger protein bound to DNA. *Biochemistry*, 42(46):13401–9, 2003.

[204] Mark M Garner and Arnold Revzin. A gel electrophoresis method for quantifying the binding of proteins to specific DNA regions: application to components of the *Escherichia coli* lactose operon regulatory system. *Nucleic Acids Res*, 9(13):3047–60, 1981.

[205] Catherine H. Schein. Production of soluble recombinant proteins in bacteria. *Nat Biotech*, 7(11):1141–1149, 1989.

[206] Jun Kuwahara and Joseph E Coleman. Role of the zinc(II) ions in the structure of the three-finger DNA binding domain of the Sp1 transcription factor. *Biochemistry*, 29(37):8627–31, 1990.

[207] Deepali Sachdev and John M Chirgwin. Order of fusions between bacterial and mammalian proteins can determine solubility in *Escherichia coli*. *Biochem Biophys Res Commun*, 244(3):933–7, 1998.

[208] Rachel B Kapust and David S Waugh. *Escherichia coli* maltose-binding protein is uncommonly effective at promoting the solubility of polypeptides to which it is fused. *Protein Sci*, 8(8):1668–74, 1999.

[209] Pascale Duplay, Hugues Bedouelle, Audree Fowler, Irving Zabin, William Saurin, and Maurice Hofnung. Sequences of the *malE* gene and of its product, the maltose-binding protein of *Escherichia coli* K12. *J Biol Chem*, 259(16):10606–13, 1984.

[210] Odile K Kellermann and Thomas Ferenci. Maltose-binding protein from *Escherichia coli*. *Methods Enzymol*, 90 Pt E(0076-6879 (Print)):459–63, 1982.

[211] Paul Riggs. Expression and purification of recombinant proteins by fusion to maltose-binding protein. *Mol Biotechnol*, 15(1):51–63, 2000.

[212] A Ullmann and D Perrin. *The lactose operon*, chapter VII. Cold Spring Harbor Laboratory, 1970.

[213] Brendan P Cormack, Raphael H Valdivia, and Stanley Falkow. FACS-optimized mutants of the green fluorescent protein (GFP). *Gene*, 173(1 Spec No):33–8, 1996.

[214] Alan D Frankel, Jeremy M Berg, and Carl O Pabo. Metal-dependent folding of a single zinc finger from transcription factor IIIA. *Proc Natl Acad Sci USA*, 84(14):4841–5, 1987.

[215] Grace Parraga, Suzanna J Horvath, Arri Eisen, Wayne E Taylor, Leroy Hood, Elton T Young, and Rachel E Klevit. Zinc-dependent structure of a single-finger domain of yeast ADR1. *Science*, 241(4872):1489–92, 1988.

[216] Hal Blumberg, Arri Eisen, Andrzej Sledziewski, Dorit Bader, and Elton T Young. Two zinc fingers of a yeast regulatory protein shown by genetic evidence to be essential for its function. *Nature*, 328(6129):443–5, 1987.

[217] Takashi Miura, Tamami Satoh, and Hideo Takeuchi. Role of metal-ligand coordination in the folding pathway of zinc finger peptides. *Biochim Biophys Acta*, 1384(1):171–9, 1998.

[218] Damian J McColl, Cynthia D Honchell, and Alan D Frankel. Structure-based design of an RNA-binding zinc finger. *Proc Natl Acad Sci USA*, 96(17):9521–6, 1999.

[219] Robert S Cox III, Michael G Surette, and Michael B Elowitz. Programming gene expression with combinatorial promoters. *Mol Syst Biol*, 3(1744-4292 (Electronic)):145, 2007.

[220] Scot A Wolfe, Lena Nekludova, and Carl O Pabo. DNA recognition by Cys_2His_2 zinc finger proteins. *Annu Rev Biophys Biomol Struct*, 29:183–212, 2000.

[221] Reshma P Shetty, Drew Endy, and Thomas F Knight Jr. Engineering BioBrick vectors from BioBrick parts. *J Biol Eng*, 2(1):5, 2008.

