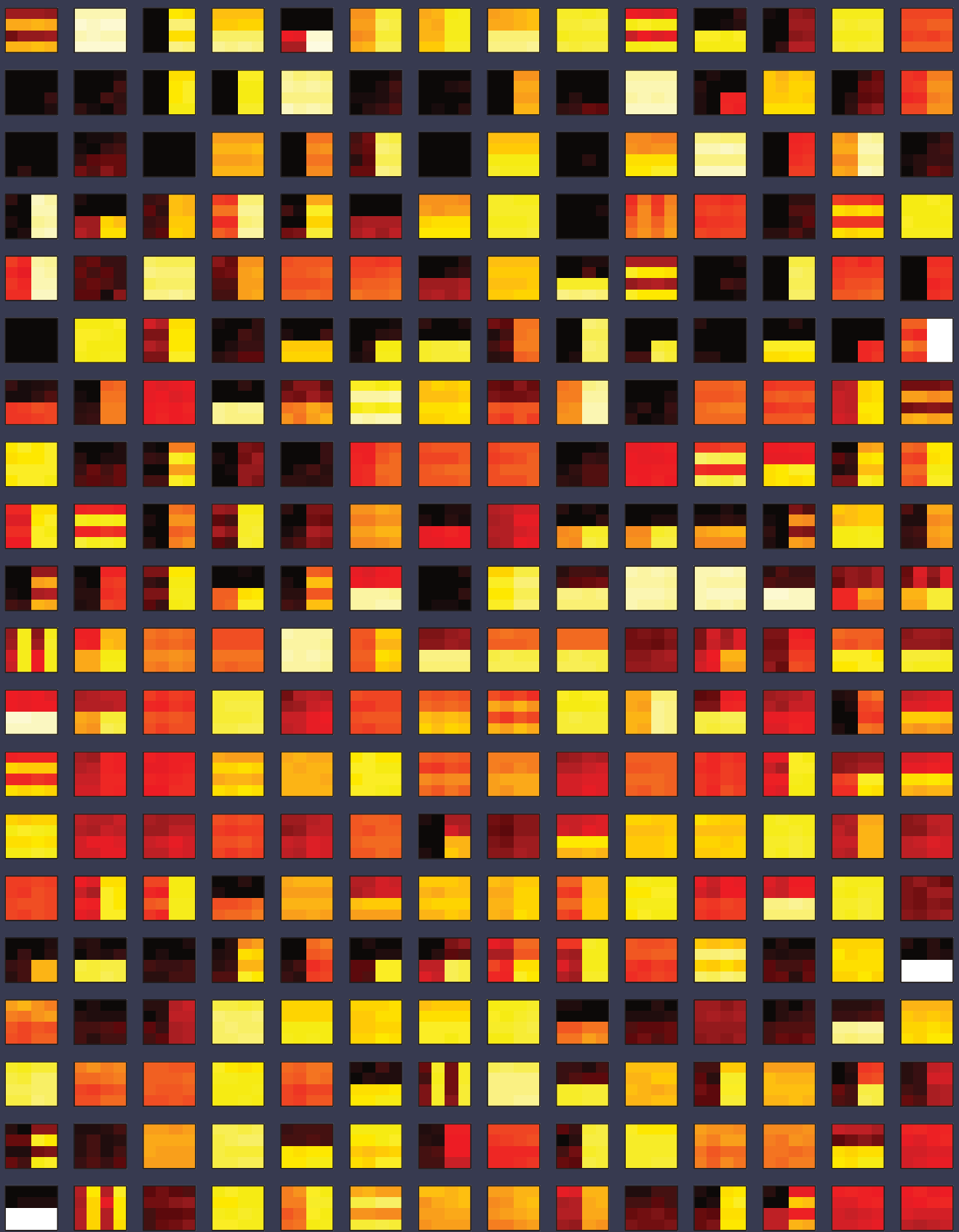
Transcriptional regulation and combinatorial genetic logic in synthetic bacterial circuits



TRANSCRIPTIONAL REGULATION AND COMBINATORIAL GENETIC LOGIC IN SYNTHETIC BACTERIAL CIRCUITS

Thesis by
Robert Sidney Cox III

In Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy

California Institute of Technology
Pasadena, California
2008
(Defended October 12, 2007)

Acknowledgements

My three mentors: Frances Arnold, Mike Surette, and Michael Elowitz, provided much of the inspiration and motivation for every aspect of the work presented here. Frances taught me "what" to do, Mike taught me "how" to do it, and Michael taught me "why" it was worth doing.

Mike Surette provided unfailing guidance for the experimental work of Chapters 1 and 2 and was always a bastion of calm assurance. Aaron White helped immensely with the design and construction of the DNA library presented in Chapter 1 in Mike's lab at the University of Calgary. I would like to thank Mike for letting me spend three amazing winter months working in his lab with Aaron.

Mercedes Paulino did the layout of the entire book, along with every table and every figure. She carefully chose each font and header, arranged the various figures, and made everything easy to find and read. Most of all, she transformed my dissertation into a living work of art.

My undergraduate mentors at New College, the Honors College of Florida: Pat McDonald (Mathematics), George Ruppeiner (Physics), and Paul Scudder (Chemistry) prepared me for the academic hurdles of graduate school so well that I barely noticed them. Christof Koch, Erik Winfree, Chris Voigt, Yohei Yokobayashi, Manish Raizada, Jared Leadbetter, Christina Smolke, Paul Sternberg, Barbara Wold, Danny Rintoul, David Womble, Ron Weiss, Drew Endy, Stuart Kauffman, and Michael Savageau provided helpful advice and guidance in my development as a scientist.

Mary Dunlop, working in Richard Murray's group, developed the cross correlation analysis method used for Figure 4 in Chapter 3. C. Davidson, J. Yang, C. Vizcarra, Y. Wang, R. Georgescu, S. Thiberge, F. Balagadde, C. Collins, S. Maerkle, and C. Kooi provided excellent technical assistance and helpful discussions. Georg Seelig and Jonmin Kim taught me to image DNA gels in Erik Winfree's lab (Chapter 1). Georg also assisted with the construction of the phage P22 and phage 434 circuits in Appendix D. Avigdor Eldar was invaluable for inventing the data representations for two key figures (Figure 2, Chapter 3 and Figure 4, Chapter 2). Jordi Garcia-Ojalvo lead the theoretical charge for the synchronized repressiliator circuits in Appendix C, and became (patiently) my first student of experimental science. Takeshi Irie, Graham Anderson, Dylan Morris, Andrea Choe, Beth Orcutt, Robert Sidney Cox Jr., James Locke, and Chiraj Dalal provided useful feedback on multiple versions of this document.

As my thesis advisor, Michael Elowitz guided every aspect of my advancement as a scientist. He introduced me to quantitative fluorescence microscopy, and spent countless hours teaching me how to assemble, use, and fix the microscopes and other experimental equipment. He also taught me to write in a focused and clear manner (and I hope this will be apparent in the following pages). Above all, he imprinted in me the ability to identify which scientific questions I was in which I was most truly interested. He taught me to constantly revise and focus my questions, and how to conduct research in an independent manner. For these things I cannot thank him enough.

This book was written in memory of Robert Sidney Cox Sr., and is dedicated to my high school biology teacher, Jim "Snakeman" Stevenson.

Abstract

We engineered several synthetic regulatory circuits to study transcriptional regulation in bacteria. We developed a new technique for DNA construction, built and characterized *in vivo* a library of genetic logic gates, examined the role of genetic noise transcriptional regulation using a fluorescent multi-reporter system, and characterized a synthetic circuit for the control of population density.

We used synthetic duplex DNA fragments and very short cohesive overhangs to direct ordered assemblies of diverse combinatorial libraries. Multiple DNA fragments were simultaneously ligated in a single step to produce random concatemers, without the need for amplification or product purification. We characterized the assembly process to identify optimal cohesive overhangs. We showed that the method was 97% efficient for assembling 100 base-pair concatemers from three duplex fragments.

We constructed a library of 10,000 prokaryotic promoters, containing over 1,000 unique 100 base-pair sequences. These promoters responded to up to three inputs, and contained diverse architectural arrangements of regulatory sequences. We characterized the logical input functions of 288 promoters in *Escherichia coli*, and analyzed the relationship between promoter function and architecture. We defined promoter function in terms of regulatory range, logic type, and input symmetry; and identified general rules for combinatorial programming of gene expression.

We built a synthetic three-color fluorescent reporter framework. This construct was non-toxic and extensible for synthetic and systems biology applications. Three spectrally distinct and genetically isolated reporter proteins allowed independent monitoring of genetic signals at the single-cell level. We showed that the simultaneous measurement of multiple genes can exploit genetic noise to sensitively detect transcriptional co-regulation.

We built and characterized a 'population control' circuit that autonomously regulated the density of an *E. coli* population. Cell density was broadcasted and detected by elements from a bacterial quorum sensing system, which regulated the death rate. The stable cell density steady-state was tuned by varying the stability of the quorum signal. This synthetic circuit coupled transcriptional regulation with population-level behavior.

Table of Contents

Introduction

We recognize life by its ability to adapt to stimulus. The critical information for each adaptive function is stored within DNA. With the possible exception of the very simplest obligate parasites (Perez-Brocal et al. 2006), a cell must choose which genes to express at any given time. Some genes can directly interfere with each other, or create toxic compounds when co-expressed. More importantly, a cell must make efficient use of the metabolic resources needed to express its genes. When, where, and how much of each gene must take into account the cell's internal representation of its environment—along with its prediction of how the environment will change. Every free-living organism can adapt to countless environmental stimuli by controlling the expression of between about 1,000 *(Aquifex aeolicus)* and 45,000 *(Populus trichocarpa)* genes.

Transcription initiation is the key step for genetic regulation in prokaryotes (Browning and Busby, 2004). Every gene is a candidate for transcriptional regulation, since all are transcribed from DNA to RNA, and genes controlled at this step will only expend the cell's metabolic resources (for transcription, translation, etc.) as they are needed. The extremely low copy number of the chromosome (~2.1 in exponentially dividing *Escherichia coli*) permits a very large range of transcriptional regulation (e.g., LacI, present at ~10 copies per cell, can regulate expression over 1,000-fold). This regulation takes place at the promoter region, upstream of the coding and translational sequences, where it is possible to control transcription independently of other forms of regulation.

In this thesis, I study how multiple specific transcription factors (TFs) combine to regulate prokaryotic gene expression. The approach taken is that of "synthetic biology": using synthetically constructed combinations of genetic sequence elements to probe cellular behaviors. I present a *promoter shuffling* method, described below and in Chapter 1, and use it to construct a random library of genetic control elements. These simple circuits integrate responses to a handful of environmental signals (inducers) that do not interact in nature, and allow us to focus on the general rules that limit combinatorial regulation at promoters (Chapter 2). Promoters regulated

by one or two specific TFs are used to characterize a synthetic reporter scaffold fit to study the expression of three genes simultaneously (Chapter 3). Finally, I describe a different form of combinatorial regulation—cell-cell communication—in the context of a synthetic transcriptional regulation circuit that dynamically controls cell density (Chapter 4).

Here we employ specific TFs, as opposed to the global TFs and σ factors that regulate transcription at a higher level (similar to the chromatin remodeling systems in eukaryotes). The catabolite regulation protein CRP, a ubiquitous global TF, regulates a third of the 533 genes in *E. coli* for which transcriptional regulation has been characterized (Salgado et al. 2006). A perturbation of CRP activity represents a major stress to the cell (Perrenoud and Sauer, 2005), and understanding its regulatory interactions would involve decomposing hundreds of interactions. Since the specific TFs are less important to the cell, we can manipulate and modify their behaviors without massively perturbing the organism. Global TFs are often exceptions to the rules that govern specific TFs, but as we shall see below, some of the general principles will still apply.

The scientific contribution of this thesis is contained almost wholly in Chapter 2, and is the focus of this introduction. Chapters 1, 3, and 4 are contributions for biological engineering. After a brief glossary of transcriptional regulation, I discuss the current state of engineering and measurement of gene expression. This highlights the need for an integrated system for measuring gene expression, which is presented in Chapter 3. Chapter 4 is an example of a synthetic regulatory circuit. I review several other examples from the synthetic biology literature, to motivate the utility of the combinatorial promoters presented in Chapter 2.

Combinatorial promoters are the focus of the last two sections of this introduction. There I give a general description of combinatorial regulation, and show how the results of Chapter 2 can be used to understand natural combinatorial promoters. The DNA construction method of Chapter 1 provides the tools to extend this analysis to other combinations of TFs, and other organisms.

Glossary: The Building Blocks of Prokaryotic Transcriptional Regulation

In this glossary, I define the principle genetic elements of prokaryotic transcriptional regulation. Examples and statistics are also given for each element, to provide context for the reader. This section can be skipped by those already familiar with prokaryotic transcription.

Promoter: Region of regulatory DNA just upstream of a gene. The promoter region is 100 bp extending from -75 to +25, including the first transcribed base (+1). Promoters contain multiple sequence features that interact with the RNA polymerase subunits and the TF regulators. There are between 3,000 and 12,000 functional promoter sequences in *E. coli*, and countless "cryptic," or non-functional, ones (Huerta and Collado-Vides, 2003). Promoters comprise about 10% of the total genome size. When a gene has no known TF regulators, or the TFs are not present in the cell, it is said to be constitutive. Combinatorial promoters are regulated by two or more TFs. 220 combinatorial promoters have been annotated in *E. coli* (Figure 1), but there undoubtedly many more. It is common for more than one promoter to regulate the same gene, such as when two closely spaced promoters direct transcription in the same direction (a "tandem" promoter).

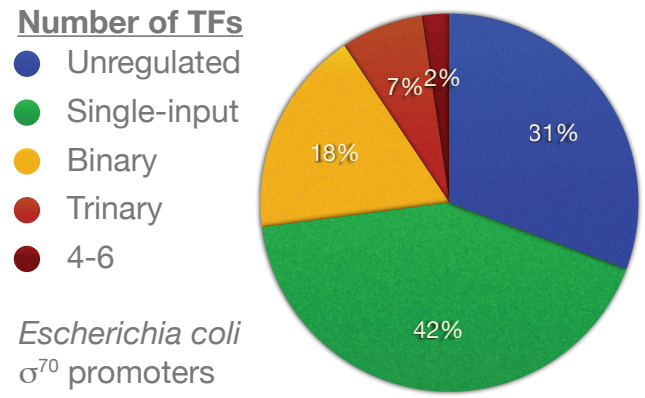


Figure 1. Combinatorial promoters are common in bacteria. The frequencies of combinatorial regulation are shown for the 554 σ^{70} promoters of *E. coli* as annotated in RegulonDB 5.0.

Terminator: Region of regulatory DNA just downstream of a gene. This variable-length region is usually less than 100 bp in length, and always contains a self-complementary hairpin region followed by an AT rich region. In very compact genomes (phages), this AT rich region may

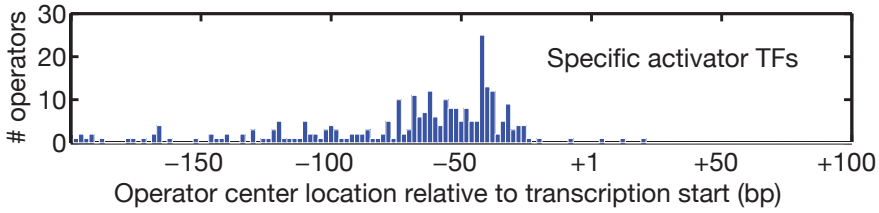
contain the promoter of a downstream gene. Very strongly transcribed genes, such as the rRNA genes, often contain stacks of three or more terminators to insure efficient termination. The activity of terminators is sometimes regulated by protein factors such as ρ, though this is not nearly as common as TF regulation at promoters.

Operon: Transcriptional unit containing genes, promoters, and terminators. This is usually extended to include regulatory sites which lie may lie outside of the promoter, coding, and termination regions. Multiple (tandem) promoters that transcribe the same gene(s) are considered part of the same operon, even when one promoter is located between two genes. A minimal operon contains just one promoter, gene, and terminator.

σ *factor*: Subunit of bacterial RNA polymerase which recognizes the promoter region. Bacteria typically contain between 3 and 30 σ factors (Rodionov, 2007). About half of all promoters are typically recognized by the "housekeeping" σ factor, called σ^{70} in *E. coli*. These promoters express the genes necessary for normal growth: polymerases, ribosomes, membrane synthesis, DNA replication and repair, respiration, amino acid metabolism, etc. Most combinatorial promoters are recognized by this σ factor. Every promoter contains a -10 box for σ binding (σ^{70} consensus TATAAT) and most promoters contain a second -35 box also bound by σ (σ^{70} consensus TTGACA). The spacing between the -10 and -35 boxes in σ^{70} promoters is strongly conserved at 17 ± 1 bp. The σ factor is released from the RNA polymerase complex during transcription elongation.

TF: Transcription factor. A protein that binds to DNA, usually near the promoter, to regulate the initation of transcription. While the σ's determine which genes can be expressed, the TFs decide which genes are actually needed. TFs usually act by regulating recognition of the promoter sequence by RNA polymerase. The seven global TFs (CRP, IHF, FNR, ArcA, Fis, NarL, Lrp) regulate about half of all characterized σ^{70} promoters, and most combinatorial promoters. In *E. coli* there are about 150 specific (non-global) TFs known, and another 150 putative TFs based on

Figure 2. Activator operators cluster upstream of -35.



The operator location for 298 specific activator TFs. The histogram plots the number of operators centered at each location, relative to the start site of transcription (+1). 60% of activator operators occur between -30 and -75.

homology. In free-living bacteria, TFs typically comprise 5-10% of the total number of genes (Rodionov, 2007).

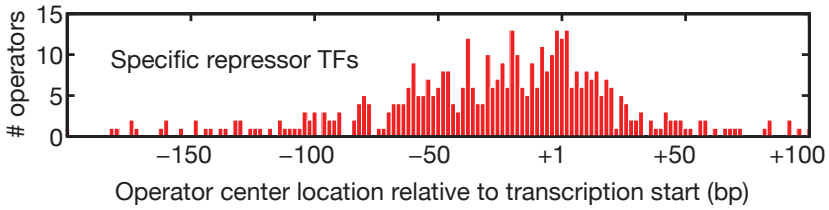
Operator: TF binding site. These usually occur within the promoter region. Most operators are 12-24 bp in length, so a maximum of about 6 (non-overlapping) operators can fit within the promoter region. In special cases, multiple TFs can recognize the same operator (e.g., λ cI and cro repressor (Ptashne, 2004).

Activator: TF that increases gene expression, usually by contacting and "recruiting" the RNA polymerase to the promoter. Specific activators normally bind between -35 and -75 (Figure 2), and increase the rate of promoter recognition by binding to the σ and α RNA polymerase subunits. Most activators respond positively to environmental inducers or co-factors; which modulate DNA binding, protein folding, stability, or conformation. Some activators (e.g., λ cI) act by increasing RNA polymerase isomerization and DNA melting. All seven global TFs are activators.

Repressor: TF that decreases gene expression, usually by steric occlusion, where the repressor blocks the RNA polymerase from binding to the promoter. Specific repressors normally bind between +25 and -75 (Figure 3), and decrease the rate of promoter recognition. Most repressors respond negatively to environmental inducers or co-factors, which modulate DNA binding. Some repressors act by blocking transcription elongation (i.e., "roadblock"), or by changing the conformation of the DNA. All global TFs are also repressors.

Figure 3. Repressor operators span the promoter region.

The operator location for 479 specific repressor TFs. The histogram plots the number of operators centered at each location, relative to the start site of transcription (+1). 70% of repressor operators occur between +25 and -75.



Measuring Gene Expression: Choice of Reporter and System

In this section, I describe the measurement of gene expression and regulation. I present a summary of qualitative differences between each measurement system, and focus on how to select the appropriate tool for a given application. This section can be skipped by those familiar with genetic reporter systems.

Reporter genes

To begin any discussion of how we might *control* gene expression, it is necessary to understand how we measure it. Unfortunately, there is no direct way to compare the diverse forms of gene expression data. As we shall see below, even ratios of expression measurements are unreliable. Still, we can obtain insight by comparing measurements within the same experimental system. The primary tool for monitoring gene expression is the genetic reporter. Three types of genetic reporters are employed in this work: the β-galactosidase enzyme from the lactose metabolism operon (LacZ), the light-producing bacterial luciferase operon from *Photorhabdus luminescens*, the Green Fluorescent Protein (GFP) and its relatives. Each reporter is naturally suited to different applications (Table I). Detailed descriptions of each reporter can be found in their associated reviews (Meighen, 1991; Miller, 1972; Tsien, 1998).

LacZ is the traditional reporter of gene expression taken from bacteria, and for good reason. It is an extremely stable enzyme, even when expressed in eukaryotes, and is well-characterized biochemically.

LacZ is extremely sensitive, and level of signal measured is linearly proportional to the amount of LacZ present. Typically, the catalytic activity of LacZ is measured in a cell culture lysate by the fluorescence of a cleaved substrate molecule. This results in signal amplification (one LacZ enzyme can catalyze many cleavage reactions), while averaging the signal across the entire cell population. The sensitivity is ultimately determined by the background fluorescence level of the measurement condition. In optimal conditions, a single LacZ molecule can be observed inside individual living cells (Cai et al. 2006). The upper end of the dynamic range is limited by the toxicity of LacZ overexpression. LacZ is a large gene (which are often toxic at high expression levels), and is slow to mature. In one comprehensive study, LacZ was unable to distinguish differing promoter response times revealed by GFP and luciferase measurements (Zaslaver et al. 2004). Commonly, quantitative measurements can be made over three to four orders of magnitude (Oehler et al. 1994). LacZ is an

Table I. Genetic reporter properties

Reporter gene family	*gfp*	*lacZ*	*luciferase*
In vivo reporter	Yes	No[1]	Yes
Requires exogenous substrate	No	Yes	No[2]
Single gene (protein fusions)	Yes	Yes	No[3]
Signal maturation	Fast[4]	Slow	Fast
Toxicity	Moderate	High	Low
Max sensitivity (molecules per cell)	~50	~1	< 1
Dynamic range	Moderate	Moderate	High
Single-cell measurements[5]	Yes	No	No
Multiple color variants	Yes	No	No
Linearity	Good	Good	Unknown
Assay type	Fluorescence	Fluoresence	Luminescence
Signal Stability[6]	High	Very high	Very Low
Protein activity	Intrinsic	Catalytic	Catalytic
Oligomerization[7]	Monomer	Homotetramer	Heterodimer

1 *In vivo* measurements are possible with special apparatus (Cai et al. 2006).
2 Bacterial luciferase requires an aldehyde substrate when only *luxAB* are expressed.
3 A protein fusion of *luxA* and *luxB* has been developed which could be used for protein tagging and expression in eukaryotic organisms (Lim, 1996).
4 GFP maturation varies dependent on the gene, organism, and growth condition; from minutes (Venus YFP), to many hours (DsRed).
5 LacZ and luciferase expression can be measured in single cells with special apparatus.
6 Refers to protein or catalytic acitivty, as appropriate. Signal stability can be decreased (along with toxicity) by the addition of ssrA degradation tags to the end of the protein(s). GFP photobleaching rates are highly variable.
7 GFP is naturally a homodimer, and DsRed is naturally a homotetramer. Stably fluorescent monomers have been engineered for most color variants.

optimal reporter for measuring low to moderate levels of steady-state gene expression across populations of cells, or for single-cell measurements of very low levels of expression.

The luciferase operon is ideal for measuring fast responses and large dynamic ranges. No excitation needed for luminometry: the sensitivity is determined by the catalytic efficiency of luciferase, the efficiency (and "dark-current") of the detection apparatus, and the measurement interval. The luciferase genes are remarkably non-toxic. Full induction of very strong promoters has little effect on the growth of *E. coli*. It is possible to measure gene expression over five orders of magnitude with this system (Bjarnason et al. 2003). Since the luciferase operon consists of five genes, it is possible for the rate-limiting step in light production to change at different expression levels, resulting in nonlinear response. For population-level measurements, luciferase expression can be detected at less than one enzyme per cell. As with LacZ, specific systems have been developed to maximize sensitivity for single-cell measurements of gene expression (Balaban et al. 2004). Luciferase is an unstable reporter, and highly dependent on growth conditions. It is therefore important to carefully design measurement assays to maximize repeatability. For example, in Chapter 2, growing cells at 25 °C greatly decreased replicate variation when compared to similar measurements of cells growing at 37 °C. With a well-designed assay, luciferase is an ideal reporter for measuring gene expression over extreme ranges of response.

The fluorescent protein reporters, showcased by the GFP family from *Aequorea victoria*, have several practical advantages over catalytic reporters such as luciferase and LacZ. The abundance of natural (Matz et al. 2002) and engineered (Campbell et al. 2002; Nagai et al. 2002; Rizzo et al. 2004) fluorescent proteins enables simultaneous monitoring of multiple genes inside living cells. We show in Chapter 3 that it is possible to spectrally separate red, yellow, and cyan color variants (RFP, YFP, and CFP, respectively) with minimal (< 0.1%) spectral crosstalk. Fluorescence is an inherently localized property; so GFP can be observed at the level of protein dynamics and diffusion (Elowitz et al. 1999), or localized to cellular compartments. Typical maturation times vary from 10

minutes to 10 hours, and require only the presence of oxygen. The response of certain GFP variants, such as Venus YFP (Nagai et al. 2002), can be observed within minutes of gene induction. GFP is active as a monomer (though naturally a dimer), and can give a quantitative linear response over four orders of magnitude (Zaslaver et al. 2006). GFP is a very stable protein, and it is possible to extract promoter activity directly from the rate of accumulation in both single-cells (Rosenfeld et al. 2005) and populations (Setty et al. 2003). Though it is sensitive to environmental and cellular conditions such as pH, engineered variants have been made to resist (or exploit) these effects along with decreased photobleaching response. GFP is moderately toxic. The sensitivity of detection is limited ultimately by the autofluorescence background levels, which can be decreased by performing measurements in non-fluorescent media. For many applications GFP is the obvious choice: it is only inappropriate for measuring very high and very low levels of expression.

Reporter gene systems

It is impossible to measure a genetic system without perturbing it, and the impact on the system studied should be kept in mind. The primary system used in this thesis is the promoter fusion. Promoter fusions employ a copy of the target promoter to control the genetic reporter directly. Very weak signals can be observed by using high-copy plasmids, or by stacking multiple copies of the reporter gene in an operon. The main disadvantage of promoter fusions is that the additional copies of the promoter can titrate TF protein molecules away from their natural target, resulting in perturbed regulation. In most cases, high copy number reporter plasmids effectively reduce the range of gene expression (Lutz and Bujard, 1997). In Chapters 2 and 3 we minimize this difficultly by using sensitive reporters, a very low copy plasmid, and high concentrations of the TF regulators. Promoter fusions are the obvious choice for measuring synthetic promoters.

Other common reporter systems are protein fusions and operon fusions. Unlike promoter fusions, no additional copies of the promoter need to be introduced. Protein fusions require the reporter protein sequence to be expressed in the same frame as the gene of interest, usually connected to

the N- or C-terminal ends of the target protein by a flexible ("linker") peptide sequence. Protein fusions ensure quantitative control of the protein copy number, and the reporter is localized along with the protein sequence. The reporter tag can interfere with the activity of the protein target, and the increased gene size will often decrease the translation level. At times, this can be a benefit. In Chapter 4, we used a protein fusion of the catalytic domain of LacZ and the control of cell death protein ccdB. We wished to observe slow, population level changes, and the decreased ccdB-LacZ activity made it easier to control: variants with higher ccdB activity were genetically unstable. The last type of reporter system is an operon fusion: the reporter gene is placed within the operon containing the target gene. The primary disadvantage here is that the introduced coding sequence can change the stability and translation level of the mRNA transcript. The advantages and drawbacks of each reporter system should be weighed to determine the minimal perturbation for a given experiment.

Engineering Gene Expression: Components and Circuits

In this section I discuss our ability to control gene expression. Here I focus on engineered interactions between promoters and TFs, and highlight some limitations of the rational design approach. I review the field of synthetic biology, which has produced a multitude of regulatory circuits from a relatively small number of components. I believe that understanding synthetic combinatorial promoters will allow us to engineer even more complex synthetic circuits, and to better understand their natural counterparts.

The range of gene expression

The same promoter-TF interaction can produce wildly different quantitative values when measured with different reporter systems. Here we look at the *fold-change*, which we represent as *r*. For inducible promoters, *r* is simply the ratio of the fully induced versus uninduced levels of promoter activity. In Chapter 3, we measure *r* for three tightly regulated promoters with a different

fluorescent protein for each. When measured under similar conditions, these three promoters gave different values for r depending on the genetic reporter used (Table II). For the fluorescent reporters, only the brightest (*yfp*) gave a similar value for r. These differences illustrate the difficulty in ascribing absolute numbers to measurements of gene expression.

Table II. Measured fold-change depends on genetic reporter

Promoter	r (firefly luciferase[1])	r (fluorescence[2])	r (bacterial luciferase[3])
p*(tet)*	5050	30 (*cfp*)	6085
p*(lac)*	1185	1000 (*yfp*)	935
p*(lac/ara)*	620	60 (*rfp*)	971

1 (Lutz and Bujard, 1997)
2 Chapter 3, Figure 2. The fluorescent protein color variant gene is given in parentheses. Reporter gene sequences are in Appendix B.
3 Values computed from similar promoter sequences in Chapter 2.

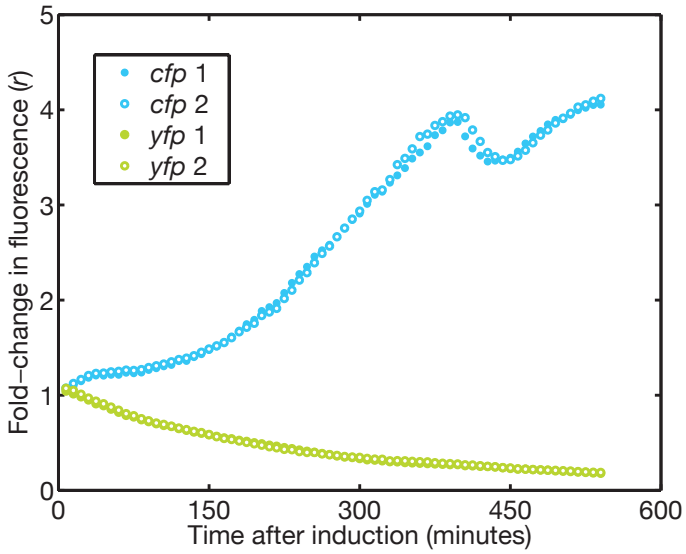
Very large fold changes in gene expression are possible. In Chapter 2, we will characterize promoters that exhibit very large fold-changes—up to $r = 10^5$. Assuming a linear correspondence, this change is equivalent to the difference between promoter activities of one transcript per second versus one transcript per day. In terms of proteins, this is the difference between 10% of the total cell mass (half of which is protein) versus ~5 copies per cell. Such strong regulation is highly desirable for engineering and synthetic biology applications. Though we cannot measure definitive values of r for every promoter, we can make relative comparisons within the same experimental system.

Genetic isolation

With the ability to engineer such huge changes in gene expression, we must be mindful of the perturbation this represents to the cell. As an illustrative example, consider two fluorescent proteins CFP and YFP whose expression levels can be independently induced by the exogenous inputs I_c and I_y. Since the regulation is independent, input I_c should not affect the level of YFP.

Independence can break down when the levels of expression are large (Figure 4). In this experiment, we track the fold-change in fluorescence as a function of time after I_c (IPTG) is added. Strikingly, the level of YFP *decreases* as CFP increases. This points to a saturation mechanism, in which the two genes compete for the metabolic resources of gene expression. For this experiment, we added

Figure 4. Over-expression results in spurious co-regulation.



a strong ssrA degradation tag to both fluorescent proteins (Gottesman et al. 1998) so that they would not accumulate. If we had saturated the degradation machinery instead of expression (Buchler et al. 2005), we would see YFP increase when CFP is induced. This simple example shows how over-expression can induce spurious co-regulation between two independently controlled genes.

In order to build a reliable synthetic biology discipline, we want to engineer systems in which each interaction is prescribed. To address this, Chapter 3 presents a synthetic scaffold for synthetic and systems biology. We show that the expression levels of the three reporter proteins can be varied independently when the copy number of the scaffold is kept low. Transcriptional terminators are common in the genome for a reason, and the strongest rRNA promoters are often flanked by stacks of three or more terminator sequences. Mirroring this, we isolated each gene in the scaffold with multiple terminator sequences. In the low copy context, we were able to detect correlations between the three reporters and derive regulatory information from them. This scaffold is a quantitative framework for measuring and engineering regulatory circuits, such as the examples reviewed below.

Synthetic biology: models of regulatory circuits

Promoters and TFs have been used to build a host of synthetic biological circuits. Each of these model circuits can illustrate a principle of regulation. Cascades of repressors demonstrate the principle of ultra-sensitivity, whereby a small change in an upstream input is amplified at each step of the cascade (Hooshangi et al. 2005). Two classic studies used repressors to engineer bi-stable and oscillatory responses. When two repressors are connected such that each represses the other, a positive feedback loop results. Here, only one of the two promoters can be active at a time, so the initial state of the system determines which of the two stable states it eventually reaches. For two stable repressors, where the TF concentration can only decrease from dilution due to cell growth, the state of this "toggle switch" can persist indefinitely (Gardner et al. 2000). This model circuit exemplifies one of the simplest forms of differentiation, whereby genetically identical cells can persistently express different sets of genes. Three repressors connected circularly form a "ring oscillator" circuit (Thomas and D'Ari, 1990). This system demonstrates negative feedback, and three unstable repressors connected in this way generate persistent oscillations in living cells (Elowitz and Leibler, 2000). An even simpler example of negative feedback is autorepression, where a repressor controls its own promoter (Rosenfeld et al. 2002). Though this system does not exhibit oscillations, it does allow a "speed-up" in the accumulation of the repressor gene: an auto-repressed gene will reach its steady-state level faster than a constitutively expressed gene with the same steady-state level. Conversely, autoactivation is a positive feedback system that can also display bi-stability (Xiong and Ferrell, 2003). These simple regulatory circuits are minimal models of the regulatory phenomena found in natural systems.

Combinatorial promoters are commonly found in natural circuit "motifs" such as the feed-forward loop (Shen-Orr et al. 2002). Here, a TF X controls the expression of a second TF Y. The two TFs X and Y then combine to regulate the promoter of a third (often metabolic) gene Z. The combinatorial promoter plays a key role in feed-forward loop function: the signs of regulation (activation or repression) along with the *logic* of the promoter (e.g., AND or OR) determines the

Table III. Combinatorial circuits can generate unexpected logic

Logic class	Inputs (1 / 2)				Predicted?
	-/-	+/-	-/+	+/+	
NOR	ON	OFF	OFF	OFF	Yes
NOT IF	OFF	OFF	ON	OFF	Yes
IMPLIES	ON	OFF	ON	ON	No
NAND	ON	ON	ON	OFF	No

dynamical behavior of the system (Mangan and Alon, 2003). In one synthetic feed-forward loop, the activator LuxR controlled the expression of the repressor cI (Basu et al. 2004). Both LuxR and cI regulated the expression of the output (GFP) at a combinatorial promoter. This circuit behaved as a "pulse-generator": a sudden increase in signal level activates GFP and cI simultaneously. As the level of cI repressor increases, the combinatorial promoter is shut off resulting in a burst of GFP production. This combinatorial promoter has also been used to create synthetic band-pass detector and a cellular pattern formation circuit (Basu et al. 2005). Such combinatorial promoters are powerful tools for engineering novel synthetic circuits.

There are many regimes for combinatorial control of gene expression besides the combinatorial promoter. One method is to integrate two inputs with a genetic circuit. With three specific TFs (LacI, TetR, and λ cI) and five single-input promoters, (Guet et al. 2002) built and characterized the responses of 30 combinatorial circuits containing random connections between the three TFs. Using two inducers as inputs (of LacI and TetR) along with a GFP output (controlled by cI), they found a variety of Boolean-like circuit functions. Four logic classes were found when a single output threshold was applied to each gate (Table III). Surprisingly, two of these logic classes could not be explained by the connections in their network diagrams. This work shows that a large diversity of combinatorial functions can be computed with randomly assembled networks. However, the method is not modular: a given logic function—especially the unexpected ones—cannot be built from of a different set of TFs and promoters connected in the same way.

In a recent study, (Anderson et al. 2007) constructed a logical AND gate by combining transcriptional

and translational regulation. Here each input controls a separate promoter. One promoter controls the transcription of a T7 RNA polymerase gene with two internal amber stop codons that block translation. The second promoter controls an amber suppressor tRNA. With both promoters active, the T7 RNA polymerase activates a T7 output promoter. This circuit is highly modular: any two promoters can be used as inputs, and the output expression level is the logical AND of their activities. However, only one copy of the T7 RNA polymerase and amber suppressor gene can be used within a single cell. This circuit achieves modularity at the expense of reusability.

Combinatorial promoters are the most direct way to engineer signal integration. In a systematic "bottom-up" study; (Guido et al. 2006) engineered a circuit composed of an activator (cI), a repressor (LacI), and a combinatorial promoter. The authors examined the response of the combinatorial promoter to each input separately, the combinatorial two-input response, and then characterized the two-input response within a simple feedback circuit. A mathematical model explained both the mean response and single-cell variation as a function of two input inducer concentrations. This model quantitatively predicted the effect of a large perturbation—change in circuit copy number—on the behavior of the feedback circuit. By characterizing each component separately, one could hope to build the same circuit out of two different TFs. This approach is modular and reusable: a second version of the circuit with different TFs (e.g., LuxR and TetR) could be used simultaneously inside a single cell.

We have a strong understanding of how each TF works (Browning and Busby, 2004), but lack the ability to quantitatively engineer the response of combinatorial promoters. In the next section I present a simple formalism for describing combinatorial regulation, and use it to examine several example combinatorial promoters. In Chapter 2, I report an initial set of rules for programming simple combinatorial promoters. By applying this knowledge, we can hope to build ever more complex and robust transcriptional networks.

Logic-Symmetry Space: The Combinatorial Promoter Phenotype

In this section, I introduce an intuitive description of combinatorial promoter phenotype. I first describe the "logic" of a promoter in terms of its relative output levels, and show how to calculate a set of three phenotype parameters (see Chapter 2 Methods for a condensed mathematical discussion). As we shall see, these parameters provide a way to compare combinatorial promoters—even between different measurement systems. With these simple equations, I analyze several example combinatorial promoters chosen from the literature and from Chapter 2. I show how one of the rules for programming combinatorial promoters (presented in Chapter 2) can provide insight into natural regulation.

Promoter logic is not Boolean

The response of a promoter is a peculiar kind of logic gate. We consider the form of a generic promoter response at the population level. For a promoter regulated by two inducible TFs, we take the gate inputs to be the corresponding inducer concentrations. A typical promoter response is shown (Figure 5). Here, the promoter activity is approximately constant at four different levels, called "plateaus." To extend the idea of a logic gate, we take the presence of these plateaus as a general axiom. Since the response to each input is sharp, it makes sense to consider the inputs as digital (0 or 1). In contrast, the output level is continuous. One way to conceptualize the output would be to set an arbitrary threshold to distinguish an OFF response from an ON (Figure 6). The problem with this type of Boolean reasoning is that you get a different function dependent on where the threshold is set: in the example promoter (Figure 5) a low

Figure 5. Combinatorial promoter responses exhibit plateaus.

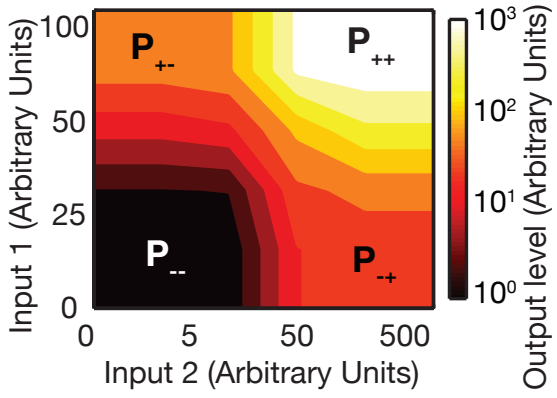
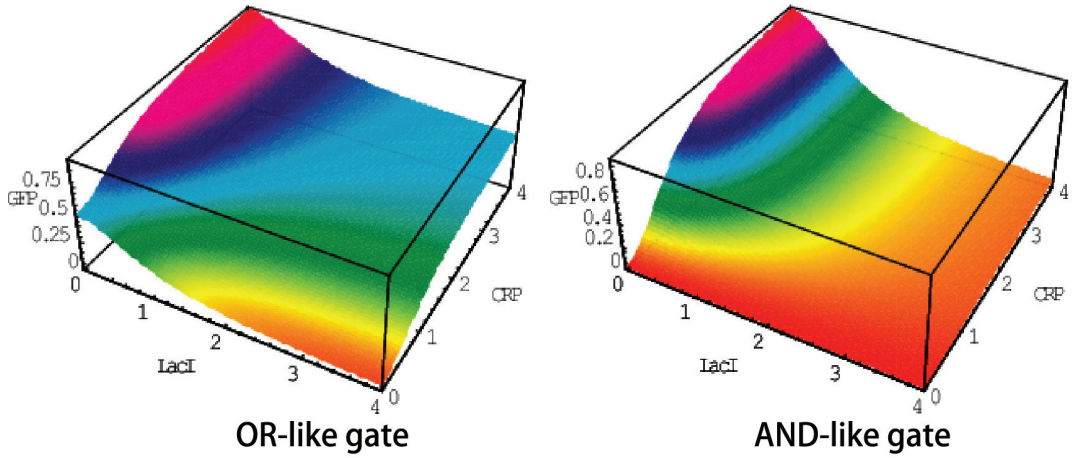


Figure 6. Mathematical models of the average regulation functions of AND-like and OR-like combinatorial promoters.



The AND-like gate is modeled here as the product of two Hill functions, and the OR-like gate as a sum of two Hill functions. If we consider an arbitrary threshold of 0.5 (cyan bands), we can see that the OR-like gate is "ON" whenever CRP activity is high or LacI activity is low. Similarly, the AND-like gate is "ON" only when CRP activity is high and LacI activity is low.

threshold ($\sim 10^1$) would make the response an OR gate, whereas a high threshold ($\sim 10^2$) would be an AND gate. Now, 23 out of the 74 well-characterized TFs of *E. coli* (Salgado et al. 2006) are themselves regulated by combinatorial promoters. For these 23 cases, a threshold may make sense: a combinatorially regulated TF can compute different functions depending on the TF's own target promoters. In the remaining 200 or so known cases, the gene product of a combinatorial promoter usually serves a metabolic function for the cell. Here the concept of a threshold does not make any sense, e.g., the amount of LacZ made by a bacterium determines how *much* lactose it can metabolize. Instead of enforcing a threshold *a priori*, we want a general way to compare the four plateau levels.

We will assume that the response is monotonic (increasing inducer concentration always increases the output level). This lets us define the four output plateau levels: $P_{++} \geq P_{+-} \geq P_{-+} \geq P_{--}$ (Figure 5). The fold-change *r*, or range, of the promoter is the same as above:

$$r = \frac{P_{++}}{P_{--}}$$

As discussed above, r is often difficult to measure accurately and depends on the reporter system. For this reason, we will often assume a value of $r = 10^3$ to enable comparisons between promoters measured with different systems. This "baseline assumption" will give us an intuitive feel for the relative plateau levels. So when a specific value of r is not given, assume $r = 10^3$.

We want to capture the relationships of the two intermediate plateaus, P_{+-} and P_{-+}. Clearly, their ratio provides a measure of how asymmetric the regulation is. When the two inputs are completely symmetric ($P_{+-} = P_{-+}$), their ratio is 1. In the other extreme, when the first input is completely dominant over the second ($P_{++} = P_{+-} \geq P_{-+} = P_{--}$), this ratio is the maximum possible: r. We define a new parameter a to capture the symmetry between the two inputs:

$$a = \log_r \frac{P_{+-}}{P_{-+}}$$

Note that by taking the logarithm of the ratio (with base r), we restrict a to lie between 0 (completely symmetric) and 1 (completely dominant). In the example promoter (Figure 5), the regulation is almost (but not quite) completely symmetric, here $a = 0.10$. When $r = 10^3$, this value of a would correspond to a case where the higher intermediate plateau (P_{+-}) is twice the value of the lower one (P_{-+}). What about AND versus OR? Both Boolean extreme responses are symmetric, but differ in the relationship between the intermediate and extreme plateaus. To capture this property, we compare the output of both intermediate plateaus to the fully induced one:

$$l = \log_r \frac{P_{++}}{\sqrt{P_{+-} P_{-+}}}$$

Again, we take logarithms base r so that the quantity ranges from 0 to 1. Since we are using logarithms, it will be mathematically convenient to compare the geometric mean of the two intermediate states to that of the fully induced state. The parameter l quantifies the "AND-ness" of the promoter logic. For an ideal OR gate ($P_{++} = P_{+-} = P_{-+} \geq P_{--}$) the geometric mean of the two intermediate states is equal to the fully induced state, so the ratio is 1 and $l = 0$. For the ideal AND

gate ($P_{++} \geq P_{+-} = P_{-+} = P_{--}$) the ratio between the intermediate states and the fully induced state is as large as it can possibly be, r, and $l = 1$. For the promoter in Figure 5, the logic-parameter is almost halfway between the perfect AND and OR, here $l = 0.49$. To illustrate, in a symmetric baseline promoter with intermediate logic ($r = 10^3$, $a = 0.0$, $l = 0.5$) the two intermediate plateaus are 32× higher than the minimal state P_{--}. The logic parameter l is orthogonal to a: it separates the effects of asymmetry (since we take the geometric mean of the intermediate states) along with the overall fold-change (since we take the logarithm base r). Together, the three parameters (r, a, l) give us a useful way to compare different promoter logic functions, without the need to invoke an artificial threshold or measure the entire response function.

Independent regulation creates intermediate logic

To see when we might expect regulation with $l = 0.5$, we shall explore a thought experiment. Suppose a promoter is regulated by two TFs X and Y (we use x and y to represent their respective activities), with X dominant. If these two TFs regulate the same promoter *independently*, we can write the full promoter regulation function as a product $P(x,y) = s(x)t(y)$ of two functions which depend (arbitrarily) on only one input each. Clearly, the output value of the highest plateau P_{++} is the product of the two maxima of $s(x)$ and $t(y)$, while the lowest plateau P_{--} is the product of the two minima. Now we can define the intermediate plateau, P_{+-}, as the product of the maximum value of $s(x)$ and the minimum value of $t(y)$ (and vice versa for P_{-+}). What does the product of these two intermediate plateaus look like? It is just the maximum of $s(x)$, times the minimum of $t(y)$, times the minimum of $s(x)$, times the maximum of $t(y)$—but this is numerically identical to the product of the fully induced and the uninduced plateaus ($P_{+-}P_{-+} = P_{++}P_{--}$)! With this property, we can quickly calculate the logic parameter l for any promoter exhibiting intependent regulation (separation of variables):

$$l = \log_r \frac{P_{++}}{\sqrt{P_{+-}P_{-+}}} = \log_r \frac{P_{++}}{\sqrt{P_{++}P_{--}}} = \log_r \sqrt{\frac{P_{++}}{P_{--}}} = \log_r \sqrt{r} = 0.5$$

Any time two TFs regulate a promoter independently, $l = 0.5$.

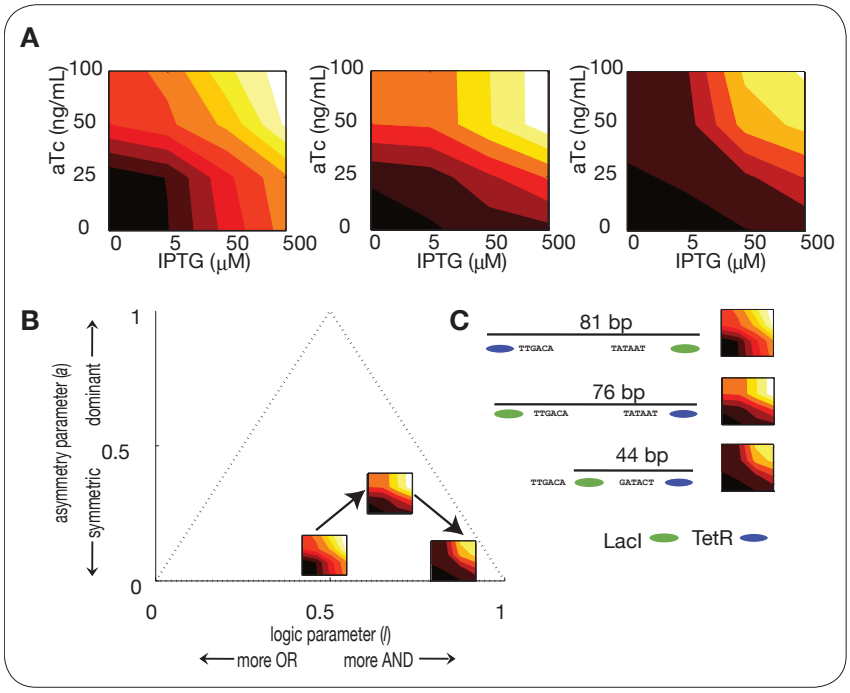
For an unknown promoter, the assumption of independent regulation is a good null hypothesis. We shall examine cases where the promoter logic is highly dependent, but in the absence of evidence to the contrary, $l = 0.5$ is the simplest assumption. Note that the above proof holds even when the regulation function is not monotonic—we just need the *existence* of plateaus for the entire concept of promoter logic to be useful. Also, the correspondence between independence and $l = 0.5$ requires the quantity in the dominator of the ratio to be the geometric mean of the intermediate plateaus (this is how we know it is a good definition). The converse is not generally true, since l is a property of the extreme responses to each input and does not capture all of the information contained in the regulation function. Still, we will often refer to $l = 0.5$ as "independent regulation."

Example promoters: two activators

To test out the phenotype parameters, we will start with two combinatorial promoters regulated by the ubiquitous global TF—CRP. The first is the natural *luxI* promoter of *V. fischeri* (Shadel and Baldwin, 1992). Here the activator LuxR binds just upstream of the -35 box, while CRP activates further upstream. Both activators can work simultaneously because they bind to different subunits of the RNA polymerase. The phenotype parameters of this promoter (measured with the native luciferase reporter) are (r = 48, a = 0.07, l = 0.39). It is almost symmetric (CRP is slightly dominant), and just on the OR side of an intermediate logic response.

The second promoter is a synthetic construct (Joung et al. 1994) based on the P_{RM} promoter of phage λ. In this promoter cI binds to a special operator just upstream of the -35 box as an activator TF (with the repressive operators removed), while the CRP operator was introduced further upstream. The phenotype parameters of this promoter (measured with a LacZ reporter) are (r = 37, a = 0.03, l = 0.44). So it is almost symmetric (here, cI is slightly dominant), and just on the OR side of an intermediate logic response. Considering these two phenotypes for our baseline ideal ($r = 10^3$); the ratios of the intermediate plateaus would be 1.6 for CRP/LuxR and 1.2 for cI/CRP, while the intermediate plateaus are about 68× above $P_{\_}$ for CRP/LuxR and 49× higher for

Figure 7. Promoter architecture determines promoter logic.



(A) Three combinatorial promoters controlled by LacI (IPTG) and TetR (aTc). Promoter activities are shown as a relative log-heat scale, as in Figure 5, with black representing minimal and white representing maximal activity. (B) Promoter response coordinates in logic-symmetry space. Each regulation icon from A is placed at the corresponding value of a and l. (C) Promoter architecture of the three promoters. Shown are the regulation icons from A, the -10 (right) and -35 boxes (left), and the relative positions of the TetR (blue) and LacI (green) operators. The inclusive distance between the two operators is shown with each promoter architecture. The experimentavl details are described in Chapter 2.

CRP/cI (compared to 32× when l = 0.5). Thus, two very different promoters can generate functions at roughly the same location in logic-phenotype space.

Example promoters: two repressors

Now consider three of the combinatorial promoters characterized in Chapter 2, each of which is repressed by both LacI and TetR (Figure 7). These promoters differ in the location of the operators for LacI and TetR. The difference in operator position results in three very different regulation functions (Figure 7A). Focusing on parameters l and a, we plot these three promoter functions as points in logic-symmetry space (Figure 7B). This space forms a triangle. At the apex, single-input regulation is always independent. Points on the base of the triangle are always symmetric ($P_{+-} = P_{-+}$), points on the right leg of the triangle have $P_{--} = P_{-+}$, and points on the left leg have $P_{+-} = P_{++}$.

Let us examine the promoter responses in terms of two operator locations (Figure 7C). The first promoter, with the two operators 81 bp (inclusive) apart, gives a near symmetric (a = 0.10) response. Switching the LacI and TetR operators and moving them 5 bp closer together has two

effects: a increases from 0.10 to 0.33 (here TetR is the dominant repressor), while l increases from 0.49 to 0.67. Going from the second promoter to the third: the LacI operator moves closer to TetR (44 bp inclusive), the response becomes symmetric again ($a = 0.07$), and the logic parameter approaches AND ($l = 0.86$). The changes in a have to do with the particulars of moving operators around the promoter (and are discussed in-depth in Chapter 2), but the change in the logic parameter l follows a simple trend: l increases as the distance between the operators decreases. Here we see that combinatorial promoters composed of the same components, measured in the same system, can generate very different phenotypes.

Operator proximity rule

Recall the differences between the CRP/LuxR and CRP/cI combinatorial promoter experiments above: marine bacterium versus phage TF, natural versus synthetic promoters, and luciferase versus lacZ reporters. Yet, the logic-symmetry parameters are almost identical. Unlike the three LacI/TetR promoters, these two have a very similar arrangement of operators, with LuxR/cI at -40 and CRP further upstream. In both cases the logic parameter is close to $l = 0.5$. Actually, the (exclusive) distance between the LuxR operator and CRP is 35 bp, while for the cI operator it is 40 bp. Both cases are *close* to fully independent regulation ($l = 0.5$), but the case with CRP 40 bp upstream is more independent ($l = 0.44$) than when CRP is placed 5 bp closer ($l = 0.39$) to the second activator. This illustrates the usefulness of assuming $l = 0.5$ as a null hypothesis when the two operators are far apart.

