Regulatory systems for the robust control of engineered genetic programs

by

Thomas Hale Segall-Shapiro

B.S.B., B.S., Rice University (2010)

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

Doctor of Philosophy in Biological Engineering

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 2017

Signature redacted

Author .....
Department of Biological Engineering
August 21, 2017

Signature redacted

Certified by .
Christopher A. Voigt
Professor of Biological Engineering
Thesis Supervisor

Signature redacted

Accepted by ..
Mark Bathe
Chair, Graduate Program Committee

This doctoral thesis has been examined by a Committee of the Department of Biological Engineering as follows:

Timothy K. Lu, MD, PhD
Thesis Committee Chair
Associate Professor of Electrical Engineering & Computer Science
Associate Professor of Biological Engineering

Christopher A. Voigt, PhD
Theisis Supervisor
Professor of Biological Engineering

Amy E. Keating, PhD
Professor of Biology
Professor of Biological Engineering

Regulatory systems for the robust control of engineered genetic programs

by

Thomas Hale Segall-Shapiro

Submitted to the Department of Biological Engineering
on August 21, 2017, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Biological Engineering

Abstract

The ability to engineer complex genetic programs could have a huge impact on many industries, yielding organisms that can respond to their environment and perform functions relevant to manufacturing, agriculture, and medicine. However, such engineering efforts have proven difficult, in part because these programs often require precise levels of gene expression for proper function. It is especially tough to build programs that have robust activity, as any changes to the host cells can perturb the context of the genetic system and disrupt carefully tuned expression levels. Additionally, genetic programs often place high demands on host resources, which can adversely affect cell growth and further upset the intended function.

In this thesis, we describe two regulatory systems in *Escherichia coli* that could serve to separate synthetic genetic programs from their host context, potentially leading to more robust activity. First, we build a 'resource allocator' by fragmenting T7 RNA polymerase variants into a conserved fragment and a set of variable fragments. The resource allocator limits the total number of polymerases that can be active in a genetic program, with the aim of protecting the host from being overburdened. This transcriptional budget can be allocated to different elements of the genetic program as necessary and further regulated using additional protein fragments.

Second, we demonstrate a set of stabilized promoters that can maintain a level of gene expression independent of their genetic context. These promoters utilize a noncooperative incoherent feedforward loop to buffer differences in gene expression caused by changes in copy number. We demonstrate that stabilized promoters can be moved between plasmids and different locations on the genome with little change in expression. Further, they minimize the effects of other perturbations that can affect copy number, such as genome mutations and media composition.

Thesis Supervisor: Christopher A. Voigt
Title: Professor of Biological Engineering

Preface

This thesis is primarily comprised of three manuscripts, all of which are the result of the hard work of a number of people. It would not have been possible without their contributions.

Chapter 1 is primarily composed of portions of the review article "Advances in genetic circuit design: novel biochemistries, deep part mining, and precision gene expression", which Alec Nielsen and I co-authored with Christopher Voigt. It was published in *Current opinion in chemical biology* in December 2013.

Chapter 2 and parts of Chapter 1 are composed of the article "A resource allocator for transcription based on a highly fragmented T7 RNA polymerase", which was published in *Molecular Systems Biology* in July 2014. Adam Meyer, Andrew Ellington, Eduardo Sontag, and Christopher Voigt contributed to this project. It was funded by the United States Office of Naval Research, the United States National Institutes of Health, and the US National Science Foundation Synthetic Biology Engineering Research Center.

Chapter 3 and parts of Chapter 1 are composed of a manuscript titled "Constant gene expression at any copy number using feedforward stabilized promoters", which is currently under review. Eduardo Sontag and Christopher Voigt were integral in this work. It was funded by an Office of Naval Research Multidisciplinary University Research Initiative grant, a National Institutes of Health National Institute of General Medical Sciences Center for Integrated Synthetic Biology grant, a National Institutes of Standards and Technology grant, and a United States National Science Foundation Synthetic Biology Engineering Research Center grant.

During my graduate career I have been supported by a National Defense Science and Engineering Graduate Fellowship and a Fannie and John Hertz Foundation Fellowship. I am grateful for their support.

Acknowledgments

First of all I would like to acknowledge my advisor, Chris Voigt, for providing a fantastic environment to work and learn in for the past ~7 years. I am constantly amazed by the sheer number of interesting ideas you have and directions that you are pursuing. Thank you for both giving me the freedom to explore so many cool things that are possible in synthetic biology and focusing me on the most important targets.

I would also like to acknowledge my committee members, Tim Lu and Amy Keating. Every interaction that I had with you was valuable – thank you for your time and advice.

Eduardo Sontag has contributed a tremendous amount of work to both of the projects discussed in this thesis and they would not remotely resemble the form that they are in now without his input. But more than that, our many discussions have had a huge impact on the way that I think about science and the overall direction that my research has taken. Thank you for all of your effort, insight, and enthusiasm.

The Voigt lab has been a fantastic place to spend far too much of my time over the years, in a large part due to the unbelievably stellar scientists who populate it. Thank you especially to Virgil Rhodius and Karsten Temme, who mentored me when I was just starting out. I can't list all the people in the lab who have had a positive impact on me – it would truly be everyone I've overlapped with. But I would like to acknowledge a few people who have contributed to or particularly helped out with the work I describe in this thesis: Alec Nielsen, Adam Meyer, YongJin Park, Emerson Glassey, and Andrew King. And thank you Terry King and Barbara Karampalas for just making things work.

I would not be writing this thesis if it were not for my experiences as an undergraduate on the Rice University iGEM team and afterwards pursuing research in the Silberg lab. I had the great fortune to interact early on with Joff Silberg and Beth Beason, who truly care about kindling scientific enthusiasm in undergraduates and had a profound impact on my scientific progression. Peter Nguyen was an irreplaceable mentor for me in the lab and Taylor Stevenson was a great colleague during the iGEM years and beyond. Thank you also to my summer internship mentors at JCVI, Prabha Iyer and Ray-Yuan Chuang, for exposing me to new aspects of synthetic biology and helping me grow as a scientist.

I have had the good fortune to be a part of three fantastic groups of graduate students over the course of my PhD: the BEAST 2010 cohort, the MIT BE class of 2011, and the community of Hertz fellows. You have been great friends, scientific sounding boards, and sources of inspiration. Grad school would not have been the same without you.

To my family: you have been a source of constant encouragement throughout, and I am immensely grateful for your love and support. Mom – I can't fully put into words what you mean to me. Thank you for always enthusiastically supporting my dreams and giving me the resources to pursue them. Thank you for always believing in me. Dad, thank you for your interest and excitement in my work – it has been a powerful motivation.

Thank you to all of my friends for dragging me out of lab and keeping me sane.

Thank you Kendal for everything.

Contents

1 Introduction **17**

1.1 Overview 17

1.2 Classes of regulators available for genetic programming of bacteria 20

1.2.1 Protein regulators of transcription 21

1.2.2 RNA regulators of translation 22

1.2.3 RNA regulators of transcription 23

1.2.4 Proteins that modify DNA 23

1.3 Precision gene expression in bacteria 24

1.3.1 Tuning knobs for expression 25

1.3.2 Insulators to buffer the impact of genetic context 26

1.4 Conclusions 28

1.5 Figures & Tables 29

2 A 'resource allocator' for transcription based on a highly fragmented T7 RNA polymerase **35**

2.1 Abstract 35

2.2 Introduction 36

2.3 Results 38

2.3.1 Bisection mapping of T7 RNA polymerase 38

2.3.2 Division of T7 RNAP into multiple fragments 40

2.3.3 Construction of σ fragments with different promoter specificities . . . 40

2.3.4 Setting and sharing the transcriptional budget 42

2.3.5 Positive and negative regulation of the core fragment 44

2.3.6 Coupling RNAP activity to the concentration of arbitrary α fragment tagged proteins . . . 45
2.3.7 Application of the α fragment to compensate for differences in copy number . . . 46
2.4 Discussion . . . 47
2.5 Technical Information . . . 50
2.5.1 Splitposon method for bisection mapping proteins . . . 50
2.5.2 Bisection mapping of T7 RNA polymerase . . . 52
2.5.3 Supporting experiments . . . 54
2.6 Methods . . . 56
2.6.1 Strains and media . . . 56
2.6.2 Plasmids and parts . . . 56
2.6.3 Bisection mapping T7 RNA polymerase . . . 57
2.6.4 Assay protocol . . . 58
2.6.5 Flow cytometry characterization . . . 59
2.6.6 Measuring the growth impact of split polymerase expression . . . 59
2.6.7 Error-prone PCR of σ fragment variants . . . 60
2.6.8 Tuning α fragment expression to compensate for copy number . . . 60
2.7 Modeling . . . 61
2.7.1 Kinetic model of the resource allocator . . . 61
2.7.2 Uniqueness and stability of steady states in resource allocator model . 63
2.7.3 Modeling σ fragment competition data . . . 69
2.8 Figures . . . 72
2.9 Tables . . . 94

3 Constant gene expression at any copy number using feedforward stabilized promoters 101

3.1 Abstract . . . 101
3.2 Introduction . . . 102
3.3 Results . . . 103

3.4 Discussion . . . 107
3.5 Methods . . . 108
3.5.1 Strains and Media . . . 108
3.5.2 Plasmids and Genetic Parts . . . 108
3.5.3 Cytometry Assays . . . 109
3.5.4 Cytometry Analysis . . . 109
3.5.5 qPCR measurements . . . 110
3.5.6 Plate Reader Assays . . . 110
3.5.7 Plate Reader Analysis . . . 111
3.5.8 Tn5 Transposition . . . 111
3.5.9 Arbitrary PCR Sequencing . . . 112
3.5.10 Error-prone PCR . . . 112
3.6 Modeling . . . 113
3.6.1 Mathematical model of iFFL stabilized promoter . . . 113
3.6.2 Comparison to feedback stabilization . . . 119
3.7 Figures . . . 122
3.8 Tables . . . 138

Bibliography **143**

List of Figures

1-1 The diversity of genetic regulatory parts available for building bacterial genetic circuits. 29

1-2 A modern expression cassette comprising genetic tuning knobs and insulators. 31

2-1 The resource allocator. 72

2-2 Bisection mapping of T7* RNAP. 73

2-3 Activation of the core fragment via σ fragments. 75

2-4 Competition between σ fragments to bind the core fragment. 76

2-5 Positive and negative post-transcriptional regulation of the core fragment. . . 78

2-6 Outcomes of a splitposon library. 80

2-7 Vectors used for bisection mapping of T7 RNA polymerase. 81

2-8 Split seams mapped onto T7 RNAP structure. 82

2-9 Surface model of three-piece T7 RNAP. 82

2-10 Detailed σ fragment orthogonality results. 83

2-11 The null fragment lacks σ fragment activity. 84

2-12 Activity of GFP : α fragment fusions. 84

2-13 P_{Tac} activity measurements. 85

2-14 Plasmids used for P_{Tac} activity measurements. 86

2-15 Growth impact of split polymerase systems. 87

2-16 Growth impact of σ fragment competition systems. 88

2-17 Growth impact of highly expressed multiply split polymerase. 89

2-18 Plasmid used for full-length T7* RNAP toxicity measurement. 90

2-19 Reporter plasmids. 90

2-20 Inducible expression plasmids. 91
2-21 Core fragment expression plasmids. 92
2-22 Constitutive σ expression plasmids. 92
2-23 σ fragment competition plasmids. 92
2-24 Reporter plasmids with α fragment compensation. 93

3-1 Stabilized promoter design. 122
3-2 Stabilized promoters compensate for changes in copy number. 123
3-3 Stabilized promoters reduce the impact of genome mutations and growth media. 125
3-4 Promoter sequences of TALE repressible promoters 126
3-5 Expression levels for TALE repressible promoters 127
3-6 Predicted evolutionary stability of TALE genes. 128
3-7 Gene expression from the P_{Tac} promoter 129
3-8 Orthogonality of TALE repressors 129
3-9 pSC101 backbones with varying copy number 130
3-10 Tuning stabilized promoters 131
3-11 Plasmid backbones based on different origins of replication 132
3-12 Design of stabilized promoters with different strengths 133
3-13 Genome position dependent gene expression 134
3-14 Design of stabilized and insulated promoters 135
3-15 Stabilized promoter variants 136
3-16 Other plasmid diagrams 137

List of Tables

1.1 Characteristics of regulatory part families currently available for constructing genetic circuits in bacteria 33

2.1 Statistics of T7 RNA polymerase bisection mapping. 94
2.2 Improved activity clones from the K1F σ fragment ePCR library. 95
2.3 Improved activity clones from the N4 σ fragment ePCR library. 96
2.4 Comparison of null fragment variants. 97
2.5 Modeling parameters. 97
2.6 New plasmids used in this work. 98

3.1 New plasmids used in this work 138
3.2 Plasmid sets used in this work 140
3.3 Genomic insertion locations 142

Chapter 1

Introduction

1.1 Overview

Cells naturally control gene expression using a variety of RNA, protein, and DNA-modifying regulators [1–3]. It was recognized early on that interactions between these regulators could lead to computational operations that are analogous to electronic circuits [4–6]. Genetic engineers have since attempted to build synthetic circuits that would implement artificial programs of gene expression. This could have a revolutionary impact on biotechnology, such as programming bacteria to individually respond to transient conditions in a bioreactor [7], designing therapeutic cells to sense and respond to diseased states within the human body [8–12], or engineering smart plants that can respond to conditions in the environment [13].

Recent progress in synthetic biology has lead to large advances in the technology for engineering the components of such genetic programs. New biosensors are being engineered and discovered that enable programming organisms to respond to a wide variety of conditions [14, 15]. Better algorithms and architectures have been developed that support the rapid construction of complicated genetic logic circuits [16–18]. Increasing numbers of gene clusters are being explored and re-engineered as the potential outputs of genetic programs [19, 20]. Tying these improvements together are cheap gene synthesis, sequencing, and improved high-throughput assembly strategies [21].

Despite these advances, engineering complex genetic programs remains a tremendous challenge. Large genetic engineering projects often require finely balanced gene expression,

but achieving and maintaining this balance is difficult [22,23]. For example, when engineering metabolic pathways there is a need to tune the expression of enzymes to optimize flux to the product, and mismatched expression could lead to the diversion of carbon to other pathways or the accumulation of toxic intermediates [24–26]. Similarly, the construction of intracellular molecular machines requires obtaining the correct ratios of components [27–29].

Balancing expression requires changing the genetic parts controlling each gene, for example promoters or ribosome binding sites (RBSs). Directed evolution seeks to achieve this by making random changes via mutagenesis or part substitution and then screening thousands of variants for improvements [19, 22, 30]. Much work has been devoted towards the improvement of the measurement [31–33] and reliability [34–37] of genetic parts as well as computational methods that use this information to guide a search [16, 38]. It is frustrating that, after going through these efforts to fine tune a system, any changes to the cells can disrupt the balance and require retuning to fix [39, 40]. These changes could be due to differences in media, growth phase, transfer to a new environment (bioreactor, soil, gut microbiota, etc), changing the genetic location of a system, adding more genetic systems, and strain mutagenesis [41, 42].

One major reason that changing the context of a genetic program poses such a problem is that cells control the production of RNA polymerase (RNAP) and ribosomes to balance their biosynthetic cost with the needs of cell growth and maintenance [43]. As such, RNAP and ribosome synthesis is under stringent regulatory control, both to coordinate their levels with respect to cellular and environmental cues for growth [44–52] and to balance the expression of their components for proper assembly into functional machines [43, 53–55]. This sets a resource budget that must be shared in the transcription of approximately 4,000 genes and translation of 106 nucleotides of mRNA in E. coli [56]. The budget is not large; on average, there are 2,000 RNAP and 10,000 ribosomes per cell [56–58]. Mathematical models often assume these budgets to be constant [59–61], but the numbers can vary significantly in different growth phases and nutrient conditions, ranging from 1,500 to 11,400 RNAPs and 6,800 to 72,000 ribosomes per cell [56, 62]. The fluctuations in resources can lead to global changes in expression levels and promoter activities [63, 64].

This poses a large problem when a synthetic genetic system is introduced. When it relies

on the transcription and translation machinery of the host, it becomes implicitly embedded in their regulation, making it sensitive to changes that occur during cell growth and function. As a result, the system can be quite fragile because the strengths of its component parts (promoters and ribosome binding sites) will vary with the resource budgets [7, 39, 65]. For example, changes in the RNAP concentration can impact the expression from constitutive promoters by fivefold [56, 62, 66–68]. These changes can reduce the performance of a system that requires precise balances in expression levels [7, 69, 70]. This has emerged as a particular problem in obtaining reliable expression levels and gene circuit performance during industrial scale-up, where each phase is associated with different growth and media conditions [7].

Another problem is that synthetic systems often place high demands on host transcription and translation resources and this can have global consequences in maintaining growth and responding to stress [71, 72]. Proteins and pathways expressed at very high levels place a burden on cells that can reach up to 30% of total cellular proteins and utilize 50% of translation capacity [73–75]. The competition with native genes can cause a decrease in their expression and a reduction or cessation of growth [73–76]. In addition, because of the small numbers of RNAP and ribosomes, the expression of recombinant genes can become coupled, where a high level of expression of one gene titrates a resource and reduces the expression of another gene. In the context of synthetic signaling networks, this has been referred to as retroactivity, where downstream targets can impart a load on the upstream signaling pathway [77–80].

In this thesis, we describe two regulatory systems that could serve to separate synthetic genetic programs from their host context, leading to more robust system function:

The first system, described in Chapter 2, is the 'resource allocator'. This system utilizes T7 RNA polymerase, which partially decouples the gene program from host factors (namely, host RNAPs). The main advantage of the resource allocator is that it allows one to set the total number of T7 RNA polymerases that can be active and can transcribe the synthetic genes. In this way, an 'upper bound' of transcriptional activity for a genetic program can be set and maintained, ensuring that the system does not draw on too many host resources. By limiting a genetic program in this way, the host cell could be insulated from the potentially disruptive effects of overburdening, and overall system function and predictability could be

maintained.

In Chapter 3, we describe 'stabilized promoters', which are genetic regulatory elements that buffer gene expression against changes in copy number. By using an incoherent feedforward regulatory motif to detect and compensate for different copy numbers, the stabilized promoters enable consistent gene expression over a wide range of plasmid backbones and genomic integrations. Additionally, they appear to counteract changes in gene expression caused by perturbations that affect copy number. Genetic programs built with these elements should therefore be easier to compose and alter, as well as more robust to a number of contextual perturbations.

In the remainder of this chapter, we review the state of the art (circa 2013) of gene circuit engineering in bacteria. This technology served as a starting point for the regulatory systems that are described in later chapters. Section 1.2 lays out the genetic components that can be used to build gene circuits and section 1.3 describes the ways in which gene expression can be tuned and insulated from its genetic context.

1.2 Classes of regulators available for genetic programming of bacteria

An explosion in the number of well-characterized regulators available for building genetic circuits has occurred over the past few years. Before this, there were relatively few regulators available (e.g., LacI, TetR, AraC, and CI) and these were re-used in many designs. A major goal has been to expand the number of regulatory parts that are orthogonal, that is, do not cross react with each other such that they can be used together in a circuit [81, 82]. This has been achieved via two approaches. First, bioinformatics and whole gene DNA synthesis have been used to access regulators from sequence databases (part mining) [83]. Second, families of regulators have been characterized that are conducive to the rational design of orthogonal sets (e.g. RNA antisense regulation, zinc finger proteins, transcription activator like effectors, and CRISPR-Cas9). Computational methods have played a role both to predict the orthogonality of regulators identified in databases as well as in structure-guided design.

Collectively, this has resulted in well over one hundred regulators that could theoretically be used together in a single large circuit in a bacterium.

Despite efforts to standardize and collect data surrounding biological parts [84, 85], the majority of the information on different regulators is buried in individual papers, making it difficult to choose parts for desired circuit designs. In Table 1.1, we show data comparing 16 characterized regulator families currently available for building genetic circuits. For each family, the number of characterized orthogonal regulators is shown (with a metric of crosstalk), along with the dynamic range that has been achieved with that regulator. Note that the table focuses on bacteria and there is more data for some families in eukaryotic cells.

1.2.1 Protein regulators of transcription

Proteins that directly bind DNA to regulate transcription make up the majority of the regulatory parts available for use in bacteria (Figure 1-1a). One way in which proteins can regulate transcription is by initiating transcription at promoters. The native E. coli RNA polymerase (RNAP) can be directed to new promoters by expressing σ factors from other organisms [86, 87]. A large set of orthogonal σ factors has been generated through part mining, in which σ factors from many organisms were synthesized and their activities characterized [87]. Alternatively, the phage RNAP from T7 is often used and the promoter specificity of this polymerase has been changed to generate new parts through rational design and part mining [88–90] as well as random mutagenesis [91, 92].

Activators upregulate transcription by binding to a promoter to recruit RNAP. Classically, there are a number of natural activator proteins that have been used in genetic engineering, such as CI and LuxR. A small library of CRP activators was built by using bioinformatics to direct mutations at residues responsible for operator specificity [93]. Part mining has been applied to identify activators that require a second chaperone protein for activity and this has been used as the basis for building AND gates [70].

Repressors block transcription by blocking the binding or progression of RNAP. Recently, there have been efforts to increase the number of orthogonal repressors available for circuit design. To expand the LacI family, mutations were made to specific DNA residues in the

binding site and DNA binding residues in the protein and a set of 5 orthogonal repressors was selected [94]. Part mining has been applied to expand the number of available TetR homologues and this led to the identification of an orthogonal set of 16 repressors [95].

There are several classes of transcription factors that have modular protein structures that facilitate their engineering to target particular DNA sequences. Zinc finger proteins (ZFPs) and transcription activator like effectors (TALEs) have such a structure and have been particularly successful in being used in eukaryotic cells [96–101]. It has been surprisingly difficult to get these regulators to work in bacteria, but there are published examples of ZFPs being used as repressors and activators [102] and a TALE as a repressor in E. coli [103].

1.2.2 RNA regulators of translation

RNA parts that regulate translation take advantage of the fact that RNA base pairing follows a simple code that is computationally predictable [104] (Figure 1-1b). Two parts families of this type utilize antisense sRNA binding to alter the accessibility of the ribosome binding site (RBS) controlling translation initiation. RNA-IN/OUT parts consist of a modified natural system [105] in which an RNA molecule base pairs to the 5′-end of an mRNA (including the RBS) such that the ribosome cannot initiate translation [106]. The orthogonal set of these regulators was increased through computational design and experimentally confirmed. A second part family uses trans-activating RNAs that work by disrupting a secondary structure that blocks the RBS by default, leading to translational activation [107, 108]. Finally, it has been shown that the expression of modified 16S RNA that has been engineered to bind a non-canonical Shine Delgarno sequence can recruit ribosomes and this has been used as the basis to build gates [109–111].

When designing gates, the challenge with using RNA that acts on the level of translation is that it is difficult to convert an RNA input to an RNA output of the same form. Therefore, the resulting gates cannot be layered. An approach to this problem is to use a cis element that converts a translational signal into a transcriptional one [112]. This component utilizes a modified sequence from the E. coli tna operon [113] and makes transcription of a downstream region dependent on translation of a short peptide, effectively linking the two. It has been successfully used with sRNA translational regulation to make RNA parts that

regulate transcription.

1.2.3 RNA regulators of transcription

An exciting development over the last year has been the development of RNA-based systems that can directly regulate transcription by behaving as a repressor (Figure 1-1c). This system is based on Cas9, which is a protein that uses a small guide RNA to target a DNA sequence as part of the CRISPR/Cas bacterial immunity system [114]. Normally, Cas9 functions as a nuclease and cleaves DNA, but it was shown that if the nuclease activity is mutated, then the complex blocks RNAP. This can be used either as a repressor or as an activator by fusing an activation domain to Cas9 [115, 116]. An advantage of this system is the potential to design vast numbers of orthogonal regulators by building guide RNAs that target different operator sequences. The cross reactions that may result from this approach are just beginning to be characterized and understood [117, 118]. Once it is shown that the Cas9-based systems can be layered, this will become a powerful toolbox for circuit engineering. However, a practical challenge with using this system is the acute toxicity of Cas9 when expressed in many organisms.

A more developed approach for RNA to control transcription is based on the PT181 attenuation system [119–121]. When an antisense RNA is present and binds a target sequence on a transcript RNA, the nascent transcript folds into a transcriptional terminator and attenuates the message. This part family was expanded both through part mining and random mutagenesis, including utilizing some of the orthogonal RNA pairs from the RNA-IN/OUT system. These attenuators have been shown to be fully composable into cascades and logic gates [120].

1.2.4 Proteins that modify DNA

Thus far, DNA modification has been implemented in bacteria by using recombinases to invert or excise segments of DNA (Figure 1-1d). This mechanism is commonly found in bacteria, phages, and mobile genetic elements, providing a diversity of natural parts to exploit in synthetic systems [122]. The current parts for engineering DNA flipping and excision

are all natural recombinases, which vary in a number of ways. The most commonly used recombinases in genetic engineering are the simple tyrosine recombinases Cre and Flp [123]. Tyrosine recombinases can flip the region between their recognition sites in both directions, leading to an even distribution of orientations. Additionally, if their recognition sites are oriented in the same direction, they catalyze DNA excision. These two recombinases have been shown to function properly and orthogonally in bacteria [124]. The invertases FimB and Hin have also been used together in a bacterial system [125]. These invertases are similarly bidirectional, but they lack the capability for excision. Finally, phage integrases are a class of recombinases that catalyze unidirectional flipping, which leads to the accumulation of a specific DNA orientation [126]. Three integrases (Bxb1, ϕC31, and TP901-1) have been characterized and used to build bacterial circuits [127,128]. Additionally, a number of these integrases have matching excisionases, which allow for the reversal of DNA flipping [126,129].

In addition to part mining natural recombinases, there has been progress in generating new families of recombinase parts through modifications. One such approach is to iteratively make mutations to the DNA binding region in the recombinase and select for proteins with new specificities [130,131]. Another promising method is to generate new parts by creating recombinase fusions with zinc finger and TALE DNA binding domains [132,133].

Finally, while current efforts have focused on recombinases, other mechanisms of DNA modification may also hold promise for building circuitry. Recently, there have been successful efforts to selectively methylate DNA in bacteria [134] using a modular zinc finger design. An in vivo means of reading methylation state would open up these parts for use in genetic circuits. This seems feasible, as bacteria are known to contain many regulatory systems that respond to DNA methylation [135,136].

1.3 Precision gene expression in bacteria

Building, tuning, and connecting genetic circuits requires the ability to engineer precise changes in gene expression. Further, when gates are combined to build a circuit their genetic context changes, potentially impacting function [36]. There have been many recent advances in the development of tuning knobs that allow for the fine-tuned control of transcription

and translation (Figure 1-2). These can take the form of part libraries or computational tools. Further, insulators have been developed that reduce interference between parts. This is leading to a redefinition of the classical expression cassette.

1.3.1 Tuning knobs for expression

Promoters. Libraries of constitutive promoters for different species have been built by mutating the -10 and -35 RNAP binding regions of the promoter or the region affecting DNA melting (-10 to $+2$) [31, 32, 137–140]. Advances in oligonucleotide synthesis have enabled these libraries to become very large. For example, >10,000 combinations of promoters and 5-UTR's were built from a pooled oligonucleotide library and screened by combining cell sorting and deep sequencing (flow-seq) [31]. Computational models of promoters have also been developed that are based on the free energy of RNAP binding to the $-10/-35$ sites and promoter melting [141, 142]. These show promise in predicting promoter strength, however, a complete model that balances all contributions has yet to be built.

Ribosome binding sites. The RBS is a part that is relatively simple to tune to achieve different expression levels. As a result, it has been broadly applied to tuning response functions for building genetic circuits [60, 70, 86, 95, 143, 144]. The ribosome makes contacts with the Shine-Dalgarno sequence and start codon and binding is influenced by RNA base-pairing, the spacing between these regions, and the mRNA secondary structure. The RBS Calculator is a computational tool based on a biophysical model that balances these contributions [145, 146]. There are additional terms that influence the strength of the RBS and several of these, including the role of the standby site, have been characterized and incorporated into new versions of the software [146, 147]. There is much to be learned from non-canonical RBSs [148] including leaderless RNAs [149] and a better understanding of these processes could improve biophysical models. Further, libraries of 5-UTRs that include the RBS have been measured [31, 32], and a recent technique to tune RBS strength using hypermutable sequence repeats between the Shine-Dalgarno region and start codon was used to explore expression parameters for a bistable switch [144].

Terminators. During transcription, mRNA is released when RNAP reaches a terminator. Terminators are important in circuit design for three reasons. First, they offer a means to tune expression by modulating read-through and could potentially decrease leaky expression. Second, many terminators are required when gates are combined to build circuits. Circuits can have many transcription units, each of which needs strong termination to avoid interference with other circuit elements. These terminators must be sequence diverse to avoid recombination. Third, the recombinase-based memory circuits utilize unidirectional terminators as a core part of their design. To address these needs, there have been major efforts to use part mining to glean large libraries of terminators from bacterial genomes [150, 151].

Origins. To finely control plasmid replication beyond the standard plasmid systems, tunable-copy-number plasmids can be used. Increasing the expression of trans-acting replication factors (repA and pir) increases the plasmid copy number (for ColE2 and R6K origins, respectively). Using this system, a library of DIAL strains that constitutively express replication factors from the genome has been constructed that can yield between 1 and 250 plasmid copies by transforming the corresponding plasmids into the different strains [152].

1.3.2 Insulators to buffer the impact of genetic context

Part function is impacted by genetic context, or the sequences of the neighboring parts [31, 34–37]. In turn, this can impact the response function of the entire genetic circuit. Neighboring genetic context effects can have two forms. First, there is a direct interference of one part type on another. For example, the strength of an RBS is influenced by the promoter and the first codons of the expressed gene. Second, when two parts are combined a new function can appear at their interface. For example, promoters have been inadvertently constructed through the assembly of two parts containing an LVA-degradation tag, a DNA barcode, and a BioBrick scar [153]. To overcome this issue, insulator parts have been developed to diminish the effect of genetic context. Figure 1-2 shows a conceptual re-visiting of the expression cassette, where insulators are strategically placed between key parts.

The first insulators in the diagram are based on bidirectional terminators, which flank the expression cassette to reduce transcriptional read-through into or out of the expression cas-

sette [150]. The next type of insulation addresses the issue that promoters used in synthetic biology are typically too small and only capture the -35 and -10 regions. This can cause the promoter to have different strengths depending on the up and downstream sequence. Longer promoters should be used that at least encompass the UP element (-35 to -64) that binds the -subunit of RNAP [141,154]. Taking this further, it has been shown that the addition of upstream insulator sequences (up to -105) and downstream insulator sequences (down to $+55$) further standardizes promoter activity [34].

The transcription start site of promoters is not always known and a promoter part is rarely annotated to end at the $+1$ position. This can be further complicated by the observation that there is sometimes a distribution of mRNA produced from the same promoter and single mutations to the promoter can change the start site. Additionally, the RBS is particularly sensitive to changes in the 5-UTR, even by a single nucleotide. Both self-cleaving ribozymes [36] and CRISPR RNA-processing [35] have been used to physically cut and detach the variable 5-UTR from the mRNA, thereby shortening and making constant the 5-mRNA context. These tools are particularly important when combining gates to build a transcriptional circuit, where the promoter inputs and output of each gate occur in new contexts.

A bicistronic RBS sequence has been shown to reduce the impact of the protien coding sequence on the 5-UTR secondary structure including the RBS [37]. A small leader peptide-coding sequence with its own RBS is positioned upstream from the gene of interest such that the peptide sequence overlaps the RBS for the downstream gene and the peptide's stop codon ends at the gene of interest's start codon. The natural helicase activity of ribosomes loaded at the first RBS unfolds the mRNA secondary structure near the second RBS that controls gene expression, decoupling the translation initiation rate of the second RBS from the downstream coding sequence.

Lastly, the variability in gene expression when coupled with various terminators can be reduced by encoding an RNase III site in the 3-UTR of the mRNA [151, 155]. Post-transcriptional processing of mRNAs by RNase III standardizes the 3-end of the mRNA, such that the sequence and secondary structure of the cleaved RNA can no longer contribute to mRNA stability and degradation.

1.4 Conclusions

The existing technologies for building gene regulatory circuits in bacteria laid the groundwork for the regulatory systems that we have constructed. We drew on the toolbox of existing genetic parts when constructing both the resource allocator and stabilized promoters. Based on our analysis of the available options, we selected protein regualators of transcription as the central components in these systems. This design choice allows for maximum flexibility both when constructing and using these devices. First, because the regulators are proteins, their expression level can be controlled with the full set of tuning knobs (Figure 1-2b-e). Second, since we are controlling transcription of our gene of interest, it is possible to connect our systems with a protein or RNA downstream module, and the full set of regulators that modify translation can be layered on top of our systems.

The insulators shown in Figure 1-1f-i serve as the foundation for the stabilized promoter project. While the existing insulators are integral in ensuring well-behaved genetic systems (and we used a number of them in our efforts), thay are limited to local genetic context. That is, they can only counteract the effects of the flanking DNA sequences on the function of gene expression systems. The stabilized promoters that we built were designed to function analogously to these insulators but have the capability to buffer against wider types of context.

Finally, we hope that our engineering efforts can feed back and expand the toolbox for gene circuit engineering in addition to yielding useful regulatory systems. In the course of engineering the resource allocator and stabilized promoters, we developed and characterized new regulators (split T7 RNAPs and improved TALE repressors). These parts provide unique capabilities when building genetic systems, and we hope that others will use them in their own engineering projects. Additionally, while engineering stabilized promoters, we built and characterized a number of variant plasmid backbones that can be used as a tuning knob when building new genetic systems. These join the inducible orgins shown in Figure 1-2e as a way to quickly test genetic systems at a range of copy numbers.

1.5 Figures & Tables

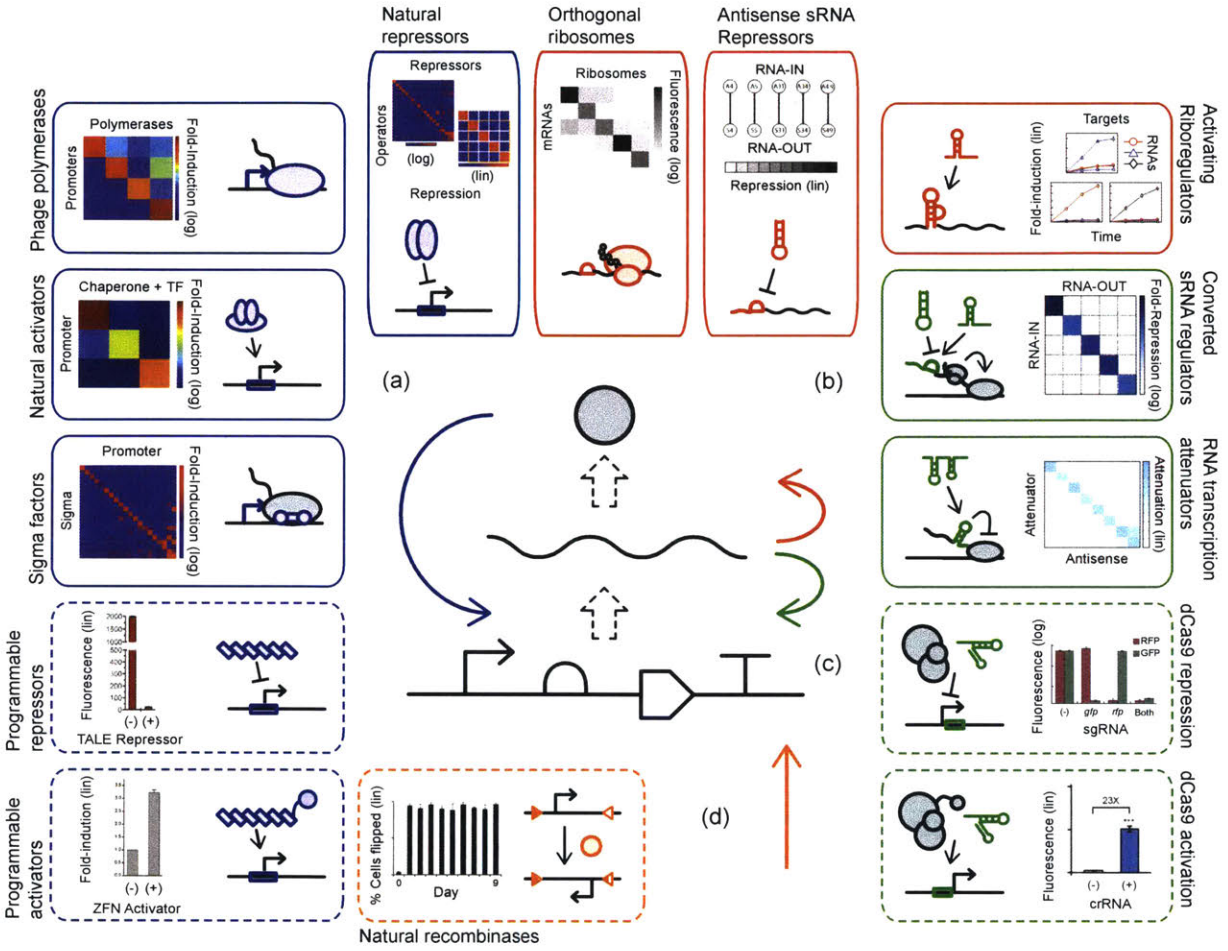


Figure 1-1: **The diversity of genetic regulatory parts available for building bacterial genetic circuits.** Schematics and representative data for a selection of different circuit-building mechanisms are shown. The colored regions in the schematic indicate the variable regions that make up each part. A solid line surrounding a part type indicates that there is a well-characterized, orthogonal set of parts of this type available for circuit building. A dashed line indicates that there is a proof-of-concept part. The data shown either demonstrates the orthogonality and size of a parts family, or if that is not available, it shows proof-of-concept activity.
(Continued on next page.)

Figure 1-1: (Continued)
a. Protein parts that act on transcription include natural repressors and activators, phage polymerases, σ factors, and repressors and activators based on programmable DNA binding proteins. Figures adapted from: Natural repressors [94,95], Phage polymerases [88], Natural activators [70], σ factors [87], Programmable repressors [103], Programmable activators [102].
b. RNA parts that act on translation include orthogonal ribosomes, antisense sRNA repressors, and activating riboregulators. Figures adapted from: Orthogonal ribosomes [109], antisense sRNA repressors [106], activating riboregulators [108].
c. RNA parts that act on transcription include sRNA regulators converted to affect transcription, antisense RNA transcriptional attenuators, dCas9 repression, and dCas9 activation. Figures adapted from: Converted sRNA regulators [112], RNA transcription attenuators [119], dCas9 repression [115], dCas9 activation [116].

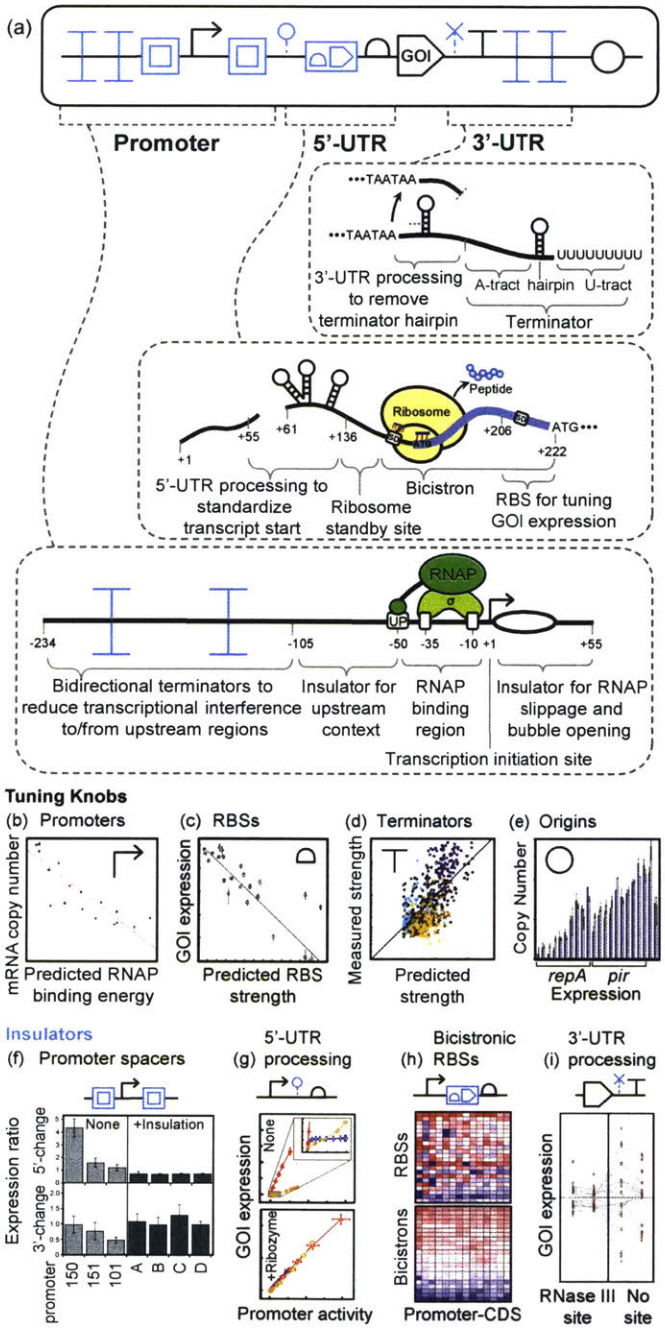


Figure 1-2: **A modern expression cassette comprising genetic tuning knobs and insulators.** a. cassette is shown for precise expression of a gene of interest (GOI). Insulating parts are highlighted in blue. The cassette is shown using symbols from the synthetic biology open language visual (SBOLv).

b. Biophysical models of transcription based on the RNAP binding energy have been constructed to complement empirically characterized promoter libraries. Figure adapted from [142].

(Continued on next page.)

Figure 1-2: (Continued)
c. The RBS Calculator provides a computational framework for designing RBS sequences of a given strength based on a thermodynamic model of translation initiation. Figure adapted from [145].
d. Terminator strength is partially informed from a biophysical model relating various aspects of the terminator sequence. Figure adapted from [150].
e. Tunable copy-number plasmids allow for a wide range of gene expression based on the repA/ColE2 and pir/R6K plasmid systems. Figure adapted from [152].
f. Promoter-insulating sequences have been shown to reduce the change in promoter activity when upstream and downstream sequences are introduced to the flanking promoter context. Figure adapted from [34].
g. Ribozymes have been used to improve the predictability of gene expression by reducing promoter/5-UTR coupling effects. Figure adapted from [36].
h. Bicistronic RBS designs cause the rank order of expression constructs to be more predictable compared to single RBSs. Figure adapted from [37].
i. RNase III sites in the 3-UTR reduce the variability in gene expression for reporters coupled with libraries of terminators. Figure adapted from [151].

Table 1.1: Characteristics of regulatory part families currently available for constructing genetic circuits in bacteria

A	B	Characterized family size	Maximum dynamic range	Largest tested parts set with minimum orthogonal range				Data references	Notes
				2x	5x	10x	50x		
DNA modification									
	Recombinases	2[d]	92	2	2	2		[128][b]	Nonlinear, implements passive memory
Transcription regulation									
Protein parts	T7 polymerases	4	270	4	4	3		[88][b]	
	TetR-family repressors	20	210	16	15	12	7	[95][a]	Nonlinear
	LacI repressors	5	45	5	5	4		[94][c]	Nonlinear
	Chaperone activators	3	69	3	2	2		[70][a]	Each chaperone-activator pair can be used as an AND gate
	ECF σ factors	27	470	17	15	13	9	[87][a]	A set of anti-σ factors enables sequestration
	TALE repressors	1+[e]	110					[103][a]	
	Zinc finger activators	1+[f]b	3					[102][b]	
	Zinc finger repressors	1+[f]	8					[102][b]	
RNA parts	Converted RNA-IN/OUT	5	880	5	5	5	5	[112][b]	
	PT181 attenuators	14	8	8	2			[119,120][a]	Attenuators can be repeated to increase signal
	dCas9 repression	1+[g]	340					[115][b]	Requires dCas9 expression
	dCas9-ω activation	1+[g]	23					[116][a]	Requires dCas9-expression
Translation regulation									
RNA parts	RNA-IN/OUT	23	10	13	6			[106][a]	Additional large sets of RNAs computationally designed
	Activating riboregulators	4	45	3	2	2		[108][b]	
	O-ribosomes	5	78	5	5	3		[109][c]	

This table presents a brief comparison of a number of regulatory part families that have been characterized and are available for use in building genetic circuits in bacteria. Each part is defined as a trans-acting element and the target of this element: for example a transcriptional repressor and its binding site. Part families were chosen to either have at least three characterized members or be based on a technology proven to be extendable (TALE, zinc finger, and CRISPR-based parts). The characteristics were derived from the cited sources as follows:

Characterized family size indicates the number of parts of this type that have been characterized for crosstalk.