[222] Ron Weiss. *Cellular Computation and Communications using Engineered Genetic Regulatory Networks*. PhD thesis, Massachusetts Institute of Technology, Department of Electrical Engineering and Computer Science;, 2001.

[223] Dmitri Rebatchouk, Nikolai Daraselia, and Jonathon O Narita. NOMAD: a versatile strategy for *in vitro* DNA manipulation applied to promoter analysis and vector design. *Proc Natl Acad Sci USA*, 93(20):10891–6, 1996.

[224] Adam Arkin and Drew Endy. A standard parts list for biological circuitry. Technical report, DARPA white paper, 1999. Available from: `http://hdl.handle.net/1721.1/29794`.

[225] Dan Ferber. Microbes made to order. *Science*, 303(5655):158–161, January 2004.

[226] Erika Check. Synthetic biology: Designs on life. *Nature*, 438(7067):417–418, 2005.

[227] Stanley N Cohen, Annie C Chang, Herbert W Boyer, and Robert B Helling. Construction of biologically functional bacterial plasmids *in vitro*. *Proc Natl Acad Sci USA*, 70(11):3240–4, 1973.

[228] Annie C Chang and Stanley N Cohen. Genome construction between bacterial species *in vitro*: replication and expression of *Staphylococcus* plasmid genes in *Escherichia coli*. *Proc Natl Acad Sci USA*, 71(4):1030–4, 1974.

[229] W Klingmuller. [Bacterial resistance factors as vectors for gene manipulation and gene therapy (author's transl)]. *MMW Munch Med Wochenschr*, 117(24):1051–60, 1975.

[230] S I Alikhanian, O I Khlebalina, A I Stepanov, M Iu Beburov, and N V Kalinina. [Preparation of functioning recombinant (hybrid) DNA molecules *in vitro* (genetic engineering experiments)]. *Genetika*, 11(11):34–45, 1975.

[231] Jeffrey Vieira and Joachim Messing. The pUC plasmids, an M13mp7-derived system for insertion mutagenesis and sequencing with synthetic universal primers. *Gene*, 19(3):259–68, 1982.

[232] Jan Norrander, Tomas Kempe, and Joachim Messing. Construction of improved M13 vectors using oligodeoxynucleotide-directed mutagenesis. *Gene*, 26(1):101–6, 1983.

[233] Celeste Yanisch-Perron, Jeffrey Vieira, and Joachim Messing. Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. *Gene*, 33(1):103–19, 1985.

[234] Alain Rambach and Pierre Tiollais. Bacteriophage λ having *Eco*RI endonuclease sites only in the nonessential region of the genome. *Proc Natl Acad Sci USA*, 71(10):3927–30, 1974.

[235] Vickers Hershfield, Herbert W Boyer, Charles Yanofsky, Michael A Lovett, and Donald R Helinski. Plasmid ColEl as a molecular vehicle for cloning and amplification of DNA. *Proc Natl Acad Sci USA*, 71(9):3455–9, 1974.

[236] Frederick R Blattner, Bill G Williams, Ann E Blechl, Katherine Denniston-Thompson, Harvey E Faber, Lesley-Anne Furlong, David J Grunwald, Delight O Kiefer, David D Moore, James W Schumm, Edward L Sheldon, and

Oliver Smithies. Charon phages: safer derivatives of bacteriophage lambda for DNA cloning. *Science*, 196(4286):161–9, 1977.

[237] F Bolivar, R L Rodriguez, P J Greene, M C Betlach, H L Heyneker, and H W Boyer. Construction and characterization of new cloning vehicles. II. a multipurpose cloning system. *Gene*, 2(2):95–113, 1977.

[238] Alan H Rosenberg, Barbara N Lade, Dao-shan Chui, Shu-Wha Lin, John J Dunn, and F William Studier. Vectors for selective expression of cloned DNAs by T7 RNA polymerase. *Gene*, 56(1):125–35, 1987.

[239] Wlodek Mandecki, Mark A Hayden, Mary A Shallcross, and Elizabeth Stotland. A totally synthetic plasmid for general cloning, gene expression and mutagenesis in *Escherichia coli*. *Gene*, 94(1):103–7, 1990.