This general rule extends to the 288 promoters examined in Chapter 2, for both activator and repressor TFs, and gives us a heuristic for inferring the regulation function of an uncharacterized promoter. The TF independence hypothesis is most likely to be valid when two TF operators are far apart, so $l = 0.5$ is a good first guess when operator spacing is known to be large (usually greater than 40-50 bp between them). This is often the case for promoters regulated by global TFs, which can activate and repress from greater distances than most specific TFs.

Example promoters: cooperative TyrR/TrpR regulation

We have seen that it is possible to achieve near perfect AND logic with two repressors (Figure 7). In Chapter 2, we will examine a promoter with perfect AND logic (within experimental error). None of the promoters from Chapter 2 display a perfect OR function, or even a strong asymmetric OR. We could predict many situations where an OR-like response would be useful to compute. For example: if a cell is using an amino acid A as a carbon source, then abruptly switches to related amino acid A', it might make sense to continue expressing some of the genes that digest A if they are also used to metabolize A'. Natural promoters should be able to compute OR-like functions, and must exploit mechanisms beyond those used by the simple synthetic promoter library of Chapter 2.

Two such related aromatic amino acids are tyrosine and tryptophan. Shikimate is third aromatic compound with an interesting property: cells which are auxotrophic for (unable to synthesize) tryptophan, tyrosine, and phenylalanine can all grow on shikimate. This is because the cell possesses enzymes to convert shikimate into all three aromatic amino acids, using a common pathway. The first step in this pathway is to convert shikimate into shikimate-3-phosphate, and the appropriately named shikimate kinases catalyze this step.

One shikimate kinase gene is encoded by the gene *aroL* and controlled by the two TF repressors TyrR and TrpR. In this case, TyrR, which responds to tyrosine, is dominant over TrpR, which responds to tryptophan. Inspection of the promoter sequence reveals an interesting architecture: three TyrR operators, one just upstream of the -35, and two stacked just downstream of the -10 (when there are multiple operators, the TF with more operators is usually dominant). The single TrpR operator is placed just 4 bp downstream of the third TyrR operator. In this case, the promoter phenotype (Heatwole and Somerville, 1992) is highly *dependent*: ($r = 29, a = 0.53, l = 0.31$). The authors of the study that measured this phenotype (LacZ assay) suggested a cooperative protein-protein contact between TyrR and Trp based solely on the *in vivo* promoter phenotype. The cooperative interaction between TrpR and TyrR was confirmed biochemically two years later by

a separate group (Lawley and Pittard, 1994), who also re-measured the promoter response with loss of function mutations (r = 47, a = 0.40, l = 0.26). Differences in the experimental design changed r (and to a lesser extent, a) but the logic parameter l remained relatively constant (± 0.05). The phenotype lies in a region of logic-symmetry space which is halfway in-between a pure OR and a single-input gate. This "asym-OR" response is due to the cooperative interactions between TyrR and TrpR. With only tryptophan present, gene expression goes up to an intermediate plateau; but when tyrosine is present gene expression is maximal, irrespective of the presence of tryptophan.

Recall that the TrpR operator is just 4 bp downstream from a TyrR operator. The authors of the second study constructed a mutant promoter by inserting an additional 4 bp between these two operators, for a total of 8 bp spacing (exclusive). Interestingly, the new phenotype (r = 46, a = 0.44, l = 0.58) differed mostly in the parameter l. Functionally, this replaced the cooperative interaction between TrpR and TyrR with a competitive interaction. This result reinforces the notion that TF regulation is more independent (l = 0.50) when the operators are far away, but can be made dependent through competitive or cooperative interaction when the operators are closer together. This important result illustrates an inherent difficulty in predicting promoter logic when the operators are adjacent, without additional knowledge of the specific TFs involved.

In Chapter 2, we show that a perfect OR gate is *thermodynamically infeasible* for a promoter controlled by two repressors. An asym-OR gate is possible, but only with a very strong cooperative interaction between them; the energy between TyrR and TrpR in the above system needs to be about the (very high) level that LacI uses to form a 92 bp DNA loop at the *lac* promoter. Apparently, TyrR and TrpR repressors coordinate to make sure the combinatorial promoter expresses exactly the amount of shikimate kinase required.

The sum regulation function

OR promoter logic is certainly possible, just not at a single promoter controlled by two repressors. Suppose we have two single-input (switch) promoters controlled by inducible TFs X and Y with regulation functions *s(x)* and *t(y)*, as before. The details of these functions aren't important, just that they are monotonic and saturating. Now suppose that X and Y regulate two separate copies of the *same* output gene (like GFP). Then the regulation function of the output is just the sum $p(x,y) = s(x) + t(y)$.

Suppose that the induced and uninduced levels of each promoter are A and 1, respectively. The logic of this sum regulation function is just:

$$\mathsf{r} = \frac{P_{++}}{P_{--}} = \frac{A+A}{1+1} = A$$

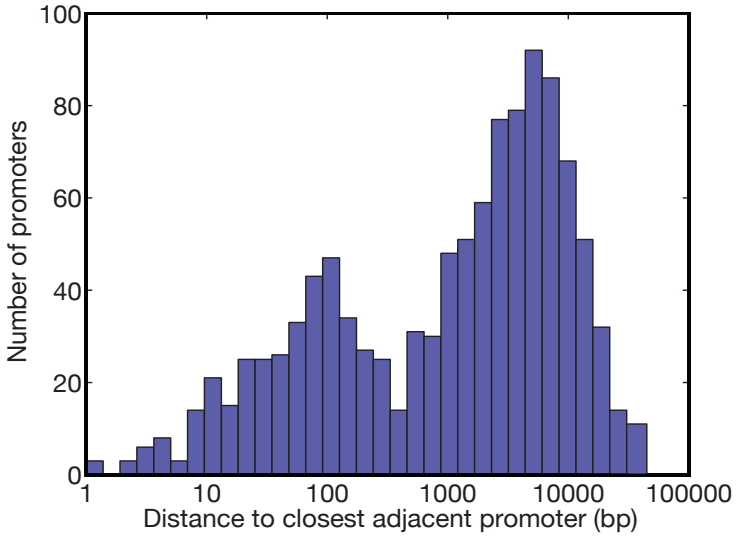
$$\mathsf{a} = \log_r \frac{P_{+-}}{P_{-+}} = \log_A \frac{A+1}{A+1} = 0$$

$$\mathsf{l} = \log_r \frac{P_{++}}{\sqrt{P_{+-}P_{-+}}} = \log_A \frac{A+A}{\sqrt{(A+1)(A+1)}} = \log_A \frac{2}{1+\frac{1}{A}} \approx \log_A 2$$

The regulation is (by design) symmetric, and the fold-change and logic depend only on the promoter strength A. This regulation function will become more OR-like the stronger it gets: for $A = 10^3$, $l = 0.10$; while for $A = 10^5$ we have $l = 0.06$.

Instead of using two single-input promoters to control two copies of a gene, we could arrange the two promoters in front of a single gene copy to form a "tandem" promoter. This arrangement is quite common in the genome: about 30% of all characterized *E. coli* promoters occur in tandem promoters (Figure 8). When the two promoters are independent, a tandem promoter generates a sum regulation function as above. For example, the tandem promoter that regulates the flagellar motor genes in *E. coli* (the *fliL* operon) exhibits exactly this type of logic function (Kalir and Alon, 2004; Kalir et al. 2005). Tandem promoters can generate more complex regulation func-

Figure 8. Tandem promoters are common in the genome.



Distribution of E. coli inter-promoter distances. For each of 1102 promoters from RegulonDB (Salgado et al. 2006), we calculated the distance to the nearest promoter that directs transcription in the same direction. The logarithm of this distribution is plotted as a histogram, revealing a bi-modal distribution. Promoters from neighboring genes are separated by ~1000 bp, while tandem promoters controlling the same gene are separated by ~100 bp or less.

tions, such as when the elongation complex of the RNA polymerase from the upstream promoter interferes with transcription initiation at the downstream promoter (i.e., "transcriptional interference" (Sneppen et al. 2005)). Other cases might involve overlapping -10/-35 boxes (common in compact genomes), or operators that regulate both promoters simultaneously. Sum regulation is the simplest assumption for the function of a tandem promoter, and can generate OR or asym-OR logic without cooperative TF-TF interactions.

Example promoters: one activator and one repressor

One particularly interesting and flexible promoter architecture responds to both an activator and a repressor. Indeed, this arrangement is quite common. Nearly 70% of the combinatorial promoters regulated by exactly two TFs (Figure 1) involve both activation and repression. As luck would have it, the first combinatorial promoter characterized—the *lac* promoter—exhibits this architecture. Two recent studies epitomize the *lac* promoter's rich history. In the first, the promoter response was measured (using both GFP and LacZ) as a function of two input inducer concentrations (Setty et al. 2003). The activator CRP is active in the presence of cAMP, while the repressor LacI is inactivated by the inducer IPTG. The authors simply measured the response of a wild type MG1655 *E. coli* strain, containing a promoter fusion reporter on a low copy plasmid, in various combinations of IPTG and cAMP concentrations. The logic parameters of this wild-type

system were ($r = 14, a = 0.26, l = 0.61$), where LacI slightly dominated CRP, forming four distinct plateaus of intermediate AND-like logic.

Several factors complicate this analysis. Under the conditions measured, the LacI tetramer forms a DNA "loop" around the CRP operator. This increases the local concentration of LacI near the operators, but might also change the nature of the LacI-CRP interaction. In this case, a LacI dimer mutant can enact significantly different regulation (Oehler et al. 1994). One of the genes in the *lac* operon is the lactose permease LacY, which also increases cell uptake of the inducer IPTG. This example of positive feedback increases the steepness of the IPTG response, and sharpens the plateau boundaries. The cAMP response is even more complicated. Since CRP is a ubiquitous global TF, its activity is highly regulated. The expression of the CRP protein, the endogenous production of cAMP, and the degradation of cAMP—are all regulated by multiple inputs. Each of these factors contribute to the "effective cooperativity" of the response.

The (Kuhlman et al. 2007) study systematically explored the results of removing these varied sources of effective cooperativity. Deleting LacY decreased the switch-like nature of the IPTG response, as did mutating LacI to be unable to form loops. Similarly, deletion of the enzymes controlling the synthesis and degradation of the cAMP inducer resulted in a completely non-cooperative (Michaelis-Menten) CRP response. The central result of this analysis was that CRP assists LacI to bend the DNA into a loop, resulting in a slight AND-like interaction. Importantly, the response function still displayed four plateaus when all of these factors (save the DNA looping) were removed. Here the phenotype parameters were ($r = 10^4, a = 0.10, l = 0.55$). Differences in the (LacZ) measurement assay produced a much larger r, and the removal of the endogenous CRP regulation system reduced the asymmetry of the response. Amazingly, the logic parameter ($l = 0.55$) remained quite similar to the previous study ($l = 0.61$). Despite many complicated and nonlinear factors, both the plateau response and the relationship between the two TFs (captured by the logic parameter l) remain robust and general properties of the promoter phenotype.

Searching for Logic Functions in Promoter Sequence Space

What determines the shape of a promoter regulation function? In this final section, I examine two methods to explore the rich diversity of combinatorial promoter logic. I review an example combinatorial promoter experiment, and discuss how libraries of combinatorial promoters can reveal the rules relating promoter sequence to promoter function.

Promoter sequence space

Promoters contain many sequence features. The -10/-35 boxes (Hawley and McClure, 1983), the extended -10 (Kumar et al. 1993), the UP elements (Ross et al. 1993), the -45 region (Czarniecki et al. 1997), the discriminator region (Haugen et al. 2006), the initiating nucleotide (Condon et al. 1995), the abortive initiation region (Hsu, 2002), cryptic sequences (Huerta and Collado-Vides, 2003), polymerase pausing sites (Nickels et al. 2004), and of course the TF operator sequences all combine to influence promoter activity. In fact, nearly every one of the 100 bp of the promoter region has been found to influence expression in one promoter or another. Not all sequence features are important in all promoters; the difficulty comes in because we do not know how to recognize important features in a given promoter sequence. 100 bp is a big sequence space: the number of possible promoter sequences (4^{100}) is about the number of protons contained in Eta Carinae, a star with 150× the mass of Sol.

The σ^{70} family of bacterial promoters (in contrast to eukaryotic and σ^{54} bacterial promoters) all posses one very interesting characteristic: the sequence features in the list above (except the operators, of course) are positioned at a fixed distance (± 1-3 bp) relative to the essential -10 box. This simple property cuts the sequence space down to a manageable, but unknown, size. Functional promoters are those that contain the "right" combination of sequence features positioned relative to -10. For example, the sequence TTTTTTTTTTTTTTTT is a (weak) functional promoter when placed almost anywhere (other than within another poly(dT) tract). Interesting regulation

forms an even smaller subset of this space.

Library example: promoter mutagenesis

One tack to explore promoter sequence space, borrowed from the field of directed evolution, is to examine the functional consequences of sequence mutagenesis. Here, a given promoter sequence is changed by randomized synthesis, site-directed mutagenesis, or mutagenic PCR. These methods mimic the evolutionary process of point mutation.

One study (Mayo et al. 2006) characterized 12 variant sequences of the *lac* promoter; each of which contained 4-9 point mutations with the two (repressor) LacI operators, the (activator) CRP operator, and the -10/-35 boxes. The phenotypes of these promoters span a range of phenotype parameters (r = 14 to 333, a = 0.00 to 0.71, l = 0.30 to 0.79), which surround the wild type (r = 14, a = 0.26, l = 0.61) in logic-symmetry space (Figure 9). This library samples the interior of the logic-symmetry space, but misses the extremes.

It is relatively simple to change a from one extreme to the other: to reach a = 1 just obliterate all operators of one TF. In the wild type promoter, LacI is the dominant TF. In two variants (blue circles), mutations in the upstream LacI operator and -35 boxes result in significant ($a \geq 0.1$) dominance by CRP. The l parameter was less variable: our canonical symmetric baseline promoter ($r = 10^3$, a = 0) with l = 0.30 would have intermediate plateaus 125× above the uninduced state, while with l = 0.79 it would have intermediate plateaus 4× higher. The most AND-like mutant promoter was near symmetry (a = 0.10), but the most OR-like mutant promoter was not (a = 0.25). These two promoters with the most extreme l values contain nine mutations each—the largest of the 12 mutants characterized. This study shows that promoters form clusters of functions within sequence space: r and a can be changed easily, but large changes in l requires several mutagenic steps.

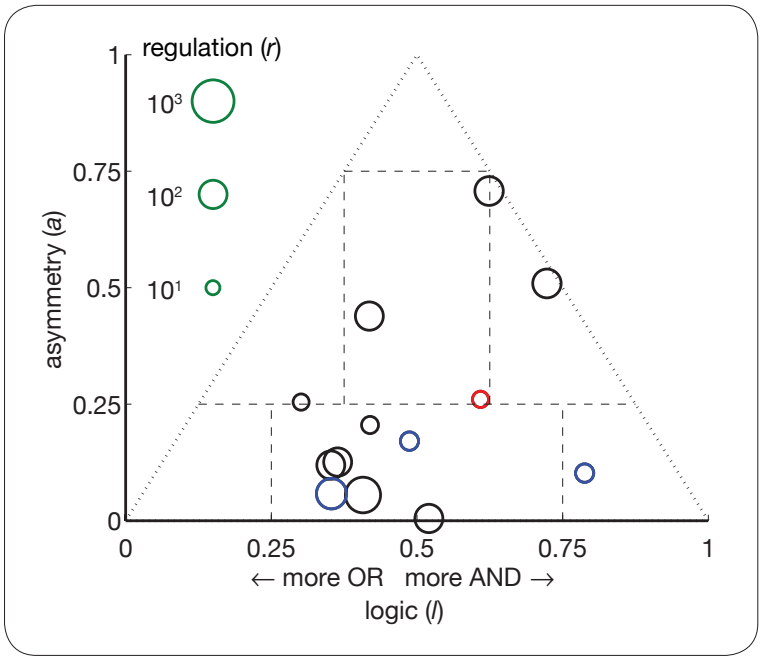


Figure 9. Promoters form clusters of functions in sequence space. Phenotype parameters of wild-type and 12 mutant *lac* promoters (Mayo et al. 2006) are re-plotted as points in logic-symmetry space. Here, the regulatory range r is represented by circle size. The wild type promoter is represented by a red circle, CRP dominant promoters by blue circles, and LacI dominant promoters by black circles.

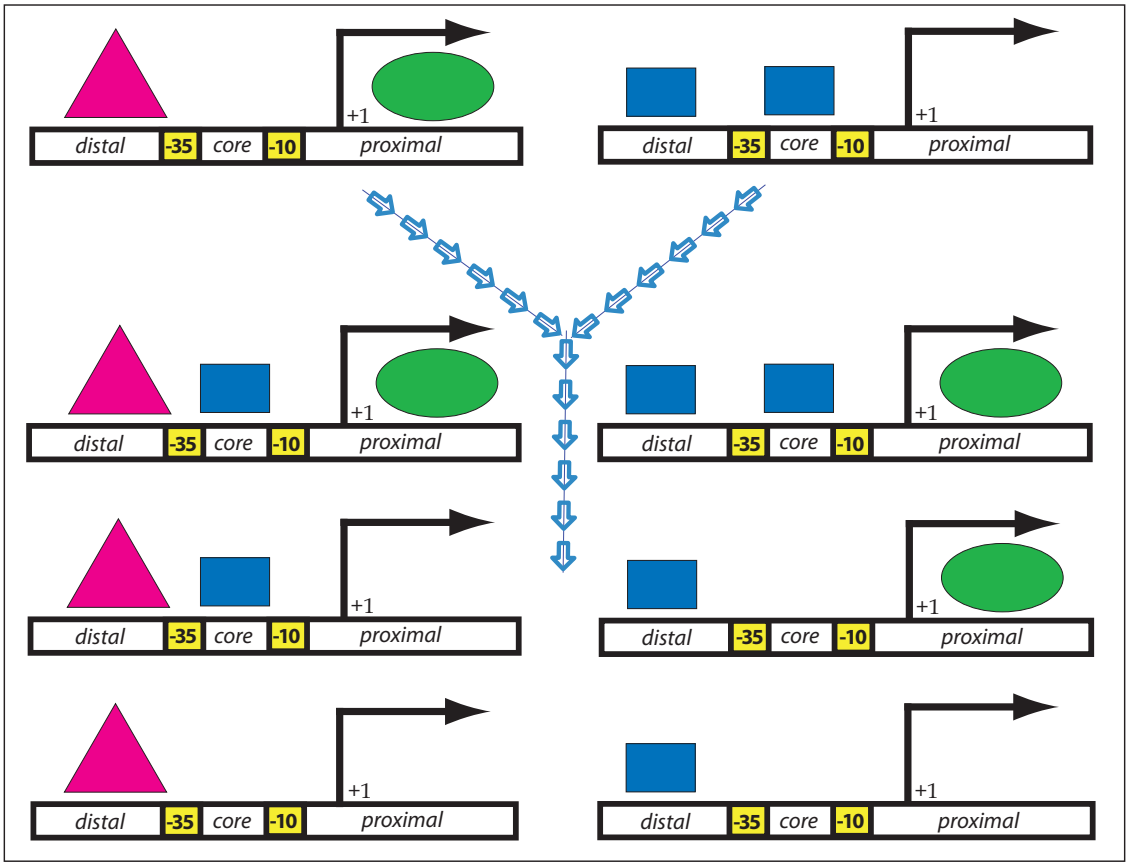
Promoter shuffling

In Chapters 1 and 2, we follow a novel path through promoter sequence space. Instead of mutating a particular promoter architecture, we take promoter sequences regulated by a small set of inducible TFs and *shuffle* them to create variants which lie between sequences known to display effective regulation. For this we wanted to control almost every nucleotide of the promoter. Since the interactions between promoter sequence features are largely unknown, we sought to preserve as much sequence diversity as possible. We developed a method (Chapter 1) that constrains 3 out of the 100 bp—and put these fixed bp at the regions of highest conservation—in the -10 and -35 boxes. This way, if an operator requires a specific -35 sequence, or the absence of an UP element, etc., to function, it is more likely to "carry" the features it needs when incorporated into a shuffled promoter. In this implementation, each promoter is split into three regions at -10 and -35. In the example shown (Figure 10), one promoter regulated by a single TF and one regulated by two TFs are shuffled to create six novel promoter variants regulated by one, two, or all three TFs. When N promoters are shuffled in this way, the number of possible promoter variants is 3^N. One could split the promoter into even smaller regions, using three allowed us to construct a highly diverse but

still manageable combinatorial library.

Several promoters in our combinatorial library compute the same function with different combinations of TFs. By comparing promoters of similar function, we were able to deduce an initial set of five rules for programming combinatorial promoters. Four of these rules detail differences between specific activator and repressor TFs, along with the most effective operator locations of each. The operator proximity rule, $l = 0.5$ when operators are far apart, applies more generally. These rules will help to engineer diverse promoter logic functions from arbitrary TF combinations, and to deduce the logic of natural promoters.

Figure 10. Promoter shuffling generates novel regulatory sequences.



Two regulatory promoters responding to different TFs are shuffled to create 6 novel regulatory promoters. Promoters are split at the conserved -10 and -35 boxes to create three fragments: distal, core, and proximal. Each shuffled promoter contains one of each fragment, selected from either of the two initial promoter sequences.

Ultimately, we aim to quantitatively predict a promoter's response function from the placement of regulatory operators, along with the minimum necessary knowledge of its TFs. Databases of transcriptional regulation such as RegulonDB (Salgado et al. 2006) have grown to contain thousands of regulatory connections, with hundreds of combinatorial promoters. In order to quantitatively predict the functions of these networks, we need to understand the combinatorial promoter functions. The recent surge in DNA synthesis technology provides the ability to cheaply build large libraries of synthetic combinatorial promoters, in order to elucidate the variables that determine combinatorial promoter functions. Armed with such knowledge, we may begin to quantitatively predict the functions of both natural and synthetic networks of TFs.

Chapter Outline

Chapter 1 introduces a simple method for assembling diverse random DNA libraries, in which the relative spacing between regulatory sequences is maintained. The primary technical advancement over similar technologies is that almost every bp of each sequence is under control: the "scars" of assembly (sequence held fixed by the assembly process) are kept at 1-2 bp. We found that 1-2 bp overhangs on the 5' end of duplex fragments were able to direct self-assembly and ligation with very low error rates. For many applications these scars would not matter, but in promoters every bit of information is critical. The library diversity grows as the cube of the number of synthetic oligonucleotides, though it is possible to extend the method to longer assemblies. The technique displays amazing efficiency: almost all (97%) of the sequenced constructs are correctly assembled, over three-quarters are unique, and over half are functional, regulatory promoter sequences.

Chapter 2 takes the constructed promoters and analyzes their responses to transcriptional regulation. The library approach illuminates trends between similar promoter architectures. Each functional promoter responds to one or two (out of a set of four) TFs. Here, I study the features of

each promoter's architecture that determine its logic. I focus only on the extreme (saturating) responses to each TF in order to classify promoters of similar logic. Mathematical analysis of the *cis*-regulatory input functions relates the TF interactions to the possible logic functions they can generate. Comparing the responses between architectures and different TFs reveals a set of heuristics for predicting promoter function from sequence signals. The heuristics support and extend known principles from the distributions of operators in natural promoters (Collado-Vides et al. 1991; Gralla and Collado-Vides, 1996), give general insight into decoding natural promoter function, and will certainly aid the synthetic biologist in constructing ever more complex genetic circuits.

In Chapter 3, I turn to single-cell measurements of transcriptional regulation, to study genetic noise. The technical contribution is a reporter construct, which can sensitively measure the single-cell response of three genes simultaneously. I show that the three reporters are spectrally and genetically isolated, which allows the analysis of correlation in genetic noise. To demonstrate the utility of the reporter system, I show how natural genetic noise can be exploited to measure the extent of transcriptional co-regulation and to infer regulatory connections. This technique can be extended to uncover new combinatorial promoters.

In Chapter 4, I describe a synthetic regulatory circuit that uses TFs and cell-cell signaling molecules to autonomously regulate the density of an *E. coli* population. This circuit is constructed such that the cells, despite the noise inherent in their gene expression at the individual cellular level, manifest a reproducible phenotype at the population level. In this case, the genetic circuit implements feedback from the population to the individual, allowing population control through feedback regulation of the suicide rate. Indeed, the robust functioning of this circuit relies on noise to choose which cells are killed and which are spared. The circuit response to an environmental signal, in this case pH, is accurately described by a simple mathematical model. Its function demonstrates a different kind of combinatorial regulation, exemplified by natural quorum sensing systems, in which multiple cells detect and respond to the average behavior over the population.

Figure Captions

Figure 4. *Over-expression results in spurious co-regulation.* *E. coli* DH5αZ1 (Lutz and Bujard, 1997) cells containing plasmids pZA3RY* and pZE12C* and expressing YFP and CFP. YFP was constitutively expressed (λ P(R) promoter, p15A plasmid) while CFP was under the control of an IPTG inducible promoter (P_{LlacO1}, ColE1 plasmid). Here, the relative levels of expression of CFP (cyan) and YFP (green) are plotted as the ratio of IPTG induced (1mM) divided by uninduced expression levels versus time after induction. This ratio reveals that IPTG induction increases CFP expression as expected, but simultaneously results in a decrease in YFP expression. For this experiment, overnight cultures were diluted 1000× into 150μL of selective LB media with and without inducer and grown at 37° C in a Wallac Victor microplate fluorimeter with continuous shaking, fluorescence, and optical density measurements. Background autofluorescence was determined by DH5αZ1 containing no plasmids and grown in similar conditions. Both fluorescent proteins were tagged with a strong ssrA degradation tag (AANDENYALA) to prevent accumulation. Two independent experiments are shown, where each experiment is averaged over eight replicates. The peak in the CFP induced case may be due to a reproducible growth effect.

Figure 5. *Combinatorial promoter responses exhibit plateaus.* An example of two input promoter response is shown. Promoter activity is plotted on a log heat scale (colorbar), and responds two inputs, where input 1 has a slightly larger effect than input 2. Most output levels lie on the four response plateaus P_{--} (no inputs), P_{+-} (input 1 saturating), P_{-+} (input 2 saturating), or P_{++} (both inputs saturating). This general monotonic response with $P_{++} \geq P_{+-} \geq P_{-+} \geq P_{--}$ is used to calculate the phenotype parameters (*r, a, l*) as a function of the four plateau levels.

References

Anderson JC, Voigt CA, Arkin AP (2007) Environmental signal integration by a modular AND gate. *Molecular Systems Biology* **3:** 133.

Balaban NQ, Merrin J, Chait R, Kowalik L, Leibler S (2004) Bacterial persistence as a phenotypic switch. *Science* **305:** 1622-1625.

Basu S, Gerchman Y, Collins CH, Arnold FH, Weiss R (2005) A synthetic multicellular system for programmed pattern formation. *Nature* **434:** 1130-1134.

Basu S, Mehreja R, Thiberge S, Chen MT, Weiss R (2004) Spatiotemporal control of gene expression with pulse-generating networks. *Proc Natl Acad Sci U S A* **101:** 6355-6360.

Bjarnason J, Southward CM, Surette MG (2003) Genomic profiling of iron-responsive genes in *Salmonella enterica serovar typhimurium* by high-throughput screening of a random promoter library. *J Bacteriol* **185:** 4973-4982.

Browning DF, Busby SJ (2004) The regulation of bacterial transcription initiation. *Nat Rev Microbiol* **2:** 57-65.

Buchler NE, Gerland U, Hwa T (2005) Nonlinear protein degradation and the function of genetic circuits. *Proc Natl Acad Sci U S A* **102:** 9559-9564.

Cai L, Friedman N, Xie XS (2006) Stochastic protein expression in individual cells at the single molecule level. *Nature* **440:** 358-362.

Campbell RE, Tour O, Palmer AE, Steinbach PA, Baird GS, Zacharias DA, Tsien RY (2002) A monomeric red fluorescent protein. *Proc Natl Acad Sci U S A* **99:** 7877-7882.

Collado-Vides J, Magasanik B, Gralla JD (1991) Control site location and transcriptional regulation in *Escherichia coli*. *Microbiol Mol Biol Rev* **55:** 371-394.

Condon C, Squires C, Squires CL (1995) Control of rRNA transcription in *Escherichia coli*. *Microbiological reviews* **59:** 623-645.

Czarniecki D, Noel RJ, Reznikoff WS (1997) The -45 region of the *Escherichia coli* lac promoter: CAP-dependent and CAP-independent transcription. *J Bacteriol* **179:** 423-429.

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

Elowitz MB, Surette MG, Wolf PE, Stock JB, Leibler S (1999) Protein mobility in the cytoplasm of *Escherichia coli*. *J Bacteriol* **181:** 197-203.

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

Gottesman S, Roche E, Zhou Y, Sauer RT (1998) The ClpXP and ClpAP proteases degrade proteins with carboxy-terminal peptide tails added by the SsrA-tagging system. *Genes & development* **12:** 1338-1347.

Gralla JD, Collado-Vides J (1996) Organization and Function of Transcription Regulatory Elements. In *Escherichia coli and Salmonella: cellular and molecular biology*, Neidhardt FC, Curtiss R (eds), 2nd edn, pp 2 v. (xx, 2822 p.). Washington, D.C.: ASM Press.

Guet CC, Elowitz MB, Hsing W, Leibler S (2002) Combinatorial synthesis of genetic networks. *Science* **296:** 1466-1470.

Guido NJ, Wang X, Adalsteinsson D, McMillen D, Hasty J, Cantor CR, Elston TC, Collins JJ (2006) A bottom-up approach to gene regulation. *Nature* **439:** 856-860.

Haugen SP, Berkmen MB, Ross W, Gaal T, Ward C, Gourse RL (2006) rRNA promoter regulation by nonoptimal binding of sigma region 1.2: an additional recognition element for RNA polymerase. *Cell* **125:** 1069-1082.

Hawley DK, McClure WR (1983) Compilation and analysis of *Escherichia coli* promoter DNA sequences. *Nucleic Acids Res* **11:** 2237-2255.

Heatwole VM, Somerville RL (1992) Synergism between the Trp repressor and Tyr repressor in repression of the aroL promoter of *Escherichia coli* K-12. *J Bacteriol* **174:** 331-335.

Hooshangi S, Thiberge S, Weiss R (2005) Ultrasensitivity and noise propagation in a synthetic transcriptional cascade. *Proc Natl Acad Sci U S A* **102:** 3581-3586.

Hsu LM (2002) Promoter clearance and escape in prokaryotes. *Biochim Biophys Acta* **1577:** 191-207.

Huerta AM, Collado-Vides J (2003) Sigma70 promoters in *Escherichia coli*: specific transcription in dense regions of overlapping promoter-like signals. *Journal of molecular biology* **333:** 261-278.

Joung JK, Koepp DM, Hochschild A (1994) Synergistic activation of transcription by bacteriophage lambda cI protein and *E. coli* cAMP receptor protein. *Science (New York, NY)* **265:** 1863-1866.

Kalir S, Alon U (2004) Using a Quantitative Blueprint to Reprogram the Dynamics of the Flagella Gene Network. *Cell*. 117: 713-720.

Kalir S, Mangan S, Alon U (2005) A coherent feed-forward loop with a SUM input function prolongs flagella expression in *Escherichia coli*. *Mol Syst Biol* **1:** 2005 0006.

Kuhlman T, Zhang Z, Saier MH, Jr., Hwa T (2007) Combinatorial transcriptional control of the lactose operon of Escherichia coli. Proc Natl Acad Sci U S A 104: 6043-6048.

Kumar A, Malloch RA, Fujita N, Smillie DA, Ishihama A, Hayward RS (1993) The minus 35-recognition region of *Escherichia coli* sigma 70 is inessential for initiation of transcription at an "extended minus 10" promoter. *Journal of molecular biology* **232:** 406-418.

Lawley B, Pittard AJ (1994) Regulation of aroL expression by TyrR protein and Trp repressor in *Escherichia coli* K-12. *J Bacteriol* **176:** 6921-6930.

Lutz R, Bujard H (1997) Independent and tight regulation of transcriptional units in *Escherichia coli* via the LacR/O, the TetR/O and AraC/I1-I2 regulatory elements. *Nucleic Acids Res* **25:** 1203-1210.

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

Matz MV, Lukyanov KA, Lukyanov SA (2002) Family of the green fluorescent protein: journey to the end of the rainbow. *BioEssays : news and reviews in molecular, cellular and developmental biology* **24:** 953-959.

Mayo AE, Setty Y, Shavit S, Zaslaver A, Alon U (2006) Plasticity of the cis-regulatory input function of a gene. *PLoS Biol* **4:** e45.

Meighen EA (1991) Molecular biology of bacterial bioluminescence. *Microbiological reviews* **55:** 123-142.

Miller JH (1972) *Experiments in molecular genetics*. [Cold Spring Harbor, N.Y.]: Cold Spring Harbor Laboratory.

Nagai T, Ibata K, Park ES, Kubota M, Mikoshiba K, Miyawaki A (2002) A variant of yellow fluorescent protein with fast and efficient maturation for cell-biological applications. *Nature biotechnology* **20:** 87-90.

Nickels BE, Mukhopadhyay J, Garrity SJ, Ebright RH, Hochschild A (2004) The sigma 70 subunit of RNA polymerase mediates a promoter-proximal pause at the lac promoter. *Nat Struct Mol Biol* **11:** 544-550.

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

Perez-Brocal V, Gil R, Ramos S, Lamelas A, Postigo M, Michelena JM, Silva FJ, Moya A, Latorre A (2006) A small microbial genome: the end of a long symbiotic relationship? *Science (New York, NY)* **314:** 312-313.

Perrenoud A, Sauer U (2005) Impact of global transcriptional regulation by ArcA, ArcB, Cra, Crp, Cya, Fnr, and Mlc on glucose catabolism in *Escherichia coli*. *J Bacteriol* **187:** 3171-3179.

Ptashne M (2004) *A genetic switch : phage lambda revisited*, 3rd edn. Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory Press.

Rizzo MA, Springer GH, Granada B, Piston DW (2004) An improved cyan fluorescent protein

variant useful for FRET. *Nature biotechnology* **22:** 445-449.

Rodionov DA (2007) Comparative genomic reconstruction of transcriptional regulatory networks in bacteria. *Chemical reviews* **107:** 3467-3497.

Rosenfeld N, Elowitz MB, Alon U (2002) Negative autoregulation speeds the response times of transcription networks. *Journal of molecular biology* **323:** 785-793.

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

Ross W, Gosink KK, Salomon J, Igarashi K, Zou C, Ishihama A, Severinov K, Gourse RL (1993) A third recognition element in bacterial promoters: DNA binding by the alpha subunit of RNA polymerase. *Science* **262:** 1407-1413.

Salgado H, Gama-Castro S, Peralta-Gil M, Diaz-Peredo E, Sanchez-Solano F, Santos-Zavaleta A, Martinez-Flores I, Jimenez-Jacinto V, Bonavides-Martinez C, Segura-Salazar J, Martinez-Antonio A, Collado-Vides J (2006) RegulonDB (version 5.0): *Escherichia coli* K-12 transcriptional regulatory network, operon organization, and growth conditions. *Nucleic Acids Res* **34:** D394-397.

Setty Y, Mayo AE, Surette MG, Alon U (2003) Detailed map of a *cis*-regulatory input function. *Proc Natl Acad Sci U S A* **100:** 7702-7707.

Shadel GS, Baldwin TO (1992) Positive autoregulation of the *Vibrio fischeri* luxR gene. LuxR and autoinducer activate cAMP-catabolite gene activator protein complex-independent and -dependent luxR transcription. *The Journal of biological chemistry* **267:** 7696-7702.

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

Sneppen K, Dodd IB, Shearwin KE, Palmer AC, Schubert RA, Callen BP, Egan JB (2005) A mathematical model for transcriptional interference by RNA polymerase traffic in *Escherichia coli*. *Journal of molecular biology* **346:** 399-409.

Thomas R, D'Ari R (1990) *Biological feedback*. Boca Raton: CRC Press.

Tsien RY (1998) The green fluorescent protein. *Annual review of biochemistry* **67:** 509-544.

Xiong W, Ferrell JE, Jr. (2003) A positive-feedback-based bistable "memory module" that governs a cell fate decision. *Nature* **426:** 460-465.

Zaslaver A, Bren A, Ronen M, Itzkovitz S, Kikoin I, Shavit S, Liebermeister W, Surette MG, Alon U (2006) A comprehensive library of fluorescent transcriptional reporters for *Escherichia coli*. *Nature methods* **3:** 623-628.

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

Short-Overhang Randomized Self-Assembly Ligation

Robert Sidney Cox III, Michael G. Surette, and Michael B. Elowitz

Abstract

We introduce Short-Overhang Randomized Self-Assembly Ligation (SORSAL): an efficient, directed method for rapidly generating diverse combinatorial DNA libraries. The method uses synthetic duplex DNA fragments and very short (1–2 bp) cohesive overhangs to direct ordered assemblies. Multiple DNA fragments are simultaneously ligated in a single step to produce random concatemers, without the need for PCR amplification or DNA purification. We describe the method, characterize the assembly process, and identify optimal 1–2 bp cohesive overhangs based on the efficiency and specificity of their ligation. We then use the method to construct a synthetic library of 10,000 prokaryotic promoters, each made of three fragments. We show, by sequencing 288 assembled constructs, that the method is 97% efficient for producing correctly assembled concatemers. Finally, we numerically simulate the assembly process and show that the library of 10,000 contains at least 1,000 unique promoter sequences. This method should be useful for any application requiring explicit control of short (< 1 kb) and highly diverse synthetic DNA libraries.

Introduction

Over the past 50 years, combinatorial synthesis of organic compounds has revolutionized the discovery of novel pharmaceuticals. Using a starting molecule, combinations of chemical substitutions are explored to generate large libraries of variants for screening (Dolle et al. 2006). Modular DNA sequences can be used to generate combinatorial libraries in an analogous or complementary fashion (Rozenman et al. 2007), with two key differences: (a) given the regularity of Watson-Crick base pairing, it is more straightforward to create libraries of DNA sequences from synthetic oligonucleotides than to create small molecule libraries with synthetic organic chemistry; and (b) the number of possible variants grows exponentially in a predictable and uniform way. For a generic DNA library of length N bp made of the four nucleotides (A, T, C, and G), the number of possible variants is 4^N.

DNA libraries have countless applications in the fields of biotechnology, directed evolution (Arnold, 1998), and synthetic biology (Sprinzak and Elowitz, 2005). Combinatorial synthesis of genetic networks can generate novel circuit functions (Guet et al. 2002). Libraries of protein antibodies can mirror the combinatorial diversity of the immune system. Protein active sites can be independently evolved (Treynor et al. 2007), or homologous proteins may be recombined (Fischbach et al. 2007; Meyer et al. 2006; Saftalov et al. 2006; Stemmer, 1994). Prokaryotic non-coding regulatory sequences (e.g., promoters, terminators, ribozymes) are generally much shorter than coding regions, and are particularly amenable to combinatorial screening (Atsumi and Little, 2006; Hammer et al. 2006). For example, varying the regions between open reading frames in an operon has been used to optimize the relative levels of protein expression in a metabolic pathway (Pfleger et al. 2006).

For many applications, combinatorial DNA libraries must preserve the relationship between multiple sequence features. For example, in *Escherichia coli* promoters, the σ^{70} polymerase subunit recognizes two hexameric sequence features called the -10 (Pribnow) box and the -35 box. Though

natural promoters typically display variation in the sequences of the -10 and -35 boxes, the spacing between them is strongly conserved at 17 ± 1 bps (Hawley and McClure, 1983). To examine promoter variants with different -10 and -35 boxes, one must vary these hexamers while constraining their relative positions. Such "ordered" combinatorial DNA libraries are powerful tools for examining the relationships between sequence features and gene regulation (Ligr et al. 2006).

We desired a method for producing diverse combinatorial DNA libraries of σ^{70} promoters, while fulfilling several criteria. First, we wanted a process that produced a very high fraction of correct assemblies. Second, we wanted the procedure to be nearly seamless (Geu-Flores et al. 2007) (e.g., minimizing the number of bps fixed), to allow for variety of -10 and -35 boxes and in placement of other promoter regulatory signals (Chapter 2). Very short (1–2 bp) fixed sequences allow a regulatory signal to be placed across them, whereas longer (4–6 bp) fixed sequences do not. Third, we wanted all correct assemblies to be nearly the same length. Fourth, we wanted to generate a "target diversity" of at least 1,000 promoter variants, while minimizing cost. Fifth, we wanted the assembly to be a modular process that could be repeated with different sets of fragments. Finally, we desired the ability to easily make a single specific sequence "to-order" from the same components.

To fulfil these aims, we introduce here a method for generating highly diverse combinatorial libraries: Short-Overhang Randomized Self-assembly Ligation (SORSAL),. As implemented here, this SORSAL employs short (12–42 bp) duplex DNA fragments with very short (1–2 bp) cohesive overhangs to generate 50–115 bp ordered assemblies of three duplex fragments. The total diversity of this scheme is p^*q^*r, where (p, q, r) represents the number of variants at each of the three positions. For equal numbers of variants at each position ($p = q = r$), the diversity increases as the cube of the number of component fragments (p^3). We show that this method is extremely efficient, optimize it, and use it to produce a diverse DNA library of σ^{70} promoters.

Results and Discussion

Optimization of ligation conditions

We initially tested single bp overhang ligation. We examined the relative rates of 3 different sets of single bp overhangs: 5' G + 5' C, 5' A + 5' T, and 3' A + 3' T. Each type of overhang was examined in a variety of conditions: with different combinations of 5' phosphorylation, in the presence of non-cognate 5' overhangs, and in the presence and absence of sodium (which can slow the rate of mispairing). In a typical experiment (Fig. 1) we tested the 5' A + 5' T ligation reaction. Reactions with the 5' A phosphorylated (Fig. 1A, Lanes 1–2) produced many more misassembled products (grey arrows) than did reactions with only the 5' T phosphorylated (Fig. 1A, Lane 1, and Fig. 1C, Lane 6). The qualitative results of all single bp overhang experiments are summarized in Table 1. The principle result was that the 5' G + 5' C reaction was fast, followed by 5' A + 5' T, while 3' A + 3' T proceeded at the slowest rate.

Several reactions produced misassembled products. For example, when a single fragment with two single bp overhangs (5' A and 5' C) was allowed to self-ligate, three different types of mispairing reactions occurred (Fig. 1B). The pattern of misassembly depended on the phosphorylation states of each end (Fig. 1C, Lanes 3–5). The phosphorylated 5' A overhang preferentially ligated to the unphosphorylated 5' C overhang. With only the 5' A overhang phosphorylated, the reaction produced a ladder of multimer misassemblies containing A/C mispairings (Fig. 1C, Lane 4). Conversely, a phosphorylated 5' C overhang primarily ligated to itself. With only the 5' C phosphorylated, the reaction produced primarily dimer misassemblies via C/C mispairings (Fig. 1C, Lane 5). When both 5' A and 5' C were phosphorylated, ligation produced an intermediate ladder of multimers, due to a mixture of A/C and C/C mispairings (Fig. 1C, Lane 3). Adding a second duplex fragment with a 5' phosphorylated T overhang reduced the rate of this multimerization (Fig. 1C, Lane 6), but had little effect when the 5' T overhang was left unphosphorylated (Fig. 1C, Lane 7). This indicated that the pairing of the 5' T overhang to

the 5' A overhang decreased the rate C/A mispairing. These results show that 5' phosphorylation controls the ligation process, but non Watson-Crick pairing can lead to misassembled reaction products.

All four single bp 5' phosphorylated overhangs were found to form "promiscuous" ligation products when the corresponding complementary overhang was absent (Table 1). Most dramatic were the phosphorylated 5' T and 5' C overhangs, that were able to pair with any 5' overhang tested. These side-reactions did not occur when a complementary (5' A or 5' G, respectively) overhang was present (i.e., though they are promiscuous, they have "high fidelity"). Strikingly, we found that a 5' A always produced misassembly products—even in the presence of a 5' T overhang (i.e., "low fidelity"). Furthermore, we detected a very low rate of non-specific ligation products from the 5' A overhang in the absence of any 5' phosphorylation. Only the 5' A was able to produce ligation products in the absence of 5' phosphorylation; the other overhangs could not react without it. Most surprisingly, though the reaction rate was much faster (Fig. 1C, Lane 1), the reaction fidelity of a 4 bp overhang control (GATC) was not qualitatively greater than for the single bp overhangs (Fig. 1D). These results show that different 5' overhangs ligate at different rates, at different levels of fidelity when the complementary overhang is present, and at different levels of promiscuity when the complement is absent.

We next assembled a 53 bp concatemer of three duplex fragments, to produce a consensus σ^{70} promoter sequence (Hawley and McClure, 1983). This reaction exhibited misassembly similar to that found for the two-fragment ligation experiments (Fig. 4). Most infidelity was generated from T and A mispairings. With all four 5' overhangs phosphorylated, the overall reaction rate was limited by the slow 5' T reaction. Phosphorylating all four 5' overhangs also produced the highest yield of correctly assembled product. Removing the 5' phosphate from the A overhang slowed the reaction, while removing the 5' phosphate from the T overhang did not significantly reduce the rate of side-reactions. These results show that phosphorylating all 5' overhangs produces the

highest yield of correctly assembled three-molecule ligation product, and that the three-molecule ligation reaction is limited by the rate of the 5' T—and the fidelity of the 5' A—reactions.

The fully assembled reaction products continued to increase in concentration over four days (not shown). We cloned the four-day ligation products into a reporter vector and sequenced 8 clones to determine the assembly products. Five of these clones were vector background, and three contained promoter assemblies. Two of the promoter assemblies contained perfectly assembled σ^{70} promoter sequences. The third was a misassembled promoter with two *core* fragments, corresponding to a dimer misassembly product (Fig. 1B). This result confirmed that single bp overhangs could direct three-molecule ligation, but the 5' T + 5' A reaction was prone to producing misassembled products.

To address these problems with yield and misassembly, we repeated the consensus promoter construction using double bp, 5' TT + 5' AA overhangs in place of the single bp 5' T + 5' A overhangs. This arrangement of one G-C pairing and two T-A pairings was designed to give similar ligation kinetics, as the free energy of one G-C pair is roughly equivalent to two T-A pairs. We cloned this assembly into a reporter vector and sequenced 12 random transformants. All 12 were correctly assembled σ^{70} consensus promoter sequences, capable of directing transcription (Fig. 5). There were no errors in all twelve sequences. This result demonstrated that, by replacing the 5' T + 5' A pairing with a double bp 5' TT + 5' AA reaction, we had overcome our previous limitations for directing three-molecule assembly.

Combinatorial promoter library design and construction

We used the results from the three-fragment promoter assemblies to design a multiplex method for generating random promoters: using one '5 G + 5' C pairing and one 5' TT + 5' AA pairing to direct and order assembly. In this scheme, an arbitrary promoter can be assembled from any combination of three duplex fragments. For each of the three fragment types, we designed sixteen variant sequences. The sixteen fragments of each type were assembled by SORSAL (Fig. 2) to

generate approximately 22,000 independent assemblies. This provided 5× coverage of the 16^3 = 4,096 possible sequences. We picked 10,000 of these constructs for storage and future characterization.

Diversity analysis

There are several ways in which an assembled sequence could become biased towards particular fragments, or incorrectly assembled. Imperfections in vector preparation could allow the vector to ligate without incorporating the duplex fragments (vector background). Even though the method used Watson-Crick base pairing to restrict the ways in which fragments could assemble, mispairing (Fig. 1, Table 1) could produce an incorrectly assembled sequence. In light of these complications, we examined the expected versus actual diversity of the assembled library.

To test the diversity of the SORSAL method, we sequenced a set of 288 randomly chosen clones. Of these, 2 (1%) were vector background and 6 (2%) were misassembly products. The remaining clones were correctly assembled sequences, giving an overall assembly yield of 280 (97%). Of the 280 correctly assembled clones, 217 (78%) were found to be unique combinations of three fragments. We expect that many of the 63 replicate sequences were introduced by handling error during plate picking and transport, as 32 of these replicates occurred in neighboring wells. These results show that, despite sources of error in handling and assembly, SORSAL generates a highly diverse set of correctly assembled sequences.

Within the 217 unique clones, we found 47 out of the 48 duplex fragments used in the assembly. We compared the observed distributions to a null hypothesis assuming equal probability of incorporation for all fragments (uniform distribution). In 217 unique assemblies, we would expect to find each fragment approximately 14 times on average. The observed frequency of each of the 48 fragments in the unique set of 217 experimental sequences is shown in Fig. 3A. The most over-represented fragments of each type were sampled 48, 31, and 32 times, respectively. We

note that the average relative deviation from a uniform distribution is greater for the 43 bp *distal* fragments (110%) than for the smaller 34 bp *proximal* (60%) and 23–24 bp *core* (60%) fragments. This suggests that the incorporation bias increases with duplex fragment length. Under the null hypothesis, maximum frequencies this great or greater would be observed by chance only 0.6% of the time (p-value of 0.006). Together, these results show that the library assembly is significantly—but not severely—biased, and represents a highly diverse set of sequences.

To estimate the expected diversity of the full library of 10,000 sequences, we simulated random assemblies under three sets of assumptions (Fig. 3B). As an extreme upper bound, we considered an ideal assembly in which each fragment was sampled independently from a uniform distribution (no bias). As an intermediate case, we sampled each fragment using the observed frequencies of Fig. 3A. Note that this "biased/unique" assumption precludes the possibility of finding the single missing *distal* fragment (out of the 48) within any of the 10,000 promoters. For an extreme lower bound, we used the frequency distributions from the 288 sequenced clones, taking into account all possible sources of error: handling, vector background, and misassembly. This lower bound indicates that the target diversity of 1,000 unique variants was achieved well within the 10,000 clones we picked.

The differences between the three simulations in Fig. 3B shows that the fragment bias (Fig. 3A) primarily limits the library diversity. All errors due to vector preparation, misassembly, colony picking, and plate handling become insignificant as the number of characterized clones grows large. This bias could be compensated for, by adjusting the initial concentrations of each fragment. An optional extension of the SORSAL method is to first generate a preliminary assembly using equal concentrations of fragments, sequence a small set of constructs at the Liquid Library step (Fig. 2A), and then use the initial fragment distributions to adjust the corresponding fragment concentrations for a larger-scale iteration. This strategy should allow for even more highly diverse constructions than the one presented here.

The complexity of a SORSAL library can be further increased to assemblies of 4 or more duplex fragments. Additional 2 bp overhangs (such as 5' TG + 5' CA) can be used to direct these

assemblies. In such cases, the total library diversity would be q^M for q variants each of M fragment types. Furthermore, synthetic oligonucleotides are currently commercially available for arbitrary sequences up to 200 bps. The acceleration of DNA synthesis technology will facilitate the extension of SORSAL towards assemblies of 1 kb and beyond. To extend this method even further, a robotic liquid handling system could be used to assemble each sequence explicitly and array them on 96-, 384-, or 1536-well microplates. The high efficiency of the assembly method determines the identity of each variant with 97% efficiency, without the need to explicitly sequence every construct.

Materials and Methods

Reagents

All chemicals were purchased from Sigma. Concentrations of antibiotics were 50 μg/mL kanamycin and 100 μg/mL ampicillin. All ligation reactions were carried out with 1.25 units of T4 DNA ligase and 0.1 mg/mL BSA, in 20 μL of T4 ligase buffer (Invitrogen). The restriction enzymes XhoI and BamHI used to prepare the vector plasmid were purchased from New England Biolabs. Library transformations used DH10B electromax electrocompotent cells (Invitrogen).

Duplex fragment preparation

For the single bp overhang ligation, each fragment was annealed separately. For the promoter library, a total of 48 duplex fragments were annealed from 96 synthetic oligonucleotides. Fragment sequences are provided in Chapter 2, Table S1. All synthetic oligonucleotides were prepared by total DNA synthesis and PAGE purification (University of Calgary DNA synthesis and sequencing center). The purification gels were imaged to ensure quality. Each synthetic oligonucleotide was mixed with its complement and normalized to 1 μM in 25 μL of T4 ligase buffer. To anneal the duplexes, each mixture was heated to 99° C for 10 min and then cooled to 0° C over one hour. Annealed duplexes were visualized by non-denaturing PAGE.

Single bp overhang ligation

We tested single bp overhang ligation in the presence and absence of 5' phosphorylation (Table 1). Combinations of fragments (200 nM each) were ligated at 4° C for up to four days. The ligation products were visualized after 24, 48, and 96 hours of ligation by 15% denaturing PAGE. Correct and incorrect assemblies were identified by ligation product length. To determine the relative reaction rates for each overhang type, variants containing all combinations of 5' phosphorylation states were compared on the same gel (not shown).

Randomized assembly ligation and cloning

To create the random library, all 48 duplex library fragments were mixed in equal proportions (50 nM each). This mixture was ligated for one week at 4° C. The crude ligation product was diluted 20× and combined with the plasmid vector pCS26 (Bjarnason et al. 2003), cut with XhoI and BamHI, to match the (unphosphorylated) 5' terminal overhangs on the assembled sequence ends. The vector-insert mixture was ligated for one week at 4° C, and transformed by electroporation (2.48 kV, 0.2 cm gap, 200 μF) into Electromax DH10B cells (Invitrogen). A fraction of the recovered transformation mix was spread onto selective plates, grown overnight, and counted. These colony counts provided the estimate of 22,000 independent assembly events.

Library handling and sequence characterization

The transformation mixes were directly inoculated into 50 mL LB cultures containing antibiotics and grown for 8 hours at 37° C. Harvested cells were used to prepare liquid libraries of midi-prep DNA (Qiagen) which were then re-transformed. Approximately 10,000 of these transformants were plated on selective media and picked into 35 384-well plates with a colony-picking robot (Norgren Systems).

We randomly selected 288 clones for sequencing. We first obtained mini-prep DNA from each clone (Qiagen Turboprep). Plasmid DNA was amplified by Accuprime supermix PCR (Invitrogen, 30 s

annealing at 55° C, 30 s extension at 68° C, 25 cycles) with primers pZE05 (CCAGCTGGCAATTCCGA) and pZE06 (AATCATCACTTTCGGGAA). The amplified DNA was digested for 1 hour with 5 units DpnI (NEB) to remove plasmid and genomic DNA, then PCR purified and commercially sequenced using primer pZE06 (Laragen Inc., Los Angeles, CA).

