Fine-tuning Engineered Gene Regulatory Networks Expressed in *Escherichia coli* using Hypervariable Simple Sequence Repeats

Robert G. Egbert

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington

2012

Reading Committee:

Eric Klavins, Chair

Georg Seelig

Paul Wiggins

Program Authorized to Offer Degree:
Electrical Engineering

University of Washington

Abstract

Fine-tuning Engineered Gene Regulatory Networks Expressed in *Escherichia coli* using Hypervariable Simple Sequence Repeats

Robert G. Egbert

Chair of the Supervisory Committee:
Associate Professor Eric Klavins
Department of Electrical Engineering

Synthetic biology aims to borrow from the vast diversity of living systems shaped by evolutionary processes to create synthetic biological systems with comparable functional complexity to natural systems that meet pressing needs in health, energy, and the environment. Construction of these systems is both aided by the richness of this evolutionary toolkit and hindered by its complexity. This dissertation presents a methodology to fine-tune engineered gene networks in *Escherichia coli* that accelerates the realization of functionally complex behaviors using focused variation to thoroughly sample gene expression levels for a target network. This tuning approach exploits errors that occur during replication of tandem DNA repeats to predictably vary gene expression. Using this approach, we have generated DNA libraries that vary in repeat length to predictably tune the expression of a fluorescent protein over a large range with high resolution. We have demonstrated the utility of the approach by tuning the expression of two transcription factors to optimize three functional behaviors of a bistable genetic switch. Finally, to extend the reach of the approach, we have investigated methods to control mutation rates of the repeats to rapidly optimize gene networks in vivo via directed evolution. This tuning methodology is extensible to biological mechanisms that affect other gene network parameters, is compatible with computational strategies for tuning networks, and should advance the field of synthetic biology by enabling timely realization of functionally complex behaviors in cells.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I thank the following group of intelligent and inspiring people for their advice and assistance throughout my graduate school career: Eric Klavins for, among countless other things, the opportunity and freedom to explore and create in the lab, which gave me confidence as an independent researcher; Josh Bishop for teaching me to clone and for having insightful comments on any topic of conversation; Georg Seelig for regular advice and perspective about doing high-quality work; Ben Kerr for an introduction to big questions in evolutionary biology; Paul Wiggins for encouraging me to think critically about choosing my engineering substrates; Jennifer Nemhauser for treating me as a peer; David Thorsley for guidance on birth-death processes; fellow iGEM advisors Ingrid Swanson Pultz and Justin Siegel for their examples as dedicated student mentors and creative collaborators; Kyle Havens for reminding me about controls; Chris Takahashi for building amazing continuous culture contraptions; Angelina Smith for cloning support on the first rbSSR constructs; Rahul Brito for his work on building evolvable strains; iGEM students Michael Brasino, Alicia Wong, and Rashmi Ravichandran for helping me learn to lead a research team (and to play with magnetosome genes!); and Ted Chavkin and David Zong for their work on library generation for the self-destructive altruism project.

Above all, I thank my dear wife, Sara Jean: without her unwavering support despite many late nights and weekends in the lab, and boundless patience and energy nurturing Jasper, Gideon, and our third child *in utero*, none of this work was possible.

Funding for this work was provided by a National Science Foundation Graduate Research Fellowship and NSF grants 0832824 (Molecular Programming Project) and 1002220 (Estimation & Observation of Stochastic Biochemical Networks).

DEDICATION

To my mother, Yvonne.

Despite only 14 years together on earth,

your love and encouragment continue to drive me forward.

To my father, Lawrence.

You are stable as an oak; constant as a rock; dependable as the sun.

Chapter 1

INTRODUCTION

1.1 Engineering living systems, borrowing from evolution

Living systems have evolved over billions of years to adapt and specialize in dynamically changing environments to ensure survival [1]. Through these evolutionary processes, genes that code for useful proteins and gene networks that encode useful behaviors have been selected for to aid organisms, especially bacteria (the focus of work in this dissertation), in adapting to nearly any conceivable environment. Two remarkable examples are bacteria that survive kilometers beneath the earth's surface, utilizing radiation as an energy source instead of sunlight [2]; and deep sea microbes that subsist on organic matter deposited in the age of the dinosaurs [3, 4].

Synthetic biology is an engineering discipline that aims to design and implement synthetic biological systems at time scales of months to years with complex behaviors comparable to natural systems that have evolved over millions to billions of years [5, 6]. These synthetic systems are engineered to address problems in health, energy and the environment, focusing efforts on identifying, extracting, and properly interconnecting collections of biological parts to realize increasingly complex behaviors in living systems, from *E. coli* [7, 8, 9, 10] to mammalian cell lines [11, 12, 13, 14].

Evolution's contribution to synthetic biology is a double-edged sword. Its positive contribution is a vast toolkit of functional proteins [15] and gene networks [16] that can be used to build model systems [7, 17] and studied [18, 19] (or reverse engineered, depending on one's perspective) to inform biological design principles. Access to this toolkit accelerates the realization of functional behaviors. From an engineering perspective , one major negative contribution is the broad interconnection among evolved proteins and networks [20] that often confound both the understanding of existing networks and the construction of novel ones. This complexity hinders progress in the field.

1.2 Challenges for scaling up functional complexity

In more specific terms, engineering reliable and predictable synthetic gene networks presents unique challenges because genetic parts such as promoters, ribosome binding sites, and protein coding regions often behave unexpectedly when used in novel designs. Biochemical noise [21, 22, 23], metabolic load [24], poorly characterized interactions with the host [5, 20], and general uncertainty about the detailed functionality of parts conspire to limit the complexity of synthetic gene networks to a small number of interacting genes [25, 6].

As a result, a complex gene network that has been predicted analytically to perform well may in fact perform poorly, if it works at all, when ultimately implemented in cells. Furthermore, even if a gene network can be made to perform well in a particular strain and environment, there is no guarantee that the same network will perform well if ported to a new strain or environment. More generally, synthetic networks often operate in substantially different parameter regimes than expected during the design process and must, therefore, be *tuned* [22, 26, 27, 28, 29] before they function properly, and even *retuned* when used in new environments or with different hosts.

To give a brief example, one complex behavior of interest to synthetic biologists is the biomineralization of magnetite crystals from soluble iron in prokaryotic organelles known as magnetosomes[30, 16]. Magnetotactic bacteria use chains of vesicles, each vesicle containing one of these nanoscale magnets, to sense the earth's magnetic field, simplifying the search for optimal growth environments. Proper biosynthesis of these crystals requires the coordinated expression of at least ten proteins responsible for the serial processes of forming empty vesicles, biomineralization of the magnetite crystals, and aligning the vesicles into chains.

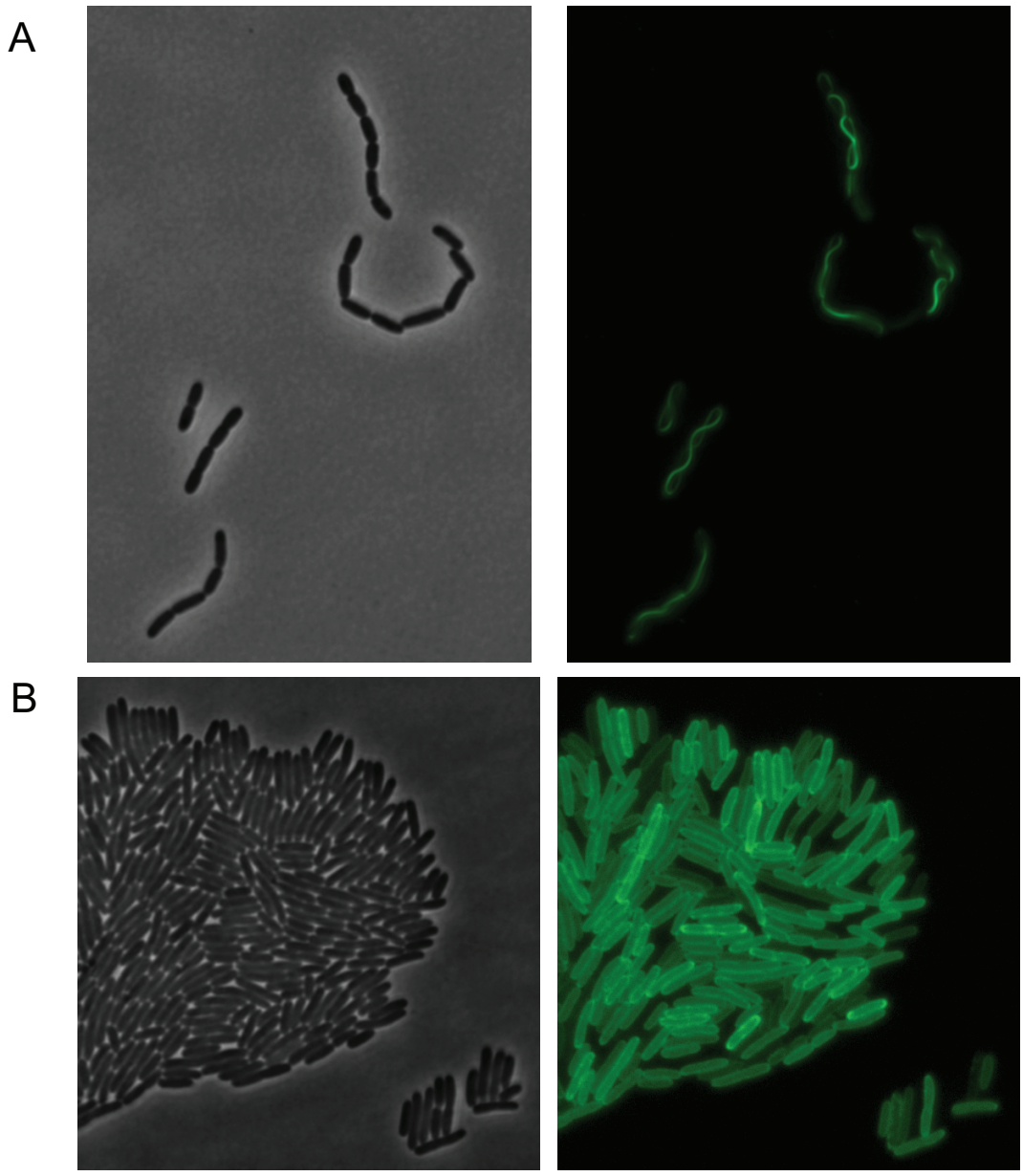

Figure 1.1: Characterization of essential genes for magnetosome formation. Magnetosome proteins MamK, which polymerizes into long actin-like filaments (A), and membrane-bound MamI (B) were translationally fused to superfolder GFP to visualize protein localization properties. This work was performed in the Klavins lab by undergraduate students working on the Magnetosome Toolkit, an effort to forward engineer magnetosome-forming organelles in *E. coli*, as members of the 2011 University of Washington iGEM – International Genetically Engineered Machines – team.

Magnetite crystals have several potential biotechnological applications in medicine and industry, and producing these magnetite crystals in an industrial strain (the minimum doubling time of the native organisms is over 10 hours) could enable magnetite biofabrication at large scales. However, due to the limitations described above, the functional composition of gene networks is generally limited to characterizing simple systems, often one gene at a time (see Fig. 1.1). While the necessary and sufficient set of genes is not yet known for magnetosome formation, this is a prime example of the type of complex behaviors synthetic biologists would like to borrow from natural system to reoptimize in synthetic contexts.

1.3 Overview of thesis contributions

This dissertation presents a novel approach to fine-tune gene networks in *Escherichia coli* that accelerates the implementation of functionally complex cell behaviors by thoroughly exploring gene expression parameters for target genes of the network. We explore the parameter space by focusing genetic variation to the spacer region of the ribosome binding site using hypervariable simple sequence repeats; we call these sequence motifs rbSSRs. We introduce multiple methods to generate rbSSR expression libraries that vary in repeat number, using them to evaluate this tuning approach against a set of criteria for effective tuning practices.

We found that these libraries incrementally and predictably sample gene expression levels over a 1,000-fold range, and that the range of expression can be expanded by coarsely tuning promoter strength. We demonstrate the utility of the approach by fine-tuning three functional behaviors of a bistable switch built with dual rbSSRs, and illustrate the need for tuning by showing that the genomic context of a host strain can have profound effects on the switch's behavior.

We also show that rbSSR sequences are stable over more than 200 generations, but that destabilization of the repeats in a mutator strain focuses mutations to the spacer region, which could be used to tune and select for optimized gene networks in vivo. These results are built upon by the construction of an inducible mutator strain to enable alternating periods of hypervariation and competition in selective environments.

Finally, we briefly describe two applications of this tuning methodology and describe

some possible extensions of it. The first application is a self-destructive altruism circuit, based on the rbSSR bistable switch, to enable *E. coli* to grow on cellulose as the sole carbon source. The second application is to direct the evolution of the lactose utilization network native to *E. coli* using rbSSR spacers. The dissertation concludes with a discussion of possible extensions of this work to tune other parameters of gene networks such as transcription factor binding efficiency or intron splicing in higher organisms. These results are broadly applicable to rapidly engineering functional gene circuits and scaling up circuit complexity by enabling the creation of expression libraries that thoroughly and predictably sample the parameter space of a gene network. A major portion of this work has been accepted for publication as "Fine-tuning gene networks using simple sequence repeats" in the Proceedings of the National Academy of Sciences of the United States of America [31].

Chapter 2

TUNING GENE NETWORKS: SAMPLING PARAMETER SPACE

2.1 Constructing synthetic gene networks

Every engineering discipline needs design principles that relate theory to implementation to manage complexity and enable timely realization of desired system behaviors. Synthetic biology is no different [25]. Engineering biological systems requires an understanding of the underlying biophysical mechanisms that drive gene expression, such as transcription and translation [32], as well as of the interconnections, or network topologies, that have been selected for through evolutionary processes [33]. In its current state, engineering biological systems also requires long hours of dedication in the laboratory environment either constructing DNA-encoded biochemical programs or culturing and interrogating the behaviors of the cells running those programs.

To design a gene network and realize its expected behaviors, it is helpful to consider the following five elements: components; connectivity; copy number; code; and chassis. The first three elements – components, connectivity and copy number – primarily concern the theoretical or architectural design of the network that can be reasoned about through mathematical models and simulation. The last two – code and chassis – primarily concern the ultimate implementation of the gene network, which are strongly affected by the environment in which they are executed. The implementation process for many gene networks often requires extensive trial and error due to the underlying complexity of biology and the rigidity of a serial design process.

Components. Structural proteins, enzymes, transcription factors, and fluorescent reporters all represent components of gene networks. Components are generally genes but can also be regulatory RNAs [34]. In some cases specific components are required for a designed system, but they can often be chosen from a class of components with similar properties.

Connectivity. The regulation and timing of gene expression via transcription fac-

tors and regulatory RNAs, the interconnection among circuit modules and with host strain networks, as well as intercellular signaling could all be considered aspects of engineering connectivity [35]. The design of connectivity is generally done in conjunction with component selection.

Copy number. The expression level of each component in a network when actively produced can profoundly affect the network's behavior. The sensitivity of a gene network's behavior to the expression level of each component is important to consider in the design process. For instance, regulatory proteins such as transcription factors are often needed only in small quantities, while the final product of a of an engineered metabolic network is generally optimized for maximal expression. For fluorescent reporters, copy number can also be extremely important, as different measurement modalities – microscope, cytometer, plate reader – have varying sensitivities. Overexpression of proteins can stress cells by reducing growth rate which can increase the likelihood of mutations that can disable the circuit or the measurement of its activity [36]. An approach to fine-tune gene expression levels is the major theme of this thesis.

Code. The corrolary to binary code for computer programs is the quaternary DNA code maintained by all living systems as the primary information carrier. All components of a gene network must be encoded in DNA. Information about connectivity is encoded by components and binding sites for regulatory proteins. Copy number is also programmed via DNA as gene expression control sequences. Many of these functional sequences can be equivalently coded by multiple DNA sequences. Examples include multiple codons for most amino acids, a range of sequence motifs that recruit transcription and translation machinery (including operator sites for transcription factors), or seemingly arbitrary sequences used as spacers between critical sequences (see Ref. [37, 38, 39]). In some cases the sequence choice is arbitrary; in other cases the identity of a single base is critical for certain properties or functions.

Chassis. The host strain and its native gene networks have unpredictable effects on synthetic gene networks [20]. Often, some aspects of system design specify which host should be used, especially if the gene network is designed to study, or work with, a particular native network. In other cases, gene networks are prototyped in one strain that is easy to

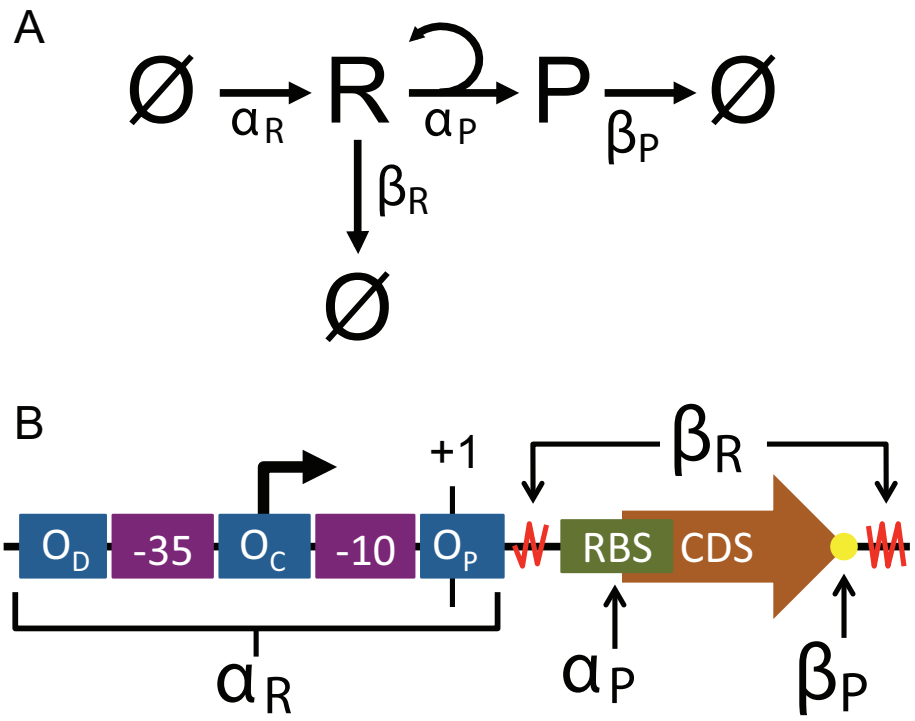

Figure 2.1: A model of gene expression and its DNA implementation. (A) Simple model for transcription rate (α_R) of mRNA (R) with associated degradation rate (β_R) and translation rate (α_P) of protein (P) with associated degradation rate (β_P). (B) Cartoon of tunable elements for a single gene. Each element implemented in DNA or RNA primarily affects one parameter from the model in (A), as shown. The black curved arrow represents the promoter with the following elements: blue boxes for the distal, core, and proximal operator sites [35]; purple boxes for the docking site of RNA polymerase, and +1 for the transcriptional start site. The remaining symbols represent functional regions of the messenger RNA (mRNA): red zigzags for 5' and 3' untranslated regions (UTR); green box for the ribosome binding site (RBS), which is also technically part of the 5' UTR; brown box for coding sequence (CDS); and yellow circle for active degradation tag[40].

manipulate, then the same circuit is transferred, or ported, to another strain. This process often has undesired consequences on gene network performance due to subtle differences between even variants of the same species.

2.2 Criteria for effective tuning strategies

Parameters that affect the expression level of an unregulated gene are depicted in Figure 2.1(A). For this simplified model, mRNA transcripts (R) are produced at some rate α_R and are degraded at some rate β_R. Similarly protein (P) is produced at some rate α_P and each protein is degraded at some rate β_P (see Ref. [41]). Each of these parameters can be controlled by adding, deleting, or altering specific DNA sequences (Fig. 2.1(B)). These parameters are affected by the code chosen to implement the components, control copy number, and choose the chassis. The complexity then scales as connectivity is added to this simple model to implement multi-gene networks.

Synthetic biologists have traditionally been constrained by the DNA on hand, acquired by a neighbor lab, or from an unfortunate bacterium to choose which code to use for a particular gene network. Circuit behaviors generally suffer from these constraints by rarely working when first assembled. In the last five years DNA synthesis costs have dropped significantly [42] while DNA assembly techniques have improved significantly [43, 44]. This trend is likely to continue, enabling, in only a few years, control over every nucleotide of a 20 kilobase gene circuit via direct synthesis for just a few hundred dollars. The question then becomes, "What do I order so it works?"

One way to improve the performance of a gene network that is not functioning as desired is to introduce focused variability into the design, generating a library of circuits [39, 45, 9, 46] with the same genetic components and connectivity, but with each member of the library operating in a different parameter regime. In bacteria, for example, promoters [47, 35], ribosome binding sites (RBS) [48, 49], RNA stability [50, 51], protein stability [40], and other biochemical details such as transcription factor regulation [52] or enzyme catalysis [53] can be varied to sample different regions of parameter space. Furthermore, sensitivity analysis [54] can guide the designer to parameters that, when tuned, are most likely to result in improved performance.

Subject to screening or selection, the effectiveness of a tuning library is proportional to the range of parameters it samples and is inversely proportional to its size. More specifically, a useful tuning approach explores the parameter space over a large range with high reso-

lution; results in a predictable relationship between the genetic sequence and the values of the corresponding parameters; and is scalable to complex networks. Ideally, a good tuning method should also be evolvable, forcing the host organism to focus mutations on highly tunable elements in the network, such that it complements directed evolution techniques [55, 56] by more frequently sampling mutations that enhance functionality.

Chapter 3

THE RBSSR: FINE-TUNING TRANSLATION RATES USING SIMPLE SEQUENCE REPEATS

3.1 Tunability of translation using the ribosome binding site spacer

The prokaryotic ribosome binding site (RBS) is a well-studied control sequence for the translation of mRNA and is an attractive target for tuning gene expression [57]. Changing 5' UTR sequences in bacteria can vary expression levels over five orders of magnitude using multi-copy plasmids encoding fluorescent proteins, with a predictable mapping from DNA sequence to protein expression [48].

One specific mechanism within the RBS that modulates translational efficiency is the distance in nucleotides between the Shine-Dalgarno region and the start codon of the target gene, with an optimal aligned spacing* of five nucleotides[58]. The spacer length between these two binding sequences has been shown to modulate translational efficiency for a broad range of proteins: RFP [48], β-galactosidase [37], and chloramphenicol acetyltransferase [58]. The spacer's effect appears to vary between organisms: *E. coli* ribosomes are more permissive of short spacer lengths than *Bacillus subtilis*; the converse is true for long spacer lengths [37].

Considering the criteria for effective tuning strategies in the previous chapter, incorporating evolvable tuning mechanisms should accelerate gene network optimization by focusing mutations to DNA sequences with strong and predictable effects on gene expression. Simple sequence repeats, tandem repeats of short nucleotide sequences, are known to have mutation rates 10^4–10^5 higher than arbitrary sequences of the same length that are manifest primarily as insertion/deletion mutations of individual repeat units[59, 60].

We have developed a tuning mechanism for gene networks in *E. coli* that couples the straightforward tunability of translation initiation rates via the RBS spacer region with the

*Aligned spacing refers to the distance between a best-match of the RBS to the reference 5'-UAAGGAGGU-3' Shine-Dalgarno sequence and a start codon AUG, GUG, or UUG.

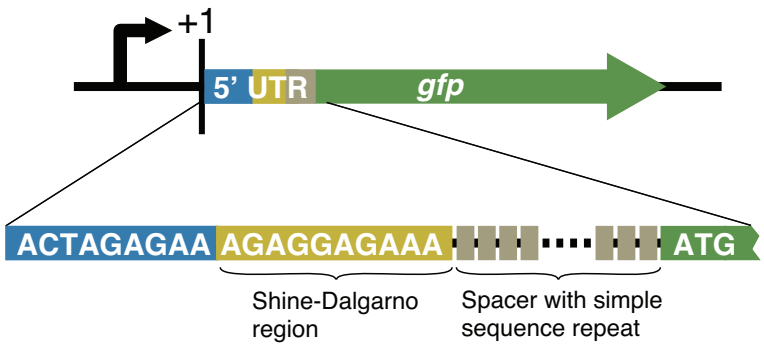

Figure 3.1: The rbSSR construct. A simple sequence repeat is embedded in the spacer region of the ribosome binding site between the Shine-Dalgarno sequence and the coding sequence of a target gene, in this case *gfp*.

Table 3.1: RBS sequences for rbSSR-GFP library parent constructs.

Parent construct	**Sequence**	**Aligned spacing [58]**
$(A)_{20}$	AGGAGAAAAAAAAAAAAAAAAAAATG	17
$(T)_{20}$	AGGAGAAATTTTTTTTTTTTTTTTTTTATG	21
$(AT)_{12}$	AGGAGAAATATATATATATATATATATATATG	22
$(AC)_8$	AGGAGAAACACACACACACACACATG	16

high mutation rate and strong bias for insertion/deletion mutations inherent to SSRs. We implement this mechanism by embedding mono- or di-nucleotide SSRs between the Shine-Dalgarno sequence and the start codon of target genes. We call this sequence motif the rbSSR (Fig. 3.1). This chapter focuses on the tunability of gene expression by generating rbSSR libraries in vitro; the in vivo mutability of these repeat sequences is covered in Chapter 5.