A '+' indicates that this parts family is based off of a technology proven to enable orthogonal, programmable DNA binding and so the parts set may be predictably extendable.

Maximum dynamic range is the largest reported fold change between the on and off states (i.e. with the trans-acting element present and absent) of a single member of the part family.

The largest tested parts sets show the largest number of parts in each family that have been shown to function above specific thresholds of orthogonal range. Orthogonal range is a conservative measurement of the orthogonality of a parts set; it represents the fold change between the on-target effect of a part and the worst off-target effect on that part. For example, the T7 polymerases family has a 3 part set that functions above 10x orthogonal range, meaning that there is a group of three polymerases where each activates its target promoter to a level more than 10x the level that either of the other two polymerases activate it.

a. Numerical data was used from this reference for the part family in this row.

b. Data computed from bar or line plots in this reference was used in this row.

c. Data was read from colored orthogonality grids in this reference. Note that the numbers in this row may be less accurate because of uncertainties in this method.

d. While only two recombinases have been tested together at a time, at least seven have been used in genetic circuits in bacterial systems [124, 125, 127–129, 156].

e. TAL repressors and activators are much more widely used in eukaryotes. A set of 8 orthogonal TAL activators has been tested in mammalian cells, and it is predicted that many more could be built [101].

f. Zinc fingers have also been widely used in eukaryotes [99, 157].

g. CRISPR repression and activation is being widely adopted in eukaryotes. Crosstalk data from these organisms suggests that many orthogonal variants could also be made in bacteria [117, 118, 158].

Chapter 2

A 'resource allocator' for transcription based on a highly fragmented T7 RNA polymerase

2.1 Abstract

Synthetic genetic systems share resources with the host, including machinery for transcription and translation. Phage RNA polymerases (RNAPs) decouple transcription from the host and generate high expression. However, they can exhibit toxicity and lack accessory proteins (σ factors and activators) that enable switching between different promoters and modulation of activity. Here, we show that T7 RNAP (883 amino acids) can be divided into four fragments that have to be co-expressed to function. The DNA-binding loop is encoded in a C-terminal 285-aa σ fragment, and fragments with different specificity can direct the remaining 601-aa core fragment to different promoters. Using these parts, we have built a resource allocator that sets the core fragment concentration, which is then shared by multiple σ fragments. Adjusting the concentration of the core fragment sets the maximum transcriptional capacity available to a synthetic system. Further, positive and negative regulation is implemented using a 67-aa N-terminal α fragment and a null (inactivated) σ fragment, respectively. The α fragment can be fused to recombinant proteins to make promoters re-

sponsive to their levels. These parts provide a toolbox to allocate transcriptional resources via different schemes, which we demonstrate by building a system which adjusts promoter activity to compensate for the difference in copy number of two plasmids.

2.2 Introduction

The challenge posed by synthetic geenetic systems sharing resources with the host was recognized early in biotechnology and a partial solution emerged by using the RNAP from T7 phage to decouple transcription from the host machinery [159–161]. Heterologous T7 RNAP was patented in 1984 [162] and since then has been the basis for expression systems across many organisms [163–166]. An advantage cited for this system was that it could achieve high expression levels by adding an inhibitor of *E. coli* RNAP, thus directing metabolic resources to recombinant protein production [167]. However, there are also some challenges with using T7 RNAP. While the polymerase itself is not toxic, when it is combined with a strong promoter, it can cause severe growth defects. The origin of this toxicity is not clear, but it could be related to the rate of transcription of T7 RNAP, which is eightfold faster than *E. coli* RNAP and could expose naked mRNA [168, 169]. Toxicity can be ameliorated by introducing a mutation near the active site and by selecting parts to lower polymerase expression [69, 88]. Beyond the RNAP from T7, many polymerases have been identified from different phage and directed evolution experiments have yielded variants that recognize different promoter sequences [88, 170, 171].

Phage polymerases are central to our organization of larger genetic systems [19, 69, 88]. We separate the regulation of a system (on a plasmid we refer to as the controller) from those genes encoding pathways or cellular functions (actuators) (Figure 2-1A). The controller contains synthetic sensors and circuits, whose outputs are phage polymerases specific to the activation of the actuators. This organization has several practical advantages. First, it avoids evolutionary pressure when manipulating the actuators because the promoters are tightly off in the absence of phage polymerase. Thus, they can be carried in an inactive state until the controller is introduced into the cell. Actuators often require many genes and assembled parts, making re-verification of their sequence expensive. Second, it allows the

regulation of the actuators to be changed quickly. Controllers can be swapped to change the conditions and dynamics of expression, so long as they produce the same dynamic range in output polymerase expression. In the same way, the controllers can also be characterized independently using surrogate fluorescent reporters prior to being combined with the actuators.

With these large and complex synthetic systems, problems can arise as the host is subjected to significant perturbation and load. Simultaneously activating a number of actuators requires expressing multiple polymerases that might collectively cross the threshold for toxicity (Figure 2-1B). While lowering expression rates throughout the system could avoid toxicity, it would needlessly constrain expression when only one actuator is active. To address this issue, we aimed to create an allocation system that allows independently setting the total desired polymerase activity and allocating this resource to the various actuators as needed. With this organization, a single actuator can be expressed to full strength, but expression of multiple actuators is attenuated to avoid overexpression (Figure 2-1C). In effect, we are proposing to add another layer to the organization of genetic designs, where a separate resource allocator is responsible for the maintenance of a desired level of orthogonal transcriptional machinery (Figure 2-1A).

Prokaryotes solve the problem of partitioning a budget of RNAP to different cellular processes through the action of σ factors, which bind to the core RNAP (α 2, β , β', and ω subunits) and direct it to promoter sequences [172, 173]. Core RNAP itself only has the ability to non-specifically bind to DNA, whereas the σ factor contains the DNA recognition domains for the 35 and 10 regions of promoters. Different σ factors bind to distinct promoter recognition sequences. In *E. coli*, there is one housekeeping σ factor (σ^{70}) that is expressed at a constant level of 500–700 molecules/cell, independent of growth phase or stress, and 6 alternate σ factors that control various stress responses (e.g., heat shock) and cellular functions (e.g., flagella assembly) [174]. σ factors can range in size; σ^{70} is 613 amino acids and the average alternative σ is 200 amino acids [87, 175, 176]. These alternative σ^{70}s can be embedded in complex regulatory networks that implement signal integration and feedback regulation that mimics engineering control architectures [177–179]. In this way, the level of core RNAP dictates the total transcriptional potential in the cell, while the relative levels

of σ factors determine how this resource is allocated between growth and stress resistance [180, 181]. Bacteria with more diverse lifestyles can have significantly more σ factors, for example, Streptomyces and Bacteroides species can have greater than 50 [177–179]. All of these σs compete to bind to the core RNAP [58, 172].

In this manuscript, we have created an analogous system by fragmenting T7 RNAP. We used a transposon method to identify five regions where the polymerase can be bisected and retain function. One of these splits produces a 285 amino acid fragment that we refer to as the σ fragment because it contains the region that binds to the promoter (Figure 2-1D). We find that variants of this fragment with different promoter specificities can bind to the remaining core fragment and direct it to different promoters. The expression level of the core fragment dictates the maximum number of active polymerases. The outputs of the controller are different σ fragments, which are used to turn on different actuators. If the pool of core fragments is saturated by σ fragments, the total number of active polymerases in the system will remain constant regardless of the levels of σ fragments being produced (Figure 2-1C). In this way, a desired transcriptional load can be specified and then dynamically allocated to different actuators as the conditions require. Negative regulators can be built by creating null σ fragments that titrate the core fragment but do not support transcription. Additionally, the core fragment can be positively regulated using the N-terminal bisection point to separate an fragment that is required for activity. These regulators could be used to implement feedback loops that control the amount of active RNAP complexes under different conditions or the dynamics of signal progression from the controller to the actuators.

2.3 Results

2.3.1 Bisection mapping of T7 RNA polymerase

Our first objective was to identify all of the places T7 RNAP could be split to yield two fragments that can be co-expressed to produce a functional protein. To do this, we developed a transposase-based method that uses a novel transposon to split proteins, which we refer to as a splitposon. Previous methods have been published to generate libraries of split proteins

or domain insertions that are based on incremental truncation [182, 183], multiplex inverse PCR [184], DNAse cleavage [185, 186], and transposon insertion [187, 188]. The transposon-based approaches are able to generate large libraries and do not require sensitive DNAse steps, but they leave ~10 added amino acids at the split site. To improve on this approach, the splitposon is a Mu transposon in which one terminal transposon recognition end is altered to contain a non-disruptive ribosome binding site (RBS) and start codon (Figure 2-2A). We further modified the transposon to add the remaining necessary regulation to divide a protein into two fragments (stop codon–P_{Tac} IPTG–inducible system–RBS–start codon). The MuA transposase efficiently yields random insertions of the splitposon throughout a DNA molecule, producing a library of split proteins flanked by just three additional amino acids for in-frame insertions (Figure 2-6).

With the splitposon, a bisection library for any protein can be generated in two steps (Figure 2-2A). First, the splitposon is transposed *in vitro* into a plasmid containing the DNA within which bisections are desired (e.g., a gene or segment of a gene). Second, the target region is digested from the plasmid backbone and size selected for fragments containing an inserted transposon. These fragments are ligated into an expression plasmid containing an upstream inducible promoter. The final library will contain only plasmids with a single transposon insertion in the region of interest and can be induced and screened for function.

The splitposon method was applied to generate a library of bisections of a variant of T7 RNAP (T7* RNAP). This gene contains the R632S mutant, which reduces host toxicity [88]. To avoid trivial truncations of the termini, we directed transposon insertions to the region of the gene corresponding to amino acids 41 through 876 of the polymerase. Both fragments are induced with IPTG from P_{Tac}. The library was co-transformed with a screening plasmid that contains a T7 RNAP dependent promoter and red fluorescent protein (RFP) [88], and 384 clones were picked by eye from agar plates, re-assayed in liquid media, and the best 192 sequenced. From these, 36 unique in-frame split sites were identified (Figure 2-2B). The split sites cluster into five distinct seams that correspond to six potential fragments if they were all implemented simultaneously. The seam around position 179 corresponds to a previously identified split site that yields a functional T7 RNAP [90, 189–191].

2.3.2 Division of T7 RNAP into multiple fragments

All of the discovered split seams occur in surface-exposed regions of the T7* RNAP, and the largest seam corresponds to a large surface-exposed loop known as the Flap in the 3-dimensional structure (Figure 2-8) [192]. This implies that additional functional domains can be inserted at these positions. We hypothesized that the addition of protein–protein interaction domains could improve the affinity of the fragments. To this end, two leucine zipper domains that bind in an antiparallel orientation were chosen from the SynZIP toolbox (variants 17 and 18) [193, 194]. Addition of either SynZIP at the 601 split site with a short flexible linker is tolerated by the split polymerase, and adding both is beneficial and improves activity by greater than tenfold at low expression levels (Figure 2-2C).

The outcome of the bisection mapping experiment also implied that it might be possible to divide T7* RNAP into more than two fragments. First, the protein was divided into three fragments based on the split points at residues 67 and 601, including the added SynZIPs at the 601 split. These three fragments were expressed as a single inducible operon and compared to versions lacking each of the single fragments. RNAP activity (4,000-fold induction) is only detected when all three fragments are expressed and there is no activity in the absence of any fragment (Figure 2-2D). We also tested a four fragment version, which includes a split at position 179 (Figure 2-2E). The expression of these four fragments yields active RNAP (900-fold induction), and there is no detectible activity if any of the fragments are not expressed.

While the four and three-piece polymerases do lead to a reduction in cell growth when expressed at high levels, this effect is more pronounced when expressing the full-length protein (Figure 2-17). Splitting the polymerase into five or six fragments was not attempted due to the attenuation of activity and growth impact of high expression with four fragments.

2.3.3 Construction of σ fragments with different promoter specificities

The C-terminal fragment generated by the split site at residue 601 (601–883) contains the DNAbinding loop that determines promoter specificity [195]. Thus, we refer to this as

the σ fragment as it functions analogously to σ factors that bind to *E. coli* RNAP and is approximately the same size. Following this analogy, the 601 amino acid N-terminal fragment is referred to as the core fragment. Note that this fragment is much smaller than the $\alpha 2/\beta/\beta'/\omega$ subunits of *E. coli* RNAP (329/1342/1407/91 amino acids) and they assemble into a very different 3-dimensional structure [196–198].

A simple resource allocator was built based on the core and σ fragments (Figure 2-3A), retaining the amino acids added by the splitposon method and the SynZIP 18 domain on the σ fragment. The core fragment is expressed from the constitutive promoter P_{J23105}, tuned to a low level such that expressing full-length polymerase in its place is not toxic. The σ fragment is expressed at varying levels using an IPTG-inducible P_{Tac} promoter. Polymerase activity is measured using P_{T7} driving green fluorescent protein (GFP) (See 2.6). The σ fragment, core fragment, and reporter are carried on three separate plasmids (p15A*, BAC, pSC101) to mimic the controller, resource allocator, and actuator organization (Figure 2-1A).

For the resource allocation scheme to function correctly, σ fragments need to saturate the core fragment, causing total RNAP activity to plateau above a certain total concentration of σ fragments. The maximum level of polymerase activity is then set by the concentration of the core fragment, independent of changes in σ fragment expression (Figure 2-1C). Core fragment expression, and thus overall maximum functional polymerase expression, can be modulated by selecting constitutive promoters and RBSs of different strengths. This saturation behavior is observed when the core fragment is fused to the SynZIP 17 domain (Figure 2-3B, red points). The RNAP activity saturates approximately fourfold below that obtained with the expression of full-length T7* RNAP in place of the core fragment, which does not change as a function of σ fragment expression (green points). Since the full-length T7* RNAP is expressed at a level equivalent to the core fragment, this indicates that the split polymerase with SynZIPs has about one quarter the activity of full-length T7* RNAP. Without the SynZIP domain on the core fragment, the σ fragment binds with much lower affinity and does not reach saturation even at high levels of expression (blue points). Because the desired saturation of the core fragment is obtained only with the SynZIPs, they were used in all further experiments.

A key feature of the allocator is to be able to direct transcriptional resources to different

actuators. This requires multiple σ fragments that can bind to the core fragment to change its promoter affinity. These σ fragments need to be orthogonal, that is, they cannot cross-react with each other's promoters. Initially, we attempted to base the orthogonal σ fragments on a set of specificity loop mutations previously shown to generate orthogonal variants of full-length T7 RNAP [88]. These specificity loops are based on polymerases from the T3, K1F, and N4 phages. We tested the corresponding σ fragments and mutated promoters. Unfortunately, of these variants, only the σ fragment containing the T3 specificity loop and corresponding promoter (Figure 2-3C) generated an activity comparable to that of the T7 σ fragment (Figure 2-3D).

The σ fragments based on the K1F and N4 specificity loops did have some residual activity. This was used as a basis to apply error-prone PCR to the σ fragments to search for mutations that increase activity (See 2.6). One mutation was found for the K1F loop (K1FR: M750R) that recovered activity to a sufficient level, but similar efforts with the N4 loop proved unsuccessful (See 2.5.3). An additional σ fragment was built based on an orthogonal T7 RNAP variant (CGG-R12-KIR) that was identified from directed evolution experiments [170]. This produced a comparable activity to the other σ fragments (Figure 2-3D). In total, four σ fragment variants (T7, T3, K1FR, and CGG) and cognate promoters were built. It is noteworthy that the σ fragments only differ in sequence by 5–10 amino acids (Figure 2-3C). Expression of each σ fragment with its cognate promoter and the same level of core fragment shows that their activities fall into a similar range with less than a fourfold difference between the strongest (T7) and weakest (T3) σ fragments (Figure 2-3D). The four σ fragments were also found to be orthogonal (Figure 2-3E), and their expression to saturation with the core fragment does not lead to growth defects (Figure 2-15).

2.3.4 Setting and sharing the transcriptional budget

The expression level of the core fragment from the resource allocator sets the maximum number of active RNAPs in the synthetic system. This budget has to be shared between σ fragments that are expressed simultaneously (Figure 2-1C). To test this, we built a plasmid where the K1FR σ fragment is expressed from P_{Tet} and the T3 σ fragment is expressed from P_{Tac} (Figure 2-4A). By inducing the system with IPTG, the level of expression of the T3

σ fragment is varied while the K1FR σ fragment is maintained at a constant level (P_{Tet} is uninduced but has leaky expression). In essence, this captures the scenario where one output of a controller is constantly on at a saturating level and then another output turns on and competes for the RNAP resource. To report how much of each type of polymerase complex is present in the system, reporter plasmids that express GFP from P_{T3} and P_{KIF} were used. The activity of the σ^{T3}:P_{T3} and σ^{K1FR}:P_{K1F} pairs are very similar (Figure 2-3D), making it possible to compare their expression levels.

Core fragment expression was driven by the P_{J23105} promoter with RBSs of different strengths. Initially, a strong RBS was chosen that sets a high expression level of the core fragment (Figure 2-4B). The K1FR σ fragment utilizes the majority of the core fragment budget before the T3 σ fragment is induced. As the T3 σ fragment is induced, it competes for the core fragment. At high concentrations, it saturates the pool of core fragment, almost completely titrating it from binding to the K1FR σ fragment. The sum of the P_{K1F} and P_{T3} promoter activities (gray points) remains constant and is independent of the expression of either σ fragment. The competition experiment was repeated with the core fragment expressed at a lower level from a weaker RBS (Figure 2-4C). Importantly, the expression level of the K1F σ fragment and the induction of the T3 σ fragment remain unchanged. As before, the sum of activities from the P_{T3} and P_{K1F} promoters remains constant. Both of these competition systems are tolerated by cells with little growth impact at the induction levels used (Figure 2-16).

The shapes of the curves are essentially identical when compared for high and low concentrations of the core fragment. The similarity is shown by plotting the P_{T3} and P_{K1F} promoter activities with low core fragment expression against their activities with high core fragment expression (Figure 2-4D). This results in a linear relationship, meaning that all promoter activities scale equally with the amount of core fragment expressed. The slope of this line indicates that the low level of core fragment yields approximately 36% of the activity compared to the high level. Hence, the budget is shared identically between the σ fragments at each core fragment expression level. This property means that the proportional outputs of the resource allocator can be set independently from the level of resource being produced.

To correct for the slight activity difference between the T3 and K1FR systems, we nor-

malized the P_{T3} and P_{K1F} activity values by the activity when each individual σ fragment is expressed to saturation with the appropriate resource allocator (Figure 2-4E). Assuming that promoter activity is linearly proportional to the number of active polymerases, these normalized values represent the proportion of the available core fragment bound by each of the σ fragments. A mathematical model of the system was built and its dynamics analyzed (See 2.7). When the core fragment is fully saturated by σ fragments, the model predicts that the proportion of the core fragment bound by each σ fragment should depend solely on the relative expression levels of each σ fragment. The simplified model has only one parameter not measured in the normalized data set: the relative expression of the K1FR σ fragment (Equations 2.29 and 2.30). Fitting this parameter yields a good agreement between the theory and experimental data (Figure 2-4E, Equations 2.31, 2.32, 2.33).

2.3.5 Positive and negative regulation of the core fragment

The resource allocators shown in Figures 2-3 and 2-4 maintain a constant level of core fragment. It is desirable to be able to dynamically shift the budget up or down, for example, to control the maximum transcriptional capacity as a function of media or growth phase. To do this, we used additional splits and mutations to create positive and negative regulators. These regulators could also be used to design feedback or feedforward circuits to implement control algorithms that act on the signal from the controller plasmid to the actuators.

The negative regulator is based on a null σ fragment that binds to the core fragment but does not support transcription. This functions to sequester the core fragment in the same way as an active σ fragment, making less of it available to the other competing σ fragments. Sequestration has emerged as a generalizable method to tune the threshold and ultrasensitivity of genetic circuits by setting a concentration of sequestering molecule that must be outcompeted before the circuit turns on [86, 87, 199, 200]. The null fragment was identified by testing amino acid substitutions and deletions identified from the literature to disrupt T7 RNAP function [201, 202]. These mutations were selected to disrupt transcription activity without impacting the ability of the σ fragment to bind and sequester the core fragment (Table 2.4). Based on the screen, we identified the Y638A mutation in the CGG σ fragment as having the strongest effect when sequestering the core fragment. This fragment was

confirmed to carry no residual activity for its original promoter (Figure 2-11).

A system was constructed to test the ability of the null fragment to titrate the core fragment and reduce its availability to the σ fragments (Figure 2-5A). For this, the σ fragments were expressed using a constitutive promoter derived from P_{J23119} and the null fragment was placed under P_{Tac} IPTG-inducible control on a separate plasmid. When expressed with the T7 σ fragment, the null fragment decreases the activity from P_{T7} as it is induced (Figure 2-5B). The null fragment is able to compete with all of the σ fragments and reduces each of their activities by at least tenfold when fully induced (Figure 2-5C).

The positive regulator is based on further splitting the core fragment at the most N-terminal split site (Figure 2-2B and 2-2D). This divides the core fragment into two pieces: a short 67 amino acid α fragment and a larger 586 amino acid β core fragment (including the SynZIP). The α fragment can be expressed separately and is required for activity. It can be used to modulate the fraction of the polymerase pool that is active. Note that it still does not enable more transcriptional activity than is set by the amount of β core fragment that is expressed. Thus, the maximum can be set and then the α fragment used to modulate the amount that is available at any given time.

A system was constructed to assay the α fragment's ability to regulate the polymerase budget (Figure 2-5D). The β core fragment is expressed from the P_{J23105} constitutive promoter on a low copy plasmid, while the T7 σ fragment is expressed from a constitutive promoter derived from P_{J23119} on a high copy plasmid. The α fragment is expressed from P_{Tac}. There is no T7 RNAP activity without the α fragment and activity increases as it is induced (Figure 2-5E).

2.3.6 Coupling RNAP activity to the concentration of arbitrary α fragment tagged proteins

Since the α fragment is relatively small (67 aa) and required for polymerase function, we hypothesized that it would be useful as a protein tag to activate transcription proportional to the level of an arbitrary protein of interest. While the C-terminus of T7 RNAP catalyzes transcription and is highly sensitive to alteration, the N-terminus (where the α fragment is

located) is much more tolerant to modifications [203]. The α fragment was fused to proteins of interest via a GGSGG flexible linker. Fusion to either the N- and C-terminus of RFP or GFP makes polymerase activity responsive to the level of fluorescent protein expression (Figures 2-5F and 2-12). This may be used to tag proteins in a synthetic system or the host, enabling the readout of an internal or cell state.

2.3.7 Application of the α fragment to compensate for differences in copy number

A challenge in building genetic systems is that regulatory parts will change their activity depending on the copy number of the system. For example, a constitutive promoter will produce a high level of expression when it is placed on a high copy plasmid and a low level of activity with placed at single copy on a bacterial artificial chromosome [152]. The α fragment could be used to regulate the activity of the polymerase to adjust the activity of promoters and compensate for the copy number at which they are carried due to different plasmid origins (or in the genome). The idea is to combine the phage promoter(s) with an expression cassette including the α fragment that is expressed at a level inversely proportional to the copy number (Figure 2-5G). In other words, a strong promoter and RBS would be selected to drive the expression of the α fragment from a low copy plasmid and vice versa.

Plasmids were constructed on pSC101 and pUC backbones that contain a P_{T7} promoter driving GFP expression and a α fragment expression cassette. We mutagenized the RBSs and altered the promoters and start codon of the α fragment expression cassettes to identify a strong cassette that would be carried on the pSC101 plasmid and weak cassette that would be carried on the pUC plasmid (See 2.6). With these different levels of α fragment expression, we were able to achieve nearly identical activities for P_{T7} in the different plasmid contexts when they are used with the β core fragment (Figure 2-5H). In contrast, when the plasmids are used with the full core fragment, which does not need the α fragment to function, high expression is seen from the high copy pUC backbone and low expression is seen from the low copy pSC101 backbone.

One of the values of this approach is that it enables actuators that require multiple phage

promoters to be moved to different copy number contexts without having to change and rebalance each of the promoters. For example, actuators that produce deoxychromoviridans, nitrogenase, and lycopene require 2, 4, and 5 phage promoters [69, 88]. These could be moved to different copy number backbones without changing their genetics by changing the expression level of the α fragment from that backbone. One can also imagine harnessing feedback or feedforward loops that self-adjust the level of α fragment to maintain constant promoter activity independent of context, similar to systems that have been implemented in mammalian cells [204].

2.4 Discussion

As a means to organize and control large genetic engineering projects, we propose to introduce a separate resource allocator module. The allocator is responsible for providing resources that are orthogonal to those required by the host for growth and maintenance. To that end, this manuscript focuses on budgeting transcriptional resources through the control of phage polymerase activity and promoter specificity. Thinking ahead, this approach can be extended to budget additional resources. For example, translational resources could be incorporated by controlling a orthogonal rRNA [110, 111] (specific to RBSs only in the synthetic system) or even introducing an entire second ribosome. Extending this idea, it may be possible to incorporate orthogonal tRNAs [205, 206], DNA replication machinery [207], protein degradation machinery [208], carbon precursors [209], and organelle structures [210,211]. While this never completely decouples the synthetic system from the host, it systematically reduces its dependence on host resources and genetic idiosyncrasies. This approaches the concept of a virtual machine for cells, where synthetic systems would bring all of the necessary cellular machinery with them. This concept will become critical as designs become larger, moving toward the scale of genomes and requiring the simultaneous control over many multi-gene actuators.

This work demonstrates an incredible tolerance of the T7 RNAP structure for division into multiple proteins without disrupting its function. To our knowledge, this is the first time that a protein has been artificially divided into four fragments that can be function-

ally co-expressed. This tolerance is surprising because T7 RNAP is known to undergo large-scale conformational changes as it proceeds from promoter binding to transcription elongation [212, 213]. The residues involved in these conformational changes occur toward the N-terminal region but are distributed across the first three fragments of the 4-fragment polymerase (Figure 2-2E). All of the RNAP split points were discovered simultaneously using a new experimental method, which we refer to as a splitposon. This approach is faster, simpler, and produces more accurate split proteins than previous methods. Split proteins have applications in genetic circuits [90, 188], plasmid maintenance with fewer antibiotics [214], and biosensors [215–219].

The fragments of T7 RNAP are used to implement regulatory control. A C-terminal fragment contains the DNA-binding loop and we demonstrate that fragments with different specificities can direct the RNAP to different promoters. For this reason, and because of its size, we draw a loose analogy to the role of σ factors in native prokaryotic transcription. However, there are notable differences between our σ fragments compared to natural σ factors. First, core *E. coli* RNAP binds to DNA in a non-specific manner and this is titrated away by the σ factors [220, 221]. It is unlikely that our T7 RNAP core fragment binds to DNA. Second, a prokaryotic σ factor only recruits the RNAP to the promoter and once transcription initiation is complete, the σ factor dissociates during transcription [222, 223]. Thus, the ratio of σ factors to core RNAP is low (50%) because they only have to compete to bind to free (non-transcribing) polymerase [58]. Our system requires larger ratios, because the σ fragments must remain associated with the core fragment during transcription. Third, while the size of a σ factor and the σ fragment are about the same, their 3-dimensional structure and mechanism of binding to core and DNA are different [197]. Finally, recent results suggest that the B. subtilis core RNAP is shared by σ factors in time as opposed to concentration [224]. In other words, the σ factors pulse in a mutually exclusive manner to take turns fully utilizing the pool of core RNAP. In contrast, our σ fragments compete for the core fragment following mass action kinetics. This is similar to the previous understanding, where differences in σ factor binding affinities are a means that cells prioritize and order different responses [220, 225, 226].

Resource allocation also occurs in natural regulatory networks. In bacteria, alterna-

tive σ factors can redirect RNAP to different condition-specific promoters. Factors such as ppGpp and 6S RNA also regulate the pool of active free RNAP [62, 227, 228]. Using up this resource has been observed and shown to result in a slower growth rate [229]. Further, the competition between σ factors for core RNAP has been quantified [64, 220]. Keren and co-workers measured the activity of thousands of native *E. coli* and *S. cerevisiae* promoters under different environmental conditions [63]. They found that while changes in conditions have a global impact on many promoters, they shift by a linear factor that is characteristic of each condition. This factor ranges from 0.51 to 1.68 with M9 + glucose being the reference condition. They found that a simple model that treats overall promoter activity as a fixed resource explains their data. Overall promoter activity is equivalent to the total active RNAP concentration that forms the backbone of our resource allocator and the ratio of 0.36 shown in Figure 2-4D is analogous to their linear factor when moving from the high to the low resource allocator.

In the context of synthetic signaling networks, retroactivity occurs when downstream regulation impacts an upstream process. For example, the titration of ribosomes or proteases by one branch of the network can influence the network as a whole [230]. This is viewed as an undesirable effect that must be buffered against in order to maintain computational integrity [80]. In contrast, the resource allocator harnesses retroactivity in order to budget transcription to different pathways without surpassing a limit. As an allocation mechanism, retroactivity is an ideal means of distributing a budgeted resource. Currently, this is limited to dividing the core fragment among the σ fragments in a way that is proportional to their expression levels. Building on this, more complex dynamics could be introduced that implement signal processing between the output of the controller plasmid and the actuators that are being regulated. For instance, it may be desirable to control several actuators via a mutually exclusive or analog relationship, for example to slow down a metabolic pathway as a molecular machine is being built. Other actuators may require graded or ultrasensitive responses, for example the all-or-none commitment to flagellum construction versus simply changing the level of an enzyme. The toolbox presented in this paper provides a means to rationally design such control that can be implemented on the signal from the output of circuitry encoded on a controller to the actuators. [230]

2.5 Technical Information

2.5.1 Splitposon method for bisection mapping proteins

Design of the splitposon The splitposon is based on a commercial mini-Mu transposon, the HyperMu <KAN-1> transposon (previously available from Epicentre Biotechnologies). Mini-Mu transposons are a commonly used tool in molecular biology, due to their small size and easy in vitro transposition protocol [231]. In vitro transposition requires only the addition of a single transposase protein, MuA, along with a linearized mini-Mu transposon. The MuA protein binds specific sequences at the termini of the transposon (recognition ends) and catalyzes an efficient, mostly sequence-independent transposition event [232, 233]. In contrast to the native transposon, which contains 6 unique sequences in the recognition ends (L1–L2–L3 at one terminus, R1–R2–R3 at the other), mini-Mu transposons have shorter, palindromic ends consisting of two of the native sequences (R1–R2) [231].

While the R1–R2 sequence is required for transposition of a mini-Mu transposon, the sequence does not have to be perfect. The promiscuity of MuA has been studied by mutating the ends of the transposon , and a number of functional transposons with altered ends have been made. To construct the splitposon, we pooled the information from these studies to identify where the transposon could be altered and retain function. We focused on the R1 recognition sequence, since it is closest to the ends of the mini-Mu transposon, and our intention was to split proteins with as little added sequence as possible.

First, we used a consensus alignment of the six recognition sequences from the natural transposon [234] to determine where mutations are generally tolerated. However, it is unclear whether all of these alterations are tolerated specifically in the terminal recognition sites. Next, the R1 sequence was aligned with the L1 sequence, which is at the opposite terminus of the natural transposon. We referenced a mutational study [235] to determine tolerated changes to the two bases at the end of the transposon when it is used for in vitro transposition reactions. Finally, we collated the mutations in previously built transposon variants. Variants with a NotI cut site insertion and a triple stop codon insertion [236, 237] have been included in commercially available kits (F-701 and F-703 from Thermo Scientific), and have high activity. In addition, transposons with two unique mLyI cut site insertions

and two unique AarI cut site insertions are specified in publications [238, 239].

A start codon was introduced into the -4 through -2 positions in the transposon. The RBS calculator (thermodynamic model v1.0) [145] was used to evaluate a number of potential transposon ends for strong RBS activity. One variant proved to retain sufficient transposition efficiency and effectively initiate translation at the start codon. A P_{Tac} IPTG inducible promoter system from pEXT20 [240] mutated to have a symmetric LacO site (aattgtgagcgctcacaatt) was added to the splitposon to drive expression of the C-terminal protein fragment. The constitutive LacI cassette was included so that the promoter would not drive high levels of expression when in a plasmid lacking LacI expression.

The natural mini-Mu transposon contains a stop codon in-frame with the newly engineered start codon. However, out of frame insertions can lead to many additional amino acids added to the N-terminal fragment of the split protein, potentially complicating the analysis of bisection libraries. For this reason, we mutated the terminus of the splitposon opposite from the start codon to contain three staggered stop codons (one stop codon in each frame). This modification had already been successfully made in a mini- Mu transposon end to create a transposon for generating libraries of truncated proteins [237].

Library generation and characterization The splitposon can be used to split a protein of interest with two standard cloning steps (Figure 2-5A). First, MuA is used to transpose the splitposon into a target insertion plasmid (Figure 2-7B), which contains the region of the gene of interest to be targeted for bisection. This library is selected for the Kanamycin resistance gene in the transposon in addition to the resistance gene on the insertion plasmid. A sufficient number of colonies to achieve good coverage are plated, scraped, and harvested to yield an insertion library. Second, the pooled insertion library is digested using Type IIs restriction sites flanking the region of interest. The digested library is run on a gel, and the band with size corresponding to the region of interest plus a single splitposon is excised and purified. Finally, the size-selected fragments are ligated into an expression plasmid (Figure 2-7C) that has also been digested with Type IIs restriction enzymes to produce compatible overhangs. This plasmid contains an inducible expression system, as well as any flanking portions of the gene that were not in the region of interest.

The single transposition yields 6 different outcomes, depending on the orientation and position of the splitposon in the protein that is being split (Figure 2-6). The splitposon can insert in either the forward or reverse direction. If it is in the reverse direction, only the N-terminal fragment of the protein is expressed, and this fragment has a number of additional bases fused to it depending on the exact insertion location. Reverse transpositions therefore, are only seen if the protein of interest can be truncated and retain function. If transposition is targeted to a region of the protein that is not sufficient for function (i.e., by choosing a small enough region for the insertion plasmid), reverse insertions should have no function and will not be seen in a final selected library.

When the transposon is inserted in the forward direction, the frame of insertion determines what protein fragments will be made. MuA transposition duplicates 5 bp, leading to a few added amino acids on the protein fragments and complicating analysis. If the transposon inserts in frame with respect to the protein fragment at the 5 end of the transposon (frame 0), then a split protein will be expressed as desired. The N-terminal fragment contains no added amino acids, and the C-terminal fragment contains 3 added amino acids: M (for the start codon), a variable residue (coded for by A12, where 1 and 2 are the first two duplicated bp), and a duplicated residue (coded for by 345, the last three duplicated bases). If the transposon is inserted in frame +1 or +2, the C-terminal protein fragment is likely not to be expressed, leading to truncations that should not appear in a selected or screened library. Occasionally, the transposon may insert in frame +1 or +2 very close to an in-frame start codon, or it may create a start codon with the terminal A. In this case, out-of frame split proteins can be expressed, where the N-terminal fragment contains 2–3 variable/added residues (before the latter stop codons are encountered), and the C-terminal fragment contains duplications, insertions, or deletions based on the location of the start codon.

2.5.2 Bisection mapping of T7 RNA polymerase

Library design and statistics To avoid seeing any truncations in the library of bisected T7* RNAP, we chose to target transpositions to a subset of the gene. Previous studies on T7 RNAP have identified the C-terminus of the gene as having a key role in catalysis

and function. A version of the polymerase lacking the last two residues has been shown to lack productive polymerase activity [202]. We excluded the last 7 residues of the gene from our library to ensure that functional truncations would not be generated. In contrast, the Nterminal region of the gene appears less sensitive to alterations. A pilot library indicated that truncations of up to 30–35 residues were tolerated, so we conservatively excluded the first 40 residues from our bisection library. Hence, the insertion plasmid contains only residues 41-876 of T7* RNAP (Figure 2-7B). This section of the gene is flanked by BsaI Type IIs restriction sites for subcloning. We chose a ColEI backbone with Ampicillin resistance for the insertion plasmid. For the expression plasmid, the flanking portions of the polymerase (AAs 1–40 and 877–883) were placed downstream of the same P_{Tac} expression system that is in the splitposon (Figure 2-7C). BsaI restriction sited are located between these fragments to allow seamLess subcloning of the T7 RNAP* 41- 876 fragment from the insertion plasmid. Based on the size of the insertion plasmid and T7 RNAP* fragment it contains we calculated the library sizes of the insertion and final libraries. Based on the number of colonies harvested for each library, sufficient coverage was achieved at each library step to achieve a high probability of sampling all possible variants (Table 3.1) [241].

Library characterization After the final split T7* RNAP library was built and harvested, it was transformed into cells containing the plasmid Nif_489 [88]. This plasmid contains a P_{T7} driven RFP gene. Colonies were plated on selective media and 384 visually red colonies were picked (P_{Tac} is leaky enough on plates that colonies were visibly red without IPTG induction). These colonies were assayed for fluorescence in liquid media and the most active 192 selected for sequencing and further analysis. Each of the 192 selected clones was assayed four times and the mean promoter activity calculated.

The 192 active clones were each sequenced to determine the splitposon insertion location. In 180/192 clones this sequencing read gave enough information to unambiguously determine the insertion site of the splitposon. The other 12/192 clones were double splitposon insertions, other failure modes of the library, or sequencing errors, and were discarded. Of the 180 sequenced clones, 56 unique split sites were identified, with 36 in-frame and 20 out-of frame. The vast majority of the out-of-frame splits inserted in a location predicted to have

a close downstream in-frame start codon, leading to a split protein. However, due to high predicted variability in the RBS strength for out-of-frame splitposon insertions, we focused on the in-frame splits for all further analysis.

2.5.3 Supporting experiments

Directed evolution of the K1F and N4 σ fragments Error-prone PCR was applied to increase the activity of σ fragments based on the K1F and N4 RNAP variants [88]. After a visual screen for fluorescence, a number of clones with increased activity were identified for each σ fragment (Tables 3.2 and 3.3). Nearly the full K1F σ fragment (residues 610–871 in the full-length polymerase) was mutated and screened for function. 13 highly active clones from this library were assayed and sequenced, revealing that 100% contained a point mutation affecting the residue corresponding to 750 in the full polymerase sequence. Based on these results, a variant of the K1F σ fragment was created with the M750R mutation (K1FR), which exhibits activity within 4-fold that of the T7 σ fragment and was used in all remaining experiments. The error-prone PCR protocol was applied to a smaller region of the N4 σ fragment (residues 716–789 in the fulllength polymerase), and 12 improved clones were sequenced, but no sufficiently active mutations were found.

Means and error underlying the σ fragment orthogonally matrix The data used to generate the orthogonality heatmap in Figure 2-3E are shown with error bars (Figure 2-10). The promoter activity was measured for each σ fragment with each promoter, and each promoter in the absence of a σ fragment. Dividing the level of activity with each σ fragment by the level of activity without a σ fragment yields the fold induction.

Identifying the null fragment To determine the optimal null fragment, three known inactivating mutations [201, 202] were tested in the background of three σ fragments. The intention was to find a mutation that abolishes polymerase function without inhibiting the ability of the null fragment to compete with other σ fragments to bind the core fragment. A deletion of residues 882-3 and two point mutations, Y639A and H811A, were tested. These mutations were made to the T7 σ fragment, the CGG σ fragment, and a σ fragment based

on WT T7 RNAP (rather than T7* RNAP as for the T7 σ fragment). The T7 σ fragment was expressed constitutively with the core fragment and a P_{T7} reporter plasmid, and the variant null fragments were induced with IPTG. By comparing the P_{T7} promoter activity with and without induction of the null fragments, a fold repression value was calculated for each variant (Table 2.4). Based on this data, The CGG σ fragment with mutation Y639A was found to be the most active and was chosen as the null fragment. To determine whether the null fragment retains residual activity, it was expressed with the core fragment and a P_{CGG} reporter. Even at high levels of induction, this null fragment shows no P_{CGG} promoter activity when expressed with the core fragment (Figure 2-11).

Activation of the β core fragment with proteins fused to the α fragment We tested fusions of the α fragment to GFP for their ability to complement the β core fragment. Similar to the RFP- fusion assays in Figure 2-5G-H, the GFP- fusions were induced from P_{Tac} in the presence of constitutively expressed β core fragment and σ^{T7}. A reporter plasmid that produces RFP from a P_{T7} promoter was used to measure polymerase activity (Figure 2-12).

Measurement of P_{Tac} activity In order to estimate the amount of RNAP fragments produced by our inducible plasmids, we measured GFP production from similar P_{Tac} expression plasmids (Figure 2-14). The RiboJ insulator removes promoter context issues, leading to linear relationships between the expression levels of two proteins driven by identical promoters [36]. Hence, the measured values for GFP production should linearly correlate with the RNAP fragments produced in each system. P_{Tac} measurements were taken and plotted on the x-axis for the four the assays presented in Figures 2-3B, 2-4B,C,E, 2-5B, and 2-5E (Figure 2-13). In each case, the P_{Tac} measurement was taken concurrently with the other measurements, from cells growing in the same conditions.

Growth impact of split polymerase expression A number of the systems used to test split polymerase activity were measured to determine their impact on cell growth.T7 RNA polymerase is known to be toxic, especially when expressed in the presence of its promoter. Additionally, split proteins can be unstable and misfold, leading to further growth

impacts. We tested three systems of split polymerase expression for growth impacts: the four orthogonal fragment and core fragment expression systems shown in Figure 2-3A (Figure 2-15), the σ fragment competition systems shown in Figure 2-4A (Figure 2-16), and the multiply-split polymerase expression systems used in Figures 2-2D-E (Figure 2-17). The full length T7* RNAP was also tested when expressed from the same system as is used for the multiply-split polymerases (Figures 2-17 and 2-18). Each of these expression systems was induced with varying levels of IPTG and compared to a negative control containing the appropriate plasmid backbones, but not expressing the polymerase fragments or fluorescent proteins.

2.6 Methods

2.6.1 Strains and media

Escherichia coli DH10B was used for all routine cloning and characterization. ElectroMAX competent cells (Life Technologies) were used for library cloning steps as noted. LB-Miller media was used for assays and strain propagation, 2YT media was used for strain propagation, and SOC media was used for transformation recovery. Antibiotics were used as necessary for plasmid maintenance, with ampicillin at 100 g/mL, spectinomycin at 100 g/mL, kanamycin at 50 g/mL, and chloramphenicol at 17 g/mL. IPTG (isopropyl β-D-1-thiogalactopyranoside) was used as an inducer at concentrations up to 1 mM.

2.6.2 Plasmids and parts

Plasmids with the ColE1 origin were based off of the plasmid pSB1C3 from the Registry of Standard Biological Parts, which has a pUC19 [242] derived origin. Plasmids with the pUC origin were based off of a pUC19 [242] vector. Plasmids with the p15A* origin were based off of plasmid pSB3C5 [243] from the Registry. This origin appears to maintain at a higher copy number than standard for p15A. Plasmids with the pSC101 origin were based on pUA66 [244]. Plasmids with the BAC origin were based on pBACr-Mgr940 [245] (BBa_J61039), which has an F plasmid derived origin. A P_{Tac} promoter system derived from pEXT20 [240]

modified to contain a symmetric LacI binding site or a shortened version of this expression system was used in all systems that required inducible expression. Constitutive protein expression was driven by promoter P_{J23119} (BBa_J23105) or P_{J23119} (BBa_J23109), by a modified P_{Tet} expression system [70] (uninduced), and by promoters selected from libraries derived from P_{J23119} (BBa_J23119) through degenerate PCR. RBSs were either generated using the RBS calculator, taken from the Registry (BBa_B0032 and BBa_B0034 [61]), or selected from libraries generated using degenerate PCR. The RiboJ insulator [36] was used between P_{Tac} or P_{Tet} and the RBS in all constructs when titrations curves were run. mRFP1 [246] and sfGFP [247] were used as fluorescent reporters. Representative plasmid maps are shown in Figures 2-7, 2-14, and 2-18 through 2-24. A list of new plasmids is given in Table 2.6. Select constructs from this study will be made available online through Addgene (http://www.addgene.org/Christopher_Voigt/).