The sequence traces were analyzed individually for quality (4Peaks by A. Griekspoor and Tom Groothuis, mekentosj.com). Of the 288 clones sequenced, 2 vector background and 6 misassembly ligations were identified. The remaining correctly assembled promoters were genotyped by automated Smith-Waterman sequence alignment (Waterman et al. 1984) with the 48-unit sequences, and the three corresponding fragments were identified. We were not able to analyze point mutations in the multiplex library, since any observed mutations could come from SORSAL or from the PCR amplification step used for sequencing.

Table 1. Single bp overhang ligation

Overhang	Rate	Fidelity[1]	Unphosphorylated	Promiscuity[2]	Sodium[3]
5' C	Fast	High	No reaction	C, T, A	Slow
5' G	Fast	High	No reaction	T	Slow
5' T	Slow	High	No reaction	C, G, T	No reaction
5' A	Slow	Low	Slow, non-specific	C, A	No reaction
3' T	Very slow	Low	No reaction	ND	ND
3' A	Very slow	Low	No reaction	ND	ND
5' GATC[4]	Very fast	Low	ND[5]	ND	ND

1 The preference of an overhang for its complement sequence, compared to an available non-complementary overhang. For example, the 5' A (Fig. 1A, Lanes 1–2) has low fidelity, while 5' T (Lane 3) has high fidelity and produces only one product.

2 Mispairings observed in the absence of the complementary overhang. For cases where multiple promiscuous interactions were found, they are ordered by decreasing relative rate.

3 The relative reaction rate when the ligation was run in the presence of 0.2M NaCl

4 This four bp 5' overhang was analyzed as a positive control in the presence of an unphosphorylated 5' TCGA overhang.

5 ND = "not determined"

Figure 1.

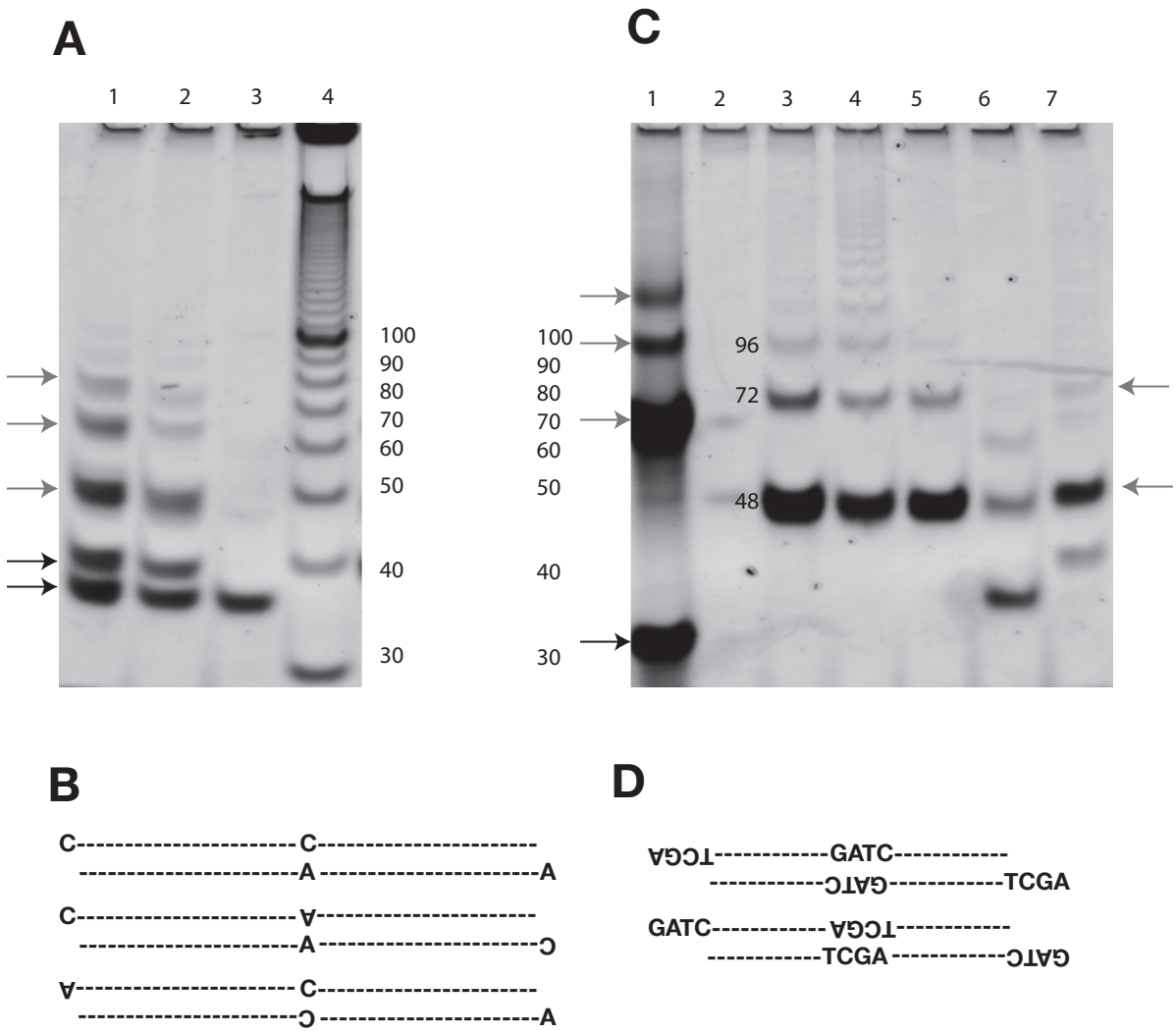


Figure 1. Single bp overhang ligation produces both correct and misassembled products. (A) The 5' T + 5' A ligation produces correct assembly products (black arrows) along with higher-molecular weight-misassemblies (grey arrows). Lane 1 contains products of a two-fragment ligation reaction. The first fragment has phosphorylated 5' A and C overhangs, while the second fragment has a single phosphorylated 5' T overhang. In Lane 2 only the 5' T and A overhangs are phosphorylated, while in Lane 3 only the 5' T overhang is phosphorylated. Lane 4 is a size ladder. (B) A single 24 bp fragment with an overhanging 5' A and 5' C can produce multiple misassembly products, depending on the relative rates of the three mispairings: 5' A + 5' A, 5' A + 5' C, and 5' C + 5' C. (C) Misassemblies formed from single molecule reactions. In Lane 1, a fragment containing two 4 bp overhangs forms both the correct product (black arrow) and high-molecular-weight-misassemblies (grey arrows). Lanes 3–5 contain self-ligation reactions of the fragment shown in (B). When only the 5' C is phosphorylated (Lane 5), multimers are found of 2–4 units (38, 72, and 96 bps). When only the 5' A is phosphorylated (Lane 4), a ladder of multimers is formed. When both 5' A and 5' C are phosphorylated (Lane 3), the behavior is intermediate. Lanes 6 and 7 show the effect on the self-assembly reaction of Lane 5, when a second molecule containing a phosphorylated (Lane 6) or unphosphorylated (Lane 7) 5' T overhang is added. (D) The correct (top) and incorrect (bottom) ligations from the 4 bp overhang control (Part C, Lane 1). For this control, only the GATC overhang is phosphorylated, though the unphosphorylated TCGA participates in promiscuous ligation.

Figure 2.

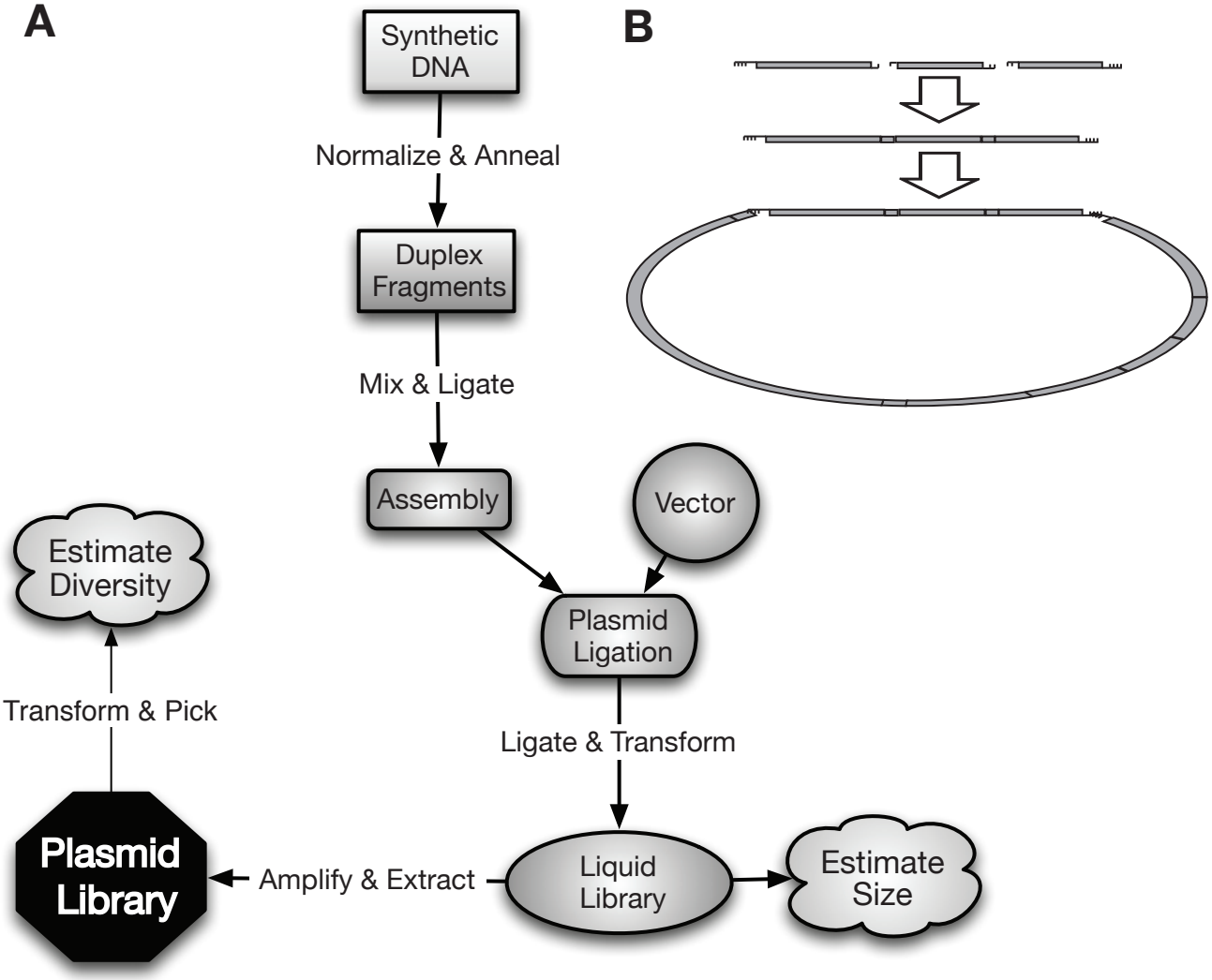


Figure 2. *The SORSAL protocol is fast and extensible.* The procedure for building the combinatorial library (Methods) is shown. (A) Purified synthetic oligonucleotides are normalized and annealed to form duplex fragments. These fragments are mixed and ligated at low temperature. The unpurified ligation mix is combined with a cut vector and transformed to create the liquid library. The plasmid DNA is then harvested, retransformed, plated, and picked into microwell plates. Individual clones can be restreaked and sequenced to determine library diversity. (B) In a single self-assembly ligation step, three synthetic fragments are assembled. In the cloning ligation step, the self-assembly products are ligated into the plasmid vector.

Figure 3.

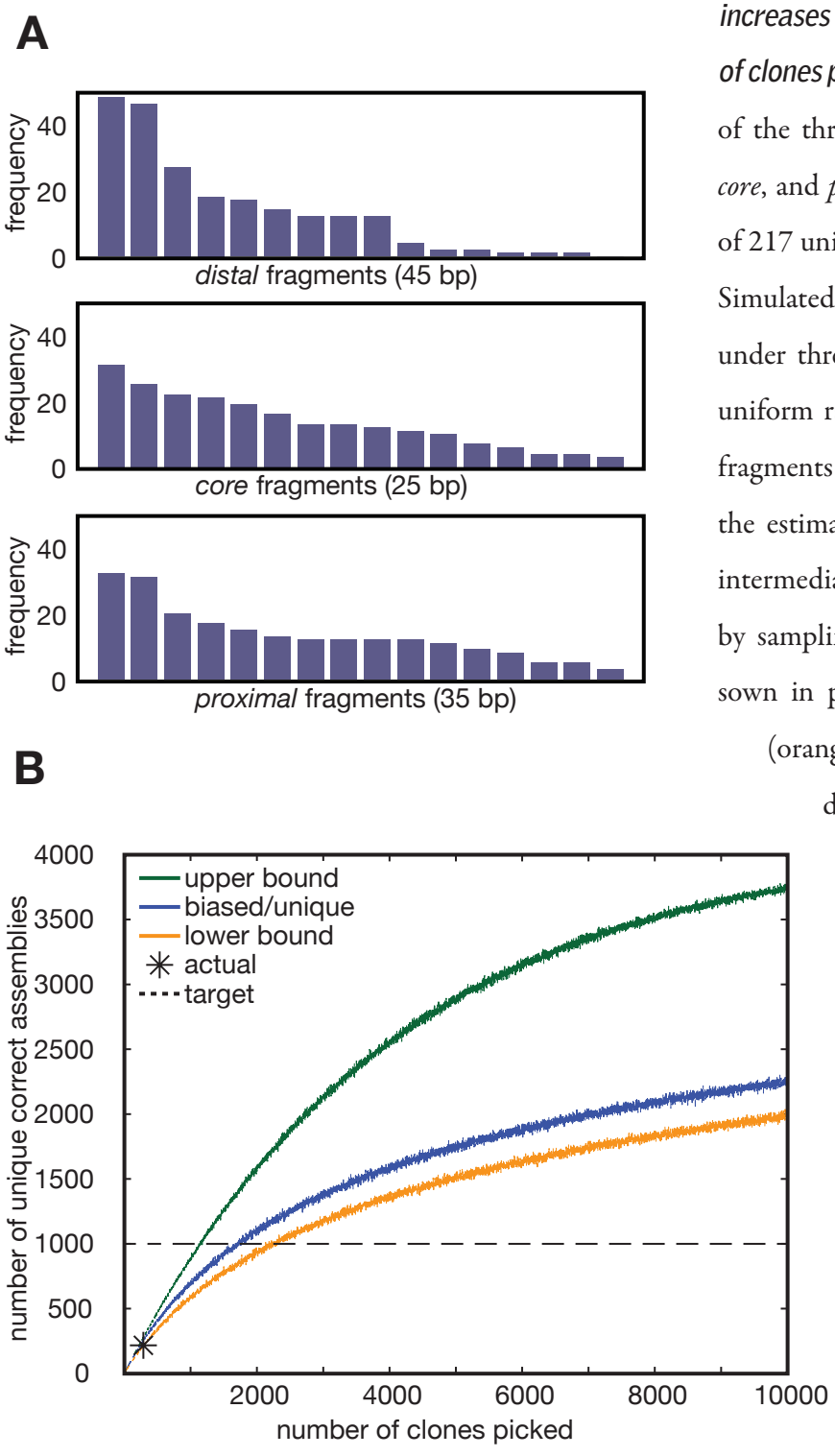
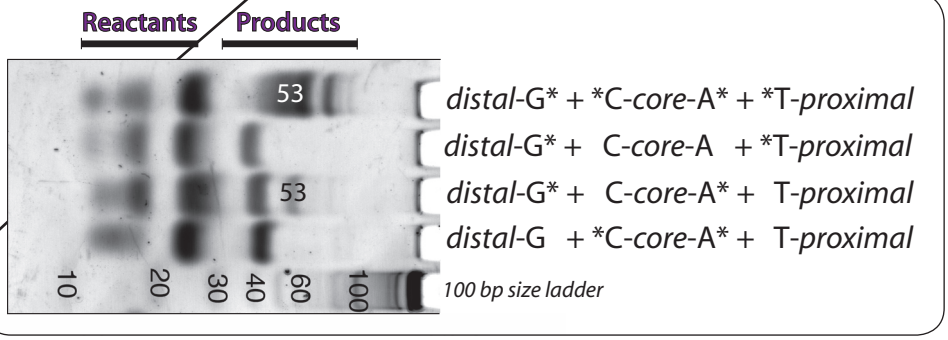


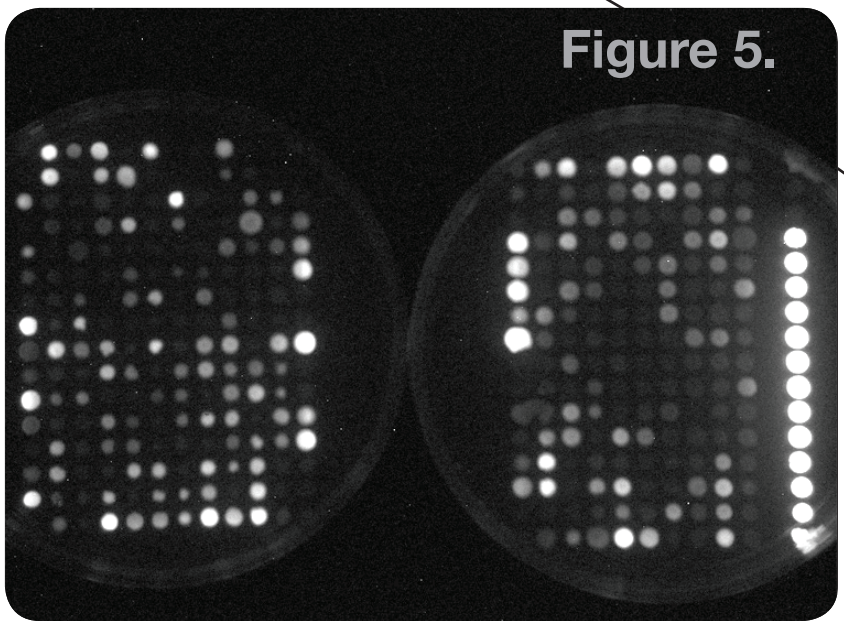
Figure 3. *Library diversity increases quickly with the number of clones picked.* (A) The frequencies of the three fragment types (*distal*, *core*, and *proximal*) found in the set of 217 unique sequenced clones. (B) Simulated diversity of the library under three assumptions: The ideal uniform random assembly of all 48 fragments gives an upper bound on the estimated diversity (green). An intermediate case (blue) is generated by sampling from the distributions sown in part A. The lower bound (orange) uses the frequency distributions in the 288 sequenced clones, and includes errors in cloning, assembly, and handling. The target diversity of 1,000 unique sequences (dashed line) is achieved in each case for differing numbers of clones picked (*x*-axis). The experimental diversity found in the sequenced set (217 out of 288) is indicated with a black star.

Figure 4. *Single bp overhangs direct the assembly of three duplex fragments.*



This gel shows the assembly of a 53bp duplex sequence from three fragments of 14 (*distal*), 24 (*core*), and 12 (*proximal*) bps. Four reactions with different combinations of 5' phosphorylation are shown, after 96 hours of ligation. The first reaction contains 5' phosphorylation at all four cohesive ends. This reaction generates the highest yield of correct 53 bp product, but also generates several higher molecular weight misassemblies. The second and fourth reactions cannot generate the full length products. The third reaction, which generates one of the two 53 bp strands, exhibits fewer side-reactions at the cost of significantly reduced product yield.

Figure 5. *SORSAL generates a large diversity of promoter strengths.*



This image shows the light generated by 350 glowing bacterial colonies in strain DH10B. Each colony contains a SORSAL reaction product promoter, cloned into a luciferase reporter plasmid. The twelve bright clones in the rightmost column are the sequenced σ^{70} promoters constructed explicitly to test the 5' TT + 5' AA ligation reaction. The remaining clones are random library constructs expressing luciferase. Background, misassembled, weak, and non-functional promoter sequences appear as dark clones.

References

Arnold FH (1998) Design by directed evolution. Accounts Chem Res 31: 125-131.

Atsumi S, Little JW (2006) A synthetic phage lambda regulatory circuit. Proc Natl Acad Sci U S A 103: 19045-19050.

Bjarnason J, Southward CM, Surette MG (2003) Genomic profiling of iron-responsive genes in Salmonella enterica serovar typhimurium by high-throughput screening of a random promoter library. J Bacteriol 185: 4973-4982.

Dolle RE, Le Bourdonnec B, Morales GA, Moriarty KJ, Salvino JM (2006) Comprehensive survey of combinatorial library synthesis: 2005. Journal of combinatorial chemistry 8: 597-635.

Fischbach MA, Lai JR, Roche ED, Walsh CT, Liu DR (2007) Directed evolution can rapidly improve the activity of chimeric assembly-line enzymes. Proc Natl Acad Sci U S A 104: 11951-11956.

Geu-Flores F, Nour-Eldin HH, Nielsen MT, Halkier BA (2007) USER fusion: a rapid and efficient method for simultaneous fusion and cloning of multiple PCR products. Nucleic Acids Res 35: e55.

Guet CC, Elowitz MB, Hsing W, Leibler S (2002) Combinatorial synthesis of genetic networks. Science 296: 1466-1470.

Hammer K, Mijakovic I, Jensen PR (2006) Synthetic promoter libraries—tuning of gene expression. Trends in biotechnology 24: 53-55.

Hawley DK, McClure WR (1983) Compilation and analysis of *Escherichia coli* promoter DNA sequences. Nucleic Acids Res 11: 2237-2255.

Ligr M, Siddharthan R, Cross FR, Siggia ED (2006) Gene expression from random libraries of yeast promoters. Genetics 172: 2113-2122.

Meyer MM, Hiraga K, Arnold FH (2006) Combinatorial Recombination of Gene Fragments to Construct a Library of Chimeras. Current Protocols in Protein Science.

Pfleger BF, Pitera DJ, Smolke CD, Keasling JD (2006) Combinatorial engineering of intergenic regions in operons tunes expression of multiple genes. Nature biotechnology 24: 1027-1032.

Rozenman MM, McNaughton BR, Liu DR (2007) Solving chemical problems through the application of evolutionary principles. Current opinion in chemical biology 11: 259-268.

Saftalov L, Smith PA, Friedman AM, Bailey-Kellogg C (2006) Site-directed combinatorial construction of chimaeric genes: general method for optimizing assembly of gene fragments. Proteins 64: 629-642.

Sprinzak D, Elowitz MB (2005) Reconstruction of genetic circuits. Nature 438: 443-448.

Stemmer WP (1994) Rapid evolution of a protein in vitro by DNA shuffling. Nature 370: 389-391.

Treynor TP, Vizcarra CL, Nedelcu D, Mayo SL (2007) Computationally designed libraries of fluorescent proteins evaluated by preservation and diversity of function. Proc Natl Acad Sci U S A 104: 48-53.

Waterman MS, Smith TF, Katcher HL (1984) Algorithms for restriction map comparisons. Nucleic Acids Res 12: 237-242.

Programming Gene Expression with Combinatorial Promoters

Robert Sidney Cox III, Michael G. Surette, and Michael B. Elowitz

Abstract

Promoters control the expression of genes in response to one or more transcription factors. The architecture of a promoter is the arrangement and type of binding sites within it. In order to understand natural genetic circuits, and to design promoters for synthetic biology, it is essential to understand the relationship between promoter function and architecture. We constructed a combinatorial library of random promoter architectures. We characterized 288 promoters in *Escherichia coli*, each containing up to three inputs from four different transcription factors. The library design allowed for multiple -10 and -35 boxes, and we observed varied promoter strength over five decades. In order to further analyze the functional repertoire we defined a representation of promoter function in terms of regulatory range, logic type, and symmetry. Using these results we identified heuristic rules for programming gene expression with combinatorial promoters.

Introduction

In many promoters, gene expression is regulated in response to two or more transcription factors (TFs). A classic example is the *lac* operon, where promoter activity depends on both the repressor LacI (Jacob and Monod, 1961) and the activator CRP (Zubay et al. 1970). Such combinatorial regulation of gene expression underlies diverse cellular programs (Ptashne, 2005), including responses to environmental conditions (Ligr et al. 2006) and multicellular development. Combinatorial promoters with multiple TF binding sites, or operators, can facilitate the integration of multiple signals. For example, a synthetic combinatorial promoter responding to LuxR and λ cI was recently used to construct a genetic pulse-generator (Basu et al. 2004), a band-pass filter, and a bulls-eye pattern formation system (Basu et al. 2005). Furthermore, circuits containing combinatorial promoters are predicted to generate robust oscillations (Hasty et al. 2002; Atkinson et al. 2003); or to create sign-sensitive filters, signal averaging, response acceleration or delay (Mangan and Alon, 2003).

Bacterial promoters typically occupy a region of 100bp or less, surrounding the start site (+1) of transcription, from approximate positions -75 to +25. This sequence includes the primary binding sites for the polymerase, the -10 and -35 boxes (Hawley and McClure, 1983), additional upstream (Chan and Busby, 1989; Ross et al. 1993) and downstream (Haugen et al. 2006; Kammerer et al. 1986) regulatory sequences, along with operators for activator and/or repressor TFs (Browning and Busby, 2004; Busby and Ebright, 1994). Operators within this region enable bound TFs to directly contact and recruit the polymerase (activation) or to sterically block polymerase contact with the -10 and -35 boxes (repression). The type and arrangement of these regulatory sequences and operators within the promoter region specify the promoter architecture.

Genome sequencing and annotation reveals the identity and placement of the TF operators in natural promoters (Collado-Vides et al. 1991; Gralla and Collado-Vides, 1996; Salgado et al.

2006). In these and related works, the distributions of TF operators in *E. coli* have highlighted trends in the operator positions relative to the polymerase box sequences. For example, it was found that activator operators occur principally around -40, whereas repressor operators were clustered from -60 to +20. These studies proposed that activation is effective only on promoters with low unregulated activity, such as in promoters containing a weak -35 box. The 'effective repression' of a promoter, defined as the ratio of expression in 'on' and 'off' states, was expected to be highest for promoters of strong unregulated activity. These results indicated that repression and activation are most effective at different promoter locations, and on different intrinsic promoter strengths.

The potential diversity of promoter architecture and functionality is large when one considers the many known mechanisms by which proteins and DNA interact. Here we focus on the simplest promoter architectures regulated by multiple TFs, and ask what types of regulation functions are possible. Classical descriptions of gene networks have used Boolean logic to describe combinatorial regulation (Kauffman, 1969; Thomas and D'Ari, 1990). However, because the output of a promoter is not a binary function of the concentrations of its regulators (Atsumi and Little, 2006; Guido et al. 2006; Mayo et al. 2006; Setty et al. 2003), a range of non-Boolean logical phenotypes are possible. Recent theoretical descriptions of transcriptional logic (Bintu et al. 2005b; Buchler et al. 2003; Hermsen et al. 2006) have focused on the effects of explicit TF-TF contacts and operator overlap, but it is not known whether such interactions are necessary to generate diverse phenotypes.

In order to better understand natural promoter function and to improve the design of new promoters for synthetic biology applications (Endy, 2005; Hasty et al. 2002; Sprinzak and Elowitz, 2005), we report a synthetic library-based approach for construction and analysis of modular combinatorial promoters. Here, we varied the placement, affinity, and sequence of known operators (Supplementary Table S1), allowing us to determine the range of functions

encoded by the simplest combinatorial promoters. This approach reveals fundamental features of the relationship between promoter architecture and function.

Results

Combinatorial library design and assembly

We developed an efficient method for assembling promoters from modular components. The method uses three classes of synthetic duplex DNA units with compatible 5' cohesive ends. These units correspond to the 45bp region upstream of the -35 box (*distal*), the 25bp region between the -35 and -10 boxes (*core*), and the 30bp region downstream of the -10 box (*proximal*). In this scheme, an arbitrary promoter can be assembled from any combination of *proximal*, *core*, and *distal* units. The internal 5' overhangs determine each unit's placement in the promoter (Fig. 1A). We assayed promoter activity using a bacterial luciferase reporter cassette on a low copy plasmid (Fig. 1B-C). Here we report all promoter activities in terms of arbitrary luminescence units (ALU).

We incorporated operators for two activators and two repressors: The activator AraC (Ogden et al. 1980; Schleif, 2003) regulates arabinose metabolism in *E. coli*, while LuxR activates luminescence genes in *Vibrio fischeri* (Fuqua et al. 1994). The repressor LacI (Jacob et al. 1961; Setty et al. 2003) controls the metabolism of lactose in *E. coli*, while TetR represses the tetracycline resistance genes in transposon Tn10 (Beck et al. 1982; Skerra, 1994). The two activators are active only in the presence of the corresponding inducers L(+)-arabinose (Lara) and oxo-C6-homoserine lactone (VAI), respectively. The repressors TetR and LacI are inactivated by the inducers anhydrotetracycline (aTc) and isopropyl β-D-1-thiogalactopyranoside (IPTG), respectively. Consequently, induction of each factor (AraC, LuxR, LacI, or TetR) is expected to increase a target promoter's activity. These four TFs bind specifically to well-defined operators, are dispensable, and can be induced by small molecules without disrupting normal cellular processes.

For each position (*distal*, *core*, and *proximal*) we designed five unregulated and eleven operator-containing units. These sequences varied the affinity, location, and orientation of operators (Fig. 1D). The design also allowed for variable -10 and -35 boxes, to encourage diverse expression levels. The sixteen units of each type were assembled by randomized assembly ligation (Methods) to generate a plasmid library containing approximately 22,000 independent assemblies, and providing 5× coverage of the $16^3 = 4{,}096$ possible promoters. We transformed the plasmid library into *E. coli* strain MGZ1X expressing LacI, TetR, AraC, and LuxR. We then sequenced a set of 288 randomly chosen transformants, and found 280 correctly assembled promoters (Supplementary Data 1). We determined 217 of these promoters to be unique. Within this set, 47 out of the 48 possible units were represented at least once. Thus, the randomized assembly ligation method produced a diverse set of correctly assembled promoters.

Library functions

We measured the expression of the 288 sequenced transformants in each of 16 combinations of the 4 chemical inducers (Fig. 1C, Methods, Supplementary Data 2). The library showed 5 decades of variation in promoter activity (Supplementary Figure S1). Promoters of high unregulated activity contained strong -10 and -35 boxes, though the presence of consensus box sequences did not predict unregulated promoter activity (Supplementary Figure S2). Of the 217 unique promoters, 83% produced measurable expression in at least one of the 16 conditions, and 49% changed expression by a factor of 10 or more. Of these 106 clones, 79 were found to respond to a single inducer and 27 responded by more than two-fold to two inducers. No promoters were found to respond more than two-fold to three or four inducers, or to decrease expression to less than half in the presence of an inducer (anti-induction). All of the dual-input promoters measured increased their activity monotonically in response to the inducer concentrations, both singly and in combination. Overall, the promoter library exhibited a diverse set of behaviors across the 16 conditions.

How does promoter architecture constrain function? For each promoter, we compared the architecture (Supplementary Data 1) to the measured response (Supplementary Data 2). We found no significant regulation without the presence of a corresponding operator (Supplementary information). The relationship between sequence and phenotype revealed several rules relating promoter architecture to promoter function, which we describe below.

Single-input gates

The simplest promoters, termed Single-Input Gates (SIGs) responded to a single inducer (Figures 2 and 3). For these switch-like gates we defined the regulatory range, r, as the ratio of the induced to uninduced activity. Within this group activated SIGs showed regulatory ranges up to $r = 10^3$, whereas the repressed SIGs exhibited higher regulatory ranges up to $r = 10^5$ (Table 1).

Activated SIGs

Activated expression level was independent of unregulated activity (Fig. 2A). The best activated SIGs (highest r) occurred at promoters with low expression in the unregulated state. Activation was ineffective for promoters with unregulated activity above approximately 10^5 ALU, which is 30-fold lower than the strongest promoter activity measured. This "activation ceiling" was the same for both AraC and LuxR activated promoters. These results show that activation is limited by the absolute expression level, and is most effective on promoters of low intrinsic activity; consistent with previous suggestions (Busby et al. 1994; Gross et al. 1998).

Activation functioned only at the *distal* position (Fig. 2), in accordance with previous studies of AraC and LuxR (Collado-Vides et al. 1991; Egland and Greenberg, 1999). We found neither inducible activation nor inducible repression by LuxR or AraC at *core* or *proximal* (Fig. 2B). In such promoters, the typical induction response was only 6% for LuxR and 11% for AraC regulation. Some of the strongest activated SIGs (Table 1) had additional activator operators at *core* or *proximal* sites, along with a functional operator at *distal*. We found that activator

binding to *core* and *proximal* did not, on average, strongly affect the maximal promoter activity (Supplementary Information, Supplementary Figure S3). These results show that AraC and LuxR have neither positive nor strong negative regulatory effects on gene expression at the *core* and *proximal* regions.

Repressed SIGs

In contrast to activation, repression occurred effectively at all three positions (Fig. 3). However, we found a clear trend between operator location and repression. Repression was most effective at *core* (Fig. 3B), followed by *proximal* (Fig. 3C) and then *distal* (Fig. 3A). Within this trend, we found that the promoters of low unregulated activity were less sensitive to operator position. This result shows that repression is effective at all three positions, with relative strength following the rule *core* ≥ *proximal* ≥ *distal*.

As with activation, the expression level in the repressed state was not determined by the unregulated level. Examples of completely repressed expression were observed at every level of unregulated promoter activity (Fig. 3). In fact, some repressed SIGs exhibited the highest activities observed ($>10^6$ ALU) upon induction (Table 1). Within the limits of detection, the effective repression (r) tended to increase with unregulated expression level.

Strikingly, the SIG showing the strongest regulation ($r = 8.9\times10^4$, Table 1, D18) had only a single TetR operator at the *core* region. Furthermore, a single repression site at any of the three positions was often enough to repress the promoter below the detection limit (Fig. 3). In general, multiple operators were not more effective at repression than single operators. We found 9 LacI-regulated and 6 TetR-regulated SIGs containing multiple repressor operators. Of these, only one LacI-regulated (Table 1, A38) and one TetR-regulated (Table 1, B19) promoter produced higher regulation than corresponding promoters containing a single operator. These results show that operator position is more important than operator multiplicity for achieving strong regulation with repressors.

Dual-input logic

We next considered dual-input gates as logic functions of their two input inducers. Because of the continuous nature of the output levels in each input state, Boolean logic does not accurately represent the space of possible functions. For example, in a recent study the natural *lac* promoter increased activity by a factor of 3.6 when induced by cAMP alone, by a factor of 7.1 when induced by IPTG alone, and by a factor of 14 when induced by both simultaneously (Setty et al. 2003). This intermediate behavior could be described as either AND-like or OR-like, depending on the activity threshold chosen.

In order to describe such "intermediate logic" phenotypes, we introduced a 3-dimensional parameterization for the space of promoter functions. In this scheme, we represented the promoter functions with three numerical parameters that quantify dynamic range, logic type, and asymmetry of inputs (Methods). As before, r is the ratio of the maximum to minimum promoter activity. The parameter l quantifies the logical behavior of the promoter: from pure OR ($l = 0$) to pure AND logic ($l = 1$). Finally, the parameter a quantifies the asymmetry of the gate with respect to its two inducers. At $a = 0$, the gate responds identically to either inducer, while at $a = 1$, the promoter responds to one input only (pure SIG). These parameters span the full range of observed phenotypes, and have intuitive interpretations. They also represent relative promoter activities, rather than absolute levels, making them less sensitive to the choice of reporter, growth media, or other experimental conditions. Therefore, they form an ideal quantitative representation for the phenotypic behavior of these promoters.

Within this logic-symmetry space, the positive monotonic response of promoters to their inputs restricts promoter logic to the triangular region shown in Figure 4. The corners of this region include three Boolean logic functions: the switch-like SIG ($l = 0.5$, $a = 1$), along with the canonical binary gates AND ($l = 1$, $a = 0$) and OR ($l = 0$, $a = 0$). The symmetric SLOPE gate ($l = 0.5$, $a = 0$) exhibits logic intermediate between AND and OR. The asymmetric asym-AND

($l = 0.75$, $a = 0.50$), asym-OR ($l = 0.25$, $a = 0.50$), and asym-SLOPE ($l = 0.50$, $a = 0.50$) gates describe idealized logic functions intermediate between SIG and AND, SIG and OR, and SIG and SLOPE, respectively (Fig. 4A). This representation provides qualitative categories for the different types of logic displayed by monotonic dual-input promoters.

We identified 50 dual-input gates (Methods). Each defined a point (r, a, l) in the logical phenotype space (Fig. 4B), revealing a range of functional behaviors. Asym-AND and SIG-like gates exhibited strong regulation up to $r = 10^5$. The AND and asym-SLOPE gates were regulated up to $r = 10^4$, while the SLOPE gates were regulated up to $r = 10^3$. Notably, we found no gates exhibiting strong OR or asym-OR logic functions. However, one class of dual-input promoters (discussed below) exhibited asym-SLOPE logic approaching an asym-OR response ($l < 0.50$). Thus we observed a wide distribution of promoter logic types.

The library contained two classes of dual-input gates. The repressor-repressor (RR)-promoters contained operators for the repressors LacI and TetR, while the activator-repressor (AR)-promoters responded to the activator AraC and one of the repressors. Due to the relative scarcity of LuxR-activated promoters, we did not find LuxR regulated AR promoters in the characterized promoter set (Fig. 2A). These two classes of dual-input gates exhibited differing, but overlapping, distributions of logical phenotypes.

Comparison of AR and RR promoter phenotypes (Fig. 4B) revealed that each has a preference for different logical categories, although both produced strong asym-AND gates. The RR promoters produced the strongest symmetric (AND and SLOPE) gates, whereas the AR promoters generated the strongest asym-SLOPE gates. This shows that RR promoters produced both symmetric and asymmetric logic, while AR promoters produced only asymmetric logic.

Mathematical model of repressor interaction

To better understand the variety of symmetric and asymmetric logic observed for the RR promoter class, we employed a simple model of promoter activity in the presence of two repressors (Methods). In this model (c_1, c_2, ω) represent the strength of repression at the stronger operator, the weaker operator, and the repressor-repressor interaction, respectively (Bintu et al. 2005b). When the repressors do not interact with each other, $\omega = 1$; whereas for exclusive interactions (only one repressor can bind at a time), $\omega = 0$. Cooperative interactions would correspond to $\omega > 1$.

The logic parameter l was tightly coupled to the model interaction parameter ω (Methods). A plot of a and l as parameterized functions of the microscopic model parameters (Supplementary Figure S4) showed that RR promoters with ω ranging from 0 (exclusive interaction) to 1 (independent interaction) can produce any logic function in the right half ($l \geq 0.5$) of the phenotype space triangle: SIG, AND, SLOPE, asym-AND, and asym-SLOPE. In particular, exclusive interaction ($\omega = 0$) approached pure AND logic ($l = 1$) whereas independent interaction ($\omega = 1$) always resulted in SLOPE-like logic ($l = 0.5$). Conversely, we found that an asym-OR gate would require extremely high cooperative interaction ($\omega = 100$); while an ideal OR gate would require infinite cooperativity. Therefore, the range of logic functions displayed by the library RR promoters (Fig. 4B) fall within the spectrum of non-cooperative interactions ($1 \geq \omega \geq 0$). This model demonstrates that a variety of logic functions can be achieved without explicit protein-protein cooperativity.

RR promoters

Dual-repression can be either symmetric or asymmetric (Fig. 4B), with either repressor dominant (Fig. 5A). As with the SIGs, even the strongest RR promoters could be fully repressed, exhibiting effective repression up to $r = 10^5$. RR promoter logic was always AND-like or SLOPE-like ($0.5 \leq l \leq 1.0$), indicating that there were no instances of strong cooperative interaction between the repressors ($\omega \leq 1$). In three cases, mutation of a repressor operator resulted in almost completely asymmetric ($a = 1$) SIG logic (Fig. 4B, top of triangle). In other cases the repression was more

balanced ($a < 0.25$), producing symmetric AND and SLOPE responses up to $r = 10^4$. Thus, RR promoters displayed a large range of dual-input regulatory logic including AND, SLOPE, asym-SLOPE, and asym-AND gates.

In principle, the logic phenotype displayed by a promoter could depend on the inducer concentrations used. Therefore, we chose three RR promoters (Fig. 5A, clones A3, D8, and D9), and measured their responses to 16 combinations of inducer concentrations (Supplementary Methods). As expected, all three promoters increased their activity monotonically with increasing concentrations of each inducer. As shown in Supplementary Figure S5, inducer concentrations primarily affected r and a, while the logic parameter l was less dependent (Supplementary information). The most AND-like gate (A3) had the highest variation in logic ($l = 0.46$ to $l = 0.86$), while the most SLOPE-like (D9) exhibited the narrowest range ($l = 0.48$ to $l = 0.53$). These results imply that r and a depend strongly on input concentration; while for l, independent (SLOPE) logic is more robust than exclusive regulation (AND).

The repressor operator location trend *core* ≥ *proximal* ≥ *distal* explains the combinatorial promoter behaviors shown in Fig. 5A. For RR promoters, the position of the operators determined whether LacI or TetR was dominant. We found only one clear exception to this trend (Fig. 5A, clone A3), where TetR acting at *proximal* slightly dominates LacI acting at *core.* Symmetric repression occurred for several architectures, such as with a TetR at *core* and two LacI operators, one at *distal* and the other at *proximal* (Fig. 5A, A28). In all other asymmetric cases core dominated proximal and distal, while proximal dominated distal. RR promoter architectures with operators at *proximal* and *distal* produced the largest range of logic behaviors including AND, SLOPE, asym-AND, and asym-SLOPE. RR promoters with operators at the *core* and *proximal* positions produced only AND and asym-AND logic. Of the 7 RRpromoters exhibiting strong AND-like logic ($l > 0.8$), 5 had operators at *core* and *proximal*. Finally, RR promoter architectures with operators at *core* and *distal* produced the most asymmetric logic functions (e.g., Fig. 5A, B83):

the repressor acting at *core* was always strongly dominant. These results show that repressor dominance in combinatorial promoters follows the trend *core* ≥ *proximal* ≥ *distal*, and that close operator proximity is consistent with AND-like logic.

AR promoters

Among AR promoters (Fig. 5B), repression always dominated activation ($0.06 \leq a \leq 0.99$). The AR promoters were regulated by AraC, in combination with LacI or TetR, and exhibited regulation up to $r = 10^4$. In all cases the activator functioned from the *distal* region, while the repressor functioned at *core* or *proximal*. We found one AR promoter that approached symmetric response ($r = 3272$, $a = 0.06$, $l = 0.81$, Fig. 5B, D61). The three most AND-like ($l > 0.8$) promoters of this class had the repressor operator at the *core*. The most OR-like (smallest l) promoter exhibited asym-SLOPE logic ($r = 9112$, $a = 0.65$, $l = 0.46$, Fig 5B, A54), with the repressor operator at *proximal*. Therefore, we found AR promoters are well represented by asym-AND when the repressor acts as *core* and asym-SLOPE when the repressor acts at *proximal*.

The AR promoters also confirmed our previous result relating activation to intrinsic promoter activity: The higher the unregulated activity of an AR promoter (+ IPTG/aTc, -Lara), the smaller the change upon activator induction (compare the last two columns in Fig. 5B). When the unregulated activity exceeded the activation ceiling, the AR promoter did not respond to AraC induction at all, resulting in SIG-like behavior (e.g. Fig. 5B, D46). This result indicates that AR promoters will depend on both inputs only when the unregulated promoter activity is below the activation ceiling.

Discussion

Combinatorial synthesis of synthetic promoters, as described here, permits systematic analysis of promoter architecture and rapid identification of promoters that implement specific functions. The spectrum of promoter functions observed in this library highlights several heuristic rules for promoter design:

1. *Limits of regulation*. Gene expression can be regulated over five orders of magnitude. Regulated promoter activity is independent of unregulated activity. As a result, effective repression tends to increase with unregulated activity, while activation tends to decrease. Activation is limited by an absolute level of expression, at around 2.5% the level of the strongest unregulated promoter activities.
2. *Repressor operator location*. The effectiveness of repression depends on the operator location with *core* ≥ *proximal* ≥ *distal*. Dual-repression may be symmetric or asymmetric, with the dominant repressor predicted by operator locations.
3. *One is enough*. Full repression is possible with a single operator between -60 and +20 at high repressor concentrations. Activators function only upstream of -35 (*distal*), and have little positive or negative effect downstream at *core* or *proximal*.
4. *Repression dominates activation*, producing asymmetric logic.
5. *Operator proximity*. Independent regulators generate SLOPE-like logic. Operator proximity increases competitive interactions, making the logic more AND-like.

For both activation and repression, the activity of the promoter in the regulated (activated/repressed) state is not determined by the activity in the unregulated state (Rule 1). Intuitively, activation has higher r when the unregulated activity is low, and repression has higher r when the unregulated activity is high. Furthermore, as predicted by recent theoretical work (Bintu et al. 2005a); repression is able to achieve extremely high levels of regulation ($r \leq 10^5$), while activated regulation is moderately strong ($r \leq 10^3$). These limits apply to both SIGs (Figs. 2-3) and dual-

input promoters (Fig. 5). AR promoters are a special case, and exhibit a trade-off: Increasing the unregulated activity increases the regulatory range (*r*), at the expense of greater asymmetry (*a*). For example, compare the first and last promoter in Fig. 5B.

Rules 2 and 3 summarize the operator position and multiplicity effects for both activators and repressors. The repression trend (Rule 2) has been previously reported for promoters regulated by LacI (Elledge and Davis, 1989; Lanzer and Bujard, 1988). The authors of the first paper proposed a mechanistic model involving two competing effects: *Core* and *proximal* sites more effectively block polymerase binding, while *core* and *distal* sites bind repressor more rapidly (are more accessible) as the polymerase initiation complex clears the -10 and -35 boxes. We confirmed the operator location trend for SIGs regulated by LacI and TetR alone, and found that this heuristic also holds for RR promoters of both repressors. Of course, differences in operator affinity, repressor concentration, and repressor structure can overcome these rules.

We compared Rules 2 and 3 with the distribution of known *E. coli* operators compiled from 1,102 natural promoters in the database RegulonDB (Salgado et al. 2006) (Figure 6). In agreement with analysis made on earlier versions of the database (Collado-Vides et al. 1991; Gralla et al. 1996), we found that activator operators are most common in the *distal* region (Fig. 6A), while repressor operators cluster around all three promoter regions (Fig. 6B). Fig. 6C shows the operator density of the 554 promoters which are recognized by the polymerase subunit σ^{70}. The small regulatory effect observed for activator operators in the *core* and *proximal* regions (Rule 3) appears consistent with the general scarcity of natural activator sites in these regions. Similarly, the density of repressor operators found in σ^{70} promoters is significantly enriched for *core* sites over *distal* and *proximal* locations, consistent with the repressor operator location trend (Rule 2).

The sufficiency of one operator for repressing promoter activity up to five orders of magnitude (Rule 3) raises the classic question of why natural promoters are so often regulated by redundant

operators (Collado-Vides et al. 1991). Our study used high concentrations of repressors in the range of 2-4 μM (Lutz and Bujard, 1997), paired with strong operators (Table S1). At lower repressor concentrations and operator affinities, the presence of multiple binding sites can increase the effective repression *r* through looping (Becker et al. 2005; Vilar and Leibler, 2003), cooperativity (Oehler et al. 1994; Ptashne, 2004; Rosenfeld et al. 2005), or even without explicit TF-TF interactions (Bintu et al. 2005a). These effects can also increase the steepness of response to repressor concentration (Ptashne, 2004), or engender exceptions to the dominance of repression (Rule 4). Finally, the presence of multiple operators might increase the mutational plasticity of promoter functions (Mayo et al. 2006).

Rule 5 provides insight for both AR and RR promoters: Operators at neighboring sites will tend to generate more AND-like logic (higher *l*) than non-neighboring sites (i.e. *distal* and *proximal*). In AR promoters, repression at *core* produces more AND-like logic than at *proximal*. This effect can be understood intuitively for RR promoters: If operators are closely spaced, binding of one repressor can inhibit the binding of the other. Removing one repressor has two conflicting effects: it increases expression due to its reduced occupancy, but it simultaneously decreases expression by allowing binding of the other repressor. This makes the overall logic more AND-like. In terms of the mathematical model, AND-like ($l > 0.8$) RR promoters correspond to strong balanced repression ($c_1 \approx c_2 >> 1$) and exclusive interaction ($\omega \approx 0$).

The library described here represents a starting point for systematic investigation of the functional repertoire of prokaryotic promoters. These simple promoters cannot include all of the complex effects found in natural promoters, including those dependent on DNA bending or specific protein-protein interactions. Nevertheless, they provide a view of what is possible with the simplest genetic elements and interactions. Within this context, the heuristics described above allow the design of particular promoter functions controlled by arbitrary TF regulators. The assembly method allows for construction of any specific promoter. Other promoter architectures

could be generated with this method to provide more diverse logic phenotypes, or to explore regulatory DNA in eukaryotic organisms (Ligr et al. 2006). For example, the *lac* promoter architecture, regulated by a *distal* activator and multiple repressor operators (including upstream sites), can exhibit phenotypes not found in our library, such as asym-OR (Mayo et al. 2006). In another case, a synthetic activator-activator (AA) promoter has been constructed which exhibits near-symmetric SLOPE logic (Joung et al. 1994). Tandem promoters are expected to generate additive logic functions more closely representing OR logic, and in fact many natural promoters are found in tandem repeats (Collado-Vides et al. 1991). If our heuristic rules apply to natural combinatorial promoters, we may begin to elucidate complicated functions by inspection of these non-coding DNA sequences. In this regard, effective parameterizations of logic such as the one shown in Fig. 4 can provide a more intuitive understanding of the computations performed by promoters.

Materials and Methods

Reagents

All inducers and chemicals were purchased from Sigma. Concentrations (unless otherwise stated) were 50 μg/mL kanamycin, 100 μg/mL ampicillin, 500 μM isopropyl β-D-1-thiogalactopyranoside (IPTG), 100 ng/mL anhydrotetracycline (aTc), 0.1% L(+)-arabinose (Lara), 1μM oxo-C6-homoserine lactone (VAI). LB growth media (Lennox) was used for all experiments. All ligation reactions were carried out with 1.25 units of T4 DNA ligase (Invitrogen) and 0.1 mg/mL BSA (Invitrogen) in 20μL of T4 ligase buffer (Invitrogen) at 4°C.

Randomized assembly ligation

Promoters were constructed by total synthesis and ligation (Chapter 1, Methods). Each promoter was constructed from three duplex DNA fragments comprising the *distal*, *core*, and *proximal* regions. An overhanging phosphorylated G on the downstream 5' end of *distal* is compatible with a phosphorylated overhanging C on the upstream 5' end of *core*. Likewise, an overhanging phosphorylated AA on the downstream 5' end of *core* is compatible with an overhanging phosphorylated TT on the 5' upstream end of *proximal*. The terminal ends of the fully assembled promoters had mutually incompatible XhoI and BamHI 4bp 5' overhangs, which remained unphosphorylated. A total of 48 duplex units (Supplementary Table S1) were annealed out of 96 PAGE-purified synthetic DNA oligonucleotides (University of Calgary DNA synthesis and sequencing center) at 1μM in T4 ligase buffer. All 48 duplex units were mixed together in equal 50nM proportions and ligated for one week, then cloned into bacterial luciferase reporter plasmid pCS26 (Bjarnason et al. 2003). We purified the plasmids using the Qiagen Plasmid Midi kit, and transformed the library into strain MGZ1X (reference MG1655 (Riley et al. 2006) containing the native *ara* operon, the LacI- and TetR-overexpressing Z1 cassette (Lutz et al. 1997) and the medium-copy plasmid pCD136 which constitutively expresses LuxR). We picked 10,000 clones and chose 288 randomly for sequencing (Bjarnason et al. 2003) and functional characterization.

Luminescence measurements

The library was assayed in 16 inducer conditions corresponding to all saturating combinations of the four inducers: VAI, IPTG, Lara, and aTc. Cells were grown in 96-well plates to stationary phase (16-22 hours at 37°C) and inoculated into triplicate 96-well plates containing LB media, antibiotics, and each inducer combination. These were grown at 25°C for 18 hours in the dark. Luminescence measurements were obtained using a Tecan Safire plate reader (100ms integration, default settings). To determine the background, we took the median measurements of non-functioning clones in each condition. All data reported are the median of triplicate measurements.

To assess the luminescent crosstalk between neighboring wells, we inoculated a constitutively bright clone into every other row and column of a 96-well plate (24 wells total) and measured it continuously during growth over 18 hours. This data was used to compute the horizontal/vertical (j_1) and diagonal (j_2) neighbor crosstalk. We assumed (linear) crosstalk of the form O = AX, where O is the observed data, A the actual luminescence of each well, and X the crosstalk matrix. We computed $A = OX^{-1}$ for combinations (j_1, j_2) and then took the total variance of all empty wells as a metric. This metric reached a minimum of 0.017% horizontal/vertical and 0.002% diagonal crosstalk. This was a very small effect compared to other sources of error (below), and only resulted in an appreciable difference for wells neighboring the very brightest clones ($\sim 10^6$ ALU). The vector background level (~10 ALU) was subtracted from all data points. We set each datum to a minimum level of 10, corresponding to one count/100ms.

To assess the plate-to-plate variation, we calculated the standard error between triplicates and divided by the mean. We found an average replicate error of 24%. To assess day-to-day error, we measured one set of 96 clones on two consecutive days and computed standard relative errors by a linear fit of the second day's data to the first (44%). Similarly, we computed the well-to-well error on the same plate by identifying clones with the same sequence genotype and doing a linear fit

between them (54%). Together these data provide an upper limit of ~50% on repeatability.