3.2 Characterization of multiple rbSSR-GFP libraries

To understand the resolution and limits of translational control with rbSSRs, we experimentally examined four rbSSR spacer motifs: $(A)_n$, $(T)_n$, $(AT)_n$, and $(AC)_n$ – that is, n repeats of either a single or a pair of nucleotides. For each motif we constructed a parent plasmid with a constitutive promoter, a strong Shine-Dalgarno region, and an initial rbSSR spacer,

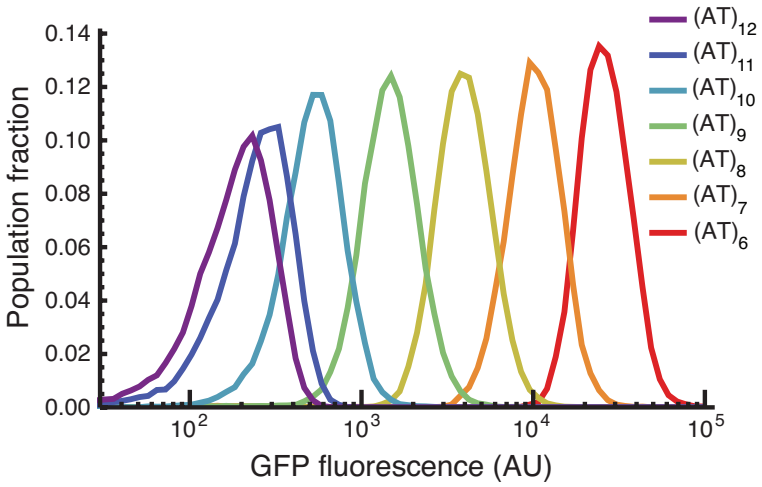

Figure 3.2: Fluorescence distributions of an rbSSR library expressing GFP. The cytometry traces shown consist of the $(AT)_6$–$(AT)_{12}$ rbSSR-GFP library expressed from constitutive promoter J23100 and measured by flow cytometry. Varying the spacer length evenly samples a large expression range over three orders of magnitude with uniform noise properties.

driving the expression of a *gfp* gene (Fig. 3.1). Taking advantage of the inherent instability of repeats during replication, especially in PCR [61] (see also Fig. 3.7(B)), we generated plasmid libraries by amplifying a region of each parent plasmid flanking the rbSSR sequence and re-inserting the mutated amplicons into pre-cut plasmid backbones (see Fig. 3.7(A)). The resulting plasmid libraries were transformed into *E. coli* and screened visually and via cytometry for unique fluorescence levels to produce a strain library (rbSSR-GFP) of repeat lengths for the four spacer motifs.

We measured the fluorescence output of rbSSR-GFP library strains via flow cytometry of exponentially growing cells. The mean intensity decreased linearly in log fluorescence over a 100-1000 fold range as the number of rbSSR repeats increased (Fig. 3.2). The cell-cell variation in GFP levels for each strain is considerably more uniform than what is observed for more noise-prone tuning approaches, such as the dose-response of an inducible promoter [62] (see Fig. 3.5). The rate at which fluorescence intensity decreases for each repeat motif depends on the nucleotide composition of the spacer, with the steepest decline for $(A)_n$ and the most gradual decline for $(T)_n$ (Fig. 3.3). The overall trend of the decline

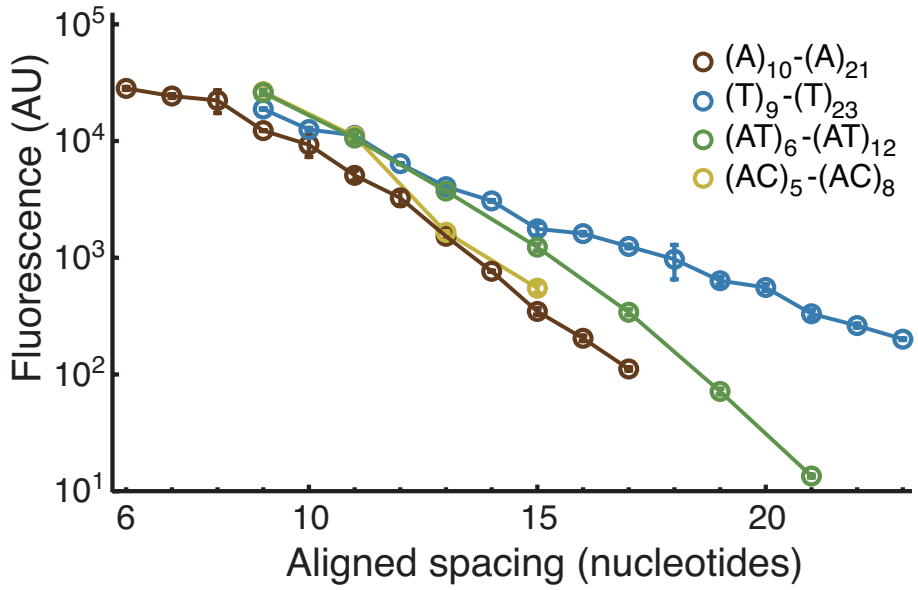

Figure 3.3: Mean GFP expression levels for rbSSR-GFP libraries generated from parent plasmids $(A)_{20}$, $(T)_{20}$, $(AT)_{12}$, and $(AC)_8$. Increasing the spacer length generally reduces gene expression, though the trend of the decline is sensitive to the nucleotide composition of the repeats. Each library uses identical promoter (J23100) and 5' UTR sequences. Error bars represent standard error from three colonies.

roughly corresponds to computational predictions [48], but with increasing disparity as the nucleotide composition of the spacer deviates from poly-(A) residues (see Fig. 3.6).

3.3 Characterization of rbSSR-GFP varying promoter and 5' UTR

The large range and high resolution in gene expression sampled by rbSSR libraries can be expanded further by altering other gene expression control sequences. Specifically, we built rbSSR-GFP libraries with two weaker promoters and an alternative 5' UTR (Fig. 3.4). Coarsely tuning transcription rate through promoter strength alone enables sampling of gene expression levels spanning nearly five orders of magnitude. Interestingly, using the alternate 5' UTR sequence (Table 3.2) which is predicted to have a stronger translation initiation rate indeed shows higher expression, but only for shorter spacer lengths. Longer spacer lengths show the inverse relationship. Shine-Dalgarno sequences that bind strongly to the 16s rRNA have been reported to cause stalling of translation initiation[63, 64]. It

Table 3.2: Sequences for promoter and 5' UTR variants of $(A)_n$ library.

Promoter	**Strength**	**Sequence**
J23100	1x	TTGACGGCTAGCTCAGTCCTAGGTACAGTGCTAGC
J23108	0.3x	CTGACAGCTAGCTCAGTCCTAGGTATAATGCTAGC
J23114	0.01x	TTTATGGCTAGCTCAGTCCTAGGTACAATGCTAGC
5' UTR variant	**Source RBS**	**Sequence**
Original	Biobrick (B0034)	ACTAGAGAAAGAGGAGAAA
Alternate	Bujard	GAATTCATTAAAGAGGAGAAA

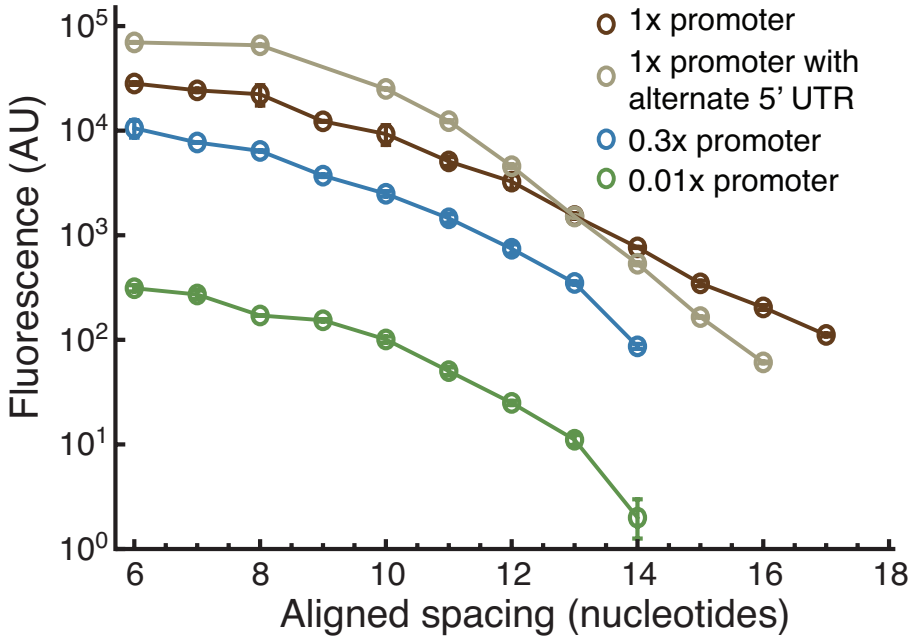

Figure 3.4: Mean GFP expression levels for multiple $(A)_n$ rbSSR-GFP libraries, as in Figure 3.3. Tuning translational and transcriptional efficiencies through regulatory sequences in the 5' UTR or promoter regions (Table 3.2), respectively, can be used in conjunction with rbSSR libraries to more broadly sample the expression space.

is possible that this effect is more pronounced as spacer length increases, resulting in the inversion of relative translation strengths for the two 5' UTR sequences when varying spacer length. We have observed the same inversion effect for two rbSSR libraries with different Shine-Dalgarno sequences for the lactose utilization project (Chapter 6).

3.4 Comparing rbSSR libraries to an inducible promoter

A conventional approach to tuning gene expression, often to study gene dosage effects, is to clone a gene of interest onto a multi-copy plasmid under control of a regulated, inducible promoter. The gene under study is often removed from its chromosomal locus and its expression is complemented by the inducible plasmid. One drawback to this approach is the non-linear dose response to inducer levels which generally complicates experimental sampling of inducer levels over more than a 10-fold range. Perhaps more importantly, cultures grown at intermediate inducer concentrations exhibit much higher cell-to-cell variability in gene expression [62]. Negative autoregulation of a transcriptional repressor has been shown to reduce expression noise and to linearize the inducer dose response in yeast [65]. An rbSSR library expressed in *E. coli* has similar characteristics, evenly sampling the expression space with uniform noise properties (Fig. 3.5). Embedding variable length spacers such as rbSSRs in RBS sequences at chromosomal loci using single-stranded DNA recombination (see Ref. [45]) may enable high-resolution gene dose response measurements with minimal modifications to wild-type DNA sequences in natural systems, thus minimizing cell-cell variability.

3.5 Comparing rbSSR library expression to computational predictions

The RBS Calculator is a software package developed to predict protein translation rates from the mRNA sequence surrounding initiation codons [48]. We used its reverse engineering mode to generate a set of translation rate predictions from the sequences of the rbSSR-GFP libraries, using the entire 5' UTR and the first 50 bases of the *gfp* coding sequence as input. The thermodynamic model of the RBS Calculator includes a free energy penalty for the RBS spacer sequence between the Shine-Dalgarno region and the start codon, termed $\Delta G_{spacing}$. For aligned spacing [58] greater than 5 nucleotides $\Delta G_{spacing}$ decreases quadradically with increased spacer length. The spacing parameters for the model are empirically fit to a set of experimental constructs with a poly-(A) spacer ranging from 0 to 15 nucleotides, which correspond to spacers $(A)_4$ to $(A)_{19}$ in the poly-(A) rbSSR-GFP library.

Comparing the predicted translation rates for each sequence to the mean fluorescence

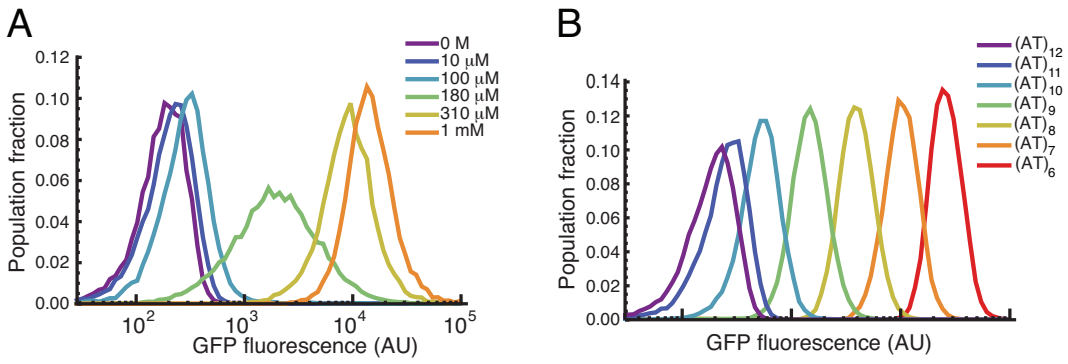

Figure 3.5: Tunability and noise comparison between traditional inducible system and rbSSR library with comparable ranges of expression. (A) Fluorescence distributions of GFP expression with a repressible promoter at multiple inducer levels. MG1655 cells were transformed with a mini-F plasmid encoding the $lacI^q$ allele and a high copy plasmid (pSB1A2, colE1 origin) encoding *gfp* under the control of a LacI-repressible promoter (BioBrick R0011) and medium-strength $(AT)_8$ rbSSR. Individual colonies were grown to saturation at IPTG concentrations of 0 M, 10 μM, 100 μM, 180 μM, 310 μM, and 1 mM in triplicate. Each culture was then diluted 200:1 in prewarmed media with unchanged IPTG levels and grown to mid-log phase at 37°C, diluted in PBS, and sampled in a cytometer. When grown at intermediate levels of IPTG, the cell-cell variability of GFP expression increases, which is consistent with other inducible systems expressed from plasmids [66] or the genome [21] and results in a system with low tunability at intermediate levels of expression. (B) Fluorescence distributions of seven strain $(AT)_6$-$(AT)_{12}$ rbSSR-GFP library, as in Figure 3.2. Covering the same range of expression levels as the system shown in (A), the rbSSR strain library samples the expression space evenly with uniform noise properties.

for each SSR motif of the rbSSR-GFP library, the qualitative trend of decreased expression with increased spacer length is captured by the model. However, as shown in Figure 3.6, while the mean GFP expression levels and the predicted translation rates (each in arbitrary units) match the values for the $(A)_n$ library well with no data transformation, significant quantitative differences in slope, spacer length offset, or fluorescence offset are apparent for the other rbSSR repeats. Specifically, the predictions fail to match the decrease in slope as adenine residues are replaced by uracil in the spacer region. It is possible that the algorithm could be refined by further characterization of spacer length variation while changing nucleotide composition, preferably with multiple coding sequences to verify that the dependence of translation rate of space nucleotide identity is not specific to the *gfp*

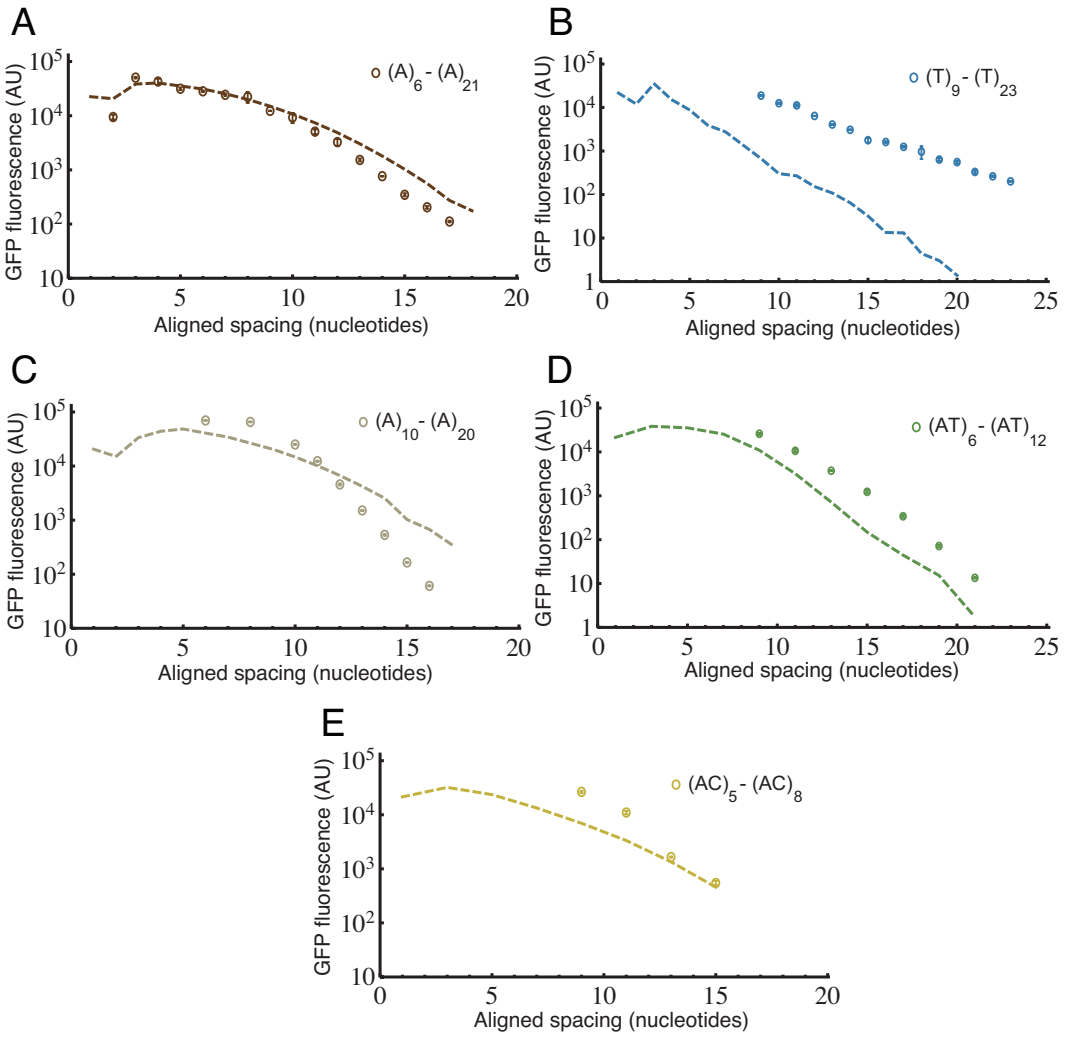

Figure 3.6: Comparison of RBS Calculator predictions to experimental results for each SSR motif of the rbSSR-GFP library. Each sub-library has an identical strong constitutive promoter (J23100) driving the expression of GFP. (A) $(A)_6$-$(A)_{21}$ rbSSR library. (B) $(A)_{10}$-$(A)_{20}$ with alternative 5' UTR and Shine-Dalgarno region. (C) $(T)_9$-$(T)_{23}$ rbSSR library. (D) $(AT)_6$-$(AT)_{12}$ rbSSR library. (E) $(AC)_5$-$(AC)_8$ library.

coding sequence. Nevertheless, the RBS Calculator appears to be accurate within at least one order of magnitude for the majority of sequences, suggesting it may be used to coarsely set a target expression level via 5' UTR sequence that can be fine-tuned using an rbSSR library.

3.6 Discussion

Gene expression libraries built with rbSSRs incrementally and predictably sample gene expression levels up to a 1,000-fold range by varying the number of repeats in the RBS spacer region. This large-range and high-resolution tuning approach can be expanded by coarsely tuning transcription rate through promoter strength or adjusted by altering the 5' UTR to change ribosome binding affinity. Coupling rbSSR tuning with coarse tuning approaches enables sampling of gene expression levels spanning nearly five orders of magnitude. These libraries can pinpoint gene expression levels with higher resolution than inducible promoters and with consistent cell-cell variability. Given these characteristics, rbSSR libraries are useful tools for fine-tuning gene expression.

Creating rbSSR libraries is compatible with existing combinatorial or compuational tuning approaches for transcription [39, 35] and translation [48, 49] rates, as well as for RNA stability [51], since none of these tuning strategies specifically utilize the RBS spacer. Note that through PCR mutagenesis of the rbSSR, we generated no fewer than 9 bases of repeat sequence, which – depending on the nucleotide composition of the Shine-Dalgarno region – results in spacing near the optimum of five bases (Fig. 3.6(A)) and suggests a practical limit for recovering mutations generated by slipped-strand mispairing.

3.7 Methods

Strains and media: rbSSR-GFP library construction and assays were carried out in MG1655. M9 minimal media (M8000, Teknova) supplemented with 50 μg/mL kanamycin was used for fluorescence assays.

GFP library generation via PCR assembly: Generation of rbSSR libraries via PCR is illustrated in Figure 3.7. The backbone for the rbSSR-GFP libraries was generated from a parent plasmid with a p15A replication origin containing $(A)_{20}$, $(T)_{20}$, $(AC)_8$, or $(AT)_{12}$ (see Table 3.1 for sequence details) by digestion with endonucleases XbaI and NdeI to excise the promoter, the rbSSR, and the first 230 bases of the GFP coding sequence. The digestion was followed by gel extraction and purification of the backbone fragment. Spacer variation for the rbSSR-GFP libraries was generated via PCR of a 450 bp fragment

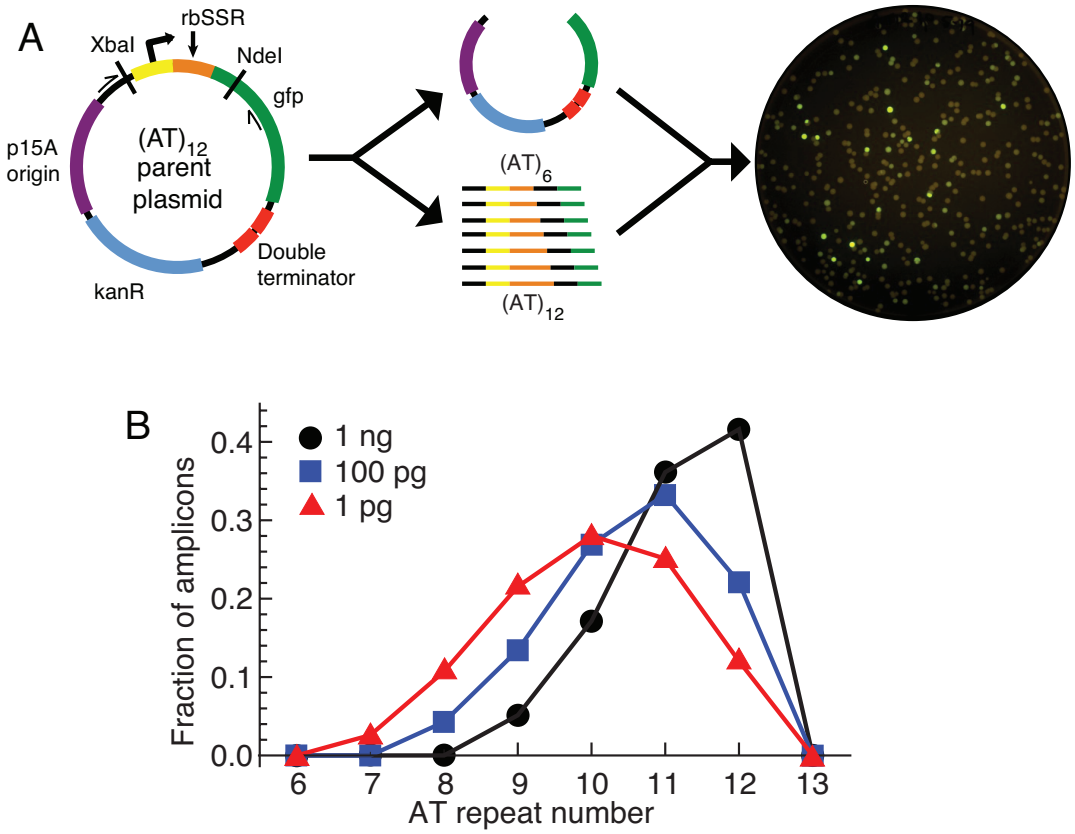

Figure 3.7: Construction of rbSSR libraries via PCR assembly, with the $(AT)_{12}$-GFP parent plasmid as example. (A) A parent plasmid is digested to excise its regulatory region and separately amplified to create spacer length variation. The spacer variants are cloned into the digested backbone and screened via flow cytometry. (B) Sample distributions of rbSSR repeats processed from sequence trace data of PCR reaction. For each distribution $(AT)_{12}$-GFP was used as the PCR template at levels from 1 ng to 1 pg. The plate image in (A) is a sample of the variety of transformants that result from a Gibson assembly reaction with a digested parent plasmid and rbSSR amplicons from a 1 pg template PCR reaction.

using Phusion Flash master mix (F548, Finnzymes) and primers Vf2 and rbSSR-GFP_rev, with the parent plasmids as templates. These PCR amplicons, which have homologies to the backbone fragment, were also purified by gel extraction. The backbone and rbSSR amplicon libraries were joined using Gibson assembly [43], and transformed directly into expression strain MG1655 for screening. Repeat lengths with fewer than nine base pairs were not observed from PCR reactions; plasmids for repeats $(A)_6$–$(A)_9$ (see Fig. 3.6) were

thus constructed by ordering oligos encoding each spacer, followed by PCR amplification and Gibson assembly to a PCR-amplified vector backbone. All constructs were verified by sequencing.