2.6.3 Bisection mapping T7 RNA polymerase

The splitposon was generated by modifying the HyperMu <KAN-1> transposon (Epicentre Biotechnologies). Examining previously described variants of the MuA transposon system [234, 236–239], a number of terminal bases were identified that could be altered while maintaining transposition activity. The RBS calculator [146] was used to design a strong terminal RBS and start codon while staying within these alterations. This modified end was combined with a previously built end containing terminal stop codons [237]. A P_{Tac} promoter and constitutive LacI expression cassette were inserted into the transposon to drive transcription at the end with the RBS and start codon. Finally, point mutations were made to remove restriction sites that would interfere with downstream cloning steps. A region of the T7* RNA polymerase CDS encoding aa 41–876 was flanked by BsaI sites in a ColE1 AmpR backbone. The splitposon (KanR) was transposed into this plasmid with MuA transposase (300 ng target DNA, 200 ng transposon, MuA buffer, 1.1 U HyperMuA transposase (Epicentre Biotechnologies), 30°C 8 h, 75°C 10 min), DNA clean and concentrated (Zymo), electroporated into ElectroMAX cells and plated on LB + Kan/Amp plates to obtain > 700,000 colonies. The colonies were scraped from the plates, pooled, and miniprepped to obtain DNA of the transposon insertion library. The transposon insertion library was digested

with BsaI, run on an agarose gel, and a band of ~5.7 kb (representing the section of the T7 CDS plus transposon) was excised, gel-purified (Zymo), and DNA clean and concentrated. A plasmid containing an inducible P_{Tac} system and the remainder of the T7 CDS (aa 1–40 and 877–883) with internal BsaI sites on a p15A* SpecR backbone was digested with BsaI and the size-selected fragment ligated into it. This reaction was DNA clean and concentrated, electroporated into ElectroMAX cells plated on LB + Spec/Kan plates to obtain > 600,000 colonies, and the colonies were scraped, pooled, and miniprepped as before to obtain the bisected library. This library was electroporated into *E. coli* DH10B cells with a plasmid containing a P_{T7}-RFP cassette on a pSC101 CamR backbone (Nif_489 [69]), plated on LB + Spec/Kan/Cam, and visually red colonies were picked after 16 h of growth for analysis in liquid media. More information on the splitposon method and T7 RNAP bisection mapping are included in 2.5.1 and 2.5.2.

2.6.4 Assay protocol

All promoter activity assays except the initial assay of T7 bisection mapping were performed as follows. Cells containing the plasmids of interest were inoculated from glycerol stocks into 0.5 mL LB-Miller media plus antibiotics in a 2-mL 96-deepwell plate (USA Scientific) sealed with an AeraSeal film (Excel Scientific) and grown at 37°C, 900 rpm overnight (~14–16 h) in a deepwell shaker. These overnights were diluted 200-fold into 150 μL LB-M with antibiotics plus varying concentrations of IPTG in 300-μL 96-well V-bottom plates (Thermo Scientific Nunc) sealed with an AeraSeal film and grown at 37°C, 1,000 rpm for 6 h. 5 μL of each sample was removed and diluted in 195 μL PBS + 2 mg/mL kanamycin to halt protein production. Cells diluted in PBS were either characterized immediately with flow cytometry or stored at 4°C until characterization. The initial T7 bisection mapping assays were performed similarly except the overnight cultures were grown in 2YT, and the overnight cultures were diluted 1:10 into 150 μL induction media.

2.6.5 Flow cytometry characterization

All fluorescence characterization was performed on a BD LSR Fortessa flow cytometer with HTS attachment and analyzed using FlowJo vX (TreeStar). Cells diluted in PBS + kanamycin were run at a rate of 0.5 μL/s until up to 100,000 events were captured (at least 50,000 events were recorded in all cases). The events were gated by forward scatter and side scatter to reduce false events and by time to reduce carry-over events. Gating was determined by eye and was kept constant for all analysis within each triplicate experiment. For all assays except the initial characterization of T7 bisection mapping, the geometric mean value of fluorescence was calculated for each sample, using a biexponential transform with a width basis of 10.0 to allow calculations with negative values. Finally, white-cell fluorescence measured concurrently from cells lacking fluorescent protein was subtracted from measured fluorescence to yield the Promoter activity (AU) values presented in the figures. The initial T7 bisection mapping assay was characterized identically, except that white-cell values were not subtracted.

Where fold induction calculations were required, fluorescence measurements were made of cells containing the appropriate reporter construct and lacking a functional polymerase, grown in the same conditions as the test cells. The fold induction is reported as the ratio of the white-cell-corrected test cell fluorescence to the white-cell-corrected fluorescence of the reporter-only cells.

To obtain relative expression levels for the polymerase fragments driven by P_{Tac}, constructs were made that express GFP after P_{Tac} and RiboJ (Figures 2-14). For each assay, cells with this construct were induced under the same conditions as the test cells, and their fluorescence measured (Figures 2-13). The P_{Tac} activity value in each plot represents the geometric mean white-cell-corrected fluorescence of these cells for that assay, and the horizontal error bars show the standard deviation of those measurements.

2.6.6 Measuring the growth impact of split polymerase expression

Cells containing the plasmids of interest were inoculated from colonies on agar plates into 0.5 mL LB-Miller media plus antibiotics in a 2-mL 96-deepwell plate, sealed with an AeraSeal

film, and grown at 37°C, 900 rpm overnight (~14–16 h) in a deepwell shaker. These overnights were diluted 200-fold into 150 μL LB-M with antibiotics plus varying concentrations of IPTG in 300-μL 96-well V-bottom plates, sealed with an AeraSeal film, and grown at 37°C, 1,000 rpm for 6 h. 20 μL of each sample were added to 80 μL LB in a 96-well optical plate (Thermo Scientific Nunc), and the OD600 of each diluted sample was measured using a BioTek Synergy H1 plate reader. These measurements were normalized by dividing by the OD600 of samples containing plasmids with the same backbones but expressing none of the proteins of interest (polymerase fragments or GFP) at each level of IPTG induction. Growth data are shown in Figures 2-15, 2-16 and 2-17.

2.6.7 Error-prone PCR of σ fragment variants

Sections of the K1F and N4 T7 RNAP variants [88] were amplified using GoTaq (Promega) in 1x GoTaq buffer plus $MgCl_2$ to a final concentration of 6.5 mM Mg^{2+}. The amplified fragments were cloned into a σ fragment expression plasmid including any necessary flanking RNAP sequence and the N-terminal SynZIP 18 domain. These mutated σ fragments were expressed with the core fragment and the appropriate promoter driving GFP. Colonies with visually improved GFP production were picked from plates, re-assayed to confirm activity, and sequenced to identify their mutations (Tables 3.2 and 3.3). Promising variants were reconstructed to isolate their effects and the resulting new σ fragments assayed for activity.

2.6.8 Tuning α fragment expression to compensate for copy number

An α fragment expression cassette consisting of the constitutive promoter P_{J23119}, RiboJ, and B0032 RBS driving the α fragment was inserted in the reverse direction before the P_{T7}: GFP cassette on a pSC101 reporter plasmid. These two cassettes were also inserted into a pUC19 backbone, with the weaker constitutive promoter P_{J23119} and start codon (GTG instead of ATG) in the α fragment cassette. Degenerate PCR was used to randomize the RBS in each plasmid at five nucleotides, and the resulting libraries were screened for fluorescence in the presence of the σ^{T7} and either core or β core fragments. Sets of pSC101 and pUC

plasmids were selected that had similar levels of activity with the β core fragment, but retained different levels of activity with the core fragment. These plasmids were isolated, sequenced, re-assayed, and the pair of pSC101 and pUC plasmids with the closest levels of expression in the presence of the β core fragment was selected.

2.7 Modeling

2.7.1 Kinetic model of the resource allocator

We used a kinetic model to examine the contrasting outcomes on total active RNAPs in a system with or without the resource allocator (Figures 2-1B-C). In each case, four promoters on the controller are modeled driving either four RNAPs or σ fragments. The promoters are switched between fully off and fully on states at different time points.

In the model with expression of full-length RNAPs (Figure 2-1B), only two reactions per RNAP were considered, yielding one equation per RNAP:

$$\dot{r}_i = u_i - \gamma r_i \qquad i = 1 - 4 \tag{2.1}$$

where the dot indicates a time derivative, and:

- $r_i = r_i(t) \geq 0$ is the concentration of the ith full-length RNAP,
- u_i is the lumped transcription and translation rate of the ith RNAP, and
- γ is the degradation rate (assumed equal) of the RNAPs.

For the model involving the resource allocator (Figure 2-1C), a number of additions were made. A core polymerase fragment is produced at a fixed rate equal to RNAP production in the previous model, while the four promoters now drive σ fragments of the polymerase. The σ fragments can bind the core fragment to form full-length RNAP complexes which can dissociate back into σ and core fragments. Again, all degradation rates are assumed to be

equal. This yields the following three equations:

$$\dot{\sigma}_i = u_i - \gamma\sigma_i + k_d r_i - k_a \sigma_i c \qquad i = 1 - 4 \tag{2.2}$$

$$\dot{r}_i = -\gamma r_i - k_d r_i + k_a \sigma_i c \qquad i = 1 - 4 \tag{2.3}$$

$$\dot{c} = v - \gamma c + k_d(\textstyle\sum_{i=1}^{k} r_i) - k_a(\sum_{i=1}^{k} \sigma_i c) \tag{2.4}$$

where dots indicate time derivatives, and:

- $\sigma_i = \sigma_i(t) \geq 0$ is the concentration of (unbound) σ fragment i,
- $r_i = r_i(t) \geq 0$ is the concentration of the ith full-length RNAP complex,
- $c = c(t) \geq 0$ is the concentration of core fragment,
- u_i, v are the lumped transcription and translation rates of the ith σ fragment and the core fragment, respectively,
- γ is the degradation rate (assumed equal) of the σ fragments, full-length RNAP complexes, and the core fragment,
- k_a is association rate of the σ fragments and the core fragment (assumed equal), and
- k_d is the dissociation rate of full-length RNAP's into σ fragments and the core fragment (again assumed equal)

We simulated time courses of RNAP concentration using these systems of equations and a set of estimated parameters (Table 2.5). The degradation rate, γ , was assumed to be dominated by dilution through cell growth and equal for all species in the system. The lumped transcription and translation rate of the full-length RNAPs was set to yield a steady-state concentration of 0.1 μM when they are expressed, the rate for the core fragment was set to be the same, and the rate for the σ fragments was set to yield 0.2 μM when expressed. Finally, the rates for the σ fragments binding and unbinding the core fragment were based on an *in vitro* analysis of a heterodimeric coiled-coil interaction [248]. Simulations were performed in MATLAB using the ode45 solver.

2.7.2 Uniqueness and stability of steady states in resource allocator model

We study the resource allocator model in (2-4), with the following changes:

- There are k different σ fragments and RNAPs rather than limiting to 4: $i = 1, \ldots, k$
- The lumped transcription and translation rates, u_i and v, are assumed constant and positive.

For simplicity, we also write the system in vector form as

$$\dot{x} = f(x) \qquad (S)$$

where

$$x(t) \;=\; (\sigma_1(t), \ldots, \sigma_k(t), r_1(t), \ldots, r_k(t), c(t))\,.$$

In general, a system of nonlinear ODE's (S) might have multiple stable states or persistent oscillations, or even exhibit chaotic behavior. It is thus of interest to show mathematically that our model has none of these, and, as a matter of fact, has the property that all solutions converge to a unique steady state, independently of initial concentrations. This is proved in the following result.

Theorem. There is a unique non-negative steady state of (S), which we will denote as

$$\bar{x} = (\bar{\sigma}_1, \ldots, \bar{\sigma}_k, \bar{r}_1, \ldots, \bar{r}_k, \bar{c})\,.$$

Moreover, every solution of (S) with $x(t) \geq 0$ satisfies $x(t) \to \bar{x}$ as $t \to \infty$.

Proof. It is convenient to introduce, for any given solution $x(t)$, the following combinations of variables:

$$s_i(t) := \sigma_i(t) + r_i(t)\,, \quad i = 1, \ldots, k \quad \text{(total } \sigma \text{ fragment } i\text{, bound and unbound)}$$

$$\sigma(t) := \sum_{i=1}^{k} \sigma_i(t)\,, \quad i = 1, \ldots, k \quad \text{(total unbound } \sigma \text{ fragments)}$$

$$s(t) := \sum_{i=1}^{k} s_i(t) \quad \text{(total } \sigma \text{ fragments, bound and unbound).}$$

$$r(t) := \sum_{i=1}^{k} r_i(t) \quad \text{(total RNAP complexes, without unbound core fragments).}$$

Observe that

$$\sigma(t) \;=\; s(t) - r(t)$$

for all t, or equivalently $s(t) = \sigma(t) + r(t)$. Since $\sigma(t) \geq 0$, it holds that

$$r(t) \leq s(t) \tag{2.5}$$

for all t. We also introduce

$$R(t) := c(t) + r(t) \quad \text{(total core fragments, bound and unbound).}$$

Since $c(t) \geq 0$, it holds that

$$r(t) \leq R(t) \tag{2.6}$$

for all t.

For each $i \in \{1, \ldots, k\}$, we have:

$$\dot{s}_i \;=\; \dot{\sigma}_i + \dot{r}_i \;=\; u_i - \gamma\sigma_i - \gamma r_i \;=\; u_i - \gamma s_i \,.$$

Therefore, along any solution,

$$\lim_{t\to\infty} s_i(t) \;=\; \bar{s}_i \;:=\; \frac{u_i}{\gamma} \tag{2.7}$$

and so also

$$\lim_{t\to\infty} s(t) \;=\; \bar{s} \;:=\; \frac{1}{\gamma}\sum_{i=1}^{k} u_i \,. \tag{2.8}$$

Similarly, for R we have:

$$\dot{R} \;=\; \dot{c} + \sum_{i=1}^{k} \dot{r}_i \;=\; v - \gamma c - \gamma(\textstyle\sum_{i=1}^{k} r_i) \;=\; v - \gamma R$$

and therefore, along any solution,

$$\lim_{t\to\infty} R(t) \;=\; \bar{R} \;:=\; \frac{v}{\gamma}\,. \tag{2.9}$$

Consider now an arbitrary steady state $\hat{x} = (\hat{\sigma}_1, \ldots, \hat{\sigma}_k, \hat{r}_1, \ldots, \hat{r}_k, \hat{c})$. Let $\hat{s}_i := \hat{\sigma}_i + \hat{r}_i$ $(i = 1, \ldots, k)$, $\hat{\sigma} := \sum_{i=1}^k \hat{\sigma}_i$, $\hat{s} := \sum_{i=1}^k \hat{s}_i$, $\hat{r} := \sum_{i=1}^k \hat{r}_i$, and $\hat{R} := \hat{c} + \hat{r}$. Because of the above remarks, it must hold that $\hat{s}_i = \bar{s}_i$ $(i = 1, \ldots, k)$, $\hat{s} = \bar{s}$, and $\hat{R} = \bar{R}$.

Along any trajectory, r satisfies the following differential equation:

$$\dot{r} \;=\; -(\gamma + k_d)r + k_a \sigma c \;=\; -(\gamma + k_d)r + k_a(s - r)(R - r)\,. \tag{2.10}$$

Note that this is a quadratic differential equation with time-dependent coefficients (since R and s are time-dependent functions). We study its stability behavior below, but first note that, at any steady state, since $R = \bar{R}$ and $s = \bar{s}$, the steady state value $\hat{r}$ must satisfy:

$$(\gamma + k_d)\hat{r} \;=\; k_a(\bar{s} - \hat{r})(\bar{R} - \hat{r})\,. \tag{2.11}$$

It is convenient to introducing the following constant, which can be thought of as an effective dissociation constant for RNAP complexes:

$$K \;=\; \frac{\gamma + k_d}{k_a}\,,$$

we can rewrite (2.11) as

$$K\hat{r} \;=\; (\bar{s} - \hat{r})(\bar{R} - \hat{r})\,. \tag{2.12}$$

As a function of $\hat{r}$, the left-hand side of (2.11) is a linear function with positive slope which vanishes at zero, and the right-hand side is a parabola opening up, with roots at $\bar{R}$ and $\bar{s}$. Thus, there is exactly one solution of (2.11), which we call $\bar{r}$, that is less than $\max\{\bar{R}, \bar{s}\}$, and, in fact, is less than $\min\{\bar{R}, \bar{s}\}$. An explicit formula for $\bar{r}$ (not required for the proof) is:

$$\bar{r} \;=\; \frac{1}{2k_a}(B - \sqrt{D}) \quad \text{where} \quad B = \gamma + k_d + k_a\bar{R} + k_a\bar{s}\,,\; D = B^2 - 4k_a^2\bar{R}\bar{s}\,.$$

By (2.5) and (2.6), $r(t) \leq s(t)$ and $r(t) \leq R(t)$ along all solutions (including constant solutions), so certainly $\hat{r} \leq \min\{\bar{R}, \bar{s}\}$, and thus $\hat{r} = \bar{r}$. Therefore

$$\hat{c} = \hat{R} - \hat{r} = \bar{c} := \bar{R} - \bar{r}\,. \tag{2.13}$$

Using that $\sigma_i = s_i - r_i$, we have, for each r_i:

$$\dot{r}_i \;=\; -(\gamma + k_d)r_i + k_a(s_i - r_i)c \;=\; k_a s_i c - (\gamma + k_d + k_a c)r_i\,. \tag{2.14}$$

So, at any steady state, since $s_i = \bar{s}_i$ and $c = \bar{c}$:

$$\hat{r}_i \;=\; \bar{r}_i \;:=\; \frac{k_a \bar{s}_i \bar{c}}{\gamma + k_d + k_a \bar{c}} \;=\; \frac{\bar{s}_i \bar{c}}{K + \bar{c}}\,, \tag{2.15}$$

which is formally analogous to a Michaelis-Menten product formation law. Notice that, as a consequence of (2.15), $\hat{r}_i/\hat{r}_j = \bar{s}_i/\bar{s}_j$ for any two i, j, and, in view of (2.7),

$$\frac{\hat{r}_i}{\hat{r}_j} \;=\; \frac{u_i}{u_j} \tag{2.16}$$

for all $i, j \in \{1, \ldots, k\}$, which means that the RNAP complexes are produced in the same proportion as the proportion between the respective inputs. It also follows that

$$\hat{\sigma}_i = \hat{s}_i - \hat{r}_i = \bar{\sigma}_i := \bar{s}_i - \bar{r}_i\,. \tag{2.17}$$

Defining $\bar{x}$ by the formulas in (2.17), (2.15), (2.13), we conclude that $\hat{x} = \bar{x}$, and the steady state is indeed unique.

Next, we show that $x(t) \to \bar{x}$ as $t \to \infty$, for every solution. If we assume that $s(t)$ and $R(t)$ are already at their steady states given by (2.8) and (2.9), the differential equation (2.10) becomes:

$$\dot{r} \;=\; -(\gamma + k_d)r + k_a(\bar{s} - r)(\bar{R} - r)\,. \tag{2.18}$$

(A justification for the assumption that R and s can be assumed to be at steady state will be given later.) The right-hand side of this ODE is the difference between the two sides

in (2.11), and thus is positive on $0 \leq r < \bar{r}$ and negative on $\bar{r} < r \leq \bar{R}$. Recall that we are only interested in solutions for which $r(t) \leq \bar{R}$. Therefore $r(t) \to \bar{r}$ as $t \to \infty$. Since $c(t) = R(t) - r(t)$, it follows that from the definition $\bar{c} = \bar{R} - \bar{r}$ that

$$\lim_{t\to\infty} c(t) \;=\; \bar{c}\,. \tag{2.19}$$

If we assume (justified later) that $s(t)$ and $c(t)$ are already at their steady states given by (2.8) and (2.19), the differential equation (2.14) becomes:

$$\dot{r}_i \;=\; k_a \bar{s}_i \bar{c} - (\gamma + k_d + k_a \bar{c}) r_i\,. \tag{2.20}$$

for each $i = 1, \ldots, k$. This is a stable linear constant-coefficient differential equation, so

$$\lim_{t\to\infty} r_i(t) \;=\; \bar{r}_i \tag{2.21}$$

for every i. Finally, from $\sigma_i(t) = s_i(t) - r_i(t)$, the definition $\bar{\sigma}_i = \bar{s}_i - \bar{r}_i$, together with (2.7) and (2.21), we have that

$$\lim_{t\to\infty} \sigma_i(t) \;=\; \bar{\sigma}_i \tag{2.22}$$

for every i. We have thus proved that $x(t) \to \bar{x}$ as $t \to \infty$.

Since (2.16) says that $\bar{r}_i/\bar{r}_j = u_i/u_j$ for all $i, j \in \{1, \ldots, k\}$, we have then that, for any arbitrary $j \in \{1, \ldots, k\}$:

$$\bar{r} \;=\; \sum_{i=1}^{k} \bar{r}_i \;=\; \sum_{i=1}^{k} \frac{u_i}{u_j}\bar{r}_j \;=\; \frac{\sum_{i=1}^{k} u_i}{u_j}\bar{r}_j$$

or equivalently:

$$\bar{r}_j \;=\; \bar{r}\left(\frac{u_j}{\sum_{i=1}^{k} u_i}\right) \tag{2.23}$$

which means that the relative expression of the jth RNAP complex is directly proportional to the fraction of its respective control input. For example, suppose that $k = 2$, and u_1 is maintained constant. Then the expression of the second RNAP complex at steady state has the hyperbolic Michaelis-Menten form $\bar{r}_2 \;=\; \frac{V u_2}{u_1 + u_2}$, where $V = \bar{r}$.

Justification of quasi-steady state assumption

It only remains to justify the hypotheses made at two points that variables already shown to approach steady state can be replaced by their steady state values in other equations (this is sometimes called the "CICS" or "convergent input to convergent state property"). One way to prove this is to appeal to the theory of asymptotically autonomous systems: we view (2.10) as a non-autonomous differential equation which, as $t \to \infty$, approaches the autonomous equation (2.18). Since this latter equation has $\bar{r}$ as a globally asymptotically stable state (for initial conditions in, for example, the interval $[0, \max\{\bar{R}, \bar{s}\}]$), it follows that solutions of (2.18) also approach $\bar{r}$. (See the last section in [249] for details of this technique and further references.) Similar considerations apply to the linear ODE (2.14) and its limit equation (2.20).

Simplifications when $K \ll 1$

For realistic degradation and association and dissociation constants, K is very small, typically $\approx 10^{-9}$M. In that case, the formulas for steady state values can be simplified considerably. We will assume that $v < \sum_{i=1}^{k} u_i$ (the core fragment is the limiting factor), in which case $\bar{R} = v/\gamma < (\sum_{i=1}^{k} u_i)/\gamma = \bar{s}$, and thus $\min\{\bar{R}, \bar{s}\} = \bar{R}$. When $K \approx 0$, the unique steady state value $\bar{r} \leq \bar{R}$ that solves $(\bar{s} - \bar{r})(\bar{R} - \bar{r}) = K\bar{r} \approx 0$ is $\bar{r} \approx \bar{R}$. This means that (2.23) is, more explictly:

$$\bar{r}_j \approx \bar{R} \left(\frac{u_j}{\sum_{i=1}^{k} u_i} \right) = \frac{v}{\gamma} \left(\frac{u_j}{\sum_{i=1}^{k} u_i} \right) \tag{2.24}$$

It is important to note, however, that informal approximation arguments are not mathematically rigorous, and can easily lead to paradoxical conclusions. For example, (2.13) implies that $\bar{c} = \bar{R} - \bar{r} \approx 0$ (since we had $\bar{R} \approx \bar{r}$), and this, combined with (2.15) gives that $\bar{r}_i = \frac{\bar{s}_i \bar{c}}{K + \bar{c}} \approx \frac{\bar{s}_i \times 0}{K+0} = 0$! (The fallacy in this case comes from the approximation "$x/(K+x) \approx 0$ when $x \approx 0$" which is false if $K \ll x$.)

To make the argument mathematically precise, let us think of the unique steady state value $\bar{r} \leq \bar{R}$ that solves $K\bar{r} = (\bar{s} - \bar{r})(\bar{R} - \bar{r})$ as a function $\bar{r}(K)$, and take its limit as

$K \to 0$ while keeping $\bar{R}$ and the $\bar{s}_i$'s fixed. Keeping these values fixed is valid for example if $k_a \to \infty$, or if $k_d \to 0$ and $\gamma \to 0$ at the same time that the control inputs (v and the u_i's) are proportionally increased. Using implicit differentiation, and primes to indicate derivative with respect to K, we have that $\bar{r} + K\bar{r}' = -\bar{r}'(\bar{R} - \bar{r}) - \bar{r}'(\bar{s} - \bar{r})$. Since $\bar{r} = \bar{R}$ when $K = 0$, the derivative at $K = 0$ is $\bar{r}' = \bar{R}/(\bar{R} - \bar{s})$ and thus we obtain the first-order Taylor expansion

$$\bar{r}(K) \;=\; \bar{r}(0) + \bar{r}'(0)K \,+\, o(K) \;=\; \bar{R} \,+\, \frac{\bar{R}}{\bar{R} - \bar{s}}K \,+\, o(K)\,.$$

Then, $\bar{c} = \bar{R} - \bar{r} = \frac{\bar{R}}{\bar{s} - \bar{R}}K + o(K)$, and now substituting into $\bar{r}_j = \frac{\bar{s}_j \bar{c}}{K + \bar{c}}$, we conclude that:

$$\bar{r}_j \;=\; \frac{\bar{s}_j \bar{c}}{K + \bar{c}} \;=\; \bar{s}_j \frac{\bar{R}}{\bar{s}} + O(K) \;=\; \frac{v}{\gamma}\left(\frac{u_j}{\sum_{i=1}^{k} u_i}\right) + O(K)\,,$$

which recovers (2.24) as $K \to 0$.

2.7.3 Modeling σ fragment competition data

Using the simplified steady-state equations presented in (2.24), we can model the σ fragment competition data shown in Figure 2-4. In the context of the experiments shown in Figure 2-4, there are only two σ fragments, T3 and K1FR, yielding the equations:

$$\bar{r}_{T3} \;\approx\; \bar{R}\left(\frac{u_{T3}}{u_{T3} + u_{K1F}}\right) \tag{2.25}$$

$$\bar{r}_{K1F} \;\approx\; \bar{R}\left(\frac{u_{K1F}}{u_{T3} + u_{K1F}}\right) \tag{2.26}$$

If the P_{T3} and P_{K1FR} promoter activities are linearly proportional to the concentration of the appropriate RNAP complex, these equations immediately predict the result shown in Figure 2-4D; changing the resource allocator results in an identical linear scaling of the promoter outputs. Changing the expression of the core fragment from the resource allocator changes the value of $\bar{R}$, which linearly scales $\bar{r}_{T3}$ and $\bar{r}_{K1FR}$ identically for any constant values of u_{T3} and u_{K1FR}.

In Figure 2-4E, we normalize the promoter activities of P_{T3} and P_{K1FR} by the maximum promoter activities obtained when the appropriate σ fragments are expressed to saturate

the core fragment. Assuming that the promoter activities are linearly proportional to the amount of corresponding RNAP present in the system, these normalized values represent the fraction of the core fragment bound by each σ fragment. That is $\bar{r}_{T3}/\bar{R}$, $\bar{r}_{K1F}/\bar{R}$, for the normalized activity values of P_{T3} and P_{K1FR}, respectively. Therefore, we have:

$$N_{T3} = \bar{r}_{T3}/\bar{R} \approx \left(\frac{u_{T3}}{u_{T3} + u_{K1F}}\right) \tag{2.27}$$

$$N_{K1F} = \bar{r}_{K1F}/\bar{R} \approx \left(\frac{u_{K1F}}{u_{T3} + u_{K1F}}\right) \tag{2.28}$$

where N_{T3} and N_{K1F} are the normalized P_{T3} and P_{K1FR} promoter activities shown in Figure 2-4E.

Finally, we have a relative measurement for the expression of the T3 σ fragment: the P_{Tac} expression level with the appropriate amount of inducer. Assuming that this value is linearly proportional to the true expression level of the T3 σ fragment, we can say: $u_{T3} = cP_{Tac}$, where c is a scaling factor to relate the P_{Tac} expression level to the σ_{T3} expression level. Substituting this into the model yields:

$$N_{T3} \approx \left(\frac{P_{Tac}}{P_{Tac} + \frac{u_{K1F}}{c}}\right) \tag{2.29}$$

$$N_{K1F} \approx \left(\frac{\frac{u_{K1F}}{c}}{P_{Tac} + \frac{u_{K1F}}{c}}\right) \tag{2.30}$$

As the N_{T3}, N_{K1F}, and P_{Tac} values are all measured, there is only one remaining free variable: $\frac{u_{K1F}}{c}$, which represents the constant expression level of the K1FR σ fragment in the same units as the P_{Tac} expression value. This parameter was determined by simultaneously fitting (29) and (30) to the N_{T3} and N_{K1F} data shown in Figure 2-4E, using a least-squares algorithm (lsqnonlin) in MATLAB. This yields a value of 617 for $\frac{u_{K1F}}{c}$. Hence, the final models shown in Figure 2-4E are:

$$N_{T3} \approx \left(\frac{P_{Tac}}{P_{Tac} + 617}\right) \tag{2.31}$$

$$N_{K1F} \approx \left(\frac{617}{P_{Tac} + 617}\right) \tag{2.32}$$

And the sum of those two equations:

$$N_{Sum} = N_{T3} + N_{K1F} \approx 1 \tag{2.33}$$

2.8 Figures

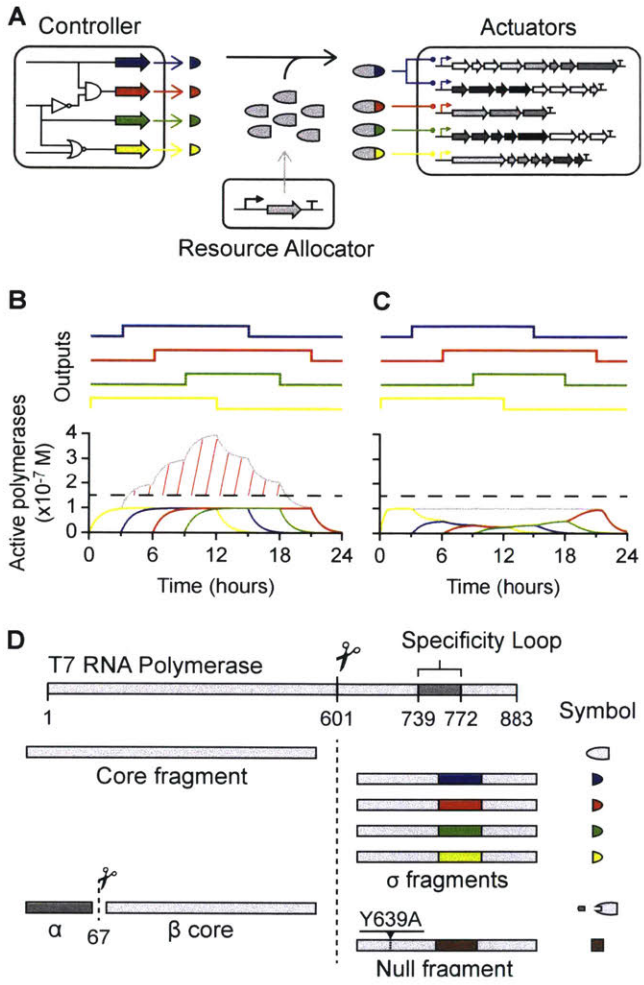


Figure 2-1: **The resource allocator.**
A. Complex synthetic genetic systems are broken down into three modules. The core fragment of RNAP is expressed from the resource allocator. Each output from the controller results in the expression of a different σ fragment (colored half-circles), which share the core fragment and turn on different actuators.
B. Dynamic simulations of resource allocation are shown, where the outputs from the controller are turned on and off at different times (colored lines) (Section 2.7.1). A hypothetical toxicity threshold is shown with the dashed horizontal line. When the outputs of the controller are complete RNAPs, their sum crosses the threshold (gray line and red hash).
C. With resource allocation, the outputs of the controller are σ fragments that must share the core fragment, thus ensuring that their sum transcriptional activity does not cross the threshold.
D. The complete toolbox of phage RNAP fragments is shown.

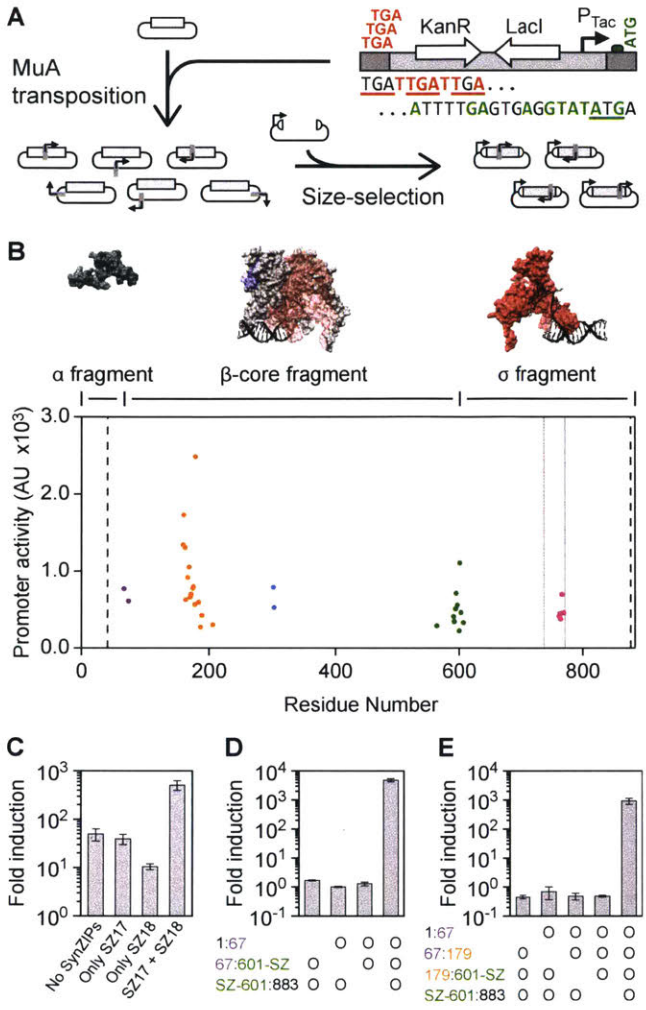


Figure 2-2: **Bisection mapping of T7* RNAP.**
A. The splitposon is based on a modified mini-Mu transposon mutated to contain staggered stop codons in one recognition end (red) and an RBS & start codon in the other (green). An internal inducible system (LacI and P_{Tac}) has been added. Bisection mapping includes two cloning steps. First, the splitposon is transposed randomLy into a gene using MuA transposase. Second, the library is size selected for inserts that contain one transposon insertion and cloned into an expression plasmid.
B. Each point represents a unique in-frame split location in T7* RNAP, where the residue number is the final residue in the N-terminal fragment. The promoter activity is the mean P_{T7} activity for all recovered clones at each split point, from four independent assays (10 μM IPTG induction). Bisection points are clustered into five seams, which are color-coded. The vertical dashed lines show the region where bisections were allowed in the library, and the gray vertical lines show the location of the promoter specificity loop. Surface models are shown for the three fragments used for the resource allocator (PDB:1QLN [195], visualized using UCSF Chimera [250]). The model for the β core fragment shows the position of the α and σ fragments in transparent blue and red, respectively. More views of the surface model are shown in Figure 2-9.
(Continued on next page)

Figure 2-2: (Continued)
C. The fragments created from splitting T7 RNAP at residue 601 were assayed with and without SynZIP domains at low expression levels (4 μM IPTG). When SynZIP 17 (SZ17) is fused to the N-terminal fragment and SynZIP 18 (SZ18) is fused to the C-terminal fragment, a large increase in the induction of P_{T7} is observed. Fold induction is calculated as the P_{T7} promoter activity in induced cells divided by the promoter activity of cells that contain the reporter plasmid but no polymerase fragments.
D. Data are shown for the expression of the three fragments corresponding to the α fragment (1:67), β core fragment (67:601-SZ), and σ fragment (SZ-601:883). An o indicates the presence of a fragment in an operon that is expressed with 100 μM IPTG.
E. Data are shown for the induction of four fragments, as in (D), with an additional split of the β core fragment at residue 179.
Data information: For the graphs in (CE), the mean is shown for three independent assays performed on different days, with error bars showing standard deviation.

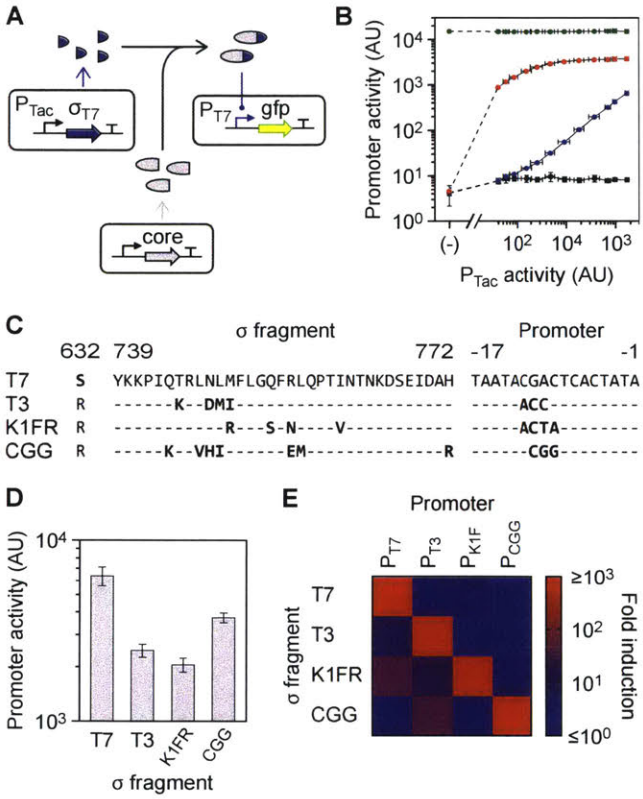


Figure 2-3: **Activation of the core fragment via σ fragments.**
A. A schematic of the induction system is shown; the core fragment is expressed at a constant level from a constitutive promoter.
B. The T7 σ fragment (SZ-601:883) is induced in the presence of different core fragments, and the activity of P_{T7} is measured. Red and blue points show the induction in the presence and absence of the SynZIP, respectively (core fragments 1:601-SZ and 1:601). The activity of full-length T7* RNAP is shown as a positive control (green). A negative control with no core fragment is shown (black). The leftmost point (marked '(−)') represents cells that did not encode the T7 σ fragment. From left to right, the remaining points represent induction levels of: 0, 1, 2, 4, 6.3, 10, 16, 25, 40, 63, 100, and 1,000 μM IPTG.
C. The variations between the σ fragments and promoters are shown. Position 632 indicates the mutation made in T7* RNAP that reduces toxicity, and positions 739–772 show the DNA-binding loop.
D. The activities of each of the four σ fragments are shown with their cognate promoters when expressed to saturation (100 μM IPTG) with the core fragment.
E. The cross-reactivity of each σ fragment with each promoter is shown (100 μM IPTG induction of the σ fragments and constant core fragment expression). The underlying activity levels and variation for this assay are shown in Figure 2-10.
Data information: For all graphs, the mean is shown for three independent assays performed on different days, with error bars showing standard deviation.

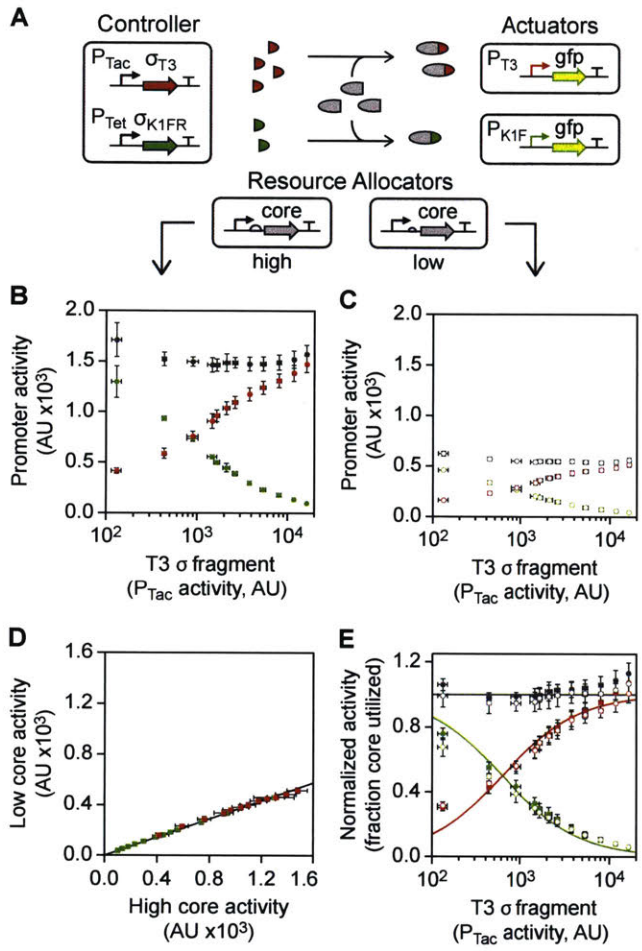


Figure 2-4: **Competition between σ fragments to bind the core fragment.**
A. The genetic system used for the competition assays is shown. Two resource allocator plasmids were built that generate high and low core fragment expression levels via a strong or weak RBS and constitutive promoter.
B. Data for the high resource allocator are shown. The K1FR σ fragment was expressed at a constant level (no induction of P_{Tet}), and the T3 σ fragment was induced with 0, 2, 4, 6.3, 7.4, 8.6, 10, 13, 16, 20, 25, and 32 μM IPTG. The activities of P_{T3} (red circles) and P_{K1F} (green circles) were measured, and the sum of their activities computed (gray circles).
C. Data for the low resource allocator are shown, as in (B).
D. Each point represents promoter activity (red: P_{T3}, green: P_{K1F}) at a specific level of inducer. The x and y values show the activity with high and low levels of core fragment expression, respectively. The line shows a linear regression, with the intercept fixed to 0.
(Continued on next page.)

Figure 2-4: (Continued)
E. Each σ fragment was expressed to saturation (100 μM IPTG) with the high and low resource allocators, and the measured promoter activities were used to normalize the data shown in (B) and (C) (solid and hollow circles, respectively). The fraction core utilized represents the proportion of the core fragment present in the system that is bound by either σ fragment, assuming a linear correlation with promoter activity. The solid lines show a simplified model of competition fit to the normalized data.
Data information: For all graphs, the mean is shown for three independent assays performed on different days, with error bars showing standard deviation.

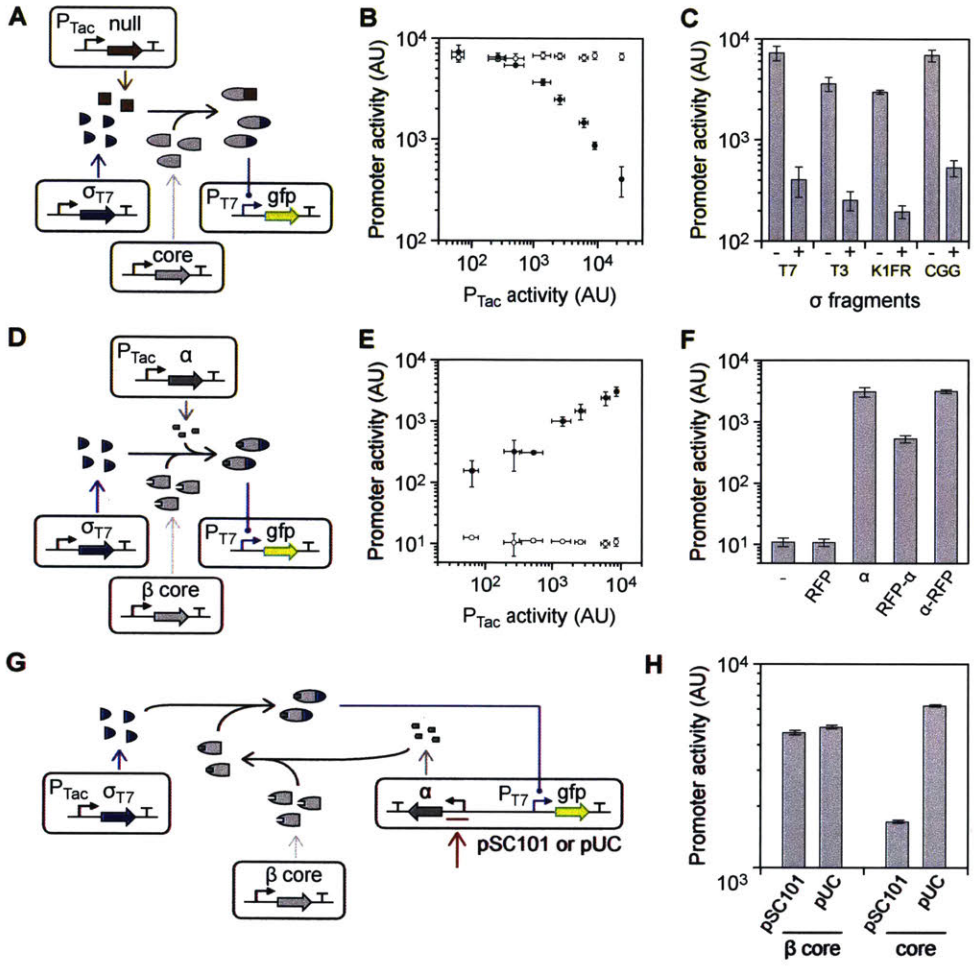


Figure 2-5: **Positive and negative post-transcriptional regulation of the core fragment.**

A. Null fragment sequestration of the core fragment.

B. The core fragment and T7 σ fragment are expressed constitutively, while null fragment expression is induced from P_{Tac} (induction from left to right is: 0, 2, 4, 10, 16, 25, 40, and 1000 μM IPTG). The effect of the expression of the null fragment on P_{T7} activity is shown as black circles. The activity of P_{T7} under the same conditions lacking the inducible null fragment cassette is shown as white circles.