Promoter function analysis

To calculate the expression levels for dual-input promoters (or SIGs) we first identified the two (one) primary inducers of each promoter. We then averaged the luminescence data over the four (eight) background conditions. Standard errors were computed from these values, and the median of the triplicate measures gave the four (two) expression levels of the gate. We then computed the regulatory ratio *r,* defined as the maximum expression level divided by the minimum. The error in regulation (Table 1) was computed from the relative errors for each state. For SIGs with expression levels b_1 (off) and b_2 (on), the error in r is:

$$\Delta r = r\sqrt{\left(\frac{\Delta b_1}{b_1}\right)^2 + \left(\frac{\Delta b_2}{b_2}\right)^2}.$$

We identified SIGs and dual-input promoters from their sequences (Supplementary Data 1). Functional activator operators were found at *distal*, and functional repressor operators occurred at all three positions. With one exception (discussed in Supplementary information), significant (2×) regulation by a TF occurred only with one or more corresponding operators in the promoter sequence. The presence of an operator did not always guarantee regulation: non-functioning SIGs lie on the diagonal lines of Figs. 2 and 3, and dual-input promoters responding to only one input occur at the apex ($a \approx 1$) of the triangle in Fig. 4B.

Logic-symmetry space

In addition to the regulation r, the two-input gates displayed a variety of relative expression levels. For the dual-input promoters, we defined four measured response values (b_1, b_2, b_3, b_4) such that $b_4 \geq b_3 \geq b_2 \geq b_1$. Since repression always dominated activation, for AR promoters b_2 corresponded to the activator induced state and b_3 corresponded to the repressor induced state. Similarly, for RR promoters, b_2 corresponded to the expression level when the weaker repressor is induced and b_3 to induction of the stronger. In order to represent the range of logical functionality we defined

the three phenotypic parameters (*r, a, l*) in terms of these response values:

$$r \equiv \frac{b_4}{b_1}, a \equiv \frac{\log b_3 - \log b_2}{\log r}, l \equiv \frac{2\log b_4 - \log(b_2 b_3)}{2\log r}$$

Specifically, l quantifies the logic type ranging from a perfect AND ($b_3 = b_2 = b_1 \Rightarrow l = 1$) to a perfect OR ($b_3 = b_2 = b_4 \Rightarrow l = 0$). The parameter a quantifies the asymmetry with respect to the two inputs, ranging from perfectly symmetric ($b_2 = b_3 \Rightarrow a = 0$) to the completely asymmetric SIG ($b_3 = b_4$ and $b_2 = b_1 \Rightarrow a = 1$).

SLOPE theorem: separation of variables in combinatorial gene regulation

Consider a dual-input promoter regulated by two TFs: X and Y (we use x and y to represent their respective activities). If these TFs regulate the promoter independently with single-input functions $s(x)$ and $t(y)$, the variables of the regulation function $p(x,y)$ separate: $p(x,y) = s(x)t(y)$. Suppose (without loss of generality) that regulator X is dominant. Then the four logical output states of the promoter are:

$$b_1 = p(\downarrow\downarrow), b_2 = p(\downarrow\uparrow), b_3 = p(\uparrow\downarrow), b_4 = p(\uparrow\uparrow)$$

The arrows signify the high ($\Uparrow$) and low ($\Downarrow$) states of the promoter with respect to each input (e.g. induced and uninduced, respectively). The logic parameters of the promoter are then, by definition:

$$r \equiv \frac{b_4}{b_1} = \frac{p(\uparrow\uparrow)}{p(\downarrow\downarrow)}, r^a \equiv \frac{b_3}{b_2} = \frac{p(\uparrow\downarrow)}{p(\downarrow\uparrow)}, r^l \equiv \frac{b_4}{\sqrt{b_2 b_3}} = \frac{p(\uparrow\uparrow)}{\sqrt{p(\downarrow\uparrow)\,p(\uparrow\downarrow)}}$$

Considering the logic parameter l, the separation of variables requires that:

$$r^l = \frac{p(\uparrow\uparrow)}{\sqrt{p(\downarrow\uparrow)\,p(\uparrow\downarrow)}} = \frac{s(\uparrow)t(\uparrow)}{\sqrt{s(\downarrow)t(\uparrow)s(\uparrow)t(\downarrow)}} = \sqrt{\frac{s(\uparrow)t(\uparrow)}{s(\downarrow)t(\downarrow)}} = \sqrt{\frac{p(\uparrow\uparrow)}{p(\downarrow\downarrow)}} = r^{\frac{1}{2}} \Rightarrow l = \frac{1}{2}$$

Therefore, separation of variables—regardless of the TF regulation functions—implies that the promoter logic is always SLOPE or asym-SLOPE (or in the case that one of the regulators is nonfunctional, SIG). The converse is not generally true, but it does hold for the model of dual-repression discussed below.

Model of RR promoter logic

We employed a previously defined model of RR promoter activity under dual-repression (Bintu et al. 2005b).

$$P(R_1, R_2) = \frac{A}{1 + c_1 R_1 + c_2 R_2 + \omega c_1 R_1 c_2 R_2}$$

The maximal promoter activity is A, and the normalized repressor concentrations (R_1, R_2) range from 0 to 1. Here c_1 and c_2 represent the effectiveness of each repressor at excluding polymerase from the promoter. The term ω represents interactions between repressors: $\omega < 1$ corresponds to competitive binding, $\omega = 0$ represents exclusive binding, and $\omega > 1$ represents cooperative binding. When $\omega = 1$ the repressors are said to act independently.

We solved for the three logic-symmetry parameters (r, a, l) in terms of the three microscopic parameters (c_1, c_2, ω):

$$r = 1 + c_1 + c_2 + \omega c_1 c_2, a = \frac{1}{\log(r)} \log\left(\frac{1 + c_1}{1 + c_2}\right), l = \frac{\log((1 + c_1)(1 + c_2))}{2\log(r)}$$

By the SLOPE theorem, independent interaction ($\omega = 1$) produces SLOPE-like logic ($l = 0.5$). The converse is also true here: when $l = 0.5$, RR promoters ($c_1 \geq c_2 > 0$) are regulated by the two repressors independently ($\omega = 1$).

$$\frac{1}{2} = \frac{\log((1 + c_1)(1 + c_2))}{2\log(1 + c_1 + c_2 + \omega c_1 c_2)} \Rightarrow 1 + c_1 + c_2 + \omega c_1 c_2 = (1 + c_1)(1 + c_2) \Rightarrow \omega = 1$$

For symmetric RR promoters ($c = c_1 = c_2 \Rightarrow a = 0$), the independently interacting RR promoter is an ideal SLOPE gate ($a = 0$, $l = 0.5$). When the interaction is symmetric but dependent ($\omega \neq 1$), the logic l is described by:

$$l = \frac{\log(1 + c)}{\log(1 + 2c + c^2\omega)}$$

For exclusive interaction ($\omega = 0$), the logic depends only on the operator strength c. As c grows large, the logic approaches pure AND ($l = 1$):

$$I \approx \frac{1}{1+\frac{1}{\log_2(c)}}$$

In the opposite extreme, pure OR logic ($l = 0$) is only approached in the limit $\log_c \omega \to \infty$:

$$I \approx \frac{1}{2+\log_c(\omega)}$$

RegulonDB analysis

Following prior analysis of transcription factor binding sites (Collado-Vides et al. 1991), we examined 1,102 *E. coli* regulatory promoter sequences from RegulonDB 5.0 (Salgado et al. 2006). Operator binding sites for activators and repressors in each promoter were identified. The TF operators annotated as "dual" were removed from this list. For each operator, we determined the middle of the annotated binding sequence; calculated the distance to the annotated transcription start, and calculated the number of repressor and activator operators centered at each base pair in the region (~400bp total). These distributions were plotted as histograms for activators and repressors (Fig. 6 A-B). We also calculated the distribution of operators for 554 promoters recognized by σ^{70} (Fig. 6C). In this histogram, the relative fraction at each region was weighted by its length in bp. This weighting was necessary to observe the enrichment of repressor operator density in the *core* region.

Table 1. SIG promoters

TF	Expression				Genotype[1]				
	Uninduced (ALU)	Induced (ALU)	Regulatory Range (r)	ID	distal	-35	core	-10	proxi-mal
TetR	26 ± 8	2.3 ± 0.2×10^6	8.9 ± 0.3×10^4	D18	con0	TTGACA	tet1	GATACT	con1
	14 ± 4	1.7 ± 0.1×10^5	1.2 ± 0.4×10^4	B10	con0	TTGACA	tet1	GATACT	con3
	14 ± 3	4.7 ± 0.2×10^5	3.3 ± 0.7×10^4	B19	ara3	TTGACA	tet2	GATACT	tet1
	23 ± 7	9.9 ± 0.8×10^4	4.4 ± 1.4×10^3	A9	con0	TTGACA	tet2	TAGAGT	ara2
	15 ± 5	6.9 ± 1.1×10^4	4.7 ± 1.6×10^3	B22	con4	TTGACA	con4	TAGATT	tet1
LacI	12 ± 0	3.0±0.2×10^5	2.5±0.2×10^4	B4	lac1	TTGACA	lac2	GATACT	con0
	18 ± 3	2.7±0.3×10^4	1.5±0.2×10^4	A81	lac2	TTGACT	con1	GATACT	lux1
	86 ± 25	3.1±0.2×10^5	3.6±1.1×10^3	A38	lac1	TTGACA	con0	TATAAT	lac4
	273 ± 92	7.2±0.2×10^4	2.6±0.9×10^2	A87	con0	TTGACA	ara1	GATACT	lac4
	58 ± 17	3.6±0.3×10^4	6.2±1.9×10^2	A52	lac1	TTGACA	ara1	GATACT	con1
LuxR	130 ± 50	1.4±0.1×10^5	1.1±0.5×10^3	D49	lux1	TTTACT	con2	GATAAT	lux2
	420 ± 310	9.7±2.3×10^4	230 ± 180	D80	lux3	TTTACA	lux2	TATAAT	con3
AraC	310 ± 60	7.3±0.9×10^4	230 ± 50	A79	ara2	TAGACA	ara1	GATACT	lux2
	250 ± 30	2.8±0.6×10^4	110 ± 30	A1	ara2	TAGACT	con2	GATAAT	con1

1 The genotype refers to the three units that make up each promoter and the -10 and -35 polymerase boxes. Here "con," "tet," "lac," "ara," and "lux" refer to no operator, TetR, LacI, AraC, and LuxR operators, respectively. In each case, the number refers to the operator variant. Full sequences for each unit are available in Supplementary Table S1. Functional operators are highlighted in bold.

Figure 1.

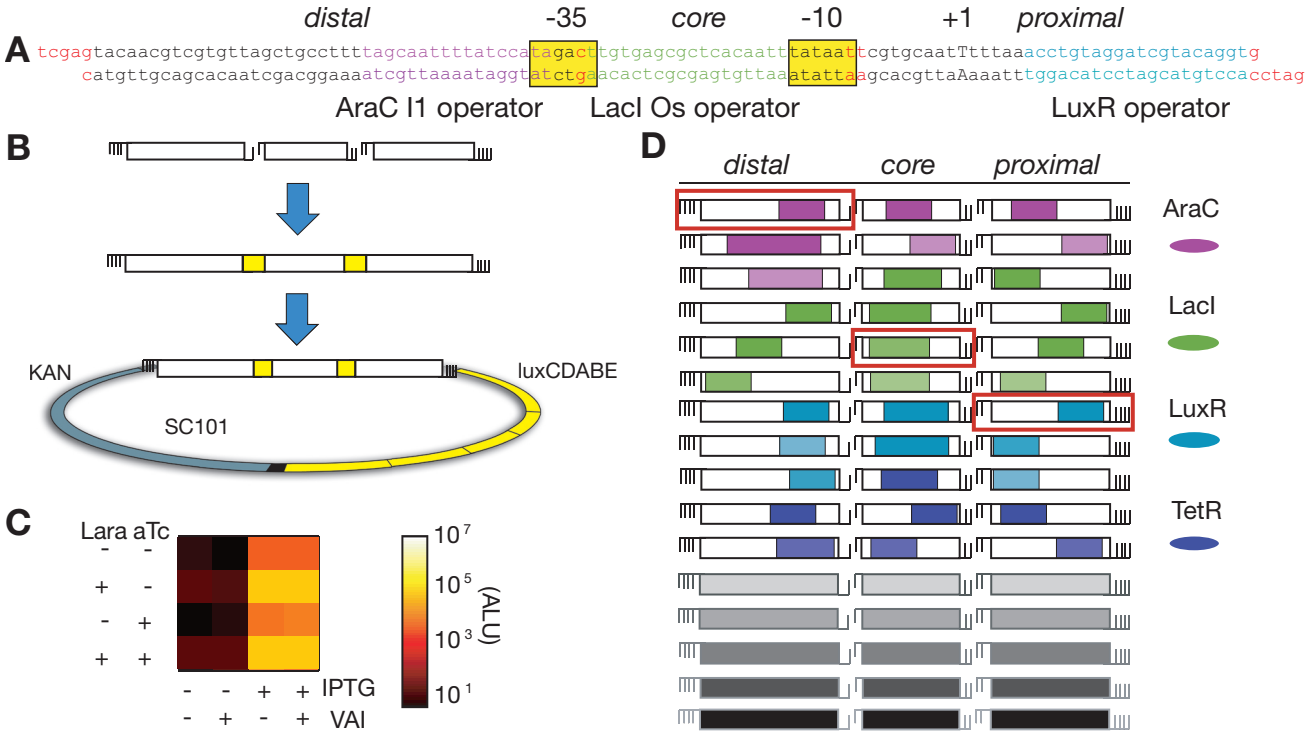


Figure 1. *Random assembly ligation generates a diverse promoter library.* Promoters can be assembled out of modular sequence units. (A) The assembled sequence of an example promoter. The 5' overhangs of each unit are shown in red. The RNA polymerase boxes (-10 and -35) are highlighted in yellow, and the predicted start site of transcription (+1) is capitalized. Operator colors are consistent throughout the figure. (B) Steps in promoter assembly and ligation into the luciferase reporter vector: Promoters are assembled by mixed ligations using 1- or 2-bp cohesive ends, and then ligated into a luciferase reporter plasmid. (C) Luminescence measurements in 16 inducer conditions (± each of 4 inducers, as indicated) for the promoter shown in part A. The output levels determine promoter logic. (D) The 48 unique units used in the library contain operators responsive to the four TFs (indicated by color) in the regions *distal*, *core*, and *proximal* (Sequences in Table S1). The promoter fragments corresponding to (A) are boxed in red.

Figure 2.

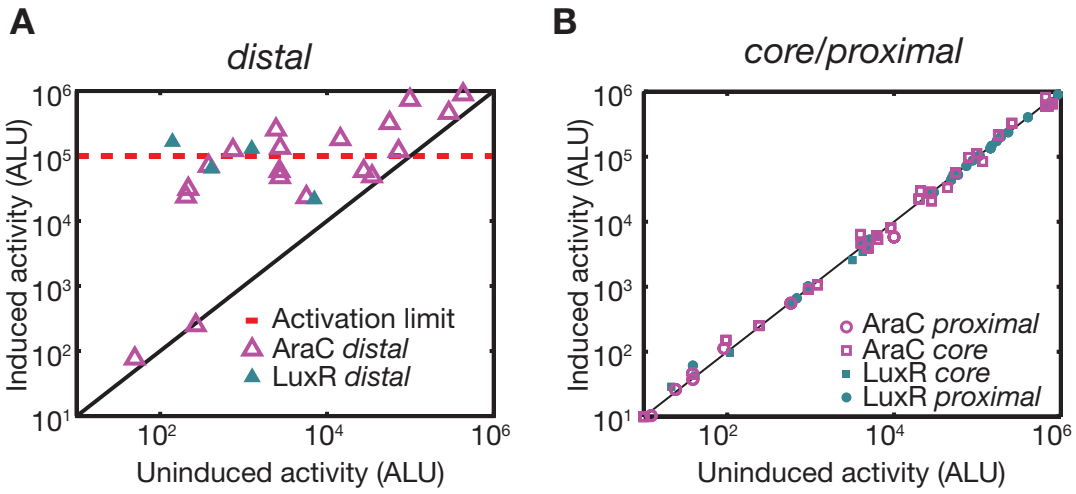


Figure 2. *Activation functions at distal, and is attenuated by intrinsic promoter strength.* (A) Measurements of promoters activated at *distal* operators. These promoters respond only to LuxR (solid triangles) or AraC (open triangles) induction. Some promoters fail to respond even though they contain a functional operator (points on the solid line). The activation ceiling (red dashed line) represents the maximal observed activation, and does not depend on the unregulated expression level. (B) Promoters containing operators at *core* or *proximal* do not respond to induction.

Figure 3.

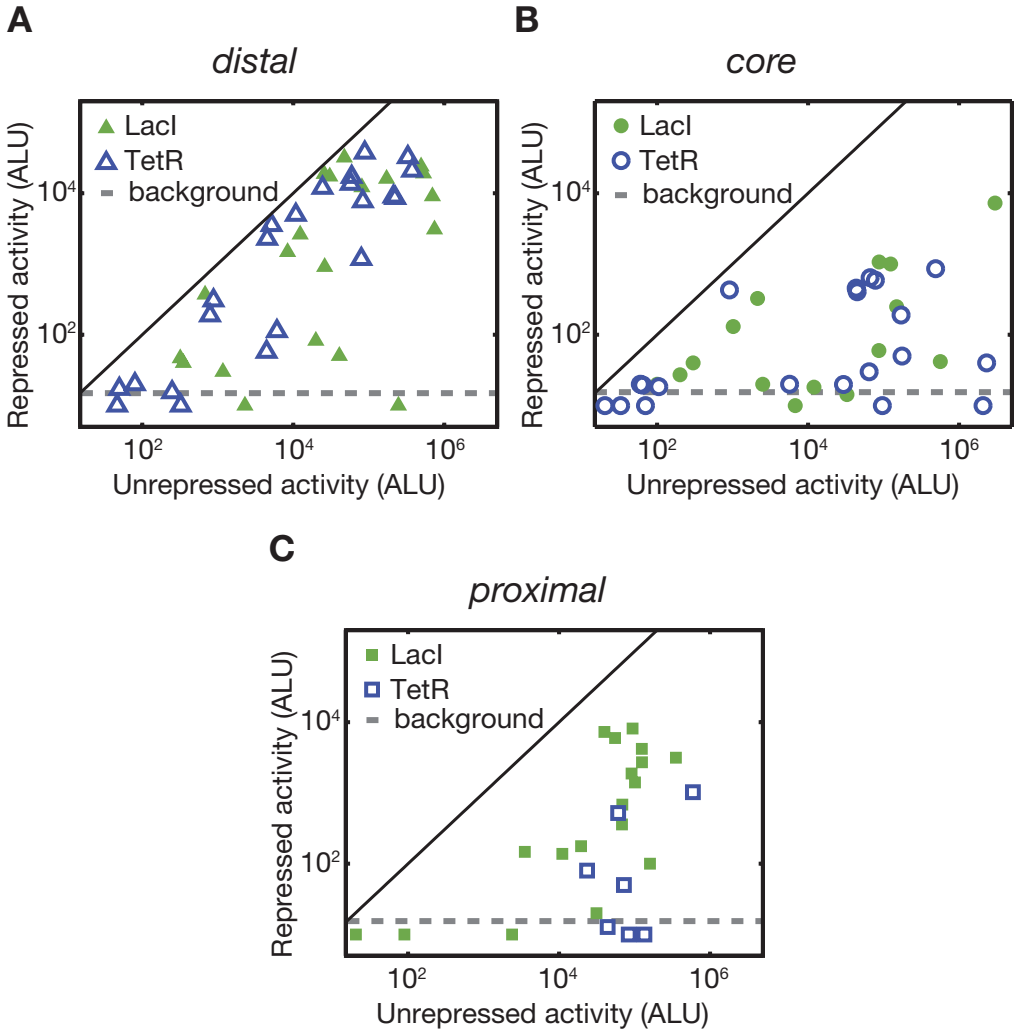


Figure 3. ***Repression is effective at all three positions, following the trend core ≥ proximal ≥ distal.*** Measurements of repressed single-input promoters. Responses are colored according to the repressor: LacI (filled) or TetR (open). Each promoter contains a single operator located at *distal* (A), *core* (B), or *proximal* (C) positions. Single-input activities are plotted in the induced (unregulated) versus uninduced (repressed) states. In some promoters, operators do not effectively repress the promoter (points located near solid black line). Luciferase detection limits are shown with grey dashed lines.

Figure 4.

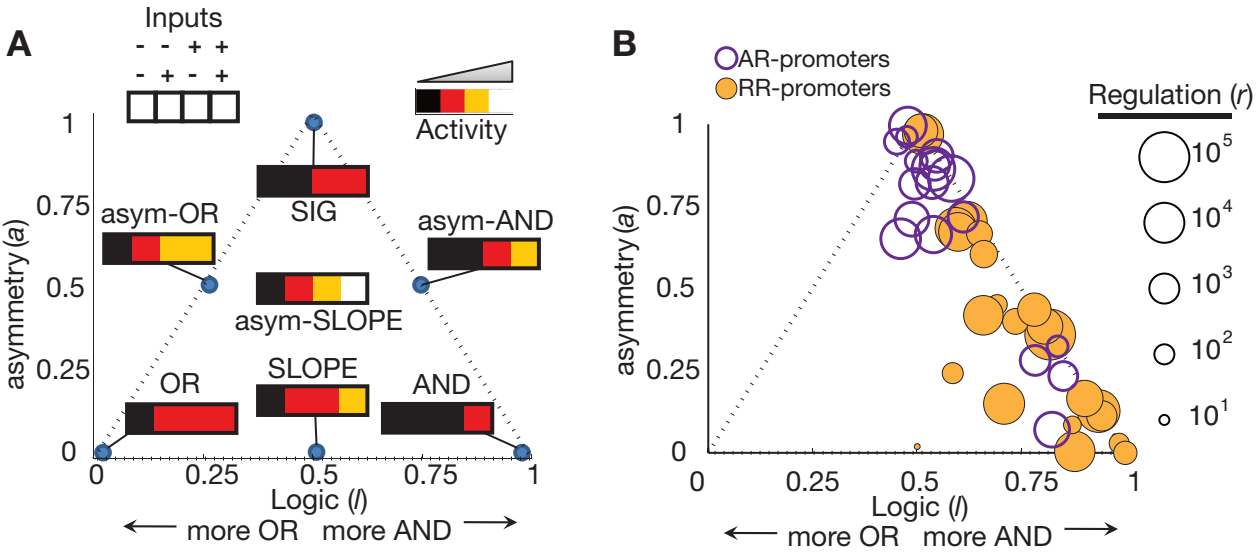


Figure 4. *Dual-input gates exhibit diverse functions in logic-symmetry space.* Promoter response phenotypes can be represented by their asymmetry, a (y-axis), logic type, l (x-axis), and regulatory range, r. (A) Diagram showing the space of allowed logical phenotypes, with the locations of ideal logic gates indicated. The SIG gate responds completely to one inducer and not at all to the other. The SLOPE gate represents an intermediate logical function between AND and OR, while the asymmetric gates represent intermediate between SIG and the corresponding symmetric gate. Intermediate logical behavior is represented between these ideal locations. The logic-symmetry parameterization is defined in the Methods. Points outside of the dashed triangle are not accessible if promoters respond monotonically to each input. (B) The logical phenotypes of 50 dual-input promoters exhibiting strong regulation ($r > 10$). AR promoters are shown as purple circles, RR promoters are shown as gold disks. The diameter of each disk is proportional to the logarithm of its regulatory range, r.

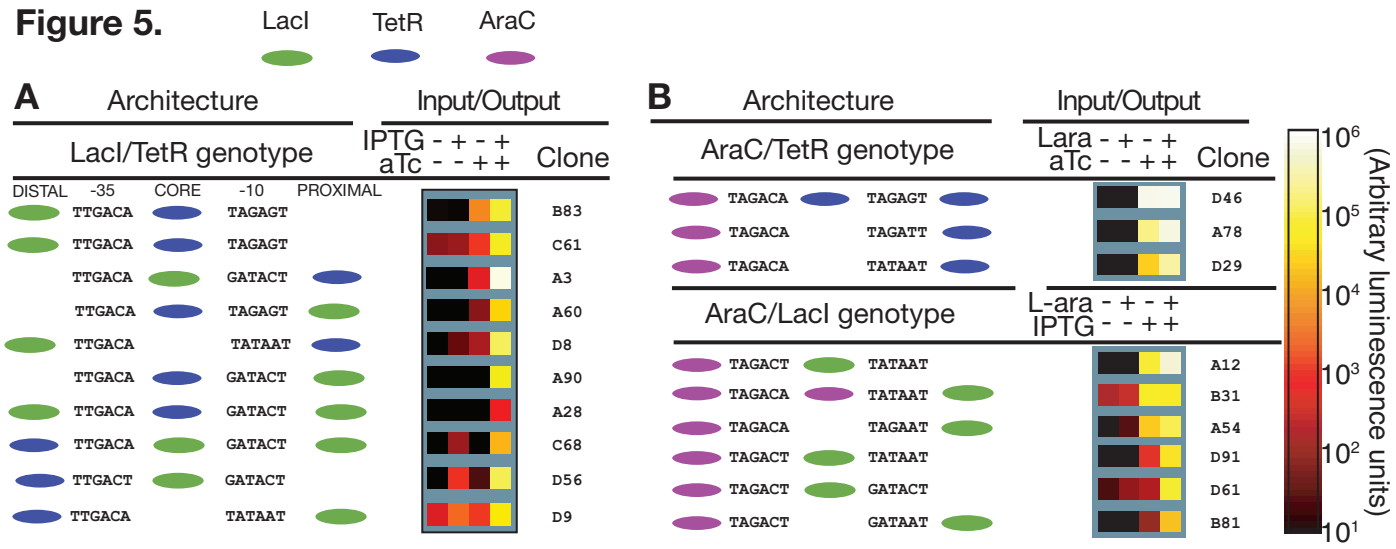


Figure 5. *Combinatorial promoter architecture reveals rules for programming gene expression.* The architecture and function of dual-input promoters. The architecture of each promoter (colored according to Fig. 1) is shown with its functional operators and -10 and -35 boxes. Promoter functions are shown as in Fig. 4A. (A) RR promoters respond to both LacI and TetR. The fourth induction column (+ IPTG, +aTc) corresponds to the unregulated state. (B) AR promoters respond to AraC and one of the two repressors, as indicated. Here, the third column (+IPTG/aTc, -Lara) corresponds to the unregulated state.

Figure 6.

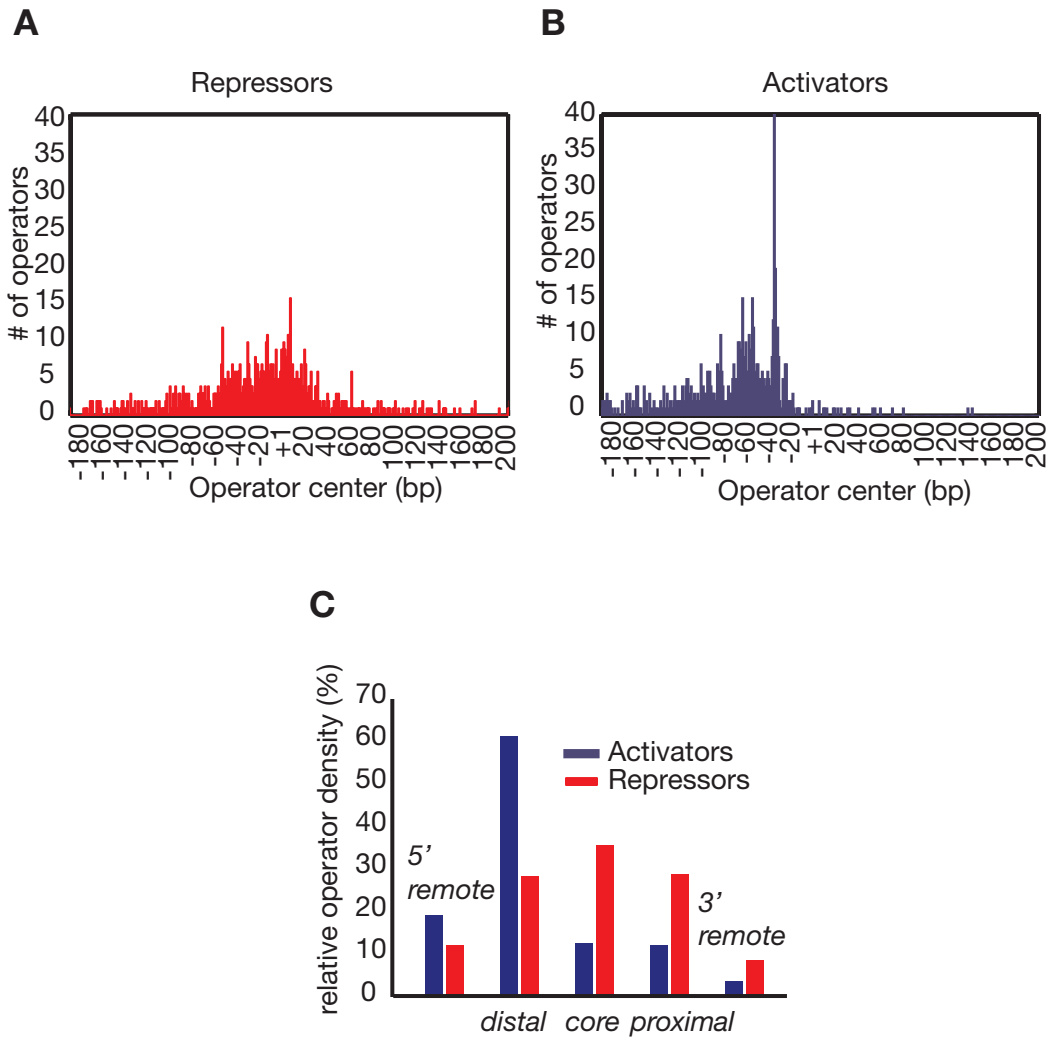


Figure 6. *The distribution of operator locations in natural promoters reflects functional trends of synthetic promoters.* Operator locations are as annotated in RegulonDB 5.0 (Salgado et al. 2006). Distributions of repressor (A) and activator (B) operators found in 1,102 *E. coli* promoters. The number of operators centered at each position relative to the start site of transcription (+1) is plotted. (C) The density of operators found in 554 σ^{70} promoters broken down into three promoter regions, *distal, core,* and *proximal*, as well as regions upstream (5' *remote*) and downstream (3' *remote*) of the promoter. The density is shown as the fraction of sites in each position weighted by the relative size (bp) of each region.

Supplementary Results

-10 and -35 polymerase box strength

The 288 promoters exhibited five decades of variation in unregulated promoter activity (Supplementary Figure S1). These sequences contained twelve -35 boxes which differed from the consensus TTGACA at up to three positions, and six -10 boxes which differed from the consensus TATAAT at up to two positions. The distributions of unregulated promoter activity for the -35 and -10 boxes were highly variable and overlapping (Supplementary Figure S2). We found that three of the twelve -35 boxes (TTGACA, TTGACT, and TAGACA) and five of the six -10 boxes (TATAAT, TAGATT, TAGAGT, GATACT, and GATAAT) produced sets of relatively strong promoters (~90% of the distributions were higher than 10^3 ALU). All of the strongest promoters in the library (~10^6 ALU) contained two of these 'strong boxes.' We used the median promoter activity of the -35 and -10 box distributions to predict the unregulated promoter activity of each promoter (Supplementary Methods). The predicted promoter activities were weakly correlated (Pearson coefficient = 0.19, Kendall $\tau = 0.32$) with the measured promoter activities, and exhibited the best agreement for the strongest promoters (Fig. S2C). Thus, strong promoters contained strong polymerase boxes; but the presence of strong polymerase boxes did not guarantee high promoter activity.

Activator operators at core and proximal

We examined the effect of activator operators at *core* and *proximal* on maximum promoter activity (fully induced). Supplementary Figure S3 shows cumulative histograms of activity for four classes of promoters: no activator operator, an activator operator at *proximal*, an activator operator at *core*, and an activator operator at *distal*. For LuxR (Fig. S3A) the presence of an operator had no effect on median promoter activity. For AraC (Fig. S3B) we found two notable effects. First, the distribution of maximal promoter activities was higher when AraC acted at *distal*. This revealed that activation increased promoter activity on average, and that the maximal

expression in the presence of the activator was uniform (near the 10^5 ALU activation ceiling). This narrow distribution of activated promoter levels is consistent with the LuxR distribution, though many fewer LuxR activated promoters were measured. Second, we found that promoters with an AraC operator only at *proximal* exhibited lower average promoter activity. Half of these promoters had a maximum activity of less than 200 ALU, and all of them exhibited activity less than 10^5 ALU. Conversely, the median strength of promoters without an AraC operator (or with an AraC operator at *core* only) was 20,000 ALU, and their maximal activity was 10^6 ALU. We note that the natural repressor activity of AraC is mediated by looping, not by steric exclusion (Hamilton and Lee, 1988), so this unexpected result is still consistent with previous work. From this analysis we infer that AraC can enact mild (10-100×) arabinose-independent repression at the *proximal* region only, and neither AraC nor LuxR can be transformed into a strong (≥10×) inducible repressor simply by moving its operator.

Spurious regulation by TetR

We found 7 promoters whose activity was induced 2-3× by aTc, without the presence of an operator for TetR. Units containing a λ cI operator (Supporting Methods) have up to 10 out of 14 conserved positions of the TetR consensus operator. Every one of the 7 spurious TetR regulated promoters contained at least one such cryptic site. These results suggest that TetR may repress weakly (3×) by binding to λ cI operators.

Dual-repressor interaction in RR promoters

We used the model of RR promoters (Methods) to analyze the relationship between logical phenotype and the repressor interaction parameter ω. Fixing *r*, we plotted lines of equal ω, varying *a* (Supplementary Figure S4). The logic parameter *l* did not depend strongly on *r*, though an increase in *r* was found to increase *l* at the extremes (near $l \sim 0$ and $l \sim 1$; e.g., compare different marker sizes in Fig. S4). We found that the logic parameter *l* did not depend strongly on *a* when $a < 0.25$. This means that logic and symmetry are 'decoupled' for near symmetric responses. As

a result, the logic parameter l depends only on ω. Asym-OR logic was possible only when r was relatively low ($r \leq 10^3$) and ω was high ($\omega \geq 100$), in agreement with the analytical results (Methods). Conversely, perfect AND logic required r to be high ($r \approx 10^5$) and ω to be low ($\omega \approx 0$).

Logic robustness to inducer concentrations

We examined the logical phenotypes of promoters with intermediate inducer concentrations. We chose three RR promoters from Fig. 5A, and measured their response to 16 combinations of inducer concentrations (Supplementary Methods). These three promoters exhibited diverse logic: AND (clone A3), asym-AND (clone D8), and SLOPE (clone D9). We found that all three promoters increased their activity monotonically with increasing concentrations of each inducer, both singly and in combination.

For 16 different combinations of inducer inputs, we calculated the logic parameters (r, a, l) corresponding to the fully induced and 8 partially induced states (Supplementary Figure S5). As expected, the parameters r and a were highly sensitive to inducer concentrations. The range r of each promoter decreased when either of the inducer concentrations was lowered. Lowering the concentration of only one inducer significantly below its threshold predictably resulted in asymmetric behavior ($a \sim 1$). Conversely, lowering the concentration of a dominant inducer could make the response more symmetric ($a \sim 0$).

The logic parameter l was less dependent on inducer concentrations, and varied differentially for the three promoters. Partial induction reduced l for the AND and asym-AND gates. The AND gate A3, with the largest l, had the highest variation in l ($l = 0.46$ to $l = 0.86$); while the SLOPE gate D9, with the smallest l, exhibited the least variation ($l = 0.48$ to $l = 0.50$). These results show that the SLOPE gate logic parameter l is extremely robust to different input concentrations, while the AND-like gates are more sensitive.

Supplementary Methods

Library fragments with λ cI operators

Each unit sequence was designed either from a consensus sequence (strong) or a sequence known to be responsive to one of five transcription factors (AraC, λ cI, LacI, TetR, LuxR), with variations in consensus signal strength, transcription factor binding site strength, spacing, and orientation (Table S1). We did not assay the response to λ cI (labeled con1-con4 for each unit, Supplementary Data 1, Supplementary Table S1, and Table 1), although 68% of the sequenced promoters contained at least one λ cI operator.

Library construction and handling

The crude randomized assembly ligation mix (Methods) was diluted 20× and combined with the bacterial luciferase reporter plasmid pCS26 (Bjarnason et al. 2003). This vector was cut with XhoI and BamHI, to match the 5' terminal overhangs on the *distal* and *proximal* ends. The vector-insert mixture was again ligated for one week, and transformed by electroporation (2.48kV, 0.2cm gap, 200uF) into Electromax DH10B cells (Invitrogen). A fraction of the recovered transformation mix was plated onto selective plates, grown overnight, and counted. These colony counts provided an estimate of 22,000 independent assembly events.

The remaining transformants were directly inoculated into LB containing antibiotics and grown for 8 hours at 37°C. Harvested cells were used to prepare liquid libraries of Midi prep DNA (Qiagen) which were re-transformed into *E. coli* K12 strain MG1655 (Blattner et al. 1997; Riley et al. 2006) containing the native *ara* operon, the LacI- and TetR-overexpressing Z1 cassette (Lutz and Bujard, 1997), and the medium-copy plasmid pCD136 which constitutively expresses LuxR).

Approximately 10,000 transformants were plated on selective media and picked into 35 384-well plates with a colony-picking robot (Norgren Systems). Each clone of the first 384-well plate was re-streaked on selective media and inoculated from a single colony into 96-well plates. 288 clones

were selected randomly for commercial sequencing (Laragen Inc., Los Angeles, CA), amplified with primers pZE05 (CCAGCTGGCAATTCCGA) and pZE06 (AATCATCACTTTCGGGAA) using the Accuprime PCR System (Invitrogen), and sequenced from the purified PCR products with primer pZE05. Sequence traces were analyzed by hand for quality (4Peaks by A. Griekspoor and Tom Groothuis, mekentosj.com).

Library measurements

Each set of 96 clones was assayed in LB Lennox media made from a single 1200mL batch. Cells were grown in 96-well plates to saturation (16-22 hours at 37°C) and inoculated into 3 replicate plates of each of 16 inducer conditions using a steel 96-pin replicator (V & P Scientific). The library was assayed in these 16 inducer conditions corresponding to all combinations of the four inducible factors: VAI (1μM), IPTG (500μM), L(+)-arabinose (0.1%), and aTc (100ng/mL). Plates were prepared by filling 96-well plates with 150μL of media and inducers on a Genetix QFill2 plate-filler (5% precision), triple-washing the apparatus to prevent inducer-carryover.

These concentrations of inducers did not significantly inhibit cell growth in the conditions used (not shown). The 48 plates were grown at 25°C without shaking for 18 hours in the dark. This growth condition minimized evaporation and sample handling time, while providing nearly uniform culture optical densities (not shown). Luciferase activity was assayed by luminescence counts using a Tecan Safire plate reader (default settings, 100ms integration time) after 30s at 30°C. Three reads of each clone were taken to assure temperature equilibration. To insure stringent control, all 16 conditions were read for one replicate before starting the next replicate.

Polymerase box strength prediction

For each -10 and -35 box in the library, we calculated the distributions of unregulated promoter activity (Fig. S2AB). We took the median of each distribution to represent the -10 and -35 box 'strength.' For each of the 288 promoters, we calculated a predicted promoter activity as the geometric mean of its -35 and -10 box strengths and plotted each against the measured unregulated

promoter activity (Fig. S2C). Alternative functions of the two box strengths (arithmetic mean, product, etc.) produced similar results.

Partial induction experiment

We measured three RR promoters (A3, D8, and D9) in sixteen inducer conditions. Each clone was grown in selective media to saturation at 37°C, and then diluted 60,000× and inoculated into a 96-well plate. Each well contained 150μL of selective media at 100, 50, 25, or 0 ng/mL aTc and 500, 50, 5, or 0 μM IPTG. We did not explore higher inducer concentrations, to avoid growth effects. This plate was grown at 25°C for 18 hours without shaking. Luminescence was measured as described above. The minimally induced case (5 μM IPTG and 25 ng/mL aTc) often produced outlying behavior, and was discarded from the phenotype-parameter analysis.

Supplementary Figures

Figure S1.

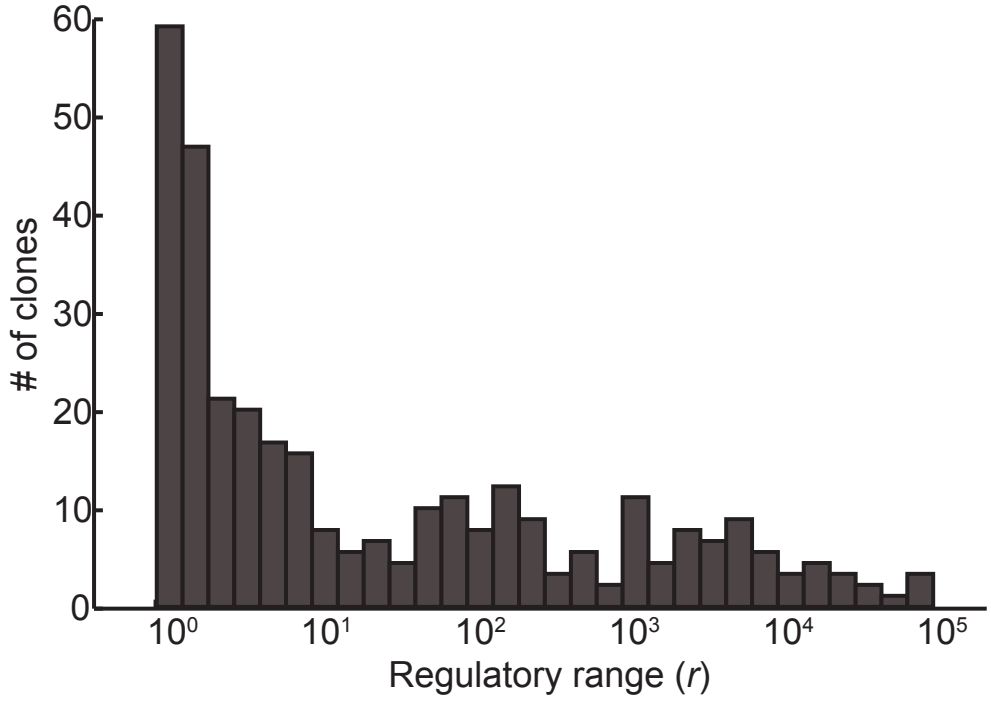


Figure S1. *The 288 characterized clones exhibit diverse regulatory ranges (r).* The characterized promoters exhibited regulation up to $r = 10^5$. Approximately half of the library promoters are regulated at least 10×.

Figure S2.

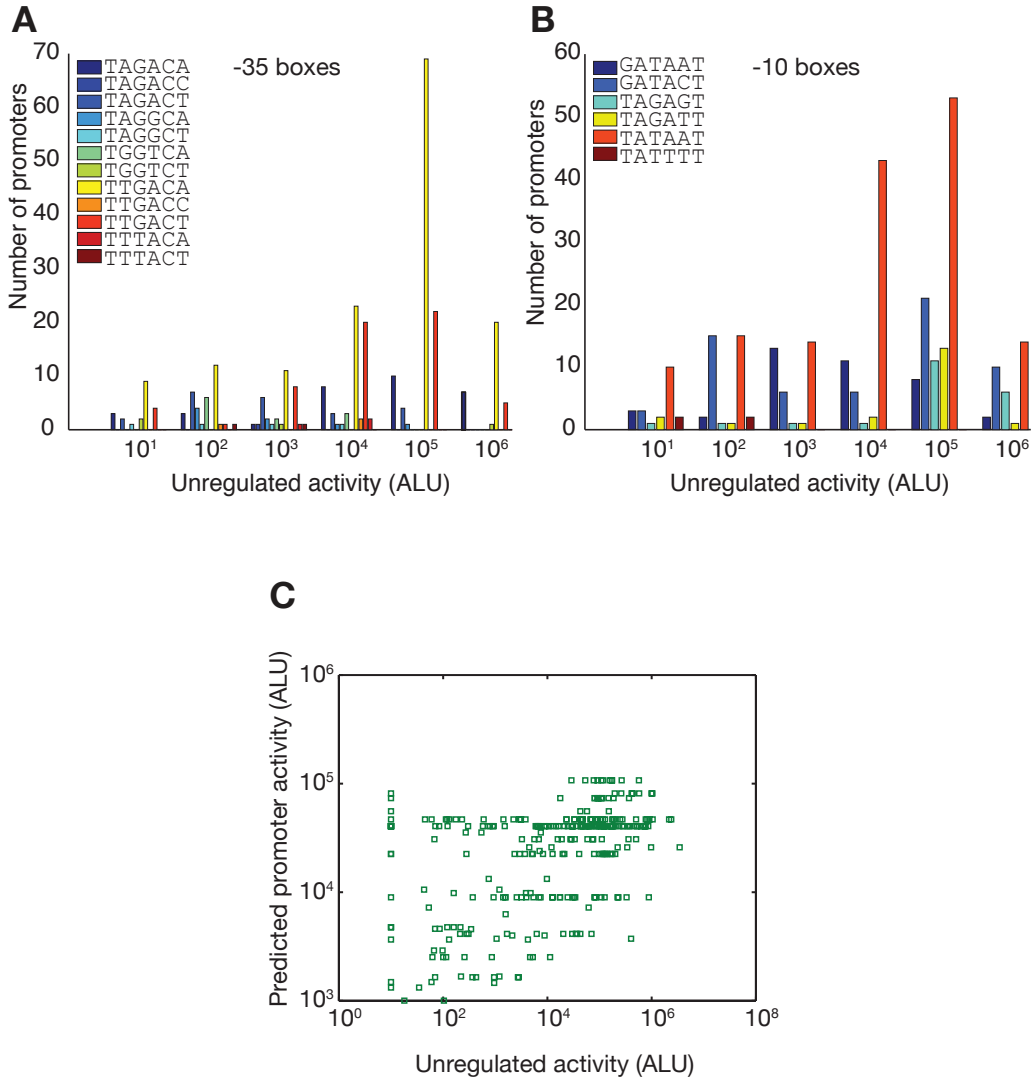


Figure S2. *Many factors contribute to promoter strength.* (A) Histograms of unregulated promoter activity for each -35 box reveal large variations in promoter strength. Three strong -35 boxes: TAGACA, TTGACT, and TTGACA (consensus) exhibit higher activities than the other nine. (B) Histograms of unregulated promoter activity for each -10 box reveal highly variable, overlapping distributions for five -10 boxes. The sixth -10 box (TATTTT) requires an activator to achieve high expression. (C) The median strength of each -35 box and -10 box distribution is used to predict the strength of each promoter. For each promoter, the geometric mean of the -10 and -35 box strengths are plotted against the unregulated activity.

Figure S3.

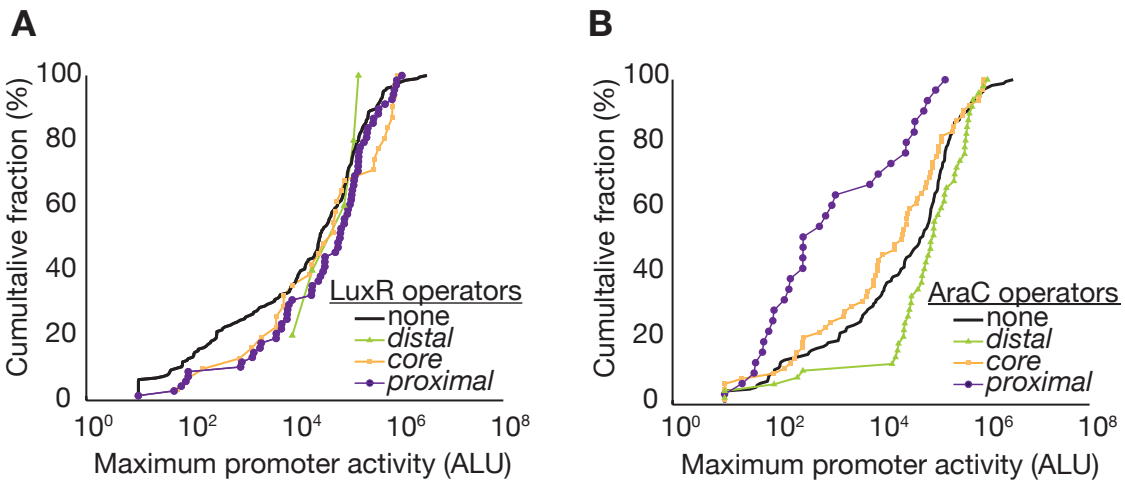


Figure S3. *Activators have small effects at core and proximal.* The cumulative histograms of maximal promoter activity for LuxR (A) and AraC (B). The maximal activity of promoters with activator operators at the *distal* position (where activation is effective) are shown for comparison.

Figure S4.

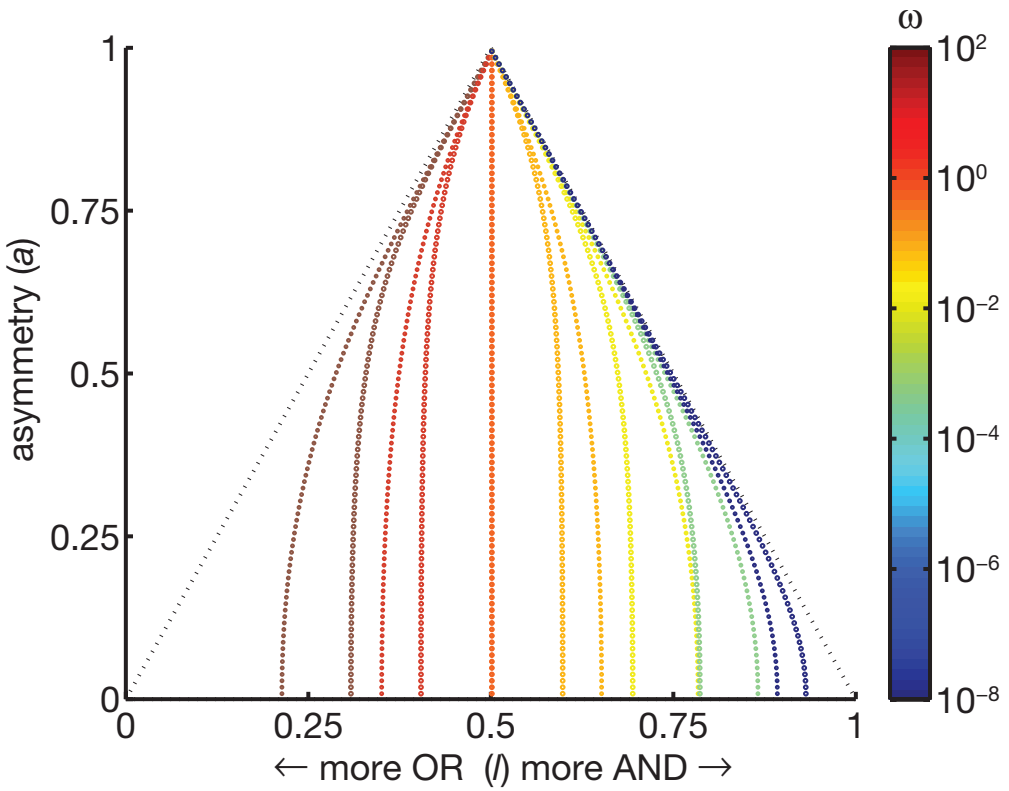


Figure S4. *Operator interactions determine logic in RR promoters.* Parametric plots of the logic parameter l as a function of the asymmetry a and repressor interaction ω. Each point is colored corresponding to ω, from ω = 100 to ω = 0, as shown on the color bar. For each value of ω, we numerically computed the logic l as a function of a for both $r = 10^3$ (smaller circles) and $r = 10^5$ (larger circles).

Figure S5.

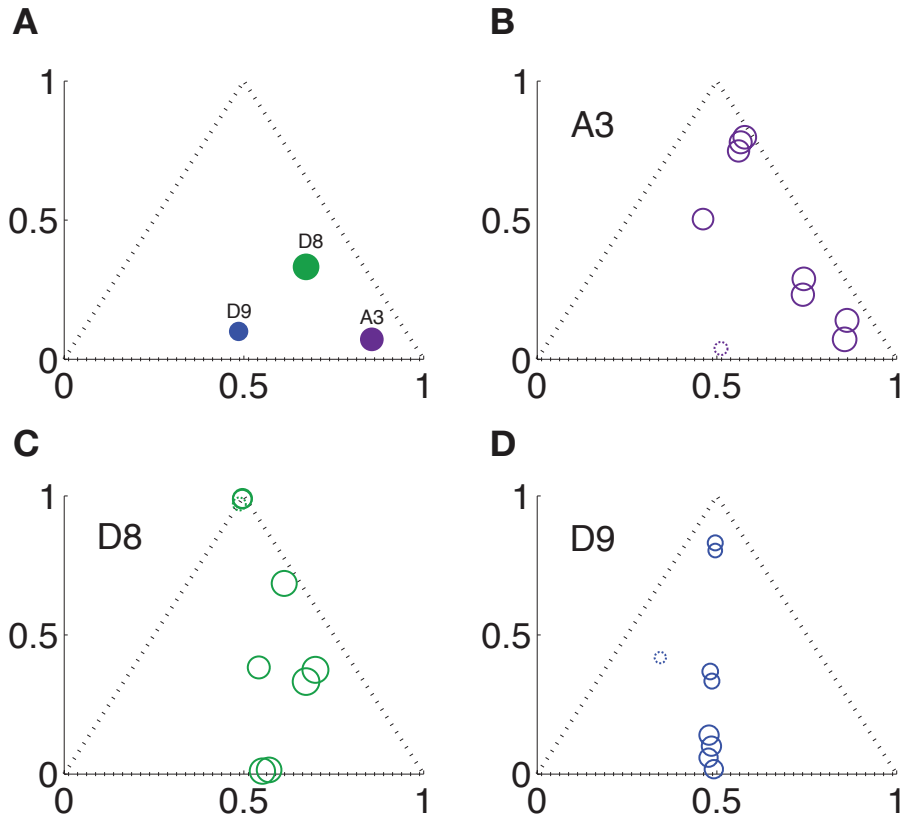


Figure S5. *RR promoters respond differentially to partial induction.* For each promoter, we measured the response in 16 different inducer conditions (Supplementary Methods). The radius of the circles is proportional to the logarithm of the regulatory range r, as in Figure 4B. The minimally induced case (5 µM IPTG and 25 ng/mL aTc) often produced outlying behavior (dashed circles). (A) The logic phenotype space coordinates of 3 RR promoters with respect to fully saturated inducer conditions. (B) The AND gate A3 exhibited differential logic when the inducer concentrations were changed. (C) The asym-AND gate D8 varied in both range r asymmetry a, and to a lesser extent, the logic parameter l. (D) The SLOPE gate D9 varied only in the range r and asymmetry a, while the logic parameter remained approximately constant ($l = 0.5$).