Due to the inherent instability of repeats, PCR – even with a high fidelity polymerase – results in insertions and deletions of SSR repeat units with a strong bias for deletions [61] (see Chapter 5 for more details). The distribution of spacer lengths is dependent on the number of PCR cycles before the reaction reaches saturation and, thus, the concentration of plasmid template. To monitor the number of cycles required to run PCR reactions to completion, amplifications were carried out in a quantitative PCR machine (CFX96, BioRad) using Phusion Flash master mix (Finnzyme) and evaGreen dye (Biotium, Inc.) at reaction volumes of 20 μL. In each reaction we varied the amount of plasmid template from 1 ng to 100 fg by factors of 10. We sequenced gel-extracted PCR amplicons to determine the distributions of rbSSR lengths, which are shown in Figure 3.7(B).

Cell growth and plate reader measurements. Freshly streaked colonies were transferred in triplicate to 200 μL M9 minimal media in 96-well plates (Costar 3795) and grown to saturation overnight at 37° C in a shaker. The cultures were then diluted 1:100 in 200 μL prewarmed fresh broth in 96-well plates (Costar 3904), grown at 37°C to OD600 0.15 to 0.2 in a plate reader (Biotek) with shaking. Optical densities (600 nm) and GFP measurements (485 nm excitation, 510 nm emission) were taken every 10 minutes. When grown to target density, 10 μL of each culture was transferred to 100 μL 1x PBS (Gibco) with 34 μg/mL chloramphenicol chilled at 4°C.

Flow cytometry measurements. Diluted cultures from the plate reader measurements were transferred to a flow cytometer (C6 with CSampler, Accuri). To prevent well-well contamination, blank wells containing PBS were sampled after each sample well. GFP measurements (488 nm excitation, 533 nm emission) were recorded for 50,000 events per sample. Cells were gated using a rectangular gate in forward scatter and side scatter. Background fluorescence levels from cells containing an empty vector without *gfp* were subtracted from the geometric mean of GFP expression for each sample culture.

Chapter 4

SCALING UP COMPLEXITY: USING RBSSRS TO FINE-TUNE THE FUNCTIONAL MODES OF A BISTABLE SWITCH

To demonstrate that rbSSRs can be used to fine-tune functional gene networks we built an rbSSR-enhanced bistable switch using the same circuit architecture as the mutually inhibitory switch described by Gardner, Cantor, and Collins [17]. For this switch, the two states are controlled by transcriptional repressor proteins LacI and TetR, which are expressed bicistronically with GFP and RFP, respectively (Fig. 4.1). Cells can be forced to either switch state using the chemical inducers IPTG and aTc, which bind and deactivate LacI and TetR, respectively.

The dominant state and spontaneous switching rate between states for this circuit depend on the initial state of the system, the expression strength of each repressor gene, the stability the associated proteins and mRNAs, the rate of leaky transcription for each repressor/promoter combination (see Fig. 4.2), plasmid copy number, and circuit-host interactions (referred to here as context) that affect global expression dynamics and growth rate. As a result, it is difficult to predict a priori if one state will dominate or if each state will be equally likely when the switch is expressed in a given strain. The architectural simplicity of this circuit make it an ideal candidate for investigating the tunability of functional behaviors when using rbSSRs.

4.1 A bistable switch architecture, untuned

The initial bistable switch construct, as in Figure 4.1, was built without rbSSRs using "off-the-shelf" parts from the Registry of Standard Biological Parts (see `http://partsregistry.org/`), encoding the switch on a p15A-origin plasmid (15-30 copies per cell). In the first assays with this construct, expressed in *E. coli* strain DH5α, the switch was only stable in the TetR-dominant, or red, state. When observing fresh transformants for this construct on agar plates, red colonies, LacI-dominant green colonies as well as bimodal colonies emerged,

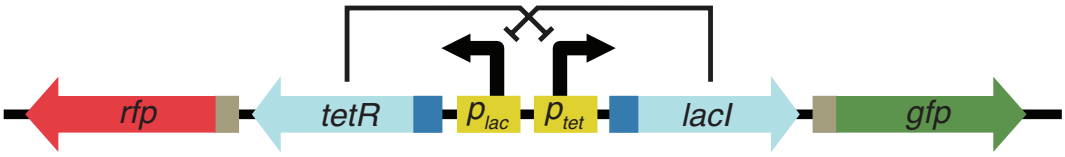

Figure 4.1: Circuit architecture for a mutual inhibitory genetic switch. Genes coding for two transcriptional repressors, TetR and LacI, are each placed under control of the other repressor. RFP and GFP reporters are coexpressed with TetR and LacI, respectively, to observe cell state.

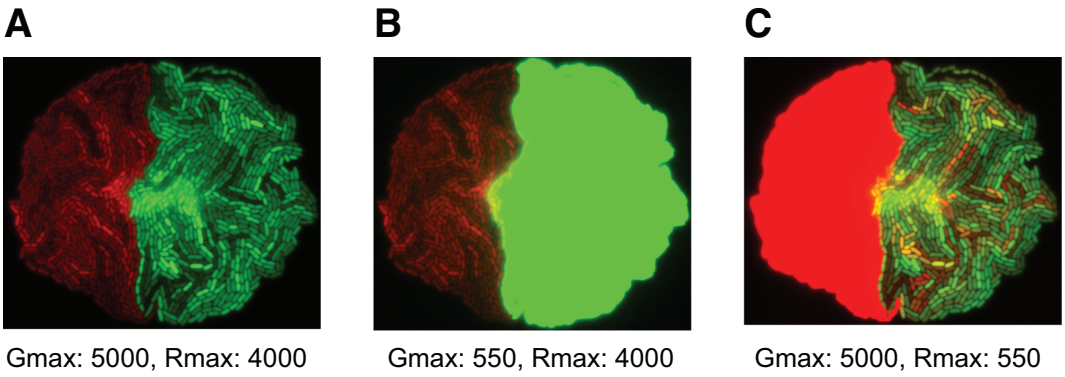

Figure 4.2: Noisy gene expression for a bistable switch circuit variant in the LacI-dominant state. TetR-dominant and LacI-dominant microcolonies growing next to each other are shown at multiple contrast levels. (A) Standard contrast settings. (B) High contrast in green channel showing little leaky expression from the TetR-dominant microcolony. (C) High contrast in red channel illustrating noisy gene expression from the LacI-dominant microcolony. The contrast levels for the 12-bit images are shown in red and green for each panel.

though unimodal green colonies appeared only after 2-3 days of incubation at 37°C.

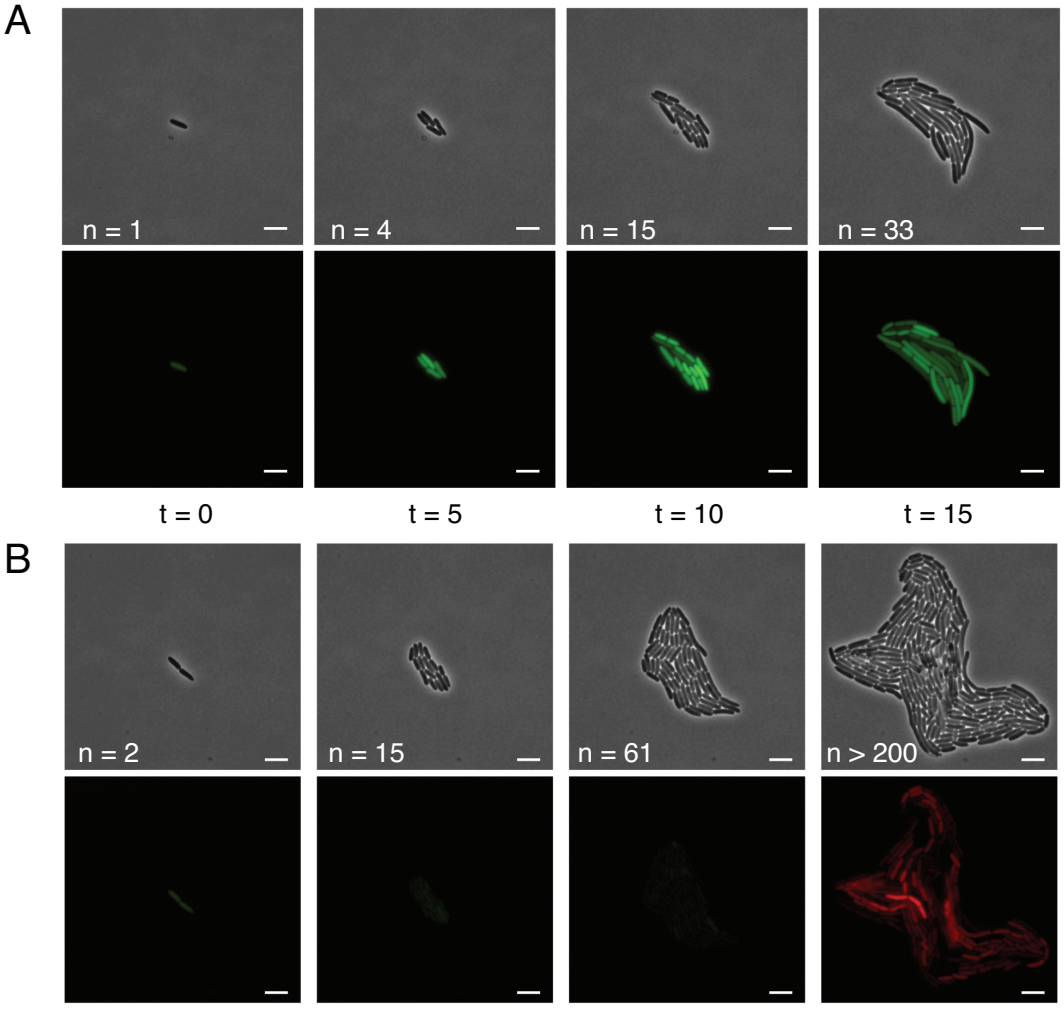

Figure 4.3: Time-lapse microscopy images of the unoptimized bistable switch circuit showing imbalanced growth rates and spontaneous switching to a more stable state. (A) Unhealthy LacI-dominant microcolony growing slowly, with elongated cells and multiple lysis events. (B) A spontaneous switcher microcolony, initially green, flips to the TetR-dominant state, resulting in a much higher growth rate. The microcolonies originated from the same culture that had been forced to the LacI state and were incubated on the same microscope slide under identical conditions. In each image frame, n is the number of cells in the microcolony and t is in hours. The scale bar is approximately 5 μm.

To investigate the dynamics of the initial circuit more closely, we recorded the growth and fluorescence dynamics of cells forced to the LacI state with aTc via time-lapse microscopy in an uninduced growth environment to confirm the growth defect and observe any switching dynamics. We found that while nearly all cells initially expressed GFP, some cells spontaneously switched to the TetR state once actively growing and dividing. As shown in Figure 4.3, cells that remained in the LacI-dominant state grew slowly and were prone to lyse, likely due to overexpression of LacI. On the other hand, cells that spontaneously switched to the TetR-dominant state grew much faster with no apparent negative effects. For this DNA implementation, the switch architecture is not well-tuned: the two switch states are poorly balanced and the LacI-dominant state is likely not mutationally robust. This implementation of the switch is not bistable over long periods covering many generations.

4.2 Tuning the bistable switch with dual rbSSRs

To make the switch easily tunable, we incorporated poly-(T) and poly-(A) rbSSRs between identical, unoptimized 5' UTR sequences and the *tetR* and *lacI* genes, respectively. From this design we built a 36-strain combinatorial library (rbSSR-BSS) using an oligo assembly method modified from Ref. [44] (see Fig. 4.11) using six single-stranded DNA fragments that encode the regulatory sequences for the two operons, including rbSSR spacers of length 10 to 20 repeats in steps of two (Fig. 4.4). This library represents a coarse sampling (30%) of the reachable rbSSR expression space from 10 to 20 repeats. We transformed sublibrary assemblies that contained all six poly-(A) variants for each poly-(T) rbSSR into $lacI^-$ expression strain 2.320 for screening via flow cytometry and sequence verification.

We initially characterized the behaviors of the rbSSR-BSS strain library by observing the natural bias of each switch variant in fresh, uninduced 2.320 transformants. We used flow cytometry to capture a scatter plot of red and green fluorescence for the cells during exponential growth (Fig. 4.5). These results were combined to produce a grid of scatter plots for the entire library (Fig. 4.6) to reveal a suprisingly broad range of switch distributions. As expected, a strong rbSSR for one repressor coupled with a weak rbSSR for the opposing repressor results in a natural bias towards the strong rbSSR state and unimodal

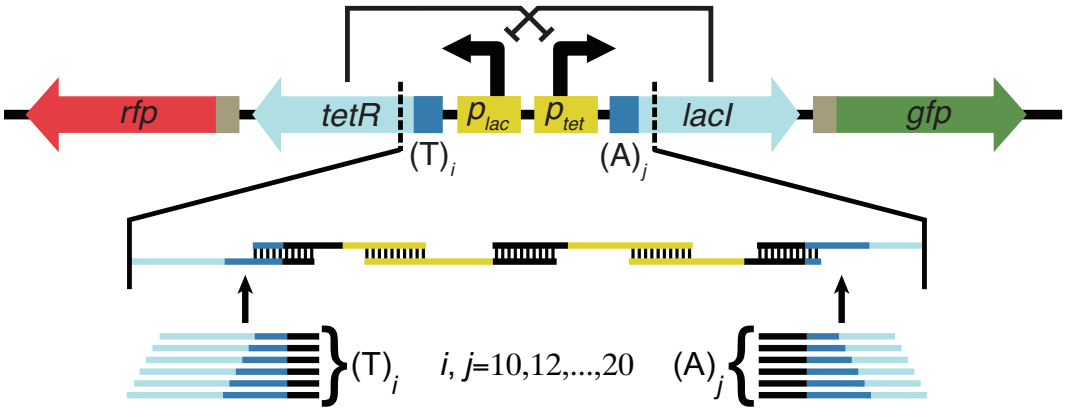

Figure 4.4: Approach to fine-tuning bistable switch behaviors with dual rbSSRs. The RBS spacer for each repressor gene is replaced with a variable-length SSR spacer to generate a library of plasmids with varied expression strengths. Specifically, a $(T)_i$ rbSSR controls TetR expression and an $(A)_i$ rbSSR controls LacI expression.

cell populations. However, the distribution of switch states is bimodal for the majority of the library variants and the fraction of cells in the green state, while consistent for a single variant, shifts predictably as rbSSR strength is varied.

4.3 Gene network context matters

The context in which a gene network such as the bistable switch is expressed can dramatically affect the performance of the network [20, 67], but its effect is, in general, poorly understood and not easily incorporated into modeling gene network performance. To investigate the effect of host strain context for the bistable switch, we transformed each variant of the rbSSR-BSS library into a second *lacI*$^-$ strain of *E. coli* – BW25113 Δ*lacI* – and performed the same initial characterization assay described for the 2.320 strain library.

The genotypes of the two strains are different, yet they might be expected to behave similarly with respect to the switch circuit since neither strain expresses active LacI. Strain 2.320 is a Monod strain with the following genotype: F-, *lacZ13*(Oc), *lacI22*, *rpsL135*(strR), *malT1*(λ^R), *xyl-7*, *mtl-1*, *thi-1* [68]. Strain BW25113 Δ*lacI* is from the Keio collection and has the following genotype: F-, Δ*(araD-araB)567*, Δ*lacZ4787*(::rrnB-3), Δ*lacI785::kan*, λ^-, *rph-1*, Δ*(rhaD-rhaB)568*, *hsdR514* [69]. It is difficult to predict how any of these genomic

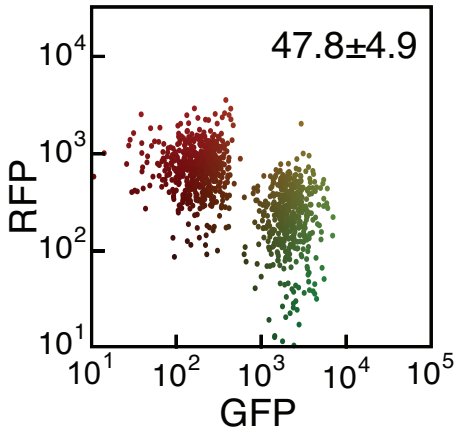

Figure 4.5: Scatter plot showing fluorescence distribution of $(T)_{20}/(A)_{20}$ rbSSR-BSS variant in strain 2.320. The horizontal axis is the GFP fluorescence level and the vertical axis is the RFP fluorescence level. The color indicates whether the cell is in the red, green, or mixed sub-population and the number in the upper right represents the mean and standard error of the percent of cells in the green state, measured from three colonies. The scatter plot shows 1,000 points selected at random from the associated cytometry data.

variations may affect circuit performance, and engineers rarely model the effects of the strain context on the gene network.

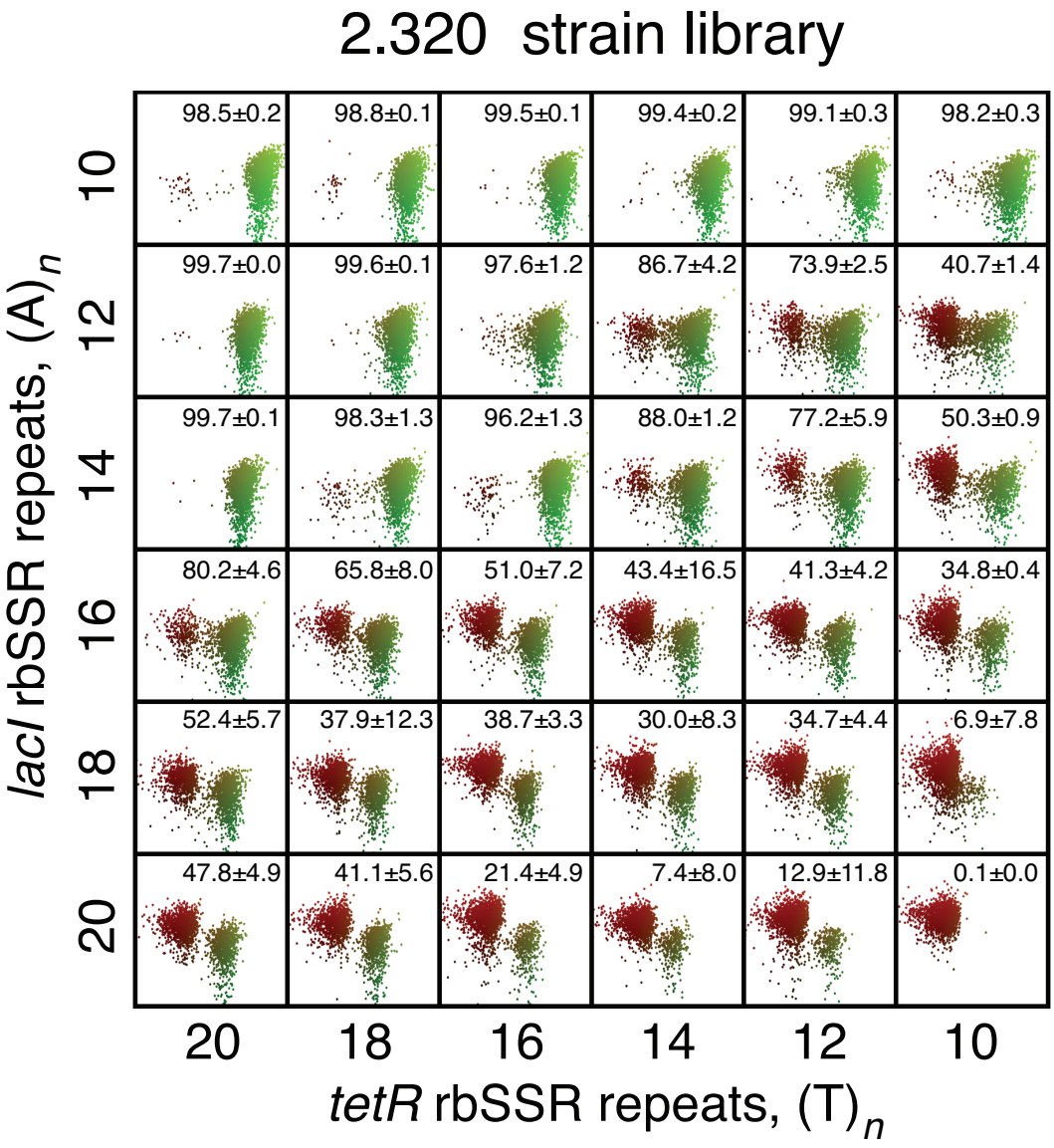

Figure 4.6: Grid of scatter plots for the 2.320 strain library plotted as in Figure 4.5. The comparative strength of rbSSR pairs determine the distribution of cells in the LacI- and TetR-dominant states. The number in the upper right represents the mean and standard error of the percent of cells in the green state, measured from three colonies. The scatter plot shows 1,000 points selected at random from the associated cytometry data.

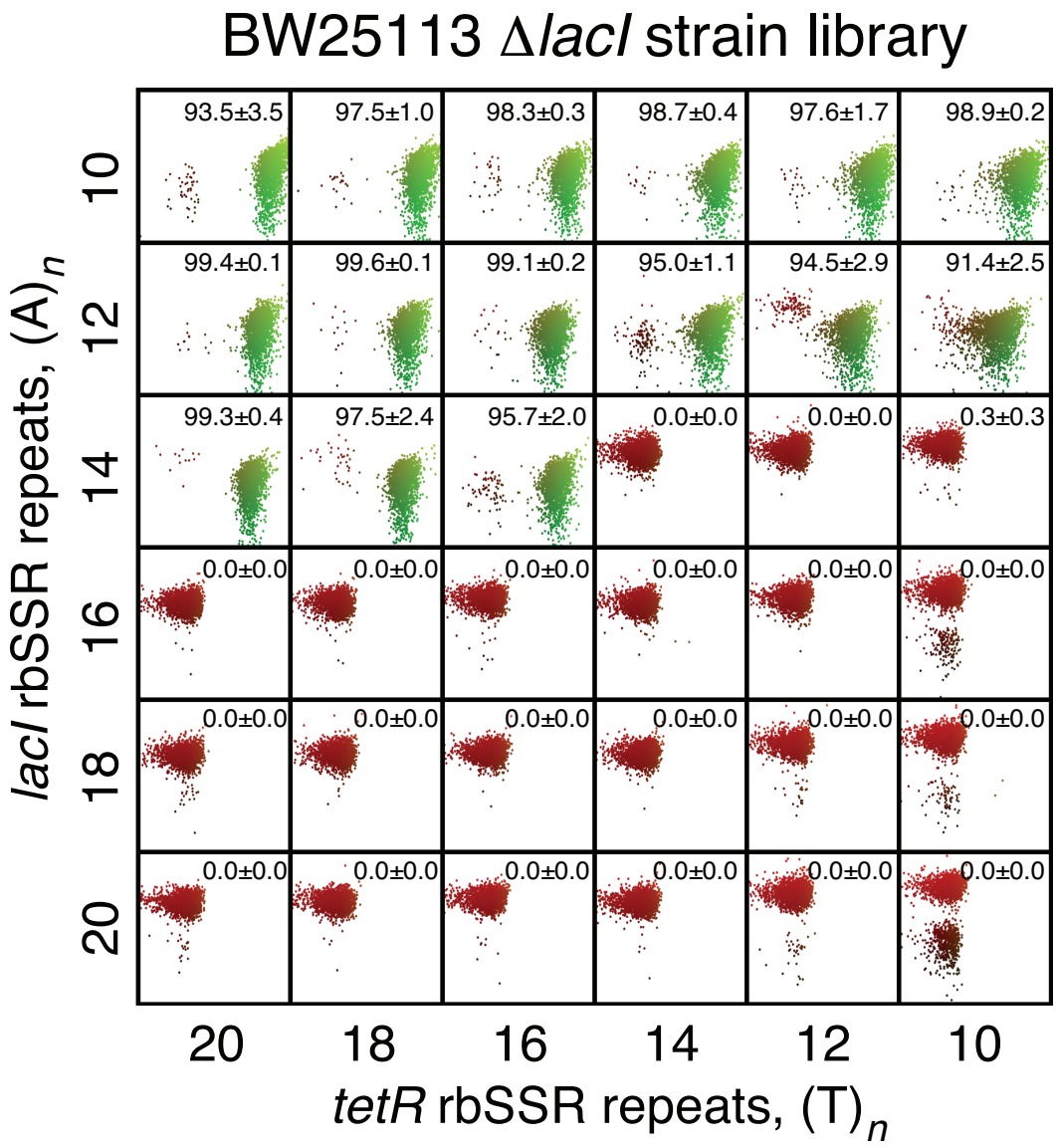

Figure 4.7: Grid of scatter plots for the BW25113 Δ*lacI* strain library, as in Figure 4.5, showing the fluorescence distributions of the majority colony type. The host strain affects the behavior of the switch library primarily by sharpening the boundary between strains in which one or the other state dominates and results in fewer bimodal constructs.