[240] James L Hartley, Gary F Temple, and Michael A Brasch. DNA cloning using *in vitro* site-specific recombination. *Genome Res*, 10(11):1788–95, 2000.

[241] Michael A Brasch, James L Hartley, and Marc Vidal. ORFeome cloning and systems biology: standardized mass production of the parts from the parts-list. *Genome Res*, 14(10B):2001–9, 2004.

[242] Albertha J Walhout, Gary F Temple, Michael A Brasch, James L Hartley, Monique A Lorson, Sander van den Heuvel, and Marc Vidal. GATEWAY recombinational cloning: application to the cloning of large numbers of open reading frames or ORFeomes. *Methods Enzymol*, 328(0076-6879 (Print)):575–92, 2000.

[243] Jérôme Reboul, Philippe Vaglio, Jean-François Rual, Philippe Lamesch, Monica Martinez, Christopher M Armstrong, Siming Li, Laurent Jacotot, Nicolas Bertin, Rekin's Janky, Troy Moore, James R Jr Hudson, James L Hartley, Michael A Brasch, Jean Vandenhaute, Simon Boulton, Gregory A Endress, Sarah Jenna, Eric Chevet, Vasilis Papasotiropoulos, Peter P Tolias, Jason Ptacek, Mike Snyder, Raymond Huang, Mark R Chance, Hongmei Lee, Lynn Doucette-Stamm, David E Hill, and Marc Vidal. *C. elegans* ORFeome version 1.1: experimental verification of the genome annotation and resource for proteome-scale protein expression. *Nat Genet*, 34(1):35–41, 2003.

[244] Jean-François Rual, Tomoko Hirozane-Kishikawa, Tong Hao, Nicolas Bertin, Siming Li, Amélie Dricot, Ning Li, Jennifer Rosenberg, Philippe Lamesch, Pierre-Olivier Vidalain, Tracey R Clingingsmith, James L Hartley, Dominic Esposito, David Cheo, Troy Moore, Blake Simmons, Reynaldo Sequerra, Stephanie Bosak, Lynn Doucette-Stamm, Christian Le Peuch, Jean Vandenhaute, Michael E Cusick, Joanna S Albala, David E Hill, and Marc Vidal. Human ORFeome version 1.1: a platform for reverse proteomics. *Genome Res*, 14(10B):2128–35, 2004.

[245] David L Cheo, Steven A Titus, Devon R N Byrd, James L Hartley, Gary F Temple, and Michael A Brasch. Concerted assembly and cloning of multiple DNA segments using *in vitro* site-specific recombination: functional analysis of multi-segment expression clones. *Genome Res*, 14(10B):2111–20, 2004.

[246] Qi-Jun Chen, Hai-Meng Zhou, Jia Chen, and Xue-Chen Wang. A Gateway-based platform for multigene plant transformation. *Plant Mol Biol*, 62(6):927–36, 2006.

[247] E S Anderson. The ecology of transferable drug resistance in the enterobacteria. *Annu Rev Microbiol*, 22(0066-4227 (Print)):131–80, 1968.

[248] Joseph Sambrook and David W Russell. *Molecular Cloning: A Laboratory Manual*, volume 1-3. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 3rd edition, 2001.

[249] Philippe Bernard, Philippe Gabant, El Mustapha Bahassi, and Martine Couturier. Positive-selection vectors using the F plasmid *ccdB* killer gene. *Gene*, 148:71–4, 1994.

[250] Philippe Bernard. New *ccdB* positive-selection cloning vectors with kanamycin or chloramphenicol selectable markers. *Gene*, 162:159–60, 1995.

[251] P Bernard. Positive selection of recombinant DNA by CcdB. *Biotechniques*, 21:320–3, 1996.

[252] Felipe Cabello, Kenneth Timmis, and Stanley N Cohen. Replication control in a composite plasmid constructed by *in vitro* linkage of two distinct replicons. *Nature*, 259(0028-0836 (Print)):285–90, 1976.