Table S1. The 48 synthetic units used to generate the library.

***Proximal* units**					
Name[1]	5'left	Duplex Sequence[2]	5'right	Length	Reference[3]
ara1	Tt	cgtggtccatattgcatcagacattgtacccaac	ggatc	34	*
ara2	Tt	cgtgcatagcatttttatccatacgttacccaac	ggatc	34	*
con0	Tt	cgtgcaatttttaaaattaaaggcgttacccaac	ggatc	34	(Crooks et al. 2004)
con1	Tt	gaatacatctggcggtgataaggcgttacccaac	ggatc	34	(Basu et al. 2004)
con2	Tt	gaatacctctggcggtgataaggcgttacccaac	ggatc	34	(Basu et al. 2004)
con3	Tt	cgtgcaatttttatatcaccgccaggggtacaac	ggatc	34	(Hochschild and Ptashne, 1986)
con4	Tt	cgttatcaccgccaggggtaaggcgttacccaac	ggatc	34	(Hochschild and Ptashne, 1986)
lac1	Tt	tgtggaattgtgagcggataacaatttcacacag	ggatc	34	(Lanzer and Bujard, 1988)
lac2	Tt	agattcaattgtgagcggataacaatttcacaca	ggatc	34	(Lanzer and Bujard, 1988)
lac3	Tt	gattcaattgtgagcggataacaatttcacacag	ggatc	34	(Lutz and Bujard, 1997)
lac4	Tt	cgtgcaatttaaatgtgagcgagtaacaaccaac	ggatc	34	(Becker et al. 2005)
lux1	Tt	cgtgcaatttttaaacctgtaggatcgtacaggt	ggatc	34	(Egland and Greenberg, 2000)
lux2	Tt	cttgcgacaaacaataggtaaggcgttacccaac	ggatc	34	*
lux3	Tt	cctgtaggatcgtacaggtaaggcgttacccaac	ggatc	34	*
tet1	Tt	ccacccctatcagtgatagagagcgttacccaac	ggatc	34	(Sizemore et al. 1990)
tet2[4]	Tt	aactctatcaatgaTAGAGTgtcaacaaaaaaac	ggatc	34	(Sizemore et al. 1990)
***Core* units**					
Name[1]	5'left	Duplex Sequence[2]	5'right	Length	Reference[3]
ara1[4]	C	AatcaatgTGGATTttctGATAC	Aa	23	(Hamilton and Lee, 1988)
ara2	C	AtagcggatacttcctgaTATAA	Aa	23	*
con0	C	AtttatgcttccggctcgTATAA	Aa	23	(Crooks et al. 2004)
con1	C	Ataaataccactggcggtgatac	Aa	23	(Ptashne, 2004)
con2	C	Tattttacctctggcggtgataa	Aa	23	(Ptashne, 2004)
con3	C	Ttttatcccttgcggtgatataa	Aa	23	(Michalowski et al. 2004)
con4	C	Atttatcccttgcggtgatagat	Aa	23	(Michalowski et al. 2004)
lac1	C	Attgtgagcggataacaagatac	Aa	23	(Lutz and Bujard, 1997)
lac2	C	Ttgtgagcggataacaatgatac	Aa	23	(Lanzer and Bujard, 1988)
lac3	C	Ttgtgagcggataacaattataa	Aa	23	(Lanzer and Bujard, 1988)
lac4	C	Ttgtgagcgctcacaattataa	Aa	23	(Lanzer and Bujard, 1988)
lux1	C	Cctgtaggatcgtacaggtataa	Aa	23	*
lux2[5]	C	Acctgtaggatcgtacaggtataa	Aa	24	(Egland and Greenberg, 2000)
tet1	C	Atccctatcagtgatagagatac	Aa	23	(Lutz and Bujard, 1997)
tet2	C	Aaataactctatcaatgatagag	Aa	23	(Sizemore et al. 1990)
tet3[5]	C	Actctatcattgatagagttattt	Aa	24	(Sizemore et al. 1990)

1 The labels "tet", "lac", "ara", and "lux" refer to TetR, LacI, AraC, and LuxR operators, respectively. The units named con1–con4 contain λ cI operators, and the units named con0 contain the consensus sequence with no operators.

2 The -10 and -35 boxes are capitalized. Approximate binding site locations are colored (TetR: blue, LuxR: cyan, LacI: green, AraC: magenta, cI: brown). 5' overhangs are shown for both the left and right sides of the duplex, cloning sites are highlighted in red.

3 A (*) refers to units designed for this study.

4 This unit contains an internal -10 box, capitalized.

5 This core unit has 1 bp extra space between -10 and -35 boxes.

***Distal* units**					
Name[1]	**5'left**	**Duplex Sequence[2]**	**5'right**	**Length**	**Reference[3]**
ara1	tcgag	aacatagcatttttatccataagattagcggatctaaccTTTA	G	43	(Lutz and Bujard, 1997)
ara2	tcgag	tacaacgtcgtgttagctgcctttagcaattttatccaTAGA	G	43	(Zhang et al. 1996)
ara3[6]	tcgag	gtaacaaaagtgtctataatcacggcagaaaagtccacaTTGA	G	43	(Hamilton and Lee, 1988)
con0	tcgag	tacaacgtcgtgttagctgcctttcgtcttcaataattcTTGA	G	43	(Crooks et al. 2004)
con1	tcgag	cagataaccatctgcggtgataaattatctctggcggtgTTGA	G	43	(Lanzer and Bujard, 1988)
con2	tcgag	tatcaccgccagaggtaaaatagtcaacacgcacggtgtTAGG	G	43	(Ptashne, 2004)
con3	tcgag	tatcaccgccagaggtaaaatagtcaacacgcacggtgtTAGA	G	43	(Ptashne, 2004)
con4	tcgag	tacaacgtcgtgttagctgtatcaccgccagaggtaagaTTGA	G	43	(Hochschild and Ptashne, 1986)
lac1	tcgag	tacaacgtcgtgttagctgcaattgtgagcggataacaaTTGA	G	43	(Lutz and Bujard, 1997)
lac2	tcgag	tacaacgtcgtgttaaattgtgagcggataacaatttagTTGA	G	43	(Lanzer and Bujard, 1988)
lac3	tcgag	tacaattgtgagcgctcacaatttcgtcttcaataattcTTGA	G	43	(Becker et al. 2005)
lux1	tcgag	tacaattgtttaacataagtacctgtaggatcgtacaggTTTA	G	43	(Egland and Greenberg, 1999)
lux2	tcgag	tacaattgtttaacataagtgaatggatcattttgcaggTTTA	G	43	(Shadel and Baldwin, 1992)
lux3[6]	tcgag	acatagcatttttatccataacctgtaggatcgtacaggTTTA	G	43	*
tet1	tcgag	tacaacgtcgtgttagctgctccctatcagtgatagagaTTGA	G	43	(Lutz and Bujard, 1997)
tet2[7]	tcgag	tacaacgtCatttcacttTTCTCTatcactgatagggagTGGT	G	43	(Sizemore et al. 1990)

6 This unit contains a non-functional AraC site.
7 This unit contains an internal -35 box, capitalized.

References

Atkinson MR, Savageau MA, Myers JT, Ninfa AJ (2003) Development of genetic circuitry exhibiting toggle switch or oscillatory behavior in *Escherichia coli*. *Cell* **113:** 597-607.

Atsumi S, Little JW (2006) A synthetic phage lambda regulatory circuit. *Proc Natl Acad Sci U S A* **103:** 19045-19050.

Basu S, Gerchman Y, Collins CH, Arnold FH, Weiss R (2005) A synthetic multicellular system for programmed pattern formation. *Nature* **434:** 1130-1134.

Basu S, Mehreja R, Thiberge S, Chen MT, Weiss R (2004) Spatiotemporal control of gene expression with pulse-generating networks. *Proc Natl Acad Sci U S A* **101:** 6355-6360.

Beck CF, Mutzel R, Barbe J, Muller W (1982) A multifunctional gene (tetR) controls Tn10-encoded tetracycline resistance. *J Bacteriol* **150:** 633-642.

Becker NA, Kahn JD, Maher LJ, 3rd (2005) Bacterial repression loops require enhanced DNA flexibility. *Journal of molecular biology* **349:** 716-730.

Bintu L, Buchler NE, Garcia HG, Gerland U, Hwa T, Kondev J, Kuhlman T, Phillips R (2005a) Transcriptional regulation by the numbers: applications. *Curr Opin Genet Dev* **15:** 125-135.

Bintu L, Buchler NE, Garcia HG, Gerland U, Hwa T, Kondev J, Phillips R (2005b) Transcriptional regulation by the numbers: models. *Curr Opin Genet Dev* **15:** 116-124.

Bjarnason J, Southward CM, Surette MG (2003) Genomic profiling of iron-responsive genes in *Salmonella enterica serovar typhimurium* by high-throughput screening of a random promoter library. *J Bacteriol* **185:** 4973-4982.

Browning DF, Busby SJ (2004) The regulation of bacterial transcription initiation. *Nat Rev Microbiol* **2:** 57-65.

Buchler N, Gerland U, Hwa T (2003) On schemes of combinatorial transcription logic. *PNAS* **100:** 5136-5141.

Busby S, Ebright RH (1994) Promoter structure, promoter recognition, and transcription activation in prokaryotes. *Cell* **79:** 743-746.

Chan B, Busby S (1989) Recognition of nucleotide sequences at the *Escherichia coli* galactose operon P1 promoter by RNA polymerase. *Gene* **84:** 227-236.

Collado-Vides J, Magasanik B, Gralla JD (1991) Control site location and transcriptional regulation in *Escherichia coli*. *Microbiol Mol Biol Rev* **55:** 371-394.

Egland KA, Greenberg EP (1999) Quorum sensing in *Vibrio fischeri*: elements of the luxl promoter. *Mol Microbiol* **31:** 1197-1204.

Elledge SJ, Davis RW (1989) Position and density effects on repression by stationary and mobile DNA-binding proteins. *Genes & development* **3:** 185-197.

Endy D (2005) Foundations for engineering biology. *Nature* **438:** 449-453.

Fuqua WC, Winans SC, Greenberg EP (1994) Quorum sensing in bacteria: the LuxR-LuxI family of cell density-responsive transcriptional regulators. *J Bacteriol* **176:** 269-275.

Gralla JD, Collado-Vides J (1996) Organization and Function of Transcription Regulatory Elements. In *Escherichia coli and Salmonella : cellular and molecular biology*, Neidhardt FC, Curtiss R (eds), 2nd edn, pp 2 v. (xx, 2822 p.). Washington, D.C.: ASM Press.

Gross CA, Chan C, Dombroski A, Gruber T, Sharp M, Tupy J, Young B (1998) The functional and regulatory roles of sigma factors in transcription. *Cold Spring Harbor symposia on quantitative biology* **63:** 141-155.

Guido NJ, Wang X, Adalsteinsson D, McMillen D, Hasty J, Cantor CR, Elston TC, Collins JJ (2006) A bottom-up approach to gene regulation. *Nature* **439:** 856-860.

Hasty J, McMillen D, Collins JJ (2002) Engineered gene circuits. *Nature* **420:** 224-230.

Haugen SP, Berkmen MB, Ross W, Gaal T, Ward C, Gourse RL (2006) rRNA promoter regulation by nonoptimal binding of sigma region 1.2: an additional recognition element for RNA polymerase. *Cell* **125:** 1069-1082.

Hawley DK, McClure WR (1983) Compilation and analysis of *Escherichia coli* promoter DNA sequences. *Nucleic acids research* **11:** 2237-2255.

Hermsen R, Tans S, Wolde PR (2006) Transcriptional Regulation by Competing Transcription Factor Modules. *PLoS Comput Biol* **2:** e164.

Jacob F, Monod J (1961) Genetic regulatory mechanisms in the synthesis of proteins. *Journal of molecular biology* **3:** 318-356.

Joung JK, Koepp DM, Hochschild A (1994) Synergistic activation of transcription by bacteriophage lambda cI protein and E. coli cAMP receptor protein. *Science* **265:** 1863-1866.

Kammerer W, Deuschle U, Gentz R, Bujard H (1986) Functional dissection of *Escherichia coli* promoters: information in the transcribed region is involved in late steps of the overall process. *The EMBO journal* **5:** 2995-3000.

Kauffman S (1969) Homeostasis and Differentiation in Random Genetic Control Networks. *Nature* **224:** 177-178.

Lanzer M, Bujard H (1988) Promoters largely determine the efficiency of repressor action. *Proc Natl Acad Sci U S A* **85:** 8973-8977.

Ligr M, Siddharthan R, Cross FR, Siggia ED (2006) Gene expression from random libraries of yeast promoters. *Genetics* **172:** 2113-2122.

Lutz R, Bujard H (1997) Independent and tight regulation of transcriptional units in *Escherichia coli* via the LacR/O, the TetR/O and AraC/I1-I2 regulatory elements. *Nucleic acids research* **25:** 1203-1210.

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

Mayo AE, Setty Y, Shavit S, Zaslaver A, Alon U (2006) Plasticity of the cis-regulatory input function of a gene. *PLoS Biol* **4:** e45.

Oehler S, Amouyal M, Kolkhof P, von Wilcken-Bergmann B, Muller-Hill B (1994) Quality and

position of the three lac operators of *E. coli* define efficiency of repression. *The EMBO journal* **13:** 3348-3355.

Ogden S, Haggerty D, Stoner CM, Kolodrubetz D, Schleif R (1980) The *Escherichia coli* L-arabinose operon: binding sites of the regulatory proteins and a mechanism of positive and negative regulation. *Proc Natl Acad Sci U S A* **77:** 3346-3350.

Ptashne M (2004) *A genetic switch : phage lambda revisited*, 3rd edn. Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory Press.

Ptashne M (2005) Regulation of transcription: from lambda to eukaryotes. *Trends in biochemical sciences* **30:** 275-279.

Riley M, Abe T, Arnaud MB, Berlyn MK, Blattner FR, Chaudhuri RR, Glasner JD, Horiuchi T, Keseler IM, Kosuge T, Mori H, Perna NT, Plunkett G, 3rd, Rudd KE, Serres MH, Thomas GH, Thomson NR, Wishart D, Wanner BL (2006) *Escherichia coli* K-12: a cooperatively developed annotation snapshot--2005. *Nucleic acids research* **34:** 1-9.

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

Ross W, Gosink KK, Salomon J, Igarashi K, Zou C, Ishihama A, Severinov K, Gourse RL (1993) A third recognition element in bacterial promoters: DNA binding by the alpha subunit of RNA polymerase. *Science* **262:** 1407-1413.

Salgado H, Gama-Castro S, Peralta-Gil M, Diaz-Peredo E, Sanchez-Solano F, Santos-Zavaleta A, Martinez-Flores I, Jimenez-Jacinto V, Bonavides-Martinez C, Segura-Salazar J, Martinez-Antonio A, Collado-Vides J (2006) RegulonDB (version 5.0): *Escherichia coli* K-12 transcriptional regulatory network, operon organization, and growth conditions. *Nucleic Acids Res* **34:** D394-397.

Schleif R (2003) AraC protein: a love-hate relationship. *Bioessays* **25:** 274-282.

Setty Y, Mayo AE, Surette MG, Alon U (2003) Detailed map of a cis-regulatory input function. *Proc Natl Acad Sci U S A* **100:** 7702-7707.

Skerra A (1994) Use of the tetracycline promoter for the tightly regulated production of a murine

antibody fragment in *Escherichia coli*. *Gene* **151:** 131-135.

Sprinzak D, Elowitz MB (2005) Reconstruction of genetic circuits. *Nature* **438:** 443-448.

Thomas R, D'Ari R (1990) *Biological feedback*. Boca Raton: CRC Press.

Vilar JM, Leibler S (2003) DNA looping and physical constraints on transcription regulation. *Journal of molecular biology* **331:** 981-989.

Zubay G, Schwartz D, Beckwith J (1970) Mechanism of activation of catabolite-sensitive genes: a positive control system. *Proc Natl Acad Sci U S A* **66:** 104-110.

Supplementary References

Basu S, Mehreja R, Thiberge S, Chen MT, Weiss R (2004) Spatiotemporal control of gene expression with pulse-generating networks. *Proc Natl Acad Sci U S A* 101: 6355-6360.

Becker NA, Kahn JD, Maher LJ, 3rd (2005) Bacterial repression loops require enhanced DNA flexibility. *Journal of molecular biology* 349: 716-730.

Bjarnason J, Southward CM, Surette MG (2003) Genomic profiling of iron-responsive genes in *Salmonella enterica serovar typhimurium* by high-throughput screening of a random promoter library. *J Bacteriol* 185: 4973-4982.

Blattner FR, Plunkett G, 3rd, Bloch CA, Perna NT, Burland V, Riley M, Collado-Vides J, Glasner JD, Rode CK, Mayhew GF, Gregor J, Davis NW, Kirkpatrick HA, Goeden MA, Rose DJ, Mau B, Shao Y (1997) The complete genome sequence of *Escherichia coli* K-12. *Science* 277: 1453-1474.

Crooks GE, Hon G, Chandonia JM, Brenner SE (2004) WebLogo: a sequence logo generator. *Genome research* 14: 1188-1190.

Egland KA, Greenberg EP (1999) Quorum sensing in *Vibrio fischeri*: elements of the luxl promoter. *Mol Microbiol* 31: 1197-1204.

Egland KA, Greenberg EP (2000) Conversion of the *Vibrio fischeri* transcriptional activator, LuxR, to a repressor. *J Bacteriol* 182: 805-811.

Hamilton EP, Lee N (1988) Three binding sites for AraC protein are required for autoregulation of araC in *Escherichia coli. Proc Natl Acad Sci U S A* 85: 1749-1753.

Hochschild A, Ptashne M (1986) Cooperative binding of lambda repressors to sites separated by integral turns of the DNA helix. *Cell* 44: 681-687.

Lanzer M, Bujard H (1988) Promoters largely determine the efficiency of repressor action. *Proc Natl Acad Sci U S A* 85: 8973-8977.

Lutz R, Bujard H (1997) Independent and tight regulation of transcriptional units in *Escherichia coli* via the LacR/O, the TetR/O and AraC/I1-I2 regulatory elements. *Nucleic acids research* 25:

1203-1210.

Michalowski CB, Short MD, Little JW (2004) Sequence tolerance of the phage lambda PRM promoter: implications for evolution of gene regulatory circuitry. *J Bacteriol* 186: 7988-7999.

Ptashne M (2004) *A genetic switch : phage lambda revisited*, 3rd edn. Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory Press.

Riley M, Abe T, Arnaud MB, Berlyn MK, Blattner FR, Chaudhuri RR, Glasner JD, Horiuchi T, Keseler IM, Kosuge T, Mori H, Perna NT, Plunkett G, 3rd, Rudd KE, Serres MH, Thomas GH, Thomson NR, Wishart D, Wanner BL (2006) *Escherichia coli* K-12: a cooperatively developed annotation snapshot--2005. *Nucleic acids research* 34: 1-9.

Shadel GS, Baldwin TO (1992) Identification of a distantly located regulatory element in the luxD gene required for negative autoregulation of the *Vibrio fischeri* luxR gene. *The Journal of biological chemistry* 267: 7690-7695.

Sizemore C, Wissmann A, Gulland U, Hillen W (1990) Quantitative analysis of Tn10 Tet repressor binding to a complete set of tet operator mutants. *Nucleic acids research* 18: 2875-2880.

Zhang X, Reeder T, Schleif R (1996) Transcription activation parameters at ara pBAD. *Journal of molecular biology* 258: 14-24.

A Synthetic Three-Color Reporter Scaffold for Monitoring Genetic Regulation and Noise

Robert Sidney Cox III, Mary J. Dunlop, and Michael B. Elowitz

Abstract

Biologists require accurate, distinguishable, non-toxic reporters for multiple genes in the same organism. Despite recent improvements in fluorescent proteins, there does not exist a single vector with which one can conveniently employ multiple reporters. Therefore, we designed and built such a system using total DNA synthesis. This scaffold will be useful for analyzing natural genetic circuits—as well as assembling synthetic circuits—in many organisms. Here we characterize the scaffold in *Escherichia coli.* Three spectrally distinct reporters allow independent monitoring of genetic signals and analysis of genetic noise. As an application, we show that the scaffold is a sensitive detector of transcriptional co-regulation.

Introduction

Cells contain genetic circuits, composed of interacting genes and proteins, which control cellular functions. Although these circuits are traditionally studied on average in large populations, they actually operate in individual living cells. As a result, they are subject to substantial variation, both from stochastic effects in circuit components (intrinsic noise), and from the substantial cell-cell variability that exists in all cellular components (extrinsic noise). A critical problem in understanding such circuits is determining how both types of noise, together with circuit structure, determine the dynamics of gene expression and thereby affect cellular behavior. A complementary question is what fluctuations in specific genes, and the correlations between them, might tell us about circuit connectivity.

Recently the interaction between noise and circuit structure has been approached from two new directions: First, several studies have followed the dynamics of endogenous circuit components in individual cells using one or more fluorescent reporter proteins (Elowitz and Leibler, 2000; Rosenfeld et al. 2005; Suel et al. 2006). These dynamics can be interpreted in terms of circuit structure, cell-cell variability, and noise. Second, researchers have begun constructing simple synthetic genetic circuits composed of well-characterized genes and proteins. These circuits are designed to implement particular functions, such as oscillation or memory (bistability) (Elowitz and Leibler, 2000; Gardner et al. 2000). Like their natural counterparts, these synthetic circuits—and the noise propagating through them (Pedraza and van Oudenaarden, 2005)—can be monitored at the single cell level using fluorescent reporter genes.

Synthetic biology involves the assembly of regulatory circuits from genetic components. Typically, circuit design is based on qualitative models of the individual components, because accurate quantitative descriptions of genetic components and their *in vivo* interactions are lacking. In many cases, the behavior of designed genetic circuits differs significantly from predictions. In order to

construct predictable circuits, it is important to isolate gene expression: the expression of one gene should not inadvertently affect the expression of another, except in ways determined by the intended circuit diagram. Decoupling the circuit components, such as by placing them on separate plasmids, can facilitate the 'debugging' of synthetic circuit function. For example, to measure the gene-regulation function of a transcription factor (Rosenfeld et al. 2005), the factor was placed on a plasmid and the target promoter in the chromosome. Independent single-cell measurements of multiple reporters will give synthetic biologists a sensitive measure of circuit function.

Experiments with both natural and synthetic circuits suffer from a lack of systematic control over reporter genes. For example, synthetic circuits are often encoded on plasmids whose copy numbers vary significantly in unpredictable ways. Similarly, multiple fluorescent reporters of natural circuits have not generally been optimized for high signal and low crosstalk. Both types of experiments will benefit from a well-characterized platform for expressing reporter genes and synthetic circuit components in a reproducible, non-toxic fashion. Many technology applications, such as metabolic engineering (Farmer and Liao, 2000; Khosla and Keasling, 2003), would benefit from continuous and independent control of multiple operons with quantitative outputs. For these reasons, one would like a general chassis into which one could insert promoters from natural genes or components of synthetic circuits, with optimized fluorescent protein reporters to independently monitor the corresponding genes.

Here we describe such an integrated platform. This system exploits our knowledge of genetic regulation and is an ideal framework for both the measurement of gene expression and the construction of synthetic networks. To create the system we exploited recent developments in total DNA synthesis to design a highly optimized sequence. We show how the scaffold can be used to analyze fluctuations of a transcription factor regulating two target promoters. In this example, we find that the co-regulation of the two promoters can be inferred from the simultaneous analysis of fluctuations of three fluorescent proteins. These results show that noise is not just an unavoidable fact of life in the single cell environment, but can be exploited to infer properties of genetic elements.

Results and Discussion

We designed a three-color fluorescent reporter scaffold (Fig. 1) to fulfill several criteria: (1) *Biocompatability:* The scaffold was genetically stable and non-toxic to cells carrying it (not shown). This was accomplished by keeping the scaffold small (4kb), using a low copy plasmid origin of replication (SC101), moderate strength ribosome binding sites (RBSs), optimizing the fluorescent proteins to remove 'toxic' codons, and placing them under the control of tightly regulated promoters (Lutz and Bujard, 1997). To discourage mutation, we explicitly avoided homologous or repeated sequences. (2) *Distinctness:* We chose fast maturing, monomeric proteins with minimally overlapping spectra to maximize linearity of response (Campbell et al. 2002; Nagai et al. 2002; Rizzo et al. 2004; Shu et al. 2006). We determined the spectral crosstalk in our microscopy set-up to be extremely low (Methods), and correctable. (3) *Sensitivity* and *independence:* We wished to detect both strong and weak genetic signals simultaneously, with the ability to watch them change over time in single living cells (Rosenfeld et al. 2005). We used multiple genetic terminator sequences (Brendel et al. 1986; Wilson and von Hippel, 1995) along with empty 'spacer' regions to insure that changing the expression level of one protein did not affect the level of another, except when genetically co-regulated (see below). (4) *Tunability:* The scaffold was designed to allow for easy tuning of reporter parameters: promoter strengths, RBSs, and degradation tags, as well as the construction of fluorescent fusions. (5) *Modularity, portability,* and *extensibility:* Restriction sites were strategically placed to allow genetic elements to be easily swapped, inserted, or deleted. The fluorescent proteins were codon-optimized for both gram-positive and gram-negative bacteria. The system can be moved between different plasmids (Lutz and Bujard, 1997), chromosomes, and organisms. Additional restriction sites were included to add the option of a fourth reporter operon (Ai et al. 2007).

We characterized the scaffold in single *E. coli* cells using quantitative fluorescence microscopy (Fig. 2). In our strain, all three fluorescent proteins were repressed. Cells carrying the plasmid

showed very weak (5%), but detectable, expression of the *yfp* and *cfp* genes compared to the mean cellular autofluorescence (Fig. 2A and B). We tested whether the transcriptional units in the scaffold plasmid could be induced independently, by using combinations of saturating inducer concentrations (Methods). We found the mean expression of *cfp* and *rfp* were independent, indicating that there was no significant transcriptional read-through from *rfp* into *cfp* (Fig. 2F and H). Similarly, the mean expression of *cfp* and *yfp* was independent (Fig. 2B through E). These results showed that the design of the scaffold provided sufficient genetic isolation for independent control of the three reporter genes.

Combinatorial promoters that accept multiple genetic inputs are ubiquitous in genomes and useful for creating synthetic circuits. The combinatorial LacI/AraC regulated promoter (Lutz and Bujard, 1997) controlling *rfp* behaved as an asym-AND (as defined in Chapter 2) gate (Fig. 2B, D, F, and H), with the repressor LacI acting as the dominant transcription factor. Expression was undetectable when induced with Lara alone (Fig. 2F). In the absence of Lara, IPTG induced expression only slightly above the autofluorescence level (Fig. 2D-E). In the presence of Lara, IPTG induced expression up to 40× the autofluorescent background level (Fig. 2H-I). These results show that the LacI/AraC promoter controls expression combinatorially, with three distinct output expression levels.

Fluctuations in plasmid copy number, overall transcriptional or translational activity, or growth rate, would be expected to enhance or reduce expression of all genes simultaneously. For each color—*cfp, yfp,* and *rfp*—the total genetic noise was affected only by the appropriate inducers (Fig. 4). Under conditions in which all fluorescent proteins were induced (Fig. 2I), we calculated correlation coefficients for each pair of fluorescent reporters (Table I). In all cases, the linear (Pearson) correlation coefficient agreed well with the rank (Spearman) correlation. We also calculated the partial correlation coefficients (Methods), which indicated the degree of extra correlation between two variables relative to the third. We found that the correlation of genetic

noise between *rfp* and *yfp* was consistently higher than the correlation of either *rfp* or *yfp* with *cfp*. This increased correlation was expected, due to their co-regulation by LacI. These results show—using multiple correlation metrics—that noise in gene expression can reveal transcriptional co-regulation.

We confirmed this method of detecting co-regulation with three additional experiments (Table I). First, we measured the same system in a strain background (MG1655) containing approximately 100× lower endogenous levels of the repressor LacI (~30 copies/cell as opposed to ~3,000). In this strain, induction was not necessary to observe fluorescence significantly above background (not shown). The level of correlation between *rfp* and *yfp* was larger than the previous case (ρ = 0.94), as expected, due to the increased noise in LacI concentration at low copy number. Second, we measured the correlation between *yfp* and *rfp* in a strain containing a deletion of the lac operon *(ΔlacI)*. Third, we switched the LacI/AraC controlled *rfp* promoter with a constitutive one containing no lac operators *(ΔlacO)*. In both of these cases, the extra correlation between *yfp* and *rfp* disappeared when the regulatory connections were broken (Table I). These results show that the increased noise correlation of *yfp* and *rfp* was due to their transcriptional co-regulation by LacI, and that this co-regulation is observable even in the absence of the corresponding inducer IPTG.

These correlation results are consistent with a model of co-regulation by the noisy (Elowitz et al. 2002; Swain et al. 2002) transcription factor LacI. Since LacI regulates *yfp* and *rfp* simultaneously, the correlation should be symmetric in time; that is, the cross-correlation function should be symmetric with its maximum near zero. To test this prediction, we analyzed the three-color fluctuations in a growing microcolony using time-lapse microscopy (Figure 3). We used the scaffold in the wild type MG1655 strain (containing no TetR, and low LacI), and grew microcolonies with arabinose as the carbon source to ensure AraC induction. When averaged over the cells present in the microcolony (Fig. 3A), the time-series revealed strong correlation

between *yfp* and *rfp*, consistent with the measurements of Table I (Fig. 3B). The cross correlation function (Fig. 3C) between *yfp* and *rfp* showed that the peak correlation occurs near zero lag, indicating that the co-regulation is mediated instantaneously by a separate element: LacI. These results show that the cross correlation function can reveal details into transcriptional regulation, such as whether two genes interact directly or are themselves regulated by an unobserved (and possibly unknown) factor.

By varying the induction of the reporter genes, we found that correlations in fluctuation were sensitive to shared regulatory inputs, suggesting that measurements of fluctuations might provide insight into regulatory structure of natural genetic circuits. We also found that this correlation could be detected even without inducing the transcription factor LacI, when the basal expression level was high enough to observe. Because the system is present on a plasmid of ~12 copies/cell, noise in expression of reporter genes is relatively small, allowing upstream fluctuations (Pedraza and van Oudenaarden, 2005) to be inferred. In particular, fluctuations in wild type (MG1655) LacI are large enough to cause correlated variation in two different LacI-regulated promoters. The presence of the third color acts as a control to compensate for extrinsic global and plasmid level fluctuations. Time-lapse microscopy adds to this analysis the relative delays in transcriptional regulation.

Given two uncharacterized promoters, one could test for the possibility of co-regulation by comparing their mutual correlation to that of a third, constitutively regulated promoter. This method could be used to confirm regulation by a global transcription factor with many putative binding sites, such as CRP (Brown and Callan, 2004). Alternatively, a pairwise comparison of all active promoters in *E. coli* (Zaslaver et al. 2006) could reveal co-regulation by as-yet uncharacterized transcription factors: about 150 transcription factors—nearly half—remain to be characterized in *E. coli* (Riley, 1993). Fluctuations and the correlations therein could even be used to identify unknown transcription factors or, when coupled with our prior knowledge of the natural

distributions of regulatory network motifs (Shen-Orr et al. 2002), to infer regulatory structures of unknown networks. We note that correlation in the promoter activities does not imply transcriptional co-regulation. Rather, this method could be coupled with time-lapse analysis of the correlation lags, along with traditional biochemical and genetic characterization, to quickly identify cases worthy of detailed study.

The synthetic biologist can use our system to aid in genetic circuit design and tuning. Multiple outputs monitor the state components of the system, and tell us when expression levels are not what we expect. A noise source (such as LacI) can be used as an input of a synthetic regulatory network, and the propagation of noise (Pedraza and van Oudenaarden, 2005) can be monitored to confirm the intended network structure.

It is now possible to design extremely complex genetic sequences with well-controlled behavior, by exploiting total DNA synthesis. This quantitative reporter system permits unprecedented quantitative analysis of natural and synthetic gene networks. Simultaneously, this system will test and expand our DNA sequence-level knowledge of the regulation of gene expression.

Materials and Methods

Reagents

All inducers and chemicals were purchased from Sigma. Concentrations (unless otherwise stated) were 50 μg/mL kanamycin, 100 μg/mL ampicillin, 500 μM isopropyl β-D-1-thiogalactopyranoside (IPTG), 100 ng/mL anhydrotetracycline (aTc), 0.1% L(+)-arabinose (Lara), 1μM oxo-C6-homoserine lactone (VAI). LB growth media (Lennox) was used for culture growth. All enzymes for plasmid construction were obtained from New England Biolabs.

Synthetic DNA design

The 3,518 bp scaffold was constructed by total synthesis by DNA2.0 (Menlo Park, CA). The sequence was cloned into the plasmid vector pDrive, and confirmed by automated sequencing. The full sequence of the scaffold in plasmid pFS2-123 is supplied in the Supplementary information, with features of the scaffold described here and bp coordinates listed in brackets. Restriction sites NotI [3], BbsI [235], BglII [1006], SalI [1088], AscI [1197], XhoI [1310], BamHI [1401], BsmBI [2165], PacI [2373], BsrGI [2590], AvrII [3348], XmaI [3462], and NheI [3511] allow for insertion and swapping of genetic elements between the promoters, open reading frames (ORFs), and terminators of the construct. The three fluorescent protein ORFs were codon optimized for expression in both gram-negative *E. coli* and gram-positive *B. subtilis*. These sequences were optimized by sampling each codon independently, in proportion to the average of the codon-usage tables for these two organisms (Nakamura et al. 2000). To explicitly avoid codons known to be toxic (Zahn, 1996), we used modified versions of the referenced codon-usage tables, with the rarest (less than 10% usage) codons removed. After performing the codon optimization of each ORF the restriction sites from the above list were removed, by choosing the mutation which introduced the most common codon. For the *cfp* sequence [260-976], we used the amino acid sequence of the Cerulean *cfp* variant (Rizzo et al. 2004) for this backtranslation

and optimization. For the *yfp* sequence [1434-2150], we used the amino acid sequence of the Venus *yfp* variant (Nagai et al. 2002), and incorporated the mutations of the Citrine *yfp* variant (Griesbeck et al. 2001). For the *rfp* sequence [2584-3294], we used the amino acid sequence of the Cherry *rfp* variant (Shu et al. 2006). Double stop codons (TAATAA) were used at the end of all three ORFs to ensure efficient termination of translation. We used promoters $P_{LtetO-1}$ [1013-1086] to control *cfp*, $P_{LlacO-1}$ [1318-1397] to control *yfp*, and $P_{lac/ara-1}$ [3353-3455] to control *rfp* (Lutz and Bujard, 1997). To control translation, we used the moderate strength SD8 RBSs (Ringquist et al. 1992) for *cfp* [977-996] and *yfp* [1414-1433], and the stronger RBS from gene 10 of phage T7 (Olins et al. 1988) for *rfp* [3296-3336]. Terminators RNAI [19-64] and TSAL [65-238] (Reynolds et al. 1992) terminated the transcriptional unit containing *cfp*. Terminators TR2-17 [2166-2288] (Wilson and von Hippel, 1995), TL17 [2289-2365] (Wright et al. 1992), BS7 [2378-2430] (Reynolds et al. 1992), and T7TE+ [2451-2577] (Uptain and Chamberlin, 1997) terminated the transcriptional units containing *yfp* and *rfp*.

Plasmids

The initial synthetic construct was subcloned into the modular pZ* expression vector system (Lutz and Bujard, 1997) using the NotI and NheI restriction sites on each end of the scaffold construct. This plasmid system allows easy swapping of the origin of replication (SC101, ColE1, or p15A) and antibiotic resistance markers. We designated plasmids containing the scaffold as pF*, and adopted an extended version of the pZ* naming system. Data for Figs. 2-3 is from measurement of plasmid pFS2-123, containing: a kanamycin resistance marker; the SC101 origin of replication, and the promoters described above. To measure the correlation between *rfp* and *yfp* with the LacI regulation of *rfp* removed (Table I, ΔlacO), we placed the *rfp* gene under the control of the P(R) promoter from phage λ (Ptashne, 2004). This promoter sequence was synthesized and cloned in between the restriction sites XmaI and AvrII to create plasmid pFS2-12R.

Strains

We used wild-type *E. coli* MG1655 (Blattner et al. 1997) for the time-lapse experiment (Figure 3), and to measure the correlation between *rfp* and *yfp* from plasmid pFS2-123 in the presence of low levels of LacI (Table I, MG1655). This strain does not contain the TetR repressor, so *cfp* is expressed constitutively. The MG1655Z1 strain was constructed from the wild type MG1655 strain and the TetR and LacI over-expressing DH5aZ1 (Lutz and Bujard, 1997) strain by P1 general transduction (Miller, 1972). We used MG1655Z1, which over-expresses LacI from the lacIq cassette, for the characterization of the scaffold response to induction in the presence of the (Fig. 2). We also measured the correlation between *rfp* and *yfp* in strain MC4100, which does not contain the LacI gene (Table I, *ΔlacI*) .

Microscopy

Single-cell measurements were acquired on an Olympus IX-81 inverted fluorescence microscope at 100× magnification, with a Hammamatsu Orca ER CCD camera (2x2 binning) using custom microscope acquisition software. Phase-contrast images were acquired to measure cell morphology, position, and image quality. Fluorescent exictation was performed with a Lambda LS Xenon lamp (Sutter Instruments, Inc.) with a liquid lightguide and fluoresent filter cubes (Chroma, Inc.) for Cyan/*cfp* (Chroma, #31044v2), Yellow/*yfp* (Chroma, #41028), and Crimson/*rfp* (Chroma, #41027). To prevent bleaching, all images were collected as ordered exposures of (*rfp*, *yfp*, *cfp*), with minimal light exposure. We verified the fluorescent field provided by the Lamda LS light source and liquid light guide with fluorescent slides (Spherotech, Inc.). The field was found to be extremely flat (std/mean ~ %3) when centered in all three colors.

In order to check and correct for spectral crosstalk between fluorescent proteins, we constructed plasmids each containing each individual fluorescent protein. We measured cells expressing only one of the fluorophores *cfp*, *yfp*, and *rfp* in all three filter cubes (Table II). Crosstalk was very small in all cases. The highest magnitude was *rfp* fluorescence in the Yellow/*yfp* channel, which amounted

to 0.1% of the detection level in the *rfp* channel. The crosstalk of *cfp* into the Crimson/*rfp* cube was undetectable on our system. All reported data are corrected for this crosstalk (using the inverse matrix of Table II). Errors in crosstalk measurement could conceivably introduce false correlations into Table I. To control for this possibility, we repeated all data analysis without the crosstalk correction and found no change in any of the results of Table I. These results confirm good spectral separation.

Induction experiment

MG1655Z1 cells containing the plasmid pFS2-123 were grown to saturation overnight in LB at 37°C and diluted 100× into non-fluorescent M9 minimal glycerol media containing combinations of the three inducers. Cells were grown for 3 hours at 32°C to an OD600 of ~0.2. For cells induced with aTc, an additional 50ng/mL of aTc was then added to insure complete induction. Cells were allowed to grow to a final OD600 of ~0.3, placed on ice, and measured on 1.5% low melting temperature agarose phosphate-buffered saline slabs in the microscope. For each condition, we acquired approximately 20 fields of cells. These measurements contained 500-1000 cells per condition measured.

Time-lapse experiment

MG1655 cells containing the plasmid pFS2-123 were grown to saturation overnight in LB at 37°C and diluted 1000× into non-fluorescent M9 arabinose media. Cells were grown for 3 hours at 32°C, then diluted 100× and transferred to M9 arabinose media pads made with 1.5% low melting point agarose. These pads were placed inside a glass Wilco dish chamber and sealed. Time-lapse images were acquired as described below, at 10 minute intervals in a 32°C temperature-controlled chamber.

Image processing

We used custom software and the Matlab (The Mathworks, Inc.) Image Processing Toolbox to

segment the phase contrast images and collect corresponding pixels from each of the three fluorescent images. The program identified individual cells on phase contrast images by progressive watershed thresholds. Shapes were filtered based on morphological properties to eliminate non-cell objects, clumps of cells, and misshapen cells. For each cell, a background value of the minimum pixel contained in the bounding box was recorded for each color. Collected fields of cells were examined by eye to check for errors in segmentation and acquisition. For the time-lapse experiment, we used custom software to identify cell division events, and track lineages of cells (lines of daughter and parent cells) during microcolony growth (Rosenfeld et al. 2006).

Data analysis

We extracted the data from the segmentation program, subtracted background value for each cell, and normalized each color with respect to the camera's exposure time for that image. After extracting the data structure from the segmentation, we collected the autofluorescence measurements as a daily control. The autofluorescent values were normally distributed (not shown). We then corrected for the spectral crosstalk measured above by multiplying the 3-color data from each strain by the inverse of the spectral crosstalk matrix (Table II). We also tested many corrections for sources of systematic or experimental error. Parabolic fluorescent field correction did not change the qualitative relationships or reduce variation. The overall fluorescence variation between fields of cells remained small: Each frame analyzed was within one standard deviation of the mean over all frames of the same color. Normalization to account for morphological factors such as size and shape did not qualitatively change our results or decrease the observed variation. As a final correction, we removed outlying cells and non-cell objects from the processed data that were more than three standard deviations from the median of the (typically 500-1000) cells. Previous noise measurements have used similar corrections (Elowitz et al. 2002; Pedraza and van Oudenaarden, 2005)

Correlation analysis

For each processed data set, we calculated the normalized Pearson and Spearman correlation coefficient between each pair of colors (Table I). Using these three pairwise correlations, we calculated the three partial correlation coefficients:

$$\rho_{xylz} = \frac{\rho_{xy} - \rho_{xz} * \rho_{yz}}{\sqrt{1-\rho_{xz}^{2}}\sqrt{1-\rho_{yz}^{2}}}$$

To calculate the errors in correlation and partial correlation coefficients, we uniformly re-sampled 1,000 data sets (bootstrap sampling with replacement) of the same size and recomputed the correlation coefficients for each sample. The errors reported in Table I and Fig. 2B are determined by the 90% confidence intervals of this bootstrap procedure (i.e., by taking the 100th and 900th values of the sorted list of resampled correlation coefficients).

For the time-lapse movie, we calculated the Pearson correlation coefficient for each pair of colors over the cells present in the entire microcolony at each time point (Fig. 3B). Since the number of cells increases with time, the error bars on each correlation coefficient become smaller with time. We were able to resolve the correlation coefficients after about 7 hours of growth, corresponding to ~100 cells per microcolony. Using the same movie, we calculated the cross-correlation function between each cell lineage for all three pairs of colors. The cross correlation function measures the degree of correlation between two signals, as a function of the delay between them. As such, the cross correlation function reaches a maximum at the time-delay for which the correlation in signals is highest (Fig. 3C). The details of calculating cross correlation analysis over a branching tree will be described in an upcoming publication by Mary J. Dunlop et al. The result presented here—that the cross correlations peak near zero delay—does not depend on the particular method used to calculate the cross-correlation.

Figure 1.

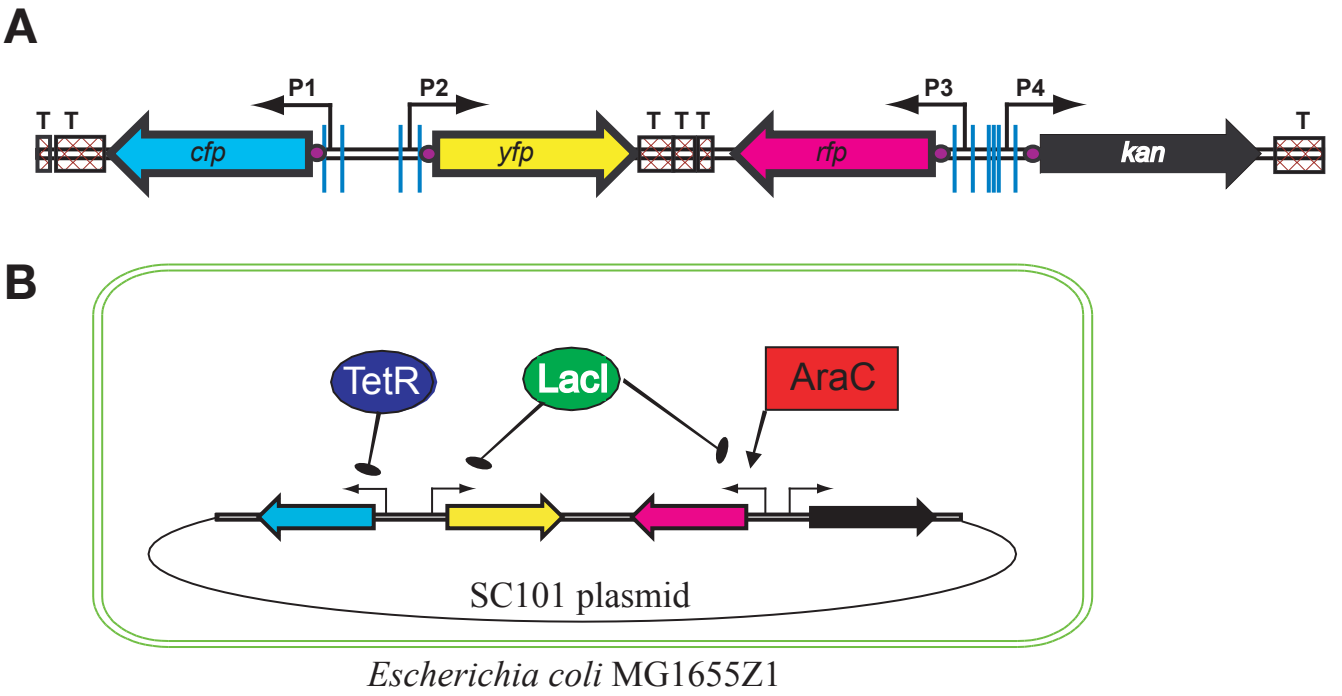


Figure 1. *The framework design.* (A) Terminators are hatched boxes. RBSs are purple circles. Restriction sites are blue bars. The promoters are small black arrows. Each fluorescent operon is shown as a colored block arrow, while the (kanamycin) antibiotic resistance is shown as a black block arrow. (B) The framework is measured in the low-copy plasmid in wild type *E. coli* strain containing the native ara operon, and a LacI and TetR over-expressing cassette. The three reporters are controlled by promoters responsive to: tetracycline/aTc (*cfp*), lactose/IPTG (*yfp*), and both lactose/IPTG and arabinose/Lara (*rfp*).

Figure 2.

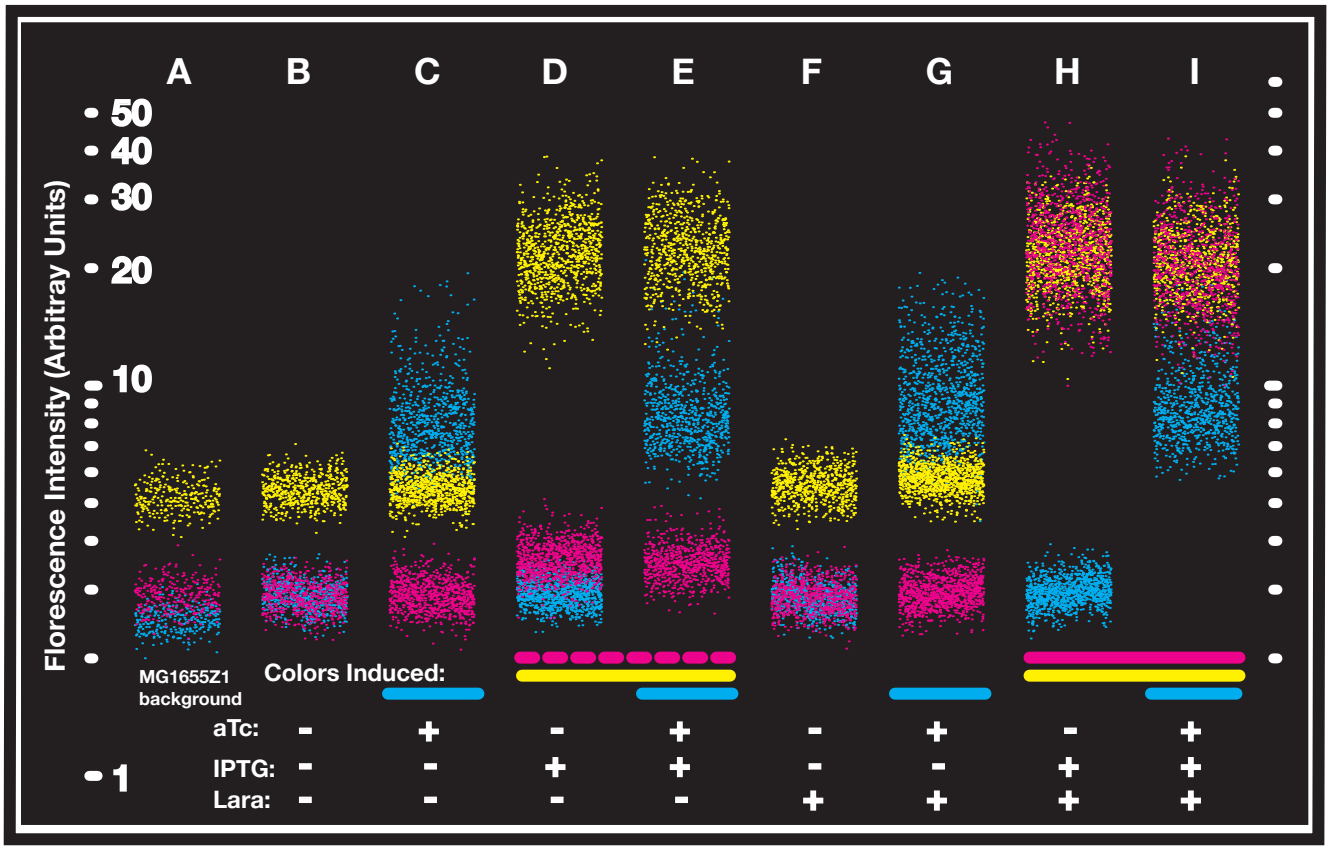


Figure 2. *Operons provide independent control of gene expression.* Fluorescence microscopy snapshots were taken of 500-1000 cells under each combination of saturating inducer concentrations. This plot shows the response of each reporter to different combinations of these three inducers (each column is one condition, the color expected is shown as a bar below). Each cell within the population is represented by three dots—one of each color—in order to show the genetic noise in each condition. Note that for the LacI/AraC regulated promoter (*rfp*) the expression is only slightly increased by induction of LacI only, and not at all (0%) by induction of AraC alone.

Figure 3.

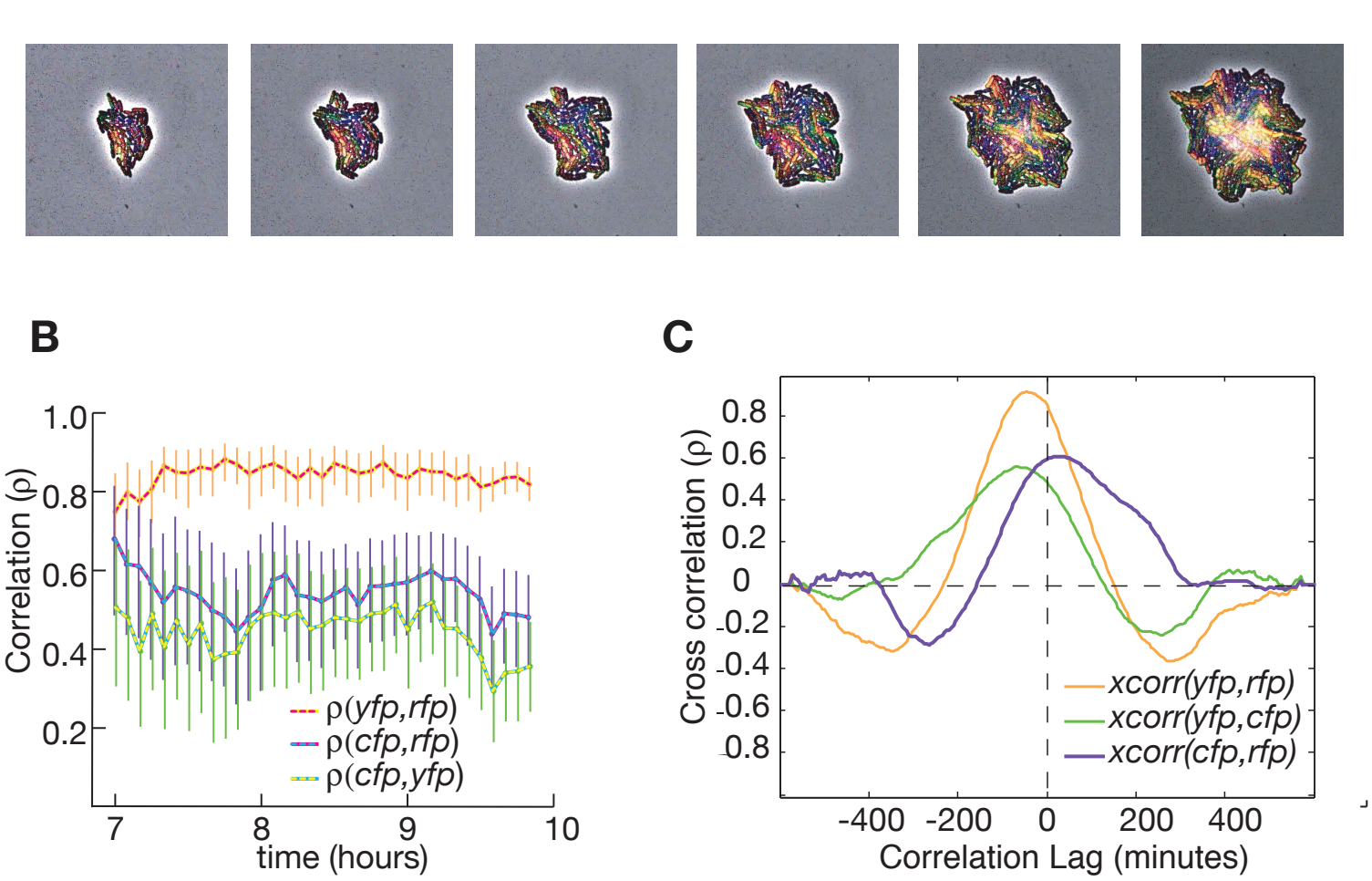


Figure 3. *The cross-corelation function reveals regulatory connections.* We monitored the levels of *cfp*, *yfp*, and *rfp* expression from the plasmid shown in Fig. 1B during growth in a microcolony of *E. coli* MG1655. This strain did not contain the TetR, leaving *cfp* constitutively active. We grew the microcolonies on agarose pads, using arabinose as a carbon source to ensure AraC induction. (A) Time-lapse images of 3-color expression, where the pseudo-colors indicate the expression levels of *cfp* (blue), *yfp* (green), and *rfp* (red), respectively. Cells appearing yellow reveal the correlation between *yfp* and *rfp* due to LacI co-regulation. (B) The average correlation between *yfp* and *rfp* persists over several hours of microcolony growth. (C) The cross correlation between *yfp* and *rfp* reveals that the co-regulation by LacI has zero-lag, i.e., it is instantaneous.