C. The null fragment is shown in competition with each of the four σ fragments. Data are shown when the null fragment is uninduced (−, 0 μM IPTG) and induced (+, 1000 μM IPTG).

D. Activation of the β core fragment through the expression of the α fragment.

E. The impact of expressing the α fragment from the P_{Tac} promoter is shown. The black and white circles show induction in the presence and absence of the α fragment cassette, respectively (from left to right: 0, 2, 4, 10, 16, 25, and 40 μM IPTG). The high level for uninduced is due to leaky expression from P_{Tac}.

(Continued on next page.)

Figure 2-5: (Continued)
F. The ability of α fragment : RFP fusions to complement the β core fragment (with the T7 σ fragment) is shown. From left to right: $(-)$, no inducible cassette; RFP, expression of unmodified RFP; α , expression of free α fragment; RFP-α , expression of a C-terminal fusion of α fragment to RFP; α-RFP, expression of an N-terminal fusion. Each system was induced with 40 μM IPTG.
G. A genetic system is shown that uses α fragment expression from a constitutive promoter to compensate for the effects of differences in copy number. A strong constitutive promoter and RBS controlling α expression (red arrow) are selected at low copy (pSC101), while a weaker promoter and RBS are used at high copy (pUC).
H. Data are shown for a pair of pSC101 and pUC plasmids carrying tuned α fragment cassettes and a P_{T7} promoter driving GFP. β core indicates that the β core fragment and T7 σ fragment are co-expressed. core indicates that the core fragment and T7 σ fragment are co-expressed.
Data information: For all graphs, the mean is shown for three independent assays performed on different days, with error bars showing standard deviation.

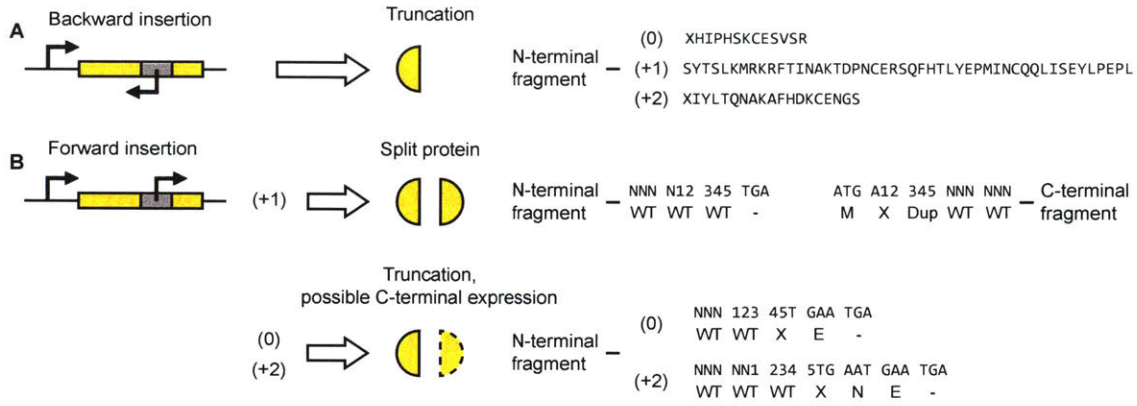


Figure 2-6: **Outcomes of a splitposon library.**
A. If the splitposon inserts in the reverse direction, only the N-terminal fragment of the protein is expressed. Additionally a number of amino acids are fused to this fragment depending on the frame of insertion (as judged by protein fragment at the 5 end of the transposon). X indicates a variable residue that depends on the sequence of the insertion site.
B. If the splitposon inserts in the forward direction, a split protein or truncation is expressed depending on the frame of insertion. If the splitposon inserts in-frame (0), a split protein is expressed with 3 AAs added to the C-terminal fragment. The DNA sequence and encoded AAs directly flanking the splitposon are shown. For DNA (top row), Ns indicate bases in the original coding sequence of the protein, 1-5 indicates the 5 bps of DNA duplicated during MuA transposition, and other letters indicate the sequence of the splitposon. For AAs (bottom row), WT indicates a residue in the split protein, X indicates a variable residue (i.e. one coded for by bps both from the splitposon and original protein coding sequence), Dup indicates a WT residue that is present in both the N and C-terminal fragments, and other letters represent the appropriate AAs. If the splitposon inserts in the (+1) or (+2) frames, the N-terminal fragment will be expressed with a few added AAs and the C-terminal fragment may be expressed by an in-frame start codon. The residues added to the N-terminal fragment are shown in the same manner as for the (0) frame.

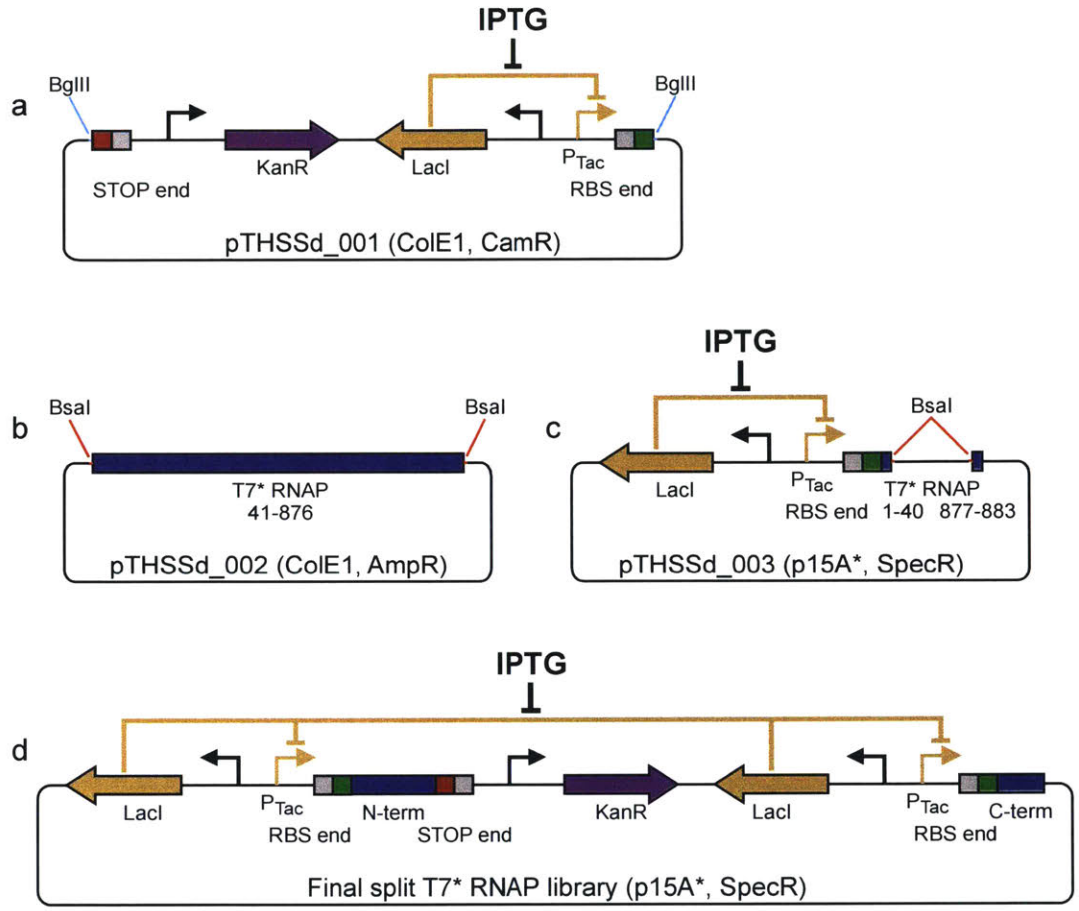


Figure 2-7: **Vectors used for bisection mapping of T7 RNA polymerase.**
A. The splitposon is carried in a high copy ColE1 plasmid with chloramphenicol resistance. It is excised with BglII and purified from an agarose gel to produce the cleaved linear transposon substrate for an in vitro transposition reaction.
B. The insertion plasmid carries the coding sequence for residues 41-876 of T7* RNAP flanked by BsaI sites on a high copy ColE1 backbone with ampicillin resistance.
C. The expression plasmid contains an inducible P_{Tac} expression system and the coding sequences for residues 1-40 and 877-883 of T7* RNAP. The P_{Tac} expression system and RBS are identical to those in the splitposon.
D. An example of a clone in the final bisection library. In this case, the splitposon is inserted in the forward direction into the T7* RNAP CDS. Plasmids pTHSSd_4-7, which were used to re-verify the 601 split and test the effect of adding SynZIPs (Figure 2-2E) look identical (plus the added SynZIPs). Both the expression plasmid and final library contain the p15A* origin and are spectinomycin resistant. Because of the splitposon, the final library is also kanamycin resistant.

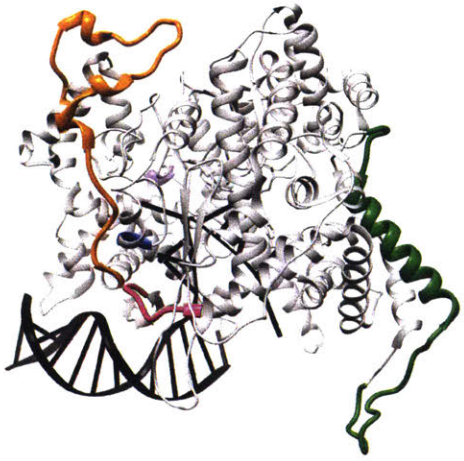


Figure 2-8: **Split seams mapped onto T7 RNAP structure.**
The five seams identified in Figure 2-2B are shown on the T7 RNAP elongation structure (PDB# = 1QLN) [195], visualized using UCSF Chimera [250]) using the same color scheme: Purple = 67–74, Orange = 160–206, Blue = 301–302, Green = 564–607, Pink = 763–770. DNA and the nascent RNA strand are shown in black.

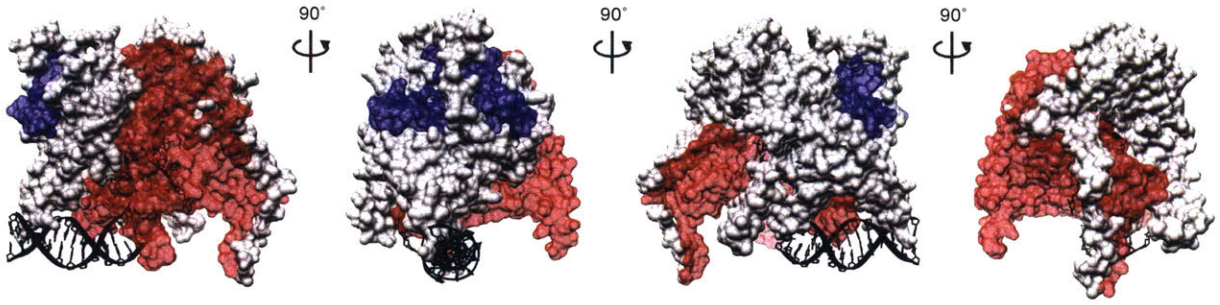


Figure 2-9: **Surface model of three-piece T7 RNAP.**
A surface model of the T7 RNAP elongation structure (PDB# = 1QLN, visualized using UCSF Chimera) is shown, with the α fragment colored blue, the β core fragment colored grey, and the σ fragment colored red. The leftmost view shows transcription from left to right, and each subsequent image is rotated 90° around the y axis. DNA and the nascent RNA strand are shown in black.

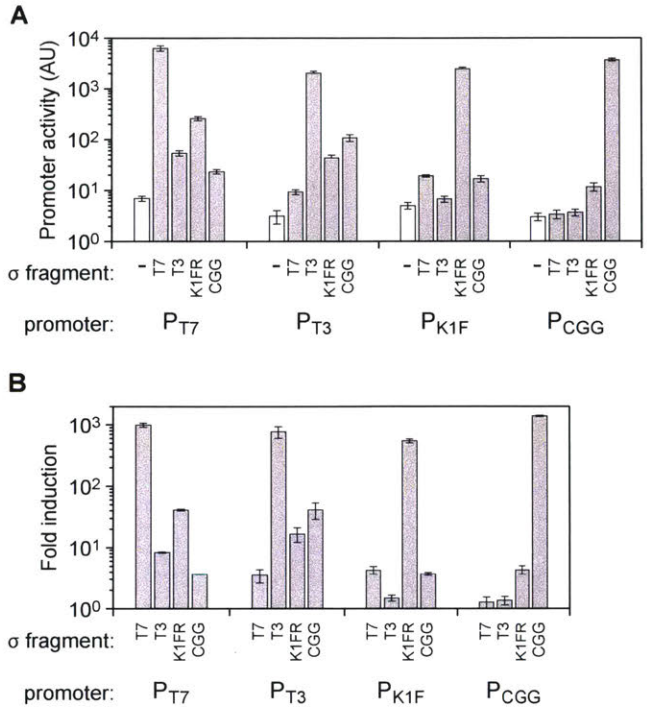


Figure 2-10: **Detailed σ fragment orthogonality results.**
A. Each of the σ fragments and a negative control were induced with 10 μM IPTG in the presence of the core fragment and each of the four promoters. Grey bars represent promoter activity with expressed σ fragments, white bars indicate the promoter activity of negative controls with no expressed σ fragment.
B. The fold induction of each σ fragment in combination with each promoter is shown. Each bar represents the mean value of three independent assays performed on different days, with error bars showing standard deviation.

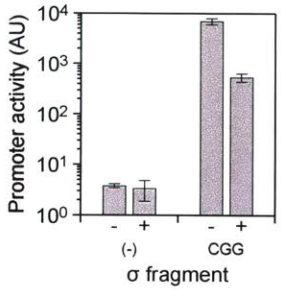


Figure 2-11: **The null fragment lacks σ fragment activity.**
The null fragment is induced from P_{Tac} with 0 (−) or 1000 μM (+) IPTG in the presence of the core fragment, a P_{CGG} reporter and either the CGG σ fragment (CGG) or no σ fragment (−). The mean promoter activity from three independent assays is shown, with error bars showing standard deviation.

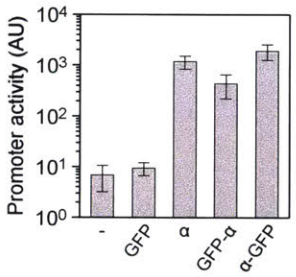


Figure 2-12: **Activity of GFP : α fragment fusions.**
The ability of α : GFP fusions to complement constitutively expressed β core fragment and σ^{T7} is shown by the activity of a P_{T7} promoter driving RFP. (-): no inducible cassette, GFP: expression of unmodified GFP, α : expression of unmodified α fragment, GFP-α : expression of an N-terminal fusion of GFP to the α fragment, α-GFP: expression of a C-terminal fusion of sfGFP1 to the α fragment. Each system was induced with 40 μM IPTG. Bars show the mean level of activity from three independent assays, and error bars show the standard deviation.

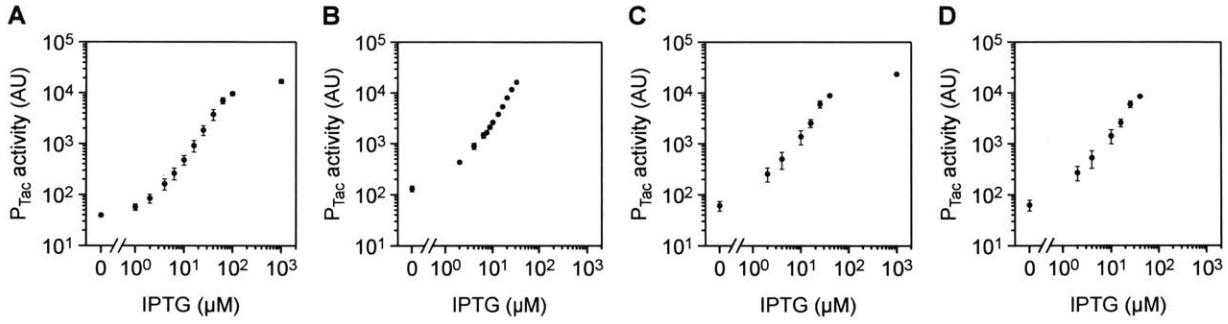


Figure 2-13: **P_{Tac} activity measurements.**
Measurements of GFP production by P_{Tac} were taken under different conditions to determine relative expression levels in a number of assays.
A. P_{Tac} measurements for the assay in Figure 2-3B with plasmid pTHSSd_34. From left to right, expression was induced with 0, 1, 2, 4, 6.3, 10, 16, 25, 40, 63, 100, and 1000 μM IPTG.
B. P_{Tac} measurements for the assay in Figures 2-4A-C with plasmid pTHSSd_51. From left to right, expression was induced with 0, 2, 4, 6.3, 7.4, 8.6, 10, 13, 16, 20, 25, and 32 μM IPTG.
C. P_{Tac} measurements for the assay in Figure 2-5B with plasmid pTHSSd_34. From left to right, expression was induced with 0, 2, 4, 10, 16, 25, 40, and 1000 μM IPTG.
D. P_{Tac} measurements for the assay in Figure 2-5E with plasmid pTHSSd_34. From left to right, expression was induced with 0, 2, 4, 10, 16, 25, and 40 μM IPTG. For all graphs, the mean is shown for three independent assays performed on different days, with error bars showing standard deviation.

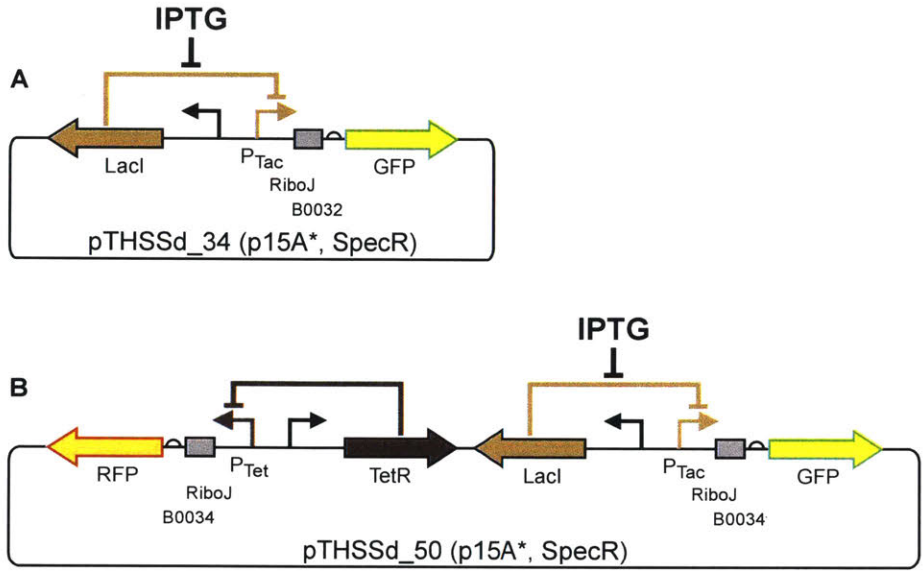


Figure 2-14: **Plasmids used for P_{Tac} activity measurements.**
A. pTHSSd_34 was used to characterize P_{Tac} expression in Figures 2-3B, 2-5B, and 2-5E. It expresses GFP under control of P_{Tac}, with RiboJ and a BBa_B0032 RBS.
B. pTHSSd_50 was used to characterize P_{Tac} expression in the σ fragment competition assay (Figure 2-4). It expresses GFP under the control of P_{Tac} with RiboJ and a BBa_B0034 RBS. Additionally, RFP is expressed under the control of P_{Tet}, with RiboJ and a BBa_B0034 RBS. Both plasmids have a p15A* origin with spectinomycin resistance.

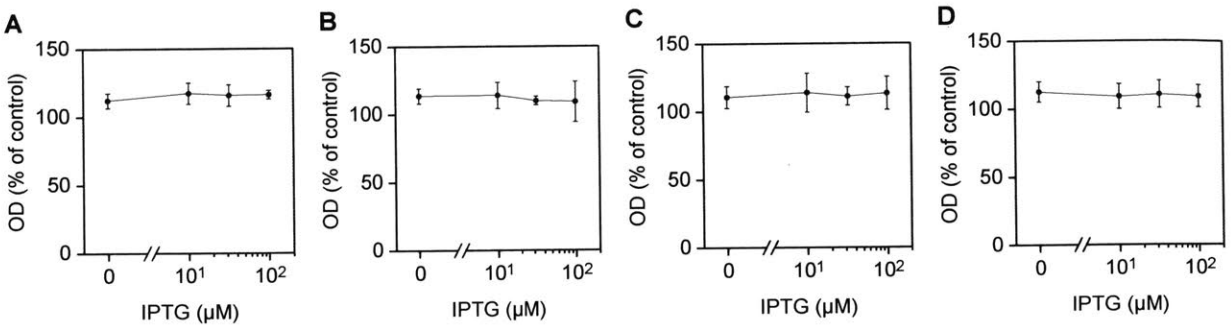


Figure 2-15: **Growth impact of split polymerase systems.**
The growth impact of the split polymerase expression systems from Figures 2-3B-C is shown. The four orthogonal σ fragments were expressed with IPTG (induction from left to right: 0, 10, 32, and 100 μM) in the presence of the core fragment (pTHSSd_38) with the appropriate reporter plasmid and the OD600 after 6 hours of growth compared to a control strain carrying plasmids that do not express the polymerase fragments or GFP (pTHSSd_36, pTHSSd_44, pTHSSd_13).
A. T7 σ fragment and reporter (pTHSSd_23 and pTHSSd_8).
B. T3 σ fragment and reporter (pTHSSd_24 and pTHSSd_9).
C. K1FR σ fragment and reporter (pTHSSd_25 and pTHSSd_10).
D. CGG σ fragment and reporter (pTHSSd_26 and pTHSSd_11). For all graphs, the mean is shown for three independent assays performed on different days, with error bars showing standard deviation.

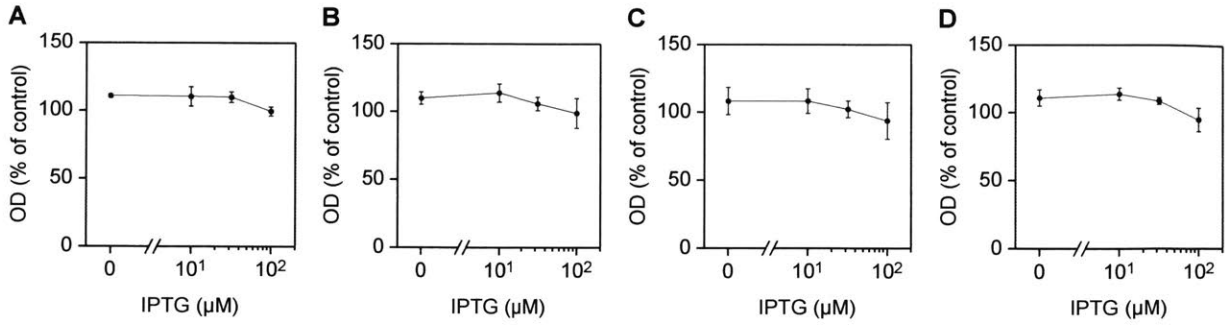


Figure 2-16: **Growth impact of σ fragment competition systems.**
The growth impact of the competition systems from Figure 2-4 is shown. The T3 σ fragment was expressed with IPTG (induction from left to right: 0, 10, 32, and 100 μM) in the presence of the K1FR σ fragment (pTHSSd_49) and high or low levels of the core fragment with either the T3 or K1FR reporter plasmid and the OD600 after 6 hours of growth compared to a control strain carrying plasmids that do not express the polymerase fragments or GFP (pTHSSd_36, pTHSSd_44, pTHSSd_13).
A. Higher level of core fragment expression with the T3 reporter (pTHSSd_38, pTHSSd_9).
B. Higher level of core fragment expression with the K1FR reporter (pTHSSd_38, pTHSSd_10).
C. Lower level of core fragment expression with the T3 reporter (pTHSSd_39, pTHSSd_9).
D. Lower level of core fragment expression with the K1FR reporter (pTHSSd_39, pTHSSd_10). For all graphs, the mean is shown for three independent assays performed on different days, with error bars showing standard deviation.

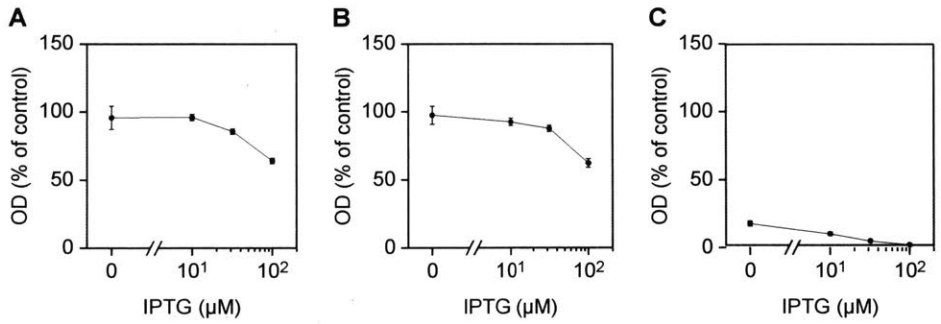


Figure 2-17: **Growth impact of highly expressed multiply split polymerase.**
The growth impact of the three and four fragment polymerases from Figures 2-2D-E is shown with a full-length polymerase control. The split or full T7* polymerases were expressed with IPTG (induction from left to right: 0, 10, 32, and 100 μM) in the presence of a T7 reporter plasmid (pTHSSd_8) and the OD600 after 6 hours of growth compared to a control strain carrying plasmids that do not express the polymerase fragments or GFP (pTHSSd_35, pTHSSd_13).
A. Three piece T7* RNA polymerase (pTHSSd_14).
B. Four piece T7* RNA polymerase (pTHSSd_18).
C. Full-length T7* RNA polymerase (pTHSSd_37). For all graphs, the mean is shown for three independent assays performed on different days, with error bars showing standard deviation.

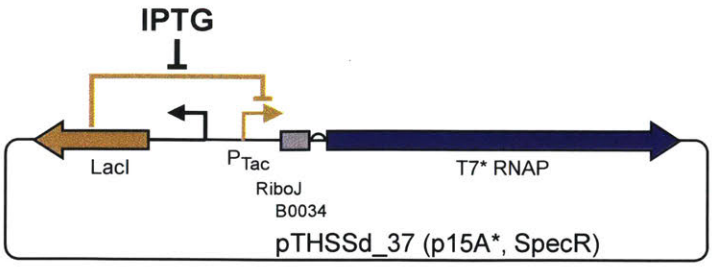


Figure 2-18: **Plasmid used for full-length T7* RNAP toxicity measurement.** pTHSSd_37 was used to characterize the growth impact of expressing T7* RNAP from the same expression system used to drive the thee-and four-piece polymerases. It contains the full T7* RNAP CDS driven by a P_{Tac} expression system with RiboJ and the B0034 RBS.

Figure 2-19: **Reporter plasmids.**
The reporter constructs used in this work are based on plasmid pUA66 [244], which has a pSC101 origin of replication. The GFPmut2 gene is replaced with sfGFP [247], and the kanamycin resistance cassette is replaced with an ampicillin resistance cassette. Variants were created with the P_{T7}, P_{T3}, P_{K1F}, and P_{CGG} promoters driving expression of GFP (pTHSSd_8-11). A strong RBS (RBS_GFP: TGTCAATTTCCGCGATAGAGGAGGTAAAG) was generated using the RBS calculator and used to control translation of GFP. For assaying GFP : α fragment fusions, a reporter variant was built with the P_{T7} mRFP1 expression cassette from NiF_489 (Temme2012) (pTHSSd_12). A negative control plasmid lacking the GFP expression cassette was also generated (pTHSSd_13).

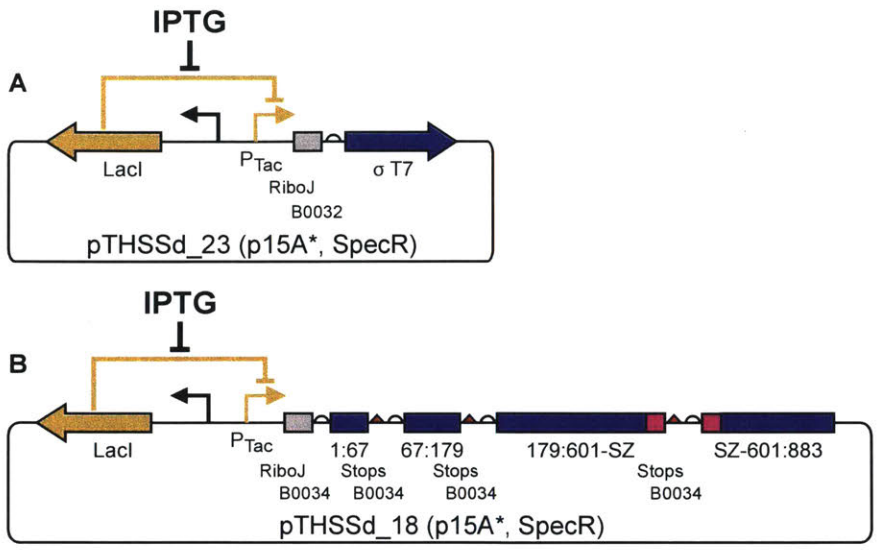


Figure 2-20: **Inducible expression plasmids.**
Plasmids for the inducible expression of genes from P_{Tac} are built from pSB3C5 [243], which has a p15A origin. This origin appears to maintain at a higher copy number than standard, so we refer to it as p15A*. The chloramphenicol resistance cassette is replaced with a spectinomycin resistance cassette, and a modified section from pEXT20 [240] containing a LacI expression cassette, a random spacer, and short P_{Tac} promoter is inserted into the plasmid. The lacO binding site in P_{Tac} is mutated to be symmetric (AATTGTGAGCGCTCACAATT), and is followed by RiboJ [36]. (A) In most systems, only one coding sequence is expressed under the control of P_{Tac} and the B0032 RBS (BBa_B0032) is used. A number of proteins were expressed from plasmids similar to this, including σ fragments (pTHSSd_23-26), the null fragment (pTHSSd_27), the α fragment (pTHSSd_29), and α: FP fusions (pTHSSd_30-33). (B) To test T7* RNAP fragmented into three or four fragments, vectors were constructed that express the fragments or a subset of them on one cistron (pTHSSd_14-22). The B0034 RBS (BBa_B0034) is used for each fragment, and a double stop codon terminates each fragment coding sequence. Two negative control plasmids were made that lack any inducible gene but contain LacI (pTHSSd_38, 39). pTHSSd_38 contains the LacI cassette and P_{Tac} promoter system found in the splitposon and bisection library, while pTHSSd_39 only contains the LacI expression cassette.

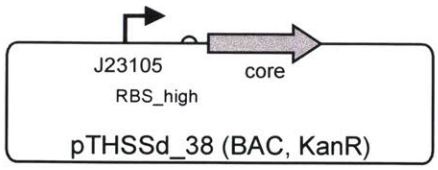


Figure 2-21: **Core fragment expression plasmids.**
The core and β core fragments are expressed from plasmids based on pBACr-Mgr940 [245] (BBa_J61039), which carries kanamycin resistance and an F plasmid derived origin. The constitutive P_{J23105} promoter (BBa_J23105) is used to drive expression of the core fragment, core fragment variants, the full T7* RNAP, or β core fragment (pTHSSd_38-42), using different ribosome binding sites to control the strength of expression. The main RBSs used were derived from a degenerate library based on B0032: RBS_high (TACTAGAGTCATTTATGAAAGTACTAG) is used for most constructs, RBS_low: (TACTAGAGTCAGCCAAGAAAGTACTAG) is used for the lower level of core fragment expression. A negative control of this plasmid was constructed that lacks an RBS and coding sequence (pTHSSd_43).

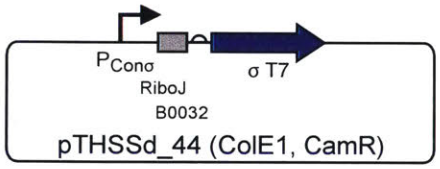


Figure 2-22: **Constitutive σ expression plasmids.** For the null fragment and α fragment assays, σ fragments were constitutively expressed from plasmids based on pSB1C3 [243], which has a ColE1 origin and chloramphenicol resistance. A variant of the constitutive promoter P_{J23105} ($P_{Con\sigma}$: TTGACAGCTAGCTCAGTCCTAGGCTATAGGCTAG), RiboJ, and the B0032 RBS are used to drive expression of each of the four σ fragments (pTHSSd_44-47). A negative control was made that lacks any piece of the expression cassette (pTHSSd_48).

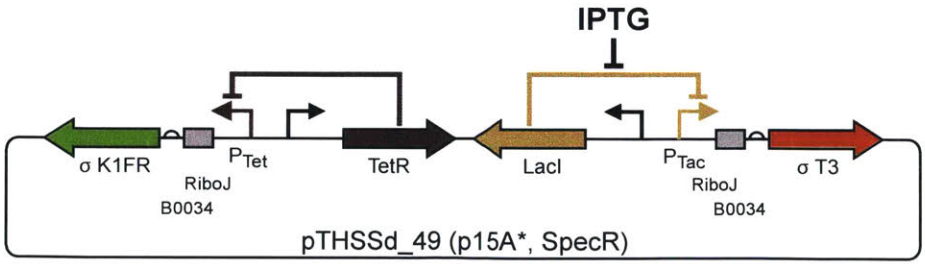


Figure 2-23: **σ fragment competition plasmids.** A variant of the T3 σ fragment inducible expression plasmid was built to test σ fragment competition. A modified P_{Tet} expression system [70] is added behind the P_{Tac} expression system facing in the reverse direction. The P_{Tet} promoter is followed by RiboJ and drives expression of the K1FR σ fragment. Both σ^{T3} and σ^{K1FR} use the B0034 RBS.

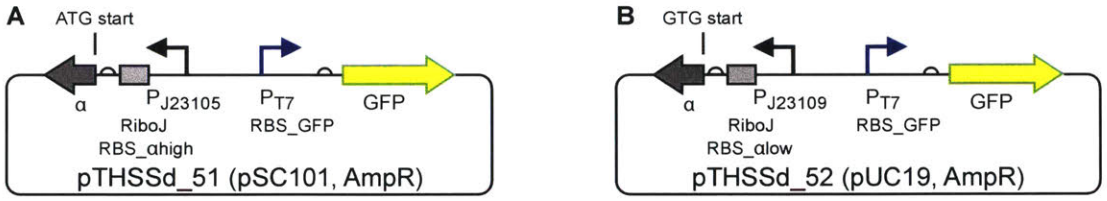


Figure 2-24: **Reporter plasmids with α fragment compensation.**
A. A constitutive α fragment expression cassette is added in the reverse direction to the P_{T7} reporter plasmid before the P_{T7} promoter to make pTHSSd_51. This cassette drives production of the α fragment with BBa_J23105, RiboJ, and a RBS derived from B0032 (RBS_αhigh: TCAACCACGAAAGTACTAG).
B. pTHSSd_52 has the same two cassettes as pTHSSd_51, inserted into a pUC19 [242] ampicillin resistant backbone. The α fragment cassette is changed to lower its expression level: the promoter is switched to P_{J23109} (BBa_J23109), a different RBS is used (RBS_αlow: CTAGTACTTTCGTTCATGA), and the α fragment start codon is changed to a GTG from ATG.

2.9 Tables

Table 2.1: **Statistics of T7 RNA polymerase bisection mapping.**

Library	Library variants[a]	Harvested colonies[b]	Library coverage[c]
Transposon Insertion	9026	7.8×10^5	87
Final	4564	6.0×10^5	132

a. The number of possible variants in the insertion and final split T7* RNAP libraries. Equal to 2x the size of he insertion vector and 2x the size of the T7* RNAP 41-876 fragment, respectively.
b. The approximate number of colonies scraped and pooled for the two libraries.
c. The harvested clones divided by the number of variants in each library.

Table 2.2: **Improved activity clones from the K1F σ fragment ePCR library.**

Clone #[a]	Mutations[b]		
1	**M151K**	Y247H	
2	K122R	**M151K**	
3	E95G	**M151K**	
4	**M151K**		
5	Q70R	**M151K**	
6	**M151K**		
7	Q154R	**M151V**	H173Y
8	**M151V**	H173R	E256K
9	**M151V**	H173R	E256K
10	**M151K**	Q187H	
11	E53K	**M151K**	
12	Q70R	**M151R**	
13	**M151R**	K227R	

a. The clones are ordered from least to most active.
b. Residues are numbered by their position in the full-length T7 RNAP sequence.
c. Mutations affecting residue 750 (the residue mutated in K1FR) are shown in bold.

Table 2.3: **Improved activity clones from the N4 σ fragment ePCR library.**

Clone #[a]	Mutations[b]	
1	H755R	
2	H755R	
3	H755R	
4	H755R	
5	H755R	
6	H755R	
7	H755R	
8	H755R	
9	H755R	
10	V725A	H772R
11	H755R	
12	H755R	

a. The clones are ordered from least to most active.
b. Residues are numbered by their position in the full-length T7 RNAP sequence.
c. Additional silent mutations were found in #1 and #4.

Table 2.4: **Comparison of null fragment variants.**

Null variant	Fold-repression[a]
σ^{T7} Δ882-3	9
σ^{T7} Y639A	12
σ^{T7} H811A	11
σ^{CGG} Δ882-3	14
σ^{CGG} Y639A	18
σ^{CGG} H811A	16
σ^{T7WT} Δ882-3	12
σ^{T7WT} Y639A	13
σ^{T7WT} H811A	13

a. Fold repression was calculated as the activity of a P_{T7} promoter with constitutive σ^{T7} expression and no null fragment induction (0 μM IPTG) divided by the activity of the P_{T7} promoter with constitutive σ^{T7} expression and high null fragment induction (1000 μM IPTG). Values are the mean fold repression from three independent assays performed on different days.

Table 2.5: **Modeling parameters.**

Parameter	RNAP model	Resource allocator model
γ	$3 \times 10^{-4}\,s^{-1}$	$3 \times 10^{-4}\,s^{-1}$
$u_i(off)$	$0\,M\,s^{-1}$	$0\,M\,s^{-1}$
$u_i(on)$	$3 \times 10^{-11}\,M\,s^{-1}$	$6 \times 10^{-11}\,M\,s^{-1}$
v	-	$3 \times 10^{-11}\,M\,s^{-1}$
k_a	-	$4.5 \times 10^{5}\,M\,s^{-1}$
k_d	-	$2 \times 10^{-4}\,s^{-1}$

Table 2.6: **New plasmids used in this work.**

Name	Origin[a]	Marker[b]	Description
pTHSSd_1	ColE1	K/C	Splitposon in KanR ColE1 backbone
pTHSSd_2	ColE1	A T7*	RNAP 41-876 transposition target
pTHSSd_3	p15A*	S T7	RNAP size-selection & expression vector
pTHSSd_4	p15A*	S/K	P_{Tac} expression of T7 RNAP* split at 601
pTHSSd_5	p15A*	S/K	P_{Tac} expression of T7 RNAP* split at 601 with SZ17
pTHSSd_6	p15A*	S/K	P_{Tac} expression of T7 RNAP* split at 601 with SZ18
pTHSSd_7	p15A*	S/K	P_{Tac} expression of T7 RNAP* split at 601 with SynZIPS
pTHSSd_8	pSC101	A	P_{T7} sfGFP reporter
pTHSSd_9	pSC101	A	P_{T3} sfGFP reporter
pTHSSd_10	pSC101	A	P_{K1F} sfGFP reporter
pTHSSd_11	pSC101	A	P_{CGG} sfGFP reporter
pTHSSd_12	pSC101	A	P_{T7} mRFP reporter
pTHSSd_13	pSC101	A	reporter negative control
pTHSSd_14	p15A*	S	Triple split (67, 601-SZ)
pTHSSd_15	p15A*	S	Triple split no fragment 1:67
pTHSSd_16	p15A*	S	Triple split no fragment 67:601
pTHSSd_17	p15A*	S	Triple split no fragment 601:883
pTHSSd_18	p15A*	S	Quad split (61, 179, 601-SZ)
pTHSSd_19	p15A*	S	Quad split no fragment 1:67
pTHSSd_20	p15A*	S	Quad split no fragment 67:179
pTHSSd_21	p15A*	S	Quad split no fragment 179:601
pTHSSd_22	p15A*	S	Quad split no fragment 601:883
pTHSSd_23	p15A*	S	P_{Tac} T7 σ fragment expression
pTHSSd_24	p15A*	S	P_{Tac} T3 σ fragment expression
pTHSSd_25	p15A*	S	P_{Tac} K1FR σ fragment expression
pTHSSd_26	p15A*	S	P_{Tac} CGG σ fragment expression
pTHSSd_27	p15A*	S	P_{Tac} CGG_Y639A null fragment expression
pTHSSd_28	p15A*	S	P_{Tac} RFP expression
pTHSSd_29	p15A*	S	P_{Tac} α fragment expression
pTHSSd_30	p15A*	S	P_{Tac} GFP-α expression
pTHSSd_31	p15A*	S	P_{Tac} α-GFP expression
pTHSSd_32	p15A*	S	P_{Tac} RFP-α expression
pTHSSd_33	p15A*	S	P_{Tac} α-mRFP1 expression
pTHSSd_34	p15A*	S	P_{Tac} GFP expression
pTHSSd_35	p15A*	S	inducible expression negative control v1
pTHSSd_36	p15A*	S	inducible expression negative control v2
pTHSSd_37	p15A*	S	Inducible full length T7* RNAP control
pTHSSd_38	BAC	K	High core fragment expression
pTHSSd_39	BAC	K	Low core fragment expression
pTHSSd_40	BAC	K	High core fragment expression without SynZIP
pTHSSd_41	BAC	K	High full length T7 RNAP* expression
pTHSSd_42	BAC	K	β core fragment expression
pTHSSd_43	BAC	K	core fragment expression negative control
pTHSSd_44	ColE1	C	constitutive expression of T7 σ fragment
pTHSSd_45	ColE1	C	constitutive expression of T3 σ fragment
pTHSSd_46	ColE1	C	constitutive expression of K1FR σ fragment
pTHSSd_47	ColE1	C	constitutive expression of CGG σ fragment
pTHSSd_48	ColE1	C	constitutive expression negative control
pTHSSd_49	p15A*	S	P_{Tac} T3 σ fragment, P_{Tet} K1FR σ fragment expression
pTHSSd_50	p15A*	S	P_{Tac} GFP, P_{Tet} RFP expression
pTHSSd_51	pSC101	A	pSC101 α fragment compensated reporter
pTHSSd_52	pUC19	A	pUC19 α fragment compensated reporter

a. ColE1: derived from pSB1C3, p15A*: derived from pSB3C5, appears to maintain at a higher copy number than p15A, pSC101: derived from pUA66, BAC: derived from pBACr-Mgr940, pUC19: derived from a pUC19 vector.

b. A: ampicillin, K: kanamycin, C: chloramphenicol, S: spectinomycin.

Chapter 3

Constant gene expression at any copy number using feedforward stabilized promoters

3.1 Abstract

The internal environment of growing cells is variable and dynamic, making it difficult to introduce reliable parts, such as promoters, for genetic engineering. Precision measurements under one set of conditions are unreliable because part performance breaks down in different genetic backgrounds and growth conditions. Here, we apply concepts from control theory to design promoters that maintain constant levels of expression, whether they are carried at single copy in the genome or on high-copy plasmids. A surprising prediction is that independence to copy number can be achieved with an incoherent feedforward loop (iFFL) if the negative regulation is perfectly non-cooperative (i.e., repression with Hill coefficient n = 1). To achieve this, iFFLs are implemented in Escherichia coli using Transcription Activator Like Effectors (TALEs). The resulting stabilized promoters generate near-identical expression across different genome locations and plasmid backbones (pSC101, p15a, ColE1, pUC). They also provide robustness to strain mutations and growth media. Further, their strength is tunable and can be used to maintain constant ratios between proteins.