[253] Panayiotis A Ioannou, Chris T Amemiya, Jeffrey Garnes, Peter M Kroisel, Hiroaki Shizuya, Chira Chen, Mark A Batzer, and Pieter J de Jong. A new bacteriophage P1-derived vector for the propagation of large human DNA fragments. *Nat Genet*, 6(1):84–9, 1994.

[254] Reiner Gentz, Annette Langner, Annie C Chang, Stanley N Cohen, and Hermann Bujard. Cloning and analysis of strong promoters is made possible by the downstream placement of a RNA termination signal. *Proc Natl Acad Sci USA*, 78(8):4936–40, 1981.

[255] Dietrich Stueber and Hermann Bujard. Transcription from efficient promoters can interfere with plasmid replication and diminish expression of plasmid specified genes. *EMBO J*, 1(11):1399–404, 1982.

[256] Richard P Bowater, Adam Jaworski, Jacquelynn E Larson, Pawel Parniewski, and Robert D Wells. Transcription increases the deletion frequency of long CTG.CAG triplet repeats from plasmids in *Escherichia coli*. *Nucleic Acids Res*, 25(14):2861–8, 1997.

[257] Ronald Godiska, Melodee Patterson, Tom Schoenfeld, and David A Mead. Beyond pUC: Vectors for cloning unstable DNA. In Jan Kieleczawa, editor, *DNA Sequencing: Optimizing the Process and Analysis*, pages 55–76. Jones and Bartlett Publishers, Inc., Sudbury, Massachusetts, 1st edition, 2005.

[258] Philippe Bernard and Martine Couturier. Cell killing by the F plasmid CcdB protein involves poisoning of DNA-topoisomerase II complexes. *J Mol Biol*, 226(3):735–45, 1992.

[259] Takeyoshi Miki, Jeong A Park, Kiminori Nagao, Nobuhiro Murayama, and Tadao Horiuchi. Control of segregation of chromosomal DNA by sex factor F in *Escherichia coli*. mutants of DNA gyrase subunit A suppress letD (ccdB) product growth inhibition. *J Mol Biol*, 225(1):39–52, 1992.

[260] T Watanabe and T Fukasawa. "Resistance transfer factor", an episome in *Enterobacteriaceae*. *Biochem Biophys Res Commun*, 3:660–665, 1960.

[261] T Watanabe and T Fukasawa. Episome-mediated transfer of drug resistance in *Enterobacteriaceae*. I. transfer of resistance factors by conjugation. *J Bacteriol*, 81(0021-9193 (Print)):669–78, 1961.

[262] T Watanabe. Infective heredity of multiple drug resistance in bacteria. *Bacteriol Rev*, 27(0005-3678 (Print)):87–115, 1963.

[263] E S Anderson. Origin of transferable drug-resistance factors in the Enterobacteriaceae. *Br Med J*, 2(5473):1289–91, 1965.

[264] Stanley N Cohen and Annie C Chang. Revised interpretation of the origin of the pSC101 plasmid. *J Bacteriol*, 132(2):734–7, 1977.

[265] Annie C Chang and Stanley N Cohen. Construction and characterization of amplifiable multicopy DNA cloning vehicles derived from the P15A cryptic miniplasmid. *J Bacteriol*, 134(3):1141–56, 1978.

[266] Reshma P Shetty and Randy Rettberg. Synthetic Biology:BioBricks/3A assembly, March 2003. Available from: `http://openwetware.org/index.php?title=Synthetic_Biology:BioBricks/3A_assembly&oldid=78119` [cited December 13, 2007].

[267] Austin Che. BioBricks++ assembly scheme. Technical report, MIT Synthetic Biology Working Group Technical Reports, February 2004. Available from: `http://hdl.handle.net/1721.1/39832`.

[268] Marlene Belfort and Richard J Roberts. Homing endonucleases: keeping the house in order. *Nucleic Acids Res*, 25(17):3379–88, 1997.

[269] Jurate Bitinaite, Michelle Rubino, Kamini Hingorani Varma, Ira Schildkraut, Romualdas Vaisvila, and Rita Vaiskunaite. USER™ friendly DNA engineering and cloning method by uracil excision. *Nucleic Acids Res*, 35(6):1992–2002, 2007.