We found that the behaviors of strain BW25113 Δ*lacI*, shown in Figure 4.7, are profoundly different from the first strain in Figure 4.6. The same rbSSR pairs that generate bimodality in the original strain produce unimodal behaviors for most variants of the new strain. Also, while we observed little colony-to-colony variation in the distributions of switch states for the first strain library, nearly half of the strains in the second library exhibited two distinct colony phenotypes – one biased towards LacI-dominant cells, and the other biased towards TetR-dominant cells. For these strains, the distribution of colony types shifted predictably with rbSSR strength (Fig. 4.8).

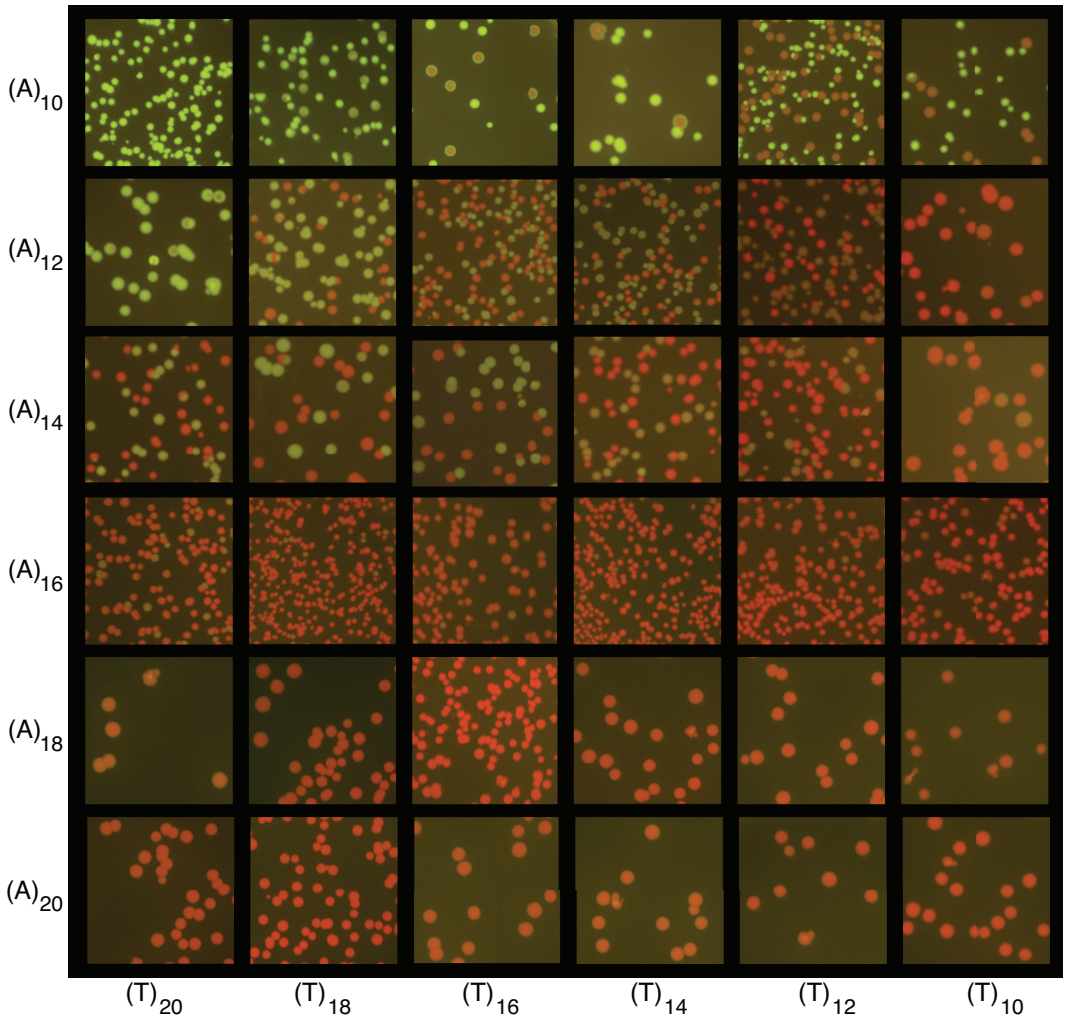

Figure 4.8: Cropped plate images for BW25113 Δ*lacI* strain library showing distributions of colony type which vary predictably with rbSSR lengths. This broadly dual-phenotype behavior was not observed for strain 2.320. To enhance fluorescence levels each plate was imaged after 36 hours of incubation at room temperature following an initial 12-14 hour incubation at 37°C after transformation.

4.4 Functional modes of a bistable switch

Depending on the values of the network parameters, the mutual inhibition circuit architecture for the bistable switch can generate a range of useful behaviors. If both states are highly stable, cells can act as molecular detectors, responding to environmental signals and retaining the detected state over tens of generations in the signal's absence [17, 70]. If one state is sufficiently less stable than the other, cells can act as programmable timers when initialized in the less stable state [39]. If the two switch states are well-balanced, cells could perform coin-flipping behaviors as a noisy switch. A synthetic bistable switch encoding coin-flipper behaviors could be a useful foundational circuit for implementing synthetic multicellular systems that emerge from individual cells, such as bet-hedging, pattern formation, and division of labor [71].

The bistable switch as a memory device or genetic timer. To further characterize the behavior of the switch library, we performed induction and microscopy assays on subsets of the strains. First, to observe the long term stability of each switch, we forced strains to the TetR- and LacI-dominant states using the chemical inducers IPTG and aTc, respectively. After a period of induced growth, we removed the inducers by washing and monitored fluorescence distributions over 96 hours of continuous growth (Fig. 4.9).

We found that a few library variants – notably, $(\mathrm{T})_{12}/(\mathrm{A})_{12}$ in strain 2.320 – remained stable in either induced state throughout the duration of the experiment. Other populations, such as $(\mathrm{T})_{16}/(\mathrm{A})_{14}$, slowly drifted away from the induced state. A few variants with reduced growth rates for one of the states – $(\mathrm{T})_{12}/(\mathrm{A})_{20}$ in strain BW25113 $\Delta lacI$ and $(\mathrm{T})_{18}/(\mathrm{A})_{10}$ in both strains (see Table 4.1) – behaved as monostable timer circuits, eventually shifting completely to the opposite state, results similar to the yeast timer circuit from Ref. [39].

It should be noted that for variants with slow growth rates, the short-term bias of the switch as observed in Figures 4.6 and 4.7 is not maintained over long periods of continuous growth, as cells that switch to the state with a higher growth rate sweep the population. In addition, it appears that the number of generations under induced conditions was not sufficient to fully dilute repressor levels in variants expected to have very high repressor concentrations: $(\mathrm{T})_{12}/(\mathrm{A})_{20}$ in BW25113 $\Delta lacI$ (aTc-induced) and $(\mathrm{T})_{18}/(\mathrm{A})_{10}$ in both

Table 4.1: Doubling times, in minutes, for 2.320 (white) and BW25113 Δ*lacI* (gray) bistable switch strains.

$(A)_{10}$	372 ± 38	229 ± 27	280 ± 20	294 ± 30	265 ± 25	153 ± 5
	304 ± 37	283 ± 6	223 ± 18	249 ± 28	215 ± 22	164 ± 7
$(A)_{12}$	200 ± 13	167 ± 11	148 ± 9	133 ± 5	136 ± 6	134 ± 2
	177 ± 26	144 ± 6	125 ± 1	121 ± 4	119 ± 3	120 ± 3
$(A)_{14}$	148 ± 6	155 ± 21	111 ± 8	100 ± 6	131 ± 9	136 ± 13
	122 ± 10	129 ± 10	118 ± 5	122 ± 6	135 ± 4	174 ± 7
$(A)_{16}$	125 ± 7	131 ± 4	126 ± 4	128 ± 3	117 ± 5	133 ± 7
	114 ± 2	118 ± 11	115 ± 6	126 ± 3	112 ± 7	184 ± 12
$(A)_{18}$	117 ± 13	131 ± 4	137 ± 6	117 ± 12	141 ± 10	149 ± 6
	119 ± 7	115 ± 1	128 ± 3	118 ± 6	148 ± 9	182 ± 47
$(A)_{20}$	134 ± 3	139 ± 6	146 ± 4	138 ± 3	149 ± 5	185 ± 9
	130 ± 2	111 ± 8	121 ± 4	185 ± 5	312 ± 57	386 ± 137
	$(T)_{20}$	$(T)_{18}$	$(T)_{16}$	$(T)_{14}$	$(T)_{12}$	$(T)_{10}$

strains (IPTG-induced). These variants exhibited timer or pulse-like behaviors that may behave more predictably if allowed to grow in the induced conditions for a longer period.

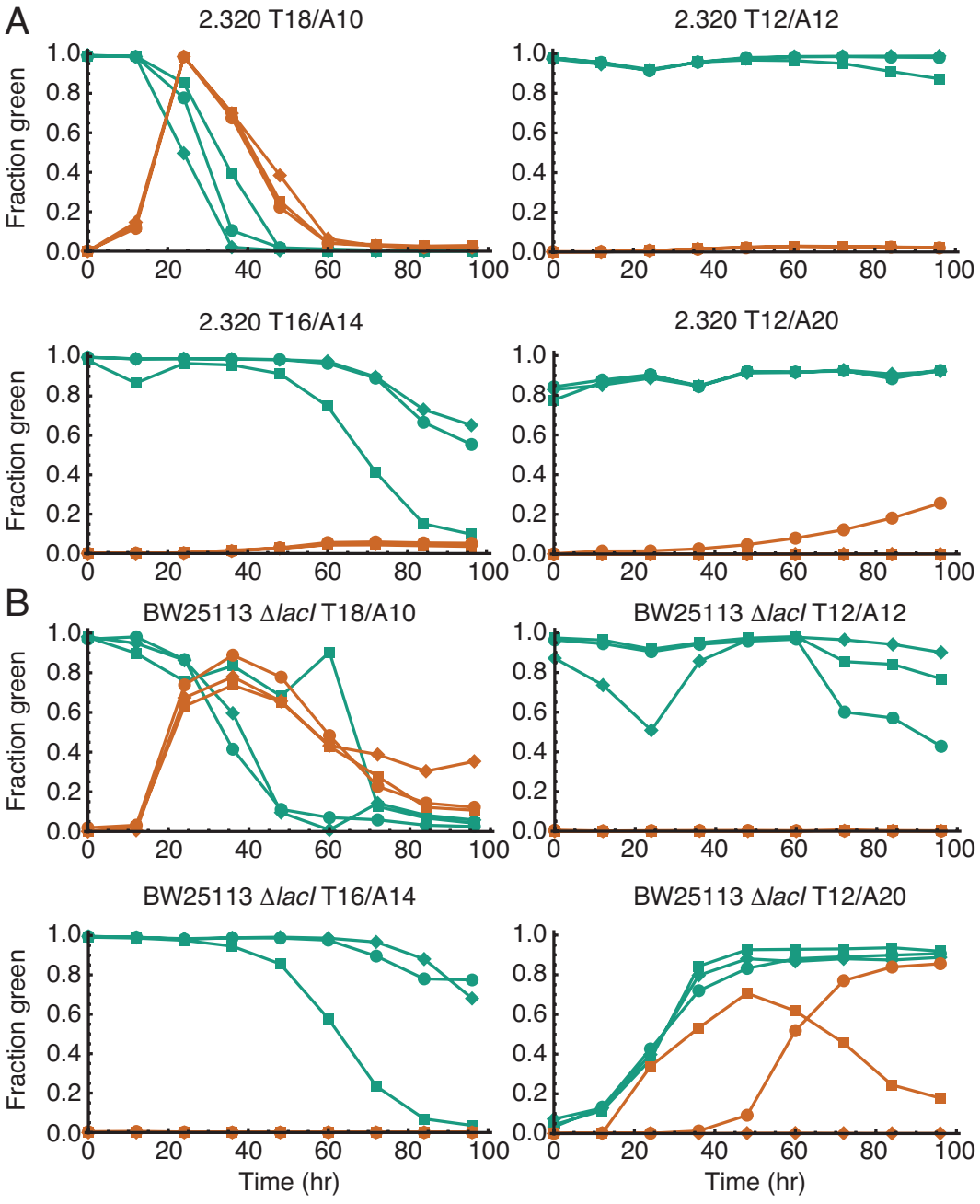

Figure 4.9: Stability of rbSSR-BSS sublibrary in strain backgrounds 2.320 (A) and BW25113 Δ*lacI* (B) over 96 hours. Aqua traces are for cultures induced with aTc; orange traces are for cultures induced with IPTG. Matching symbols correspond to the same starting population prior to induction.

Table 4.2: Fraction of green (top left), mixed (top center), and red (top right) microcolony types with sample size (bottom) for a subset of the 2.320 rbSSR-BSS strain library.

	$(T)_{20}$			$(T)_{18}$			$(T)_{16}$			$(T)_{14}$			$(T)_{12}$			$(T)_{10}$		
$(A)_{10}$																1.00	0.00	0.00
	n=0			n=0			n=0			n=0			n=0			n=40		
$(A)_{12}$													0.94	0.06	0.00	0.29	0.43	0.29
	n=0			n=0			n=0			n=0			n=16			n=35		
$(A)_{14}$										0.88	0.12	0.00	0.81	0.15	0.04	0.45	0.00	0.55
	n=0			n=0			n=0			n=16			n=27			n=29		
$(A)_{16}$				0.72	0.17	0.10	0.89	0.11	0.00	0.50	0.11	0.39	0.26	0.15	0.59	0.40	0.10	0.50
	n=0			n=29			n=19			n=28			n=27			n=20		
$(A)_{18}$	0.29	0.38	0.33	0.23	0.13	0.64	0.23	0.07	0.70	0.24	0.08	0.68				0.00	0.00	1.00
	n=21			n=47			n=43			n=25			n=0			n=14		
$(A)_{20}$	0.42	0.16	0.42	0.45	0.04	0.51												
	n=19			n=47			n=0			n=0			n=0			n=0		

The bistable switch as a noisy stochastic switch To investigate the switching dynamics of the uninduced switch, we performed microcolony growth assays from single cells for a subset of the 2.320 strain library (Table 4.2). For a given rbSSR-BSS library variant, we observed some combination of three distinct microcolony phenotypes: green (LacI-dominant), red (TetR-dominant), and mixed. Green microcolonies are characterized by strong GFP expression. Red microcolonies are characterized by strong RFP expression. Mixed microcolonies for a given strain are characterized by a mixture of green cells and red cells, but expressed at lower fluorescence levels that the green or red microcolonies. This suggests that cells in mixed microcolonies may be expressing non-negligible levels of both transcription factors. While we observed no leaky GFP expression for TetR-dominant microcolonies, the LacI-dominant state is generally much noisier (see Fig. 4.2), which is consistent with the reported leakiness of synthetic promoters with *lac* operator sites [72].

Along the diagonal of Table 4.2 – for roughly equal rbSSR lengths – a few rbSSR pairings exhibited a significant fraction of cells in transition between states, which resulted in mixed-state microcolonies. We further characterized the switching behaviors for a set of three strains: $(T)_{10}/(A)_{12}$ (strong), $(T)_{14}/(A)_{16}$ (medium), and $(T)_{20}/(A)_{20}$ (weak), using time-lapse microscopy of microcolonies originating from individual cells. All three strain variants generated microcolonies with each of the three microcolony phenotypes (Fig. 4.10). While

mixed microcolonies for the strong and medium switcher constructs expressed fluorescent proteins at lower levels than the respective red and green microcolonies, those from the weak construct were nearly as fluorescent as the red and green microcolonies, suggesting that leaky expression from the opposing transcription factor is reduced for this weak pairing of rbSSRs. To view high-resolution time-lapse movies for each of these colony types, please visit the URL `http://depts.washington.edu/soslab/rbSSR`.

It is likely that expression levels lower than those for $(\mathrm{T})_{20}/(\mathrm{A})_{20}$ may retain bistability. One result of weaker repressor expression appears to be a tradeoff of increased switching rates for lower protein production costs. Strategies outside rbSSR tuning, such as reducing the binding affinity of the ribosome to the Shine-Dalgarno region, adding active protein degradation tags [40], or reducing plasmid copy number, could be used to reduce gene expression. However, not all strategies have the same effect. For instance, reducing plasmid copy number will likely affect other network parameters, such as switching rates [73]. Further investigation to find minimum expression levels for the repressors to stably maintain a state may reveal more interesting and useful behaviors.

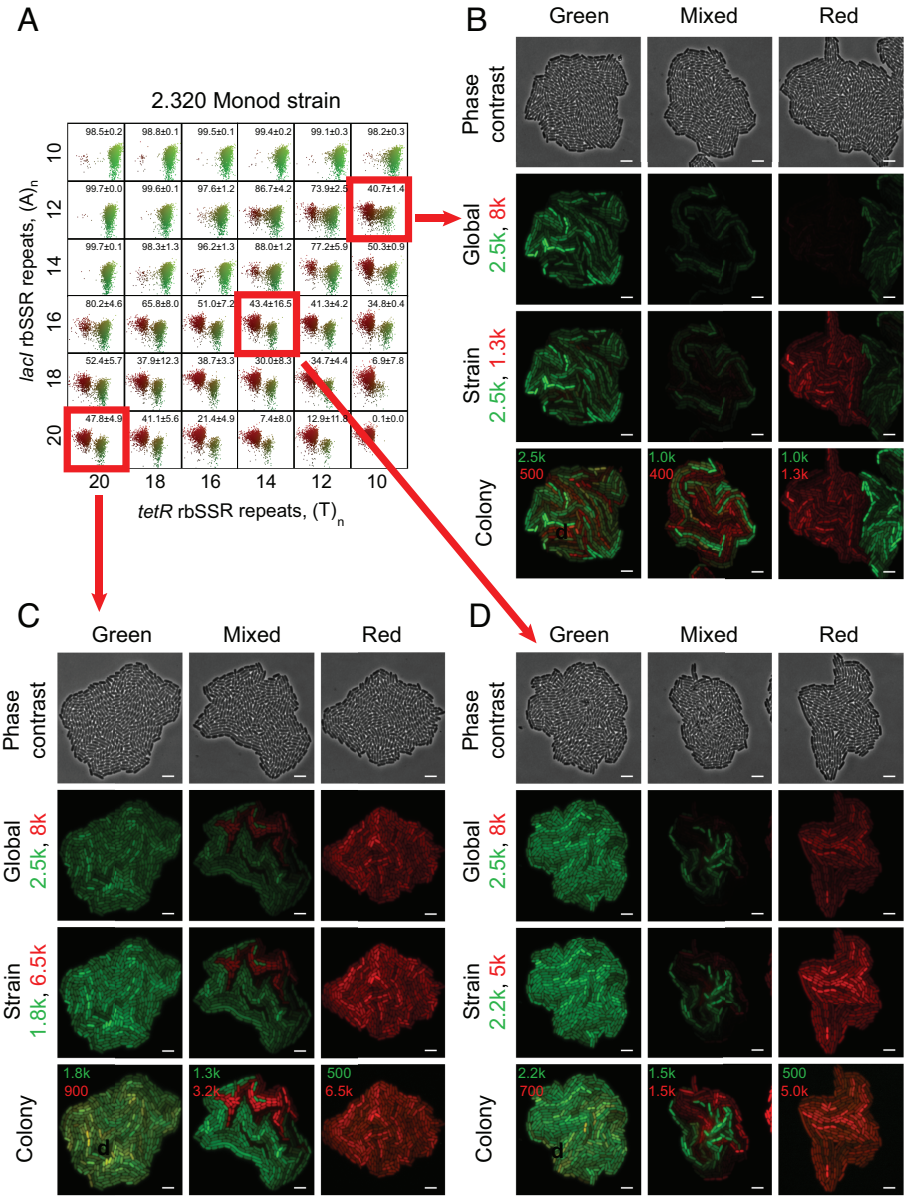

Figure 4.10: Well-balanced rbSSR pairs from different regions of the switch library exhibit fast switching dynamics with three distinct microcolony types. (A) rbSSR-BSS library, as shown in Figure 4.6, with the "switcher" strains highlighted by red boxes. (B) Still images of green, mixed, and red microcolonies from strain $(T)_{10}/(A)_{12}$ in phase contrast and fluorescence at three contrast levels: overall contrast settings (global); settings matching the maximum red and green levels observed for each strain (strain); and settings matched to the fluorescence levels of the microcolony (colony). (C) Still images of green, mixed, and red microcolonies for strain $(T)_{20}/(A)_{20}$ (D) Still images of green, mixed, and red microcolonies for strain $(T)_{14}/(A)_{16}$. The scale bar is approximately 5 μm.

4.5 Discussion

The difference between Figure 4.6 and Figure 4.7 and the range of functional behaviors observed when varying rbSSR pairings both underscore the importance of tuning and highlight limitations of tuning a single network parameter. In the initial characterization assay, roughly 40 generations pass from the initial plating of transformants to the cytometry assay. Over that period, switch variants in strain 2.320 appear to approach steady-state distributions relatively quickly, evidenced by consistent fluorescence distributions across colonies.

By contrast, the same constructs in BW25113 $\Delta lacI$ appear to maintain some initial state, resulting in two distinct colony phenotypes. For this circuit, it is likely that adjusting an additional network parameter, such as circuit copy number [73] or promoter leak [35] could compensate for the contextual differences between the two strains. We show some evidence for this result via plasmid copy number reduction in the section on engineering a self-destructive altruism circuit in Chapter 6. Although it is unlikely that these differences could be predicted for two arbitrary strains, it is reasonable to expect that two arbitrary strains could be tuned to behave the same way – suggesting that building tuning into system is an extremely important design consideration.

This work also demonstrates that the behaviors for a gene network library, like a bistable switch, can be highly sensitive to the host context. We show that cells expressing the switch can act as a molecular detector or a timer, with possible applications in medical diagnostics or industrial bioprocessing. However, growth inhibition strategies for the timer that rely on protein overexpression are not mutationally robust, so other strategies may be required to increase circuit stability. We also show that the same switch architecture, tuned properly, produces noisy coin-flipping behaviors that could be used as a core circuit element to initiate cell differentiation for synthetic multicellular systems. Further investigation of bistable switches could enable construction of synthetic networks to generate behaviors observed in natural bistable networks [74], such as bet-hedging [75, 18].

4.6 Methods

Strains and media: Bistable switch library constructs were screened in 2.320 (CGSC accession number 6440) and assayed in strains 2.320 and BW25113 Δ*lacI*. Strains BW25113 Δ*lacI* was generated from JW0336-1 (CGSC accesstion number 8528) using plasmid pFLP2 [76] to excise the FRT-flanked kanamycin cassettes, followed by sucrose counterselection to cure the pFLP2 plasmid [69]. M9 minimal media (M8000, Teknova) supplemented with 50 μg/mL kanamycin was used for growth and fluorescence assays.

Microscopy for initial switch construct. We prepared the cells by inducing the culture to the green state in M9CA broth with 100 ng/mL aTc, incubated the culture overnight at 37°C, then removed the inducer via multiple washes, diluted the cells 1:100 in fresh M9CA media, and incubated them for two hours at 37°C. A small sample of this culture was added to a microscope slide and sandwiched to the surface with a 1% nutrient agar pad prepared with M9CA broth and 50 μg/mL kanamycin to observe single-cell dynamics.

rbSSR library generation via oligo assembly. The backbone for rbSSR-BSS library construction was prepared from a parent plasmid with a p15A replication origin by digestion with endonucleases AclI and SnaBI, followed by gel extraction and purification. A set of six primers (O1R, O2F, O3R, O4F, O5R, and O6F) were ordered (IDT) to encode the regulatory region and introduce rbSSR sequences. Six variant oligos encoding rbSSRs $(T)_{10,12,...,20}$ and $(A)_{10,12,...,20}$ were ordered for primers O1R and O6F, respectively. A short PCR amplicon from *tetR* was used to join the digested backbone to the oligos. Combinatorial libraries of all $(A)_n$ repeats for each $(T)_n$ repeat were generated using Gibson assembly with the backbone fragment, the PCR amplicon, and the proper mix of oligos, and transformed directly into expression strain 2.320 for screening. The regulatory regions between digestion sites for each library clone were verified by sequencing. A more detailed description of the method is found in the SI text.