Figure 4.

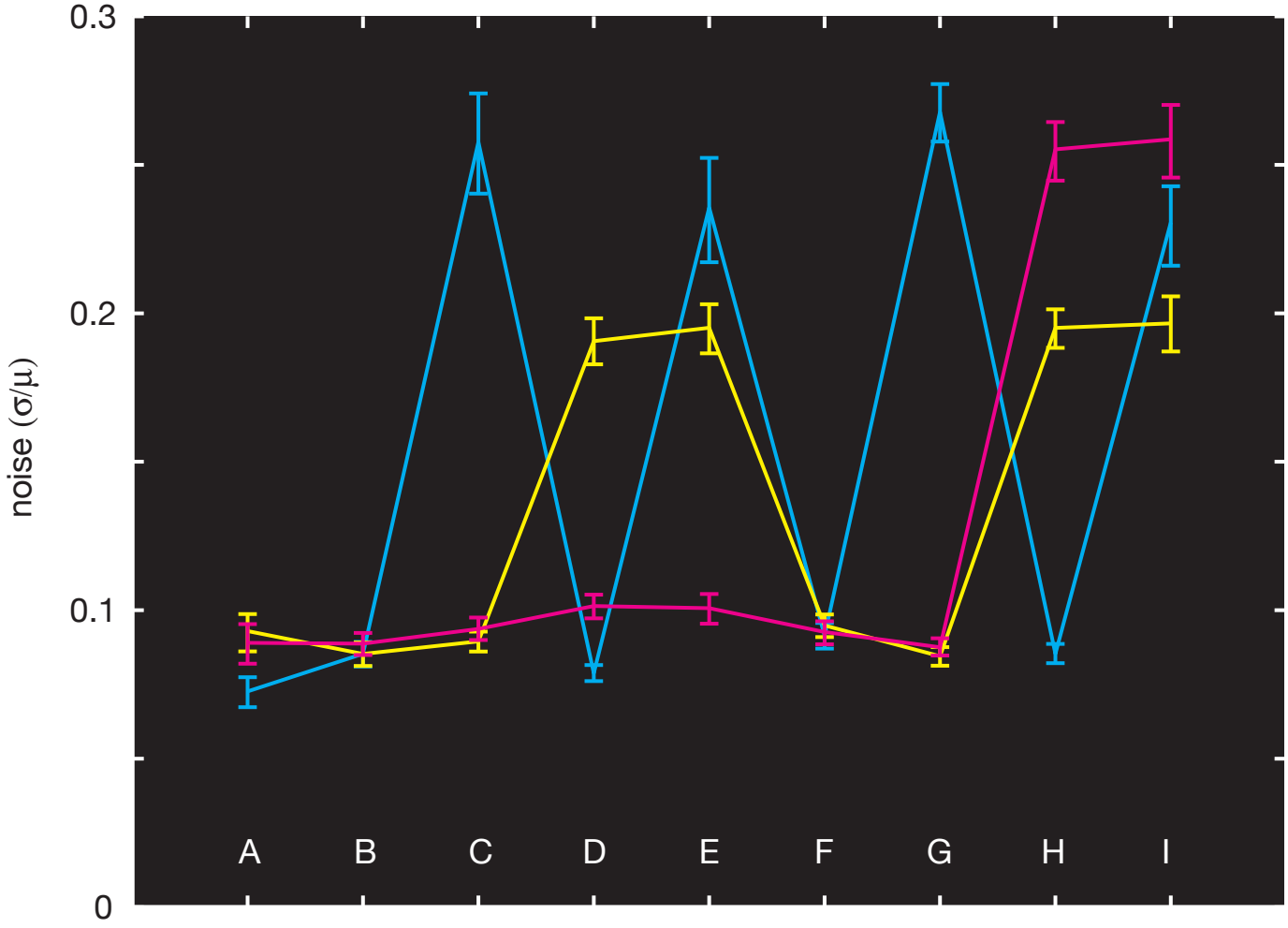


Figure 4. *Total genetic noise is controlled by induction.* The total genetic noise, calculated as the standard error divided by the mean, is plotted for each of the conditions in Fig. 2. Here cyan corresponds to noise in *cfp*, yellow to noise in *yfp*, and red to noise in *rfp*. In each case, the noise is maximal in the fully induced state. Notably, the noise of each color is only affected by the inducer(s) which control it: aTc for *cfp*, IPTG for *yfp*, and both IPTG and Lara for *rfp*.

Table I. Multi-color noise correlations reveal co-regulation.

	CORRELATION	PARTIAL CORRELATION	RANK CORRELATION	PARTIAL RANK CORRELATION
ρ(*cfp*,*yfp*)	0.43 ± 0.06	0.23 ± 0.10	0.40 ± 0.06	0.19 ± 0.07
ρ(*cfp*,*rfp*)	0.37 ± 0.07	-0.07 ± 0.11	0.36 ± 0.06	-0.04 ± 0.08
ρ(*yfp*,*rfp*) **lacI^q**	0.85 ± 0.02	0.83 ± 0.02	0.85 ± 0.01	0.71 ± 0.03
ρ(*yfp*,*rfp*) **MG1655**	0.94 ± 0.01	0.93 ± 0.01	0.94 ± 0.01	0.93 ± 0.01
ρ(*yfp*,*rfp*) **Δ*lacI***	0.48 ± 0.13	0.02 ± 0.18	0.51 ± 0.14	0.16 ± 0.18
ρ(*yfp*,*rfp*) **Δ*lacO***	0.38 ± 0.12	-0.14 ± 0.09	0.19 ± 0.07	-0.14 ± 0.07

Table II. Three fluorescent reporters exhibit spectral separation[1].

Cube \ *gene*	*cfp*	*yfp*	*rfp*
Cyan	1.0×10^{0}	1.5×10^{-4}	1.0×10^{-4}
Yellow	5.0×10^{-4}	1.0×10^{0}	1.1×10^{-3}
Crimson	0.0×10^{0}	1.3×10^{-5}	1.0×10^{0}

1 The crosstalk is calculated as the ratio of the intensity in the given channel divided by the intensity in the primary channel (*e.g* the crosstalk of *yfp* into the Yellow cube is 1).

References

Ai HW, Shaner NC, Cheng Z, Tsien RY, Campbell RE (2007) Exploration of new chromophore structures leads to the identification of improved blue fluorescent proteins. *Biochemistry* **46:** 5904-5910.

Blattner FR, Plunkett G, 3rd, Bloch CA, Perna NT, Burland V, Riley M, Collado-Vides J, Glasner JD, Rode CK, Mayhew GF, Gregor J, Davis NW, Kirkpatrick HA, Goeden MA, Rose DJ, Mau B, Shao Y (1997) The complete genome sequence of *Escherichia coli* K-12. *Science* **277:** 1453-1474.

Brendel V, Hamm GH, Trifonov EN (1986) Terminators of transcription with RNA polymerase from *Escherichia coli*: what they look like and how to find them. *J Biomol Struct Dyn* **3:** 705-723.

Brown CT, Callan CG, Jr. (2004) Evolutionary comparisons suggest many novel cAMP response protein binding sites in *Escherichia coli*. *Proceedings of the National Academy of Sciences of the United States of America* **101:** 2404-2409.

Campbell RE, Tour O, Palmer AE, Steinbach PA, Baird GS, Zacharias DA, Tsien RY (2002) A monomeric red fluorescent protein. *Proceedings of the National Academy of Sciences of the United States of America* **99:** 7877-7882.

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

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

Farmer WR, Liao JC (2000) Improving lycopene production in *Escherichia coli* by engineering metabolic control. *Nature biotechnology* **18:** 533-537.

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

Griesbeck O, Baird GS, Campbell RE, Zacharias DA, Tsien RY (2001) Reducing the environmental sensitivity of yellow fluorescent protein. Mechanism and applications. *The Journal of biological chemistry* **276:** 29188-29194.

Khosla C, Keasling JD (2003) Metabolic engineering for drug discovery and development. *Nature reviews* **2:** 1019-1025.

Lutz R, Bujard H (1997) Independent and tight regulation of transcriptional units in *Escherichia coli* via the LacR/O, the TetR/O and AraC/I1-I2 regulatory elements. *Nucleic acids research* **25:** 1203-1210.

Miller JH (1972) *Experiments in molecular genetics*. [Cold Spring Harbor, N.Y.]: Cold Spring Harbor Laboratory.

Nagai T, Ibata K, Park ES, Kubota M, Mikoshiba K, Miyawaki A (2002) A variant of yellow fluorescent protein with fast and efficient maturation for cell-biological applications. *Nature biotechnology* **20:** 87-90.

Nakamura Y, Gojobori T, Ikemura T (2000) Codon usage tabulated from international DNA sequence databases: status for the year 2000. *Nucleic acids research* **28:** 292.

Olins PO, Devine CS, Rangwala SH, Kavka KS (1988) The T7 phage gene 10 leader RNA, a ribosome-binding site that dramatically enhances the expression of foreign genes in *Escherichia coli*. *Gene* **73:** 227-235.

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

Ptashne M (2004) *A genetic switch : phage lambda revisited*, 3rd edn. Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory Press.

Reynolds R, Bermudez-Cruz RM, Chamberlin MJ (1992) Parameters affecting transcription termination by *Escherichia coli* RNA polymerase. I. Analysis of 13 rho-independent terminators. *Journal of Molecular Biology* **224:** 31-51.

Riley M (1993) Functions of the gene products of *Escherichia coli*. *Microbiol Rev* **57:** 862-952.

Ringquist S, Shinedling S, Barrick D, Green L, Binkley J, Stormo GD, Gold L (1992) Translation initiation in *Escherichia coli*: sequences within the ribosome-binding site. *Molecular microbiology* **6:** 1219-1229.

Rizzo MA, Springer GH, Granada B, Piston DW (2004) An improved cyan fluorescent protein variant useful for FRET. *Nature biotechnology* **22:** 445-449.

Rosenfeld N, Perkins TJ, Alon U, Elowitz MB, Swain PS (2006) A fluctuation method to quantify *in vivo* fluorescence data. *Biophys J* **91:** 759-766.

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

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

Shu X, Shaner NC, Yarbrough CA, Tsien RY, Remington SJ (2006) Novel chromophores and buried charges control color in mFruits. *Biochemistry* **45:** 9639-9647.

Suel GM, Garcia-Ojalvo J, Liberman LM, Elowitz MB (2006) An excitable gene regulatory circuit induces transient cellular differentiation. *Nature* **440:** 545-550.

Swain PS, Elowitz MB, Siggia ED (2002) Intrinsic and extrinsic contributions to stochasticity in gene expression. *Proceedings of the National Academy of Sciences of the United States of America* **99:** 12795-12800.

Uptain SM, Chamberlin MJ (1997) *Escherichia coli* RNA polymerase terminates transcription efficiently at rho-independent terminators on single-stranded DNA templates. *Proceedings of the National Academy of Sciences of the United States of America* **94:** 13548-13553.

Wilson KS, von Hippel PH (1995) Transcription termination at intrinsic terminators: the role of the RNA hairpin. *Proceedings of the National Academy of Sciences of the United States of America* **92:** 8793-8797.

Wright JJ, Kumar A, Hayward RS (1992) Hypersymmetry in a transcriptional terminator of *Escherichia coli* confers increased efficiency as well as bidirectionality. *The EMBO journal* **11:** 1957-1964.

Zahn K (1996) Overexpression of an mRNA dependent on rare codons inhibits protein synthesis and cell growth. *Journal of bacteriology* **178:** 2926-2933.

Zaslaver A, Bren A, Ronen M, Itzkovitz S, Kikoin I, Shavit S, Liebermeister W, Surette MG, Alon U (2006) A comprehensive library of fluorescent transcriptional reporters for *Escherichia coli*. *Nature methods* **3:** 623-628.

Contribution for Chapter 4

As second author of this paper[1], Cox directly participated in all theoretical and experimental aspects of the project. The construction of the synthetic circuit was performed by the first author You, while Cox assisted with the initial characterization of individual circuit components: the LuxI quorum-signal synthesis gene, the stability of AHL at various pH levels and growth media, several killer gene variants, and the promoter controlling the individual quorum sensing genes. Cox and You contributed equally to the mathematical modeling, including the analytical solutions describing the damped oscillations precededing the rise to steady level cell densities at low pH. You collected most of the primary data presented in the figures, including the LacZ gene expression assays; Cox optimized the growth curve assays and measurements of cell density. Cox further characterized the circuit response to endogenous signal, measured long-term circuit function to establish reliable limits on its function, and assessed the single-cell variability of cell death.

1 You L, Cox RS, 3rd, Weiss R, Arnold FH (2004) Programmed population control by cell-cell communication and regulated killing. *Nature* **428:** 868-871.

Programmed Population Control by Cell-Cell Communication and Regulated Killing

Lingchong You, Robert Sidney Cox III, Ron Weiss, and Frances H. Arnold

Major challenges confront the *de novo* engineering of gene circuits inside cells (Atkinson et al. 2003; Becskei and Serrano, 2000; Chen et al. 1993; Elowitz and Leibler, 2000; Farmer and Liao, 2000; Gardner et al. 2000; Isaacs et al. 2003; Weiss et al. 1999; Yokobayashi et al. 2002), where efforts to realize predictable and robust performance must deal with noise in gene expression and cell-to-cell variation in phenotype (Blake et al. 2003; Elowitz et al. 2002; Ozbudak et al. 2002). Here we demonstrate that coupling gene expression with cell survival and death using cell-cell communication enables us to program the dynamics of a population despite variability in the behavior of individual cells. Specifically, we have built and characterized a "population control" circuit that autonomously regulates the density of an *E. coli* population. The cell density is broadcasted and detected by elements from a bacterial quorum sensing system (Fuqua et al. 1994; Miller and Bassler, 2001), which in turn regulate the death rate. As predicted by a simple mathematical model, the circuit can set a stable steady state in terms of both cell density and gene expression, which is easily tunable by varying the stability of the cell-cell communication signal. Incorporating a mechanism for programmed death in response to changes in the environment, this minimal synthetic construct allows us to probe the design principles of its more complex natural counterparts.

The circuit (Fig. 1A) programs a bacterial population to maintain a cell density that is lower than the limits imposed by the environment (e.g., nutrient supply). The LuxI protein of the well-characterized LuxI/LuxR system from the marine bacterium *Vibrio fischeri* (Fuqua et al. 1994; Miller and Bassler, 2001) synthesizes a small, diffusible acyl-homoserine lactone (AHL) signaling molecule. The AHL accumulates in the medium and inside the cells as the cell density increases. At sufficiently high concentrations, it binds and activates the LuxR transcriptional regulator, which in turn induces expression of a killer gene (*E*) under the control of a *luxI* promoter (p_{luxI}) (Egland and Greenberg, 1999). Sufficiently high levels of the killer protein cause cell death. We implemented the circuit using a two-plasmid system (Fig. 1B), where pLuxRI2 expresses LuxI and LuxR upon induction by isopropyl-beta-D-thiogalactopyranoside (IPTG), and p*lux*CcdB3 responds to activated LuxR (at high enough cell density) and causes cell death. The *lacZ*α*-ccdB* killer gene codes for a fusion protein of LacZα and CcdB. The LacZα portion of the fusion protein retains the ability to complement LacZΔM15 in appropriate cell strains (e.g., Top10F' cells), allowing the measurement of fusion protein levels by LacZ assay (Methods). The CcdB portion retains the toxicity of native CcdB, which kills susceptible cells by poisoning the DNA gyrase complex (Engelberg-Kulka and Glaser, 1999).

A simple mathematical model predicted that the system would reach a stable cell density for all realistic parameter values (Methods), although it might go through damped oscillations while approaching the steady state. Experiments confirmed our predictions. Fig. 2A shows the growth of Top10F' cells containing the population control circuit at pH 7.0. As anticipated, the uninduced culture (circuit OFF) grew exponentially and reached stationary phase upon nutrient exhaustion. The induced culture (circuit ON) grew almost identically to the OFF culture until its density reached a threshold (at 7 hr). It then deviated sharply from the OFF culture and briefly went through a damped oscillation (between 7 hr and 24 hr) of at least one cycle before reaching a steady-state density ~ 10 times lower than that of the OFF culture. The measured peak density (at 10 hr) was two-fold higher than the measured floor (12.5 hr), a difference significantly greater

than measurement variations. The steady state was maintained for more than 30 hours: the variation in cell density between 28 hr and 62 hr was < 5%, which was smaller than typical measurement variations of individual data points. The transient dynamics of circuit-ON growth was captured well in simulation, where the AHL degradation rate constant (d_A) was adjusted for the simulation to match the experiment in the steady-state circuit-ON cell density. Intracellular levels of LacZα-CcdB for the ON culture, measured in terms of LacZ activity, reached a steady state after an overshoot (Fig. 2B), again predicted by simulation. The basal level of LacZα-CcdB expression in the OFF culture was negligible for all time points.

The growth curves and the LacZα-CcdB time courses illustrate the tight yet dynamic coupling between population dynamics and intracellular expression of the LacZα-CcdB killer protein. A lower than steady-state level of the killer protein allows the population to grow; conversely, its excessive production decreases cell density. After some delay, the decline in cell density leads to a decrease in the AHL concentration, which in turn leads to reduced levels of the killer protein, allowing the population to recover. Continuous production and degradation (or death) of each circuit component are essential for the observed homeostasis, and they are closely coupled. Any perturbation that decouples or overrides these processes will disrupt circuit function. For example, the circuit without *luxI* (thus lacking a communication link) could not control growth (not shown). Also, we observed that 200 nM exogenous AHL, which was not toxic to cells without the circuit or with the circuit OFF, completely prevented growth with the circuit ON (not shown). This is expected, because a high level of AHL would activate LuxR and lead to overproduction of the killer protein. This observation also verifies that the killer protein production rate was limited by AHL synthesis in circuit-ON growth, a key assumption in our mathematical model.

Circuit function could also be delicately manipulated in a predictable fashion. Our model predicted that the steady-state cell density would increase nearly proportionally with the AHL

degradation rate constant (Methods). Thus, AHL serves as an external "dial" to operate the circuit: AHL degradation affects cell-cell communication, and rapid AHL breakdown can disrupt it completely. Degradation of AHLs is facilitated by hydrolytic enzymes (Dong et al. 2001; Leadbetter and Greenberg, 2000) or by increasing the medium pH (Schaefer et al. 2000). Confirming the model prediction, a moderate increase in the medium pH (6.2 to 7.8) significantly increased (~ 4-fold) steady-state cell densities of the ON cultures, but caused only minor changes in those of the OFF cultures (Fig. 3A–E, Table 1). For each pH value, circuit-ON populations reached a steady state after 28 hours, and the variation of cell density afterwards was smaller than typical measurement variations. Similar to the pH 7.0 case, the ON cultures grew almost identically to the OFF cultures at low cell density, but deviated from the latter at high density. Again, simulations captured the experimental behavior (Fig. 3A–D), with adjustment only of the AHL degradation rate constant (Table 1).

Our model predicted that, unlike cell density, intracellular levels of the killer protein would remain roughly constant as pH varied. At steady state, $E_s = k/d\ (1\text{-}N_s/N_m) \approx k/d$, the ratio of the growth and killing rate constants (assuming that the circuit-ON cell density is far below the carrying capacity, i.e., $N_s/N_m << 1$). Thus if pH did not significantly affect the growth or killing rate constants, the killer protein would reach concentration k/d, independent of pH. Experimental results showed that LacZα-CcdB for different pH reached similar levels after about 28 hours (Fig. 3F–I, Table 1). Nevertheless, the pH changes affected other measured parameters, including k, N_m, and more notably N_s (Table 1). To also account for effects of these changes on killer protein expression, we normalized LacZ activity with respect to $k(1\text{-}N_s/N_m)$. According to the model, the normalized LacZ activity should remain constant if pH does not affect the killing rate constant (because $E_s/(k(1\text{-}N_s/N_m)) = 1/d$). Supporting the prediction, normalized LacZ activities were nearly identical for different pH (Fig. 3J). Except for pH 8.05, these values were statistically indistinguishable ($p = 0.26$). The slightly but significantly higher ($p = 8.6\times10^{-4}$) normalized LacZ activity at pH 8.05 suggests that the killer protein was less toxic (smaller d).

Construction of *de novo* gene circuits has recently emerged as an approach to decoding "design principles" of biological systems (Hasty et al. 2002; Wall et al. 2004; Weiss et al. 2003). Studies have demonstrated the feasibility of building gene circuits that lead to stable (Becskei and Serrano, 2000), bi-stable (Atkinson et al. 2003; Gardner et al. 2000; Isaacs et al. 2003), or oscillatory (Atkinson et al. 2003; Elowitz and Leibler, 2000) gene expression, act as a digital logic inverter (Weiss et al. 1999; Yokobayashi et al. 2002) or a metabolic controller (Farmer and Liao, 2000), or provide better-regulated gene expression systems (Chen et al. 1993). They have also revealed major hurdles to achieving reliable circuit behavior, such as noise in cellular processes and cell-to-cell variation (Blake et al. 2003; Elowitz et al. 2002; Ozbudak et al. 2002). Here we addressed these issues by using cell-cell communication to coordinate behavior across the population. By coupling gene expression with population dynamics, cell-cell communication integrates the entire population as an essential component of the population control circuit. This coupling enables the circuit to function reliably at the population level by exploiting cell heterogeneity in terms of their size, age, plasmid copy number, gene expression, and response to the killer protein. Such phenotypic variations, which interfere with the functioning of circuits lacking a communication mechanism, are actually required for the population control circuit to work. The outcome of circuit function is binary for any given cell: it lives or dies (judged by the ability to form a colony). If all cells had the same phenotype, the circuit would fail to achieve a stable cell density because the population would go extinct once the killer protein concentration reached a critical threshold. Also due to the coupling, the circuit can only function at the population level. This is supported by the experimental data: under each set of conditions, growth of the circuit-ON culture only deviated from that of the OFF culture when cell density was sufficiently high (Figs. 2 and 3). If the circuit had functioned inside each cell autonomously, growth of the two cultures would have deviated from the beginning.

Like multicellular organisms, bacteria possess sophisticated suicide machinery triggered by stress and starvation or by "addiction modules" during post-segregational killing (Ameisen, 2002;

Engelberg-Kulka and Glaser, 1999; Lewis, 2000). Natural systems, such as the signaling network dictating the lysis of a subpopulation of *S. pneumoniae* at high cell density, employ a mechanism similar to our circuit of regulated killing coordinated by quorum sensing (Lewis, 2000; Steinmoen et al. 2002). Due to selection pressure, mutants that escape regulation by the synthetic population control circuit usually arise three to six days after circuit activation (tested by dilution into fresh medium; data not shown). Although there are no direct measurements, escaping the regulation of natural systems is likely much rarer. The greater genetic stability of natural systems may be due to more sophisticated regulation or their coupling to other physiological processes, or both. For example, cell lysis is an essential stage of the development of natural transformation in *S. pneumoniae*, where survivors assimilate DNA released by lysed cells (Steinmoen et al. 2002). Thus there may be selection pressure favoring the overall signaling network (with the lysis regulation network as a subset) responsible for this developmental process. Design concepts such as these can be tested by introducing additional regulatory modules into the population control circuit or coupling the circuit to functions that are beneficial to the cells.

The population control circuit lays the foundation for using cell-cell communication to program interactions among bacterial communities, allowing the concept of communications-regulated growth and death to be extended to engineering synthetic ecosystems. The rich repertoire of natural quorum-sensing modules (Fuqua et al. 1994; Miller and Bassler, 2001), supplemented by engineered counterparts (Bulter et al. 2004; Gerchman and Weiss, 2004), will facilitate construction of multi-channel feedback systems where population densities are coupled to communications.

Materials and Methods

Plasmids

Plasmid pLuxRI2 (ColE1 origin, chloramphenicolR) was constructed by inserting PCR-amplified luxI from pSND-1 (Weiss and Knight Jr., 2000) into pLuxR2, downstream of LuxR. pLuxR2 was made in two steps: 1) pLuxR encoding LuxR under the control of plac/ara-1 was constructed by inserting PCR-amplified LuxR from pKE705 (Egland and Greenberg, 2001) into pPROLar.A122 (BD Biosciences Clontech); 2) pLuxR2 was made by inserting a fragment containing $p_{lac/ara-1}$-LuxR from restriction-digested pLuxR into pPROTet.E133 (BD Biosciences Clontech). Plasmid pluxCcdB3 (p15A origin, kanamycinR) was constructed by PCR-fusing lacZα-ccdB from pZErO-2 (Invitrogen) to pluxI from pluxGFPuv, and inserting the fused DNA into pPROLar.A122. Plasmid pluxGFPuv contains pluxI cloned from pKE555 (Egland and Greenberg, 1999). Both pLuxRI2 and pLuxCcdB3 were confirmed by sequencing. Versions of the circuit constructed using different combinations of replication origins, promoters, and CcdB variants with different lethality demonstrated similar phenotypes to those reported here.

Strains, growth conditions, and media

Unless otherwise stated, Top10F' cells (Invitrogen) were used throughout this study. Standard LB medium was used for cell growth to probe qualitative behavior. For quantitative measurements, cells were grown in pH-buffered TBK medium (10 g tryptone and 7 g KCl per liter buffered with 100 mM weak acids). Medium pH (measured with Accumet® pH Meter 925, Fischer Scientific) was adjusted by adding 5M KOH. 3-(N-morpholino) propanesulfonic acid (MOPS) was used to buffer pH between 7 and 8, and Piperazine-N,N'-bis(2-ethanesulfonic acid) (PIPES) was used for pH between 6.2 and 6.8. The buffered media were able to maintain pH well: variations in pH were < 0.3 and mostly occurred within first 24 hrs of growth. Plasmids were maintained with 100 μg/ml of chloramphenicol and 50 μg/ml of kanamycin. 1 mM IPTG was used to activate the circuit.

To measure circuit function, 3 ml LB in a 12 ml culture tube was inoculated from a single colony and incubated overnight at 30˚ C and 250 rpm. At an optical density (OD) between 0.1~0.3 (measured at 600 nm with a Spectra MAX250 microplate reader, Molecular Devices), the overnight culture was diluted 1000fold into 50ml of fresh buffered TBK in 250 ml flasks supplemented with antibiotics and IPTG when applicable. The flask cultures were incubated at 34C and 250rpm. Antibiotics and inducer were replenished about 50 hrs after inoculation with additional 50 μg/ml chloramphenicol, 25 μg/ml of kanamycin, and 0.25 mM IPTG when applicable. Samples were drawn at different time points to measure the number of viable cells by serial dilution and plating (in triplicate), LacZ activity (see below), and OD.

LacZ assay

LacZα-CcdB levels were measured in triplicate using a FluoReporter® LacZ/Galactosidase Quantitation kit (Molecular Probes). To permeate cells, 5 μL of chloroform was added to 200 μL culture, which was vortexed vigorously for 30 seconds. Then 10 μL cell lysate was added to 100 μl 1.1 mM 3-carboxyumbelliferyl β-D-galactopyranoside (CUG) solution in each well of a 96-well microplate. The reaction mixture was incubated at room temperature for 30 minutes before fluorescence was measured (excitation: 390 nm; emission: 460 nm). Measured LacZ activity was corrected by subtracting the background fluorescence, measured in control wells containing 100 μL CUG solution but no cell lysate. The corrected value was normalized to culture volume and OD600 and expressed in fluorescence/(mL*OD600). We observed that LacZα-CcdB caused cell death but did not destroy dead cells. Thus both live and dead cells contributed to LacZ activity. Normalizing the measured LacZ activity to total cell mass would thus give a good estimate of the intracellular killer protein concentration.

Mathematical modeling

We model the major kinetic events dictating the circuit function, including cell growth and death (Eq. A1) and production and degradation of the killer protein (Eq. A2) and the AHL signal (Eq.

A3). We assume that, (1) without the circuit, changes in viable cell density (N, ml^{-1}) follow logistic kinetics with an intrinsic per capita growth rate of k (hr^{-1}) and a carrying capacity of N_m (ml^{-1}); (2) for circuit-regulated growth, the cell death rate is proportional to the intracellular concentration of the killer protein (E, nM) with a rate constant of d ($nM^{-1}hr^{-1}$); (3) production rate of E is proportional to AHL concentration (A, nM, assumed to be the same inside and outside the cells) with a rate constant of k_E (hr^{-1}); (4) AHL production rate is proportional to N with a rate constant of ν_A (nM ml hr^{-1}); and (5) degradation of the killer protein and AHL follows first-order kinetics with rate constants of d_E (hr^{-1}) and d_A (hr^{-1}).

$$\frac{dN}{dt} = kN(1 - N/N_m) - dEN \qquad \text{(A1)}$$

$$\frac{dE}{dt} = k_E A - d_E E \qquad \text{(A2)}$$

$$\frac{dA}{dt} = \nu_A N - d_A A \qquad \text{(A3)}$$

In experiments, N and N_m are measured as colony forming units per ml (CFU/ml). The production term ($k_E A$) in Eq. A2 lumps several intermediate steps: activation of LuxR by AHL, binding of activated LuxR to p_{luxI}, and expression of the killer gene. It implies that the LuxR-AHL interaction is limited by the concentration of AHL, which is valid for our system (see main text). Although active LuxR functions as a dimer, we assume the cooperativity of AHL action to be 1, based on measurements for a closely related quorum sensing system (Zhu and Winans, 1999). Assuming a greater cooperativity does not significantly affect model predictions. Eq. A3 implies that the diffusion of AHL is fast compared with other processes and that the production rate of AHL from each viable cell is the same on average.

When $N << N_m$, Eq. A1 reduces to $dN/dt = (k - dE)N$. Then the simplified model will have two steady-state solutions: (N_s=0, E_s=0, A_s=0) and ($N_s = \frac{d_A d_E k}{\nu_A k_E d}$, $E_s = k/d$, $A_s = \frac{d_E k}{k_E d}$), where subscript "s" represents steady state. Linear stability analysis shows that the trivial steady state is

unstable for all positive parameters, and the non-trivial steady state is stable if and only if $d_A+d_E > k$. Since the effective degradation rate constant of the killer protein inside the cell is at least its dilution rate constant caused by cell growth, we have: $d_A+d_E > d_E > k$. Thus, the second steady state is stable for all biologically feasible parameters.

The analytical solution of the non-trivial steady state for the full model (Eqs. A1-A3) can also be solved, in particular, $N_s = \frac{N_m d_A d_E k}{N_m v_A k_E d + d_A d_E k}$. This equation was used to deduce d_A (Table 1). In simulations (carried out using Dynetica (You et al. 2003)), the following parameters were kept at their base values: d=4×10^{-3} nM^{-1}hr^{-1}, k_E =5 hr^{-1}, d_E = 2 hr^{-1}, v_A = 4.8×10^{-7} nM ml hr^{-1}. The others (k, N_m, d_A) were computed from our experimental data (Table 1).

Figure 1.

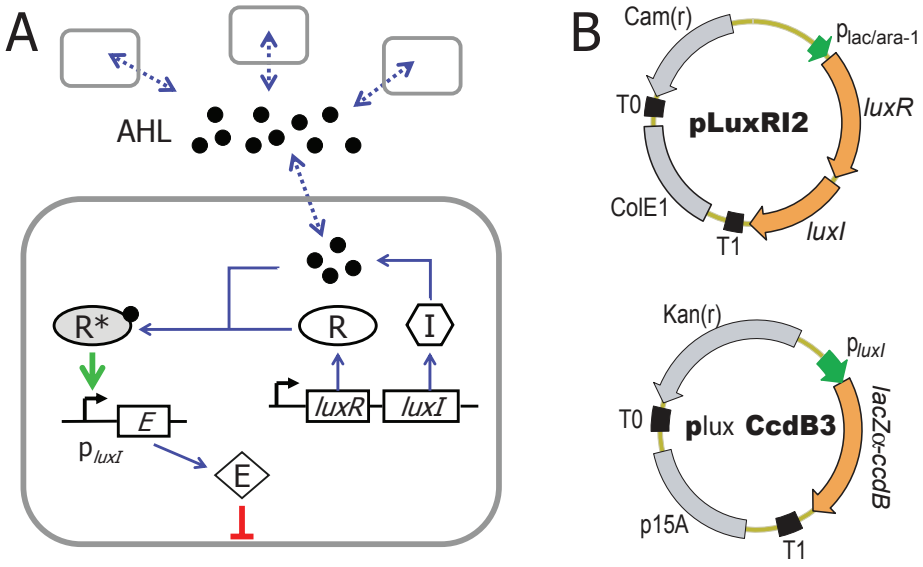


Figure 1. *A population control circuit programs population dynamics by broadcasting, sensing, and regulating the cell density using cell-cell communication and negative feedback.* (a) Schematic diagram of the circuit. E is a "killer" gene. I, R, and R* represent LuxI, LuxR, and active LuxR, respectively. Filled circles represent AHL. (b) Experimental implementation with two plasmids, pLuxRI2 and pLuxCcdB3. LuxI and LuxR are under the control of a synthetic promoter $p_{lac/ara-1}$[32], and the killer gene (lacZα-ccdB) is under the control of p_{luxI}[15]. T0 and T1 are transcription terminators. See text for details.

Figure 2.

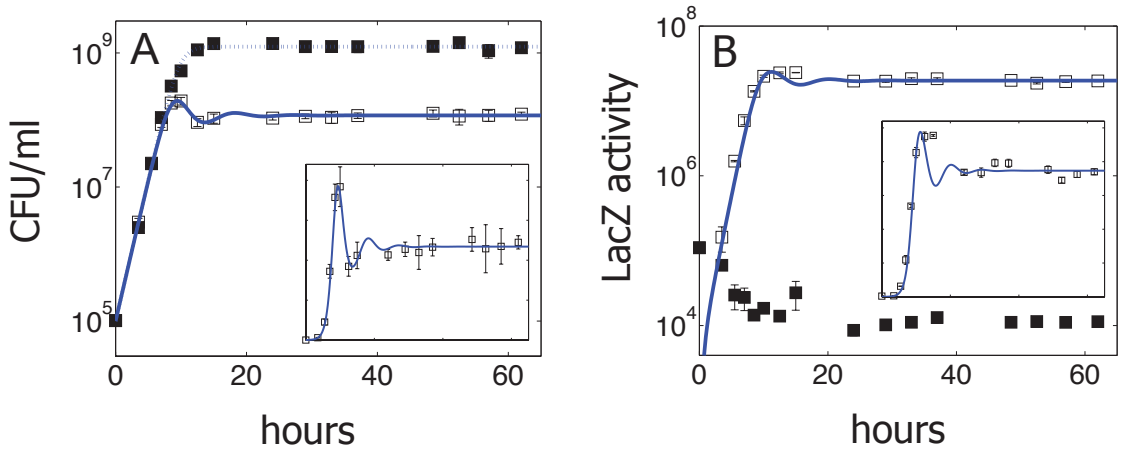


Figure 2. Experimentally measured (a) growth curves and (b) corresponding levels of LacZα-CcdB of Top10F′ cells with the population control circuit OFF (filled squares) and ON (open squares), for pH 7.0. Model predictions are shown in solid (ON) and dotted (OFF) lines, except for the OFF case of LacZ activity, where the killer protein concentration is always zero in the model. The simulated LacZ activity is obtained by multiplying the simulated killer protein concentration (in nM) by a constant factor so that the experiment and simulation are at the same scale. Insets show the growth curves and the LacZ activity in linear scale for the ON case. The similar growth of two cultures at low cell density is an intrinsic feature of the circuit and is not caused by a lag in circuit activation: when the ON culture at steady state was diluted into fresh medium with and without the inducer, the resulting cultures again grew similarly at low density but deviated at high density.

Figure 3.

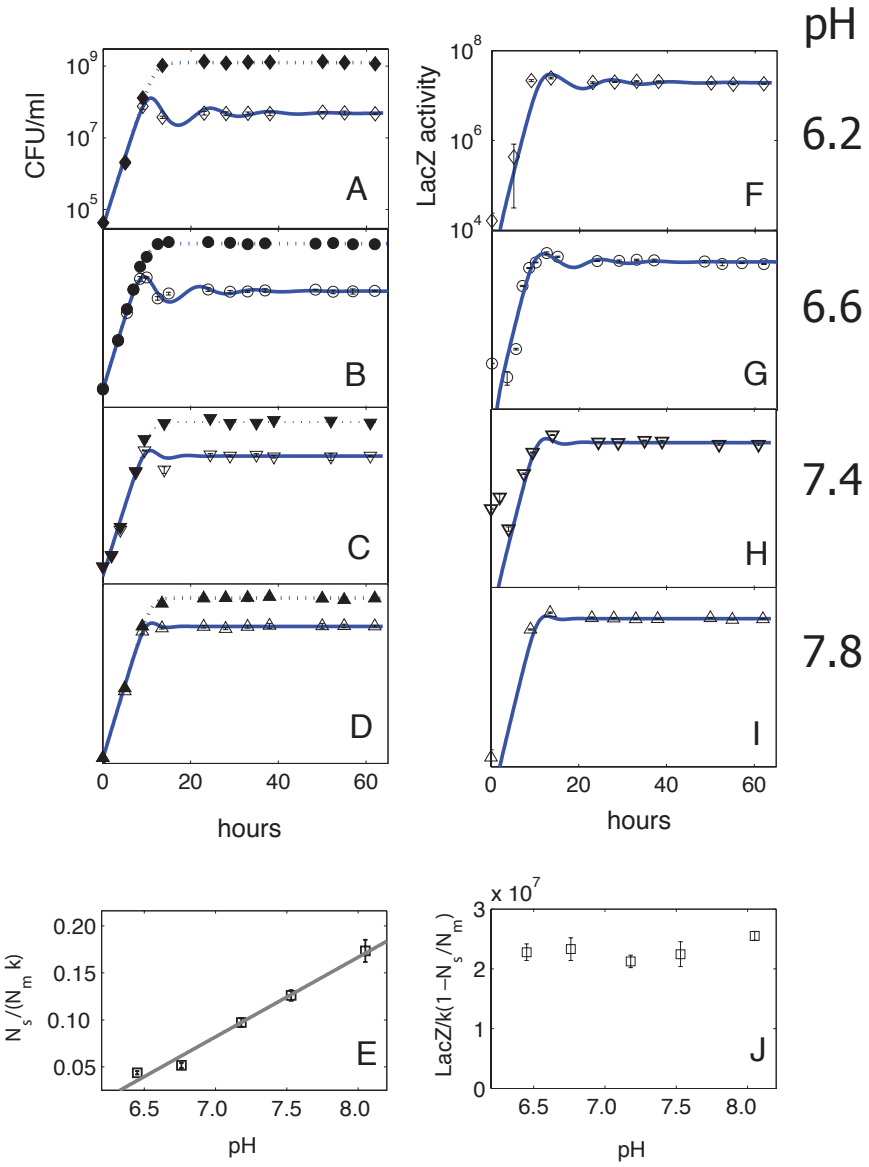


Figure 3. *Effects of pH on circuit behavior.* (a–d) Cell growth with the circuit OFF (filled symbols) and ON (open symbols). (e) Dependence of $N_s/(N_m{*}k)$ on pH. (f–i) Time courses of LacZ activity with the circuit ON. (j) Dependence of $^{(\text{LacZ activity})}/_{k(1-N_s/N_m)}$ on pH. Panels (a–d) have the same scale in both x- and y- axes, as do panels (f–i). Simulated growth curves (ON=solid line, OFF=dotted line) and killer protein time courses (ON=solid line) are shown in (a–d) and (f–i). The killer protein concentration for the OFF cases is always zero in the model. Simulated LacZ activity in (f–i) is obtained by multiplying the simulated killer protein concentration by a constant factor so that the experiment and simulation are at the same scale in each panel. In (e) and (j), steady-state pH values (Table 1) are used along the x-axes. In (e), N_s is normalized with respect to N_m and k to account for minor variations in these variables (Table 1). The dependence of $N_s/(N_m{*}k)$ on pH is nearly linear ($R^2 = 0.98$).

Figure 4.

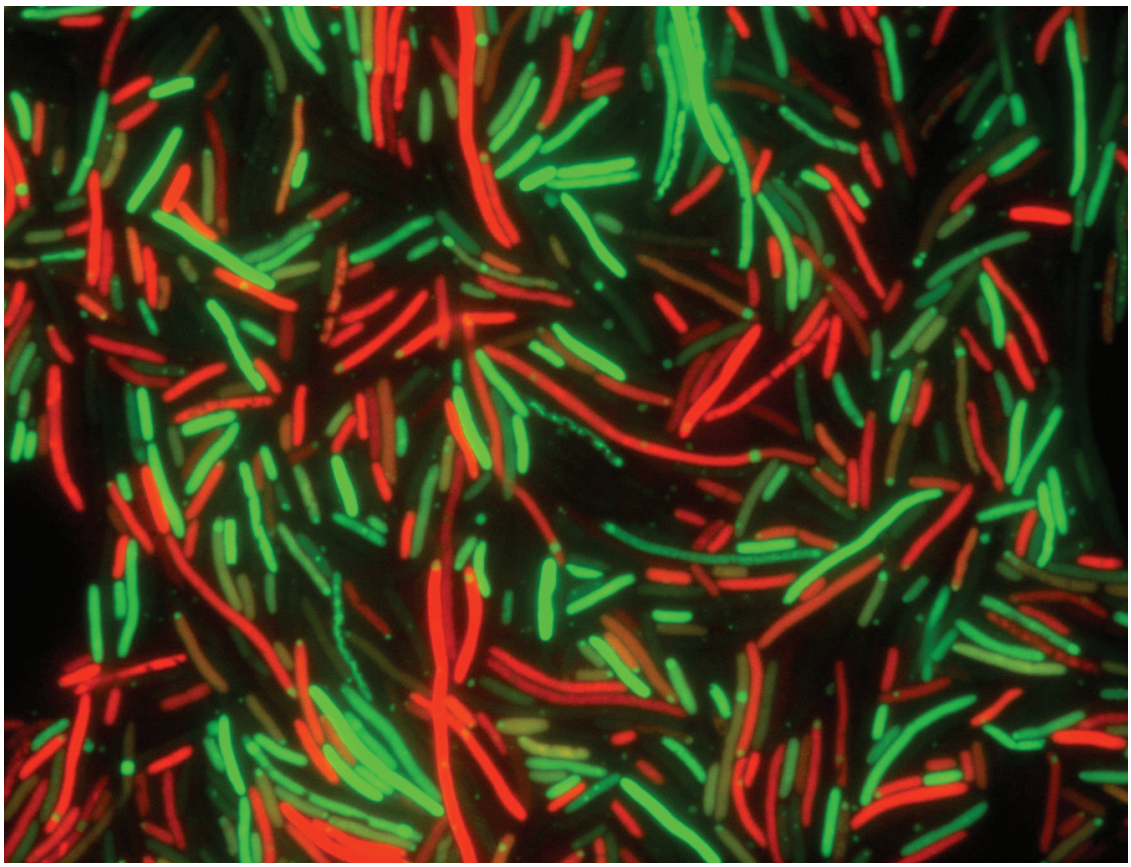


Figure 4. *Living (green) and dead (red) cells regulated by the population control circuit.* These DH5αZ1 cells were more susceptible to the killer gene than TOP10F' cells, due to the absence of the F plasmid addiction system. Noise in circuit regulation produces this individual variability in cell death.

Table 1. Effects of pH on circuit parameters

Medium pH	Steady state culture pH[a]	k [b] (hr^{-1})	$N_m/10^9$ [c] (CFU/ml)	$N_s/10^7$ [d] (CFU/ml)	d_A [e] (hr^{-1})	LacZ activity/10^7 [f] (fluorescence/(ml*OD600))
6.2	6.45	0.885	1.25±0.06	4.86±0.02	0.274	1.94±0.12
6.6	6.76	0.928	1.17±0.05	5.59±0.03	0.304	2.02±0.17
7.0	7.18	0.970	1.24±0.10	11.7±0.6	0.639	1.87±0.09
7.4	7.53	0.897	1.16±0.10	13.1±0.6	0.791	1.79±0.16
7.8	8.05	0.936	1.20±0.07	19.5±1.3	1.19	2.00±0.06

[a] measured at about 50 hrs after growth initiation in ON cultures.

[b] obtained by fitting the exponential phase of growth curves of OFF cultures.

[c] average of the stationary phase cell density of OFF cultures between 28hr and 62 hr.

[d] average of the circuit-ON cell density between 28hr and 62hr.

[e] obtained by solving equation $N_s = \frac{N_m d_A d_E k}{N_m v_A k_E d + d_A d_E k}$ with d_A as the only unknown.

[f] average of LacZ activity of ON cultures between 28hr and 62hr.

References

Ameisen JC (2002) On the origin, evolution, and nature of programmed cell death: a timeline of four billion years. *Cell Death Differ* **9:** 367-393.

Atkinson MR, Savageau MA, Myers JT, Ninfa AJ (2003) Development of genetic circuitry exhibiting toggle switch or oscillatory behavior in Escherichia coli. *Cell* **113:** 597-607.

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

Blake WJ, M KA, Cantor CR, Collins JJ (2003) Noise in eukaryotic gene expression. *Nature* **422:** 633-637.

Bulter T, Lee SG, Woirl WWC, Fung E, Connor MR, Liao JC (2004) Design of artificial cell-cell communication using gene and metabolic networks. *Proc Natl Acad Sci U S A* **101:** 2299-2304.

Chen W, Kallio PT, Bailey JE (1993) Construction and characterization of a novel cross-regulation system for regulating cloned gene expression in *Escherichia coli*. *Gene* **130:** 15-22.

Dong YH, Wang LH, Xu JL, Zhang HB, Zhang XF, Zhang LH (2001) Quenching quorum-sensing-dependent bacterial infection by an N-acyl homoserine lactonase. *Nature* **411:** 813-817.

Egland KA, Greenberg EP (1999) Quorum sensing in *Vibrio fischeri*: elements of the luxl promoter. *Mol Microbiol* **31:** 1197-1204.

Egland KA, Greenberg EP (2001) Quorum sensing in *Vibrio fischeri*: analysis of the LuxR DNA binding region by alanine-scanning mutagenesis. *J Bacteriol* **183:** 382-386.

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

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

Engelberg-Kulka H, Glaser G (1999) Addiction modules and programmed cell death and

antideath in bacterial cultures. *Annu Rev Microbiol* **53:** 43-70.

Farmer WR, Liao JC (2000) Improving lycopene production in *Escherichia coli* by engineering metabolic control. *Nature biotechnology* **18:** 533-537.

Fuqua WC, Winans SC, Greenberg EP (1994) Quorum sensing in bacteria: the LuxR-LuxI family of cell density-responsive transcriptional regulators. *J Bacteriol* **176:** 269-275.

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

Gerchman Y, Weiss R (2004) Teaching bacteria a new language. *Proc Natl Acad Sci U S A* **101:** 2221-2222.

Hasty J, McMillen D, Collins JJ (2002) Engineered gene circuits. *Nature* **420:** 224-230.

Isaacs FJ, Hasty J, Cantor CR, Collins JJ (2003) Prediction and measurement of an autoregulatory genetic module. *Proc Natl Acad Sci U S A* **100:** 7714-7719.

Leadbetter JR, Greenberg EP (2000) Metabolism of acyl-homoserine lactone quorum-sensing signals by *Variovorax paradoxus*. *J Bacteriol* **182:** 6921-6926.

Lewis K (2000) Programmed death in bacteria. *Microbiol Mol Biol Rev* **64:** 503-514.

Lutz R, Bujard H (1997) Independent and tight regulation of transcriptional units in *Escherichia coli* via the LacR/O, the TetR/O and AraC/I1-I2 regulatory elements. *Nucleic Acids Res* **25:** 1203-1210.

Miller MB, Bassler BL (2001) Quorum sensing in bacteria. *Annu Rev Microbiol* **55:** 165-199.

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

Schaefer AL, Hanzelka BL, Parsek MR, Greenberg EP (2000) Detection, purification, and structural elucidation of the acylhomoserine lactone inducer of *Vibrio fischeri* luminescence and other related molecules. In *Bioluminescence and Chemiluminescence, Pt C*, Vol. 305, pp 288-301.

Steinmoen H, Knutsen E, Havarstein LS (2002) Induction of natural competence in *Streptococcus pneumoniae* triggers lysis and DNA release from a subfraction of the cell population. *Proc Natl Acad Sci U S A* **99:** 7681-7686.

Wall ME, Hlavacek WS, Savageau MA (2004) Design of gene circuits: lessons from bacteria. *Nat Rev Genet* **5:** 34-42.

Weiss R, Basu S, Hooshangi S, Kalmbach A, Karig D, Mehreja R, Netravali I (2003) Genetic circuit building blocks for cellular computation, communications, and signal processing. *Natural Computing* **2:** 47-84.

Weiss R, Homsy GE, Knight Jr. T (1999) In *Dimacs Workshop on Evolution as Computation* pp 275-295. Springer, Princeton.

Weiss R, Knight Jr. T (2000) In *6th International Workshop on DNA-Based Computers*, Rozenberg G (ed) pp 1-16, Leiden, The Netherlands.

Yokobayashi Y, Weiss R, Arnold FH (2002) Directed evolution of a genetic circuit. *Proc Natl Acad Sci U S A* **99:** 16587-16591.

You L, Hoonlor A, Yin J (2003) Modeling biological systems using Dynetica--a simulator of dynamic networks. *Bioinformatics (Oxford, England)* **19:** 435-436.

Zhu J, Winans SC (1999) Autoinducer binding by the quorum-sensing regulator TraR increases affinity for target promoters in vitro and decreases TraR turnover rates in whole cells. *Proc Natl Acad Sci U S A* **96:** 4832-4837.

Appendix A: Promoter library data

This appendix contains the promoter sequences characterized in Chapters 1 and 2, the luminescence data of Chapter 2. See Chapter 2 methods for experimental protocols and Chapter 2 Table S1 for the definitions of the 48 duplex unit sequences.