3.2 Introduction

DNA copy number is a major source of uncertainty when designing genetic systems [251]. While often treated as a constant, the copy number of plasmids can vary widely. Within a clonal population, plasmid copy number varies over time and between cells due to stochastic fluctuatiuons [252, 253]. Many changes to the cell or environment can alter plasmid copy number, including host strain [254–257], media [257–261], temperature [262], and growth rate [263]. Changing the size of a plasmid or genes transcribed also can impact its copy number [264, 265]. Even inserting a system into the genome does not protect it from copy number effects. Fast-dividing bacteria initiate genome replication more than once per cell division and contain partially replicated genomes, leading to an enrichment of genes closer to the origin of replication [266]. Due to this, the average copy number of different locations on the genome can span up to 8-fold, depending on the rate of cell division [267,268]. Therefore, stabilizing a genetic system with respect to copy number is hypothesized to improve robustness and enable it to be modified or transferred between genetic locations and strains with less chance of disruption.

The total level of gene expression produced by constitutive promoters closely reflects the copy number at which they are present in the cell [269]. Minimal constitutive bacterial promoters consist of DNA sequences recognized by the factor of the RNA polymerase holoenzyme complex and a downstream transcriptional start site (Figure 3-1a). The performance of minimal promoters can be altered by changing the surrounding genetic sequence, so they are often flanked by insulators [36, 150, 270, 271] (Figure 3-1a).

Stabilized promoters incorporate additional elements that decouple gene expression from copy number. Stabilizing a promoter requires the introduction of regulation that detects changes in copy number and compensates the promoter activity accordingly. This could be achieved by autoregulatory feedback, integral feedback, or incoherent feedforward loops [272–274]. In comparing approaches, one has to consider design feasibility and the ability to achieve perfect adaptation (convergence to a constant level of expression regardless of copy number). While autoregulation is straightforward to implement [272], it cannot achieve perfect adaptation without infinitely cooperative repression (See 3.6.2) and it has the potential

to cause oscillations [275]. In contrast, integral feedback can achieve perfect adaptation but is complicated to implement [276].

3.3 Results

We chose to implement an iFFL because it is both simple and predicted to achieve perfect adaptation. Feedforward loops are commonly found in natural transcription networks [277, 278], and have been shown to compensate for differences in chromosomal copy number in yeast [274]. An iFFL has been used for a similar purpose in mammalian cells, where it was shown to partially decouple gene expression from the amount of DNA transfected [204]. A feedforward loop is labeled incoherent when an input signal is split and both positively and negatively controls the output [278]. In this case the input signal is copy number, which is expected to affect gene expression from all promoters in a construct equally. Our design centers on a stabilized promoter, the core of which is the same as the insulated promoter in Figure 3-1a. On its own, this promoter is positively correlated with copy number, which functions as the input signal. Negative regulation is introduced by making the promoter responsive to a protein repressor, which itself is controlled by an insulated promoter (Figure 3-1b). Thus, increased copy number leads to increased repressor expression, which interacts with the stabilized promoter to cancel out the change in expression caused by copy number.

The promoter output y can be modeled at steady-state using a simplified power law for the repressor transfer function (Figure 3-1c),

$$y = c/x^n \tag{3.1}$$

where c is the copy number, x is the concentration of repressor, and n is the cooperativity of repression. Because repressor expression is controlled with an unregulated promoter, x scales with the copy number and

$$y = c^{(1-n)} \tag{3.2}$$

Hence, when $n = 1$ (repression is non-cooperative), the output promoter activity is predicted to be independent of copy number (Figure 3-1d). Transcriptional repression generally

follows the Hill equation, which can approach this ideal power law transfer function as the repressor concentration or binding affinity increases (See 3.6.1). However, there is a tradeoff between output strength and adaptation. Increasing expression of the repressor lowers the stabilization error (the relative change in gene expression as copy number is increased) but also decreases the absolute output strength (Figure 3-1e). The key parameter is the expression level of the repressor at the lowest copy number; expression must be sufficient to ensure good adaptation but not so high as to render the output too weak to be useful.

We chose to build stabilized promoters using TALE proteins because they can be programmed to tightly bind arbitrary DNA sequences [97,98] and have been previously shown to achieve greater than 100-fold repression in *E. coli* [103,279]. Importantly, they bind as monomers to a single operator, which supports their use to achieve non-cooperative repression. To identify the optimal site for the operator, a TALE from the literature [280] was used and the location and orientation of the operator was varied in the backbone of the very strong constitutive promoter P_{T7A1} [31] (Figures 3-1f, 3-4, and 3-5). Then, the SIFTED suite [281] was used to design TALEs that bind 18 bp operators and are predicted to not bind to the *E. coli* genome. The TALE genes were recoded to avoid predicted evolutionary instability [282] (Figure 3-6). Two TALE-promoter pairs (TALEsp1 and TALEsp2) were identified that are orthogonal to each other and generate at least ~90 and ~230-fold repression respectively (Figures 3-1g, and 3-8). Importantly, the response functions approach the power-law approximation of Equation 3.1 at higher repressor expression levels, with $n = 0.95 \pm 0.04$ and 1.01 ± 0.02 (mean of three replicates $\pm$ S.D.), respectively.

We characterized the effect of copy number on an insulated promoter by moving it onto a set of variant pSC101 backbones that are maintained at a range of copy numbers [283] (Figure 3-9). These plasmids are nearly identical, but contain up to four mutations in the repA gene of the pSC101 origin that alter copy number regulation. In exponential growth, the copy number of these plasmid backbones spanned a range from 3 to 100 copies per cell and expression from the insulated promoter varied by 20-fold when carried on these backbones (Figure 3-2a, black points).

We built a stabilized promoter using TALEsp1 and characterized it with the pSC101 backbones (Figure 3-2a, blue points). The promoter and RBS controlling TALE expression

were chosen to express a level of repressor sufficient for good adaptation at copy numbers matching pSC101 (~3/cell) or higher while maintaining high output expression (Figure 3-10). This stabilized promoter successfully buffered against most of the effects of copy number, showing near identical expression levels. Remarkably, the population distributions measured by cytometry nearly collapsed onto a single distribution (Figure 3-2a, inset).

The stabilized promoter was then tested across different plasmid backbones. Five commonly used origins of replication were chosen: incW, pSC101, p15A, ColE1, and pUC. Like the pSC101 set, they vary widely in copy number (from ~2 to 38 copies/cell in exponential growth, Figure 3-11), but they also vary in size, genes expressed, and possibly cellular localization [284, 285]. When the insulated promoter was used on each backbone, expression varied 16-fold (Figure 3-2b). In contrast, the stabilized promoter was able to eliminate most of the variability.

Stabilized promoters have to be tunable so that the output level can be set depending on the optimal expression level required for a gene. Additionally, it is undesirable that the promoter controlling each gene in a genetic system has its own corresponding TALE, because this increases the constructs DNA size and number of characterized repressors required. Thus, we built a series of stabilized promoters that generate varied expression levels, all of which are controlled by the same TALE (Figure 3-2c). We made mutations to the UP element and -10 box of the TALEsp1 output promoter (P_{UPsp1}) to generate four variant promoters with a range of strengths (Figure 3-12). The promoters were combined with the TALEsp1 expression cassette to create a set of four variant stabilized promoters with different output levels. All promoters are able to maintain consistent expression levels across a wide range of copy numbers.

With these promoters, one TALE can be used to stabilize multiple operons at different ratios, a feat required for the stabilization of more complex genetic systems, such as metabolic pathways and molecular machines [19, 69]. The P_{UPsp1} and P_{DNsp1} output promoters were used with the TALEsp1 expression cassette in a single cell to maintain GFP and RFP at high and low levels, respectively (Figure 3-2d). This multi operon stabilized system maintained a nearly constant ratio of gene expression across the pSC101 plasmid set.

In order to use stabilized promoters at lower copy numbers than pSC101, we built a second

stabilized promoter with the TALEsp2 repressor. For this promoter, TALE expression was increased so that good stabilization can be achieved at copy numbers below pSC101 (Figure 3-10). The higher level of TALE expression leads to a lower level of expression from the stabilized promoter.

The TALEsp2 stabilized promoter was tested in the genome for its ability to buffer against copy number differences caused by rapid cell division. The promoter-reporter fusions were flanked by insulators and inserted randomLy in the genome using a Tn5 transposon system [286] (Methods). Single insertion events were isolated, and the insertion sites determined using arbitrary PCR. This yielded strains with 35 insertions of the stabilized promoter distributed across the genome. This process was repeated with the unstabilized insulated promoter to create 35 additional strains with distributed insertions. As expected, the unstabilized promoter showed a clear trend of maximal expression near the origin (Figures 3-2e and 3-13). In contrast, the stabilized library showed almost no position-dependent differences in gene expression. Further, the levels of expression from the stabilized promoter match that obtained when the promoters are carried on plasmids. This demonstrates that a stabilized promoter can allow a system to be moved from plasmids to arbitrary locations in the genome without impacting the level of expression.

The stabilized promoters are designed to primarily buffer against changes in DNA copy number. However, a number of perturbations can impact gene expression through DNA copy number as an intermediate. It is expected that stabilized promoters would additionally buffer genetic systems against these changes. For example, mutations to the host genome can impact plasmid replication systems [254–257]. In previous work, we serendipitously observed this effect when a host strain dealt with a toxic genetic circuit carried on a plasmid by making a genomic mutation that reduced the copy number [287]. This occurred due to a mutation to pcnB, which codes for a poly(A) polymerase that has been shown to affect global RNA stability [288] and as a result affects the copy number of plasmids that rely on RNA regulation [254] (e.g., p15A, ColE1, and pUC). We sequenced and maintained this strain and tested here whether the stabilized promoter could ameliorate the impact of this mutation. Indeed, the stabilized promoter was able to achieve similar levels of expression across both strains for all the plasmid backbones (Figure 3-3a). This will lead to improved evolutionary

robustness as the use of a stabilized promoter eliminates some mutational paths that could disrupt function. This is analogous to increasing the stability of a genetic construct by eliminating repeats that provoke homologous recombination or removing transposons from a strain [289, 290].

Changes in media and growth conditions can also change copy number and commonly break genetic circuits and metabolic pathways [7, 291]. We tested whether a stabilized promoter could reduce the variation in gene expression caused by different growth media (Figure 3-3b). Four variants of M9 media were made with different carbon sources (glucose or glycerol) and amino acids (casamino acids or leucine) (Methods). These compositions lead to large differences in the growth rate, with average doubling times between 40 and 140 minutes. Further, media composition is known to impact plasmid copy number differentially depending on the origin of replication [262]. When the unstabilized promoter is compared amongst all media conditions and origins, there is a 90-fold spread in expression levels. The TALEsp1 stabilized promoter eliminates most of this effect.

3.4 Discussion

This work started with a simple question. Could we design a promoter that produces that same protein concentration no matter where it is placed? Applying principles from control theory, we were able to design a new class of stabilized promoter that maintains the same level of gene expression irrespective of the plasmid backbone or location in the genome. This is achieved by harnessing the feedforward loop, a common motif in natural regulatory networks that have been implicated in broad functions, including maintaining homeostasis between proteins, implementing dynamic ordering, and producing a pulse of gene expression [143, 278, 292–295]. While the stabilized promoter was designed to buffer gene expression against the effects of changing DNA copy number, our results demonstrate a wider robustness to genome mutations and media. Collectively, robustness to these conditions eliminates much of the context dependence that plagues precision genetic engineering.

3.5 Methods

3.5.1 Strains and Media

Escherichia coli DH10B (taxid: 316385) was used for all routine cloning and characterization. LB-Miller media (BD, Franklin Lakes, NJ - USA) was used for strain propagation and all cytometry assays. SOC media [SOB media (Teknova, Hollister, CA - USA) plus 20 mM Glucose] was used for transformation recovery. The Mix & Go *E. coli* Transformation Kit (Zymo Research, Irvine, CA - USA) was used to prepare competent cells, with SOB used for the outgrowth step. M9 minimal media [1xM9 salts (Teknova, Hollister, CA - USA), 2 mM MgSO4, 0.1 mM CaCl2] supplemented with a carbon source and amino acids was used for the plate reader assays. *E. coli* S17-1 pir (Biomedal, Seville - Spain) was used for Tn5 transpositions. *E. coli* DH10B broken colony 6A [287] was used when testing host context.

3.5.2 Plasmids and Genetic Parts

lasmids were initially derived from pSB1A3 [243], pSB3C5 [243], or pUA66 [244]. pSC101 origins were derived from pUA66 and pCL1921 [296], p15A origins were derived from pSB3C5, ColE1 and pUC origins were derived from pSB1A3, incW origins were derived from Nif_249 [88]. Origins were sequenced, corrected for errors, and problematic restriction sites removed. pSC101 variant origins were derived from the pCL1921 origin through targeted mutations and error prone PCR. Resistance cassettes were derived from pSB1A3 and pSB1C3, with the resistance gene swapped to streptomycin adenyltransferase where necessary. Promoters P_{J23101} (parts.igem.org: BBa_J23101), P_{J23105} (BBa_J23105), P_{T7A1} [31], $P_{TacSymO}$ (P_{Tac} [240] modified to have a symmetric lacO site), and modified forms of P_{T7A1} were used for gene expression. Ribozymes RiboJ [36] and SarJ [36] were used to insulate transcripts from promoters. B0032 (BBa_B0032), B0034 (BBa_B0034) or RBSs created through degenerate PCRs were used to drive protein translation. Terminators B0062 (BBa_B0062), B0015 (BBa_B0015), L3S2P1 [150], and ECK120029600 [150] were used to stop transcription. Gene expression was measured by expressing sfGFP [247] (GFP) or mRFP [246] (RFP). TALEs were based on TALE1 [280] truncated to residues 153–840 and modified to remove predicted evolution-

arily unstable regions in the N-terminal sequence. New TALE DNA binding repeats were built using the Platinum Gate TALE assembly kit [297], and synthesized by Gen9 after codon shuffling to reduce evolutionary instability. Plasmid pBAMD1-4 [286] was modified and used for genomic integration. A LacI expression cassette from pEXT20 [240] modified to have extra insulation and a symmetric lac operator was used for inducible gene expression. See Figures 3-9, 3-11, and 3-14 to 3-16 for construct and plasmid schematics, and Tables 3.1 and 3.2 for a list of all plasmids used in this work. Selected plasmids will be made available on Addgene (https://www.addgene.org/Christopher_Voigt/).

3.5.3 Cytometry Assays

All measurements shown except for when media is varied were taken by cytometry of cells in exponential growth as follows. Glycerol stocks of strains containing the plasmids of interest were streaked on LB + 2% Agar (BD, Franklin Lakes, NJ USA) plates and grown overnight at 37°C. Plates were stored at 4°C after growth. Single colonies were inoculated into 0.5 mL LB with antibiotics in 2-mL 96-deepwell plates (USA Scientific, Orlando, FL - USA) sealed with an AeraSeal film (Excel Scientific, Victorville, CA - USA) and grown at 37°C, 900 rpm overnight in a deepwell shaker. The overnight growths were diluted 1:225 into 150 uL LB with antibiotics in 300-μL 96-well V-bottom plates (Thermo Scientific Nunc, Waltham, MA - USA) sealed with an AeraSeal film and grown at 37°C, 1000 rpm. After 2 h the growths were diluted 1:900 into 150 μL LB with antibiotics plus inducer where necessary in 300-μL 96-well V-bottom plates, sealed with an AeraSeal film, and grown at 37°C, 1000 rpm for 6 h. 20 μL of each sample was removed and diluted in 180 μL PBS + 2 mg/mL kanamycin to halt protein production. Cells diluted in PBS were either characterized immediately with flow cytometry or stored at 4°C until characterization.

3.5.4 Cytometry Analysis

Fluorescence characterization with cytometry was performed on a BD LSR Fortessa flow cytometer with HTS attachment (BD, Franklin Lakes, NJ USA). Cells diluted in PBS + kanamycin were run at a rate of 0.5 μL/s. The events were gated by forward scatter height

and side scatter height to reduce false events. Gating was determined by eye and was kept constant for all analyses within each triplicate experiment. After gating, thousands of events were used for analysis. For each sample, the median GFP (and RFP, where needed) fluorescence was calculated. The median fluorescence of a sample with *E. coli* DH10B carrying no fluorescent protein was subtracted from these values to correct for background fluorescence.

3.5.5 qPCR measurements

Copy number was measured with qPCR using the same growth protocol as used for cytometry. After the 6hr growth, 10μL of each sample was boiled for 10:00 to lyse cells, spun down, and 2μL of the lysate used in qPCR reactions. Primers [298] that amplify a region of the AmpR cassette on the plasmid backbones and the dxs gene on the genome were used to quantify plasmid copy number relative to the genome. qPCR was run in a LightCycler 96 (Roche Diagnostics Corporation, Indianapolis, IN - USA) using SsoFast EvaGreen Supermix with Low ROX (BioRad, Hercules, CA USA) in a 10μL reaction volume with the protocol: [98C-3:00, (98C-0:10, 60C-0:10, 72C-0:10)x40]. The relative plasmid copy number was calculated using the $\Delta\Delta C_t$ formula with a plasmid containing one copy of the AmpR cassette and dxs gene as the calibrator and assuming an efficiency of 100%.

3.5.6 Plate Reader Assays

The assays involving growth in different media were performed as follows. Overnights of strains containing the plasmids of interest were grown from colonies in the same manner as for the cytometry assays. The overnight growths were diluted 1:225 into 150 μL M9 minimal media plus supplements (0.4% Glucose + 0.2% casamino acids / 0.4% Glycerol + 0.2% casamino acids / 0.4% Glucose + 0.5 mM leucine / 0.4% Glycerol + 0.5 mM leucine) with antibiotics in 300-μL 96-well V-bottom plates sealed with an AeraSeal film and grown at 37°C, 1000 rpm. After 2 h the growths were diluted 1:900 (media containing casamino acids) or 1:100 (media containing Leucine) into 150 μL of the same media used previously in 300-μL 96-well black walled optical bottom plates (Thermo Scientific Nunc, Waltham, MA USA) sealed with a BreathEasy film (Sigma-Aldrich, St. Louis, MO USA), and grown in

a Synergy H1 plate reader (BioTek, Winooski, VT USA) at 37°C, 1000 rpm. OD600 and GFP fluorescence were measured every 20 minutes over 18 h of growth.

3.5.7 Plate Reader Analysis

OD600 and fluorescence reading were also taken from wells containing media with no cells, and the first 5 readings from four such wells were averaged and subtracted from the appropriate sample measurements to remove background. Data points that were followed by a large drop in OD600 (>=10%) were removed to eliminate erroneous measurements with inflated OD readings. OD600 values were converted to equivalent 1cm path length measurements. To calculate doubling times, the last measurement with OD600 < 0.04 and the first measurement with OD600 > 0.16 were identified, and the doubling time was calculated assuming exponential growth between those two points. To calculate the GFP expression of each sample in exponential growth, the two closest measurements above and below OD600=0.1 were identified for each sample and their average fluorescence value was calculated.

3.5.8 Tn5 Transposition

E. coli S17-1 λpir containing Tn5 transposition vectors (pTHSSe_52 or pTHSSe_53) and DH10B cells containing an empty AmpR vector (pTHSSe_54) were inoculated into LB-Miller with antibiotics and grown at 37°C, 250 rpm overnight. The overnights were diluted 1:100 into LB with antibiotics and grown at 37°C, until at approximately 0.3-0.4 OD600. An aliquot containing an equivalent number of cells to 100 μL at OD600=1.0 of each of the transposition vector growths and the same number of aliquots of empty vector cells were spun down at 13,000 rpm for 1:00, resuspended with 1mL LB, and spun down again. Each transposition vector aliquot was resuspended in 20 μL LB and mixed with an aliquot of the empty vector cells. These mixes were spotted onto a LB plate without anibiotics and grown at 37°C for approximately 4h. After growth, the spots were scraped, resuspended in 1 mL LB, plated on LB agar plates with Ampicillin and Spectinomycin, and grown overnight at 37°C.

3.5.9 Arbitrary PCR Sequencing

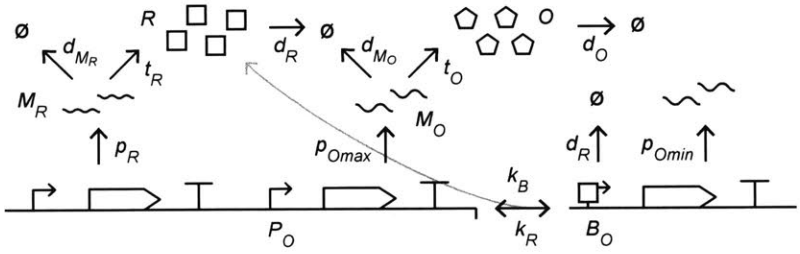
Arbitrary PCR following Das et. al. [299] was used to locate the position of genomic integrations. Briefly, cells containing integrations were boiled, spun down, and 1μL of cell lysate was used as the template in a PCR reaction. A section of genome before or after the transposition was amplified using Phusion polymerase in HF buffer (Thermo Scientific, Waltham, MA - USA): [98C-0:30, (98C-0:10, 30C-0:30, 72C-1:30)x6, (98C-0:10, 45C-0:30, 72C-1:30)x30, 72C-5:00]. 1μL of these reactions were used as templates for secondary amplifications: [98C-0:30, (98C-0:10, 60C-0:30, 72C-1:30)x28, 72C-5:00]. These reactions were sequenced to determine the genomic insertion location.

3.5.10 Error-prone PCR

Error-prone PCR was used to introduce copy number-altering mutations into the repA gene of the pSC101 origin. GoTaq Polymerase (Promega, Madison, WI USA) was used in its standard buffer with $MgCl_2$ added to a final concentration of 11.5 mM Mg^{2+} and 31 amplification cycles were run. The resulting products were cloned into fresh plasmid backbones and assayed for the desired GFP production. Promising clones were fully sequenced to identify the mutations and ensure that mutations were only found in the plasmid origin.

3.6 Modeling

3.6.1 Mathematical model of iFFL stabilized promoter



The iFFL regulatory system can be modeled with the following set of reactions:

$\rightarrow M_R$	,	cp_R	Production of the mRNA for the repressor
$M_R \rightarrow \varnothing$	,	d_{M_R}	Degradation of the mRNA for the repressor
$M_R \rightarrow M_R + R$	,	t_R	Translation of the repressor
$R \rightarrow \varnothing$	,	d_R	Degradation of repressor
$P_O \rightarrow P_O + M_O$	,	p_{Omax}	Production of output mRNA from an unrepressed promoter
$B_O \rightarrow B_O + M_O$	,	p_{Omin}	Leak of output mRNA from a repressed promoter
$nR + P_O \rightarrow B_O$	,	k_B	Binding of the repressor to the output promoter
$B_O \rightarrow nR + P_O$	,	k_R	unbinding of the repressor from the output promoter
$B_O \rightarrow P_O + (n-1)R$	,	d_R	Degradation of the repressor bound to the promoter
$M_O \rightarrow \varnothing$	,	d_{M_O}	Degradation of the mRNA for the output protein
$M_O \rightarrow M_O + O$	,	t_O	Translation of the output protein
$O \rightarrow \varnothing$	,	d_O	Degradation of output protein

and:

$P_O + B_O = c$	Conservation of promoters

Where:

- M_R, M_O are the mRNAs coding for the repressor and output proteins,
- R, O are the the repressor and output proteins,
- P_O is the promoter driving production of the output,
- B_O is the output promoter bound to the repressor protein,

- p_R is the strength of the promoter driving the repressor,
- d_{M_R}, d_{M_O} are the degradation rates of the repressor and output mRNAs,
- t_R, t_O represent the strength of the RBSs driving the repressor and output,
- p_{Omax} represents the strength of the unrepressed output promoter,
- p_{Omin} represents leaky production from the repressed output promoter,
- k_B is the rate of binding of the repressor to the output promoter,
- k_R is the rate of unbinding of the repressor from the output promoter,
- n is the cooperativity of repressor binding,
- c is the copy number of the system

Combining these reactions into a system of equations yields

$$\dot{M}_R = cp_R - d_{M_R} M_R \tag{3.3}$$

$$\dot{R} = t_R M_R - d_R R - nk_B R^n P_O + nk_R B_O + (n-1) d_R B_O \tag{3.4}$$

$$\dot{P}_O = k_R B_O - k_B R^n P_O + d_R B_O \tag{3.5}$$

$$\dot{B}_O = k_B R^n P_O - k_R B_O - d_R B_O \tag{3.6}$$

$$\dot{M}_O = p_{Omax} P_O + p_{Omin} B_O - d_{M_O} M_O \tag{3.7}$$

$$\dot{O} = t_O M_O - d_O O \tag{3.8}$$

$$c = P_O + B_O \quad . \tag{3.9}$$

In order to model the performance of the iFFL system, we make a number of assumptions. First, we assume that the amount of the repressor protein present greatly outnumbers the amount of the output promoter, such that the promoter does not titrate away a significant amount of the free repressor. That is,

$$R \gg (P_O + B_O) \tag{3.10}$$

and therefore,

$$R \approx R + B_O \quad . \tag{3.11}$$

This assumption is reasonable because the concentration of the genome or plasmids (and therefore promoters) in the cell should be relatively small: on the order of $1 - 100$nM. Next, we assume that the repressor binding and unbinding from the promoter occurs more rapidly than protein production and degradation, and therefore exists at a quasi-steady-state.

With these assumptions, Equation 3.4 can be simplified and combined with Equations 3.5-3.7 and 3.9, resulting in

$$\dot{R} = t_R M_R - d_R R \tag{3.12}$$

$$\dot{M_O} = c\left(p_{Omin} + (p_{Omax} - p_{Omin})\left[\frac{K_d^n}{K_d^n + R^n}\right]\right) - d_{M_O} M_O \tag{3.13}$$

where

$$K_d^n = k_R/k_B \quad . \tag{3.14}$$

Using equations 3.3, 3.8, 3.12, and 3.13, we can solve for the steady state, yielding

$$M_R = c\left[\frac{p_R}{d_{M_R}}\right] \tag{3.15}$$

$$R = \frac{t_R M_R}{d_R} = c\left[\frac{t_R p_R}{d_R d_{M_R}}\right] \tag{3.16}$$

$$\begin{aligned} M_O &= \frac{c}{d_{M_O}}\left(p_{Omin} + \left[\frac{(p_{Omax} - p_{Omin})K_d^n}{K_d^n + R^n}\right]\right) \\ &= \frac{c}{d_{M_O}}\left(p_{Omin} + \left[\frac{(p_{Omax} - p_{Omin})K_d^n}{K_d^n + \left(\frac{ct_R p_R}{d_R d_{M_R}}\right)^n}\right]\right) \end{aligned} \tag{3.17}$$

$$O = \frac{t_O M_O}{d_O} = \left[\frac{ct_O}{d_O d_{M_O}}\right]\left(p_{Omin} + \left[\frac{(p_{Omax} - p_{Omin})K_d^n}{K_d^n + \left(\frac{ct_R p_R}{d_R d_{M_R}}\right)^n}\right]\right) \quad . \tag{3.18}$$

This set of equations represents the full model of the iFFL system when repression is modeled as a Hill function.

To further simplify the model, the Hill function is reduced to a power law. This simplification can occur when there is an excess of the repressor present

$$R^n \gg K_d^n \quad , \tag{3.19}$$

such that

$$R^n + K_d^n \approx R^n \quad , \tag{3.20}$$

and the output of the repressible promoter has a negligible amount of leak when fully saturated by the repressor, ie.

$$p_{Omin} \approx 0 \quad . \tag{3.21}$$

Note that this second assumption can also be satisfied if a defined range of copy numbers is being considered, and the promoter leak is minimal compared to the output of the partially-repressed promoter at those copy numbers, that is:

$$p_{Omin} \ll p_{Omax}\left[\frac{K_d^n}{K_d^n + R^n}\right] \quad , \tag{3.22}$$

for the maximum amount of repressor present at the considered copy numbers.
In this case, equation 3.13 can be further simplified to

$$\dot{M_O} = c\left[\frac{p_O}{R^n}\right] - d_{M_O} M_O \quad , \tag{3.23}$$

where

$$p_O = p_{Omax} K_d^n \quad . \tag{3.24}$$

Equations 3.17 and 3.18 can be simplified to

$$M_O = \frac{cp_O}{R^n d_{M_O}} = \left(\frac{cp_O}{d_{M_O}}\right)\left(\frac{d_R d_{M_R}}{ct_R p_R}\right)^n \tag{3.25}$$

$$O = \frac{t_O M_O}{d_O} = \left(\frac{ct_O p_O}{d_O d_{M_O}}\right)\left(\frac{d_R d_{M_R}}{ct_R p_R}\right)^n \quad . \tag{3.26}$$

Finally, rearranging Equation 3.26 gives

$$O = \left(\frac{c}{c^n}\right)\left(\frac{p_O t_O d_R^n d_{M_R}^n}{p_R^n t_R^n d_O d_{M_O}}\right) \quad . \tag{3.27}$$

Hence, when the only variable in the system is copy number (assuming all other parameters are constant),

$$O \propto \frac{c}{c^n} \propto c^{1-n} \quad . \tag{3.28}$$

Therefore, when repression is completely noncooperative (that is, $n = 1$), the concentration of the output protein will be entirely decoupled from the copy number of the iFFL. This simplified power law model for repression and the iFFL is shown in Figures 1c-d.

However, since the assumptions required for the power law model cannot be strictly met in a biological system, the actual function of an iFFL stabilized promoter will deviate the ideal case of noncooperative power law repression.

To model this deviation, we use a noncooperative simplified version of the Hill function model, assuming that $n = 1$ and there is no leaky expression from repressed promoters ($P_{Omax} \gg P_{Omin}$). These conditions can be nearly obtained in biological systems since nooncooperative repressors should have $n = 1$, and there

are repression systems that can achieve $\left(\frac{p_{Omax}}{p_{Omin}}\right) > 10^3$. With these assumptions, Equation 3.18 can be simplified to

$$O = \left(\frac{t_O p_{Omax}}{d_O d_{M_O}}\right)\left(\frac{c}{1+\frac{R}{K_d}}\right) = \left(\frac{t_O p_{Omax}}{d_O d_{M_O}}\right)\left(\frac{c}{1+c\left[\frac{t_R p_R}{d_R d_{M_R} K_d}\right]}\right) \quad . \tag{3.29}$$

For clarity we will use the parameters:

$$\alpha = \frac{t_O p_{Omax}}{d_O d_{M_O}} \tag{3.30}$$

$$\beta = \frac{t_R p_R}{d_R d_{M_R} K_d} \quad . \tag{3.31}$$

α encompasses the constants relating to transcription, translation, and degradation of the output mRNA and protein. It scales the steady state expression of the output. β encompasses the constants relating to transcription, translation, and degradation of the repressor mRNA and protein, scaled by the K_d of the repressor binding to the output promoter. It scales the concentration of the repressor relative to its K_d.

We can now calculate the relative change in output expression as copy number changes. Since $O \to 0$ as $c \to 0$, it is necessary to pick a minimum copy number at which the stabilized promoter can be used. The ability of a stabilized promoter to cancel changes in copy number can be measured as how the output changes as copy number is increased from the minimum chosen value. Specifically, the stabilization error E is calculated as the relative increase in output expression as the copy number increases from c_{min} to ∞,

$$E = \frac{\lim_{c\to\infty} O(\alpha,\beta)}{O_{c=c_{min}}(\alpha,\beta)} - 1 = \frac{\lim_{c\to\infty}\left(\frac{\alpha c}{1+\beta c}\right)}{\left(\frac{\alpha c_{min}}{1+\beta c_{min}}\right)} - 1 = \frac{\left(\frac{\alpha}{\beta}\right)}{\left(\frac{\alpha c_{min}}{1+\beta c_{min}}\right)} - 1 = \frac{\alpha+\alpha\beta c_{min}}{\alpha\beta c_{min}} - 1 = \frac{1}{\beta c_{min}} \quad . \tag{3.32}$$

Hence, stabilization of output expression improves as βc_{min}, and therefore the concentration of the repressor relative to its K_d at the lowest copy number increases. However, there is a trade off as increasing the amount of repressor expression will decrease output strength. This can be analyzed by calculating the output strength at the lowest copy number relative to an unrepressed promoter at the lowest copy number S,

$$S = \frac{O_{c=c_{min}}(\alpha,\beta)}{O_{c=c_{min}}(\alpha,0)} - 1 = \frac{\left(\frac{\alpha c_{min}}{1+\beta c_{min}}\right)}{\alpha c_{min}} = \frac{1}{1+\beta c_{min}} \quad . \tag{3.33}$$

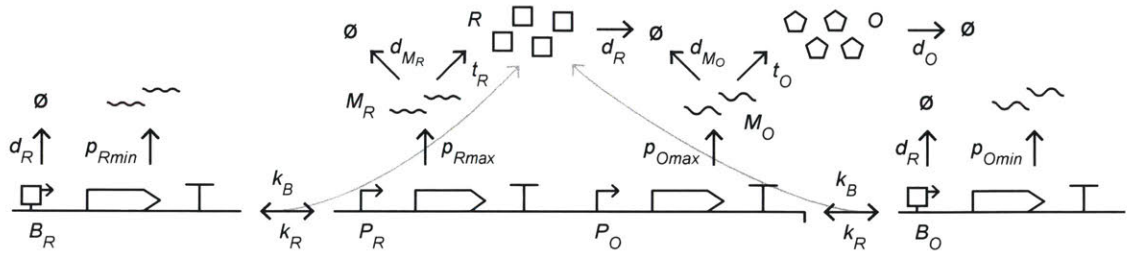
Hence, the amount of repressor produced at the lowest copy number relative to its K_d mediates both the output strength and stabilization quality of an iFFL stabilized promoter. Increasing this parameter, either by increasing the amount of repressor produced (for example, by increasing its transcription or translation) or

by decreasing K_d, will lead to a weaker stabilized promoter that stabilizes output expression more effectively and can function at lower copy numbers.

This tradeoff is shown in Figure 1e. Specifically, the relationship between S and E is shown,

$$S = \frac{1}{1+\beta c_{min}} = \frac{1}{1+\left(\frac{1}{E}\right)} = \frac{E}{E+1} \quad . \tag{3.34}$$

3.6.2 Comparison to feedback stabilization



A feedback regulatory system where one promoter drives repressor production and one drives the output(s) can be modeled with the following set of reactions. Note that it is assumed that the repressor interacts with both promoters identically, though the promoters may have different strengths.

$P_R \rightarrow P_R + M_R$	,	p_{Rmax}	Production of repressor mRNA from an unrepressed promoter
$B_R \rightarrow B_R + M_R$	,	p_{Rmin}	Leak of repressor mRNA from a repressed promoter
$nR + P_R \rightarrow B_R$	,	k_B	Binding of the repressor to the repressor promoter
$B_R \rightarrow nR + P_R$	,	k_R	unbinding of the repressor from the repressor promoter
$B_R \rightarrow P_R + (n-1)R$	,	d_R	Degradation of the repressor bound to the promoter
$M_R \rightarrow \varnothing$	,	d_{M_R}	Degradation of the mRNA for the repressor
$M_R \rightarrow M_R + R$	,	t_R	Translation of the repressor
$R \rightarrow \varnothing$	,	d_R	Degradation of repressor
$P_O \rightarrow P_O + M_O$	,	p_{Omax}	Production of output(s) mRNA from an unrepressed promoter
$B_O \rightarrow B_O + M_O$	,	p_{Omin}	Leak of output(s) mRNA from a repressed promoter
$nR + P_O \rightarrow B_O$	,	k_B	Binding of the repressor to the output promoter
$B_O \rightarrow nR + P_O$	,	k_R	unbinding of the repressor from the output promoter
$B_O \rightarrow P_O + (n-1)R$	,	d_R	Degradation of the repressor bound to the promoter
$M_O \rightarrow \varnothing$	,	d_{M_O}	Degradation of the mRNA for the output(s)
$M_O \rightarrow M_O + O$	,	t_O	Translation of the output(s)
$O \rightarrow \varnothing$	,	d_O	Degradation of output(s)

and:

$P_R + B_R = c$	Conservation of repressor promoters
$P_O + B_O = c$	Conservation of output promoters

Where:

- M_R, M_O are the mRNAs coding for the repressor and output proteins,
- R, O are the the repressor and output proteins,
- P_R, P_O are the promoters driving production of the repressor and output,
- B_R, B_O are the repressor or output promoter bound to the repressor protein,
- d_{M_R}, d_{M_O} are the degradation rates of the repressor and output mRNAs,
- t_R, t_O represent the strength of the RBSs driving the repressor and output(s),
- p_{Rmax}, p_{Omax} represent the strength of the unrepressed repressor and output promoters,
- p_{Rmin}, p_{Omin} represents leaky production from the repressed repressor and output promoters,
- k_B is the rate of binding of the repressor to the promoters,
- k_R is the rate of unbinding of the repressor from the promoters,
- n is the cooperativity of repressor binding,
- c is the copy number of the system

Combining these reactions into a system of equations yields

$$\dot{P}_R = k_R B_R - k_B R^n P_R + d_R B_R \tag{3.35}$$

$$\dot{B}_R = k_B R^n P_R - k_R B_R - d_R B_R \tag{3.36}$$

$$\dot{M}_R = p_{Rmax} P_R + p_{Rmin} B_R - d_{M_R} M_R \tag{3.37}$$

$$\dot{R} = t_R M_R - d_R R - k_B R^n (P_R + P_O) + n k_R (B_R + B_O) + (n-1) d_R (B_R + B_O) \tag{3.38}$$

$$\dot{P}_O = k_R B_O - k_B R^n P_O + d_R B_O \tag{3.39}$$

$$\dot{B}_O = k_B R^n P_O - k_R B_O - d_R B_O \tag{3.40}$$

$$\dot{M}_O = p_{Omax} P_O + p_{Omin} B_O - d_{M_O} M_O \tag{3.41}$$

$$\dot{O} = t_O M_O - d_O O \tag{3.42}$$

$$c = P_R + B_R \tag{3.43}$$

$$c = P_O + B_O \tag{3.44}$$

Using the same assumptions as with the iFFL model, we can solve for the steady state simplified power-

law form:

$$M_R = \left(\frac{cp_R}{d_{M_R}}\right) R^{-n} = \left(\frac{cp_R}{d_{M_R}}\right)^{\left(\frac{1}{1+n}\right)} \left(\frac{t_R}{d_R}\right)^{\left(\frac{-n}{1+n}\right)} \tag{3.45}$$

$$R = \frac{t_R M_R}{d_R} = \left(\frac{ct_R p_R}{d_R d_{M_R}}\right) R^{-n} = \left(\frac{cp_R t_R}{d_{M_R} d_R}\right)^{\left(\frac{1}{1+n}\right)} \tag{3.46}$$

$$M_O = \left(\frac{cp_O}{d_{M_O}}\right) R^{-n} = c^{\left(\frac{1}{1+n}\right)} \left(\frac{p_O}{d_{M_O}}\right) \left(\frac{p_R t_R}{d_{M_R} d_R}\right)^{\left(\frac{-n}{1+n}\right)} \tag{3.47}$$

$$O = \frac{t_O M_O}{d_O} = c^{\left(\frac{1}{1+n}\right)} \left(\frac{t_O p_O}{d_O d_{M_O}}\right) \left(\frac{p_R t_R}{d_{M_R} d_R}\right)^{\left(\frac{-n}{1+n}\right)} \tag{3.48}$$

with:

$$K_d^n = k_R/k_B$$

$$p_R = p_{Rmax} K_d^n$$

$$p_O = p_{Omax} K_d^n$$

Hence, when the only variable in the system is copy number (assuming all other parameters are constant),

$$O \propto c^{\left(\frac{1}{1+n}\right)} \quad . \tag{3.49}$$

Therefore, a feedback loop achieves the best stabilization when the cooperativity of the repressor is as high as possible. There are many cooperative repressors to use in bacteria, but the most cooperative examples that we are aware of top out at $n \approx 4$. Hence, a feedback loop could at best obtain $O \propto c^{0.2}$. While this is not a unacceptable amount of stabilization (a 10-fold change in copy number would be a 60% change in output expression), it is inferior to what is theoretically possible with an iFFL.

3.7 Figures

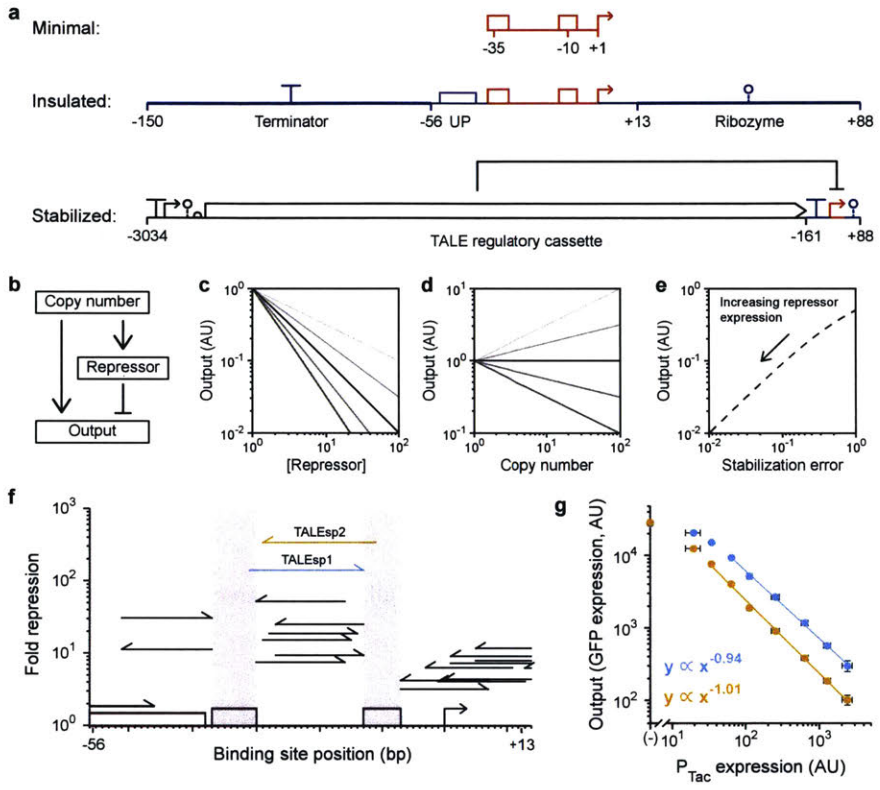


Figure 3-1: **Stabilized promoter design.**
a. The stabilized promoter is shown along with nested promoters (colors).
b. The schematic of the iFFL is shown.
c. The predictions from a mathematical model are shown (See 3.6.1). The theoretical power law response functions for the repression of the promoter are colored by the cooperativity (n = 1 is black and from top to bottom, n = 0.5, 0.75, 1.25, and 1.5).
d. The predicted stabilization by the response functions from part c.
e. The predicted trade-off between stabilization error and output strength (Equation 3.34).
f. The ability of TALEs to repress transcription from bacterial promoters based on the position of their binding site. Lines indicate the location of the TALE binding site and arrows indicate its orientation. Black lines: TALE1 binding sites, Blue line: TALEsp1 binding site, Orange line: TALEsp2 binding site.
g. The response function for the repression of the output promoter by TALEsp1 (blue) and TALEsp2 (orange). The TALEs are expressed from the IPTG-inducible P_{Tac} (left to right, 0, 5, 10, 16, 25, 40, 63, and 100 μM) that was characterized separately in parallel using a fluorescent reporter to generate the x-axis values (Figure 3-7). On the x-axis, (−) indicates a control with no TALE. For all data, the average of three biological replicates from different days is shown, and error bars indicate the standard deviation.