The distributions of PCR-generated rbSSR libraries are not uniform, which may not be useful for some tuning applications. Encoding variable repeat lengths in oligonucleotides allows uniform, or arbitrary, sampling of rbSSR libraries. To uniformly explore the two-

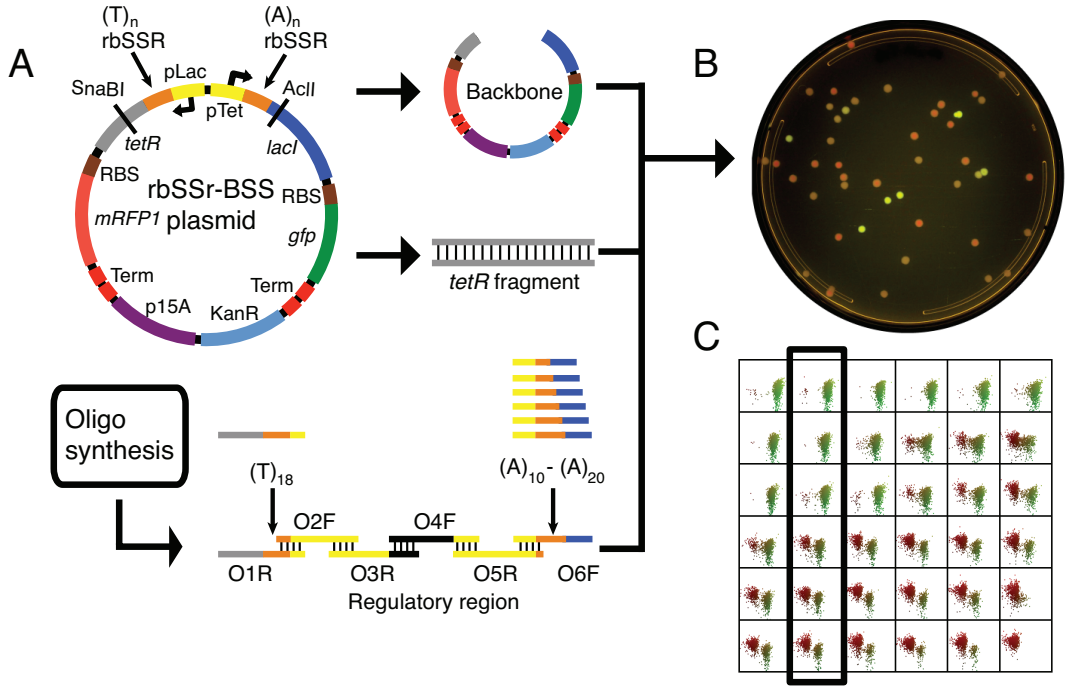

Figure 4.11: Construction of rbSSR-BSS library using oligo assembly, with the six-member $(T)_{18}$ rbSSR-BSS sub-library as an example. (A) A digested backbone, a fragment of the *tetR* gene, and the regulatory region–encoded by oligos–are assembled into plasmids using the Gibson assembly method. (B) Plate of $(T)_{18}$ sublibrary from Gibson assembly reaction with O1R-$(T)_{18}$ and all O6F oligo variants. Individual rbSSR variants were identified from this plate via flow cytometry and DNA sequencing by grouping colony types with similar green-to-red fluorescence distributions, followed by sequencing clones from each group. (C) $(T)_{18}$ strains from plate in (B) highlighted in grid of bistable switch library for strain 2.320.

dimensional rbSSR space for the bistable switch we used a method adapted from [44], here called oligo assembly, to reconstruct the regulatory regions of the two operons. To rapidly generate the 36-strain library, three parts were prepared: a backbone, a short segment of *tetR*, and the regulatory region containing the promoters and both RBS regions (see Fig. 4.11(A)). The backbone was prepared by digesting the bistable switch plasmid at SnaBI and AclI to remove the region encoding the promoters for each operon as well as the RBS sequences and portions of the *lacI* and *tetR* genes. We amplified a 293 bp fragment of the *tetR* gene from the parent plasmid using primers ssrBSS_cut and tetR_T10.

To reconstruct the regulatory regions and introduce rbSSR variation into the RBS se-

quences, we designed a set of six single-stranded DNA oligos: O1R, O2F, O3R, O4F, O5R, and O6F. Each oligo has twenty-nucleotide 5' and 3' overlaps to the *tetR* fragment, its neighbor oligos, or the digested parent plasmid. For oligos O1R and O6F, we ordered six variants encoding a set of rbSSR repeats: $(T)_{10}$, $(T)_{12}$, ..., $(T)_{20}$ for O1R to drive TetR sexpression; and $(A)_{10}$, $(A)_{12}$, ..., $(A)_{20}$ for O6F to drive LacI expression. Gibson assembly reactions were prepared at 20 μL volumes with 10-20 ng of the digested backbone, 10-40 ng of the *tetR* PCR product, and a mixture of the oligos each at a final concentration from 10-40 μM. These assemblies were transformed directly into the expression strain 2.320. To simplify the screening process, we assembled and transformed the plasmid library as sets of the six $(A)_n$ rbSSR constructs that correspond to one column of the grid plot in Figure 4.6. To accomplish this, we added all six O6F oligo variants to the Gibson assembly reaction at equimolar concentrations to match the concentration the other five oligos. This process is illustrated for the set of $(T)_{18}$ variants in Figure 4.11.

Oligo assemblies resulted in low transformation efficiencies compared to equivalent PCR assemblies for the bistable switch. To maximize the likelihood of sampling all expected rbSSR variants when performing oligo assemblies, Gibson reaction products were purified after incubation (Qiaquick PCR purification column, Qiagen), then chilled and added in full to a 40 μL aliquot of electrocompetent cells. Ice-cold molecular grade water was then added to a final volume of 160 μL. After mixing, the cell-DNA slurry was transformed in 40 μL aliquots, each rescued in 250 μL LB broth, and incubated at 37° C for 45 minutes. The individual transformations were then pooled and plated on selective media.

Cell growth and plate reader measurements. Freshly transformed colonies were selected at random from plates after incubation for 12-14 hours at 37°C, transferred in triplicate to 200 μL M9 minimal media in 96-well plates (Costar 3795), and grown to saturation overnight at 37° C in a shaker. The cultures were then diluted 1:100 in 200 μL prewarmed fresh broth in 96-well plates (Costar 3904), grown at 37°C to OD600 0.15 to 0.2 in a plate reader (Biotek) with shaking. Optical densities (600 nm), GFP measurements (485 nm excitation, 510 nm emission), and RFP measurements (590 nm excitation, 632 nm emission) were taken every 10 minutes. When grown to target density, 10 μL of each culture was

transferred to 100 μL 1x PBS (Gibco) with 34 μg/mL chloramphenicol chilled at 4°C.

Flow cytometry measurements. Diluted cultures from plate reader measurements were transferred to a flow cytometer (C6 with CSampler, Accuri). To prevent well-well contamination, blank wells containing PBS were briefly read after each sample well. GFP measurements (488 nm excitation, 533 nm emission) and RFP measurements (488 nm excitation, 610 nm emission) were recorded for 50,000 events per sample. Cells were gated using a rectangular gate in forward scatter and side scatter. Fluorescence levels in both the GFP and RFP channels were used to generate scatter plots for the rbSSR-BSS variants. Cytometry measurements were analyzed using custom Matlab scripts (see SI text). For color compensations, 4.5% of the red channel fluorescence was subtracted from the green channel fluorescence, and 7.8% of the green channel fluorescence was subtracted from the red channel fluorescence. These percentages were determined using experimental data from strains with plasmids expressing only RFP or GFP, respectively. Cutoffs from color-compensated levels of 500 and 750 were used for classifying cells as "red" or "green", respectively.

Long-term stability of rbSSR-BSS. To assess the stability of rbSSR-BSS library variants, we measured fluorescence distributions over 96 hours via flow cytometry for multiple constructs expressed in each strain background. We transformed rbSSR-BSS variants $(T)_{18}/(A)_{10}$, $(T)_{12}/(A)_{12}$, $(T)_{16}/(A)_{14}$, and $(T)_{12}/(A)_{20}$ each into strains 2.320 and BW25113 $\Delta lacI$, isolated three colonies from each plate after 12 hours of growth on selective LB media, and suspended each colony in 50 μL sterile PBS. We induced cultures into the TetR- and LacI-dominant switch states by seperately inoculating 10 μL from each PBS culture into 0.5 mL selective LB media with 1 mM IPTG and 100 ng/mL aTc, respectively. These cultures were incubated overnight in a 96×2 mL deep-well plate (Eppendorf), spun down and washed two times with 1.2 mL sterile PBS to remove the inducers, then diluted 100:1 into 200 μL M9 2.0% glucose media supplemented with kanamycin. Cultures were sampled immediately ($t = 0$), then sampled every 12 hours to observe fluorescence distributions and diluted between 50:1 and 200:1 to maintain the cultures in exponential growth. The percentage of green cells for each sample is determined as described in the flow cytometry methods. The results are shown in Figure 4.9.

Fluorescence microscopy for rbSSR-BSS variants. I prepared cells for characterizing switching behaviors of the bistable switch and for the supplementary movies by transforming strain 2.320 with a plasmid from the rbSSR-BSS plasmid library as noted, distributing the rescue culture on LB agar plates, and incubating the plates at 37° C for 12-14 hours. Individual colonies were then resuspended in 500 μL M9CA broth (M8010, Teknova) supplemented with 50 μg/mL kanamycin, vortexed to break up cell aggregates, then 1 μL of the culture was added to a glass-bottomed microscopy dish (GWSB-3512, WillCo Wells BV). The cells were sandwiched to the dish surface by a 1% nutrient agar slab prepared with M9CA broth and 50 μg/mL kanamycin. Cell samples were then transferred to a Nikon Ti-S inverted microscope (Nikon Instruments) equipped with a Coolsnap-HQ camera (Roper Scientific) and an environmental chamber (In Vivo Scientific). For microcolony snapshots, the cells were incubated in microscopy dishes at 32°C for 10-12 hours and then imaged in phase contrast, GFP, and RFP channels. For movies, frames for 25 individual microcolonies were recorded every 12 minutes in the same channels while the samples were incubated at 32°C for 10-12 hours.

Time-lapse switch movies. The three time-lapse movies submitted electronically with the dissertation show the growth trajectories and switch states for microcolonies generated from a single colony of the $(T)_{20}/(A)_{20}$ rbSSR-BSS variant in strain 2.320. The movies demonstrate that both states are stable, and that for a few rbSSR-BSS variants in strain 2.320 switching rates were high enough to observe switching events within microcolonies. The scale bar for each movie is approximately 5 μm. These movies can be viewed at higher resolution at `http://depts.washington.edu/soslab/rbSSR`.

Movie 1: Green microcolony for (T)20/(A)20 rbSSR-BSS variant in strain 2.320.

Movie 2: Divergent microcolony for (T)20/(A)20 rbSSR-BSS variant in strain 2.320.

Movie 3: Red microcolony for (T)20/(A)20 rbSSR-BSS variant in strain 2.320.

Genotypes of *lacI*$^-$ strains. Strain 2.320 is a Monod strain with the following genotype: F-, *lacZ13*(Oc), *lacI22*, *rpsL135*(strR), *malT1*(λ^R), *xyl-7*, *mtl-1*, *thi-1* [68]. Strain

BW25113 Δ*lacI* is a Keio collection strain with the following genotype: F-, Δ*(araD-araB)567*, Δ*lacZ4787*(::rrnB-3), Δ*lacI785::kan*, λ^-, *rph-1*, Δ*(rhaD-rhaB)568*, *hsdR514* [69].

Growth rates for rbSSR-BSS library in expression strains. To observe the effect of a given rbSSR-BSS plasmid on the growth rates for each strain, we fit plate reader absorbance data to an exponential function to derive doubling times. We used Mathematica function "FindFit" to determine the least-squares fit of the absorbance data to the growth model $OD_{600} = \alpha + \gamma e^{\beta t}$ where α is an offset, β is the growth rate, and $\gamma = 0.033$ is a conversion factor to map the plate reader absorbance measurements to OD600. The doubling time fits for the entire rbSSR-BSS library in both strains, calculated as Log[2]/β, are summarized in Table 4.1. While we observed reduced growth rates for several LacI-dominant 2.320 strains with short $(A)_n$ repeats, we found many of the TetR-dominant BW25113 Δ*lacI* strains with short $(T)_n$ repeats grew slowly. The reduced growth rate in these states is likely due to the toxicity of over-expressed repressor proteins.

Chapter 5

ENHANCING EVOLVABILITY: CONTROLLING SSR MUTATION RATES IN VIVO

Although simple sequence repeats embedded in RBS spacers have not been reported in any natural context, these repeat sequences are known to accelerate evolutionary adaptation when situated within coding sequences and other regulatory regions [77]. SSRs undergo insertion/deletion mutations at rates four to five orders of magnitude higher than arbitrary sequences of the same length [59, 60].

Some bacteria utilize SSR variability to strictly control protein expression via frameshift errors in coding sequences [78] or to alter transcription rates via insertions or deletions to promoter spacers [79]. Repeats embedded in promoter and gene coding sequences are also responsible for environmental adaptations in many higher organisms [77], including tuning circadian rhythms from fungi [80] to flies [81].

When used in synthetic gene circuits, the instability of SSRs could be detrimental if it caused the performance of a circuit to degrade or de-tune over time. On the other hand, repeat variation could be a powerful tool to optimize circuit performance through directed evolution by focusing mutations to regions of the circuit that have strong and predictable effects on gene expression.

One simple example of how these repeats enhance evolvability in natural systems is the *lac* repressor, which contains the tandem repeat $(CTGG)_3$ in its coding sequence. Insertion/deletion mutations of a single repeat unit occur at this site at rates 1000-fold higher than arbitrary mutations [82] and result in frame shift errors that produce inactive, truncated LacI monomers. Cell populations grown in lactose minimal media over hundreds of generations are highly enriched for these mutations [83]. Expanding from the strict on/off switch illustrated here, the combination of focused variation for SSRs with predictable results when utilized at tunable elements of gene networks, has potential to significantly

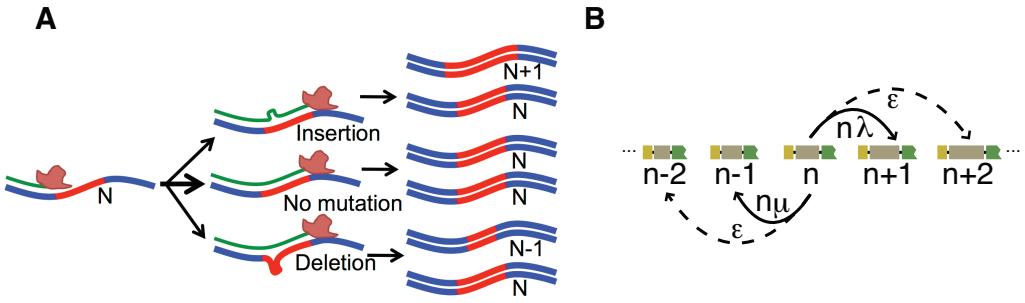

Figure 5.1: Mechanism and model for slipped-strand mispairing mutations. (A) Cartoon depicting insertion/deletion mutations via slipped-strand mispairing through DNA looping and the absence of DNA repair. Replication events are represented by black arrows, DNA polymerase is represented by the pink shape, the nascent DNA replicon is the green line, the DNA template strand is blue with the repeat sequence in red, and the number of repeats is indicated for the initial strand and the final strands. (B) Simple birth-death process model for SSR variation through slipped-strand mispairing. The probability of repeat unit insertion is $n\lambda$ and repeat unit deletion is $n\mu$ per repeat unit per generation, given a replication template of repeat unit length n. The probability ϵ that two deletions or insertions occur in one generation is assumed to be zero.

accelerate the optimization of gene networks through directed evolution.

5.1 Slipped-strand mispairing mechanism and model

Simple sequence repeat mutations primarily occur via an error-prone process in DNA replication termed slipped-strand mispairing (SSM) [84]. As DNA polymerase replicates tandem repeat sequences, the nascent strand can hybridize with complementary bases on the template strand in multiple locations. An unpaired DNA loop is generated when the nascent strand binds an unaligned complementary sequence, which is propagated as an insertion or deletion mutation to subsequent generations if the unpaired loop is not corrected before the next replication event by the cell's mismatch repair machinery (Fig. 5.1(A)).

To estimate the rates of insertion and deletion mutations at rbSSR sequences in vivo, we modeled the repeat variation as a birth-death process with insertion rate λ and deletion rate μ as shown in Figure 5.1(B). For n repeat units the rbSSR insertion rate per replication event is $n\lambda$ and the associated deletion rate is $n\mu$. Each rate has units of "mutations

per replication event per base pair of SSR sequence." The distribution of repeats after m replication cycles for a given initial repeat distribution π_0 is

$$\pi_m = \pi_0^T \mathbf{P}^m$$

where

$$\mathbf{P} = \frac{1}{2}\begin{pmatrix} 1-\mu & \mu & 0 & \dots \\ 2\lambda & 1-2\lambda-2\mu & 2\mu & \dots \\ 0 & 3\lambda & 1-3\lambda-3\mu & \dots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix} + \frac{1}{2}\begin{pmatrix} 1 & 0 & 0 & \dots \\ 0 & 1 & 0 & \dots \\ 0 & 0 & 1 & \dots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

and π_0 is the initial distribution.

5.2 Long-term sequence drift for $(A)_{15}$ rbSSR-GFP

To examine the long-term stability of rbSSR sequences, we quantified the sequence drift of the rbSSR-GFP gene circuit using DNA sequence trace analysis of serially-passaged wild-type and mutator strains over approximately 220 generations. Specifically, we transformed the $(A)_{15}$ rbSSR-GFP plasmid into wild-type strain BW25113 and mutator strain BW25113 Δ*mutS*, which is mismatch repair deficient, and cultured each strain in triplicate over 16 serial passages. We extracted plasmid DNA from each overnight culture for sequence trace analysis (see Fig. 5.6). For the wild-type strain, there were no mutations in the *gfp* gene, including the regulatory and rbSSR sequences (Fig. 5.2(A)), which demonstrates that rbSSR sequences can remain stable for very long periods. However, a strong bias for SSR deletions was observed when the plasmid was propagated in the mutator strain, as the fraction of $(A)_{14}$ plasmids increased steadily over time (Fig. 5.2(B)).

5.3 Estimating mutation rates from the birth-death model

We fit a model of insertion/deletion mutations based on the birth-death process to the data in Figure 5.2 with an initial π_0 of zero except for the 15^{th} entry, which is 1, for $(A)_{15}$ rbSSR-GFP. From this fit, we inferred the mutation rates to be 2.6×10^{-4} deletions and 5.1×10^{-5} insertions per base pair of repeat sequence per generation which are within the

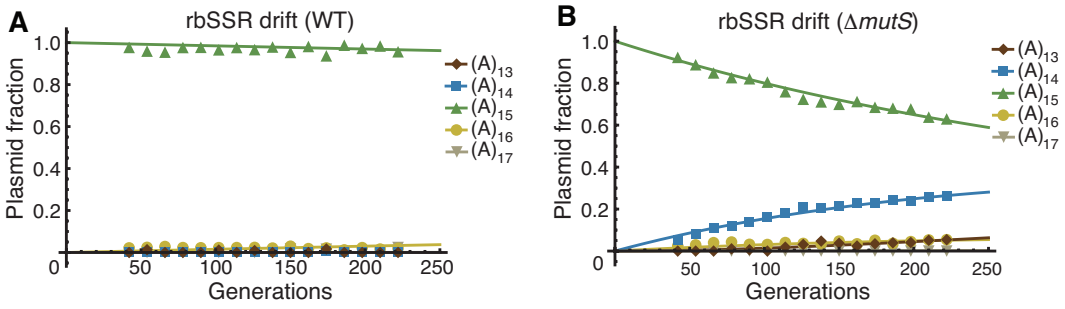

Figure 5.2: SSR mutational drift for $(A)_{15}$ rbSSR-GFP in wild-type and mutator strains. (A) Distributions of $(A)_{13}$ to $(A)_{17}$ rbSSR repeats in the plasmid population as a function of estimated cell generation for wild-type strain BW25113. The distribution of rbSSR repeats was essentially unchanged after 220 generations. Curves are least-absolute-deviations fits to the simple birth-death model. (B) Distributions as in (A) for strain BW25113 Δ*mutS*. The fraction of plasmids with $(A)_{15}$ steadily decreased as the fraction of plasmids with a single-unit deletion increased. The fraction of plasmids with a single-unit insertion also increased, though at a slower rate.

range of reported rates [59]. While we observed no mutations in the promoter or the *gfp* coding sequence for the mutator strain, we did find that rbSSR instability varied among replicates (Fig. 5.4) with one replicate resulting in a final distribution of plasmids with repeats $(A)_{13}$ through $(A)_{17}$. Note that these experiments were performed without any intentional selection, and the results are likely primarily due to drift.

Mismatch repair mutants are reported to, in general, have mutation rates 10 to 100-fold higher than wild-type cells [85]. Since no SSR mutations were observed for the wild-type cells in the rbSSR-GFP drift assay, we used the parameter fits from our model to estimate the minimum difference in mutation rates between the *mutS*-knockout and the wild-type strain 5.3. From these predictions, we inferred the SSR mutation rate in the mutator strain to be at least 20 times greater than in the wild-type strain, which is consistent with previous work [60].

5.4 Building and characterizing an inducible mutator strain

A goal of this work is to optimize genomically encoded gene networks in place using selective pressure. Doing this work in mutator strains is not desirable, since deleterious mutations

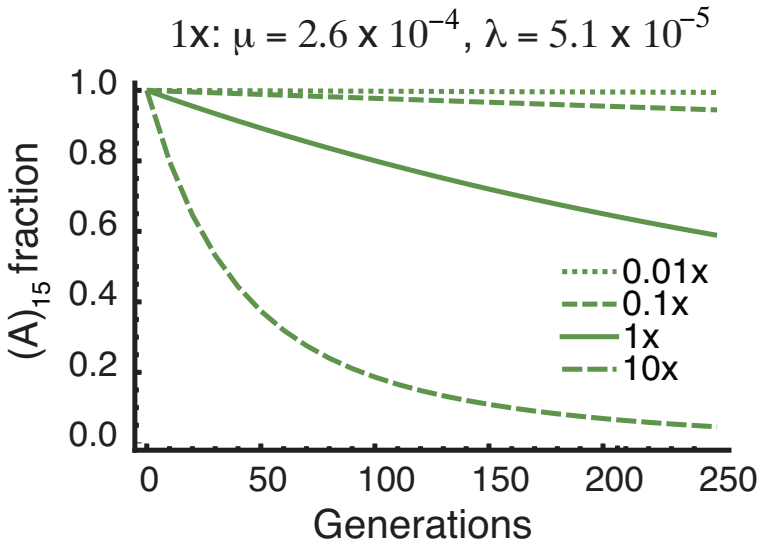

Figure 5.3: Predictions for the fraction of $(A)_{15}$ plasmids over 250 generations for multiple mutation rates, with 1x corresponding to the rates obtained from a fit of the data from Figure 5.2(B) to the birth-death model. Mutation rates of 0.05x or below fit the data in (A) for the wild-type strain, suggesting that the mutator strain is at least 20 times more likely than wild-type to insert or delete a repeat.

are likely to build up over time and hypervariation should cease once an evolved strain with the desired behaviors emerges. A better scenario is to tune networks in vivo using an inducible mutator. An inducible mutator would enable temporal control over mutation rate to allow dynamic switching between induced periods of hypervariation and uninduced periods of competition.

As an alternative to knocking out mismatch repair genes, mutation rates can also be increased by overexpressing DNA-adenine methyltransferase (Dam). Cells overexpressing Dam methylate newly replicated DNA at a faster rate than wild-type cells, upsetting the balance between methylation and mismatch repair[86] (Fig. 5.5(A)). Overproduction of Dam has been shown to generate mutator phenptypes with mutation rates 18 [87] to 30 [86] fold higher than wild-type cells. Therefore, SSR mutations generated via replication errors are more likely to become protected from mismatch repair when Dam is overexpressed.

We built an inducible mutator plasmid by cloning the native *dam* gene from *E. coli*,

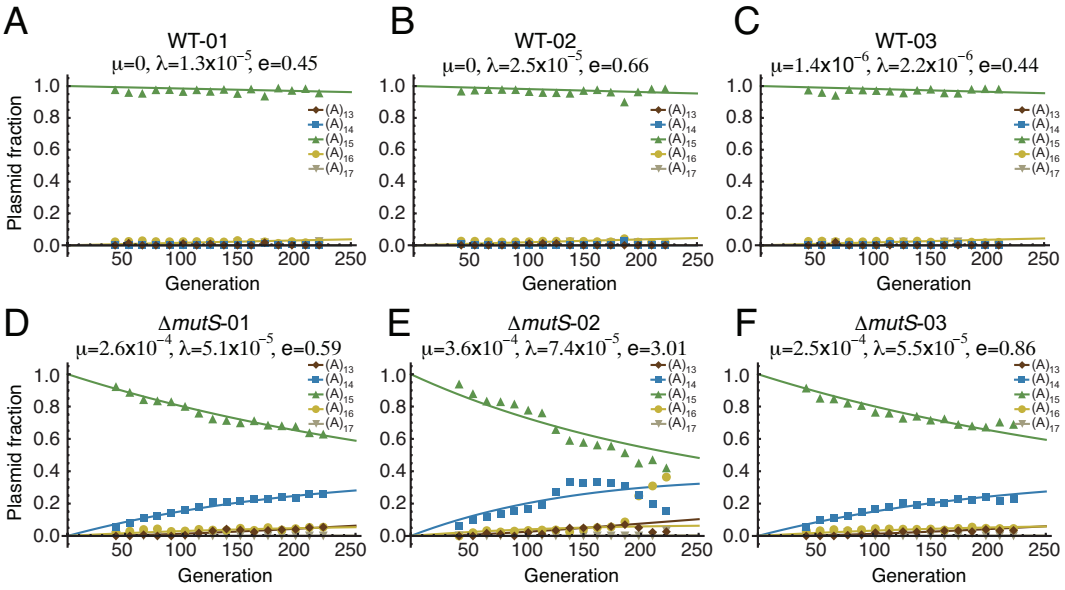

Figure 5.4: Distributions of rbSSR sequences in evolution study for wild-type and mutator strains. Panels (A), (B), and (C) show the rbSSR distributions over 16 serial passages for the three replicates of the wild-type strain. Panels (D), (E), and (F) show the rbSSR distributions for the three replicates of the mutator strain. The best fits for the mutation rates are shown for each replicate along with the error of the fit to the data. The 16 passages correspond to approximately 220 generations.