Appendix A1. *Promoter library sequence data.*

clone	Promoter sequence (between cloning sites XhoI and BamHI)	*distal*	*core*	*proximal*
A1	TACAACGTCGTGTTAGCTGCCTTTTAGCAATTTTATCCATAGACTATTTTACCTCTGGCGGTGATAATTGAATACATCTGGCGGTGATAAGGCGTTACCCAAC	ara2	con2	con1
A11	TACAACGTCGTGTTAGCTGCCTTTCGTCTTCAATAATTCTTGACTATTTTACCTCTGGCGGTGATAATTCTTGCGACAAACAATAGGTAAGGCGTTACCCAAC	con0	con2	lux2
A12	TACAACGTCGTGTTAGCTGCCTTTTAGCAATTTTATCCATAGACTTGTGAGCGGATAACAATTATAATTCGTGCAATTTTTAAACCTGTAGGATCGTACAGGT	ara2	lac3	lux1
A2	TACAACGTCGTGTTAGCTGCTCCCTATCAGTGATAGAGATTGACATTTATCCCTTGCGGTGATAGATTTGAATACATCTGGCGGTGATAAGGCGTTACCCAAC	tet1	con4	con1
A3	TACAACGTCGTGTTAGCTGCCTTTCGTCTTCAATAATTCTTGACATTGTGAGCGGATAACAAGATACTTAACTCTATCAATGATAGAGTGTCAACAAAAAAAC	con0	lac1	tet2
A4	TACAACGTCGTGTTAGCTGCAATTGTGAGCGGATAACAATTGACATAGCGGATACTTCCTGATATAATTCGTGCAATTTTTAAACCTGTAGGATCGTACAGGT	lac1	ara2	lux1
A5	TACAACGTCGTGTTAGCTCAATTGTGAGCGGATAACAATTGACTTTTATCCCTTGCGGTGATATAATTGAATACATCTGGCGGTGATAAGGCGTTACCCAAC	lac1	con3	con1
A6	TACAACGTCATTTCACTTTTCTCTATCACTGATAGGGAGTGGTCTTTTATCCCTTCGCGGTGATATAATTCGTGCAATTTTTATATCACCGCCAGGGGTACAAC	tet2	con3	con3
A7	TACAACGTCGTGTTAGCTGCCTTTCGTCTTCAATAATTCTTGACTTTTATCCCTTGCGGTGATATAATTCTTGCGACAAACAATAGGTAAGGCGTTACCCAAC	con0	con3	lux2
A8	TACAACGTCGTGTTAGCTGCCTTTTAGCAATTTTATCCATAGACTATTTTACCTCTGGCGGTGATAATTCTTGCGACAAACAATAGGTAAGGCGTTACCCAAC	ara2	con2	lux2
A9	TACAACGTCGTGTTAGCTGCCTTTCGTCTTCAATAATTCTTGACAAATAACTCTATCAATGATAGAGTTCGTGCATAGCATTTTTATCCATACGTTACCCAAC	con0	tet2	ara2
A13	TATCACCGCCAGAGGTAAAATAGTCAACACGCACGGTGTTAGACTTGTGAGCGGATAACAATTATAATTGAATACAGCTGGCGGTGATAAGGCGTTACCCAAC	con3	lac3	con1
A23	TACAACGTCGTGTTAGTTGCCTTTCGTCTTCAATAATTCTTGACATTTATGCTTCCGGCTCGTATAATTCGTGCAATTTTTAAACCTGTAGGATCGTACAGGT	con0	con0	lux1
A14	TACAACGTCGTGTTAGCTGTATCACCGCCAGAGGTAAGATTGACAATCAATGTGGATTTTCTGATACTTCTTGCGACAAACAATAGGTAAGGCGTTACCCAAC	con4	ara1	lux2
A16	TACAACGTCGTGTTAGCTGTATCACCGCCAGAGGTAAGATTGACACTCTATCATTGATAGAGTTATTTTTCGTGGTCCATATTGCATCAGACATTGTACCCAAC	con4	tet3	ara1
A17	TACAACGTCATTTCACTTTTCTATCACTGATAGGGAGTGGTCATTTATGCTTCCGGCTCGTATAATTCGTGCAATTTTTAAACCTGTAGGATCGTACAGGT	tet2	con0	lux1
A18	TACAACGTCGTGTTAAATTGTGAGCGGATAACAATTTAGTTGACTATTTTACCTCTGGCGGTGATAATTCCACCCCTTCAGTGATAGAGAGCGTTACCCAAC	lac2	con2	tet1
A19	TACAACGTCGTGTTAAATTAGTGAGCGGATAACAATTTAGTTGACTATTTTACCTCTGGCGGTGATAATTCCACCCCTTCAGTGATAGAGAGCGTTACCCAAC	lac2	con2	tet1
A20	TACAACGTCGTGTTAGCTGTATCACCGCCAGAGGTAAGATTGACACCTGTAGGATCGTACAGGTATAATTCTTGCGACAAACAATAGGTAAGGCGTTACCCAAC	con4	lux2	lux2
A25	TATCACCGCCAGAGGTAAAATATTCAACACGCACGGTGTTAGACACTCTATCATTGATAGAGTTATTTTTCGTGCAATTTTTAAACCTGTAGGATCGTACAGGT	con3	tet3	lux1
A34	TATCACCGCCAGAGTAAAATAGTCAACACGCACGGTGTTAGGCAAATAACTCTATCAATGATAGAGTTAGATTCAATTAGTGAGCGGATAACAATTTCACACA	con2	tet2	lac2
A35	TACAACGTCGTGTTAGCTGCTCCCTATAGTGATAGAGATTGACTTGTGAGCGGATAACAATTATAATTGAATACATCTGGCGGTGATAAGGCGTTACCCAAC	tet1	lac3	con1
A26	TACAACGTCGTGTTAGCTGTATCACCGCCAGAGGTAAGATTCGACACTCTATCATTGATAGAGTTATTTTTGAATACCTCTGGCGGTGATAAGGCGTTACCCAAC	con4	tet3	con2
A27	TACAACGTCGTGTTAGCTGTATCACCGCCAGAGGTAAGATTGACATAGCGGATACTTCCTGATATAATTCGTGCAATTTTTAAAATTAAAGGCGTTACCCAAC	con4	ara2	con0
A28	TACAACGTCGTGTTAAATTGTGAGCGGATAACAATTTAGTTGACATCCCTATCAGTGATAGAGATACTTTGTGGAATTGTGAGCGGATAACAATTTCACACAG	lac2	tet1	lac1
A29	TACAACGTCGTGTTAGCTGTATCACCGCCAGAGGTAAGATTGACACCTGTAGGATCGTACAGGTATAATTCGTGCAATTTTTATATCACCGCCAGGGGTACAAC	con4	lux2	con3
A30	TACAACGTCGTGTTAGCTGTATCACCGCCAGAGGTAAGATTGACAATCAATGTGGATTTTCTGATACTTTGTGGAATTGTGAGCGGATAACAATTTCACACAG	con4	ara1	lac1

clone	Promoter sequence (between cloning sites XhoI and BamHI)	*distal*	*core*	*proximal*
A32	TACAACGTCGTGTTAAATTGTGAGCGGATAACAATTTAGTTGACATAGCGGATACTTCCTGATATAATTGAATACATCTGGCGGTGATAAGGCGTTACCCAAC	lac2	ara2	con1
A33	TATCACCGCCAGAGGTAAAATAGTCAACACGCACGGTGTTAGGCTTGTGAGCGGATAACAATTATAATTCGTGCATTTTTAAACCTGTAGGATCGTACAGGT	con2	lac3	lux1
A37	TACAACGTCGTGTTAGCTGTATCACCGCCAGAGGTAAGATTGACATCCCTATCAGTGATAGAGATACTTCGTTATCACCGCCAGGGGTAAGGCGTTACCCAAC	con4	tet1	con4
A46	TACAACGCATTTCACTTTTCTCTATCACTGATAGGGAGTGGTCTATTTTACCTCTGGCGGTGATAATTCGTGCAATTTTTATATCACCGCCAGGGGTACAAC	tet2	con2	con3
A47	TACAACGTCGTGTTAGCTGTATCACCGCCAGAGGTAAGATTGACATAGCGGATACTTCCTGATATAATTGAATACCTCTGGCGGTGATAAGCGTTACCCAAC	con4	ara2	con2
A48	TACAACGTCGTGTTAAATTGTGAGCGGATAACAATTTAGTTGACATTTATGCTTCCGGCTCGTATAATTAGATTCAATTGTGAGCGGATAACAATTTCACACA	lac2	con0	lac2
A38	TACAACGTCGTGTTAGCTGCAATTGTGAGCGGATAACAATTGACATTTATGCTTCCGGCTCGTATAATTCGTGCAATTTAAATGTGAGCGAGTAACAACCAAC	lac1	con0	lac4
A39	TACAACGTCGTGTTAGCTGCCTTTCGTCTTCAATAATTCTTGACATAGCGGATACTTCCTGATATAATTCGTGGTCCATATTGCATCAGACATTGTACCCAAC	con0	ara2	ara1
A40	TACAACGTCGTGTTAGCTGTATCACCGCCAGAGGTAAGATTGACATTTATCCCTTGCGGTGATAGATTTCGTTATCACCGCCAGGGGTAAGGCGTTACCCAAC	con4	con4	con4
A41	TATCACCGCCAGAGGTAAAATAGTCAACACGCACGGTGTTAGGCATTTATCCCTTGCGGTGATAGATTTCGTGCATAGCATTTTTATCCATACGTTACCCAAC	con2	con4	ara2
A42	TACAACGTCGTGTTAGCTGCTCCCTATCAGTGATAGAGATTGACATTTATGCTTCCGGCTCGTATAATTGAATACATCTGGCGGTGATAAGGCGTTACCCAAC	tet1	con0	con1
A43	TACAACGTCGTGTTAGCTGCCTTTTAGCAATTTTATCCATAGACATAGCGGATACTTCCTGATATAATTCGTGCAATTTTTAAACCTGTAGGATCGTACAGGT	ara2	ara2	lux1
A44	TACAACGTCGTGTTAGCTGTATCACCGCCAGAGGTAAGATTGACTATTTTACCTCTGGCGGTGATAATTTGTGGAATTGTGAGCGGATAACAATTTCACACAG	con4	con2	lac1
A45	TACAACGTCGTGTTAGCTGCAATTGTGAGCGGATAACAATTGACACCTGTAGGATCGTACAGGTATAATTCGTGCAATTTTTAAAATTAAAGGCGTTACCCAAC	lac1	lux2	con0
A49	TACAACGTCATTTCACTTTTCTCTATCACTGATAGGGAGTGGTCATTTATGCTTCCGGCTCGTATAATTCGTTATCACCGCCAGGGGTAAGGCGTTACCCAAC	tet2	con0	con4
A59	TACAACGTCGTGTTAGCTGCCTTTCGTCTTCAATAATTCTTGACTTGTGAGCGGATAACAATGATACTTCGTGCAATTTTTAAAATTAAAGGCGTTACCCAAC	con0	lac2	con0
A60	TACAACGTCGTGTTAGCTGCCTTTCGTCTTCAATAATTCTTGACAAATAACTCTATCAATGATAGAGTTAGATTCAATTGTGAGCGGATAACAATTTCACACA	con0	tet2	lac2
A52	TACAACGTCGTGTTAGCTGCAATTGTGAGCGGATAACAATTGACAATCAATGTGGATTTTCTGATACTTGAATACATCTGGCGGTGAATAAGGCGTTACCCAAC	lac1	ara1	con1
A53	TACAACGTCGTGTTAGCTGCCTTTCGTCTTCAATAATTCTTGACACCTGTAGGATCGTACAGGTATAATTAGATTCAATTGTGAGCGGATAACAATTTCACACA	con0	lux2	lac2
A54	TACAACGTCGTGTTAGCTGCCTTTTAGCAATTTTATCCATAGACATTTATCCCTTGCGGTGATAGATTTAGATTCAATTGTGAGCGGATAACAATTTCACACA	ara2	con4	lac2
A56	TACAACGTCATTTCACTTTTCTCTATCACTGATAGGGAGTGGTCATTTATCCCTTGCGGTGATAGATTTCGTGCAATTTTTAAAATTAAAGGCGTTACCCAAC	tet2	con4	con0
A57	TACAACGTCGTGTTAGCTGCTCCCTATCAGTGATAGAGATTGACATTTATGCTTCCGGCTCGTATAATTCCTGTAGGATCGTACAGGTAAGGCGTTACCCAAC	tet1	con0	lux3
A61	TACAACGTCGTGTTAGCTGCCTTTCGTCTTCAATAATTCTTGACTTTTATCCCTTGCGGTGATATAATTCGTGCAATTTTTATATCACCGCCAGGGGTACAAC	con0	con3	con3
A70	TACAACGTCGTGTTAGCTGTATCACCGCCAGAGGTAAGATTGACTTTTATCCCTTGCGGTGATATAATTCGTGCAATTTTTATATCACCGCCAGGGGTACAAC	con4	con3	con3
A72	ACATAGCATTTTTATCCATAACCTGTAGGATCGTACAGGTTTACATTTATCCCTTGCGGTGATAGATTTCCTGTAGGATCGTACAGGTAAGGCGTTACCCAAC	lux3	con4	lux3
A62	TACAACGTCGTGTTAGCTGTATCACCGCCAGAGGTAAGATTGACTTTTATCCCTTGCGGTGATATAATTCGTTATCACCGCCAGGGGTAAGGCGTTACCCAAC	con4	con3	con4
A63	TACAACGTCATTTCACTTTTCTCTATCACTGATAGGGAGTGGTCAATCAATGTGGATTTTCTGATACTTCGTGCAATTTTTAAAATTAAAGGCGTTACCCAC	tet2	ara1	con0
A64	TACAACGTCGTGTTAGCTGCCTTTTAGCAATTTTATCCATAGACACCTGTAGGATCGTACAGGTATAATTGAATACATCTGGCGGTGATAAGGCGTTACCCAAC	ara2	lux2	con1
A65	TACAACGTCGTGTTAGCTGCCTTTCGTCTTCAATAATTCTTGACATAAATACCACTGGCGGTGATACTTCGTGCAATTTTTATATCACCGCCAGGGGTACAAC	con0	con1	con3
A66	TACAACGTCGTGTTAAATTGTGAGCGGATAACAATTTAGTTGACACCTGTAGGATCGTACAGGTATAATTCGTGCAATTTTTATATCACCGCCAGGGGTACAAC	lac2	lux2	con3
A67	TACAACGTCGTGTTAGCTGCCTTTCGTCTTCAATAATTCTTGACATAGCGGATACTTCCTGATATAATTCGTGCATAGCATTTTTATCCATACGTTACCCAAC	con0	ara2	ara2
A68	TACAACGTCGTGTTAGCTGCCTTTTAGCAATTTTATCCATAGACATAGCGGATACTTCCTGATATAATTGAATACCTCTGGCGGTGATAAGGCGTTACCCAAC	ara2	ara2	con2
A69	TACAACGTCGTGTTAGCTGCCTTTCGTCTTCAATAATTCTTGACTTTTATCCCTTGCGGTGATATAATTAGATTCAATTGTGAGCGGATAACAATTTCACACA	con0	con3	lac2

clone	Promoter sequence (between cloning sites XhoI and BamHI)	*distal*	*core*	*proximal*
A82	TACAACGTCGTGTTAGCTGTATCACCGCCAGAGGTAAGATTGACTATTTTACCTCTGGCGGTGATAATTCGTGCAATTTTTAAAATTAAAGGCGTTACCCAAC	con4	con2	con0
A83	TACAACGTCGTGTTAGCTGCTCCCTATCAGTGTAGAGATTGACTATTTTACCTCTGGCGGTGATAATTCCACCCCTTCAGTGATAGAGAGCGTTACCCAAC	tet1	con2	tet1
A76	TACAACGTCATTTCACTTTTCTCTATCACTGATAGGGAGTGGTCATTTATGCTTCCGGCTCGTATAATTCGTGCAATTTTTAAAATTAAAGGCGTTACCCAAC	tet2	con0	con0
A77	GTAACAAAAGTGTCTATAATCACGGCAGAAAAGTCCACATTGACTTTTATCCCTTGCGGTGATATAATTGAATACATCTGGCGGTGATAAGGCGTTACCCAAC	ara3	con3	con1
A78	TACAACGTCGTGTTAGCTGCCTTTTAGCAATTTTATCCATAGACATTTATCCCTTGCGGTGATAGATTTCCACCCCTATCAGTGATAGAGAGCGTTACCCAAC	ara2	con4	tet1
A79	TACAACGTCGTGTTAGCTGCCTTTTAGCAATTTTATCCATAGACAATCAATGTGGATTTTCTGATACTTCTTGCGACAAACAATAGGTAAGGCGTTACCCAAC	ara2	ara1	lux2
A80	TATCACCGCCAGAGGTAAAATAGTCAACACGCACGGTGTTAGGATCCCTATCAGTGATAGAGATACTTCGTGCAATTTTTAAAATTAAAGGCGTTACCCAAC	con2	tet1	con0
A81	TACAACGTCGTGTTAAATTGTGAGCGGATAACAATTTAGTTGACATAAATACCACTGGCGGTGATACTTCGTGCAATTTTTAAACCTGTAGGATCGTACAGGT	lac2	con1	lux1
A85	TACAACGTCGTGTTAGCTGCTCCCTATCAGTGATAGAGATTGACTTTTATCCCTTGCGGTGATATAATTGAATACATCTGGCGGTGATAAGGCGTTACCCAAC	tet1	con3	con1
A95	TACAACGTCGTGTTAAATTGTGAGCGGATAACAATTTAGTTGACTTTTATCCCTTGCGGTGATATAATTAGATTCAATTGTGAGCGGATAACAATTTCACACA	lac2	con3	lac2
A96	TCTCACCGCCAGAGGTAAAATAGTCAACACGCACGGTGTTAGACACTCTATCATTCGATAGAGTTATTTTTCCACCCCTATCAGTGATAGAGAGCGTTACCCAAC	con3	tet3	tet1
A86	TACAACGTCGTGTTAGCTGCCTTTCGTCTTCAATAATTCTTGACATAGCGGATACTTCCTGATATAATTCGTGCAATTTTTAAACCTGTAGGATCGTACAGGT	con0	ara2	lux1
A87	TACAACGTCGTGTTAGCTGCCTTTCGTCTTCAATAATTCTTGACAATCAATGTGGATTTTCTGATACTTCGTGCAATTTAAATGTGAGCGAGTAACAACCAAC	con0	ara1	lac4
A88	TACAACGTCGTGTTAGCTGCCTTTTAGCAATTTTATCCATAGACATTTATGCTTCCGGCTCGTATAATTCGTGGTCCATATTGCATCAGACATTGTACCCAAC	ara2	con0	ara1
A89	TACAACGTCGTGTTAGCTGTATCACCGCCAGAGGTAAGATTGACACCTGTAGGATCGTACAGGTATAATTAACTCTATCAATGATAGAGTGTCAACAAAAAAAC	con4	lux2	tet2
A90	TACAACGTCGTGTTAGCTGCCTTTCGTCTTCAATAATTCTTGACATCCCTATCAGTGATAGAGATACTTTGTGGAATTGTGAGCGGATAACATTTCACACAG	con0	tet1	lac1
B10	TATAACGTCGTATTAGCTGCCTTTCGTCTTCAATAATTCTTGACATCCCTATCAGTGATAGAGATACTTCGTGCAATTTTTATATCACCGCCAGGGGTACAAC	con0	tet1	con3
B11	TACAACGTCGTGTTAGCTGCCTTTCGTCTTCAATAATTCTTGACTATTTTACCTCTGGCGGTGATAATTTGTGGAATTGTGAGCGGATAACATTCACACAG	con0	con2	lac1
B4	TACAACGTCGTGTTAGTTGCAATTGTGAGCGGATAACAATTGACTTGTGAGCGGATAACAATGATACTTCGTGCAATTTTTAAAATTAAAGGCGTTACCCAAC	lac1	lac2	con0
B5	TACAACGTCGTGTTAGCTGCCTTTTAGCAATTTTATCCATAGACAAATAACTCTATCAATGATAGAGTTAGATTCAATTGTGAGCGGATAACAATTTCACACA	ara2	tet2	lac2
B13	TATCACCGCCAGAGGTAAAATAGTCAACACGCACGGTGTTAGACTTGTGAGCGGATAACAATGATACTTCGTGCATAGCATTTTTATCCATACGTTACCCAAC	con3	lac2	ara2
B22	TACAACGTCGTGTTAGCTGTATCACCGCCAGAGGTAAGATTGACATTTATCCCTTGCGGTGATAGATTTCCACCCCTATCAGTGATAGAGAGCGTTACCCAAC	con4	con4	tet1
B23	TACAACGTCGTGTTAGCTGCTCCCTATCAGTGATAGAGATTGACAATCAATGTGGATTTTCTGATACTTGATTCAATTGTGAGCGGATAACAATTTCACACAG	tet1	ara1	lac3
B14	TACAACGTCGTGTTAGCTGTATCACCAGATAACCATCTGCGGTGATAAATTATCTCTGGCGGTGTTGACTTGTGAGCGGATAACAATGATACTTCGTG-CAATTTTTAAAATTAAAGGCGTTACCCAAC	con1	lac2	con0
B16	TACAACGTCATTTCACTTTTCTCTATCACTGATAGGGAGTGGTCATAGCGGATACTTCCTGATATAATTCGTGCAATTTTTAAACCTGTAGGATCGTACAGGT	tet2	ara2	lux1
B17	TACAACGTCGTGTTAGCTGTATCACCGCCAGAGGTAAGATTGACTTGTGAGCGCTCACAATTTATAATTGAATACATCTGGCGGTGATAAGGCGTTACCCAAC	con4	lac4	con1
B18	TACAACGTCGTGTTAGCTGTATCACCGCCAGAGGTAAGATTGACATAAATACCACTGGCGGTGATACTTCGTTATCACCGCCAGGGGTAAGGCGTTACCCAAC	con4	con1	con4
B19	GTAACAAAAGTGTCTATAATCACGGCAGAAAAGTCCACATTGACAAATAACTCTATCAATGATAGAGTTCCACCCCTATCAGTGATAGAGAGCGTTACCCAAC	ara3	tet2	tet1
B20	TACAACGTCGTGTGAGCTGCCTTTTAGCAATTTTATCCATAGACTTTTATCCCTTGCGGTGATATAATTCCCCCCTATCAGTGATAGAGAGCGTTACCCAAC	ara2	con3	tet1
B21	TACCACGCGTGTTAGCTGCCTTTTAGCAATTTTATCCATAGACACCTGTAGGATCGTACAGGTATAATTCGTGCAATTTTTATATCACCGCCAGGGGTACAAC	ara2	lux2	con3
B25	TACAACGTCGTGTTAGCTGCCTTTCGTCTTCAATAATTCTTGACATTTATCCCTTGCGGTGATAGATTTCGTTATCACCGCCAGGGGTAAGGCGTTACCCAAC	con0	con4	con4
B35	TACAACGTCGTGTTAGCTGCTCCCTATCAGTGATAGAGATTGACACCTGTAGGATCGTACAGGTATAATTCGTGCATAGCATTTTTATCCATACGTTACCCAAC	tet1	lux2	ara2

clone	Promoter sequence (between cloning sites XhoI and BamHI)	*distal*	*core*	*proximal*
B26	TACAACGTCGTGTTAAATTGTGAGCGGATAACAATTTAGTTGACATTTATGCTTCCGGCTCGTATAATTCGTGCAATTTTTATATCACCGCCAGGGGTACAAC	lac2	con0	con3
B27	TACAACGTCGTGTTAGCTGTATCACCGCCAGAGGTAAGATTGACACCTGTAGGATCGTACAGGTATAATTCGTGGTCCATATTGCATCAGACATTGTACCCAAC	con4	lux2	ara1
B29	TACAACGTCGTGTTAGCTGTATCACCGCCAGAGGTAAGATTGACTATTTTACCTCTGGCGGTGATAATTCGTGCAATTTTTAAAATTAAAGGCGTTACCCAAC	con4	con2	con0
B30	TACAACGTCGTGTTAGCTGTATCACCGCCAGAGGTAAGATTGACTATTTTACCTCTGGCGGTGATAATTCGTGCAATTTTTAACTGTAGGATCGTACAGGT	con4	con2	lux1
B31	TACAACGTCGTGTTAGCTGCCTTTTAGCAATTTTATCCATAGACATAGCGGATACTTCCTGATATAATTAGATTCAATTGTGAGCGGATAACAATTTCACACA	ara2	ara2	lac2
B33	TACAACGTCGTGTTAGCTGCCTTTTAGCAATTTTATCCATAGACAATCAATGTGGATTTTCTGATACTTGAATACATCTGGCGGTGATAAGGCGTTACCCAAC	ara2	ara1	con1
B37	TACAACGTCGTGTTAGCTGCCTTTCGTCTTCAATAATTCTTGACATAAATACCACTGGCGGTGATACTTCGTGCAATTTTTATATCACCGCCAGGGGTACAAC	con0	con1	con3
B46	TACAACGTCGTGTTAGCTGTATCACCGCCAGAGGTAAGATTGACATCCCTATCAGTGATAGAGATACTTGAATACATCTGGCGGTGATAAGGCGTTACCCAAC	con4	tet1	con1
B47	TACAACGTCGTGTTAGCTGCCTTTTAGCAATTTTATCCATAGACTTGTGAGCGGATAACAATTATAATTCGTGCAATTTTTAAACCTGTAGGATCGTACAGGT	ara2	lac3	lux1
B48	TACAACGTCGTGTTAGCTGCCTTTTAGCAATTTATCCATAGACTTGTGAGCGATAACAATTATAATTTGTGGAATTGTGAGCGGATAACAATTTCACACAG	ara2	lac3	lac1
B39	TACAACGTCATTTCACTTTTCTCTATCACTGATAGGGAGTGGTCATAAATACCACTGGCGGTGATACTTAACTCTATCAATGATAGAGTGTCAACAAAAAAAC	tet2	con1	tet2
B40	TACAACGTCGTGTTAGCTGCAATTGTGAGCGGATAACAATTGACATAGCGGATACTTCCTGATATAATTCGTTATCACCGCCAGGGGTAAGGCGTTACCAAC	lac1	ara2	con4
B42	TATCACCGCCAGAGGTAAAATAGTCAACACGCACGGTGTTAGGCATAGCGGATACTTCCTGATATAATTCGTGCAATTTTTAAAATTAAAGGCGTTACCCAAC	con2	ara2	con0
B43	TACAACGTCGTGTTAGCTGCCTTTCGTCTTCAATAATTCTTGACTTTTATCCCTTGCGGTGATATAATTGAATACATCTGGCGGTGATAAGGCGTTACCCAAC	con0	con3	con1
B45	TACAACGTCGTGTTAGCTGCCTTTCGTCTTCAATAATTCTTGACACCTGTAGGATCGTACAGGTATAATTCGTGGTCCATATTGCATCAGACATTGTACCCAAC	con0	lux2	ara1
B49	TAAACGGCGTGTTAGCTGTATCACCGCCAGAGGTAAGATTGACTATTTTACCTCTGGCGGTGATAATTGAATACATCTGGCGGTGATAAGGCGTTACCCAAC	con4	con2	con1
B60	TACAACGTCGTGTTAGCTGCCTTTTAGCAATTTTATCCATAGACAATCAATGTGGATTTTCTGATACTTCGTGCAATTTTTAAAATTAAAGGCGTTACCCAAC	ara2	ara1	con0
B50	TACAACGTCGTGTTAGCTGCTCCCTATCAGTGATAGAGATTGACTATTTTACCTCTGGCGGTGATAATTGAATACCTCTGGCGGTGATAAGGCGTTACCCAAC	tet1	con2	con2
B51	TACAACGTCGTGTTAGCTGCCTTTTAGCAATTTTATCCATAGACTTGTGAGCGGATAACAATTATAATTGAATACATCTGGCGGTGATAAGGCGTTACCCAAC	ara2	lac3	con1
B52	TACAACGTCGTGTTAGCTGTATCACCGCCAGAGGTAAGATTGACATTTATGCTTCCGGCTCGTATAATTCGTGCAATTTAAATGTGAGCGAGTAACAACCAAC	con4	con0	lac4
B56	TACAACGTCGTGTTAGCTGCCTTTCGTCTTCAATAATTCTTGACTTTTATCCCTTGCGGTGATATAATTAGATTCAATTGTGAGCGGATAACAATTTCACACA	con0	con3	lac2
B61	TACAACGTCGTGTTAGCTGCCTTTTAGCAATTTTATCCATAGACTATTTTACCTCTGGCGGTGATAATTCTTGCGACAAACAATAGGTAAGGCGTTACCCAAC	ara2	con2	lux2
B71	GTAACAAAAGTGTCTATAATCACGGCAGAAAAGTCCACATTGACTTGTGAGCGGATAACAATGATACTTGAATACCTCTGGCGGTGATAAGGCGTTACCCAAC	ara3	lac2	con2
B72	TACAACGTCGTGTTAGCTGCCTTTCGTCTTCAATAATTCTTGACTTGTGAGCGGATAACAATTATAATTCGTGCAATTTTTAAACCTGTAGGATCGTACAGGT	con0	lac3	lux1
B62	TACAACGTCGTGTTAGCTGCAATTGTGAGCGGATAACAATTGACTATTTTACCTCTGGCGGTGATAATTCGTGCATAGCATTTTTATCCATACGTTACCCAAC	lac1	con2	ara2
B63	TACAATTGTGAGCGCTCACAATTTCGTCTTCAATAATTCTTGACTATTTTACCTCTGGCGGTGATAATTCGTGCATAGCATTTTTATCCATACGTTACCCAAC	lac3	con2	ara2
B66	TACAACGTCGTGTTAGTGCTCCCTATCAGTGATAGAGATTGACTATTTTACCTCTGGCGGTGATAATTCGTGCAATTTTTAAACCTGTAGGATCGTACAGGT	tet1	con2	lux1
B73	TACAACGTCGTGTTAAATTGTGAGCGGATAACAATTTAGTTGACATAGCGGATACTTCCTGATATAATTCGTTATCACCGCCAGGGGTAAGGCGTTACCCAAC	lac2	ara2	con4
B82	TACAACGTCTGTTAGCTGCAATTGTGAGCGGATAACAATTGACAAATAACTCTATCAATGATAGAGTTCGTTATCACCGCCAGGGGTAAGGCGTTACCCAAC	lac1	tet2	con4
B83	TACAACGTCTGTTAGCTGCAATTGTGAGCGGATAACAATTGACAAATAACTCTATCAATGATAGAGTTCGTTATCACCGCCAGGGGTAAGGCGTTACCCAAC	lac1	tet2	con4
B77	TACAACGTCGTGTTAGCTGCCTTTCGTCTTCAATAATTCTTGACATCCCTATCAGTGATAGAGATACTTCGTGCATAGCATTTTTATCCATACGTTACCCAAC	con0	tet1	ara2
B78	TACAACGTCGTGTTAGCTGCCTTTTAGCAATTTTATCCATAGACTTGTGAGCGGATAACAATGATACTTCGTGCAATTTTTAAACCTGTAGGATCGTACAGGT	ara2	lac2	lux1
B79	TACAACGTCGTGTTAGCTGTATCACCGCCAGAGGTAAGATTGACATTTATCCCTTGCGGTGATAGATTTGAATACATCTGGCGGTGATAAGGCGTTACCCAAC	con4	con4	con1

clone	Promoter sequence (between cloning sites XhoI and BamHI)	*distal*	*core*	*proximal*
B80	TACAACGTCGTGTTAGCTGCAATTGTGAGCGGATAACAATTGACTATTTTACCTCTGGCGGTGATAATTCGTGCAATTTTTATATCACCGCCAGGGGTACAAC	lac1	con2	con3
B81	TACAACGTCGTGTTAGCTGCCTTTTAGCAATTTTATCCATAGACTATTTTACCTCTGGCGGTGATAATTGATTCAATTGTGAGCGGATAACAATTTCACACAG	ara2	con2	lac3
B85	TACAACGTCGTGTTAGCTGCAATTGTGAGCGGATAACAATTGACTATTTTACCTCTGGCGGTGAGAATTGAATACATCTGGCGGTGATAAGGCGTTACCCAAC	lac1	con2	con1
B95	TACAATTGTGAGCGCTCACAATTTCGTCTTCAATAATTCTTGACATTGTGAGCGGATAACAAGATACTTGAATACATCTGGCGGTGATAAGGCGTTACCCAAC	lac3	lac1	con1
B96	TACAACGTCTGTTAGCTGCAATTGTGAGCGGATAACAATTGACAAATAACTCTATCAATGATAGAGTTCGTTATCACCGCCAGGGGTAAGGCGTTACCCAAC	lac1	tet2	con4
B86	TACAACGTCGTGTTAGCTGCCTTTTAGCAATTTTATCCATAGACTTGTGAGCGGATAACAATTATAATTCGTGCAATTTTTAAACCTGTAGGATGTACAGGT	ara2	lac3	lux1
B91	TACAACGTCGTGTTAGCTGCCTTTCGTCTTCAATAATTCTTGACACCTGTAGGATCGTACAGGTATAATTAACTCTATCAATGATAGAGTGTCAACAAAAAAAC	con0	lux2	tet2
B92	TACAACGTCGTGTTAGCTGTATCACCGCCAGAGGTAAGATTGACATTGTGAGCGGATAACAAGATACTTCGTTATCACCGCCAGGGGTAAGGCGTTACCCAAC	con4	lac1	con4
B93	TACAACGACATGGTAGCTGCAATTGTGAGCGGATAACAATTGACATAGCGGATACTTCCTGATATAATTGAATACCTCTGGCGGTGATAAGGCGTTACCCAAC	lac1	ara2	con2
C1	TACAACGTCGTGTTAGCTGCCTTTTAGCAATTTTATCCATAGACAAATAACTCTATCAATGATAGAGTTCCACCCCTATCAGTGATAGAGAGCGTTACCCAAC	ara2	tet2	tet1
C11	TACAACGTCGTGTTAGCTGCCTTTGTCTTCAATAATTCTTGACAATCAATGTGGATTTTCTGATACTTCGTGCAATTTTTAAACCTGTAGGATCGTACAGGT	con0	ara1	lux1
C12	TACAACGTCGTGTTAGCTGCCTTTGTCTTCAATAATTCTTGACAATCAATGTGGATTTTCTGATACTTCGTGCAATTTTTAAACCTGTAGGATCGTACAGGT	con0	ara1	lux1
C3	TACAACGTCGTGTTAGCTGCCTTTTAGCAATTTTATCCATAGACAAATAACTCTATCAATGATAGAGTTCCACCCCTATCAGTGATAGAGAGCGTTACCCAAC	ara2	tet2	tet1
C4	TACAACGTCGTGTTAGCTGCCTTTCGTCTTCAATAATTCTTGACATCCCTATCAGTGATAGAGATACTTAGATTCAATTGTGAGCGGATAACAATTTCACACA	con0	tet1	lac2
C5	TACAACGTCGTGTTAGCTGCAATTGTGAGCGGATAACAATTGACAAATAACTCTATCAATGATAGAGTTCGTTATCACCGCCAGGGGTAAGGCGTTACCCAAC	lac1	tet2	con4
C6	TACAATTAGTTTAACATAAGTACCTAGTAGGATCGTACAGGTTTACTATTTTACCTCTGGCGGTGATAATTCTTAGCAACAAACAATAGGTAAGGCGTTACCCAAC	lux1	con2	lux2
C9	TACAACGTCGTGTTAGCTGTATCACCGCCAGAGGTAAGATTGACATTTATCCCTTGCGGTGATAGATTTTGTGGAATTGTGAGCGGATAACAATTTCACACAG	con4	con4	lac1
C13	TACAACGTCGTGTTAGCTGTATCACCGCCAGAGGTAAGATTGACTATTTTACCTCTGGCGGTGATAATTCGTGCAATTTTTAAAATTAAAGGCGTTACCCAAC	con4	con2	con0
C22	TACAACGTCATTTCACTTTTCTCTATCACTGATAGGGAGTGGTCATAGCGGATACTTCCTGATATAATTCGTTATCACCGCCAGGGGTAAGGCGTTACCCAAC	tet2	ara2	con4
C23	TACAACGTCGTGTTAGCTGCCTTTGTCTTCAATAATTCTTGACAATCAATGTGGATTTTCTGATACTTCGTGCAATTTTTAAACCTGTAGGATCGTACAGGT	con0	ara1	lux1
C24	TACAACGTCGTGTTAGCTGCCTTTCGTCTTCAATAATTCTTGACACCTGTAGGATCGTACAGGTATAATTAGATTCAATTAGTGAGCGGATAACAATTTCACACA	con0	lux2	lac2
C14	TACAACGTCGTGTTAGCTGCCTTTTAGCAATTTTATCCATAGACAAATAACTCTATCAATGATAGAGTTCCACCCCTATCAGTGATAGAGAGCGTTACCCAAC	ara2	tet2	tet1
C15	TACAACGTCGTGTTAGCTGCTCCCTATCAGTGATAGAGATTGACTATTTTACCTCTGGCGGTGATAATTCGTGCAATTTTTAAAATTAAAGGCGTTACCCAAC	tet1	con2	con0
C16	TACAACGTCGTGTTAGCTGCTCCCTATCAGTGATAGAGATTGACTATTTTACCTCTGGCGGTGATAATTCGTGCAATTTTTAAAATTAAAGGCGTTACCCAAC	tet1	con2	con0
C18	TACAACGTCGTGTTAGCTGCCTTTCGTCTTCAATAATTCTTGACATAGCGGATACTTCCTGATATAATTCGTGGTCCATATTGCATCAGACATTGTACCCAAC	con0	ara2	ara1
C19	TACAACGTCGTGTTAAATTGTGAGCGGATAACAATTTAGTTGACAAATAACTCTATCAATGATAGAGTTTGTGGAATTTGAGCGGATAACAATTTCACACAG	lac2	tet2	lac1
C20	TATCACCGCCAGAGGTAAAATAGTCAACACGCACGGTGTTAGGCTTTTATCCCTTGCGGTGATATAATTAGATTCAATTGTGAGCGGATAACAATTTCACACA	con2	con3	lac2
C21	TACAACGTCGTGTTAGCTGCTCCCTATCAGTGATAGAGATTGACATTTATCCCTTGCGGTGATAGATTTCGTGCAATTTTTAAACCTGTAGGATCGTACAGGT	tet1	con4	lux1
C34	TACAACGTCGTGTTAGCTGCCTTTCGTCTTCAATAATTCTTGACATCCCTATCAGTGATAGAGATACTTCGTGCATAGCATTTTTATCCATACGTTACCCAAC	con0	tet1	ara2
C35	TACAACGTCGTGTTAGCTGCCTTTTAGCAATTTTATCCATAGACAAATAACTCTATCAATGATAGAGTTCCACCCCTATCAGTGATAGAGAGCGTTACCCAAC	ara2	tet2	tet1
C26	TACAACGTCGTGTTAAATTGTGAGCGGATAACAATTTAGTTGACATTTATGCTTCCGGCTCGTATAATTCCACCCCTATCAGTGATAGAGAGCGTTACCCAAC	lac2	con0	tet1
C27	TACAACGTCGTGTTAGCTGTATCACCGCCAGAGGTAAGATTGACACCTGTAGGATCGTACAGGTATAATTCGTTATCACCGCCAGGGGTAAGGCGTTACCCAAC	con4	lux2	con4
C29	TACAACGTCGTGTTAGCTGCCTTTCGTCTTCAATAATTCTTGACATAGCGGATACTTCCTGATATAATTCTTGCGACAAACAATAGGTAAGGCGTTACCCAAC	con0	ara2	lux2

clone	Promoter sequence (between cloning sites XhoI and BamHI)	*distal*	*core*	*proximal*
C31	TACAACGTCATTTCTTCTCTATCACTGATAGGGAGTGGTCTTGTGAGCGGATAACAATTATAATTCGTGCATAGCATTTTTATCCATACGTTACCCAAC	tet2	lac3	ara2
C32	TACAACGTCGTGTTAGCTGCCTTTCGTCTTCAATAATTCTTGACATAGCGGATACTTCCTGATATAATTCCTGTAGGATCGTACAGGTAAGGCGTTACCCAAC	con0	ara2	lux3
C33	TACAACGTCGTGTTAGCTGCCTATCAGTGATAGAGATTGACATTTATGCTTCCGGCTCGTATAATTGAATACATCTGGCGGTGATAAGGCGTTACCCAAC	tet1	con0	con1
C37	TACAACGTCGTGTTAGCTGCCTTTTAGCAATTTTATCCATAGACACCTGTAGGATCGTACAGGTATAATTCCTGTAGGATCGTACAGGTAAGGCGTTACCCAAC	ara2	lux2	lux3
C46	TACAACGTCGTGTTAGCTGCCTTTCGTCTTCAATAATTCTTGACATTTATCCCTTGCGGTGATAGATTTGAATACATCTGGCGGTGATAAGGCGTTACCCAAC	con0	con4	con1
C47	TACAACGTCGTGTTAGCTGCAATTGTGAGCGGATAACAATTGACACCTGTAGGATCGTACAGGTATAATTCGTGCAATTTTTAAAATTAAAGGCGTTACCCAAC	lac1	lux2	con0
C48	TACAACGTCGTGTTAGCTGCCTTTTAGCAATTTTATCCATAGACAAATAACTCTATCAATGATAGAGTTCCACCCCTATCAGTGATAGAGAGCGTTACCCAAC	ara2	tet2	tet1
C38	TATCACCGCCAGAGGTAAAATAGTCAACACGCACGGTGTTAGACCCTGTAGGATCGTACAGGTATAATTCGTGCAATTTTTAAAATTAAAGGCGTTACCCAAC	con3	lux1	con0
C39	TACAACGTCGTGTTAGCTGCCTTTCGTCTTCAATAATTCTTGACTTGTGAGCGGATAACAATGATACTTTGTGGAATTGTGAGCGGATAACAATTCACACAG	con0	lac2	lac1
C40	TACAACGTCGTGTTAGCTGCCTTTCGTCTTCAATAATTCTTGACATCCCTATCAGTGATAGAGATACTTCGTGCATAGCATTTTTATCCATACGTTACCCAAC	con0	tet1	ara2
C41	TACAACGTCGTGTTAGCTGCCTTTTAGCAATTTTATCCATAGACTATTTTACCTCTGGCGGTGATAATTCTTGCGACAAACAATAGGTAAGGCGTTACCCAAC	ara2	con2	lux2
C42	TACAACGTCGTGTTAGCTGTATCACCGCCAGAGGTAAGATTGACATTGTGAGCGGATAACAAGATACTTCGTGCAATTTTTAAACCTGTAGGATCGTACAGGT	con4	lac1	lux1
C44	TATCACCGCCAGAGGTAAAATAGTCAACACGCACGGTGTTAGGCACCTGTAGGATCGTACAGGTATAATTCGTGCAATTTTTAAACCTGTAGGATCGTACAGGT	con2	lux2	lux1
C45	TACAACGTCGTGTTAGCTGCCTTTTAGCAATTTTATCCATAGACATTTATGCTTCCGGCTCGTATAATTCGTGCAATTTTTAAACCTGTAGGATCGTACAGGT	ara2	con0	lux1
C49	TATCACCGCCAGAGGTAAAATAGTCAACACGCACGGTGTTAGACATAGCGGATACTTCCTGATATAATTCGTGCAATTTTAAACCTGTAGGATCGTACAGGT	con3	ara2	lux1
C59	TACAACGTCGTGTTAGCTGTATCACCGCCAGAGGTAAGATTGACATAGCGGATACTTCCTGATATAATTCCTGTAGGATCGTACAGGTAAGGCGTTACCCAAC	con4	ara2	lux3
C51	TACAACGTCGTGTTAGCTGCCTTTCGTCTTCAATAATTCTTGACATTTATCCCTTGCGGTGATAGATTTCGTGCATAGCATTTTTATCCATACGTTACCCAAC	con0	con4	ara2
C52	TACAACGTCGTGTTAGCTGCTCCCTATCAGTGATAGAGATTGACATAGCGGATACTTCCTGATATAATTGAATACCTCTGGCGGTGATAAGGCGTTACCCAAC	tet1	ara2	con2
C53	TATCACCGCCAGAGGTAAAATAGTCAACACGCACGGTGTTAGGCTATTTTACCTCTGGCGGTGATAATTCGTGCAATTTTTAAACCTGTAGGATCGTACAGGT	con2	con2	lux1
C54	TACAATTGTTTAACATAAGTGAATGGATCATTTTGCAGGTTTACACCTGTAGGATCGACAGGTATAATTCGTGCAATTTTTAAACCTGTAGGATCGTACAGGT	lux2	lux2	lux1
C55	TACAACGTCGTGTTAGCTGTATCACCGCCAGAGGTAAGATTGACCCTGTAGGATCGTACAGGTATAATTCGTTATCACCGCCAGGGGTAAGGCGTTACCCAAC	con4	lux1	con4

clone	Promoter sequence (between cloning sites XhoI and BamHI)	*distal*	*core*	*proximal*
C56	TACAACGTCGTGTTAGCTAGCCTTTCGTCTTCAATAATTCTTAGACTTGTGAGCGGATAACAATTATAATTCGTGCAATTTTTATATCACCGCCAGGGGTACAAC	con0	lac3	con3
C61	TACAACGTCGTGTTAGCTGCAATTGTGAGCGGATAACAATTGACAAATAACTCTATCAATGATAGAGTTCGTGCAATTTTTATATCACCGCCAGGGGTACAAC	lac1	tet2	con3
C70	TACAACGTCGTGTTAGCTGCCTTTCGTCTTCAATAATTCTTGACATCCCTATCAGTGATAGAGATACTTCGTGCATAGCATTTTTATCCATACGTTACCCAAC	con0	tet1	ara2
C71	TACAACGTCGTGTTAGCTGCCTTTTAGCAATTTTATCCATAGACATAGCGGATACTTCCTGATATAATTCGTGCAATTTTTAAACCTGTGGATCGTACAGGT	ara2	ara2	lux1
C63	TATCACCGCCAGAGGTAAAATAGTCAACACGCACGGTGTTAGGCAAATAACTCTATCAATGATAGAGTTCTTGCGACAAACAATAGGTAAGGCGTTACCCAAC	con2	tet2	lux2
C64	TACAACGTCGTGTTAGCTGCTCCCTATCAGTGATAGAGATTGACATAGCGGATACTTCCTGATATAATTGAATACCTCTGGCGGTGATAAGGCGTTACCCAAC	tet1	ara2	con2
C65	TATCACCGCCAGAGGTAAAATAGTCAACACGCACGGTGTTAGACATTTATGCTTCCGGCTCGTATAATTCGTGCAATTTTTAAACCTGTAGGATCGTACAGGT	con3	con0	lux1
C66	TACAACGTCGTGTTAGCTGCCTTTCGTCTTCAATAATTCTTGACACCTGTAGGATCGTACAGGTATAATTCGTGGTCCATATTGCATCAGACATTGTACCCAAC	con0	lux2	ara1
C67	TACAACGTCGTGTTAGCTGTATCACCGCCAGAGGTAAGATTGACATAGCGGATACTTCCTGATATAATTCGTGCATAGCATTTTTATCCATACGTTACCCAAC	con4	ara2	ara2
C68	TACAACGTCGTGTTAGCTGCTCCCTATCAGTGATAGAGATTGACATTGTGAGCGGATAACAAGATACTTTGTGGAATTTGAGCGGATAACAATTTCACACAG	tet1	lac1	lac1
C69	TACAACGTCGTGTTAGCTGCCTTTTAGCAATTTTATCCATAGACTTTTATCCCTTGCGGTGATATAATTCGTGCATAGCATTTTTATCCATACGTTACCCAAC	ara2	con3	ara2
C82	TACAACGTCGTGTTAGCTGCCTTTCGTCTTCAATAATTCTTGACATCCCTATCAGTGATAGAGATACTTCGTGCATAGCATTTTTATCCATACGTTACCCAAC	con0	tet1	ara2
C83	TACAACGTCGTGTTAGCTGTATCACCGCCAGAGGTAAGATTGACATTTATGCTTCCGGCTCGTATAATTCGTGCAATTTTTATATTACCGCCAGGGGTACAAC	con4	con0	con3
C84	TACAACGTCGTGTTAGCTGTATCACCGCCAGAGGTAAGATTGACATTTATGCTTCCGGCTCGTATAATTCGTGCAATTTTTATATTACCGCCAGGGGTACAAC	con4	con0	con3
C75	TACAACGTCGTGTTAGTTGCCTTTCGTCTTCAATAATTCTTGACATTTATGCTTCCGGCTCGTATAATTCGTGCAATTTTTAAACCTGTAGGATCGTACAGGT	con0	con0	lux1
C76	TACAACGTCGTGTTAGCTGCAATTGTGAGCGGATAACAATTGACATTTATGCTTCCGGCTCGTATAATTAGATTCAATTGTGAGCGGATAACAATTTCACACA	lac1	con0	lac2
C77	TACAACGTCATTTCACTTTTCTCTATCACTGATAGGGAGTGGTCATTTATGCTTCCGGCTCGTATAATTGAATACATCTGGCGTGATAAGGCGTTACCCAAC	tet2	con0	con1
C78	TACAACGTCATTTCACTTTTCTCTATCACTGATAGGGAGTGGTCATAGCGGATACTTCCTGATATAATTGAATACCTCTGGCGGTGATAAGGCGTTACCCAAC	tet2	ara2	con2
C79	TACAACGTCGTGTTAGCTGCCTTTCGTTTTCAATAATTCTTGACTTGTGAGCGGATAACAATTATAATTGAATACATCTGGCGGTGATAAGGCGTTACCCAAC	con0	lac3	con1
C80	TACAACGTCGTGTTAGCTGTATCACCGCCAGAGATAAGATTGACTTTTATCCCTTGCGGTGATATAATTTGTGGAATTGTGAGCGGATAACAATTTCACACAG	con4	con3	lac1
C81	TACAACGTCGTGTTAGCTGCCTTTCGTCTTCAATAATTCTTGACATCCCTATCAGTGATAGAGATACTTCGTTATCACCGCCAGGGGTAAGGCGTTACCCAAC	con0	tet1	con4
C85	TACAACGTCGTGTTAGCTGCCTTTCGTCTTCAATAATTCTTGACAATCAATGTGGATTTTCTGATACTTCGTTATCACCGCCAGGGGTAAGGCGTTACCCAAC	con0	ara1	con4
C94	TACAACGTCGTGTTAGCTGCCTTTCGTCTTCAATAATTCTTGACATCCCTATCAGTGATAGAGATACTTCGTGCATAGCATTTTTATCCATACGTTACCCAAC	con0	tet1	ara2
C96	TACAACGTCGTGTTAGCTGTATCACCGCCAGAGGTAAGATTGACATAGCGGATACTTCCTGATATAATTCGTGCAATTTTTATATCACCGCCAGGGGTACAAC	con4	ara2	con3
C86	TACAACGTCGTGTTAGCTGCAATTGTGAGCGGATAACAATTGACATAGCGGATACTTCCTGATATAATTCGTTATCACCGCCAGGGGTAAGGCGTTACCCAAC	lac1	ara2	con4
C87	TATCACCGCCAGAGGTAAAATAGTCAACACGCACGGATTAGACTTTTATCCCTTGCGGTGATATAATTCGTGCAATTTAAATAGTGAGCGAGTAACAACCAAC	con3	con3	lac4
C90	TACAACGTCGTGTTAGCTGTATCACCGCCAGAGGTAAGATTGACATAGCGGATACTTCCTGATATAATTCGTGCAATTTTTATATCACCGCCAGGGGTACAAC	con4	ara2	con3
C91	TACAACGTCGTGTTAGCTGTATCACCGCCAGAGGTAAGATTGACCCTGTAGGATCGTACAGGTATAATTCGTGCAATTTTTAAACCTGTAGGATCGTACAGGT	con4	lux1	lux1
C92	TACAACGTCGTATTAGCTGCCTTTCGTCTTCAATAATTCTTGACATCCCTATCAGTGATAGAGATACTTCGTGCAATTTTTATATCACCGCCAGGGGTACAAC	con0	tet1	con3
C93	TACAACGTCGTGTTAGCTGTATCACCGCCAGAGGTAAGATTGACTTTTATCCCTTGCGGTGATATAATTCGTGCAATTTTTAAACCTTAGGATCGTACAGGT	con4	con3	lux1
D52	TACAACGCCGTGTTAGCTGCCTTTCGTCTTCAATAATTCTTGACTTTTATCCCTTGCGGTGATATAATTCGTTATCACCGCCAGGGGTAAGGCGTTACCCAAC	con0	con3	con4
D10	TACAACGTCGTGTTAAATTGTGAGCGGATAACAATTTAGTTGACTATTTTACCTCTGGCGGTGATAATTCCACCCCTATCAGTGATAGAGAGCGTTACCCAAC	lac2	con2	tet1
D11	TACAACGTCGTGTTAGCTGCTCCTATCAGTGATAGAGATTGACAAATAACTCTATCAATGATAGAGTTCTTGCGACAAACAATAGGTAAGGCGTTACCCAAC	tet1	tet2	lux2
D3	TACAACGTCGTGTTAGCTGTATCACCGCCAGAGGTAAGATTGACATAAATACCACTGGCGGTGATACTTCGTGGTTCATATTGCATCAGACATTGTACCCAAC	con4	con1	ara1