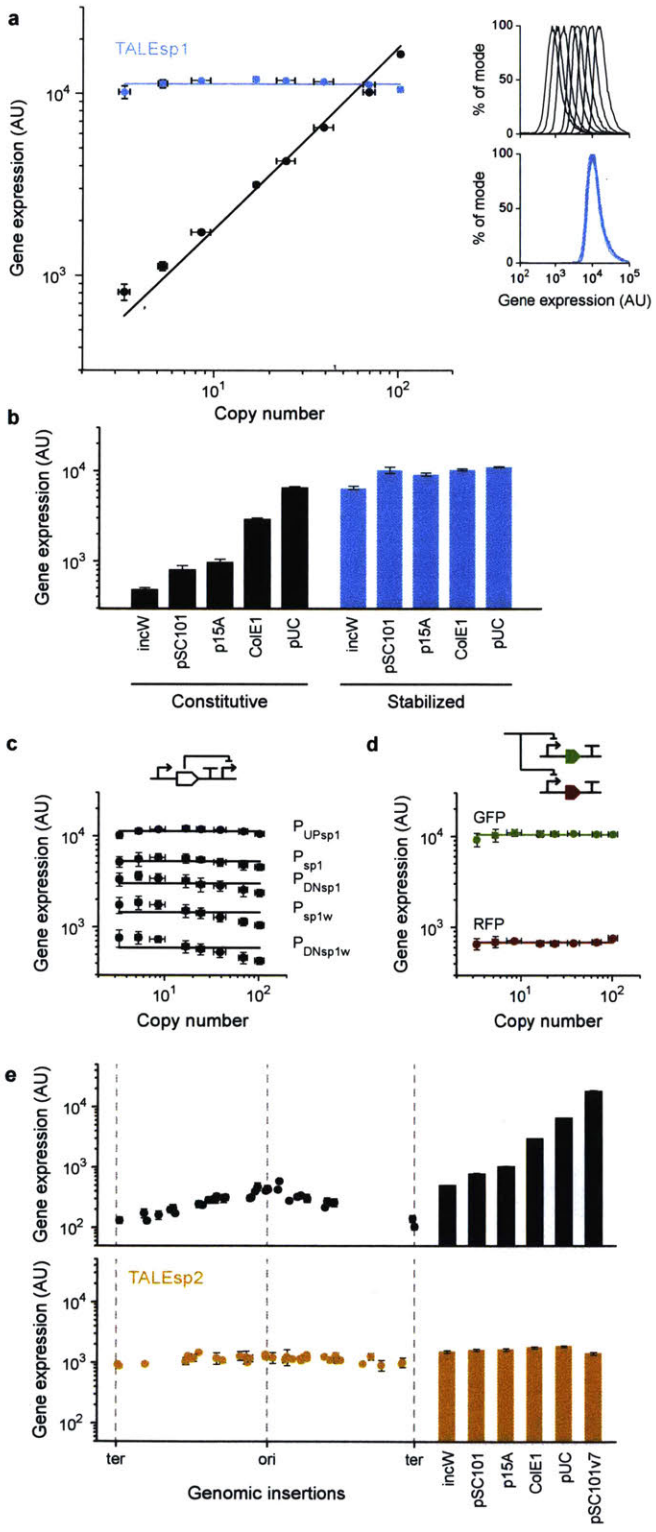


Figure 3-2: **Stabilized promoters compensate for changes in copy number.** (Caption on next page)

Figure 3-2:
a. Performance of an insulated and stabilized promoter on a set of pSC101 plasmid variants with different copy numbers. The x-axis shows the copy number of each pSC101 variant backbone measured with qPCR (Methods). Black points are the data for the insulated promoter and blue points correspond to the TALEsp1 stabilized promoter. Cytometry distributions corresponding to these data points are shown to the right. The black line is a linear fit to the insulated promoter data and the blue line shows the geometric mean of the stabilized promoter data.
b. Performance of the TALEsp1 stabilized promoter across different plasmid backbones.
c. Each line is a variant of a stabilized promoter with the output promoter set for a different expression level. The TALEsp1 system and pSC101 library are used.
d. One TALE (sp1) is used to stabilize two promoters of different strength across the pSC101 backbones. Green: GFP expression levels driven by the P_{UPsp1} promoter, red: RFP expression levels driven by the P_{DNsp1} promoter.
e. Performance of the insulated constitutive promoter (black) and TALEsp2 stabilized promoter (orange) are compared when integrated into the genome and on different plasmid backbones. For all data, the average of three biological replicates from different days is shown, and error bars indicate the standard deviation.

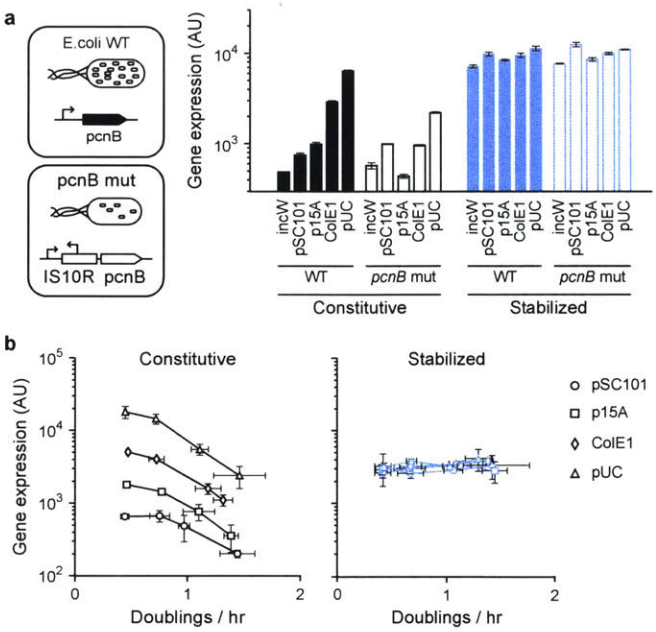


Figure 3-3: **Stabilized promoters reduce the impact of genome mutations and growth media.**

a. The performance of the TALEsp1 promoter across plasmid backbones is compared when in wild-type *E. coli* DH10B and a mutant strain. The schematic shows the reduction in copy number that occurs when an insertion element (IS10R) disrupts this gene.

b. Performance of a stabilized promoter when *E. coli* DH10B is grown in varying media compositions. Each line is a different plasmid backbone. From left to right, the media compositions are: 1. M9 with 0.4% glycerol and 0.5 mM Leucine, 2. M9 with 0.4% glucose and 0.5 mM Leucine, 3. M9 with 0.4% glycerol and 0.2% casamino acids, and 4. M9 with 0.4% Glucose and 0.2% casamino acids. The doubling time of *E. coli* in each media was determined, and gene expression measured in exponential growth (Methods). For all data, the average of three biological replicates from different days is shown, and error bars indicate the standard deviation.

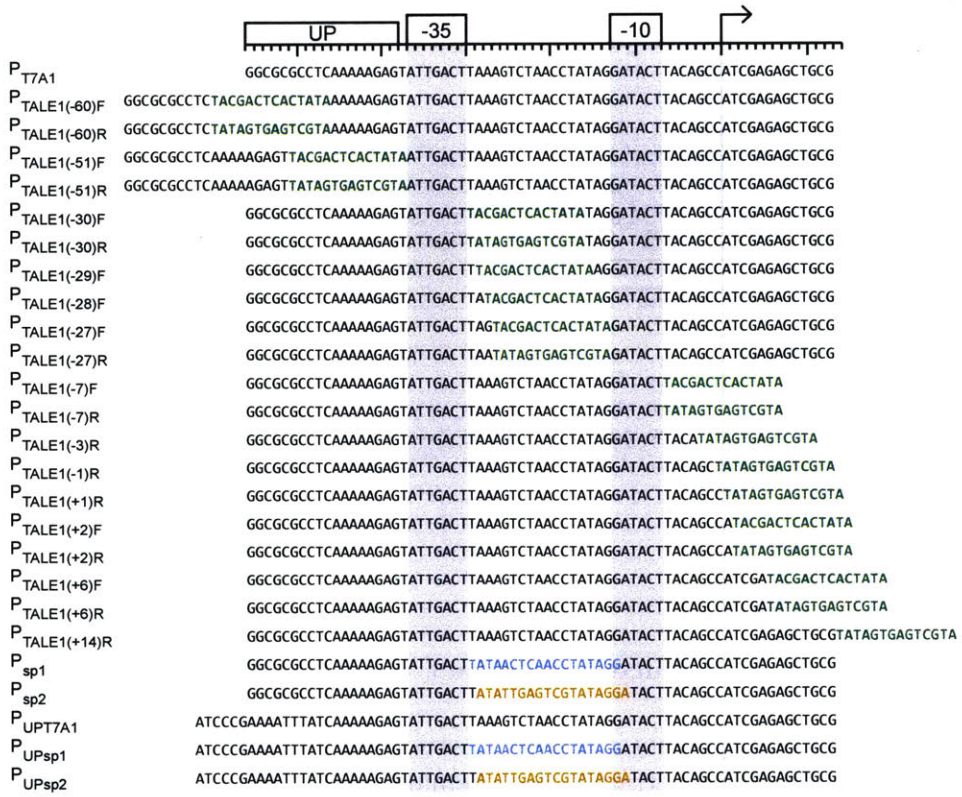


Figure 3-4: **Promoter sequences for TALE repressible promoters.** Promoter sequences of the TALE repressible promoters are shown along with the constitutive promoters from which they were constructed. $P_{TALE1(-60)F}$ through P_{sp2} were built from promoter P_{T7A1} and are used in Figure 3-1f. P_{UPsp1} and P_{UPsp2} are improved versions of P_{sp1} and P_{sp2} that include more of the native P_{T7A1} UP element and are slightly stronger. They were used in Figure 3-1g and in the strongest stabilized promoters. Green letters show TALE1 binding sites, blue letters show TALEsp1 binding sites, and orange letters show TALEsp2 binding sites.

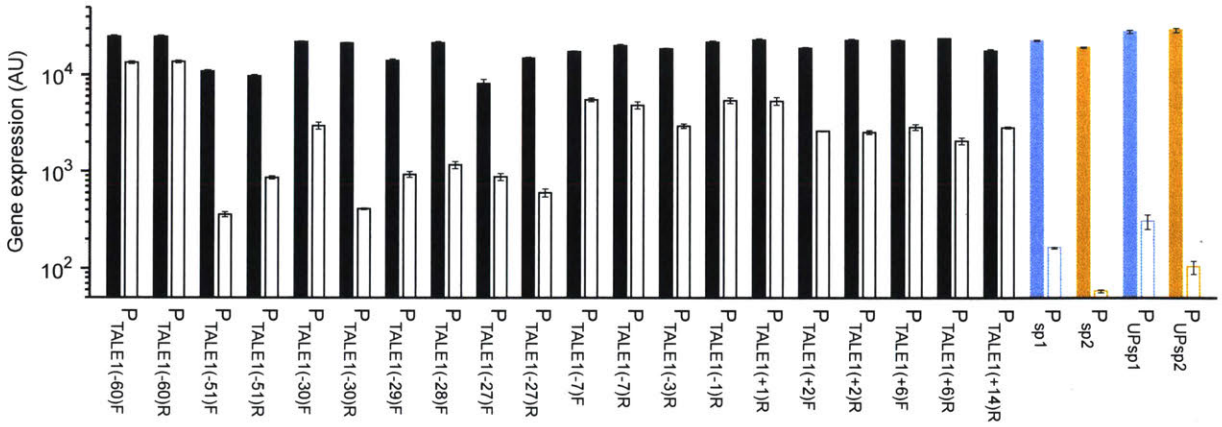


Figure 3-5: **Expression levels for TALE repressible promoters.** The fold-change values shown in Figure 3-1f were calculated by measuring gene expression from TALE repressible promoters in the presence and absence of the TALE. Black bars show TALE1 repression, blue bars show TALEsp1 repression, and orange bars show TALEsp2 repression. P_{UPsp1} and P_{UPsp2} show the extreme values from Figure 3-1g for comparison. Solid bars show gene expression in the absence of the TALE, and hollow bars show gene expression when the TALE is present and induced with 100 μM IPTG from a P_{Tac} promoter. The average of three biological replicates from different days is shown, and error bars indicate the standard deviation.

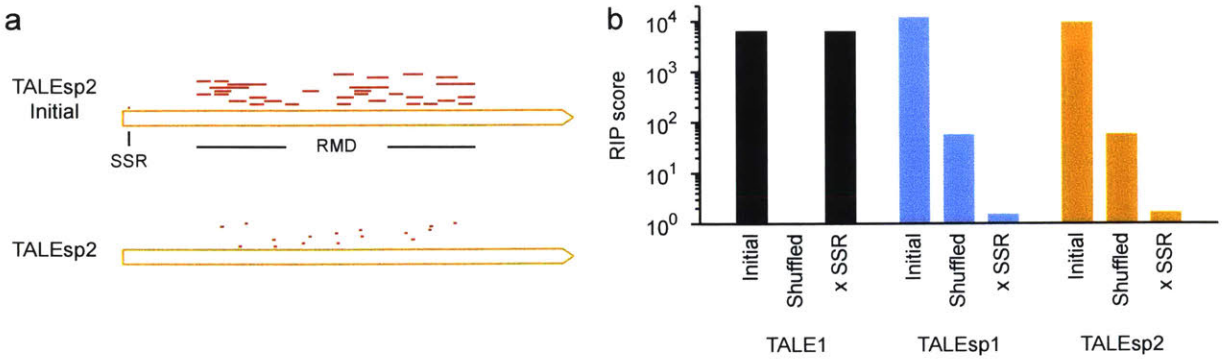


Figure 3-6: **Predicted evolutionary stability of TALE genes.** The TALE repressors were initially built using TypeIIs assembly methods, which leads to large repeated sections of DNA sequence. This can make them unsuitable for use in bacteria due to high rates of homologous recombination. The Evolutionary Failure Mode (EFM) calculator [282] predicted that the initial TALE sequences would lead to large evolutionary instabilities in bacteria, due to repeat mediated deletions (RMD) and a simple sequence repeat (SSR). We modified the TALE repressors by changing codons in the repeated sections and removing the SSR to make them more evolutionarily stable. a) The instabilities identified by the EFM calculator are shown for the initial and final versions of the TALEsp2 repressor. Only the worst 10 RMDs are shown. b) The 'relative instability prediction' (RIP) score calculated by the EFM calculator is shown for the initial and improved versions of the TALE repressors. This score indicates how much more likely a DNA sequence is to mutate compared to unavoidable random point mutations, with a value of 1 indicating no added mutational propensity. The TALEsp1 and TALEsp2 repressors were re-synthesized with different codons to remove repeats and the SSR was removed. The TALE1 repressor was not re-synthesized, but had the SSR removed so that its N-terminus matches TALEsp1 and TALEsp2.

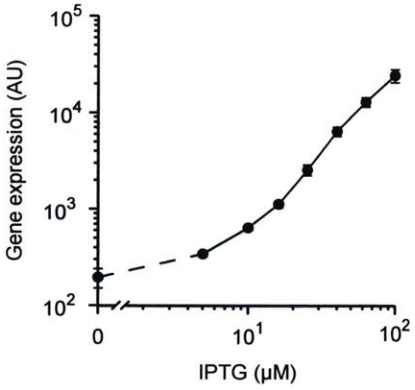


Figure 3-7: **Gene expression from the P_{Tac} promoter.** GFP expression from the P_{Tac} promoter used to induce TALE expression is shown. These values were used on the x-axis of Figure 3-1g to indicate how much of the TALEs was being expressed. Due to the use of RiboJ between the promoter and RBS the GFP expression level should be proportional to the TALE expression level. This data was collected at the same time as the data shown in Figure 3-1. The average of three biological replicates from different days is shown, and error bars indicate the standard deviation.

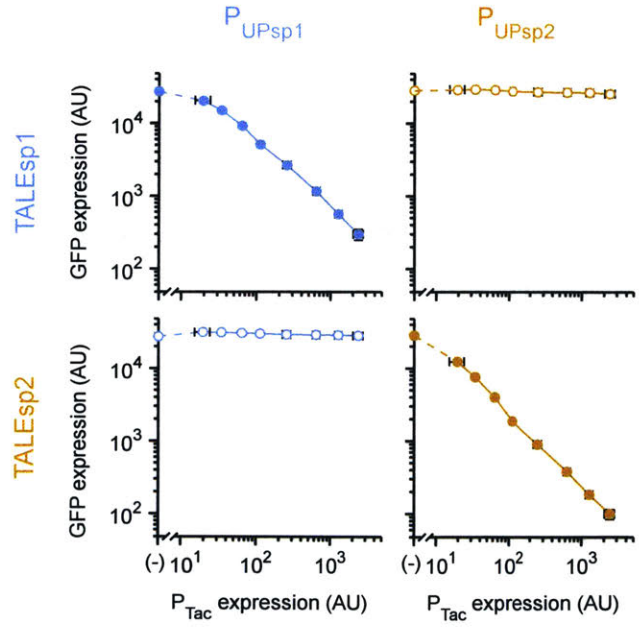


Figure 3-8: **Orthogonality of TALE repressors.** The ability of TALEsp1 and TALEsp2 to repress the P_{UPsp1} and P_{UPsp2} promoters is shown. Data was collected in the same way as in Figure 3-1g, and the on-target data is the same as what is shown in that figure. The average of three biological replicates from different days is shown, and error bars indicate the standard deviation.

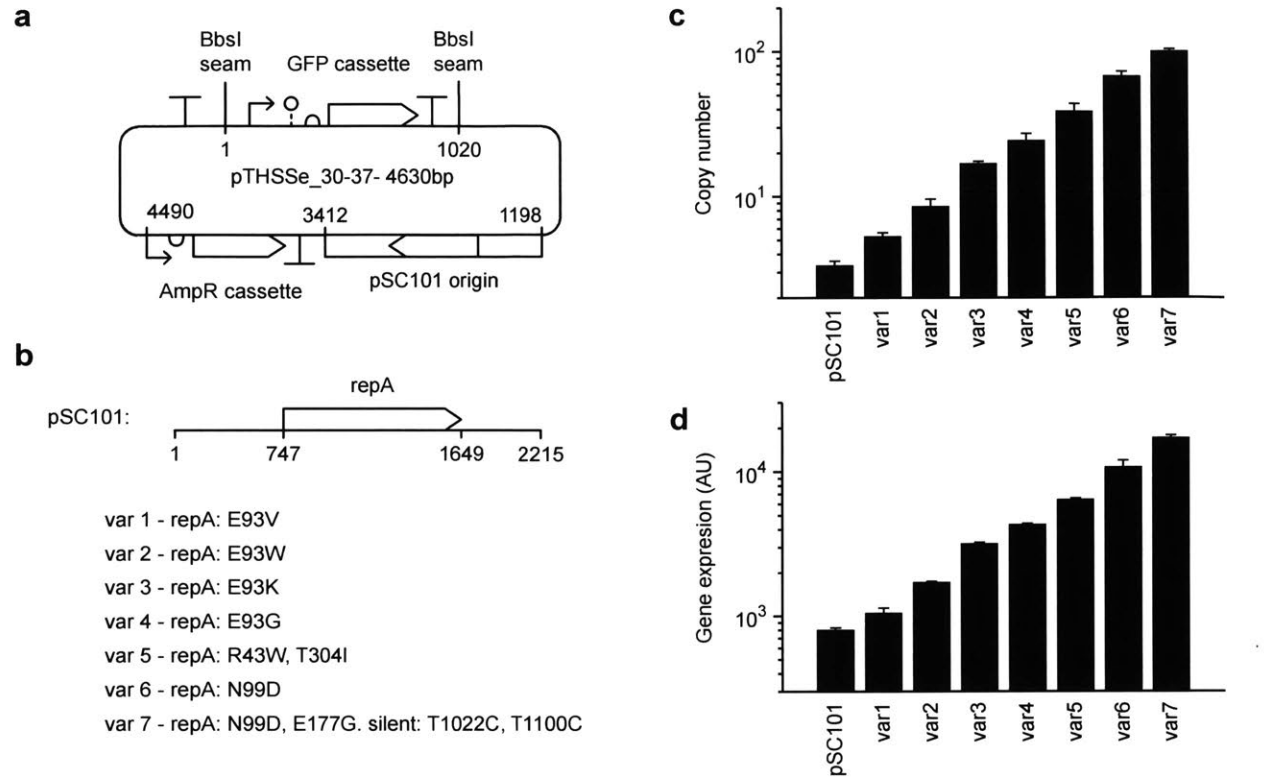


Figure 3-9: **pSC101 backbones with varying copy number.** a) Diagram of pSC101 plasmids used to characterize promoter function at different copy numbers. Each backbone contains a constitutive GFP expression cassette that can be replaced by TypeIIs BbsI cloning. The insulated and stabilized promoters were cloned into these sites for testing. b) A list of pSC101 origin variants that were used in this work. Var 1-4 were constructed by targeted mutagenesis of the E93 codon in the repA gene following a previous study [283]. Var 5-7 were constructed through error-prone PCR of the repA gene. c) Copy number of the pSC101 plasmids measured by qPCR in exponential growth. This data was used as the x-axis on Figure 3-2d) GFP expression from the constitutive GFP cassette on each plasmid backbone. For all data, the average of three biological replicates from different days is shown, and error bars indicate the standard deviation.

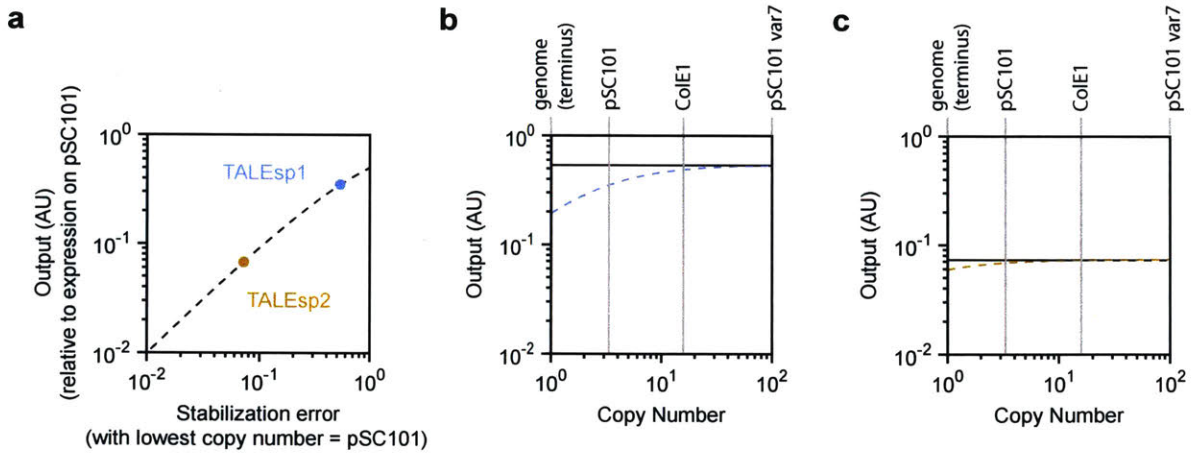


Figure 3-10: **Tuning stabilized promoters.** The TALEsp1 and TALEsp2 stabilized promoters were constructed with variable promoters and RBSs driving expression of the TALE repressors. Due to the tradeoff between stabilization and output strength, constructs were chosen that expressed the minimal level of the TALE repressors needed to provide good stabilization at the desired range of copy numbers. It is assumed that the TALEs exhibit optimal noncooperative Hill function repression for this analysis. a) The tradeoff between stabilization and output strength of the stabilized promoters characterized on pSC101. 'Stabilization error' shows the relative increase in output as copy number is raised from the level of pSC101, 'output' shows the relative level of output expression on pSC101 compared to an unrepressed promoter on pSC101. The TALEsp1 stabilized promoter has an ouput of 34% compared to the unrepressed P_{UPs1} promoter on pSC101, implying that it expresses approximately $1.9K_d$ of TALEsp1 on pSC101. This predicts an increase in gene expression of 50% as copy number increases from the level of pSC101.The TALEsp2 stabilized promoter has an ouput of 7% compared to the unrepressed P_{UPs2} promoter on pSC101, implying that it expresses approximately $14K_d$ of TALEsp2 on pSC101. This predicts an increase in gene expression of 7% as copy number increases from the level of pSC101. Note that the TALEsp2 promoter was tuned so that it can function at copy numbers below pSC101, with a predicted increase in output of 29% as copy number increases from the level of the genome. b) Predicted performance of the TALEsp1 stabilized promoter at different copy numbers (blue dashed line) compared to the ideal power law case (black line). c) Predicted performance of the TALEsp2 stabilized promoter at different copy numbers (orange dashed line) compared to the ideal power law case (black line).

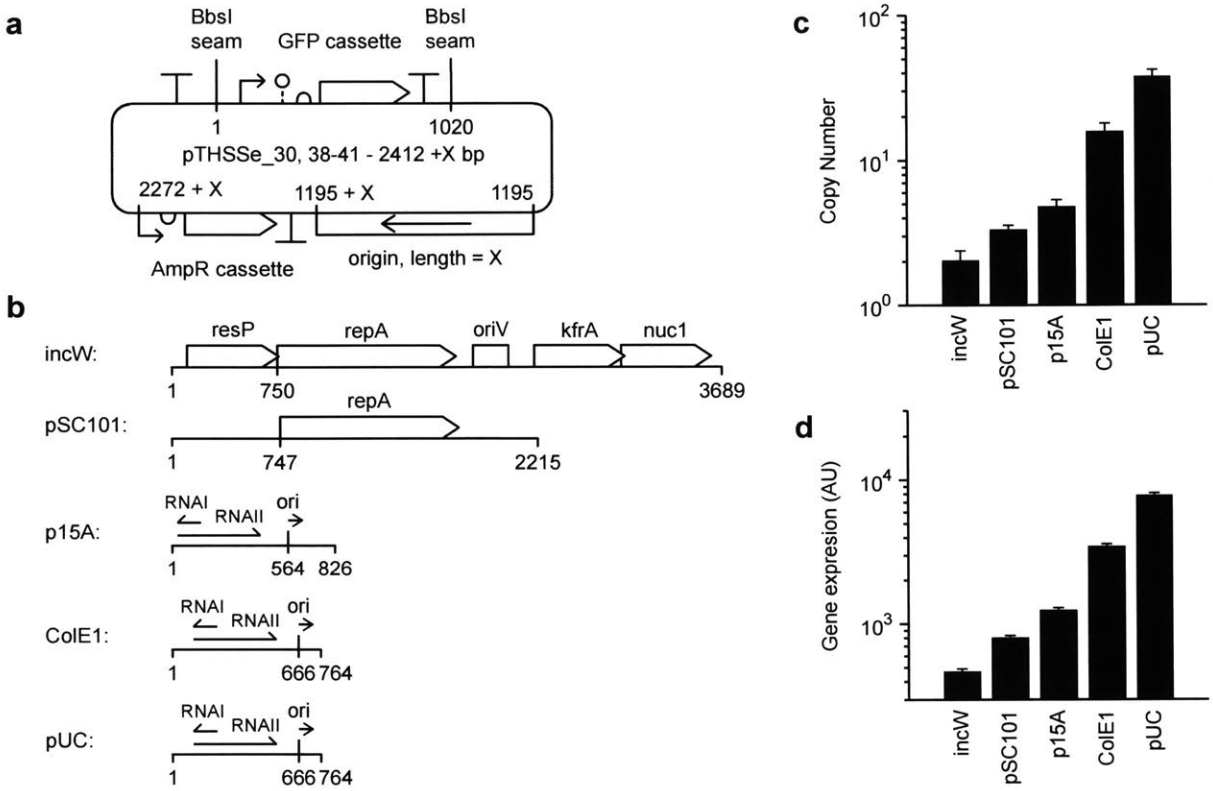


Figure 3-11: **Plasmid backbones based on different origins of replication.** a) Diagram of the set of common cloning vector backbones used in Figures 3-2b and 3-2e. Plasmid architecture is the same as for the pSC101 plasmids shown in Figure 3-9, but with different plasmid origins. The length of the origin is given as 'X'. b) Schematics of the origins used. incW and pSC101 are relatively long origins that produce proteins required for plasmid replication, p15A, ColE1, and pUC are shorter origins that use RNAs. c) Copy number of the plasmids measured by qPCR in exponential growth. d) GFP expression from the constitutive GFP cassette on each plasmid backbone. For all data, the average of three biological replicates from different days is shown, and error bars indicate the standard deviation.

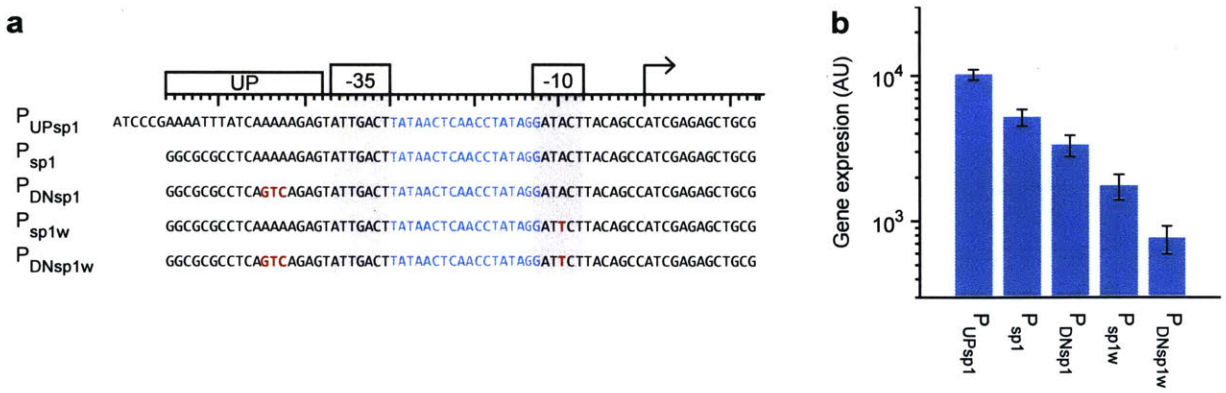


Figure 3-12: **Design of stabilized promoters with different strengths.** Variants of the TALEsp1 stabilized promoter with varying output gene expression were built by mutating the TALEsp1 repressible promoter. a) Promoter variants used. Light blue bases show the TALEsp1 binding site, red bases show mutations made to weaken promoter strength. b) Comparison of stabilized promoter output strengths. TALEsp1 stabilized promoters were constructed with each of the variant promoters and GFP expression on a pSC101 backbone is shown. This data is the same as the leftmost data points shown in Figure 3-2c. the average of three biological replicates from different days is shown, and error bars indicate the standard deviation.

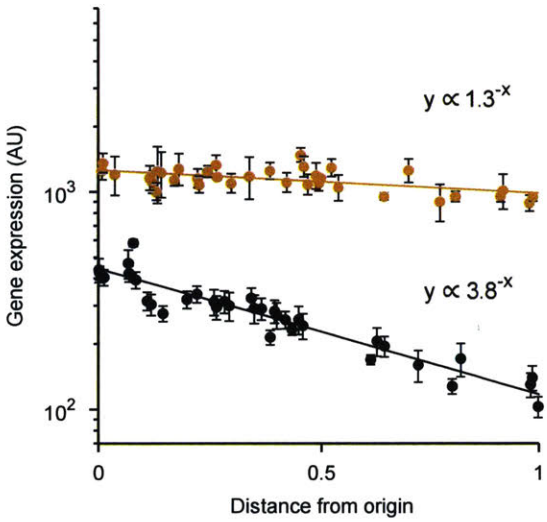


Figure 3-13: **Genome position dependent gene expression.** GFP expression data from identified insulated constitutive (black) and TALEsp2 stabilized (orange) genomic integrations is shown plotted against the distance from the chromosomal origin of replication. The average of three biological replicates from different days is shown, and error bars indicate the standard deviation. Exponential functions were fit to the data using a proportional least squares algorithm. The fit functions predict a 3.8-fold range in gene expression due to genome position for the constitutive integrations and a 1.3-fold range for the stabilized integrations.

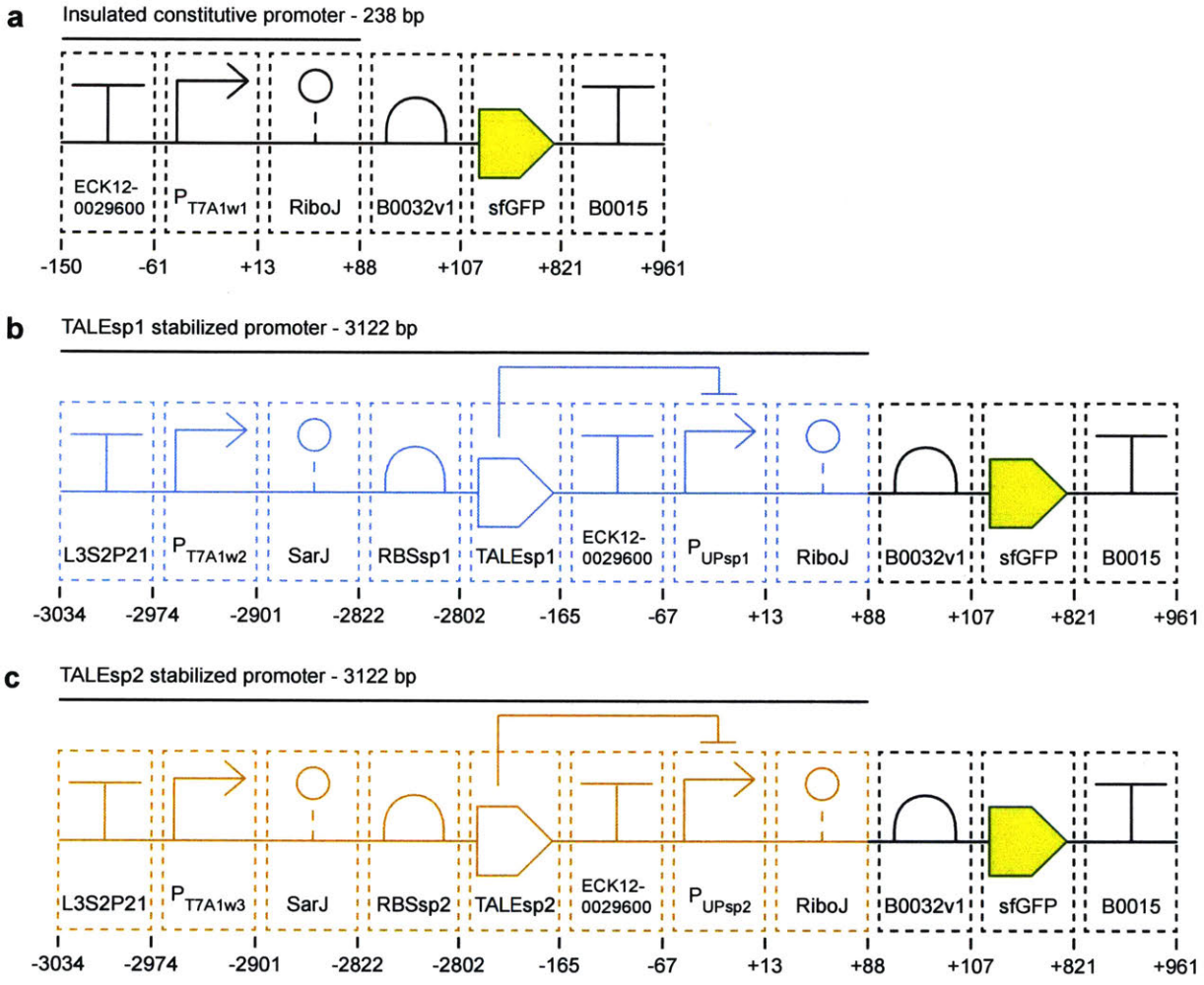


Figure 3-14: **Design of stabilized and insulated promoters.** Parts diagrams and lengths are shown for the insulated constitutive promoter (a), TALEsp1 stabilized promoter (b), and TALEsp2 stabilized promoter (c). These systems are depicted driving a GFP expression cassette, which was used for most experiments. All base pair numbers are relative to the transcription initiation site of the output promoter.

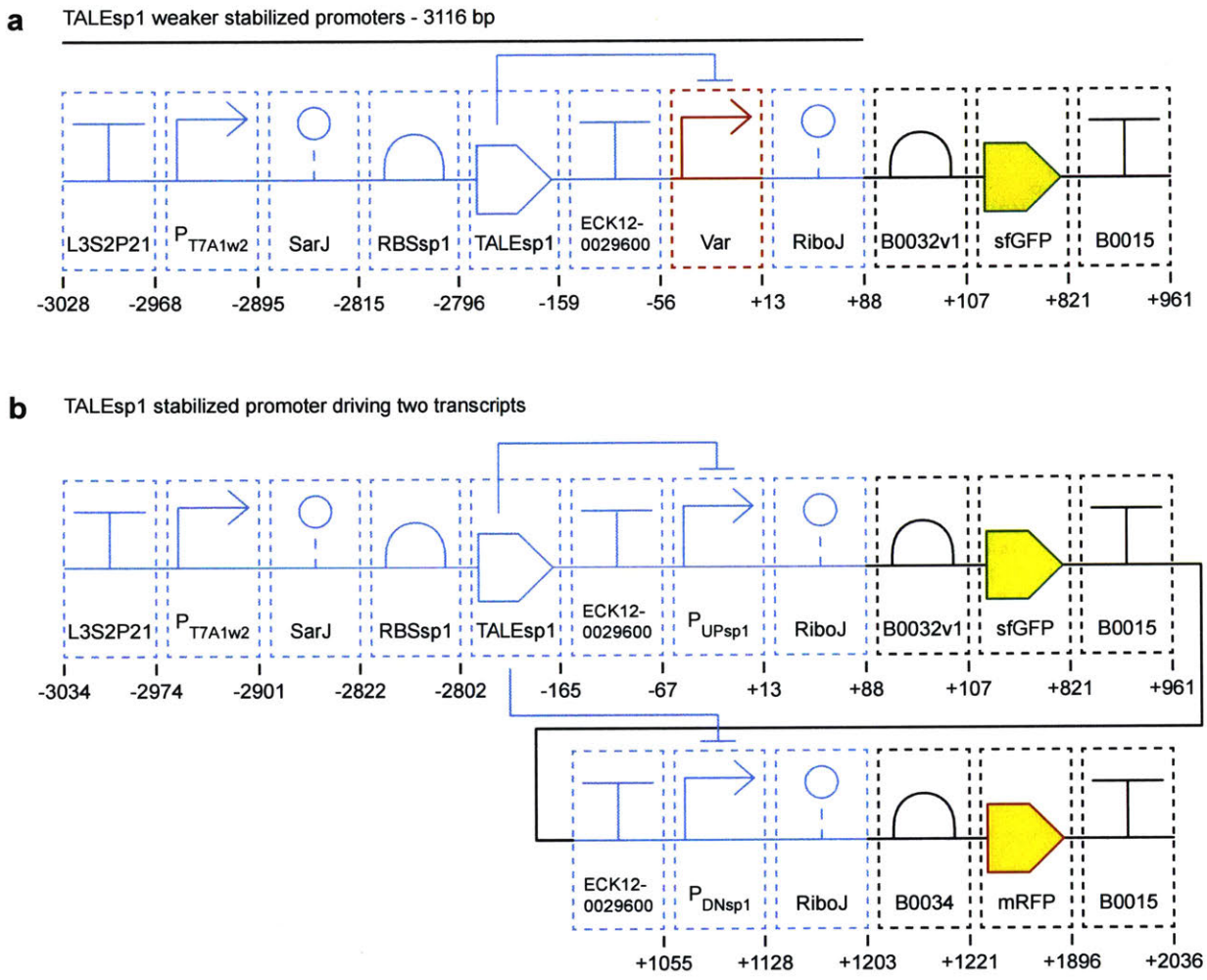


Figure 3-15: **Stabilized promoter variants.** Parts diagrams and lengths are shown for the weaker TALEsp1 stabilized promoter (a) and TALEsp1 stabilized promoter driving two transcripts (b). All base pair numbers are relative to the transcription initiation site of the first output promoter.

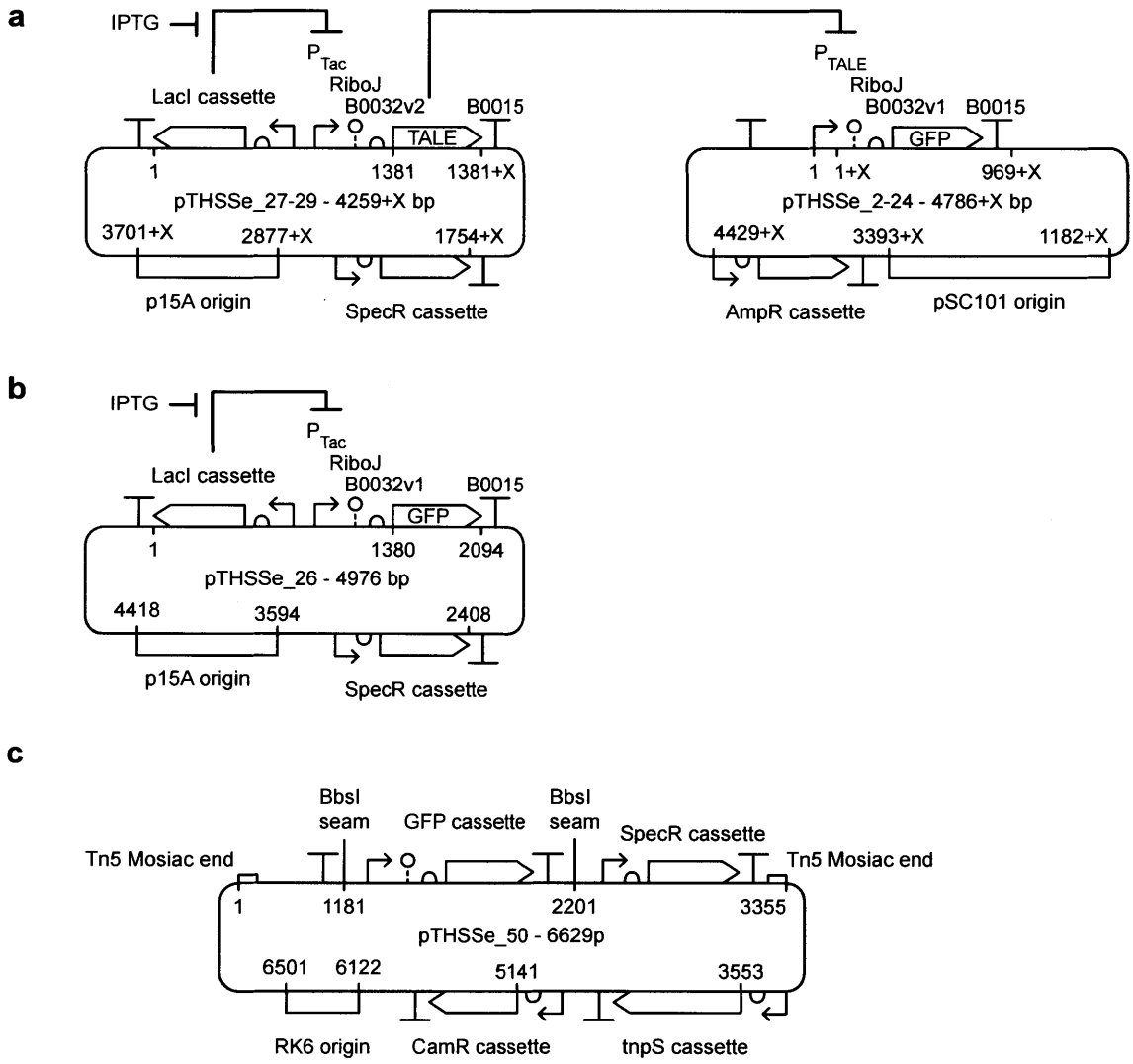


Figure 3-16: **Other plasmid diagrams.** a) Diagram of plasmids used to test TALE repression. A p15A, SpecR plasmid was used to express each TALE under the control of a P_{Tac} expression system. The TALEs vary in length, which is shown as 'X' in the plasmid diagram. A pSC101, AmpR plasmid was used to measure the output from the TALE-repressible promoters using GFP. The TALE-repressible promoters (P_{TALE}) vary in length, which is shown as 'X' in the diagram. b) Diagram of the plasmid used to measure output from the P_{Tac} promoter. This plasmid is similar to the TALE expression plasmids shown in (a), but contains GFP under the control of the inducible system. c) Diagram of the Tn5 transposition vector used to insert contitutive and stabilized promoters into the genome. The region between the Tn5 mosiac ends (bases1 - 3355) will randomLy insert into the genome. This plasmid is a modified form of pBAMD1-4 [286], with the resistance cassette changed to CamR, and the insert modified to allow type IIs cloning in a similar manner to the plasmids shown in Figures 3-9 and 3-11. An approximately 1kb-long stretch of DNA was inserted between the first Tn5 mosiac end and the BbsI cloning site to buffer genomic insertions from local context effects.