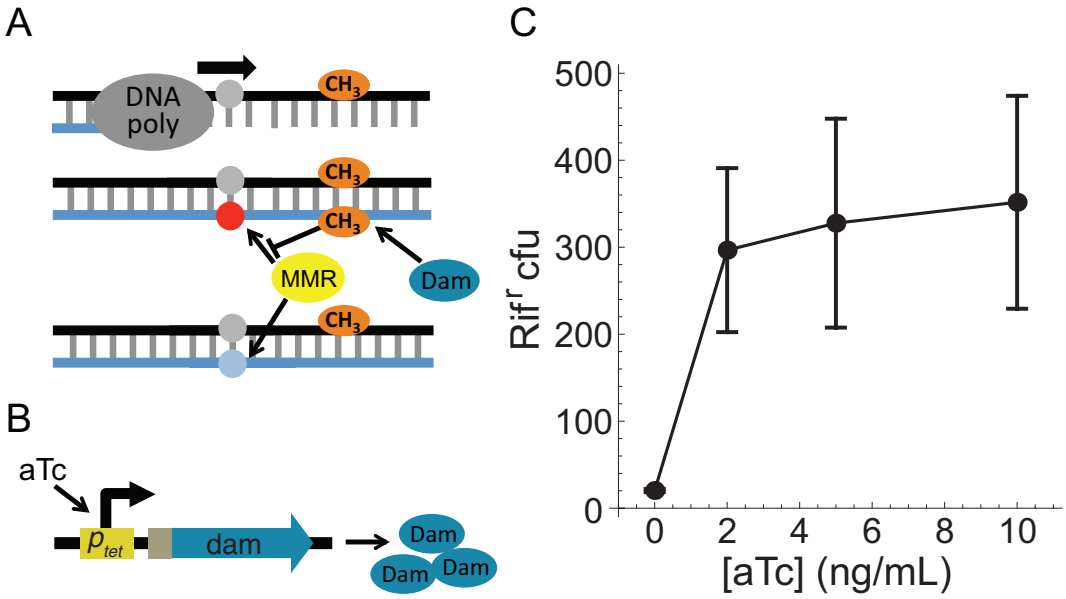

Figure 5.5: Inducible mutator construct and characterization. (A) Hypermethylation mechanism. Overexpression of DNA-adenine methyltransferase (Dam) increases methylation of newly replicated DNA, which increases mutation rates by preventing the mismatch-repair system from correcting replication errors. (B) The inducible hypermethylation construct, under TetR control, is inducible by aTc. (C) Number of rifampicin-resistant colonies as a function of inducer level for the inducible mutator strain using a low-copy plasmid. Induced cultures exhibit mutation rates approximately 15-18 fold higher than uninduced cultures, consistent with previous work [86, 87]. Error bars represent standard error from three independent cultures.

adding a strong RBS, and placing it under the control of TetR (Fig. 5.5(B)). We tested the activity of the inducible mutator gene on a high copy plasmid and a low copy plasmid in an EcNR1-derivative strain [45] that expresses TetR from the genome. Both plasmids enhanced mutation rates as indicated by rifampicin resistance assays, following Ref. [88]. However, induced cells with the high copy plasmid grew slowly, suggesting toxicity from Dam overexpression. We found that intermediate levels of induction with the low-copy plasmid enhanced mutation rates from 15 to 18 times higher than uninduced populations (Fig. 5.5(C)).

5.5 Discussion

This chapter demonstrates that rbSSR repeats are suitably stable for many gene networks and highlights the potential for using rbSSRs, and SSRs in general, to direct the evolution of gene networks in vivo. First, rbSSR sequences can be stable over very long periods in a wild-type strain. This is important for circuits such as the rbSSR-BSS library studied in Chapter 4. When expressed in mutator strains, the repeats are readily destabilized for even the GFP drift circuit. It is possible that rbSSR tuning could be used to explore parameter space and prototyping circuit designs in place with the desired expression strain. Once a suitable mutant is identified via selection, the repeat sequences could be altered by adding non-repeat nucleotides resulting in a degenerate SSR which is likely to have much lower rates of mutation.

Much of the potential of this work is the idea of harnessing the focused variation of rbSSRs with its strong effects on gene expression to rapidly tune gene networks in vivo. In Chapter 6, we briefly describe some progress in this vein of research using the inducible mutator plasmid introduced in this chapter to enhance mutation rates for cells grown in lactose minimal media that have rbSSR-derived lactose utilization deficiencies.

One additional future thrust for this research is to better characterize differential mutation rates between SSRs and arbitrary sequences. Finding genetic constructs or environmental conditions that further enhance SSR mutation rates while leaving non-repeat mutation rates as close to wild-type levels as possible would enable broad and simultaneous sampling of mutations at multiple loci in growing cell populations.

5.6 Methods

Strains and media: Serial passage experiments for in vivo mutation rate analysis were carried out in wild-type BW25113 (CGSC accession number 7636) and mutator BW25113 $\Delta mutS$ strains. Strain BW25113 $\Delta lacI$ and BW25113 $\Delta mutS$ was generated from JW2703-2 (CGSC accession number 10126) using plasmid pFLP2 [76] to excise the FRT-flanked kanamycin cassettes, followed by sucrose counterselection to cure the pFLP2 plasmid [69]. Serial passages were performed in LB broth (10 g/l tryptone, 5 g/l yeast extract, 10 g/l

NaCl) supplemented with 50 μg/mL kanamycin.

rbSSR variation in vivo. Freshly transformed colonies of wild-type and mutator strains containing the $(A)_{15}$ rbSSR-GFP plasmid were transferred in triplicate to 6 mL of LB supplemented with kanamycin and grown to saturation overnight at 37°C in a shaker. Overnight cultures were mixed and 1 μL transferred into 6 mL fresh broth, which was grown to saturation. Plasmid minipreps from overnight cultures were prepared, sequenced, and the sequencing traces were analyzed as described below. This process was followed through 16 passages.

Chromatogram processing of in vivo rbSSR-GFP passages. Chromatogram trace data from sequencing reactions were converted to csv files with the `abiparser.py` Python script (`http://www.bioinformatics.org/groups?group\_id=497`) and imported into *Mathematica* for analysis. Trace data were separated by nucleotide identity into individual channels and four isolated bases – A_{30}, T_{99}, T_{115}, and G $_{143}$ – were selected for calculating repeat distributions. Each isolated nucleotide was located at least five bases from the nearest nucleotide of the same type (Fig. 5.6). For each isolated nucleotide, peak heights over a 7-peak window were normalized to calculate a distribution of repeats $(A)_{13}$ to $(A)_{17}$. The mean of these four distributions was used to populate the serial passage dataset.

Note that an apparent 5–8% contribution from a single-unit deletion was consistently observed for the $(A)_{15}$ sequence traces. This noise, possibly an artifact of polymerase slippage in the sequencing reaction, was observed in all rbSSR $(A)_{15}$ sequencing traces, including the initial sequence verification of the $(A)_{15}$ rbSSR-GFP construct and all passages of the wild-type strain across three replicates, with a mean contribution of 6.6% ($n = 50$). To compensate for this noise we adjusted the peak height P_n for repeat length n to $\hat{P}_n = P_n - 0.066P_{n+1}$ for each target nucleotide when computing distributions.

Fitting rbSSR-GFP drift data to birth-death model parameters. We used the serial passage dataset from the mutator strain to fit rates λ and μ from the birth-death model. We fit the data using the *Mathematica* function "NMinimize" to find the least absolute deviations fit using the peak distribution data, equally weighted for all repeats from $(A)_{13}$ to $(A)_{17}$. We assume for this model that the multiple-repeat jump rate ε is zero and that 30 generations pass from plating to the inoculation of first passage.

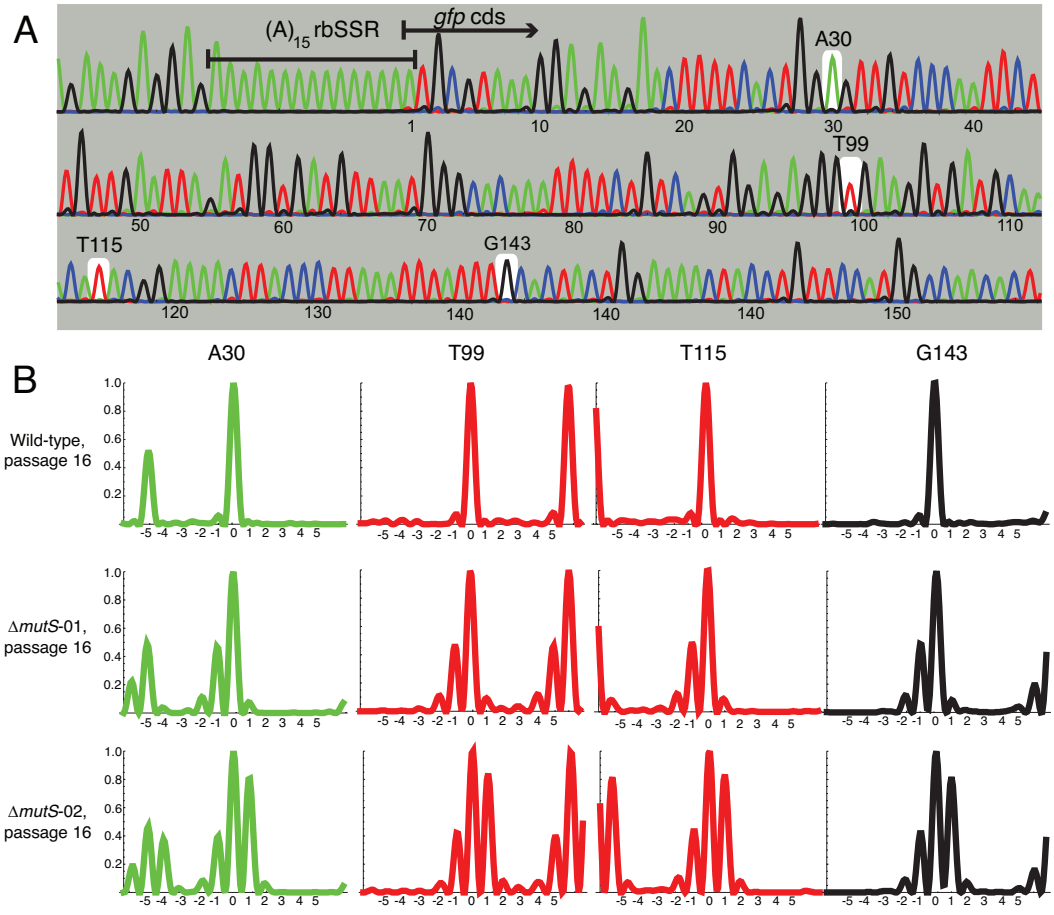

Figure 5.6: Chromatogram trace analysis for SSR drift. (A) Sequence trace of $(A)_{15}$ rbSSR-GFP, highlighting rbSSR region and the four isolated nucleotides used to calculate repeat distributions: A_{30}, T_{99}, T_{115}, and G_{143}. (B) Representative traces of the four target sites from the 16^{th} passage of the wild-type strain and replicates one and two of the mutator strain.

Chapter 6

APPLICATIONS AND EXTENSIONS OF RBSSR TUNING

6.1 Self-destructive altruism: cellulase

The self-destructive altruism: cellulase (SDAc) circuit encodes a division of labor [89] in *E. coli* between producer cells and altruist cells (Fig. 6.1). Producer cells act as immortal germ cells capable of reproduction; altruist cells act as soma cells that survive only a few generations to build up and release a protein payload that breaks down a complex nutrient source – cellulose – into digestible units – cellobiose – to feed the producer population. The system will likely work as expected if the cells switch from the producer to the altruist state at a sufficient rate, and if the altruists continue to release sufficient cellulase to support a growing population of producers. Proper implementation of this gene network should address a major challenge in consolidated bioprocessing [90, 91, 92], which is to export cellulase at levels to promote rapid growth on cellulose. This network could also be used to investigate gene networks sufficient to produce multicellular behaviors and the evolution of division of labor in these multicellular systems [93, 94].

We have created a simple model to reason about the range of behaviors for the SDAc circuit. The model captures the stochastic dynamics of switching between cell types, production of cellulase and lysis protein in the altruistic state, cellulase release when lysis protein levels cross some threshold, and nutrient production as extracellular cellulase breaks down cellulose. In order for the system to work properly, it is clear that the switch and the lysis gene must be tuned properly. If the switch is biased too strongly towards the producer state, the cell population will fail to export sufficient cellulase to convert cellulose into cellobiose. If, on the other hand, the switch is biased too strongly to the altrusist state, the cell population will fail grow robustly and is likely to crash. For the lysis gene, if the altruist cells produce too little lysis protein, the cells will not lyse and the extracellular cellulase levels will be insufficient for robust growth; excessive production of the lysis protein could

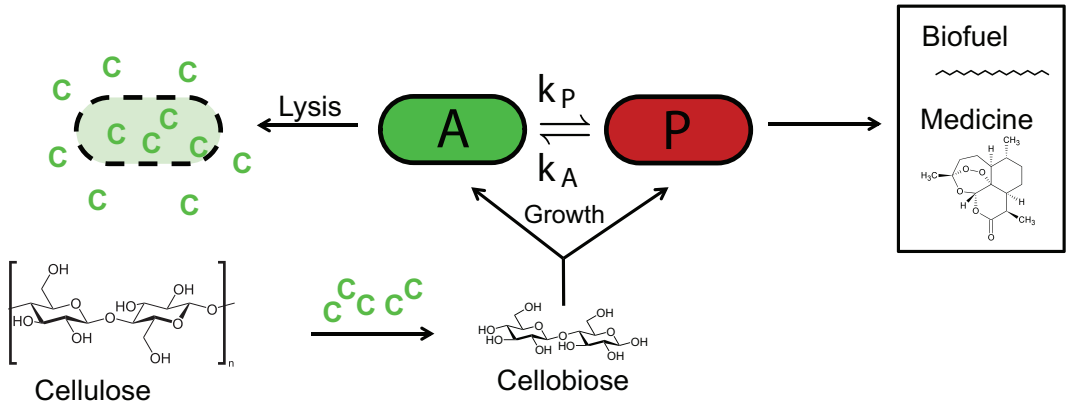

Figure 6.1: Expected behaviors for self-destructive altruism: cellulase circuit. Cells switch randomly between producer (P) and altruist (A) states. Cells in the altruist state built up cellulase payload (C) and lyse to export cellulase to the environment as a public good. Cellulase breaks down cellulose to promote growth on cellobiose. Producer cells express genes for a high-value commodity such as biofuel or pharmaceuticals.

lyse the altruist cells before substantial cellulase is produced.

We have implemented a version of this model in the simulation platform `gro` [95], which provides an environment for simulating stochastic gene expression at the single cell level – an important characteristic for observing switching events between switch states and probabilistic lysis. The switch tuning scenario is depicted in Figure 6.2 for three combinations of expression levels for the two repressor proteins, which are parameters in the network that can be controlled by rbSSRs.

The DNA implementation of the SDAc circuit is shown in Figure 6.3, which uses the rbSSR-BSS library from Chapter 4 and a second plasmid under control of TetR that coexpresses a cellulase gene with a very strong RBS and a lysis gene[96] under initial control of a weak $(AT)_{12}$ rbSSR. The behavior of this system is likely to differ from the original rbSSR-BSS system because of the titration effect of additional *tet* operator sites on the cellulase/lysis plasmid, suggesting that higher expression of TetR is required for the SDAc circuit to result in equivalent switching behaviors to the rbSSR-BSS circuit. We first built a prototype of the system to demonstrate that the cells can switch between states at a high enough rate to observe mixed-state microcolonies, and to show that cells in the altruistic

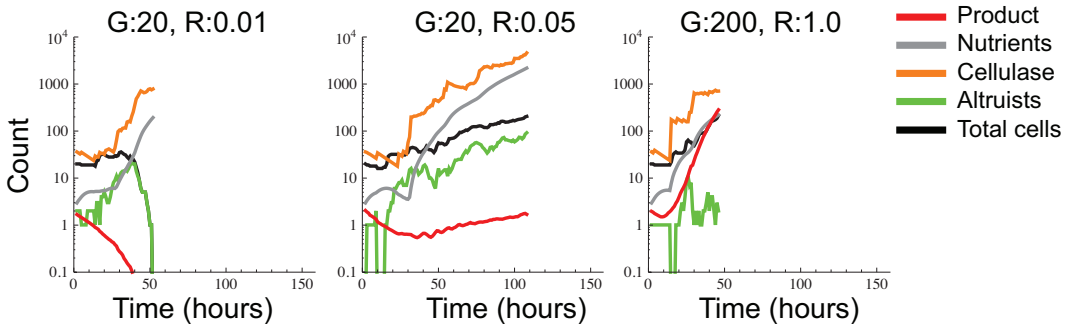

Figure 6.2: SDAc `gro` simulations for three combinations of production rates for each repressor protein. Each panel is a plot showing the time-varying populations of important species in the SDAc circuit, including the total number of cells (black), the number of altruists (green), the cellulase level (orange), nutrient level (gray), and the amount of product, represented by RFP levels (red). Each sample trajectory was initialized with 20 cells, each with random concentrations of each repressor protein. The simulation is also initialized with nominal concentrations of digestible nutrients and cellulase enzyme. The simulations ended when populations reached 200 cells, depleted all available nutrients, or all cells lysed.

state (expressing GFP) lyse after accumulating a high concentration of GFP acting as a proxy for cellulase protein. We found that rbSSR-BSS variant $(\mathrm{T})_{10}/(\mathrm{A})_{12}$ cotransformed with the cellulase/lysis plasmid produced the desired microcolony behaviors (Fig. 6.4).

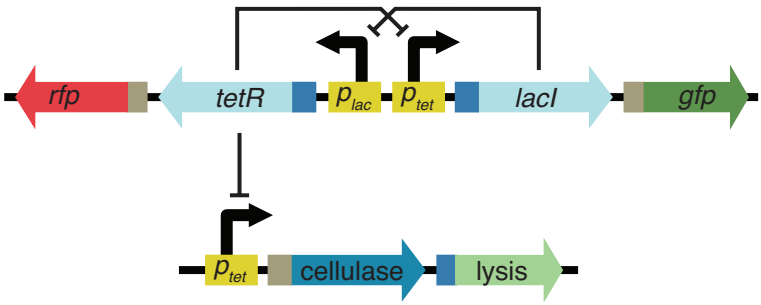

Figure 6.3: SDAc plasmid constructs. Above is the rbSSR-BSS gene network. Below is the cellulase/lysis operon under control of a TetR-regulated promoter with a strong RBS and $(AT)_n$ rbSSR driving the expression of the cellulase and lysis genes, respectively.

In order for cells transformed with the SDAc system to grow, they must digest cellobiose. To enable *E. coli* growth on cellobiose, we reengineered the native *chb* operon, which can digest cellobiose but is only expressed in the presence of chitobiose[97], by replacing the native promoter with a strong constitutive promoter using ssDNA recombination [45]. The parent strain EcNR1, however, behaves similarly to strain BW25113 $\Delta lacI$ from the rbSSR-BSS library with low switching rates for the rbSSR-BSS plasmids. In order to increase switching rates, we transferred the rbSSR-BSS library to low-copy plasmids with a pSC101 replication origin. We also synthesized three cellulase genes that have been shown to be active when produced in *E. coli*[91, 92, 98]. These genes were synthesized from 120 to 160 nucleotide oligos supplied by OligoCo, Inc.

We cotransformed the low-copy rbSSR-BSS library with the plasmid library of cellulase/lysis variants into the cellobiose-digesting strain. The switches vary the strength of each repressor protein which controls the natural bias of the switch and its switching rate; the cellulase variant defines the catalytic efficiency of cellulose degradation; the strength of lysis expression determines how quickly the cells lyse which affects extracellular cellulase levels. To sample library variants with initial bias to the altruistic state, we plated the transformants with IPTG to bind LacI and force cells to the TetR dominant state. We then grew 96 colonies overnight in uninduced media to restore switch balance and spotted cultured on LB plates with 0.1% carboxymethylcellulose to observe growth and cellulase activity. Cellulase activity can be observed by white clearings on otherwise red plates stained with Congo Red [92].

An ideal outcome for assaying a colony would be a large, red colony with a large Congo Red clearing, suggesting that cells entering the altruist state are readily lysing and exporting high concentrations of cellulase. Mixed red/green or primarily green colonies may indicate the lysis gene isn't expressed strongly enough in these variants. These mixed colonies may be more likely to produce cheater phenotypes. The results for the plate assay are shown in Figure 6.5.

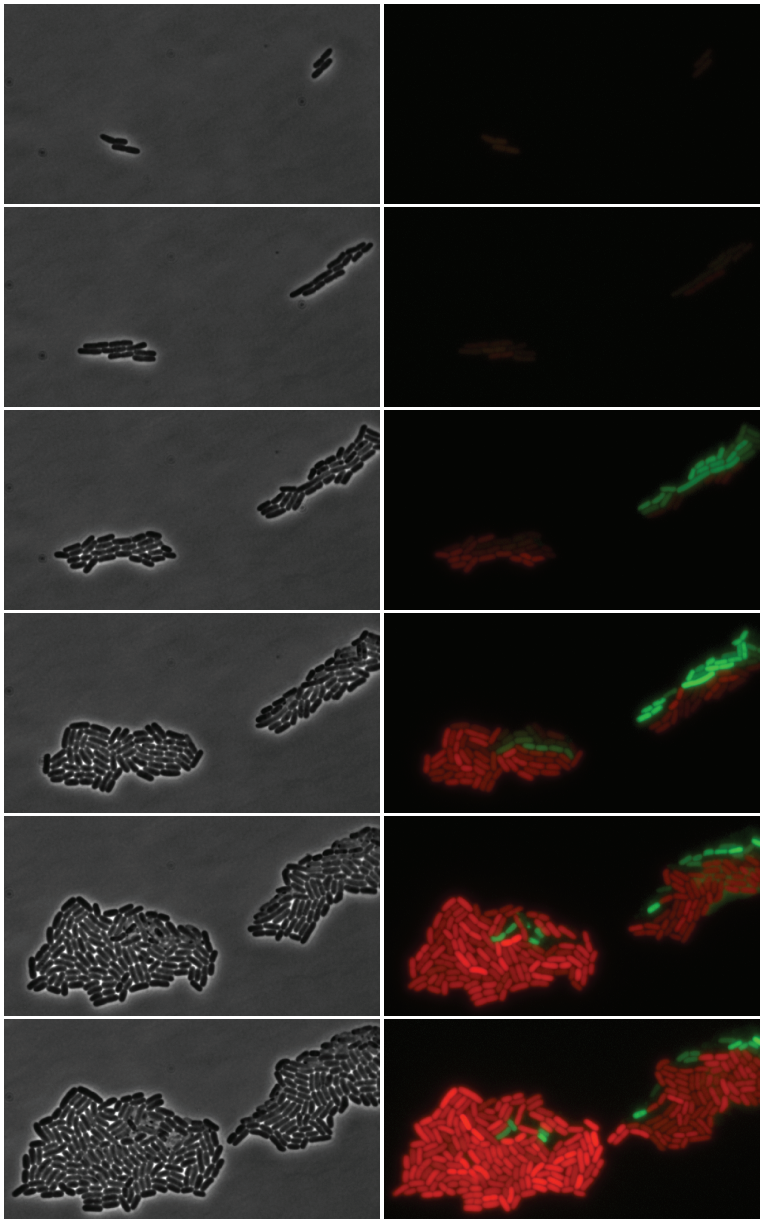

Figure 6.4: Time-lapse movie of microcolony growth, state switching, and lysis of 2.320 cells transformed with the SDAc circuit. Cells that switch to the altruistic state (green), built up GFP levels over multiple generations, followed by lysis. These microcolonies were grown in the presence of glucose, so no effects of cellulase release are expected. The observation of GFP release and perdurance in the fifth frame suggests that this system could be useful for studying self-destructive altruism in the context of local microbial communities.