[270] Sarah M Richardson, Sarah J Wheelan, Robert M Yarrington, and Jef D Boeke. Genedesign: rapid, automated design of multikilobase synthetic genes. *Genome Res*, 16(4):550–6, 2006.

[271] Austin Che and Reshma P Shetty. Knight:Preparing electrocompetent cells, May 8, 2007. Available from: `http://openwetware.org/index.php?title=Knight:Preparing_electrocompetent_cells&oldid=115509` [cited December 22, 2007].

[272] Reshma P Shetty. Knight:Electroporation, May 8, 2007. Available from: `http://openwetware.org/index.php?title=Knight:Electroporation&oldid=115508` [cited December 22, 2007].

[273] Malcolm J Casadaban and Stanley N Cohen. Analysis of gene control signals by DNA fusion and cloning in *Escherichia coli*. *J Mol Biol*, 138(2):179–207, 1980.

[274] Seth G Grant, Joel Jessee, Fredric R Bloom, and Douglas Hanahan. Differential plasmid rescue from transgenic mouse DNAs into *Escherichia coli* methylation-restriction mutants. *Proc Natl Acad Sci USA*, 87(12):4645–9, 1990.

[275] Thomas F Knight Jr. TOP10 chemically competent cells, August 31, 2007. Available from: `http://openwetware.org/index.php?title=TOP10_chemically_competent_cells&oldid=185656` [cited February 11, 2008].

[276] Stacey Thibodeau-Beganny and J Keith Joung. Engineering Cys2His2 zinc finger domains using a bacterial cell-based two-hybrid selection system. *Methods Mol Biol*, 408(1064-3745 (Print)):317–34, 2007.

[277] James Newcomb, Robert Carlson, and Stephen C Aldrich. Genome synthesis and design futures: Implications for the u.s. economy. Technical report, Bio Economic Research Associates, 2007.

[278] Douglas Hanahan, Joel Jessee, and Fredric R Bloom. Plasmid transformation of *Escherichia coli* and other bacteria. *Methods Enzymol*, 204(0076-6879 (Print)):63–113, 1991.

[279] Reshma P Shetty. Knight:NuPAGE electrophoresis/Hybrid staining, 2007. Available from: `http://openwetware.org/index.php?title=Knight:NuPAGE_electrophoresis/Hybrid_staining&oldid=144437` [cited April 22, 2008].

[280] X Zhang and H Bremer. Control of the *Escherichia coli rrnB* P1 promoter strength by ppGpp. *J Biol Chem*, 270(19):11181–9, 1995.

[281] Kevin L Griffith and Richard E Wolf Jr. Measuring β-galactosidase activity in bacteria: cell growth, permeabilization, and enzyme assays in 96-well arrays. *Biochem Biophys Res Commun*, 290(1):397–402, 2002.

[282] Sean Moore. Beta-galactosidase assay (a better miller). Available from: http://openwetware.org/index.php?title=Beta-Galactosidase_Assay_%28A_better_Miller%29&oldid=158444 [cited April 17, 2008].

[283] S R McCorkle. pSCANS vector. Available from: http://www.genome.bnl.gov/Vectors/pscans.php [cited April 6, 2008].

[284] Leon Y Chan, Sriram Kosuri, and Drew Endy. Refactoring bacteriophage T7. *Mol Syst Biol*, 1(1744-4292 (Electronic)):2005.0018, 2005.

[285] Karmella A Haynes, L Broderick, Marian, Adam D Brown, Trevor L Butner, James O Dickson, Lance Harden, Lane H Heard, Eric L Jessen, Kelly J Malloy, Brad J Ogden, Sabriya Rosemond, Samantha Simpson, Erin Zwack, Malcolm Campbell, Todd T Eckdahl, Laurie J Heyer, and Jeffrey L Poet. Engineering bacteria to solve the burnt pancake problem. *J Biol Eng*, 2(1):8, 2008.

[286] Jason R Kelly. Endy:Screening plasmid/v1.0. Available from: http://openwetware.org/index.php?title=Endy:Screening_plasmid/v1.0&oldid=117856 [cited April 6, 2008].