3.8 Tables

Table 3.1: New plasmids used in this work

Name[a]	Origin	Marker	Description
pTHSSe_01	pSC101	Amp	Empty AmpR pSC101 vector
pTHSSe_02	pSC101	Amp	$P_{TALE1(-60)F}$ GFP reporter
pTHSSe_03	pSC101	Amp	$P_{TALE1(-60)R}$ GFP reporter
pTHSSe_04	pSC101	Amp	$P_{TALE1(-51)F}$ GFP reporter
pTHSSe_05	pSC101	Amp	$P_{TALE1(-51)R}$ GFP reporter
pTHSSe_06	pSC101	Amp	$P_{TALE1(-30)F}$ GFP reporter
pTHSSe_07	pSC101	Amp	$P_{TALE1(-30)R}$ GFP reporter
pTHSSe_08	pSC101	Amp	$P_{TALE1(-29)F}$ GFP reporter
pTHSSe_09	pSC101	Amp	$P_{TALE1(-28)F}$ GFP reporter
pTHSSe_10	pSC101	Amp	$P_{TALE1(-27)F}$ GFP reporter
pTHSSe_11	pSC101	Amp	$P_{TALE1(-27)R}$ GFP reporter
pTHSSe_12	pSC101	Amp	$P_{TALE1(-7)F}$ GFP reporter
pTHSSe_13	pSC101	Amp	$P_{TALE1(-7)R}$ GFP reporter
pTHSSe_14	pSC101	Amp	$P_{TALE1(-3)R}$ GFP reporter
pTHSSe_15	pSC101	Amp	$P_{TALE1(-1)R}$ GFP reporter
pTHSSe_16	pSC101	Amp	$P_{TALE1(+1)R}$ GFP reporter
pTHSSe_17	pSC101	Amp	$P_{TALE1(+2)F}$ GFP reporter
pTHSSe_18	pSC101	Amp	$P_{TALE1(+2)R}$ GFP reporter
pTHSSe_19	pSC101	Amp	$P_{TALE1(+6)F}$ GFP reporter
pTHSSe_20	pSC101	Amp	$P_{TALE1(+6)R}$ GFP reporter
pTHSSe_21	pSC101	Amp	P_{sp1} GFP reporter
pTHSSe_22	pSC101	Amp	P_{sp2} GFP reporter
pTHSSe_23	pSC101	Amp	P_{UPsp1} GFP reporter
pTHSSe_24	pSC101	Amp	P_{UPsp2} GFP reporter
pTHSSe_25	p15A	Spec	Empty SpecR p15A vector
pTHSSe_26	p15A	Spec	P_{Tac} driving sfGFP
pTHSSe_27	p15A	Spec	P_{Tac} driving TALE1
pTHSSe_28	p15A	Spec	P_{Tac} driving TALEsp1
pTHSSe_29	p15A	Spec	P_{Tac} driving TALEsp2
pTHSSe_30	pSC101	Amp	pSC101 BbsI cloning vector
pTHSSe_31	pSC101*	Amp	pSC101 var1 BbsI cloning vector
pTHSSe_32	pSC101*	Amp	pSC101 var2 BbsI cloning vector
pTHSSe_33	pSC101*	Amp	pSC101 var3 BbsI cloning vector
pTHSSe_34	pSC101*	Amp	pSC101 var4 BbsI cloning vector
pTHSSe_35	pSC101*	Amp	pSC101 var5 BbsI cloning vector
pTHSSe_36	pSC101*	Amp	pSC101 var6 BbsI cloning vector
pTHSSe_37	pSC101*	Amp	pSC101 var7 BbsI cloning vector
pTHSSe_38	incW	Amp	incW BbsI cloning vector
pTHSSe_39	p15A	Amp	p15A BbsI cloning vector
pTHSSe_40	ColE1	Amp	ColE1 BbsI cloning vector
pTHSSe_41	pUC	Amp	pUC BbsI cloning vector

Table 3.1 continued

Name[a]	Origin	Marker	Description
pTHSSe_42	p15A	Cam	Insulated promoter driving GFP
pTHSSe_43	pSC101	Cam	TALEsp1 stabilized promoter driving GFP
pTHSSe_44	pSC101	Cam	TALEsp2 stabilized promoter driving GFP
pTHSSe_45	pSC101	Cam	TALEsp1 weak1 stabilized promoter driving GFP
pTHSSe_46	pSC101	Cam	TALEsp1 weak2 stabilized promoter driving GFP
pTHSSe_47	pSC101	Cam	TALEsp1 weak3 stabilized promoter driving GFP
pTHSSe_48	pSC101	Cam	TALEsp1 weak4 stabilized promoter driving GFP
pTHSSe_49	pSC101	Cam	TALEsp1 stabilized promoter driving GFP and RFP
pTHSSe_50	RK6	Cam / Spec	BbsI transposition vector
pTHSSe_51	RK6	Cam / Spec	Insulated, constituive control in transposition vector
pTHSSe_52	RK6	Cam / Spec	TALEsp2 stabilized promoter in transposition vector
pTHSSe_53	pSC101	Amp	pSC101 AmpR empty vector for transpositions
pTHSSe_54	ColE1	Amp	L3S2P21 terminator in GGAG-TACT BsaI assembly vector
pTHSSe_55	ColE1	Amp	TALEsp1 CDS in AATG-AGGT BsaI assembly vector
pTHSSe_56	ColE1	Amp	TALEsp2 CDS in AATG-AGGT BsaI assembly vector
pTHSSe_57	ColE1	Amp	TALEsp1 cassette in GGAT-GCTT BsaI assembly vector
pTHSSe_58	ColE1	Amp	TALEsp2 cassette in GGAT-GCTT BsaI assembly vector
pTHSSe_59	pSC101	Cam	pSC101 GGAG-CGCT BasI-BbsI cloning vector
pTHSSe_60	p15A	Cam	p15A GGAG-CGCT BasI-BbsI cloning vector

a. Plasmid numbers are preliminary. Please refer to the published manuscript for updated numbers.

Table 3.2: Plasmid sets used in this work

Name[a]	Origin	Marker	Description	Cloning[b]
pTHSSf_01	pSC101	Amp	Insulated promoter driving GFP in pSC101	44 -> 32
pTHSSf_02	pSC101*	Amp	Insulated promoter driving GFP in pSC101 var1	44 -> 33
pTHSSf_03	pSC101*	Amp	Insulated promoter driving GFP in pSC101 var2	44 -> 34
pTHSSf_04	pSC101*	Amp	Insulated promoter driving GFP in pSC101 var3	44 -> 35
pTHSSf_05	pSC101*	Amp	Insulated promoter driving GFP in pSC101 var4	44 -> 36
pTHSSf_06	pSC101*	Amp	Insulated promoter driving GFP in pSC101 var5	44 -> 37
pTHSSf_07	pSC101*	Amp	Insulated promoter driving GFP in pSC101 var6	44 -> 38
pTHSSf_08	pSC101*	Amp	Insulated promoter driving GFP in pSC101 var7	44 -> 39
pTHSSf_09	incW	Amp	Insulated promoter driving GFP in incW	44 -> 40
pTHSSf_10	p15A	Amp	Insulated promoter driving GFP in p15A	44 -> 41
pTHSSf_11	ColE1	Amp	Insulated promoter driving GFP in ColE1	44 -> 42
pTHSSf_12	pUC	Amp	Insulated promoter driving GFP in pUC	44 -> 43
pTHSSf_13	pSC101	Amp	TALEsp1 stabilized promoter driving GFP in pSC101	45 -> 32
pTHSSf_14	pSC101*	Amp	TALEsp1 stabilized promoter driving GFP in pSC101 var1	45 -> 33
pTHSSf_15	pSC101*	Amp	TALEsp1 stabilized promoter driving GFP in pSC101 var2	45 -> 34
pTHSSf_16	pSC101*	Amp	TALEsp1 stabilized promoter driving GFP in pSC101 var3	45 -> 35
pTHSSf_17	pSC101*	Amp	TALEsp1 stabilized promoter driving GFP in pSC101 var4	45 -> 36
pTHSSf_18	pSC101*	Amp	TALEsp1 stabilized promoter driving GFP in pSC101 var5	45 -> 37
pTHSSf_19	pSC101*	Amp	TALEsp1 stabilized promoter driving GFP in pSC101 var6	45 -> 38
pTHSSf_20	pSC101*	Amp	TALEsp1 stabilized promoter driving GFP in pSC101 var7	45 -> 39
pTHSSf_21	incW	Amp	TALEsp1 stabilized promoter driving GFP in incW	45 -> 40
pTHSSf_22	p15A	Amp	TALEsp1 stabilized promoter driving GFP in p15A	45 -> 41
pTHSSf_23	ColE1	Amp	TALEsp1 stabilized promoter driving GFP in ColE1	45 -> 42
pTHSSf_24	pUC	Amp	TALEsp1 stabilized promoter driving GFP in pUC	45 -> 43
pTHSSf_25	pSC101	Amp	TALEsp2 stabilized promoter driving GFP in pSC101	46 -> 32
pTHSSf_26	pSC101*	Amp	TALEsp2 stabilized promoter driving GFP in pSC101 var7	46 -> 39
pTHSSf_27	incW	Amp	TALEsp2 stabilized promoter driving GFP in incW	46 -> 40
pTHSSf_28	p15A	Amp	TALEsp2 stabilized promoter driving GFP in p15A	46 -> 41
pTHSSf_29	ColE1	Amp	TALEsp2 stabilized promoter driving GFP in ColE1	46 -> 42
pTHSSf_30	pUC	Amp	TALEsp2 stabilized promoter driving GFP in pUC	46 -> 43
pTHSSf_31	pSC101	Amp	TALEsp1 weak1 stabilized promoter driving GFP in pSC101	47 -> 32
pTHSSf_32	pSC101*	Amp	TALEsp1 weak1 stabilized promoter driving GFP in pSC101 var1	47 -> 33
pTHSSf_33	pSC101*	Amp	TALEsp1 weak1 stabilized promoter driving GFP in pSC101 var2	47 -> 34
pTHSSf_34	pSC101*	Amp	TALEsp1 weak1 stabilized promoter driving GFP in pSC101 var3	47 -> 35
pTHSSf_35	pSC101*	Amp	TALEsp1 weak1 stabilized promoter driving GFP in pSC101 var4	47 -> 36
pTHSSf_36	pSC101*	Amp	TALEsp1 weak1 stabilized promoter driving GFP in pSC101 var5	47 -> 37
pTHSSf_37	pSC101*	Amp	TALEsp1 weak1 stabilized promoter driving GFP in pSC101 var6	47 -> 38
pTHSSf_38	pSC101*	Amp	TALEsp1 weak1 stabilized promoter driving GFP in pSC101 var7	47 -> 39
pTHSSf_39	pSC101	Amp	TALEsp1 weak2 stabilized promoter driving GFP in pSC101	48 -> 32
pTHSSf_40	pSC101*	Amp	TALEsp1 weak2 stabilized promoter driving GFP in pSC101 var1	48 -> 33
pTHSSf_41	pSC101*	Amp	TALEsp1 weak2 stabilized promoter driving GFP in pSC101 var2	48 -> 34
pTHSSf_42	pSC101*	Amp	TALEsp1 weak2 stabilized promoter driving GFP in pSC101 var3	48 -> 35
pTHSSf_43	pSC101*	Amp	TALEsp1 weak2 stabilized promoter driving GFP in pSC101 var4	48 -> 36
pTHSSf_44	pSC101*	Amp	TALEsp1 weak2 stabilized promoter driving GFP in pSC101 var5	48 -> 37
pTHSSf_45	pSC101*	Amp	TALEsp1 weak2 stabilized promoter driving GFP in pSC101 var6	48 -> 38
pTHSSf_46	pSC101*	Amp	TALEsp1 weak2 stabilized promoter driving GFP in pSC101 var7	48 -> 39
pTHSSf_47	pSC101	Amp	TALEsp1 weak3 stabilized promoter driving GFP in pSC101	49 -> 32
pTHSSf_48	pSC101*	Amp	TALEsp1 weak3 stabilized promoter driving GFP in pSC101 var1	49 -> 33
pTHSSf_49	pSC101*	Amp	TALEsp1 weak3 stabilized promoter driving GFP in pSC101 var2	49 -> 34
pTHSSf_50	pSC101*	Amp	TALEsp1 weak3 stabilized promoter driving GFP in pSC101 var3	49 -> 35
pTHSSf_51	pSC101*	Amp	TALEsp1 weak3 stabilized promoter driving GFP in pSC101 var4	49 -> 36
pTHSSf_52	pSC101*	Amp	TALEsp1 weak3 stabilized promoter driving GFP in pSC101 var5	49 -> 37
pTHSSf_53	pSC101*	Amp	TALEsp1 weak3 stabilized promoter driving GFP in pSC101 var6	49 -> 38
pTHSSf_54	pSC101*	Amp	TALEsp1 weak3 stabilized promoter driving GFP in pSC101 var7	49 -> 39
pTHSSf_55	pSC101	Amp	TALEsp1 weak4 stabilized promoter driving GFP in pSC101	50 -> 32
pTHSSf_56	pSC101*	Amp	TALEsp1 weak4 stabilized promoter driving GFP in pSC101 var1	50 -> 33
pTHSSf_57	pSC101*	Amp	TALEsp1 weak4 stabilized promoter driving GFP in pSC101 var2	50 -> 34
pTHSSf_58	pSC101*	Amp	TALEsp1 weak4 stabilized promoter driving GFP in pSC101 var2	50 -> 34
pTHSSf_59	pSC101*	Amp	TALEsp1 weak4 stabilized promoter driving GFP in pSC101 var3	50 -> 35
pTHSSf_60	pSC101*	Amp	TALEsp1 weak4 stabilized promoter driving GFP in pSC101 var4	50 -> 36
pTHSSf_61	pSC101*	Amp	TALEsp1 weak4 stabilized promoter driving GFP in pSC101 var5	50 -> 37
pTHSSf_62	pSC101*	Amp	TALEsp1 weak4 stabilized promoter driving GFP in pSC101 var6	50 -> 38
pTHSSf_63	pSC101*	Amp	TALEsp1 weak4 stabilized promoter driving GFP in pSC101 var7	50 -> 39

Table 3.2 continued

Name[a]	Origin	Marker	Description	Cloning[b]
pTHSSf_64	pSC101	Amp	TALEsp1 stabilized promoter driving GFP and RFP in pSC101	51 -> 32
pTHSSf_65	pSC101*	Amp	TALEsp1 stabilized promoter driving GFP and RFP in pSC101 var1	51 -> 33
pTHSSf_66	pSC101*	Amp	TALEsp1 stabilized promoter driving GFP and RFP in pSC101 var2	51 -> 34
pTHSSf_67	pSC101*	Amp	TALEsp1 stabilized promoter driving GFP and RFP in pSC101 var3	51 -> 35
pTHSSf_68	pSC101*	Amp	TALEsp1 stabilized promoter driving GFP and RFP in pSC101 var4	51 -> 36
pTHSSf_69	pSC101*	Amp	TALEsp1 stabilized promoter driving GFP and RFP in pSC101 var5	51 -> 37
pTHSSf_70	pSC101*	Amp	TALEsp1 stabilized promoter driving GFP and RFP in pSC101 var6	51 -> 38
pTHSSf_71	pSC101*	Amp	TALEsp1 stabilized promoter driving GFP and RFP in pSC101 var7	51 -> 39

a. Plasmid numbers are preliminary. Please refer to the published manuscript for updated numbers.
b. Numbers refer to pTHSSe plasmids from Table 3.1.

Table 3.3: Genomic insertion locations

Insertion #	Insulated Location[a]	Orientation[b]	Insertion #	Stabilized Location	Orientation
1	604394	Fwd	1	-327531	Rev
2	-2300253	Fwd	2	1811424	Fwd
3	1050060.5	Rev	3	1099717	Rev
4	181487	Fwd	4	-19881	Fwd
5	898113	Fwd	5	-789422	Fwd
6	-28933	Fwd	6	-1227525	Fwd
7	-253564	Rev	7	897969	Rev
8	-1921234	Fwd	8	86844	Rev
9	-851326	Rev	9	1159469	Fwd
10	-814174	Rev	10	-1062333	Fwd
11	-1073034	Fwd	11	2142751	Rev
12	-193907	Rev	12	281986	Fwd
13	979066	Fwd	13	-420370	Fwd
14	514408	Rev	14	571851	Rev
15	6633	Rev	15	615939	Rev
16	-1471863	Rev	16	518866	Rev
17	617018	Fwd	17	-2295964	Rev
18	932461	Fwd	18	987467	Rev
19	-1511916	Fwd	19	-2315671	Fwd
20	-798985	Fwd	20	-1894334	Rev
21	-1014260	Fwd	21	307588	Fwd
22	336123	Fwd	22	-1144415	Rev
23	2340069	Rev	23	1643979	Rev
24	-155749	Rev	24	-694590	Fwd
25	159839	Fwd	25	-302635	Fwd
26	-658535	Fwd	26	-268017	Fwd
27	461690	Rev	27	527321	Rev
28	-1694070	Rev	28	-1264993	Rev
29	-273957	Fwd	29	2156245	Fwd
30	-1437949	Fwd	30	1078956	Rev
31	-3599	Fwd	31	397309	Fwd
32	-921586	Rev	32	1511815	Fwd
33	-683477	Rev	33	-24882	Rev
34	-1875522	Rev	34	-1170877	Fwd
35	2309556	Fwd	35	622300	Rev

a. The location of each genomic insertion is given relative to the center of the genomic origin of replication identified using OriDB [300]. (*E. coli* DH10B taxid: 316385, genome ref seq: NC_010473.1) The location corresponds to the middle bp of the 9bp replicated during Tn5 transposition. When sequencing reads were not available for each end of the transposition, a 9bp duplication was assumed.
b. The orientation of the integration, relative to the orientation of the GFP gene. Fwd: in the direction of leading strand replication, Rev: in the direction of lagging strand replication.
c. This integration had reads on both ends of the transposition revelaing a 10bp duplication. The position reported is between the 5th and 6th bp of the duplication.

Bibliography

[1] Mark Ptashne. Gene regulation by proteins acting nearby and at a distance. *Nature*, 322(6081):697–701, August 1986.

[2] Trey Ideker, Timothy Galitski, and Leroy Hood. A NEW APPROACH TO DECODING LIFE: Systems Biology. *Annual Review of Genomics and Human Genetics*, 2(1):343–372, 2001.

[3] Uri Alon. *An Introduction to Systems Biology: Design Principles of Biological Circuits.* CRC Press, 2007.

[4] Jacques Monod and François Jacob. General Conclusions: Teleonomic Mechanisms in Cellular Metabolism, Growth, and Differentiation. *Cold Spring Harb Symp Quant Biol*, 26:389–401, January 1961.

[5] M. Ptashne. A genetic switch: Gene control and phage. lambda. January 1986.

[6] H. H. McAdams and L. Shapiro. Circuit simulation of genetic networks. *Science*, 269(5224):650–656, August 1995.

[7] Felix Moser, Nicolette J. Broers, Sybe Hartmans, Alvin Tamsir, Richard Kerkman, Johannes A. Roubos, Roel Bovenberg, and Christopher A. Voigt. Genetic Circuit Performance under Conditions Relevant for Industrial Bioreactors. *ACS Synth. Biol.*, 1(11):555–564, November 2012.

[8] Warren C. Ruder, Ting Lu, and James J. Collins. Synthetic Biology Moving into the Clinic. *Science*, 333(6047):1248–1252, September 2011.

[9] Yvonne Y. Chen and Christina D. Smolke. From DNA to Targeted Therapeutics: Bringing Synthetic Biology to the Clinic. *Sci Transl Med*, 3(106):106ps42, October 2011.

[10] Debadyuti Ghosh, Aditya G. Kohli, Felix Moser, Drew Endy, and Angela M. Belcher. Refactored M13 Bacteriophage as a Platform for Tumor Cell Imaging and Drug Delivery. *ACS Synth. Biol.*, 1(12):576–582, December 2012.

[11] Jin H. Huh, Josh T. Kittleson, Adam P. Arkin, and J. Christopher Anderson. Modular Design of a Synthetic Payload Delivery Device. *ACS Synth. Biol.*, 2(8):418–424, August 2013.

[12] Jeff Hasty. Engineered Microbes for Therapeutic Applications. *ACS Synth. Biol.*, 1(10):438–439, October 2012.

[13] Tessa A. Bowen, Jeffrey K. Zdunek, and June I. Medford. Cultivating plant synthetic biology from systems biology. *New Phytologist*, 179(3):583–587, 2008.

[14] Justin Feng, Benjamin W. Jester, Christine E. Tinberg, Daniel J. Mandell, Mauricio S. Antunes, Raj Chari, Kevin J. Morey, Xavier Rios, June I. Medford, George M. Church, Stanley Fields, and David Baker. A general strategy to construct small molecule biosensors in eukaryotes. *eLife Sciences*, 4:e10606, December 2015.

[15] Noah D. Taylor, Alexander S. Garruss, Rocco Moretti, Sum Chan, Mark A. Arbing, Duilio Cascio, Jameson K. Rogers, Farren J. Isaacs, Sriram Kosuri, David Baker, Stanley Fields, George M. Church, and Srivatsan Raman. Engineering an allosteric transcription factor to respond to new ligands. *Nat Meth*, 13(2):177–183, February 2016.

[16] Alec A. K. Nielsen, Bryan S. Der, Jonghyeon Shin, Prashant Vaidyanathan, Vanya Paralanov, Elizabeth A. Strychalski, David Ross, Douglas Densmore, and Christopher A. Voigt. Genetic circuit design automation. *Science*, 352(6281):aac7341, April 2016.

[17] Nathaniel Roquet, Ava P. Soleimany, Alyssa C. Ferris, Scott Aaronson, and Timothy K. Lu. Synthetic recombinase-based state machines in living cells. *Science*, 353(6297):aad8559, July 2016.

[18] Benjamin H. Weinberg, N. T. Hang Pham, Leidy D. Caraballo, Thomas Lozanoski, Adrien Engel, Swapnil Bhatia, and Wilson W. Wong. Large-scale design of robust genetic circuits with multiple inputs and outputs for mammalian cells. *Nat Biotech*, 35(5):453–462, May 2017.

[19] Michael J. Smanski, Swapnil Bhatia, Dehua Zhao, YongJin Park, Lauren B. A. Woodruff, Georgia Giannoukos, Dawn Ciulla, Michele Busby, Johnathan Calderon, Robert Nicol, D. Benjamin Gordon, Douglas Densmore, and Christopher A. Voigt. Functional optimization of gene clusters by combinatorial design and assembly. *Nat Biotech*, 32(12):1241–1249, December 2014.

[20] Miryoung Song, David J. Sukovich, Luciano Ciccarelli, Julia Mayr, Jesus Fernandez-Rodriguez, Ethan A. Mirsky, Alex C. Tucker, D. Benjamin Gordon, Thomas C. Marlovits, and Christopher A. Voigt. Control of type III protein secretion using a minimal genetic system. *Nature Communications*, 8:ncomms14737, May 2017.

[21] Lauren B. A. Woodruff, Thomas E. Gorochowski, Nicholas Roehner, Tarjei S. Mikkelsen, Douglas Densmore, D. Benjamin Gordon, Robert Nicol, and Christopher A. Voigt. Registry in a tube: multiplexed pools of retrievable parts for genetic design space exploration. *Nucleic Acids Res*, 45(3):1553–1565, February 2017.

[22] Harris H. Wang, Farren J. Isaacs, Peter A. Carr, Zachary Z. Sun, George Xu, Craig R. Forest, and George M. Church. Programming cells by multiplex genome engineering and accelerated evolution. *Nature*, 460(7257):894–898, August 2009.

[23] Jay D. Keasling. Synthetic biology and the development of tools for metabolic engineering. *Metabolic Engineering*, 14(3):189–195, May 2012.

[24] William R. Farmer and James C. Liao. Improving lycopene production in Escherichia coli by engineering metabolic control. *Nat Biotech*, 18(5):533–537, May 2000.

[25] Alon Zaslaver, Avi E. Mayo, Revital Rosenberg, Pnina Bashkin, Hila Sberro, Miri Tsalyuk, Michael G. Surette, and Uri Alon. Just-in-time transcription program in metabolic pathways. *Nature Genetics*, 36(5):486–491, April 2004.

[26] Brian F. Pfleger, Douglas J. Pitera, Christina D. Smolke, and Jay D. Keasling. Combinatorial engineering of intergenic regions in operons tunes expression of multiple genes. *Nat Biotech*, 24(8):1027–1032, August 2006.

[27] Alon Zaslaver, Avi Mayo, Michal Ronen, and Uri Alon. Optimal gene partition into operons correlates with gene functional order. *Phys. Biol.*, 3(3):183, 2006.

[28] Fabienne F. V. Chevance and Kelly T. Hughes. Coordinating assembly of a bacterial macromolecular machine. *Nat Rev Micro*, 6(6):455–465, June 2008.

[29] Gene-Wei Li, David Burkhardt, Carol Gross, and Jonathan S. Weissman. Quantifying absolute protein synthesis rates reveals principles underlying allocation of cellular resources. *Cell*, 157(3):624–635, April 2014.

[30] Eric L. Haseltine and Frances H. Arnold. Synthetic gene circuits: design with directed evolution. *Annu Rev Biophys Biomol Struct*, 36:1–19, 2007.

[31] Sriram Kosuri, Daniel B. Goodman, Guillaume Cambray, Vivek K. Mutalik, Yuan Gao, Adam P. Arkin, Drew Endy, and George M. Church. Composability of regulatory sequences controlling transcription and translation in Escherichia coli. *PNAS*, 110(34):14024–14029, August 2013.

[32] Vivek K. Mutalik, Joao C. Guimaraes, Guillaume Cambray, Quynh-Anh Mai, Marc Juul Christoffersen, Lance Martin, Ayumi Yu, Colin Lam, Cesar Rodriguez, Gaymon Bennett, Jay D. Keasling, Drew Endy, and Adam P. Arkin. Quantitative estimation of activity and quality for collections of functional genetic elements. *Nat Meth*, 10(4):347–353, April 2013.

[33] Evan J. Olson and Jeffrey J. Tabor. Optogenetic characterization methods overcome key challenges in synthetic and systems biology. *Nat Chem Biol*, 10(7):502–511, July 2014.

[34] Joseph H. Davis, Adam J. Rubin, and Robert T. Sauer. Design, construction and characterization of a set of insulated bacterial promoters. *Nucl. Acids Res.*, 39(3):1131–1141, February 2011.

[35] Lei Qi, Rachel E. Haurwitz, Wenjun Shao, Jennifer A. Doudna, and Adam P. Arkin. RNA processing enables predictable programming of gene expression. *Nat Biotech*, 30(10):1002–1006, October 2012.

[36] Chunbo Lou, Brynne Stanton, Ying-Ja Chen, Brian Munsky, and Christopher A. Voigt. Ribozyme-based insulator parts buffer synthetic circuits from genetic context. *Nat Biotech*, 30(11):1137–1142, November 2012.

[37] Vivek K. Mutalik, Joao C. Guimaraes, Guillaume Cambray, Colin Lam, Marc Juul Christoffersen, Quynh-Anh Mai, Andrew B. Tran, Morgan Paull, Jay D. Keasling, Adam P. Arkin, and Drew Endy. Precise and reliable gene expression via standard transcription and translation initiation elements. *Nat Meth*, 10(4):354–360, April 2013.

[38] Iman Farasat, Manish Kushwaha, Jason Collens, Michael Easterbrook, Matthew Guido, and Howard M Salis. Efficient search, mapping, and optimization of multi-protein genetic systems in diverse bacteria. *Mol Syst Biol*, 10(6):n/a–n/a, June 2014.

[39] Joshua T Kittleson, Gabriel C Wu, and J Christopher Anderson. Successes and failures in modular genetic engineering. *Current Opinion in Chemical Biology*, 16(34):329–336, August 2012.

[40] Stefano Cardinale and Adam Paul Arkin. Contextualizing context for synthetic biology identifying causes of failure of synthetic biological systems. *Biotechnology Journal*, 7(7):856–866, 2012.

[41] Adam Paul Arkin. A wise consistency: engineering biology for conformity, reliability, predictability. *Current Opinion in Chemical Biology*, 17(6):893–901, December 2013.

[42] Stefano Cardinale, Marcin Pawel Joachimiak, and Adam Paul Arkin. Effects of genetic variation on the E. coli host-circuit interface. *Cell Rep*, 4(2):231–237, July 2013.

[43] Jonathan R Warner. The economics of ribosome biosynthesis in yeast. *Trends in Biochemical Sciences*, 24(11):437–440, November 1999.

[44] D. P. Nierlich. Amino acid control over RNA synthesis: a re-evaluation. *PNAS*, 60(4):1345–1352, August 1968.

[45] R S Hayward, I P Tittawella, and J G Scaife. Evidence for specific control of RNA polymerase synthesis in Escherichia coli. *Nature New Biol.*, 243(122):6–9, May 1973.

[46] Y Iwakura and A Ishihama. Biosynthesis of RNA polymerase in Escherichia coli. II. control of RNA polymerase synthesis during nutritional shift up and down. *Mol. Gen. Genet.*, 142(1):67–84, December 1975.

[47] David M. Bedwell and Masavasu Nomura. Feedback regulation of RNA polymerase subunit synthesis after the conditional overproduction of RNA polymerase in Escherichia coli. *Molec Gen Genet*, 204(1):17–23, July 1986.

[48] H. Bremer and P. Dennis. Feedback control of ribosome function in Escherichia coli. *Biochimie*, 90(3):493–499, March 2008.

[49] M. Schaechter, O. MaalØe, and N. O. Kjeldgaard. Dependency on Medium and Temperature of Cell Size and Chemical Composition during Balanced Growth of Salmonella typhimurium. *J Gen Microbiol*, 19(3):592–606, December 1958.

[50] Harri Lempiäinen and David Shore. Growth control and ribosome biogenesis. *Current Opinion in Cell Biology*, 21(6):855–863, December 2009.

[51] Kirsten Gausing. Regulation of ribosome production in Escherichia coli: Synthesis and stability of ribosomal RNA and of ribosomal protein messenger RNA at different growth rates. *Journal of Molecular Biology*, 115(3):335–354, September 1977.

[52] David A Schneider, Wilma Ross, and Richard L Gourse. Control of rRNA expression in Escherichia coli. *Current Opinion in Microbiology*, 6(2):151–156, April 2003.

[53] Ishihama A. Subunit of assembly of Escherichia coli RNA polymerase. *Adv Biophys*, 14:1–35, 1981.

[54] K. H. Nierhaus. The assembly of prokaryotic ribosomes. *Biochimie*, 73(6):739–755, June 1991.

[55] Alessandro Fatica and David Tollervey. Making ribosomes. *Current Opinion in Cell Biology*, 14(3):313–318, June 2002.

[56] Hans Bremer, Patrick P Dennis, et al. Modulation of chemical composition and other parameters of the cell by growth rate. *Escherichia coli and Salmonella: cellular and molecular biology*, 2:1553–1569, 1996.

[57] A Ishihama, M. Taketo, T. Saitoh, and R. Fukuda. Control of formation of RNA polymerase in Escherichia coli. *Faculty Research 1970 - 1979*, pages 485–502, January 1976.

[58] Akira Ishihama. Functional Modulation of Escherichia Coli RNA Polymerase. *Annual Review of Microbiology*, 54(1):499–518, 2000.

[59] Madeline A. Shea and Gary K. Ackers. The OR control system of bacteriophage lambda: A physical-chemical model for gene regulation. *Journal of Molecular Biology*, 181(2):211–230, January 1985.

[60] Timothy S. Gardner, Charles R. Cantor, and James J. Collins. Construction of a genetic toggle switch in Escherichia coli. *Nature*, 403(6767):339–342, January 2000.

[61] Michael B. Elowitz and Stanislas Leibler. A synthetic oscillatory network of transcriptional regulators. *Nature*, 403(6767):335–338, January 2000.

[62] Stefan Klumpp and Terence Hwa. Growth-rate-dependent partitioning of RNA polymerases in bacteria. *Proc. Natl. Acad. Sci. U.S.A.*, 105(51):20245–20250, December 2008.

[63] Leeat Keren, Ora Zackay, Maya LotanPompan, Uri Barenholz, Erez Dekel, Vered Sasson, Guy Aidelberg, Anat Bren, Danny Zeevi, Adina Weinberger, Uri Alon, Ron Milo, and Eran Segal. Promoters maintain their relative activity levels under different growth conditions. *Molecular Systems Biology*, 9(1):701, January 2013.

[64] Dirk De Vos, Frank J. Bruggeman, Hans V. Westerhoff, and Barbara M. Bakker. How Molecular Competition Influences Fluxes in Gene Expression Networks. *PLoS ONE*, 6(12):e28494, December 2011.

[65] Adam P. Arkin and Daniel A. Fletcher. Fast, cheap and somewhat in control. *Genome Biology*, 7(8):114, August 2006.

[66] S.-T. Liang, M. Bipatnath, Y.-C. Xu, S.-L. Chen, P. Dennis, M. Ehrenberg, and H. Bremer. Activities of constitutive promoters in Escherichia coli. *Journal of Molecular Biology*, 292(1):19–37, 1999.

[67] Stefan Klumpp, Zhongge Zhang, and Terence Hwa. Growth Rate-Dependent Global Effects on Gene Expression in Bacteria. *Cell*, 139(7):1366–1375, December 2009.

[68] S.-T. Liang, Y.-C. Xu, P. Dennis, and H. Bremer. mRNA composition and control of bacterial gene expression. *Journal of Bacteriology*, 182(11):3037–3044, 2000.

[69] Karsten Temme, Dehua Zhao, and Christopher A. Voigt. Refactoring the nitrogen fixation gene cluster from Klebsiella oxytoca. *PNAS*, 109(18):7085–7090, May 2012.

[70] Tae Seok Moon, Chunbo Lou, Alvin Tamsir, Brynne C. Stanton, and Christopher A. Voigt. Genetic programs constructed from layered logic gates in single cells. *Nature*, 491(7423):249–253, November 2012.

[71] Frank Hoffmann and Ursula Rinas. Stress Induced by Recombinant Protein Production in Escherichia coli. In *Physiological Stress Responses in Bioprocesses*, number 89 in Advances in Biochemical Engineering, pages 73–92. Springer Berlin Heidelberg, January 2004.

[72] S. Birnbaum and J. E. Bailey. Plasmid presence changes the relative levels of many host cell proteins and ribosome components in recombinant Escherichia coli. *Biotechnology and Bioengineering*, 37(8):736–745, 1991.

[73] H Dong, L Nilsson, and C G Kurland. Gratuitous overexpression of genes in Escherichia coli leads to growth inhibition and ribosome destruction. *J Bacteriol*, 177(6):1497–1504, March 1995.

[74] Matthew Scott, Carl W. Gunderson, Eduard M. Mateescu, Zhongge Zhang, and Terence Hwa. Interdependence of Cell Growth and Gene Expression: Origins and Consequences. *Science*, 330(6007):1099–1102, November 2010.

[75] Javier Carrera, Guillermo Rodrigo, Vijai Singh, Boris Kirov, and Alfonso Jaramillo. Empirical model and in vivo characterization of the bacterial response to synthetic gene expression show that ribosome allocation limits growth rate. *Biotechnology Journal*, 6(7):773–783, 2011.

[76] Jeffrey J Tabor, Travis S Bayer, Zachary B Simpson, Matthew Levy, and Andrew D Ellington. Engineering stochasticity in gene expression. *Mol Biosyst*, 4(7):754–761, July 2008.

[77] Peng Jiang, Alejandra C. Ventura, Eduardo D. Sontag, Sofia D. Merajver, Alexander J. Ninfa, and Domitilla Del Vecchio. Load-Induced Modulation of Signal Transduction Networks. *Sci. Signal.*, 4(194):ra67, October 2011.

[78] Shridhar Jayanthi, Kayzad Soli Nilgiriwala, and Domitilla Del Vecchio. Retroactivity Controls the Temporal Dynamics of Gene Transcription. *ACS Synth. Biol.*, 2(8):431–441, August 2013.

[79] Domitilla Del Vecchio, Alexander J Ninfa, and Eduardo D Sontag. Modular cell biology: retroactivity and insulation. *Molecular Systems Biology*, 4, February 2008.

[80] Domitilla Del Vecchio and Richard M. Murray. *Biomolecular Feedback Systems*. Princeton University Press, Princeton, New Jersey, October 2014.

[81] Julius B Lucks, Lei Qi, Weston R Whitaker, and Adam P Arkin. Toward scalable parts families for predictable design of biological circuits. *Current Opinion in Microbiology*, 11(6):567–573, December 2008.

[82] Christopher V Rao. Expanding the synthetic biology toolbox: engineering orthogonal regulators of gene expression. *Current Opinion in Biotechnology*, 23(5):689–694, October 2012.

[83] Travis S. Bayer, Daniel M. Widmaier, Karsten Temme, Ethan A. Mirsky, Daniel V. Santi, and Christopher A. Voigt. Synthesis of Methyl Halides from Biomass Using Engineered Microbes. *J. Am. Chem. Soc.*, 131(18):6508–6515, May 2009.

[84] Barry Canton, Anna Labno, and Drew Endy. Refinement and standardization of synthetic biological parts and devices. *Nat Biotech*, 26(7):787–793, July 2008.

[85] http://parts.igem.org/.

[86] Shuobing Chen, Haoqian Zhang, Handuo Shi, Weiyue Ji, Jingchen Feng, Yan Gong, Zhenglin Yang, and Qi Ouyang. Automated Design of Genetic Toggle Switches with Predetermined Bistability. *ACS Synth. Biol.*, 1(7):284–290, July 2012.

[87] Virgil A. Rhodius, Thomas H. Segall-Shapiro, Brian D. Sharon, Amar Ghodasara, Ekaterina Orlova, Hannah Tabakh, David H. Burkhardt, Kevin Clancy, Todd C. Peterson, Carol A. Gross, and Christopher A. Voigt. Design of orthogonal genetic switches based on a crosstalk map of s, anti-s, and promoters. *Mol Syst Biol*, 9(1), October 2013.

[88] Karsten Temme, Rena Hill, Thomas H. Segall-Shapiro, Felix Moser, and Christopher A. Voigt. Modular control of multiple pathways using engineered orthogonal T7 polymerases. *Nucl. Acids Res.*, 40(17):8773–8781, September 2012.

[89] C. A. Raskin, G. A. Diaz, and W. T. McAllister. T7 RNA polymerase mutants with altered promoter specificities. *PNAS*, 90(8):3147–3151, April 1993.

[90] David L. Shis and Matthew R. Bennett. Library of synthetic transcriptional AND gates built with split T7 RNA polymerase mutants. *PNAS*, March 2013.

[91] Jijumon Chelliserrykattil, George Cai, and Andrew D. Ellington. A combined in vitro / in vivo selection for polymerases with novel promoter specificities. *BMC Biotechnology*, 1(1):13, December 2001.

[92] Kevin M. Esvelt, Jacob C. Carlson, and David R. Liu. A system for the continuous directed evolution of biomolecules. *Nature*, 472(7344):499–503, April 2011.

[93] Tasha A. Desai, Dmitry A. Rodionov, Mikhail S. Gelfand, Eric J. Alm, and Christopher V. Rao. Engineering transcription factors with novel DNA-binding specificity using comparative genomics. *Nucl. Acids Res.*, 37(8):2493–2503, May 2009.

[94] Jian Zhan, Bo Ding, Rui Ma, Xiaoyu Ma, Xiaofeng Su, Yun Zhao, Ziqing Liu, Jiarui Wu, and Haiyan Liu. Develop reusable and combinable designs for transcriptional logic gates. *Mol Syst Biol*, 6(1), July 2010.

[95] Brynne C. Stanton, Alec A. K. Nielsen, Alvin Tamsir, Kevin Clancy, Todd Peterson, and Christopher A. Voigt. Genomic mining of prokaryotic repressors for orthogonal logic gates. *Nat. Chem. Biol.*, 10(2):99–105, February 2014.

[96] J. R. Desjarlais and J. M. Berg. Toward rules relating zinc finger protein sequences and DNA binding site preferences. *PNAS*, 89(16):7345–7349, August 1992.

[97] Matthew J. Moscou and Adam J. Bogdanove. A Simple Cipher Governs DNA Recognition by TAL Effectors. *Science*, 326(5959):1501–1501, December 2009.

[98] Jens Boch, Heidi Scholze, Sebastian Schornack, Angelika Landgraf, Simone Hahn, Sabine Kay, Thomas Lahaye, Anja Nickstadt, and Ulla Bonas. Breaking the Code of DNA Binding Specificity of TAL-Type III Effectors. *Science*, 326(5959):1509–1512, December 2009.

[99] Roger R. Beerli and Carlos F. Barbas. Engineering polydactyl zinc-finger transcription factors. *Nat Biotech*, 20(2):135–141, February 2002.

[100] Robert Morbitzer, Patrick Römer, Jens Boch, and Thomas Lahaye. Regulation of selected genome loci using de novo-engineered transcription activator-like effector (TALE)-type transcription factors. *PNAS*, 107(50):21617–21622, December 2010.

[101] Abhishek Garg, Jason J. Lohmueller, Pamela A. Silver, and Thomas Z. Armel. Engineering synthetic TAL effectors with orthogonal target sites. *Nucl. Acids Res.*, 40(15):7584–7595, August 2012.

[102] Ju Young Lee, Bong Hyun Sung, Byung Jo Yu, Jun Hyoung Lee, Sang Hee Lee, Mi Sun Kim, Michael D. Koob, and Sun Chang Kim. Phenotypic engineering by reprogramming gene transcription using novel artificial transcription factors in Escherichia coli. *Nucl. Acids Res.*, 36(16):e102–e102, September 2008.

[103] Mark C. Politz, Matthew F. Copeland, and Brian F. Pfleger. Artificial repressors for controlling gene expression in bacteria. *Chem. Commun.*, 49(39):4325–4327, April 2013.

[104] Nicholas R. Markham and Michael Zuker. UNAFold. In Jonathan M. Keith, editor, *Bioinformatics*, number 453 in Methods in Molecular Biology, pages 3–31. Humana Press, January 2008.

[105] J.D. Kittle, R.W. Simons, J. Lee, and N. Kleckner. Insertion sequence IS10 anti-sense pairing initiates by an interaction between the 5 end of the target RNA and a loop in the anti-sense RNA. *Journal of Molecular Biology*, 210(3):561–572, December 1989.

[106] Vivek K. Mutalik, Lei Qi, Joao C. Guimaraes, Julius B. Lucks, and Adam P. Arkin. Rationally designed families of orthogonal RNA regulators of translation. *Nat Chem Biol*, 8(5):447–454, May 2012.

[107] Farren J. Isaacs, Daniel J. Dwyer, Chunming Ding, Dmitri D. Pervouchine, Charles R. Cantor, and James J. Collins. Engineered riboregulators enable post-transcriptional control of gene expression. *Nat Biotech*, 22(7):841–847, July 2004.

[108] Jarred M. Callura, Charles R. Cantor, and James J. Collins. Genetic switchboard for synthetic biology applications. *PNAS*, 109(15):5850–5855, April 2012.

[109] Lon M. Chubiz and Christopher V. Rao. Computational design of orthogonal ribosomes. *Nucl. Acids Res.*, 36(12):4038–4046, July 2008.

[110] Oliver Rackham and Jason W. Chin. A network of orthogonal ribosomeůmRNA pairs. *Nat Chem Biol*, 1(3):159–166, August 2005.

[111] Wenlin An and Jason W. Chin. Synthesis of orthogonal transcription-translation networks. *PNAS*, 106(21):8477–8482, May 2009.

[112] Chang C. Liu, Lei Qi, Julius B. Lucks, Thomas H. Segall-Shapiro, Denise Wang, Vivek K. Mutalik, and Adam P. Arkin. An adaptor from translational to transcriptional control enables predictable assembly of complex regulation. *Nat Meth*, 9(11):1088–1094, November 2012.

[113] Feng Gong, Koichi Ito, Yoshikazu Nakamura, and Charles Yanofsky. The mechanism of tryptophan induction of tryptophanase operon expression: Tryptophan inhibits release factor-mediated cleavage of TnaC-peptidyl-tRNAPro. *PNAS*, 98(16):8997–9001, July 2001.

[114] Philippe Horvath and Rodolphe Barrangou. CRISPR/Cas, the Immune System of Bacteria and Archaea. *Science*, 327(5962):167–170, January 2010.

[115] LeiăS. Qi, MatthewăH. Larson, LukeăA. Gilbert, JenniferăA. Doudna, JonathanăS. Weissman, AdamăP. Arkin, and WendellăA. Lim. Repurposing CRISPR as an RNA-Guided Platform for Sequence-Specific Control of Gene Expression. *Cell*, 152(5):1173–1183, February 2013.

[116] David Bikard, Wenyan Jiang, Poulami Samai, Ann Hochschild, Feng Zhang, and Luciano A. Marraffini. Programmable repression and activation of bacterial gene expression using an engineered CRISPR-Cas system. *Nucl. Acids Res.*, page gkt520, June 2013.

[117] Yanfang Fu, Jennifer A. Foden, Cyd Khayter, Morgan L. Maeder, Deepak Reyon, J. Keith Joung, and Jeffry D. Sander. High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nat Biotech*, advance online publication, June 2013.