A

	1	2	3	4	5	6	7	8	9	10	11	12
A	T??/A??	T??/A??	T??/A??	T18/A??	T18/A??	T16/A??	T16/A??	T16/A??	T14/A??	T14/A??	T12/A??	T12/A??
B	T??/A??	T??/A??	T??/A??	T18/A??	T18/A??	T16/A??	T16/A??	T16/A??	T14/A??	T14/A??	T12/A??	T12/A??
C	T??/A??	T??/A??	T??/A??	T18/A??	T18/A??	T16/A??	T16/A??	T14/A??	T14/A??	T14/A??	T12/A??	T12/A??
D	T??/A??	T??/A??	T??/A??	T18/A??	T18/A??	T16/A??	T16/A??	T14/A??	T14/A??	T14/A??	T12/A??	T12/A??
E	T??/A??	T??/A??	T18/A??	T18/A??	T18/A??	T16/A??	T16/A??	T14/A??	T14/A??	T14/A??	T12/A??	T12/A??
F	T??/A??	T??/A??	T18/A??	T18/A??	T18/A??	T16/A??	T16/A??	T14/A??	T14/A??	T12/A??	T12/A??	T12/A??
G	T??/A??	T??/A??	T18/A??	T18/A??	T18/A??	T16/A??	T16/A??	T14/A??	T14/A??	T12/A??	T12/A??	T12/A??
H	T??/A??	T??/A??	T18/A??	T18/A??	T16/A??	T16/A??	T16/A??	T14/A??	T14/A??	T12/A??	T12/A??	T12/A??

B

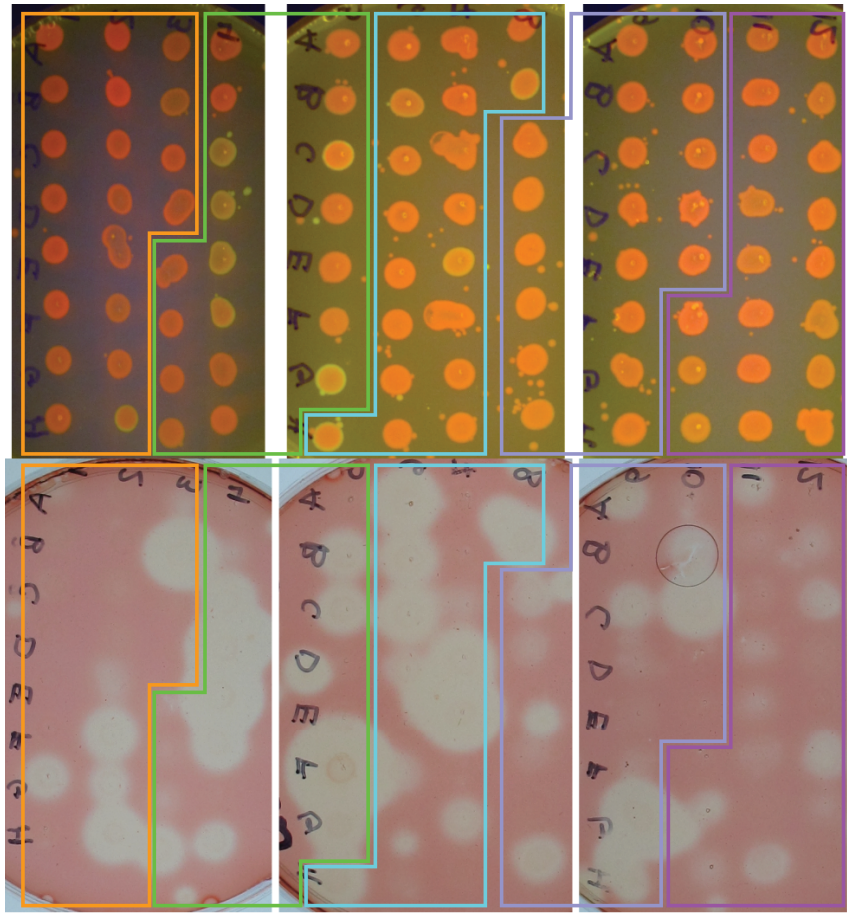

Figure 6.5: Activity assay for SDAc combinatorial rbSSR library. (A) Identities of SDAc library variants for each well of a 96 well plate. The colors indicate the rbSSR strength for the TetR gene. All variants have a random rbSSR strength for the LacI gene. The cultures in each well also have one of the three cellulases along with a random expression strength for the lysis gene. (B) Fluorescence as captured on a blue-light transilluminator (top). Red populations and mixed red-green populations are observed. The results of the Congo Red assay are shown at the bottom. Large clearings indicate strong extracellular cellulase concentrations.

A few of the colonies from the library assay have encouraging phenotypes. Colonies F5 and C10 are of particular interest. Each have primarily red cells as well as large cellulose clearings. This may indicate that the cells are switching state and releasing large amounts of cellulase, but also that the lysis state is sufficiently lethal to prevent long growth periods in the altruist state. This behavior may minimize cheaters. The next step for this project is to test how well these variants grow on cellulose as a sole carbon source. This circuit, fully implemented, may improve yields of industrial bioprocessing [90] of high-value molecules such as drop-in replacements for diesel fuel [15] and pharmaceuticals [99].

6.2 Rapid in vivo optimization: tuning lactose utilization

The lactose utilization network in *Escherichia coli*has three major genes: *lacY*, *lacZ*, and *lacI* 6.6. LacY, known as lactose permease, is an active importer of lactose. LacZ, known as β-galactosidase, is a catalytic enzyme that breaks lactose down into glucose and galactose, which are digested by the cell. LacI, know as the *lac* repressor, is a transcriptional repressor that is deactivated in the presence of a lactose analog called allolactose. LacZ enzymatically converts lactose to allolactose, which induces the production of LacY and LacZ. LacA, coexpressed with LacY and lacZ, has an ill-defined function for lactose utilization, though it is suspected that it has a role in detoxyfication.

What the lactose utilization network is optimized for is a current question among biologists. In 2005, Dekel and Alon demonstrated that the lactose utilization network is evolutionarily tunable in *E. coli* cultures grown via serial passage in constant lactose environments over 300–500 generations. [100]. They also quantified the fitness cost, in growth rate, of expressing the *lac* genes in the absence of lactose as well as the benefit in its presence. Further theoretical work [101] suggested that the network is likely optimized to respond to pulses of lactose in fluctuating environments, as the logarithmic growth states generally employed in biology experiments are rare in natural environments..

Combining the tunability of gene expression using rbSSR motifs with the inducible mutator strain should accelerate the rate of adaptation for a given environment. To demonstrate the potential for accelerating evolution with rbSSRs, we constructed a strain library6.1 in *E. coli* that disrupts the native lactose utilization network by introducing pure or degenerate

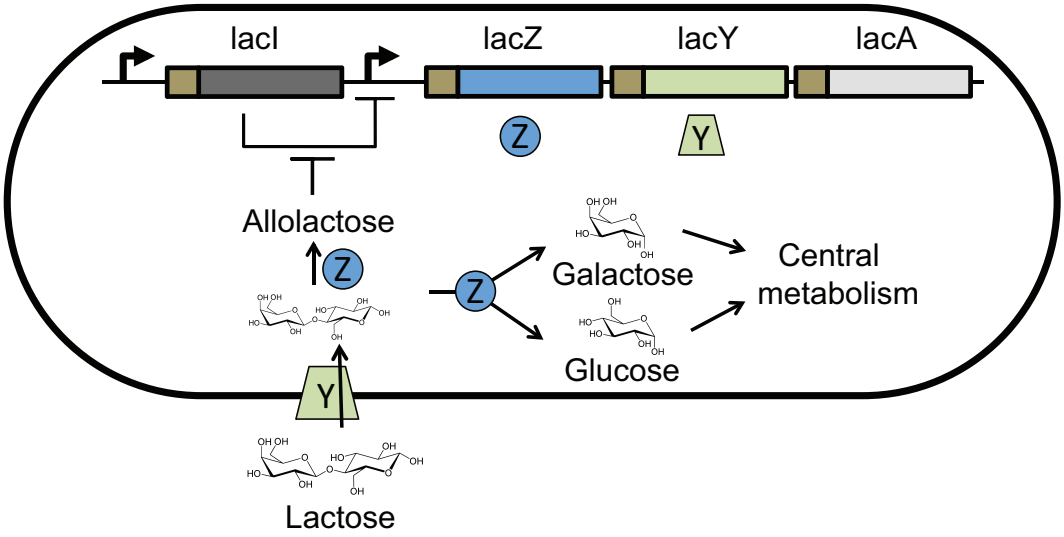

Figure 6.6: Lactose utilization gene network in *E. coli*

Table 6.1: Strain library for rbSSR directed evolution of lactose utilization. Spacer sequences are shown in red. Shine-Dalgarno sequences and start codons are shown in blue (directed mutations in green).

Construct	Short form	Sequence	Spacer length
lacZ-WT	Z-WT	TCACACAGGAAACAGCTATGACCATG	7
lacZ-$(A)_{20}$E	Z20E	TCACAGGAGGAAAAAAAAAAAAAAAAAAATGACCATG	18
lacZ-$(A)_{24}$d(CT)	Z24d(CT)	TCACACAGGAAACAAATAAACAAATAAAAAGACCATG	24
lacY-WT	Y-WT	GTAAGGAAATCCATTATGTACTAT	8
lacY-$(A)_{20}$	Y20	GTAAGGAAAAAAAAAAAAAAAAAAATGTACTAT	18
lacY-$(A)_{20}$d(CT)	Y20d(CT)	GTAAGGAAACAAATAAACAAATAAAATGTACTAT	18

SSRs to the RBS spacers of *lacY* and *lacZ*, genes primarily responsible for importing and digesting lactose, respectively. We expect the high mutation rates of pure SSRs coupled with the predictable relationship between spacer length and gene expression to enable rapid evolution when these strains are grown with lactose as the sole carbon source.

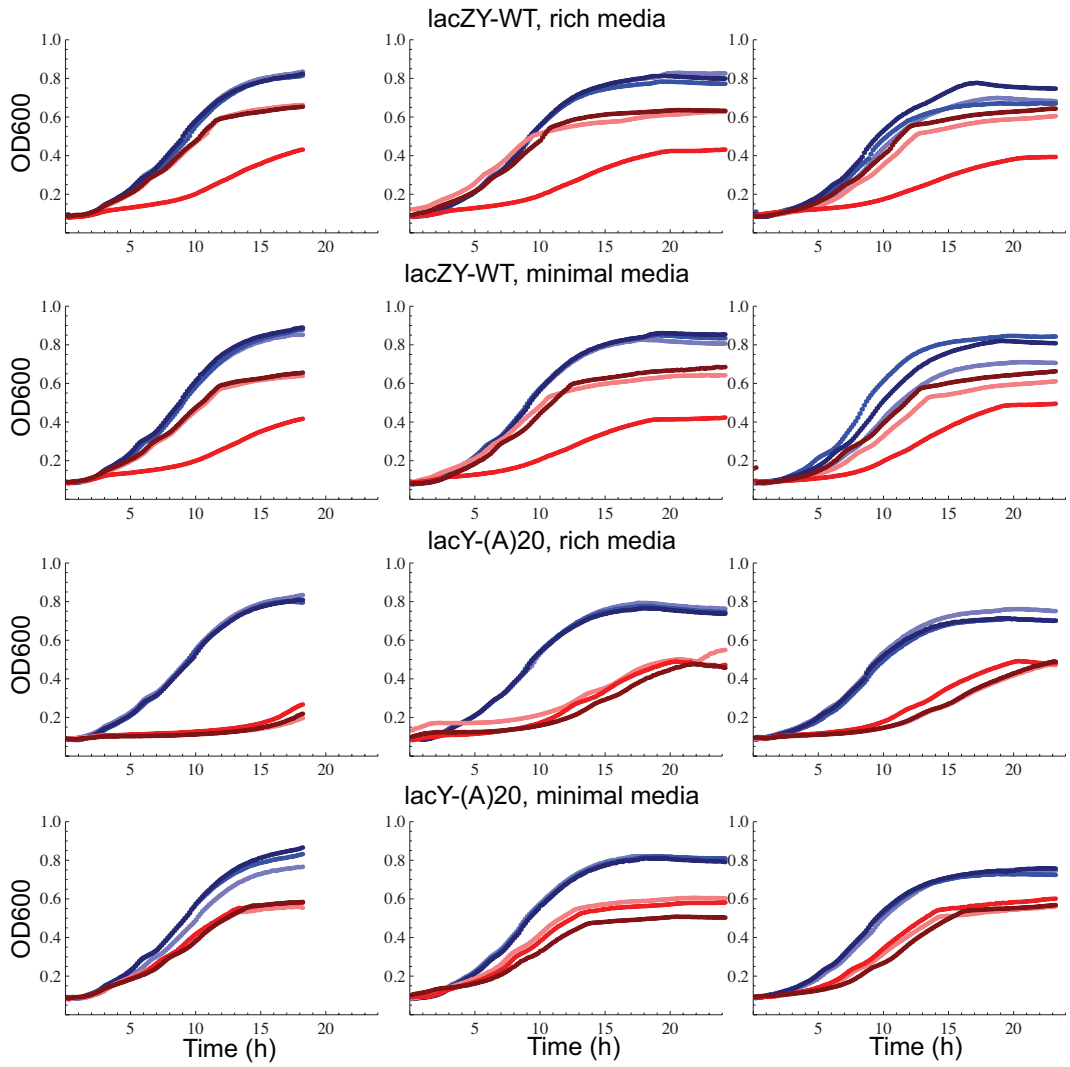

Figure 6.7: Directed evolution of *lacY* rbSSR in batch culture with wild-type control.

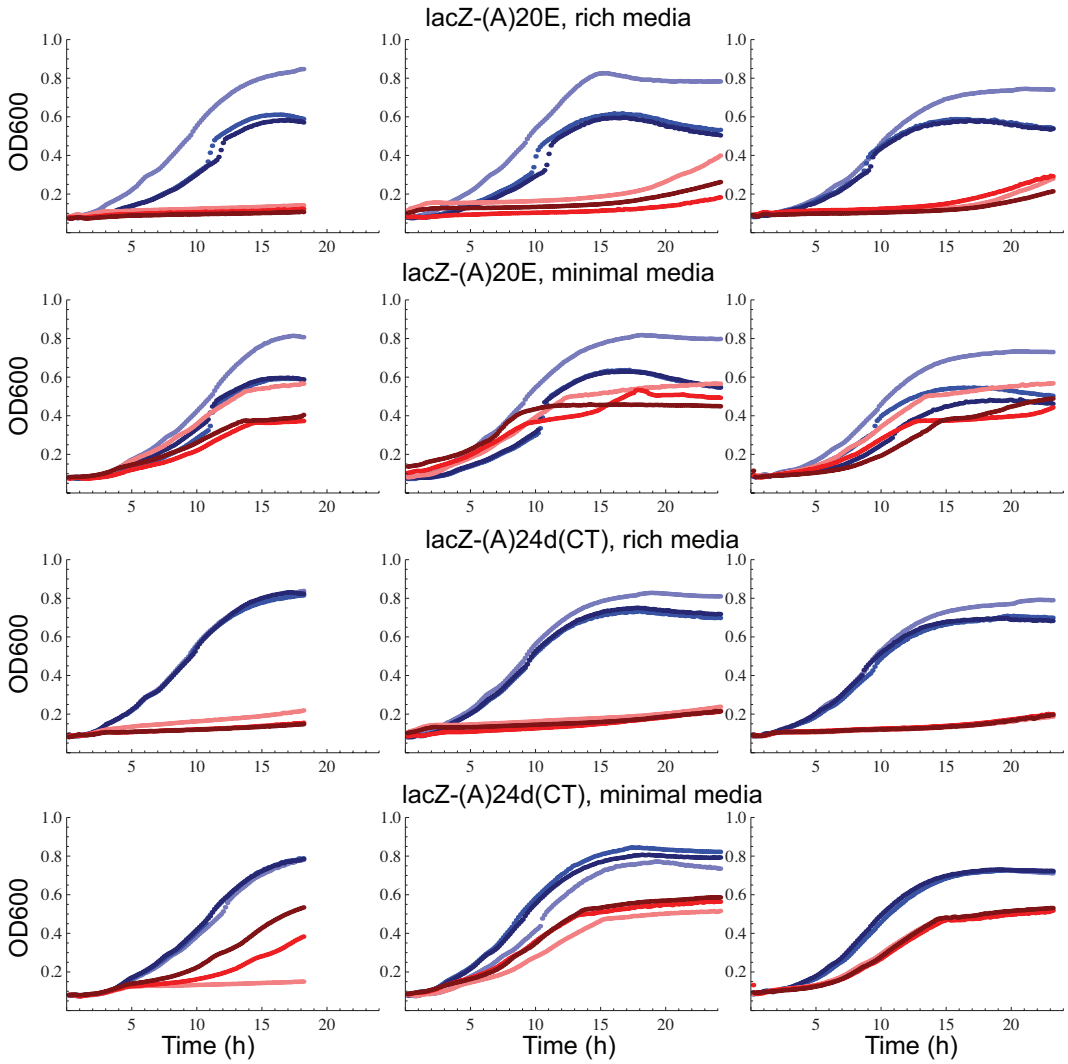

Figure 6.8: Directed evolution of *lacY* and *lacZ* rbSSRs in batch culture.

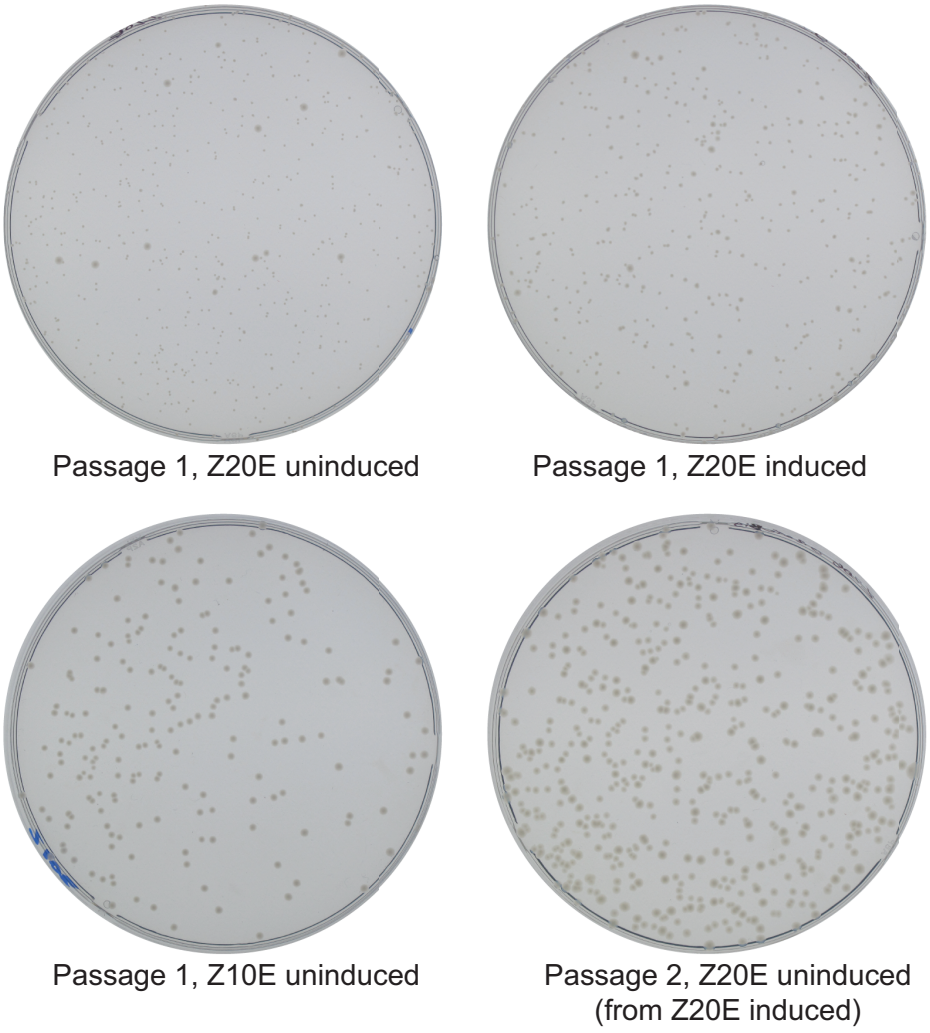

Figure 6.9: Directed evolution of *lacZ* rbSSR on minimal lactose agar plates. The *lacZ*-A10E variant produces colonies of similar size to wild-type cells.

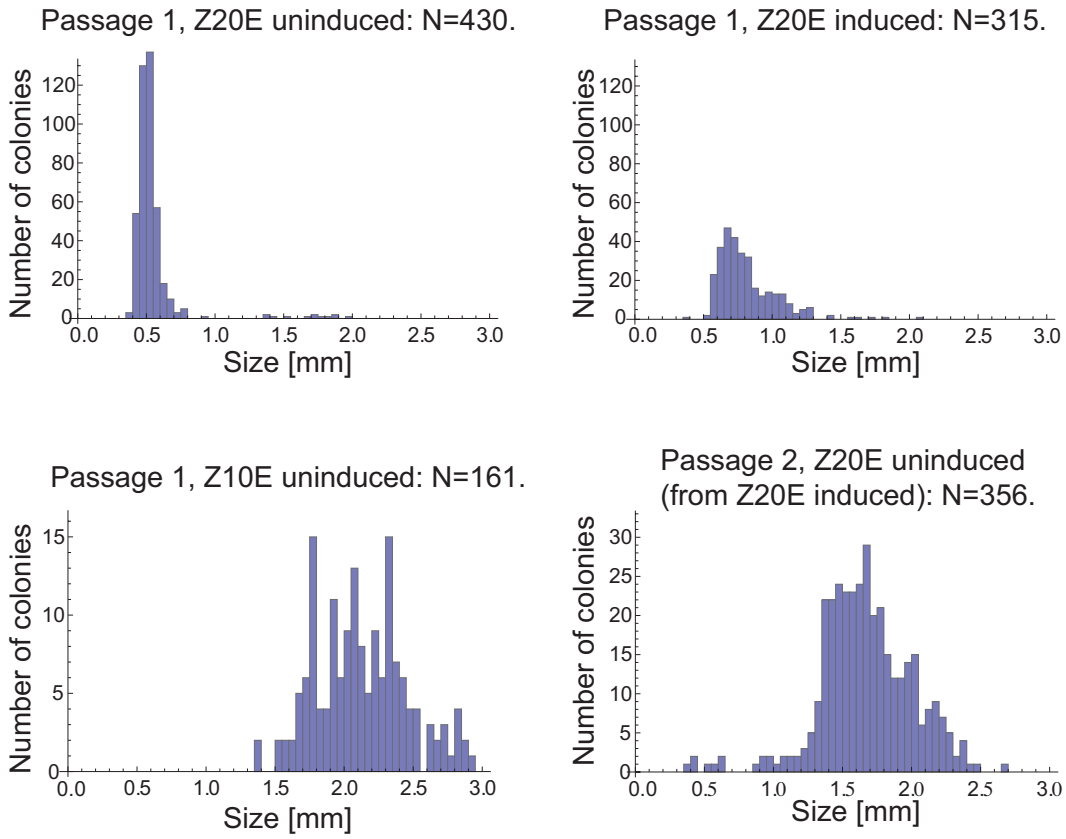

Figure 6.10: Directed evolution of *lacZ* rbSSR on minimal lactose agar plates. The *lacZ*-A10E variant produces colonies of similar size to wild-type cells.

The first characterization assay we ran was to grow cells in a rich environment (M9CA + 0.1% glycerol + 0.4% lactose + biotin) and a minimal environment (M9 + 0.4% lactose + biotin). Results of three serial passages in each environment for wild-type and *lacY*-$(A)_{20}$ cells are shown in Figure 6.7. Results of three serial passages in each environment for *lacZ*-$(A)_{20}$E and *lacZ*-$(A)_{24}$d(CT) cells are shown in Figure 6.8. All cells in the rich environment grew to saturation in one day. The detuned strains grown in the minimal environment took 2-3 days to grow to saturation.

To observe single colony dynamics, we also grew cells on minimal lactose agar plates(Fig. 6.9) and observed the variation in colony size (Fig. 6.10). To assay lactose utilization, the lacZ-(A)20E construct was grown to saturation in LB, then diluted and plated on M9 lactose plates with 0 (upper left) and 0.2 (upper right) ng/mL aTc as inducer and grown for approximately 60 hours at 37 C. A single colony from the induced plate was similarly passaged in LB, then plated on M9 lactose (lower right), resulting in colonies nearly as large as colonies generated from wild-type cells (lower left). Sequencing from a Passage 2 colony resulted in primarily lacZ-(A)17E.

This demonstration project of a well-studied model system is an ideal test system to explore the potential for in vivo tuning with focused mutagenesis.