[118] Patrick D. Hsu, David A. Scott, Joshua A. Weinstein, F. Ann Ran, Silvana Konermann, Vineeta Agarwala, Yinqing Li, Eli J. Fine, Xuebing Wu, Ophir Shalem, Thomas J. Cradick, Luciano A. Marraffini, Gang Bao, and Feng Zhang. DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat Biotech*, 31(9):827–832, September 2013.

[119] Melissa K. Takahashi and Julius B. Lucks. A modular strategy for engineering orthogonal chimeric RNA transcription regulators. *Nucl. Acids Res.*, page gkt452, June 2013.

[120] Julius B. Lucks, Lei Qi, Vivek K. Mutalik, Denise Wang, and Adam P. Arkin. Versatile RNA-sensing transcriptional regulators for engineering genetic networks. *PNAS*, 108(21):8617–8622, May 2011.

[121] Sabine Brantl and E. Gerhart H. Wagner. Antisense RNA-mediated transcriptional attenuation: an in vitro study of plasmid pT181. *Molecular Microbiology*, 35(6):1469–1482, 2000.

[122] Nobutaka Hirano, Tetsurou Muroi, Hideo Takahashi, and Mitsuru Haruki. Site-specific recombinases as tools for heterologous gene integration. *Appl Microbiol Biotechnol*, 92(2):227–239, October 2011.

[123] Andras Nagy. Cre recombinase: The universal reagent for genome tailoring. *genesis*, 26(2):99–109, 2000.

[124] Ari E. Friedland, Timothy K. Lu, Xiao Wang, David Shi, George Church, and James J. Collins. Synthetic Gene Networks that Count. *Science*, 324(5931):1199–1202, May 2009.

[125] Timothy S. Ham, Sung K. Lee, Jay D. Keasling, and Adam P. Arkin. Design and Construction of a Double Inversion Recombination Switch for Heritable Sequential Genetic Memory. *PLoS ONE*, 3(7):e2815, July 2008.

[126] Amy C. Groth and Michele P. Calos. Phage Integrases: Biology and Applications. *Journal of Molecular Biology*, 335(3):667–678, January 2004.

[127] Piro Siuti, John Yazbek, and Timothy K. Lu. Synthetic circuits integrating logic and memory in living cells. *Nat Biotech*, 31(5):448–452, May 2013.

[128] Jerome Bonnet, Peter Yin, Monica E. Ortiz, Pakpoom Subsoontorn, and Drew Endy. Amplifying Genetic Logic Gates. *Science*, 340(6132):599–603, May 2013.

[129] Jerome Bonnet, Pakpoom Subsoontorn, and Drew Endy. Rewritable digital data storage in live cells via engineered control of recombination directionality. *PNAS*, 109(23):8884–8889, June 2012.

[130] Stephen W. Santoro and Peter G. Schultz. Directed evolution of the site specificity of Cre recombinase. *PNAS*, 99(7):4185–4190, April 2002.

[131] Thomas Gaj, Andrew C. Mercer, Charles A. Gersbach, Russell M. Gordley, and Carlos F. Barbas. Structure-guided reprogramming of serine recombinase DNA sequence specificity. *PNAS*, 108(2):498–503, January 2011.

[132] Andrew C. Mercer, Thomas Gaj, Roberta P. Fuller, and Carlos F. Barbas. Chimeric TALE recombinases with programmable DNA sequence specificity. *Nucl. Acids Res.*, 40(21):11163–11172, November 2012.

[133] Russell M. Gordley, Justin D. Smith, Torbjörn Gräslund, and Carlos F. Barbas. Evolution of programmable zinc finger-recombinases with activity in human cells. *J. Mol. Biol.*, 367(3):802–813, March 2007.

[134] Brian Chaikind, Krishna Praneeth Kilambi, Jeffrey J. Gray, and Marc Ostermeier. Targeted DNA Methylation Using an Artificially Bisected M.HhaI Fused to Zinc Fingers. *PLoS ONE*, 7(9):e44852, September 2012.

[135] Anders Løbner-Olesen, Ole Skovgaard, and Martin G Marinus. Dam methylation: coordinating cellular processes. *Current Opinion in Microbiology*, 8(2):154–160, April 2005.

[136] Josep Casadesús and David Low. Epigenetic Gene Regulation in the Bacterial World. *Microbiol. Mol. Biol. Rev.*, 70(3):830–856, September 2006.

[137] Marjan De Mey, Jo Maertens, Gaspard J. Lequeux, Wim K. Soetaert, and Erick J. Vandamme. Construction and model-based analysis of a promoter library for E. coli: an indispensable tool for metabolic engineering. *BMC Biotechnol*, 7(1):34, June 2007.

[138] Peter Ruhdal Jensen and Karin Hammer. The Sequence of Spacers between the Consensus Sequences Modulates the Strength of Prokaryotic Promoters. *Appl Environ Microbiol*, 64(1):82–87, January 1998.

[139] Ida Rud, Peter Ruhdal Jensen, Kristine Naterstad, and Lars Axelsson. A synthetic promoter library for constitutive gene expression in Lactobacillus plantarum. *Microbiology*, 152(4):1011–1019, April 2006.

[140] Nicolas Seghezzi, Patrick Amar, Brian Koebmann, Peter R. Jensen, and Marie-Joëlle Virolle. The construction of a library of synthetic promoters revealed some specific features of strong Streptomyces promoters. *Appl Microbiol Biotechnol*, 90(2):615–623, April 2011.

[141] Virgil A. Rhodius, Vivek K. Mutalik, and Carol A. Gross. Predicting the strength of UP-elements and full-length E. coli E promoters. *Nucl. Acids Res.*, 40(7):2907–2924, April 2012.

[142] Robert C. Brewster, Daniel L. Jones, and Rob Phillips. Tuning Promoter Strength through RNA Polymerase Binding Site Design in Escherichia coli. *PLoS Comput Biol*, 8(12):e1002811, December 2012.

[143] Subhayu Basu, Rishabh Mehreja, Stephan Thiberge, Ming-Tang Chen, and Ron Weiss. Spatiotemporal control of gene expression with pulse-generating networks. *PNAS*, 101(17):6355–6360, April 2004.

[144] Robert G. Egbert and Eric Klavins. Fine-tuning gene networks using simple sequence repeats. *PNAS*, 109(42):16817–16822, October 2012.

[145] Howard M. Salis, Ethan A. Mirsky, and Christopher A. Voigt. Automated design of synthetic ribosome binding sites to control protein expression. *Nat Biotech*, 27(10):946–950, October 2009.

[146] Howard M. Salis. Chapter two - The Ribosome Binding Site Calculator. In Christopher Voigt, editor, *Methods in Enzymology*, volume Volume 498 of *Synthetic Biology, Part B Computer Aided Design and DNA Assembly*, pages 19–42. Academic Press, 2011.

[147] Maarten H. de Smit and Jan van Duin. Translational Standby Sites: How Ribosomes May Deal with the Rapid Folding Kinetics of mRNA. *Journal of Molecular Biology*, 331(4):737–743, August 2003.

[148] Irina V. Boni, Valentina S. Artamonova, Nina V. Tzareva, and Marc Dreyfus. Non-canonical mechanism for translational control in bacteria: synthesis of ribosomal protein S1. *EMBO J*, 20(15):4222–4232, August 2001.

[149] Brian Søgaard Laursen, Hans Peter Sørensen, Kim Kusk Mortensen, and Hans Uffe Sperling-Petersen. Initiation of Protein Synthesis in Bacteria. *Microbiol. Mol. Biol. Rev.*, 69(1):101–123, March 2005.

[150] Ying-Ja Chen, Peng Liu, Alec A. K. Nielsen, Jennifer A. N. Brophy, Kevin Clancy, Todd Peterson, and Christopher A. Voigt. Characterization of 582 natural and synthetic terminators and quantification of their design constraints. *Nat Meth*, 10(7):659–664, July 2013.

[151] Guillaume Cambray, Joao C. Guimaraes, Vivek K. Mutalik, Colin Lam, Quynh-Anh Mai, Tim Thimmaiah, James M. Carothers, Adam P. Arkin, and Drew Endy. Measurement and modeling of intrinsic transcription terminators. *Nucleic Acids Res*, 41(9):5139–5148, May 2013.

[152] Joshua T. Kittleson, Sherine Cheung, and J. Christopher Anderson. Rapid optimization of gene dosage in E. coli using DIAL strains. *J Biol Eng*, 5(1):10, July 2011.

[153] Andrew I. Yao, Timothy A. Fenton, Keegan Owsley, Phillip Seitzer, David J. Larsen, Holly Sit, Jennifer Lau, Arjun Nair, Justin Tantiongloc, Ilias Tagkopoulos, and Marc T. Facciotti. Promoter Element Arising from the Fusion of Standard BioBrick Parts. *ACS Synth. Biol.*, 2(2):111–120, February 2013.

[154] Shawn T. Estrem, Tamas Gaal, Wilma Ross, and Richard L. Gourse. Identification of an UP element consensus sequence for bacterial promoters. *PNAS*, 95(17):9761–9766, August 1998.

[155] U Schmeissner, K McKenney, M Rosenberg, and D Court. Removal of a terminator structure by RNA processing regulates int gene expression. *J. Mol. Biol.*, 176(1):39–53, June 1984.

[156] Timothy S. Ham, Sung Kuk Lee, Jay D. Keasling, and Adam P. Arkin. A tightly regulated inducible expression system utilizing the fim inversion recombination switch. *Biotechnology and Bioengineering*, 94(1):1–4, 2006.

[157] Jason J. Lohmueller, Thomas Z. Armel, and Pamela A. Silver. A tunable zinc finger-based framework for Boolean logic computation in mammalian cells. *Nucl. Acids Res.*, February 2012.

[158] Fahim Farzadfard, Samuel D. Perli, and Timothy K. Lu. Tunable and Multifunctional Eukaryotic Transcription Factors Based on CRISPR/Cas. *ACS Synth. Biol.*, 2(10):604–613, October 2013.

[159] Michael Chamberlin, Janet Mcgrath, and Lucy Waskell. New RNA Polymerase from Escherichia coli infected with Bacteriophage T7. *Nature*, 228(5268):227–231, October 1970.

[160] F.William Studier and Barbara A. Moffatt. Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *Journal of Molecular Biology*, 189(1):113–130, May 1986.

[161] W A Alexander, B Moss, and T R Fuerst. Regulated expression of foreign genes in vaccinia virus under the control of bacteriophage T7 RNA polymerase and the Escherichia coli lac repressor. *J Virol*, 66(5):2934–2942, May 1992.

[162] F. William Studier, Parichehre Davanloo, Alan H. Rosenberg, Barbara A. Moffatt, and John J. Dunn. Cloning and expression of the gene for bacteriophage T7 RNA polymerase, August 1990.

[163] O. Elroy-Stein and B. Moss. Cytoplasmic expression system based on constitutive synthesis of bacteriophage T7 RNA polymerase in mammalian cells. *PNAS*, 87(17):6743–6747, September 1990.

[164] Elaine Brunschwig and Aldis Darzins. A two-component T7 system for the overexpression of genes in Pseudomonas aeruginosa. *Gene*, 111(1):35–41, February 1992.

[165] K. E. McBride, D. J. Schaaf, M. Daley, and D. M. Stalker. Controlled expression of plastid transgenes in plants based on a nuclear DNA-encoded and plastid-targeted T7 RNA polymerase. *PNAS*, 91(15):7301–7305, July 1994.

[166] Birgit Conrad, Regina S. Savchenko, Roland Breves, and Jürgen Hofemeister. A T7 promoter-specific, inducible protein expression system forBacillus subtilis. *Molec. Gen. Genet.*, 250(2):230–236, February 1996.

[167] S. Tabor and C. C. Richardson. A bacteriophage T7 RNA polymerase/promoter system for controlled exclusive expression of specific genes. *PNAS*, 82(4):1074–1078, February 1985.

[168] I Iost, J Guillerez, and M Dreyfus. Bacteriophage T7 RNA polymerase travels far ahead of ribosomes in vivo. *J Bacteriol*, 174(2):619–622, January 1992.

[169] Bruno Miroux and John E. Walker. Over-production of Proteins inEscherichia coli: Mutant Hosts that Allow Synthesis of some Membrane Proteins and Globular Proteins at High Levels. *Journal of Molecular Biology*, 260(3):289–298, July 1996.

[170] Jared W. Ellefson, Adam J. Meyer, Randall A. Hughes, Joe R. Cannon, Jennifer S. Brodbelt, and Andrew D. Ellington. Directed evolution of genetic parts and circuits by compartmentalized partnered replication. *Nat Biotech*, advance online publication, November 2013.

[171] Jacob C. Carlson, Ahmed H. Badran, Drago A. Guggiana-Nilo, and David R. Liu. Negative selection and stringency modulation in phage-assisted continuous evolution. *Nat Chem Biol*, 10(3):216–222, March 2014.

[172] Tanja M. Gruber and Carol A. Gross. Multiple Sigma Subunits and the Partitioning of Bacterial Transcription Space. *Annual Review of Microbiology*, 57(1):441–466, 2003.

[173] H. El-Samad, H. Kurata, J. C. Doyle, C. A. Gross, and M. Khammash. Surviving heat shock: Control strategies for robustness and performance. *PNAS*, 102(8):2736–2741, February 2005.

[174] M Jishage, A Iwata, S Ueda, and A Ishihama. Regulation of RNA polymerase sigma subunit synthesis in Escherichia coli: intracellular levels of four species of sigma subunit under various growth conditions. *J. Bacteriol.*, 178(18):5447–5451, September 1996.

[175] Z Burton, R R Burgess, J Lin, D Moore, S Holder, and C A Gross. The nucleotide sequence of the cloned rpoD gene for the RNA polymerase sigma subunit from E coli K12. *Nucleic Acids Res*, 9(12):2889–2903, June 1981.

[176] Anna Staro, Heidi J. Sofia, Sascha Dietrich, Luke E. Ulrich, Heiko Liesegang, and Thorsten Mascher. The third pillar of bacterial signal transduction: classification of the extracytoplasmic function (ECF) factor protein family. *Molecular Microbiology*, 74(3):557–581, 2009.

[177] R Lange and R Hengge-Aronis. The cellular concentration of the sigma S subunit of RNA polymerase in Escherichia coli is controlled at the levels of transcription, translation, and protein stability. *Genes Dev.*, 8(13):1600–1612, July 1994.

[178] Regine Hengge-Aronis. Signal Transduction and Regulatory Mechanisms Involved in Control of the S (RpoS) Subunit of RNA Polymerase. *Microbiol. Mol. Biol. Rev.*, 66(3):373–395, September 2002.

[179] H. Kurata, H. El-Samad, T.-M. Yi, M. Khammash, and J. Doyle. Feedback regulation of the heat shock response in E. coli. In *Proceedings of the 40th IEEE Conference on Decision and Control, 2001*, volume 1, pages 837–842 vol.1, 2001.

[180] Thomas Nyström. MicroReview: Growth versus maintenance: a trade-off dictated by RNA polymerase availability and sigma factor competition? *Molecular Microbiology*, 54(4):855–862, November 2004.

[181] Ram Maharjan, Susanna Nilsson, Judy Sung, Ken Haynes, Robert E. Beardmore, Laurence D. Hurst, Tom Ferenci, and Ivana Gudelj. The form of a trade-off determines the response to competition. *Ecology Letters*, 16(10):1267–1276, 2013.

[182] Marc Ostermeier, Andrew E. Nixon, Jae Hoon Shim, and Stephen J. Benkovic. Combinatorial protein engineering by incremental truncation. *PNAS*, 96(7):3562–3567, March 1999.

[183] David E Paschon and Marc Ostermeier. Construction of Protein Fragment Complementation Libraries Using Incremental Truncation. In Dan E. Robertson and Joseph P. Noel, editor, *Methods in Enzymology*, volume Volume 388, pages 103–116. Academic Press, 2004.

[184] Manu Kanwar, R. Clay Wright, Amol Date, Jennifer Tullman, and Marc Ostermeier. Chapter Seventeen - Protein Switch Engineering by Domain Insertion. In Amy E. Keating, editor, *Methods in Enzymology*, volume Volume 523, pages 369–388. Academic Press, 2013.

[185] Gurkan Guntas and Marc Ostermeier. Creation of an Allosteric Enzyme by Domain Insertion. *Journal of Molecular Biology*, 336(1):263–273, February 2004.

[186] Yong Chen, Shuang Li, Tingjian Chen, Hui Hua, and Zhanglin Lin. Random dissection to select for protein split sites and its application in protein fragment complementation. *Protein Science*, 18(2):399–409, February 2009.

[187] Thomas H. Segall-Shapiro, Peter Q. Nguyen, Edgardo D. Dos Santos, Saurav Subedi, Justin Judd, Junghae Suh, and Jonathan J. Silberg. Mesophilic and Hyperthermophilic Adenylate Kinases Differ in Their Tolerance to Random Fragmentation. *Journal of Molecular Biology*, 406(1):135–148, February 2011.

[188] Alborz Mahdavi, Thomas H Segall-Shapiro, Songzi Kou, Granton A Jindal, Kevin G Hoff, Shirley Liu, Mohsen Chitsaz, Rustem F Ismagilov, Jonathan J Silberg, and David A Tirrell. A genetically encoded and gate for cell-targeted metabolic labeling of proteins. *J. Am. Chem. Soc.*, 135(8):2979–2982, February 2013.

[189] R. A. Ikeda and C. C. Richardson. Enzymatic properties of a proteolytically nicked RNA polymerase of bacteriophage T7. *J. Biol. Chem.*, 262(8):3790–3799, March 1987.

[190] R A Ikeda and C C Richardson. Interactions of a proteolytically nicked RNA polymerase of bacteriophage T7 with its promoter. *J. Biol. Chem.*, 262(8):3800–3808, March 1987.

[191] Daniel K. Muller, Craig T. Martin, and Joseph E. Coleman. Processivity of proteolytically modified forms of T7 RNA polymerase. *Biochemistry*, 27(15):5763–5771, July 1988.

[192] Tahir H Tahirov, Dmitry Temiakov, Michael Anikin, Vsevolod Patlan, William T McAllister, Dmitry G Vassylyev, and Shigeyuki Yokoyama. Structure of a T7 RNA polymerase elongation complex at 2.9 A resolution. *Nature*, 420(6911):43–50, November 2002.

[193] Aaron W. Reinke, Robert A. Grant, and Amy E. Keating. A Synthetic Coiled-Coil Interactome Provides Heterospecific Modules for Molecular Engineering. *J. Am. Chem. Soc.*, 132(17):6025–6031, May 2010.

[194] Kenneth Evan Thompson, Caleb J. Bashor, Wendell A. Lim, and Amy E. Keating. SYNZIP Protein Interaction Toolbox: in Vitro and in Vivo Specifications of Heterospecific Coiled-Coil Interaction Domains. *ACS Synth. Biol.*, 1(4):118–129, April 2012.

[195] G M Cheetham, D Jeruzalmi, and T A Steitz. Structural basis for initiation of transcription from an RNA polymerase-promoter complex. *Nature*, 399(6731):80–83, May 1999.

[196] Rui Sousa, Yong Je Chung, John P. Rose, and Bi-Cheng Wang. Crystal structure of bacteriophage T7 RNA polymerase at 3.3 |[angst]| resolution. *Nature*, 364(6438):593–599, August 1993.

[197] Dmitry G. Vassylyev, Shun-ichi Sekine, Oleg Laptenko, Jookyung Lee, Marina N. Vassylyeva, Sergei Borukhov, and Shigeyuki Yokoyama. Crystal structure of a bacterial RNA polymerase holoenzyme at 2.6Å resolution. *Nature*, 417(6890):712–719, June 2002.

[198] Natacha Opalka, Jesse Brown, William J. Lane, Kelly-Anne F. Twist, Robert Landick, Francisco J. Asturias, and Seth A. Darst. Complete Structural Model of Escherichia coli RNA Polymerase from a Hybrid Approach. *PLoS Biol*, 8(9):e1000483, September 2010.

[199] Nicolas E. Buchler and Matthieu Louis. Molecular Titration and Ultrasensitivity in Regulatory Networks. *Journal of Molecular Biology*, 384(5):1106–1119, December 2008.

[200] Nicolas E Buchler and Frederick R Cross. Protein sequestration generates a flexible ultrasensitive response in a genetic network. *Molecular Systems Biology*, 5(1):n/a–n/a, 2009.

[201] G Bonner, D Patra, E M Lafer, and R Sousa. Mutations in T7 RNA polymerase that support the proposal for a common polymerase active site structure. *EMBO J*, 11(10):3767–3775, October 1992.

[202] Kasim A. Mookhtiar, Paul S. Peluso, Daniel K. Muller, John J. Dunn, and Joseph E. Coleman. Processivity of T7 RNA polymerase requires the C-terminal Phe882-Ala883-COO- or "foot". *Biochemistry*, 30(25):6305–6313, June 1991.

[203] John J. Dunn, Bernd Krippl, Kenneth E. Bernstein, Heiner Westphal, and F. William Studier. Targeting bacteriophage T7 RNA polymerase to the mammalian cell nucleus. *Gene*, 68(2):259–266, September 1988.

[204] Leonidas Bleris, Zhen Xie, David Glass, Asa Adadey, Eduardo Sontag, and Yaakov Benenson. Synthetic incoherent feedforward circuits show adaptation to the amount of their genetic template. *Molecular Systems Biology*, 7, August 2011.

[205] David R. Liu, Thomas J. Magliery, Miro Pastrnak, and Peter G. Schultz. Engineering a tRNA and aminoacyl-tRNA synthetase for the site-specific incorporation of unnatural amino acids into proteins in vivo. *PNAS*, 94(19):10092–10097, September 1997.

[206] Jason W. Chin. Expanding and Reprogramming the Genetic Code of Cells and Animals. *Annual Review of Biochemistry*, 83(1):null, 2014.

[207] Arjun Ravikumar, Adrian Arrieta, and Chang C. Liu. An orthogonal DNA replication system in yeast. *Nat Chem Biol*, 10(3):175–177, March 2014.

[208] Chris Grilly, Jesse Stricker, Wyming Lee Pang, Matthew R Bennett, and Jeff Hasty. A synthetic gene network for tuning protein degradation in Saccharomyces cerevisiae. *Molecular Systems Biology*, 3(1):n/a–n/a, 2007.

[209] Blaine A. Pfeifer, Suzanne J. Admiraal, Hugo Gramajo, David E. Cane, and Chaitan Khosla. Biosynthesis of Complex Polyketides in a Metabolically Engineered Strain of E. coli. *Science*, 291(5509):1790–1792, March 2001.

[210] Tae Seok Moon, John E Dueber, Eric Shiue, and Kristala L Jones Prather. Use of modular, synthetic scaffolds for improved production of glucaric acid in engineered E. coli. *Metab. Eng.*, 12(3):298–305, May 2010.

[211] Walter Bonacci, Poh K. Teng, Bruno Afonso, Henrike Niederholtmeyer, Patricia Grob, Pamela A. Silver, and David F. Savage. Modularity of a carbon-fixing protein organelle. *PNAS*, 109(2):478–483, January 2012.

[212] Kaiyu Ma, Dmitri Temiakov, Manli Jiang, Michael Anikin, and William T. McAllister. Major Conformational Changes Occur during the Transition from an Initiation Complex to an Elongation Complex by T7 RNA Polymerase. *J. Biol. Chem.*, 277(45):43206–43215, November 2002.

[213] Qing Guo, Dhananjaya Nayak, Luis G. Brieba, and Rui Sousa. Major Conformational Changes During T7rnap Transcription Initiation Coincide with, and are Required for, Promoter Release. *Journal of Molecular Biology*, 353(2):256–270, October 2005.

[214] Calvin M. Schmidt, David L. Shis, Truong D. Nguyen-Huu, and Matthew R. Bennett. Stable Maintenance of Multiple Plasmids in E. coli Using a Single Selective Marker. *ACS Synth. Biol.*, 1(10):445–450, October 2012.

[215] N. Johnsson and A. Varshavsky. Split ubiquitin as a sensor of protein interactions in vivo. *PNAS*, 91(22):10340–10344, October 1994.

[216] André Galarneau, Martin Primeau, Louis-Eric Trudeau, and Stephen W. Michnick. -Lactamase protein fragment complementation assays as in vivo and in vitro sensors of proteinprotein interactions. *Nat Biotech*, 20(6):619–622, June 2002.

[217] Chang-Deng Hu and Tom K. Kerppola. Simultaneous visualization of multiple protein interactions in living cells using multicolor fluorescence complementation analysis. *Nat Biotech*, 21(5):539–545, May 2003.

[218] Stephen W. Michnick, Po Hien Ear, Emily N. Manderson, Ingrid Remy, and Eduard Stefan. Universal strategies in research and drug discovery based on protein-fragment complementation assays. *Nat Rev Drug Discov*, 6(7):569–582, July 2007.

[219] Karla Camacho-Soto, Javier Castillo-Montoya, Blake Tye, and Indraneel Ghosh. Ligand-Gated Split-Kinases. *J. Am. Chem. Soc.*, February 2014.

[220] Irina L. Grigorova, Naum J. Phleger, Vivek K. Mutalik, and Carol A. Gross. Insights into transcriptional regulation and competition from an equilibrium model of RNA polymerase binding to DNA. *PNAS*, 103(14):5332–5337, April 2006.

[221] Benjamin P. Bratton, Rachel A. Mooney, and James C. Weisshaar. Spatial Distribution and Diffusive Motion of RNA Polymerase in Live Escherichia coli. *J. Bacteriol.*, 193(19):5138–5146, October 2011.

[222] A. A. Travers and null Burgessrr. Cyclic re-use of the RNA polymerase sigma factor. *Nature*, 222(5193):537–540, May 1969.

[223] Marni Raffaelle, Elenita I. Kanin, Jennifer Vogt, Richard R. Burgess, and Aseem Z. Ansari. Holoenzyme Switching and Stochastic Release of Sigma Factors from RNA Polymerase In Vivo. *Molecular Cell*, 20(3):357–366, November 2005.

[224] Joe H. Levine, Yihan Lin, and Michael B. Elowitz. Functional Roles of Pulsing in Genetic Circuits. *Science*, 342(6163):1193–1200, December 2013.

[225] Matthew Lord, Daniela Barillà, and Michael D. Yudkin. Replacement of Vegetative A by Sporulation-Specific F as a Component of the RNA Polymerase Holoenzyme in Sporulating Bacillus subtilis. *J. Bacteriol.*, 181(8):2346–2350, April 1999.

[226] Hiroto Maeda, Nobuyuki Fujita, and Akira Ishihama. Competition among seven Escherichia coli subunits: relative binding affinities to the core RNA polymerase. *Nucl. Acids Res.*, 28(18):3497–3503, September 2000.

[227] K. F. Jensen and S. Pedersen. Metabolic growth rate control in Escherichia coli may be a consequence of subsaturation of the macromolecular biosynthetic apparatus with substrates and catalytic components. *Microbiol. Rev.*, 54(2):89–100, June 1990.

[228] Karen Montzka Wassarman and Gisela Storz. 6s RNA Regulates E. coli RNA Polymerase Activity. *Cell*, 101(6):613–623, June 2000.

[229] Anne Farewell, Kristian Kvint, and Thomas Nyström. Negative regulation by RpoS: a case of sigma factor competition. *Molecular Microbiology*, 29(4):1039–1051, 1998.

[230] Natalie A. Cookson, William H. Mather, Tal Danino, Octavio MondragónPalomino, Ruth J. Williams, Lev S. Tsimring, and Jeff Hasty. Queueing up for enzymatic processing: correlated signaling through coupled degradation. *Molecular Systems Biology*, 7(1), January 2011.

[231] Saija Haapa, Suvi Taira, Eini Heikkinen, and Harri Savilahti. An efficient and accurate integration of mini-Mu transposons in vitro: a general methodology for functional genetic analysis and molecular biology applications. *Nucl. Acids Res.*, 27(13):2777–2784, July 1999.

[232] M. Mizuuchi and K. Mizuuchi. Target Site Selection in Transposition of Phage Mu. *Cold Spring Harb Symp Quant Biol*, 58:515–523, January 1993.

[233] Brian Green, Christiane Bouchier, Cécile Fairhead, Nancy L. Craig, and Brendan P. Cormack. Insertion site preference of Mu, Tn5, and Tn7 transposons. *Mobile DNA*, 3(1):3, February 2012.

[234] Ilana Goldhaber-Gordon, Michael H. Early, Matthew K. Gray, and Tania A. Baker. Sequence and Positional Requirements for DNA Sites in a Mu Transpososome. *J. Biol. Chem.*, 277(10):7703–7712, March 2002.

[235] Insuk Lee and Rasika M Harshey. Importance of the conserved CA dinucleotide at mu termini. *Journal of Molecular Biology*, 314(3):433–444, November 2001.

[236] Eini Poussu, Mauno Vihinen, Lars Paulin, and Harri Savilahti. Probing the -complementing domain of E. coli -galactosidase with use of an insertional pentapeptide mutagenesis strategy based on Mu in vitro DNA transposition. *Proteins: Structure, Function, and Bioinformatics*, 54(4):681–692, 2004.

[237] Eini Poussu, Jussi Jäntti, and Harri Savilahti. A gene truncation strategy generating N- and C-terminal deletion variants of proteins for functional studies: mapping of the Sec1p binding domain in yeast Mso1p by a Mu in vitro transposition-based approach. *Nucl. Acids Res.*, 33(12):e104–e104, January 2005.

[238] Dafydd Jones. Polypeptide mutagenesis method, July 2006.

[239] Brigitte M. Hoeller, Birgit Reiter, Sandra Abad, Ina Graze, and Anton Glieder. Random tag insertions by Transposon Integration mediated Mutagenesis (TIM). *Journal of Microbiological Methods*, 75(2):251–257, October 2008.

[240] Derek M. Dykxhoorn, Rebecca St. Pierre, and Thomas Linn. A set of compatible tac promoter expression vectors. *Gene*, 177(12):133–136, 1996.

[241] Wayne M. Patrick, Andrew E. Firth, and Jonathan M. Blackburn. User-friendly algorithms for estimating completeness and diversity in randomized protein-encoding libraries. *Protein Eng.*, 16(6):451–457, June 2003.

[242] Celeste Yanisch-Perron, Jeffrey Vieira, and Joachim Messing. Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mpl8 and pUC19 vectors. *Gene*, 33(1):103–119, 1985.

[243] Reshma P. Shetty, Drew Endy, and Thomas F. Knight. Engineering BioBrick vectors from BioBrick parts. *Journal of Biological Engineering*, 2(1):5, April 2008.

[244] Alon Zaslaver, Anat Bren, Michal Ronen, Shalev Itzkovitz, Ilya Kikoin, Seagull Shavit, Wolfram Liebermeister, Michael G. Surette, and Uri Alon. A comprehensive library of fluorescent transcriptional reporters for Escherichia coli. *Nat Meth*, 3(8):623–628, August 2006.

[245] J. Christopher Anderson, Christopher A. Voigt, and Adam P. Arkin. Environmental signal integration by a modular AND gate. *Mol Syst Biol*, 3(1), August 2007.

[246] Robert E. Campbell, Oded Tour, Amy E. Palmer, Paul A. Steinbach, Geoffrey S. Baird, David A. Zacharias, and Roger Y. Tsien. A monomeric red fluorescent protein. *PNAS*, 99(12):7877–7882, June 2002.

[247] Jean-Denis Pédelacq, Stéphanie Cabantous, Timothy Tran, Thomas C. Terwilliger, and Geoffrey S. Waldo. Engineering and characterization of a superfolder green fluorescent protein. *Nat Biotech*, 24(1):79–88, January 2006.

[248] Heman Chao, Michael E. Houston, Suzanne Grothe, Cyril M. Kay, Maureen O'Connor-McCourt, Randall T. Irvin, and Robert S. Hodges. Kinetic Study on the Formation of a de Novo Designed Heterodimeric Coiled-Coil: Use of Surface Plasmon Resonance To Monitor the Association and Dissociation of Polypeptide Chains. *Biochemistry*, 35(37):12175–12185, January 1996.

[249] Eugene P. Ryan and Eduardo D. Sontag. Well-defined steady-state response does not imply CICS. *Systems & Control Letters*, 55(9):707–710, September 2006.

[250] Eric F. Pettersen, Thomas D. Goddard, Conrad C. Huang, Gregory S. Couch, Daniel M. Greenblatt, Elaine C. Meng, and Thomas E. Ferrin. UCSF ChimeraA visualization system for exploratory research and analysis. *Journal of Computational Chemistry*, 25(13):1605–1612, 2004.

[251] Ashty S. Karim, Kathleen A. Curran, and Hal S. Alper. Characterization of plasmid burden and copy number in Saccharomyces cerevisiae for optimization of metabolic engineering applications. *FEMS Yeast Res*, 13(1):107–116, February 2013.

[252] Jérôme Wong Ng, Didier Chatenay, Jérôme Robert, and Michael Guy Poirier. Plasmid copy number noise in monoclonal populations of bacteria. *Phys. Rev. E*, 81(1):011909, January 2010.

[253] Michael Jahn, Carsten Vorpahl, Dominique Türkowsky, Martin Lindmeyer, Bruno Bühler, Hauke Harms, and Susann Müller. Accurate Determination of Plasmid Copy Number of Flow-Sorted Cells using Droplet Digital PCR. *Anal. Chem.*, 86(12):5969–5976, June 2014.

[254] Jane Lopilato, Scott Bortner, and Jon Beckwith. Mutations in a new chromosomal gene of Escherichia coli K-12, pcnB, reduce plasmid copy number of pBR322 and its derivatives. *Molec. Gen. Genet.*, 205(2):285–290.

[255] Francis L. Macrina, George G. Weatherly, and Roy Curtiss. R6k Plasmid Replication: Influence of Chromosomal Genotype in Minicell-Producing Strains of Escherichia coli K-12. *J Bacteriol*, 120(3):1387–1400, December 1974.

[256] D. E. Cress and B. C. Kline. Isolation and characterization of Escherichia coli chromosomal mutants affecting plasmid copy number. *J. Bacteriol.*, 125(2):635–642, February 1976.

[257] Grzegorz Wgrzyn. Replication of Plasmids during Bacterial Response to Amino Acid Starvation. *Plasmid*, 41(1):1–16, January 1999.

[258] Ivo Angelov and Ivan Ivanov. Amplification of bacterial plasmids without blocking protein biosynthesis. *Plasmid*, 22(2):160–162, September 1989.

[259] A. Herman, A. Wgrzyn, and G. Wgrzyn. Regulation of replication of plasmid pBR322 in amino acid-starved Escherichia coli strains. *Molec. Gen. Genet.*, 243(4):374–378, July 1994.

[260] Borys Wróbel and Grzegorz Wgrzyn. Amplification of pSC101 replicons in Escherichia coli during amino acid limitation. *Journal of Biotechnology*, 58(3):205–208, December 1997.

[261] Zhijun Wang, Guowei Le, Yonghui Shi, Grzegorz Wegrzyn, and Borys Wrobel. A Model for Regulation of ColE1-like Plasmid Replication by Uncharged tRNAs in Amino Acid-Starved Escherichia coli Cells. *Plasmid*, 47(2):69–78, March 2002.

[262] Sue Lin-Chao, Wen-Tsuan Chen, and Ten-Tsao Wong. High copy number of the pUC plasmid results from a Rom/Rop-suppressible point mutation in RNA II. *Molecular Microbiology*, 6(22):3385–3393, 1992.

[263] S. Lin-Chao and H. Bremer. Effect of the bacterial growth rate on replication control of plasmid pBR322 in Escherichia coli. *Mol Gen Genet*, 203(1):143–149.

[264] U. E. Cheah, W. A. Weigand, and B. C. Stark. Effects of recombinant plasmid size on cellular processes in Escherichia coli. *Plasmid*, 18(2):127–134, September 1987.

[265] J. L. Corchero and A. Villaverde. Plasmid maintenance in Escherichia coli recombinant cultures is dramatically, steadily, and specifically influenced by features of the encoded proteins. *Biotechnol. Bioeng.*, 58(6):625–632, June 1998.

[266] Stephen Cooper and Charles E. Helmstetter. Chromosome replication and the division cycle of Escherichia coli Br. *Journal of Molecular Biology*, 31(3):519–540, February 1968.

[267] Dena H. S. Block, Razika Hussein, Lusha W. Liang, and Han N. Lim. Regulatory consequences of gene translocation in bacteria. *Nucleic Acids Res.*, 40(18):8979–8992, October 2012.

[268] Jack A. Bryant, Laura E. Sellars, Stephen J. W. Busby, and David J. Lee. Chromosome position effects on gene expression in Escherichia coli K-12. *Nucleic Acids Res*, 42(18):11383–11392, October 2014.

[269] Meghdad Hajimorad, Paul R. Gray, and Jay D. Keasling. A framework and model system to investigate linear system behavior in Escherichia coli. *Journal of Biological Engineering*, 5(1):3, April 2011.

[270] Alec AK Nielsen, Thomas H Segall-Shapiro, and Christopher A Voigt. Advances in genetic circuit design: novel biochemistries, deep part mining, and precision gene expression. *Current Opinion in Chemical Biology*, 17(6):878–892, December 2013.

[271] Shawn T. Estrem, Wilma Ross, Tamas Gaal, Z. W. Susan Chen, Wei Niu, Richard H. Ebright, and Richard L. Gourse. Bacterial promoter architecture: subsite structure of UP elements and interactions with the carboxy-terminal domain of the RNA polymerase subunit. *Genes Dev*, 13(16):2134–2147, August 1999.

[272] Attila Becskei and Luis Serrano. Engineering stability in gene networks by autoregulation. *Nature*, 405(6786):590–593, June 2000.

[273] Wenzhe Ma, Ala Trusina, Hana El-Samad, Wendell A. Lim, and Chao Tang. Defining Network Topologies that Can Achieve Biochemical Adaptation. *Cell*, 138(4):760–773, August 2009.

[274] Murat Acar, Bernardo F. Pando, Frances H. Arnold, Michael B. Elowitz, and Alexander Oudenaarden. A General Mechanism for Network-Dosage Compensation in Gene Circuits. *Science*, 329(5999):1656–1660, September 2010.

[275] Jesse Stricker, Scott Cookson, Matthew R. Bennett, William H. Mather, Lev S. Tsimring, and Jeff Hasty. A fast, robust and tunable synthetic gene oscillator. *Nature*, 456(7221):516–519, November 2008.

[276] Jordan Ang, Sangram Bagh, Brian P. Ingalls, and David R. McMillen. Considerations for using integral feedback control to construct a perfectly adapting synthetic gene network. *J. Theor. Biol.*, 266(4):723–738, October 2010.

[277] Shai S. Shen-Orr, Ron Milo, Shmoolik Mangan, and Uri Alon. Network motifs in the transcriptional regulation network of Escherichia coli. *Nat Genet*, 31(1):64–68, May 2002.

[278] S. Mangan and U. Alon. Structure and function of the feed-forward loop network motif. *PNAS*, 100(21):11980–11985, October 2003.

[279] Matthew F. Copeland, Mark C. Politz, Charles B. Johnson, Andrew L. Markley, and Brian F. Pfleger. A transcription activator-like effector (TALE) induction system mediated by proteolysis. *Nat Chem Biol*, 12(4):254–260, April 2016.

[280] Feng Zhang, Le Cong, Simona Lodato, Sriram Kosuri, George M. Church, and Paola Arlotta. Efficient construction of sequence-specific TAL effectors for modulating mammalian transcription. *Nat Biotech*, 29(2):149–153, February 2011.

[281] Julia M. Rogers, Luis A. Barrera, Deepak Reyon, Jeffry D. Sander, Manolis Kellis, J. Keith Joung, and Martha L. Bulyk. Context influences on TALE-DNA binding revealed by quantitative profiling. *Nat Commun*, 6:7440, June 2015.

[282] Benjamin R. Jack, Sean P. Leonard, Dennis M. Mishler, Brian A. Renda, Dacia Leon, Gabriel A. Suárez, and Jeffrey E. Barrick. Predicting the Genetic Stability of Engineered DNA Sequences with the EFM Calculator. *ACS Synth. Biol.*, 4(8):939–943, August 2015.

[283] G. J. Phillips, S. K. Park, and D. Huber. High copy number plasmids compatible with commonly used cloning vectors. *BioTechniques*, 28(3):400–402, 404, 406 passim, March 2000.

[284] Thanh Quoc Ho, Zhenping Zhong, Stefan Aung, and Joe Pogliano. Compatible bacterial plasmids are targeted to independent cellular locations in Escherichia coli. *EMBO J*, 21(7):1864–1872, April 2002.

[285] Rodrigo Reyes-Lamothe, Tung Tran, Diane Meas, Laura Lee, Alice M. Li, David J. Sherratt, and Marcelo E. Tolmasky. High-copy bacterial plasmids diffuse in the nucleoid-free space, replicate stochastically and are randomly partitioned at cell division. *Nucleic Acids Res.*, 42(2):1042–1051, January 2014.

[286] Esteban Martínez-García, Tomás Aparicio, Víctor Lorenzo, and Pablo I. Nikel. New Transposon Tools Tailored for Metabolic Engineering of Gram-Negative Microbial Cell Factories. *Front Bioeng Biotechnol*, 2, October 2014.

[287] Jesus Fernandez-Rodriguez, Lei Yang, Thomas E. Gorochowski, D. Benjamin Gordon, and Christopher A. Voigt. Memory and Combinatorial Logic Based on DNA Inversions: Dynamics and Evolutionary Stability. *ACS Synth. Biol.*, 4(12):1361–1372, December 2015.

[288] L. C. Raynal, H. M. Krisch, and A. J. Carpousis. Bacterial poly(A) polymerase: An enzyme that modulates RNA stability. *Biochimie*, 78(6):390–398, January 1996.

[289] György Pósfai, Guy Plunkett, Tamás Fehér, David Frisch, Günther M. Keil, Kinga Umenhoffer, Vitaliy Kolisnychenko, Buffy Stahl, Shamik S. Sharma, Monika Arruda, Valerie Burland, Sarah W. Harcum, and Frederick R. Blattner. Emergent Properties of Reduced-Genome Escherichia coli. *Science*, 312(5776):1044–1046, May 2006.

[290] Sean C. Sleight, Bryan A. Bartley, Jane A. Lieviant, and Herbert M. Sauro. Designing and engineering evolutionary robust genetic circuits. *J Biol Eng*, 4(1):1–20, December 2010.

[291] Alvaro R. Lara, Enrique Galindo, Octavio T. Ramírez, and Laura A. Palomares. Living with heterogeneities in bioreactors. *Mol Biotechnol*, 34(3):355–381, November 2006.

[292] S. Mangan and U. Alon. Structure and function of the feed-forward loop network motif. *PNAS*, 100(21):11980–11985, October 2003.

[293] Uri Alon. *An Introduction to Systems Biology: Design Principles of Biological Circuits*. Chapman and Hall/CRC, Boca Raton, FL, 1 edition edition, July 2006.

[294] John Tsang, Jun Zhu, and Alexander Oudenaarden. MicroRNA-Mediated Feedback and Feedforward Loops Are Recurrent Network Motifs in Mammals. *Molecular Cell*, 26(5):753–767, June 2007.

[295] Lea Goentoro, Oren Shoval, Marc W. Kirschner, and Uri Alon. The Incoherent Feedforward Loop Can Provide Fold-Change Detection in Gene Regulation. *Molecular Cell*, 36(5):894–899, December 2009.

[296] C G Lerner and M Inouye. Low copy number plasmids for regulated low-level expression of cloned genes in Escherichia coli with blue/white insert screening capability. *Nucleic Acids Res*, 18(15):4631, August 1990.

[297] Tetsushi Sakuma, Hiroshi Ochiai, Takehito Kaneko, Tomoji Mashimo, Daisuke Tokumasu, Yuto Sakane, Ken-ichi Suzuki, Tatsuo Miyamoto, Naoaki Sakamoto, Shinya Matsuura, and Takashi Yamamoto. Repeating pattern of non-RVD variations in DNA-binding modules enhances TALEN activity. *Sci Rep*, 3:3379, 2013.

[298] Changsoo Lee, Jaai Kim, Seung Gu Shin, and Seokhwan Hwang. Absolute and relative QPCR quantification of plasmid copy number in Escherichia coli. *Journal of Biotechnology*, 123(3):273–280, May 2006.

[299] Sankar Das, Jody C. Noe, Sehmi Paik, and Todd Kitten. An improved arbitrary primed PCR method for rapid characterization of transposon insertion sites. *Journal of Microbiological Methods*, 63(1):89–94, October 2005.

[300] Cheuk C. Siow, Sian R. Nieduszynska, Carolin A. Müller, and Conrad A. Nieduszynski. OriDB, the DNA replication origin database updated and extended. *Nucleic Acids Res*, 40(Database issue):D682–D686, January 2012.