6.3 Extending SSR tuning to other network parameters.

The same approach used to engineer highly tunable elements with simple sequence repeats can be extended to other network parameters in bacteria [79, 38] and to higher organisms [102] by tuning the spacing between known regulatory motifs such as those responsible for transcription initiation [39] or intron splicing efficiency [103]. To continue scaling up functionally complex behaviors in synthetic gene networks, these approaches will likely need to be combined with tuning methods that control parameters that are untunable by our methods, including network connectivity, protein-protein interactions and enzyme catalytic efficiencies. The present tuning approach will likely be a part of comprehensive strategies for fine-tuning gene circuits to perform optimally in a given context.

6.4 Conclusion: towards engineered systems with functional complexity comparable to natural systems

The experiments described in this work suggest that a complex gene network may require substantial tuning to function as desired. The approach we have developed to tune engineered gene networks uses a very simple construct: a tandem DNA sequence repeat in the spacer region of the ribosome binding site. Sequence repeats seem ideal for this purpose for a variety of reasons. First, the relationship between the length of the repeat and the strength of the resulting RBS is clear. Second, the range of expression obtained by coupling rbSSR libraries with other regulatory sequences, such as promoters, is large. Third, genetic instability can be focused on the RBS spacer, allowing rapid exploration of the expression space via PCR or combinatorial assembly methods. We have also demonstrated some of the potential of these repeats as tools for tuning gene networks in vivo by disrupting the native lactose utilization network in *E. coli* using long rbSSRs, and directing the evolution of the network by forcing mutations to the repeat sequences.

Each assembly method for rbSSR variation discussed in this work has its respective merits and drawbacks for creating functional gene networks. Overall, using two PCR products in Gibson assemblies to vary the rbSSR strength for each repressor gene of the bistable switch network produced 10-100 times more transformants than the equivalent oligo assembly. We observed very few mutations outside the rbSSR sequence for PCR-generated libraries; by contrast, only 50-60% of clones from oligo assembly were error-free. As suggested in [44], the relatively high error rates from oligo assembly are most likely the result of errors generated by oligo synthesis. In terms of explorability, the oligo assembly method provides a more uniform, or arbitrary, search as well as programmable sampling of the expression space. The PCR method provides local search, with a deletion bias, from a start point determined by the initial rbSSR repeats from the parent plasmid.

The rbSSR tuning methods described in this work could be combined to form a comprehensive tuning strategy for complex gene networks through coarse sampling of the expression space followed by fine-tuning. An example network might have four target genes for tuning. Assuming a sample space of 11 rbSSR spacer sequences for each target, there are

$11^4 = 14641$ possible sequence combinations, covering up to a 1000-fold range of expression for each gene. If an engineer chose to sample only three spacer lengths per target, the initial, coarsely sampled space would include 81 unique clones, spaced arbitrarily across the rbSSR design space via the oligo assembly method. Once the "best" variant of this set is determined through experimental screens, the system would be fine-tuned from this new initial point via high-fidelity PCR-based library generation and assembly, followed by additional rounds of screening. If a selection assay is available for the desired behavior, in vivo rbSSR tuning may reach optimal performance more rapidly by starting from an initial set of rbSSR combinations, or by competing a pool of transformants generated by combinatorial assembly of rbSSR PCR products. We have followed this approach at some level for the SDAc project with promising results.

As mentioned above and in Chapter 3 SSR tuning is extensible to other network parameters and is compatible with combinatorial and computational tuning strategies. Choosing the optimal combination of network parameters and tuning strategies for a given circuit architecture is an area of research that will likely produce comprehensive tuning strategies to accelerate the realization of functionally complex engineered gene networks in science, medicine, and industry.

BIBLIOGRAPHY

[1] Woese CR (1987) Bacterial evolution. *Microbiol Rev* 51:221–71.

[2] Lin LH, et al. (2006) Long-term sustainability of a high-energy, low-diversity crustal biome. *Science* 314:479–82.

[3] Schippers A, et al. (2005) Prokaryotic cells of the deep sub-seafloor biosphere identified as living bacteria. *Nature* 433:861–4.

[4] Røy H, et al. (2012) Aerobic microbial respiration in 86-million-year-old deep-sea red clay. *Science* 336:922–5.

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

[6] Purnick PE, Weiss R (2009) The second wave of synthetic biology: from modules to systems. *Nat Rev Mol Cell Biol* 10:410–422.

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

[8] Danino T, Mondragón-Palomino O, Tsimring L, Hasty J (2010) A synchronized quorum of genetic clocks. *Nature* 463:326–30.

[9] Yokobayashi Y, Weiss R, Arnold FH (2002) Directed evolution of a genetic circuit. *Proc Natl Acad Sci USA* 99:16587–16591.

[10] Tamsir A, Tabor JJ, Voigt CA (2011) Robust multicellular computing using genetically encoded NOR gates and chemical 'wires'. *Nature* 469:212–215.

[11] Tigges M, Marquez-Lago TT, Stelling J, Fussenegger M (2009) A tunable synthetic mammalian oscillator. *Nature* 457:309–312.

[12] Aubel D, Fussenegger M (2010) Mammalian synthetic biology–from tools to therapies. *Bioessays* 32:332–345.

[13] Lohmueller JJ, Armel TZ, Silver PA (2012) A tunable zinc finger-based framework for boolean logic computation in mammalian cells. *Nucleic Acids Res* 40:5180–5187.

[14] Khalil AS, et al. (2012) A synthetic biology framework for programming eukaryotic transcription functions. *Cell* 150:647–658.

[15] Schirmer A, Rude MA, Li X, Popova E, del Cardayre SB (2010) Microbial biosynthesis of alkanes. *Science* 329:559–562.

[16] Fischbach M, Voigt CA (2010) Prokaryotic gene clusters: a rich toolbox for synthetic biology. *Biotechnol J* 5:1277–1296.

[17] Gardner TS, Cantor CR, Collins JJ (2000) Construction of a genetic toggle switch in *Escherichia coli*. *Nature* 403:339–342.

[18] Acar M, Mettetal J, van Oudenaarden A (2008) Stochastic switching as a survival strategy in fluctuating environments. *Nat Genet* 40:471–475.

[19] Levine JH, Fontes ME, Dworkin J, Elowitz MB (2012) Pulsed feedback defers cellular differentiation. *PLoS Biol* 10:e1001252.

[20] Danchin A (2012) Scaling up synthetic biology: Do not forget the chassis. *FEBS Lett* 586:2129–2137.

[21] Rosenfeld N, Young JW, Alon U, Swain PS, Elowitz MB (2005) Gene regulation at the single-cell level. *Science* 307:1962–1965.

[22] Süel GM, Kulkarni RP, Dworkin J, Garcia-Ojalvo J, Elowitz MB (2007) Tunability and noise dependence in differentiation dynamics. *Science* 315:1716–1719.

[23] Thattai M, van Oudenaarden A (2001) Intrinsic noise in gene regulatory networks. *Proc Natl Acad Sci USA* 98:8614–8619.

[24] Arkin AP, Fletcher DA (2006) Fast, cheap and somewhat in control. *Genome Biol* 7:114.

[25] Lu TK, Khalil AS, Collins JJ (2009) Next-generation synthetic gene networks. *Nat Biotechnol* 27:1139–1150.

[26] O'Shaughnessy EC, Palani S, Collins JJ, Sarkar CA (2011) Tunable signal processing in synthetic MAP kinase cascades. *Cell* 144:119–131.

[27] Murphy K, Adams R, Wang X, Balazsi G, Collins J (2010) Tuning and controlling gene expression noise in synthetic gene networks. *Nucleic Acids Res* 38:2712–2726.

[28] Callura J, Dwyer D, Isaacs F, Cantor C, Collins J (2010) Tracking, tuning, and terminating microbial physiology using synthetic riboregulators. *Proc Natl Acad Sci USA* 107:15898–15903.

[29] Babiskin AH, Smolke CD (2011) A synthetic library of RNA control modules for predictable tuning of gene expression in yeast. *Mol Syst Biol* 7:471.

[30] Murat D, Quinlan A, Vali H, Komeili A (2010) Comprehensive genetic dissection of the magnetosome gene island reveals the step-wise assembly of a prokaryotic organelle. *Proc Natl Acad Sci USA* 107:5593–5598.

[31] Egbert RG, Klavins E (2012) Fine-tuning gene networks using simple sequence repeats. *Proc Natl Acad Sci USA* doi:10.1073/pnas.1205693109.

[32] Crick F (1970) Central dogma of molecular biology. *Nature* 227:561–563.

[33] Alon U (2007) Network motifs: theory and experimental approaches. *Nat Rev Genet* 8:450–461.

[34] Waters LS, Storz G (2009) Regulatory RNAs in bacteria. *Cell* 136:615–628.

[35] Cox RS, Surette MG, Elowitz MB (2007) Programming gene expression with combinatorial promoters. *Mol Syst Biol* 3:145.

[36] Sleight S, Bartley B, Lieviant J, Sauro H (2010) Designing and engineering evolutionary robust genetic circuits. *J Biol Eng* 4:12.

[37] Vellanoweth RL, Rabinowitz JC (1992) The influence of ribosome-binding-site elements on translational efficiency in *Bacillus subtilis* and *Escherichia coli in vivo*. *Mol Microbiol* 6:1105–1114.

[38] Müller J, Oehler S, Müller-Hill B (1996) Repression of *lac* promoter as a function of distance, phase and quality of an auxiliary *lac* operator. *J Mol Biol* 257:21–29.

[39] Ellis T, Wang X, Collins JJ (2009) Diversity-based, model-guided construction of synthetic gene networks with predicted functions. *Nat Biotechnol* 27:465–471.

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

[41] Ozbudak EM, Thattai M, Kurtser I, Grossman AD, van Oudenaarden A (2002) Regulation of noise in the expression of a single gene. *Nat Genet* 31:69–73.

[42] Carlson R (2009) The changing economics of DNA synthesis. *Nat Biotechnol* 27:1091–1094.

[43] Gibson DG, et al. (2009) Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat Methods* 6:343–345.

[44] Gibson DG, Smith HO, Hutchison CA, Venter JC, Merryman C (2010) Chemical synthesis of the mouse mitochondrial genome. *Nat Methods* 7:901–903.

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

[46] Bashor CJ, Horwitz AA, Peisajovich SG, Lim WA (2010) Rewiring cells: synthetic biology as a tool to interrogate the organizational principles of living systems. *Annu Rev Biophys* 39:515–537.

[47] Alper H, Fischer C, Nevoigt E, Stephanopoulos G (2005) Tuning genetic control through promoter engineering. *Proc Natl Acad Sci USA* 102:12678–12683.

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

[49] Mutalik VK, Qi L, Guimaraes JC, Lucks JB, Arkin AP (2012) Rationally designed families of orthogonal RNA regulators of translation. *Nat Chem Biol* 8:447–454.

[50] Pfleger BF, Pitera DJ, Smolke CD, Keasling JD (2006) Combinatorial engineering of intergenic regions in operons tunes expression of multiple genes. *Nat Biotechnol* 24:1027–1032.

[51] Carothers JM, Goler JA, Juminaga D, Keasling JD (2011) Model-driven engineering of RNA devices to quantitatively program gene expression. *Science* 334:1716–9.

[52] Poelwijk FJ, de Vos MGJ, Tans SJ (2011) Tradeoffs and optimality in the evolution of gene regulation. *Cell* 146:462–470.

[53] Siegel JB, et al. (2010) Computational design of an enzyme catalyst for a stereoselective bimolecular Diels-Alder reaction. *Science* 329:309–313.

[54] Feng XJ, et al. (2004) Optimizing genetic circuits by global sensitivity analysis. *Biophys J* 87:2195–2202.

[55] Wong TS, Zhurina D, Schwaneberg U (2006) The diversity challenge in directed protein evolution. *Combinatorial Chemistry #38; High Throughput Screening* 9:271–288.

[56] Haseltine EL, Arnold FH (2007) Synthetic gene circuits: Design with directed evolution. *Annu Rev Biophys Biomol Struct* 36:1–19.

[57] Barrick D, et al. (1994) Quantitative analysis of ribosome binding sites in *E. coli*. *Nucleic Acids Res* 22:1287–1295.

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

[59] Torres-Cruz J, van der Woude MW (2003) Slipped-strand mispairing can function as a phase variation mechanism in *Escherichia coli*. *J Bacteriol* 185:6990.

[60] Levinson G, Gutman GA (1987) High frequencies of short frameshifts in poly-CA/TG tandem repeats borne by bacteriophage M13 in *Escherichia coli* K-12. *Nucleic Acids Res* 15:5323–5338.

[61] Shinde D, Lai YL, Sun FZ, Arnheim N (2003) *Taq* DNA polymerase slippage mutation rates measured by PCR and quasi-likelihood analysis: $(CA/GT)_n$ and $(A/T)_n$ microsatellites. *Nucleic Acids Res* 31:974–980.

[62] Elowitz MB, Levine AJ, Siggia ED, Swain PS (2002) Stochastic gene expression in a single cell. *Science* 297:1183–1186.

[63] Komarova AV, Tchufistova LS, Supina EV, Boni IV (2002) Protein S1 counteracts the inhibitory effect of the extended Shine-Dalgarno sequence on translation. *RNA* 8:1137–1147.

[64] Vimberg V, Tats A, Remm M, Tenson T (2007) Translation initiation region sequence preferences in *Escherichia coli*. *BMC Mol Biol* 8:100.

[65] Nevozhay D, Adams RM, Murphy KF, Josic K, Balázsi G (2009) Negative autoregulation linearizes the dose-response and suppresses the heterogeneity of gene expression. *Proc Natl Acad Sci USA* 106:5123–5128.

[66] Pedraza JM, van Oudenaarden A (2005) Noise propagation in gene networks. *Science* 307:1965–1969.

[67] Bagh S, et al. (2008) Plasmid-borne prokaryotic gene expression: sources of variability and quantitative system characterization. *Phys Rev E* 77:021919.

[68] Bachmann BJ (1972) Pedigrees of some mutant strains of *Escherichia coli* K-12. *Bacteriol Rev* 36:525–57.

[69] Baba T, et al. (2006) Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: The Keio collection. *Mol Syst Biol* 2:2006.0008.

[70] Kobayashi H (2004) Programmable cells: Interfacing natural and engineered gene networks. *Proc Natl Acad Sci USA* 101:8414–8419.

[71] Eldar A, Elowitz MB (2010) Functional roles for noise in genetic circuits. *Nature* 467:167–173.

[72] Oehler S, Amouyal M, Kolkhof P, von Wilcken-Bergmann B, Müller-Hill B (1994) Quality and position of the three *lac* operators of *E. coli* define efficiency of repression. *EMBO Journal* 13:3348–3355.

[73] Loinger A, Biham O (2009) Analysis of genetic toggle switch systems encoded on plasmids. *Phys Rev Lett* 103:068104.

[74] Dubnau D, Losick R (2006) Bistability in bacteria. *Mol Microbiol* 61:564–572.

[75] Veening JW, Smits WK, Kuipers OP (2008) Bistability, epigenetics, and bet-hedging in bacteria. *Annu Rev Microbiol* 62:193–210.

[76] Hoang TT, Karkhoff-Schweizer RR, Kutchma AJ, Schweizer HP (1998) A broad-host-range Flp-FRT recombination system for site-specific excision of chromosomally-located DNA sequences: application for isolation of unmarked *Pseudomonas aeruginosa* mutants. *Gene* 212:77–86.

[77] Gemayel R, Vinces MD, Legendre M, Verstrepen KJ (2010) Variable tandem repeats accelerate evolution of coding and regulatory sequences. *Annu Rev Genet* 44:445–477.

[78] Lin WH, Kussell E (2012) Evolutionary pressures on simple sequence repeats in prokaryotic coding regions. *Nucleic Acids Res* 40:2399–2413.

[79] van Belkum A, Scherer S, van Alphen L, Verbrugh H (2006) Bacterial contingency loci: the role of simple sequence DNA repeats in bacterial adaptation. *Annu Rev Genet* 40:307–333.

[80] Michael TP, et al. (2007) Simple sequence repeats provide a substrate for phenotypic variation in the *Neurospora crassa* circadian clock. *PLoS ONE* 2:e795.

[81] Sawyer LA, et al. (1997) Natural variation in a *Drosophila* clock gene and temperature compensation. *Science* 278:2117–20.

[82] Farabaugh PJ, Schmeissner U, Hofer M, Miller JH (1978) Genetic studies of the *lac* repressor. VII. on the molecular nature of spontaneous hotspots in the *lacI* gene of *Escherichia coli*. *J Mol Biol* 126:847–57.

[83] Quan S, et al. (2012) Adaptive evolution of the lactose utilization network in experimentally evolved populations of *Escherichia coli*. *PLoS Genet* 8:e1002444.

[84] Levinson G, Gutman GA (1987) Slipped-strand mispairing: a major mechanism for DNA sequence evolution. *Mol Biol Evol* 4:203–21.

[85] Strand M, Prolla TA, Liskay RM, Petes TD (1993) Destabilization of tracts of simple repetitive DNA in yeast by mutations affecting DNA mismatch repair. *Nature* 365:274–276.

[86] Herman GE, Modrich P (1981) *Escherichia coli* k-12 clones that overproduce *dam* methylase are hypermutable. *J Bacteriol* 145:644–6.

[87] Marinus MG, Poteete A, Arraj JA (year?) Correlation of dna adenine methylase activity with spontaneous mutability in *Escherichia coli* k-12. *Gene*.

[88] Nakashima N, Tamura T (2009) Conditional gene silencing of multiple genes with antisense RNAs and generation of a mutator strain of *Escherichia coli*. *Nucleic Acids Res* 37:e103.

[89] Crespi BJ (2001) The evolution of social behavior in microorganisms. *Trends Ecol Evol (Amst)* 16:178–183.

[90] Lynd LR, van Zyl WH, McBride JE, Laser M (2005) Consolidated bioprocessing of cellulosic biomass: an update. *Curr Opin Biotechnol* 16:577–583.

[91] Bokinsky G, et al. (2011) Synthesis of three advanced biofuels from ionic liquid-pretreated switchgrass using engineered *Escherichia coli*. *Proc Natl Acad Sci USA* 108:19949–19954.

[92] Zhang XZ, Sathitsuksanoh N, Zhu Z, Zhang YHP (2011) One-step production of lactate from cellulose as the sole carbon source without any other organic nutrient by recombinant cellulolytic *Bacillus subtilis*. *Metabolic Eng* 13:364–372.

[93] Ratcliff WC, Denison RF, Borrello M, Travisano M (2012) Experimental evolution of multicellularity. *Proc Natl Acad Sci USA* 109:1595–1600.

[94] Goldsby HJ, Dornhaus A, Kerr B, Ofria C (2012) Task-switching costs promote the evolution of division of labor and shifts in individuality. *Proc Natl Acad Sci USA* doi:10.1073/pnas.1202233109.

[95] Jang S, Oishi K, Egbert R, Klavins E (2012) Specification and simulation of synthetic multi-celled behaviors. *ACS Synthetic Biology* doi:10.1021/sb300034m.

[96] Masaki H, Ohta T (1985) Colicin E3 and its immunity genes. *J Mol Biol* 182:217–227.

[97] Kachroo AH, Kancherla AK, Singh NS, Varshney U, Mahadevan S (2007) Mutations that alter the regulation of the *chb* operon of *Escherichia coli* allow utilization of cellobiose. *Mol Microbiol* 66:1382–1395.

[98] Zhang D, Lax AR, Bland JM, Allen AB (2011) Characterization of a new endogenous endo--1,4-glucanase of formosan subterranean termite (*Coptotermes formosanus*). *Insect Biochem Mol Biol* 41:211–218.

[99] Ro DK, et al. (2006) Production of the antimalarial drug precursor artemisinic acid in engineered yeast. *Nature* 440:940–943.

[100] Dekel E, Alon U (2005) Optimality and evolutionary tuning of the expression level of a protein. *Nature* 436:588–92.

[101] Kalisky T, Dekel E, Alon U (2007) Cost-benefit theory and optimal design of gene regulation functions. *Phys Biol* 4:229–45.

[102] Wierdl M, Greene CN, Datta A, Jinks-Robertson S, Petes TD (1996) Destabilization of simple repetitive DNA sequences by transcription in yeast. *Genetics* 143:713–721.

[103] Chua K, Reed R (2001) An upstream AG determines whether a downstream AG is selected during catalytic step II of splicing. *Mol Cell Biol* 21:1509–14.

[104] Lou C, et al. (2010) Push-on push-off switch. *Mol Syst Biol* 6:350.

[105] Mettetal JT, Muzzey D, Gómez-Uribe C, van Oudenaarden A (2008) The frequency dependence of osmo-adaptation in saccharomyces cerevisiae. *Science* 319:482–4.

Appendix A

SOFTWARE, PLASMIDS, AND PRIMERS

A.1 Software URL and descriptions

The scripts used to process cytometry and sequencing data in this study are available at: `http://depts.washington.edu/soslab/code/`.

List of algorithms with descriptions SSR analysis

- Mathematica script for processing rbSSR distributions from sequencing chromatogram traces
- Mathematica script for fitting data to birth-death process model

Cytometry

- Matlab script to import Accuri C6 cytometry data: read96cyto.m
- Matlab script for cytometry pre-processing (gating)
- Matlab script for calculating bistable switch distributions

Self-destructive altruism: cellulase

- SDAc:gro – gro code for simulating circuit behaviors

Lactose utilization

- Mathematica script for importing plate reader data & displaying growth curves for batch culture evolution.
- Matlab Image processing script for calculating colony size distributions.

Table A.1: Primers used for experimental work.

rbSSR-GFP

Primer Name	Sequence
rbSSR-GFP_rev	GCACTCTTGAAAAAGTCATGCTG
Vf2	TGCCACCTGACGTCTAAGAA
Vr	TGCCACCTGACGTCTAAGAA

rbSSR-BSS

Primer Name	Sequence
ssrBSS_cut	CGCGATGACTTAGTAAAGCACATC
tetR_T10	TTTTTTTTTTATGTCCAGATTAGATAAAAGTAAAGTG
ssrBSS_center_rev	GACTCTAGTAGTGCTCAGTATCTCTATCACTG
ssrBSS_center_fwd	CAAGTTAGCTGCCTTTCGTCTTC
ssrBSS_3'_digest_rev	GCGGGAAACGGTCTGATAAG
ssrBSS_readout	CCGTGTACCTAAATGTAC
ssrBSS_O1R-T20	CACAGGAAACCTTTTTTTTTTTTTTTTTTTTATGTCCAGATTAGATAAAAGTAAAGTGAT
ssrBSS_O1R-T18	CACAGGAAACCTTTTTTTTTTTTTTTTTTATGTCCAGATTAGATAAAAGTAAAGTGAT
ssrBSS_O1R-T16	CACAGGAAACCTTTTTTTTTTTTTTTTATGTCCAGATTAGATAAAAGTAAAGTGAT
ssrBSS_O1R-T14	CACAGGAAACCTTTTTTTTTTTTTTATGTCCAGATTAGATAAAAGTAAAGTGAT
ssrBSS_O1R-T12	CACAGGAAACCTTTTTTTTTTTTATGTCCAGATTAGATAAAAGTAAAGTGAT
ssrBSS_O1R-T10	CACAGGAAACCTTTTTTTTTTATGTCCAGATTAGATAAAAGTAAAGTGAT
ssrBSS_O2F	AAAAAAAAAAGGTTTCCTGTGTGACTCTAGTATGTGAATTGTTATCCGCTCACAATTCTC
ssrBSS_O3R	AGGCAGCTAACTTGACATTGTGAGCGGATAACAAGATACTGAGAATTGTGAGCGGATAAC
ssrBSS_O4F	CAATGTCAAGTTAGCTGCCTTTCGTCTTCAATAATTCTTTGACATCCCTATCAGTGATAG
ssrBSS_O5R	TTTTTGGTTTCCTGTGTGACTCTAGTAGTGCTCAGTATCTCTATCACTGATAGGGATGTC
ssrBSS_O6F-A20	ACACAGGAAACCAAAAAAAAAAAAAAAAAAAAATGAAACCAGTAACGTTATACGATGTCGC
ssrBSS_O6F-A18	ACACAGGAAACCAAAAAAAAAAAAAAAAAATGAAACCAGTAACGTTATACGATGTCGC
ssrBSS_O6F-A16	ACACAGGAAACCAAAAAAAAAAAAAAAAATGAAACCAGTAACGTTATACGATGTCGC
ssrBSS_O6F-A14	ACACAGGAAACCAAAAAAAAAAAAAATGAAACCAGTAACGTTATACGATGTCGC
ssrBSS_O6F-A12	ACACAGGAAACCAAAAAAAAAAAATGAAACCAGTAACGTTATACGATGTCGC
ssrBSS_O6F-A10	ACACAGGAAACCAAAAAAAAAATGAAACCAGTAACGTTATACGATGTCGC

A.2 Plasmids and primers for assembly and sequencing of rbSSR libraries

The plasmid used as the basis of the rbSSR-GFP constructs is ssrGFP and is available on the JBEI Public Registry as JPUB_000323. The plasmid used as the basis for the bistable switch is ssrBBS and is available on JBEI Public Registry as JPUB_000466. Both sequences can also be found at `http://depts.washington.edu/soslab/rbSSR/DNA`. The primers used for assembling constructs, performing PCR reactions, and for sequencing are shown in Table A.1.