clone	Promoter sequence (between cloning sites XhoI and BamHI)	*distal*	*core*	*proximal*
D4	TACAACGTCGTGTTAGCTGCCTTTTAGCAATTTTATCCATAGACATTTATCCCTTGCGGTGATAGATTTAGATTCAATTGTGAGCGGATAACAATTTCACACA	ara2	con4	lac2
D5	TACAACGTCGTGTTAGCTGTATCACCGCCAGAGGTAAGATTGACTTGTGAGCGGATAACAATTATAATTCGTTATCACCGCCAGGGGTAAGGCGTTACCCAAC	con4	lac3	con4
D7	TACAACGTCGTGTTAGCTGCCTTTCGTCTTCAATAATTCTTGACATTGTGAGCGGATAACAAGATACTTAACTCTATCAATGATAGAGTGTCAACAAAAAAAC	con0	lac1	tet2
D8	TACAACGTCGTGTTAAATTGTGAGCGGATAACAATTTAGTTGACATTTATGCTTCCGGCTCGTATAATTCCACCCCTATCAGTGATAGAGAGCGTTACCCAAC	lac2	con0	tet1
D9	TACAACGTCGTGTTAGCTGCTCCCTATCAGTGATAGAGATTGACACCTGTAGGATCGTACAGGTATAATTCGTGCAATTTAAATGTGAGCGAGTAACAACCAAC	tet1	lux2	lac4
D13	TACAACGTCGTGTTAGCTGCCTTTCGTCTTCAATAATTCTTGACATTTATGCTTCCGGCTCGTATAATTGATTCAATTGTGAGCGGATAACAATTTCACACAG	con0	con0	lac3
D24	TACAACGTCGTGTTAGCTGCCTTTCGTCTTCAATAATTCTTGACATAGCGGATACTTCCTGATATAATTCCTGTAGGATCGTACAGGTAAGGCGTTACCCAAC	con0	ara2	lux3
D14	TACAACGTCGTGTTAGCTGCCTTTTAGCAATTTTATCCATAGACACCTGTAGGATCGTACAGGTATAATTCCTGTAGGATCGTACAGGTAAGGCGTTACCCAAC	ara2	lux2	lux3
D16	TATCACCGCCAGAGGTAAAATAGTCAACACGCACGGTGTTAGGCATCCCTATCAGTGATAGAGATACTTGAATACATCTGGCGGTGATAAGGCGTTACCCAAC	con2	tet1	con1
D17	TACAACGTCGTGTTAGCTGCCTTTCGTCTTCAATAATTCTTCGACATAGCGGATACTTCCTGATATAATTCCTGTAGNATCGTACAGGTAAGGCGTTACCCAAC	con0	ara2	lux3
D18	TACAACGTCGTGTTAGCTGCCTTTCGTCTTCAATAATTCTTGACATCCCTATCAGTGATAGAGATACTTGAATACATCTGGCGGTGATAAGGCGTTACCCAAC	con0	tet1	con1
D19	TACAACGTCGTGTTAACTGTATCACCGCCAGAGGTAAGATTGACATTTATGCTTCCGGCTCGTATAATTGAATACATCTGGCGGTGATAAGGCGTTACCCAAC	con4	con0	con1
D20	TATCACCGCCAGAGGTAAAATAGTCAACACGCACGGTGTTAGACTTGTGAGCGGATAACAATAATACTTCGTGCAATTTTTAAACCTGTAGGATCGTACAGGT	con3	lac2	lux1
D25	TACAACGTCGTGTTAGCTGTATCACCGCCAGAGGTAAGATTGACATTGTGAGCGGATAACAAGATACTTGAATACATCTGGCGGTGATAAGGCGTTACCCAAC	con4	lac1	con1
D34	TACAACGTCGTGTTAGCTGTATCACCGCCAGAGGTAAGATTGACATAGCGGATACTTCCTGATATAATTCGTGCAATTTTTATATCACCGCCAGGGGTACAAC	con4	ara2	con3
D35	TACAACGTCGTGTTAGCTGTATCACCGCCAGAGGTAAGATTGACATTTATCCCTTGCGGTGATAGATTTCGTTATCACCGCCAGGGGTAAGGCGTTACCCAAC	con4	con4	con4
D36	TACAACGTCGTGTTAGCTGCCTTTCGTCTTCAATAATTCTTGACATTTATCCCTTGCGGTGATAGATTTCGTGCAATTTTTAAACCTGTAGGATCGTACAGGT	con0	con4	lux1
D27	TACAACGTCGTGTTAGCTGCTCCCTATCAGTGATAGAGATTGACATAGCGGATACTTCCTGATATAATTGAATACCTCTGGCGGTGATAAGGCGTTACCCAAC	tet1	ara2	con2
D28	TACAACGTCGTGTTAGCTGCCTTTCGTCTTCAATAATTCTTGACTTTTATCCCTTGCGGTGATATAATTAGATTCAATTATGAGCGGATAACAATTTCACACA	con0	con3	lac2
D29	TACAACGTCGTGTTAGCTGCCTTTTAGCAATTTTATCCATAGACACCTGTAGGATCGTACAGGTATAATTAACTCTATCAATGATAGAGTGTCAACAAAAAAAC	ara2	lux2	tet2
D31	TATCACCGCCAGAGGTAAAATAGTCAACACGCACGGTGTTAGGCTTGTGAGCGGATAACAATTATAATTGAATACATCTGGCGGTGATAAGGCGTTACCCAAC	con2	lac3	con1
D32	TACAACGTCGTGTTAGCTGTATCACCGCCAGAGGTAAGATTGACAATCAATGTGGATTTTCTGATACTTGAATACATCTGGCGGTGATAAGGCATTACCCAAC	con4	ara1	con1
D37	TACAACGTCATTTCACTTTTCTCTATCACTGATAGGGATGGTCATAGCGGATACTTCCTGATATAATTCGTGGTCCATATTGCATCAGACATTGTACCCAAC	tet2	ara2	ara1
D46	TACAACGTCGTGTTAGCTGCCTTTTAGCAATTTTATCCATAGACAAATAACTCTATCAATGATAGAGTTCCACCCCTATCAGTGATAGAGAGCGTTACCCAAC	ara2	tet2	tet1
D38	TATCACCGCCAGAGGTAAAATAGTCAACACGCACGGTTAGACATTTATCCCTTGCGGTGATAGATTTCTTGCGACAAACAATAGGTAAGGCGTTACCCAAC	con2	con4	lux2
D40	TACAACGTCGTGTTAAATTGGAGCGGATAACAATTTAGTTGACACCTGTAGGATCGTACAGGTATAATTGAATACCTCTGGCGGTGATAAGGCGTTACCCAAC	lac2	lux2	con2
D41	TACAACGTCGTGTTAGCTGCCTTTCGTCTTCAATAATTCTTGACATAGCGGATACTTCCTGATATAATTGAATACATCTGGCGGTGATAAGGCGTTACCCAAC	con0	ara2	con1
D42	TACAACGTCGTGTTAGCTGCCTTTCGTCTTCAATAATTCTTGACATAGCGGATACTTCCTGATATAATTCCTGTAGGATCGTACAGGTAAGGCGTTACCCAAC	con0	ara2	lux3
D43	TACAACGTCGTGTTAGCTGTATCACCGCCAGAGGTAAGATTGACATTTATGCTTCCGGCTCGTATAATTCGTGCAATTTTTAAACCTGTAGGATCGTACAGGT	con4	con0	lux1
D44	TACAACGTCGTGTTAGCTGTATCACCGCCAGAGGTAAGATTGACTATTTTACCTCTGGCGGTGATAATTCGTGCAATTTTTAAAATTAAAGGCGTTACCCAAC	con4	con2	con0
D45	TACAACGTCGTGTTAGCTGTATCACCGCCAGAGGTAAGATTGACAAATAACTCTATCAATGATAGAGTTGAATACCTCTGGCGGTGATAAGGCGTTACCCAAC	con4	tet2	con2
D49	TACAATTGTTTAACATAAGTACCTGTAGGATCGTACAGGTTTACTATTTTACCTCTGGCGGTGATAATTCTTGCAACAAACAATAGGTAAGGCGTTACCCAAC	lux1	con2	lux2
D60	TACAACGTCGTGTTAGCTGTATCACCGCCAGAGGTAAGATTGACATTTATGCTTCCGGCTCGTATAATTCGTGCAATTTTTAAACCTGTAGGATCGTACAGGT	con4	con0	lux1
D50	TACAACGTCGTGTTAGCTGCCTTTCGTCTTCAATAATTCTTGACATCCCTATCAGTGATAGAGATACTTCGTGCAATTTTTATATCACCGCCAGGGGTACAAC	con0	tet1	con3

clone	Promoter sequence (between cloning sites XhoI and BamHI)	*distal*	*core*	*proximal*
D51	TACAACGTCGTGTTAGCTGTATCACCGCCAGAGGTAAGATNGACTATTTTACCTCTGGCGGTGATAATTCGTGCAATTTTTAAAATTAAAGGCGTTACCCAAC	con4	con2	con0
D53	GTAACAAAAGTGTCTATAATCACGGCAGAAAAGTCCACATTGACATTGTGAGCGGATAACAAGATACTTGAATACCTCTGGCGGTGATAAGGCGTTACCCAAC	ara3	lac1	con2
D54	TATCACCGCCAGAGGTAAAATAGTCAACACGCACGGTGTTAGACTTTTATCCCTTGCGGTGATATAATTCGTGCAATTTTTAAACCTGTAGGATCGTACAGGT	con3	con3	lux1
D56	TACAACGTCGTGTTAGCTGCTCCCTATCAGTGATAGAGATTGACTTGTGAGCGGATAACAATGATACTTCGTGCAATTTTTAAAATTAAAGGCGTTACCCAAC	tet1	lac2	con0
D57	TACAACGTCGTGTTAGCTGCAATTGTGAGCGGATAACAATTGACATAGCGGATACTTCCTGATATAATTCGTGGTCCATATTGCATCAGACATTGTACCCAAC	lac1	ara2	ara1
D61	TACAACGTCGTGTTAGCTGCCTTTTAGCAATTTTATCCATAGACTTGTGAGCGGATAACAATGATACTTGAATACATCTGGCGGTGATAAGGCGTTACCCAAC	ara2	lac2	con1
D70	TATCACCGCCAGAGGTAAAATAGTCAACACGCACGGTGTTAGACTATTTTACTTCTGGCGGTGATAATTCTTGCGACAAACAATAGGTAAGGCGTTACCCAAC	con3	con2	lux2
D71	TATCACCGCCAGAGGTAAAATAGTCAACACGCACGGTGTTAGACATAGCGGATACTTCCTGATATAATTCGTTATCACCGCCAGGGGTAAGGCGTTACCCAAC	con3	ara2	con4
D72	TACAACGTCGTGTTAGCTGCCTTTCGTCTTCAATAATTCTTGACTATTTTACCTCTGGCGGTGATAATTCGTGCAATTTTTAAACCTGAGGATCGTACAGGT	con0	con2	lux1
D62	TACAACGTCGTGTTAGCTGCCTTTCGTCTTCAATAATTCTTGACTATTTTACCTCTGGCGGTGATAATTCCACCCCTATCAGTGATAGAGAGCGTTACCCAAC	con0	con2	tet1
D64	TACAACGTCGTGTTAGTTGCCTTTCGTCTTCAATAATTCTTGACATTTATGCTTCCGGCTCGTATAATTCGTGCAATTTTTAAACCTGTAGGATCGTACAGGT	con0	con0	lux1
D65	TACAACGTCGTGTTAGCTGCAATTGTGAGCGGATAACAATTGACTATTTTACCTCTGGCGGTGATAATTCGTGCATAGCATTTTTATCCATACGTTACCCAAC	lac1	con2	ara2
D66	TACAACGTCGTGTTAGCTGTATCACCGCAGAGGTAAGATTGACTTTTATCCCTTGCGGTGATATAATTCGTTATCACCGCCAGGGGTAAGGCGTTACCCAAC	con4	con3	con4
D68	TACAACGTCGTGTTAGCTGCCTTTCGTCTTCAATAATTCTTGACATAAATACCACTGGCGGTGATACTTTTGGAATTGTGAGCGGATAACAATTTCACACAG	con0	con1	lac1
D69	TACAACGTCGTGTTAGCTGTATCACCGCCAGAGGTAAGATTGACATTTATGCTTCCGGCTCGTATAATTGAATACCTCTGGCGGTGATAAGGCTTACCCAAC	con4	con0	con2
D73	TACAACGTCGTGTTAGCTGTATCACCGCCAGAGGTAAGATTGACACCTGTAGGATCGTACAGGTATAATTCCTGTAGGATCGTACAGGTAAGGCGTTACCCAAC	con4	lux2	lux3
D83	TACAACGTCGTGTTAGCTGTATCACCGCCAGAGGTAAGATTGACTATTTTACCTCTGGCGGTGATAATTCGTGCAATTTTTAAAATTAAAGGCGTTACCCAAC	con4	con2	con0
D77	TACAACGTCGTGTTAGCTGTATCACCGCCAGAGGTAAGATTGACATAGCGGATACTTCCTGATATAATTCCACCCCTATCAGTGATAGAGAGCGTTACCCAAC	con4	ara2	tet1
D78	TACAACGTCGTGTTAGCTGCCTTTCGTCTTCAATAATTCTTGACATAGCGGATACTTCCTGATATAATTCGTTATCACCGCCAGGGGTAAGGCGTTACCCAAC	con0	ara2	con4
D79	TACAACGTCGTGTTAGCTAGCCTTTCGTCTTCAATAATTCTTAGACATCCCTATCAGTGATAGAGATACTTGAATACATCTGGCGGTGATAAGGCGTTACCCAAC	con0	tet1	con1
D80	ACATAGCATTTTTATCCATAACCTGTAGGATCGTACAGGTTTACACCTGTAGGATCGTACAGGTATAATTCGTGCAATTTTTATATCACCGCCAGGGGTACAAC	lux3	lux2	con3
D81	TACAACGTCGTGTTAGCTGTATCACCGCCAGAGGTAAGATTGACAATCAATGTGGATTTTCTGATACTTCGTTATCACCGCCAGGGGTAAGGCGTTACCCAAC	con4	ara1	con4
D85	TACAACGTCGTGTTAGCTGCCTTTCGTCTTCAATAATTCTTGACTTGTGAGCGCTACAATTTATAATTCGTGCAATTTTTAAACCTGTAGGATCGTACAGGT	con0	lac4	lux1
D94	TACAACGTCGTGTTAGCTGCAATTGTGAGCGGATAACAATTGACATAGCGGATACTTCCTGATATAATTCGTGCAATTTTTATATCACCGCCGGGGGTACAAC	lac1	ara2	con3
D95	TACAACGTCGTGTTAGCTGCCTTTTAGCAATTTTATCCATAGACACCTGTAGGATCGTACAGGTATAATTCGTGCAATTTTTATATCACCGCCAGGGGTACAAC	ara2	lux2	con3
D96	TACAACGTCGTGTTAGCTGTATCACCGCCAGAGGTAAGATTGACATTTATGCTTCCGGCTCGTATAATTGAATACATCTGGCGGTGATAAGGCGTTACCCAAC	con4	con0	con1
D86	TACAACGTCGTGTTAGCTGCAATTGTGAGCGGATAACAATTGACATAGCGGATACTTCCTGATATAATTCGTTATCACCGCCAGGGGTAAGGCGTTACCCAAC	lac1	ara2	con4
D87	TACAACGTCGTGTTAGCTGCAATTGTGAGCGGATAACAATTGACACCTGTAGGATCGTACAGGTATAATTCCTGTAGGATCGTACAGGTAAGGCGTTACCCAAC	lac1	lux2	lux3
D89	TACAACGTCGTGTTAGCTGCCTTTCGTCTTCAATAATTCTTGACCCTGTAGGATCGTACAGGTATAATTCGTGGTCCATATTGCATCAGACATTGTACCCAAC	con0	lux1	ara1
D90	TACAACGTCGTGTTAGCTGCAATTGTGAGCGGATAACAATTGACTTGTGAGCGGATAACAATTATAATTTGTGGAATTGTGAGCGGATAACAATTTCACACAG	lac1	lac3	lac1
D91	TACAACGTCGTGTTAGCTGCCTTTTAGCAATTTTATCCATAGACTTGTGAGCGCTCACAATTTATAATTCGTGCAATTTTTAAACCTGTAGGATCGTACAGGT	ara2	lac4	lux1
D92	TACAACGTCGTGTTAGCTGTATCACCGCAGAGGTAAGATTGACTATTTTACCTCTGGCGGTGATAATTCCTGTAGGATCGTACAGGTAAGGCGTTACCCAAC	con4	con2	lux3
D93	TACAACGTCGTGTTAGCTGCCTTTCGTCTTCAATAATTCTTGACATAGCGGATACTTCCTGATATAATTCGTTATCACCGCCAGGGGTAAGGCGTTACCCAAC	con0	ara2	con4
A55	GTACCCGGGAATTCGATC			

Appendix A2. *Promoter library luminescence data.*

Clone	Promoter activity in triplicate (ALU)															
inducers	None	Lara	VAI	Lara VAI	IPTG	Lara IPTG	VAI ITPG	Lara VAI ITPG	aTc	Lara aTc	VAI aTc	Lara VAI aTc	IPTG aTc	Lara IPTG aTc	VAI ITPG aTc	Lara VAI ITPG aTc
A01	413	16797	274	15755	254	17158	244	19225	186	21475	276	19403	226	23793	325	25406
	193	24260	163	24020	322	27965	242	24610	133	30845	191	30346	192	32725	253	31084
	262	29883	303	29793	202	31682	342	31009	116	33337	216	32052	364	34214	282	33950
A02	32334	25931	26104	19858	25602	23962	24507	27115	251230	370381	305869	334570	367697	428286	312459	391528
	31449	42534	32485	38296	36921	48126	35452	47188	325226	488022	336910	469562	329401	493064	328203	442638
	36931	62174	34718	53762	37645	58477	35794	56780	309501	498499	309869	484921	317529	540940	329149	560239
A03	10	10	10	10	10	10	16	10	2364	714	2075	320	925578	1142542	986154	1144582
	24	10	14	10	10	10	10	10	843	270	989	253	1031996	1325067	1054722	1342502
	10	10	10	10	10	10	10	29	1055	307	675	319	1047987	1431281	1079441	1381706
A04	13340	13178	12898	9522	142645	157541	131719	162512	11505	14301	12405	12627	163884	222660	170146	206192
	16025	20488	15513	19896	170909	232688	170105	223051	15352	26490	16906	27845	177115	249644	183892	235387
	18270	28066	17738	27544	175447	242808	175488	240569	16575	29301	16956	27653	171984	285707	180289	287866
A05	20066	20278	21772	17419	67067	84940	62035	87293	19744	21383	23113	19514	99711	110233	98870	110842
	22635	27675	27033	26058	108636	127996	115574	123706	28818	38276	30768	41429	115127	151564	117193	151256
	32286	44318	36519	41328	118836	144809	118101	144266	32198	46964	33017	48579	117571	150989	122033	150587
A06	20260	15521	17015	16958	12907	15748	13511	16428	25809	239170	225740	263965	217349	289484	254611	316371
	20844	23224	25170	23837	28601	28454	37461	34003	332836	415964	389701	462618	374699	459080	392918	479585
	40163	56264	46436	51511	39696	49535	41928	53681	379339	496739	409941	547255	403526	487548	431613	499714
A07	174178	141523	142525	126497	132513	140630	117117	155084	129295	161759	157007	165980	162203	198147	174830	202883
	155808	204755	156262	197808	150201	211466	146439	208878	132063	219729	158354	197688	135565	206460	142296	208115
	145397	191217	153353	192405	130545	215449	144942	206250	122352	204106	128883	185859	131234	232548	137641	246356
A08	1186	99560	1115	69746	886	72955	1033	85100	784	105081	1068	108250	949	139434	825	123446
	1082	119910	938	125297	796	112669	912	102591	582	110123	1210	126579	753	71727	969	69855
	621	73998	833	82381	705	100395	919	103547	533	98947	714	75731	652	114680	953	122769
A09	15	10	10	10	10	34	91	54	10	95526	92400	73135	116067	108999	108385	94946
	10	12	10	10	27	25	49	38	89556	110141	95131	99547	96270	111565	91445	101561
	10	10	10	10	10	15	19	35	84625	107109	87628	96322	96066	119130	92957	112788
A11	856490	913426	778014	891068	787058	931087	777213	977956	785126	1081384	813492	960771	882714	1136974	891982	1093945
	793790	1042590	823192	1027894	781287	1029048	773809	1015158	725168	1106510	867995	985246	729445	1030242	756230	958702
	803309	1044418	762804	968853	732263	1064891	767107	1067431	728163	1117810	733053	1070664	762613	1193776	786771	1165630
A12	10	10	10	10	65825	318294	60198	333497	10	10	10	10	70117	403309	73870	370120
	10	10	10	10	65412	357101	63330	363325	10	10	10	10	61094	364851	62010	334937
	10	10	10	10	60912	363274	62655	350298	10	10	10	10	61958	409043	63984	385033
A13	10	25	10	10	128	195	118	184	43	10	42	58	191	175	222	246
	27	10	10	10	227	302	187	252	21	10	10	63	201	402	241	373
	47	10	10	21	237	281	237	351	10	32	10	10	211	530	231	490
A14	4215	2864	3554	2756	2724	2693	2513	3533	2732	3810	2982	3265	3575	4494	3092	4671
	3566	5181	3855	4912	3904	6439	4105	6559	3415	6980	4239	6965	3541	6954	4581	7242
	4164	7238	4475	7468	3986	7767	4384	7546	3757	7459	4138	8513	3804	8849	4870	9065

inducers	None	Lara	VAI	Lara VAI	IPTG	Lara IPTG	VAI ITPG	Lara VAI ITPG	aTc	Lara aTc	VAI aTc	Lara VAI aTc	IPTG aTc	Lara IPTG aTc	VAI ITPG aTc	Lara VAI ITPG aTc
A16	10	10	14	10	42	10	10	10	10	10	10	10	10	45	53	118
	10	10	10	12	10	10	10	23	10	10	10	10	10	17	10	29
	10	10	10	10	10	10	10	27	10	10	15	10	10	10	69	10
A17	10	20	10	10	10	26	41	66	50	15	10	26	66	10	115	37
	49	10	10	10	10	10	10	60	20	52	18	22	47	10	26	50
	23	10	10	10	25	10	24	10	10	47	26	47	115	71	42	111
A18	10	10	10	12	58585	61258	62100	58016	20	10	10	10	73995	89143	75761	91321
	10	39	10	10	60115	77326	56821	80080	10	10	10	10	73260	107327	80376	106549
	10	10	10	10	59234	93925	66803	95540	10	10	10	10	79696	128377	88005	137424
A19	10	10	10	10	81366	92805	79214	87835	10	10	10	10	108154	130279	102986	128692
	10	10	10	10	84083	105805	84668	111315	10	10	10	10	102645	146302	109580	150319
	10	10	10	10	84600	125800	84440	127923	10	10	10	10	108353	178774	117521	177491
A20	485447	479832	347310	342270	345252	487702	258766	383165	426615	621629	399312	440978	527378	634274	430199	525737
	472871	662302	375231	507835	444093	591223	353749	524503	464044	837049	442868	668409	433057	744803	399763	626876
	484700	688766	407995	603070	408216	682608	370152	639854	413260	812274	400682	708062	444732	840322	438072	775970
A23	169664	169658	139782	155617	126466	169102	118475	171037	127602	192869	130786	178696	133202	203168	138293	178497
	128213	169358	127219	182324	131063	167857	127069	169310	113013	161802	135940	149815	127393	146681	127082	136015
	133781	140821	123849	141581	128742	164120	130962	159394	118060	152756	114857	149711	125548	174337	141750	201322
A25	10	10	28	10	43	32	33	73	10	18	98	28	33	30	43	161
	28	48	10	10	10	101	14	22	38	10	28	38	103	51	53	81
	10	10	10	18	14	32	54	62	28	48	18	58	10	69	73	49
A26	10	10	10	10	10	58	39	10	55	46	35	10	35	82	10	72
	10	10	10	10	36	10	10	10	11	57	51	56	10	167	10	88
	10	10	10	10	10	10	10	10	10	24	23	86	10	120	99	167
A27	754790	734849	808537	698948	691434	752994	781969	713887	718878	828669	717956	812627	738368	923517	700314	894022
	636520	740892	649363	737286	643223	803923	592424	798193	567756	874886	682742	763516	670128	842747	649822	807700
	643717	758269	639816	824885	560218	844486	651860	861148	598942	836110	555015	823896	598252	1051614	713376	1119185
A28	10	30	40	10	20	10	10	20	30	50	10	10	1770	2110	1770	2100
	60	40	30	20	20	20	10	20	10	30	10	10	220	1940	2490	1370
	20	40	10	40	40	50	10	10	30	30	30	10	3110	3521	420	420
A29	38189	25441	31429	17532	26414	25631	21223	22217	26162	29627	20861	21222	29134	39029	29728	36765
	31270	43193	31370	34407	39023	53556	37342	50166	41439	63248	42224	58046	47230	64277	45204	61286
	49422	67538	48313	66476	50418	73941	50257	67050	49511	75034	48633	72148	52372	82954	53038	76629
A30	10	10	10	10	48	89	59	101	27	45	36	29	31	154	82	256
	28	10	10	10	47	111	118	241	18	20	36	43	113	109	131	140
	10	10	10	41	84	140	123	132	10	16	10	66	89	160	116	216
A32	2912	2026	2754	1682	11514	12038	11441	12109	1850	2428	1504	2130	12862	17520	13351	16549
	2305	2976	2373	3122	12522	17596	12287	17948	1955	4042	2208	4011	11439	16064	12847	16204
	2612	4529	2565	4634	12037	19704	13205	20947	2603	4795	2375	4101	12348	21763	13118	24459
A33	10	10	15	10	10	10	10	10	10	28	26	62	10	10	10	10
	13	10	10	11	10	10	10	10	10	10	10	48	10	10	10	10

inducers	None	Lara	VAI	Lara VAI	IPTG	Lara IPTG	VAI ITPG	Lara VAI ITPG	aTc	Lara aTc	VAI aTc	Lara VAI aTc	IPTG aTc	Lara IPTG aTc	VAI ITPG aTc	Lara VAI ITPG aTc
	13	10	15	39	10	10	10	10	10	16	10	49	10	10	12	10
A34	10	10	10	10	10	10	45	72	10	10	10	10	20	19	10	86
	16	60	10	10	42	59	10	40	10	10	10	41	43	19	10	39
	10	10	10	10	10	31	10	23	10	10	10	25	10	55	29	10
A35	10	10	10	10	14779	16375	13989	18149	10	10	10	10	18478	27920	19440	27804
	10	12	10	10	16382	23900	16564	24531	10	10	10	10	19001	23763	20386	25111
	10	10	10	10	16554	27073	18197	26943	10	10	10	10	19663	33453	23313	36746
A37	19	10	19	19	21	10	13	59	49	49	59	49	132	121	113	106
	19	49	29	10	27	70	80	24	79	189	109	159	106	273	75	178
	10	29	29	59	50	52	40	122	19	139	89	78	104	202	112	110
A38	143	134	192	125	278806	303405	269249	292563	74	102	84	102	275413	398232	268581	366988
	66	74	106	10	243061	343476	229239	319696	37	91	85	73	252511	387467	257269	353890
	66	73	136	102	223735	334929	226017	331673	27	132	67	162	263525	450786	272922	460726
A39	12	26	33	13	54	10	10	10	17	27	10	62	10	12	10	15
	10	10	10	10	22	10	82	11	10	10	36	16	10	46	18	28
	13	11	10	41	10	20	13	27	10	10	10	10	10	10	81	10
A40	48832	45143	49709	41780	50363	43264	47882	44495	113384	123490	112507	110509	102260	122733	94728	109968
	33861	37561	33288	35519	33758	41781	36123	41108	96543	138678	95108	131223	95805	131080	94129	129907
	33127	51020	35400	43711	35560	44641	36010	45085	100081	149710	103773	144873	109092	165081	109761	168934
A41	10	10	42	14	10	10	10	10	10	57	10	21	37	32	10	12
	10	20	10	10	37	10	10	15	10	10	10	10	53	10	10	10
	22	54	10	46	10	14	20	28	10	10	10	58	10	10	10	21
A42	14548	14442	14852	12395	13333	13259	13406	14431	72200	79380	73329	77412	68720	86902	66829	82368
	13193	17179	12333	15437	11931	18147	12261	17983	55196	74963	59064	70911	55614	70467	58649	73730
	13565	19344	13677	19382	11875	19364	12917	19670	50180	71164	53892	73115	57616	96726	63734	113756
A43	456504	778409	488162	681785	398500	813143	374866	760736	489181	911236	520431	806504	503113	866083	446201	806224
	459220	724214	387787	707697	339475	718025	338353	737730	447929	907460	440029	869454	394978	906539	408207	884031
	454034	900964	442190	861396	371601	875867	368762	859277	455762	984670	455082	997468	428673	1115063	446180	1129179
A44	10	10	10	10	1448	1100	1306	1249	10	10	10	10	1883	1864	2096	1869
	10	10	10	10	1789	2057	2039	2207	10	10	10	10	2387	3524	2855	3553
	10	10	10	10	2794	2957	2346	2983	10	10	10	10	3183	3734	3366	3598
A45	21110	23200	21341	19326	646408	809931	563263	748879	19308	25038	14574	22931	625305	952141	602017	808882
	19035	25106	17570	24514	578979	796746	522483	775483	15158	22576	15206	20089	531058	565941	531058	598077
	17489	23499	16727	23137	522669	752528	511062	676824	13213	17149	13433	16949	518761	819686	547981	1013806
A46	10	10	10	10	10	10	10	10	10	10	10	19	10	10	10	10
	10	10	10	10	10	10	10	10	10	16	10	10	10	10	10	10
	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10
A47	30032	25562	31319	22500	22770	24486	20688	25833	19489	24960	21085	26285	20501	31810	22461	31182
	19891	27590	20944	29218	21542	29223	20602	32686	16681	29027	21546	30002	20099	26373	21047	29659
	20070	27912	21034	29942	21433	32743	22293	34243	18766	29962	19198	29429	21139	36181	23923	41229
A48	10	10	10	10	8469	8063	7810	10536	35	10	10	10	8541	13933	10242	13186
	10	10	10	10	10156	13353	9950	15678	10	10	10	10	9822	13270	10111	14359

inducers	None	Lara	VAI	Lara VAI	IPTG	Lara IPTG	VAI ITPG	Lara VAI ITPG	aTc	Lara aTc	VAI aTc	Lara VAI aTc	IPTG aTc	Lara IPTG aTc	VAI ITPG aTc	Lara VAI ITPG aTc
	15	14	10	10	9762	16993	11282	17134	10	10	10	10	10783	18271	12443	18666
A49	30	10	10	49	33	43	52	70	38	47	35	16	50	115	48	64
	10	10	34	15	21	58	63	56	10	10	36	63	10	126	51	73
	17	10	14	22	34	85	30	64	28	10	36	40	62	88	98	101
A52	10	27	86	10	36845	28712	30194	28784	56	24	67	27	45023	43242	39049	36395
	50	90	70	70	30020	32346	29839	30275	99	62	29	105	36985	47090	35658	44151
	60	107	49	47	32266	36422	30024	35357	19	100	68	91	40772	56921	41391	56055
A53	5243	12363	3025	12053	333721	573768	354186	509158	3090	13793	2020	12680	328207	587455	348218	609377
	2364	9798	2482	10248	327386	536833	332130	483932	2430	7407	2508	4804	352449	479485	407566	454020
	3150	6939	2759	5725	359109	520390	354099	445815	2418	6222	2436	4360	386153	683162	408393	772494
A54	10	66	10	69	31715	82021	32235	84558	84	192	33	225	57536	164983	57518	161504
	10	67	10	10	23243	88040	22855	91955	10	144	17	145	47011	131459	49218	133607
	10	13	14	24	19634	86833	21820	85722	18	142	10	122	47970	174634	51527	197197
A56	10	10	10	10	10	37	10	10	304	326	325	333	324	261	340	485
	10	10	10	10	29	10	10	24	264	357	309	457	320	275	313	333
	10	10	10	10	10	10	17	10	186	312	285	444	343	448	406	461
A57	24318	15860	23092	14348	14542	13290	15166	16105	63438	66798	60689	61594	62371	73800	62829	71162
	15261	19900	15732	18484	15954	21520	15423	21882	52654	75722	62506	67343	56991	65660	58456	66441
	16995	24243	16873	23880	16214	23120	16966	23836	52017	71461	53267	73101	59447	96802	64926	104464
A59	1224	101	213	42	1045453	1283105	1010589	858127	72	49	10	78	978044	1139498	958678	931524
	10	10	10	10	824804	764994	627301	727532	10	10	10	24	564517	384038	592498	492615
	42	15	10	10	517585	682799	596102	615499	32	10	10	34	522508	737159	618710	939106
A60	40	10	40	20	20	10	30	20	240	230	540	220	34379	57536	25825	85459
	20	20	10	20	10	10	10	10	260	580	90	290	48738	42974	52218	63608
	30	10	10	10	20	10	10	10	340	420	270	500	53206	116713	36733	55719
A61	167203	131340	177793	116822	115313	116506	126109	131862	120627	124281	138223	133303	136709	148230	148784	158756
	134766	142288	142933	136413	131749	137814	124445	154671	109067	143321	131115	149449	114498	147747	131800	169382
	125158	150639	142697	154342	117495	156365	140813	163554	118342	151285	131576	168333	123321	187153	148578	225686
A62	6603	3557	4930	3299	2849	3339	2817	3326	2828	3888	2585	3787	2965	3804	3153	4132
	2485	3605	2884	3766	2496	3806	2677	3723	2310	3605	2646	3564	2629	3604	2646	3941
	2587	3714	2524	4313	2288	3453	2744	3702	2478	3774	2356	3541	2417	4787	3253	4932
A63	10	18	10	10	57	98	10	79	10	68	10	39	47	87	67	47
	28	39	10	10	47	10	47	78	10	19	18	10	67	48	87	10
	10	10	38	10	10	77	57	67	10	38	10	59	10	74	87	83
A64	5353	55752	4207	49083	3319	52457	2300	47061	3083	56448	2433	50908	3248	61211	2657	58475
	2774	50575	2693	53783	3031	56732	2420	54299	2453	53992	3080	45228	2668	50412	2925	56112
	3321	56452	2980	59784	2828	60515	3217	58253	2920	56328	2829	51817	2703	73801	3532	79053
A65	155595	88798	127940	91449	90498	82602	94467	96339	88216	92788	84331	96347	88535	117017	99345	116951
	81476	95897	89926	94516	88932	110531	95091	119704	88938	115900	103134	122289	99380	121067	107377	132037
	100157	123406	106303	133235	98965	131098	111248	140145	104509	141658	111698	143779	115816	161679	127652	181937
A66	4332	1551	1966	1001	589321	669868	525651	608570	2424	1889	1946	1196	769602	847330	689441	749832

inducers	None	Lara	VAI	Lara VAI	IPTG	Lara IPTG	VAI ITPG	Lara VAI ITPG	aTc	Lara aTc	VAI aTc	Lara VAI aTc	IPTG aTc	Lara IPTG aTc	VAI ITPG aTc	Lara VAI ITPG aTc
	2224	1820	1773	1020	617453	694458	528991	641043	2740	2384	1797	2532	670100	925732	615608	821862
	3081	3396	2300	2293	618526	800142	566012	727988	3088	4183	2577	3476	742105	1066294	727997	1054528
A67	216	50	247	52	53	99	108	11	81	82	82	100	62	10	121	181
	60	116	24	59	65	56	24	87	59	154	136	86	122	28	54	114
	83	186	103	250	10	173	52	50	71	91	110	55	84	102	31	190
A68	303673	377799	303073	271585	261222	329063	238101	317394	314049	544066	370628	402426	359425	516823	331227	458934
	282570	403516	259693	334269	280971	357491	262130	349201	272935	471414	291326	427476	288422	425233	271546	429750
	263494	396370	267507	377766	253424	381122	237121	356758	259317	435991	266765	423537	265210	498076	295664	501301
A69	611	181	422	129	15075	12713	16960	15774	143	94	94	158	15436	18285	18330	19365
	177	75	81	77	17374	20196	17879	22522	72	44	77	53	19548	22011	19001	23576
	60	115	109	199	18381	24285	20310	24230	34	61	10	62	21168	28893	23496	34242
A70	10	10	10	10	21	98	10	31	34	10	10	10	10	27	10	61
	10	10	10	10	10	10	10	10	10	27	11	10	10	10	10	20
	10	10	10	17	10	15	10	76	10	16	10	10	10	10	10	10
A72	3109	2744	16802	12988	3036	3013	16138	12631	6596	8177	16975	20346	7039	10184	19856	22113
	1962	3292	10169	14136	2579	3609	10433	13323	5896	9939	15242	14358	6880	9797	15774	15263
	2455	3753	9090	11713	2218	4392	10975	13747	6658	10942	14263	16675	7081	12565	16664	21613
A76	15	20	10	52	88	50	39	51	249	189	189	250	257	255	327	415
	10	41	20	10	19	28	10	48	300	308	329	380	248	297	228	316
	20	19	10	18	10	37	49	57	210	347	290	448	228	289	306	646
A77	53441	41063	46367	35528	39934	38011	36725	38221	41826	44167	37226	42131	40702	50859	39717	48629
	34012	39542	34271	38545	35174	41771	32229	42022	33015	44201	36724	42758	34659	43845	35643	46800
	34098	45489	32518	43120	29421	42400	32456	42329	30550	44268	32157	44160	34209	58367	39267	68292
A78	10	10	10	10	30	78	31	10	170950	419249	159426	371118	151182	397646	136853	364935
	31	36	10	10	10	89	31	10	134595	421167	138267	373914	128840	369516	127634	359324
	10	10	10	10	46	10	25	32	129882	412329	128983	393732	135074	488181	142616	519545
A79	482	82996	442	58519	351	49101	353	52248	323	89211	304	70978	544	76991	288	67849
	353	82488	333	71843	401	68722	313	65204	210	89892	309	71301	390	66541	282	70337
	253	83026	233	82925	231	74337	372	73862	211	83919	211	74305	318	104128	335	117193
A80	10	10	17	10	49	25	40	10	10	10	10	10	57	10	32	24
	10	10	10	10	10	10	10	10	10	10	10	10	86	10	71	149
	10	10	10	10	10	18	29	24	10	10	10	44	70	82	51	10
A81	10	10	10	10	204215	228804	164965	223041	10	27	10	29	234457	290243	227816	254993
	10	10	10	10	222253	299562	225170	261905	10	10	32	19	249418	316187	237360	314290
	10	19	22	10	227704	310753	225107	300222	10	10	12	10	247220	370082	268177	356509
A82	8511	6039	6199	5808	5963	5528	5559	6150	4958	6408	5498	6538	5427	7939	6368	7044
	5028	6109	5488	6069	5399	7317	5269	7854	4998	7780	5058	6488	5204	6733	5236	7414
	4557	6900	5378	7901	4978	8427	5659	7728	4747	7279	4537	6979	4904	9277	6171	10941
A83	4558	3393	3228	3293	2975	3238	3365	4104	5284	7225	5922	7976	5505	8926	6654	9287
	2431	3782	2790	4330	3034	4602	3155	5572	6126	9167	7106	8476	6409	8280	8139	10353
	3525	5193	3504	5263	3518	5404	4207	5835	6727	9798	7209	9788	7410	10869	9059	13590
A85	4030	3123	3572	3034	2961	3081	2652	3343	6105	7036	6665	8459	5245	8435	6215	7985
	2164	2712	2573	3393	2121	2517	2693	3229	3513	5638	4681	5991	4511	5597	5771	7930
	2270	3057	2719	3577	2100	3646	2897	3677	3630	5297	5110	6098	3770	7054	4998	10288

inducers	None	Lara	VAI	Lara VAI	IPTG	Lara IPTG	VAI ITPG	Lara VAI ITPG	aTc	Lara aTc	VAI aTc	Lara VAI aTc	IPTG aTc	Lara IPTG aTc	VAI ITPG aTc	Lara VAI ITPG aTc
A86	110158	85386	99482	84686	96432	96361	89962	86573	90662	90002	90124	88195	85952	107893	90490	102326
	80194	92765	85112	90877	93730	118678	85396	109921	87862	121934	100621	108558	101555	126606	104333	127657
	101140	113867	109343	126600	100051	130448	120116	127965	103775	128990	111604	128182	107020	153987	121063	155595
A87	830	154	982	174	65349	66127	65188	59944	503	173	353	164	68809	78737	61701	79113
	685	183	364	103	61438	69906	61268	74524	414	158	261	100	66925	85334	68321	85953
	511	119	290	147	62266	86034	70625	81035	351	87	260	147	71375	99863	80458	112322
A88	27	10	27	18	67	67	57	48	14	71	31	52	59	156	80	96
	38	38	18	10	10	66	47	35	11	39	11	39	50	34	10	84
	10	10	10	38	10	33	46	44	11	17	51	57	68	89	56	85
A89	10	12	10	12	52	22	12	52	23092	40530	38543	35258	44005	49377	41375	44883
	13	32	10	10	23	11	13	61	39913	52423	38806	49872	38356	54512	39301	51284
	13	41	13	51	10	71	13	61	41032	61984	41504	59428	45122	68427	46722	74070
A90	10	10	10	27	18	37	28	47	15	28	44	57	91894	108679	99742	104115
	18	10	10	10	10	36	28	56	48	25	68	34	99997	112951	103517	115889
	10	26	38	10	10	16	18	10	10	35	10	39	104716	137229	116631	142043
A95	10	10	10	10	3048	2508	3038	3579	28	10	10	10	2347	3528	2867	3538
	10	10	10	10	2478	2958	2438	2827	10	17	10	10	1587	1677	2147	2167
	10	10	18	10	1568	3177	2158	2267	10	10	10	10	1407	2457	2227	2966
A96	10	10	58	10	10	10	10	10	10	10	10	10	18	10	10	10
	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10
	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10
B04	10	10	10	10	368497	243538	315461	243720	10	10	15	36	339622	406690	310647	310628
	10	14	10	10	351587	225836	297852	247981	10	34	21	10	341269	320570	325836	283128
	10	13	10	14	306198	236113	309276	217828	27	10	10	32	329911	273955	340246	226322
B05	20	10	20	10	10	10	10	10	10	10	10	60	196212	250294	96323	162405
	30	10	70	20	10	10	10	10	10	60	10	40	71922	161502	153237	186105
	10	10	10	30	20	10	20	20	10	330	20	110	191378	384085	151709	22883
B10	28	18	10	10	10	10	29	10	152166	164949	139109	181334	109491	201558	191355	195105
	48	10	18	17	10	10	10	10	158167	149900	179522	145668	185039	152408	195835	172578
	69	10	10	18	20	10	50	10	170239	138141	198897	173582	188245	183266	209025	181691
B11	162	30	89	40	9677	10487	9397	12381	69	38	33	96	7995	11890	13505	11342
	79	40	91	69	8063	9515	9286	11970	66	37	104	41	11040	7420	12192	8861
	138	19	79	59	9523	6250	9995	14221	115	32	141	37	12534	7147	11058	8540
B13	39	33	10	10	10	10	10	10	24	39	14	50	10	10	10	10
	10	10	10	10	10	10	10	10	10	65	30	25	10	10	10	10
	38	10	16	10	10	10	10	34	10	36	10	36	10	10	30	10
B14	5665	3115	7746	2114	2422970	2471654	2451402	2509650	11536	3631	9041	2851	2863449	3211668	3481556	2938677
	7261	2866	6540	1954	1781722	2691458	2584312	2814416	9331	3213	7646	2594	2992939	3430892	3213402	3338558
	7512	2314	7270	2013	2198997	3023759	1614853	2754390	12758	3356	7425	2795	3456947	3939473	3107916	3421408
B16	77	27	57	27	34	74	42	83	2846	2666	3236	2536	3786	3082	4170	3049
	57	47	10	27	67	27	26	42	4028	2786	4327	3375	4436	2776	5229	3733

inducers	None	Lara	VAI	Lara VAI	IPTG	Lara IPTG	VAI ITPG	Lara VAI ITPG	aTc	Lara aTc	VAI aTc	Lara VAI aTc	IPTG aTc	Lara IPTG aTc	VAI ITPG aTc	Lara VAI ITPG aTc
	57	27	27	57	34	45	34	87	2866	2895	2835	2945	4657	3505	4345	3963
B17	38	48	10	10	7243	9964	8494	10757	37	57	27	67	10070	12860	12036	11805
	18	28	10	10	5860	9591	7786	11998	27	10	47	17	11969	12753	13041	12364
	10	28	38	38	8913	9731	6485	10794	37	47	57	57	12579	13202	12830	13759
B18	1062	868	882	1068	920	777	1185	1013	1051	1138	1348	1336	1216	1092	1601	1252
	764	1258	754	1028	775	907	1213	1029	1053	1123	1281	1610	1083	902	1502	1178
	883	898	854	1210	954	798	854	1328	879	1294	1260	1661	1434	2068	1547	1431
B19	25	10	10	10	43	10	10	10	476286	397717	378925	420400	498989	510272	589383	472911
	12	10	10	10	44	10	10	10	412135	483585	419074	453303	555260	405659	560535	467901
	10	10	10	10	10	10	10	10	483187	480231	532220	450014	570709	468687	542705	481706
B20	5542	220	330	60	60	90	40	190	8085	10990	11641	10098	23476	21840	13294	16121
	80	150	190	160	180	290	90	120	9297	17084	10168	18819	19421	19562	24761	28608
	70	130	130	130	230	330	80	170	12743	31271	29673	10379	29854	49414	16031	31322
B21	104822	599651	91830	642216	106791	606844	89987	683246	97267	632911	97003	705767	111044	757933	100037	738342
	89760	638020	82749	627760	94913	658690	86194	683122	95781	772487	85751	620804	92461	643670	79839	608707
	67802	588067	72446	590710	69536	619195	77121	626551	79985	692902	65847	615686	98827	742617	84576	731238
B22	10	10	24	10	10	10	16	10	79282	74336	96222	80855	95786	79458	113267	78582
	10	10	16	10	35	10	17	10	88764	53331	79290	61769	69357	54995	79188	63904
	28	10	10	10	83	10	10	10	55888	58737	57209	60441	83657	74006	75701	76791
B23	10	10	10	10	10	10	10	10	10	10	10	10	2917	2335	3171	2678
	10	10	10	10	10	10	10	10	10	6336	10	10	1491	1683	1646	1846
	10	10	10	10	10	10	10	10	10	10	10	10	1905	1971	2151	2800
B25	37149	35681	48400	35219	46190	33443	42395	38019	51163	46476	57489	44672	60935	48018	54727	53480
	35623	27381	39898	35109	42971	32343	49036	38878	87457	57622	80334	56125	80884	59086	78031	64718
	41943	34155	53752	37574	43680	28861	55155	45827	55893	52696	56479	52786	63245	57864	66211	68370
B26	9067	12565	10388	12364	416024	595522	502673	603417	14355	11901	12921	14527	696428	630619	841298	710182
	8977	11304	11171	9309	384408	658449	545849	540409	15501	16109	17488	18175	695500	700892	753252	772604
	10629	11454	14235	11553	547701	694813	316645	679906	18646	17685	21065	17093	861907	949452	751564	866668
B27	10	10	18	10	62	21	10	31	27	48	47	17	10	10	68	43
	10	18	10	28	10	17	10	10	37	47	37	67	71	73	56	112
	18	10	10	18	23	14	10	31	26	47	46	10	10	27	10	10
B29	5971	5421	7053	5561	8454	5869	8093	6901	8995	6712	9386	7623	10816	7550	10726	8982
	6923	4740	6402	5101	6191	5789	8403	7410	10428	8745	9897	8645	11578	7832	11808	9012
	7954	5291	8965	5982	7382	5990	7993	8252	7683	8284	9095	8705	12510	9994	12409	11066
B30	5740	6841	4468	6691	3136	6253	2827	6870	2888	5352	3109	6162	3944	5939	3493	6508
	2988	5950	3418	6651	3129	6413	3367	7992	3419	5919	3999	5450	3485	4431	3865	6040
	2127	5100	2497	4859	2449	4421	3147	7120	2378	5259	2506	5630	3804	5897	3463	7287
B31	308	658	648	518	64141	48555	61908	58403	640	669	631	607	75549	61734	72205	71157
	389	588	458	438	52710	49008	60918	58361	597	808	504	558	64834	55628	61305	64106
	359	579	409	439	46338	44071	57120	64161	363	612	193	465	68086	70623	73426	79628
B33	242	25492	214	23748	123	16171	177	18947	124	32106	209	33624	178	22225	162	23707

inducers	None	Lara	VAI	Lara VAI	IPTG	Lara IPTG	VAI ITPG	Lara VAI ITPG	aTc	Lara aTc	VAI aTc	Lara VAI aTc	IPTG aTc	Lara IPTG aTc	VAI ITPG aTc	Lara VAI ITPG aTc
	125	19745	136	18612	76	14898	78	17552	141	24380	171	25988	204	15572	146	18149
	78	16051	138	14888	154	10859	151	15836	72	26275	152	27460	272	24441	276	26044
B35	112	151	79	142	180	175	131	205	6859	5860	5189	6181	6070	5417	9828	4648
	130	192	121	130	180	106	100	145	5766	3554	4826	2527	5854	3935	5426	3053
	89	60	100	110	90	132	119	104	5477	3688	4067	3017	6737	5294	4966	3270
B37	81993	73170	83178	77100	66680	70327	90317	82191	78626	71193	86552	73421	80489	86155	81178	83484
	63703	53484	77342	73869	86725	72380	85331	85333	106161	72679	123278	85000	90056	59179	92711	95423
	70598	56410	96923	70730	77021	73082	98717	69314	84454	75282	80184	81949	86278	86364	97721	99882
B39	39	29	49	10	10	10	57	109	18	78	58	58	70	38	28	43
	19	10	10	10	79	10	55	37	10	18	37	48	15	38	97	74
	10	19	19	19	16	10	18	52	18	48	10	28	97	49	35	10
B40	2108	1729	1388	1449	4648	5801	5798	5807	1778	1899	2178	2359	7018	6469	7467	7457
	1048	1399	1369	1459	4438	5250	5981	7832	2208	2329	2408	2099	8418	7089	8147	8038
	1479	1709	1489	1559	5590	5379	5386	7679	2069	2139	1969	2149	8768	8227	9218	9687
B42	3589	4639	5340	4069	6360	4078	6110	6039	5420	5150	6051	5750	7741	5179	8482	7110
	5351	3769	4980	4029	5249	4348	5499	5288	6571	5510	7332	5750	7742	5599	7792	6200
	5241	4060	6412	3890	5770	4519	6170	7030	5511	6321	6872	6581	8833	6839	8391	8451
B43	7333	4850	6953	4220	6710	4611	6640	6620	5158	4816	7179	5707	8235	5065	8005	7084
	4290	3640	5291	3840	6032	4630	5739	6401	5637	3896	6458	4687	4944	3656	6206	5485
	4170	2699	3349	2979	3601	3581	5880	6209	3216	4407	4447	4947	6504	5643	6974	8353
B45	40	10	10	10	38	65	10	54	67	21	46	40	43	69	93	69
	10	27	20	10	18	55	28	45	17	52	46	42	84	92	84	81
	20	17	10	17	10	10	28	45	27	42	57	32	74	79	44	108
B46	10	39	19	19	10	34	10	34	87	95	107	154	175	95	103	116
	19	49	10	39	20	36	30	24	56	35	106	85	45	75	115	153
	29	10	19	10	32	22	61	37	46	74	55	84	65	70	85	58
B47	36	66	10	48	10894	98799	12217	107622	13	53	23	62	14205	116979	23253	118685
	10	35	10	28	11145	88450	10092	101078	10	61	17	73	18667	73076	17906	84465
	24	58	15	48	13035	63415	11874	76363	10	69	29	62	20157	90830	19473	104971
B48	10	10	29	10	147	44	126	285	10	18	16	11	114	199	241	122
	10	31	10	10	127	228	167	248	10	10	36	66	282	247	181	173
	10	35	10	10	157	135	117	187	66	10	25	10	231	209	251	150
B49	925	1173	845	817	949	952	1368	974	826	1146	1004	1199	1316	903	1083	1148
	949	733	926	1041	1097	780	1138	1050	1450	1438	1546	1078	1390	1655	1947	1200
	1348	1050	1103	1080	1186	1198	1376	1361	1214	1308	1233	1132	1420	1581	1418	1792
B50	868	1025	1158	956	1119	897	1277	1039	61585	60279	67458	62654	73457	60709	85426	65597
	1178	836	1108	866	1347	966	1166	1146	65141	60641	76242	71562	79661	65399	91138	69361
	1268	1056	1697	1116	1527	1288	1334	1454	66719	69974	77195	69113	83905	67812	85017	76808
B51	14	127	24	37	15064	60131	23813	71706	25	167	44	96	29191	71815	31469	89897
	10	67	16	57	20541	63607	26014	68287	64	146	42	174	32693	93157	36150	101754
	25	97	10	87	26573	70250	19897	88012	34	194	81	145	35296	106928	36816	112417

inducers	None	Lara	VAI	Lara VAI	IPTG	Lara IPTG	VAI ITPG	Lara VAI ITPG	aTc	Lara aTc	VAI aTc	Lara VAI aTc	IPTG aTc	Lara IPTG aTc	VAI ITPG aTc	Lara VAI ITPG aTc
B52	11800	5361	11460	3299	71260	59574	76036	77031	8334	4850	7964	4800	91593	74904	93124	87704
	8124	4300	6722	3490	71461	62675	67515	69793	8905	5431	9034	4280	94120	78719	98031	86849
	7834	3950	6432	3279	64541	65009	89198	75833	7063	5581	5931	4620	99546	91283	102944	104538
B56	1880	1070	2370	830	56143	65710	57244	79122	2510	1190	2350	1090	60950	78322	81291	92940
	2000	780	2010	770	47327	69452	58304	74250	2480	1170	2290	1190	91183	70940	101731	86818
	1490	760	1740	610	69483	64002	51542	82943	1910	1150	2250	1310	109223	87650	116254	103756
B60	5048	333492	3696	238029	6241	223161	6270	210843	3526	338967	3725	316942	7486	268506	5823	254887
	2995	280938	2766	262539	3968	211744	4267	199866	2151	164601	2047	114639	2044	145927	2167	164287
	1094	202623	1635	216945	1709	136559	2225	113437	1636	258783	1270	205681	2444	173275	2293	280327
B61	2800	140632	2260	113298	1950	105099	1490	93671	1139	126785	1508	107673	2308	128071	1628	111757
	1090	107125	1270	97319	1820	120732	1460	110455	2728	174365	2188	146642	2738	144706	2318	133508
	1510	119834	2030	106748	1470	127000	2340	112635	1748	162106	1688	132281	2828	173510	2708	162722
B62	76	34	26	10	165	168	155	110	26	96	65	19	192	204	150	85
	17	10	10	81	106	45	136	127	35	58	33	70	370	179	258	160
	47	18	10	10	185	134	146	76	25	38	42	83	319	283	280	173
B63	20102	12953	19149	11690	22768	14826	22456	15867	20633	13352	19881	13723	25003	16557	29221	18941
	17203	11450	15298	11029	19988	16299	20168	16749	20292	14595	20392	14706	30427	17968	29009	18868
	17243	10908	18196	10748	25970	15606	18332	18371	22329	16260	25562	15267	32306	20653	28085	21144
B66	19	49	19	19	19	37	39	17	908	1097	1387	1138	1537	974	1387	1144
	29	29	10	19	29	36	29	27	958	1008	1057	968	987	1074	1207	1125
	19	10	19	19	39	86	29	57	788	998	968	1028	1577	1073	1177	1253
B71	32	33	10	10	10853	17094	9421	18281	10	40	24	17	11985	20596	13079	21007
	10	10	10	10	9052	15632	8615	16311	10	80	10	18	6530	12112	8022	15276
	10	32	10	24	4660	11721	6856	14434	21	10	10	10	6762	12553	6170	16303
B72	499	122	459	69	102644	149103	109332	160249	386	97	286	31	101887	204848	167638	158406
	249	71	239	35	106582	161647	109271	143891	266	108	206	156	148982	140217	161820	136367
	219	85	239	102	129195	123482	107764	134131	345	40	175	70	177755	166331	168564	133492
B73	410	306	510	411	659	1021	619	943	340	318	370	322	629	987	609	1040
	300	202	270	373	659	918	629	760	400	460	420	545	679	1094	829	886
	370	300	470	412	729	807	749	1250	450	352	450	367	729	1149	949	1361
B77	10	10	10	30	10	47	10	47	20239	13790	21063	13431	23942	13410	27095	13581
	20	30	10	20	29	46	10	36	21133	12749	22538	11677	29431	10646	28346	11610
	40	30	10	30	29	36	10	36	21054	13511	23955	13001	32878	14179	30626	16085
B78	13	14	10	10	145	14835	57	15146	46	34	19	20	96	19278	151	17813
	10	25	10	14	67	19428	95	14845	10	49	10	23	214	20791	143	15161
	10	25	14	15	160	18287	78	15316	77	28	10	42	205	26574	160	22850
B79	38695	34871	36984	39510	47221	39826	40895	39575	101932	114644	113294	110387	148784	124024	121164	120810
	38856	30538	44526	32568	41942	38426	45063	39796	95853	100382	102990	89843	95769	92255	102013	99027
	25745	29653	34992	31442	28042	32532	38147	37551	78926	102769	83950	98426	116870	123011	104830	130097
B80	82	42	122	62	22671	14730	24058	16815	37	54	56	56	24537	16833	28941	19323
	92	23	32	23	22270	14701	32064	17016	69	29	69	52	17604	16675	21666	16875

inducers	None	Lara	VAI	Lara VAI	IPTG	Lara IPTG	VAI ITPG	Lara VAI ITPG	aTc	Lara aTc	VAI aTc	Lara VAI aTc	IPTG aTc	Lara IPTG aTc	VAI ITPG aTc	Lara VAI ITPG aTc
	55	34	33	33	15832	11607	19049	20557	54	48	114	90	19915	21149	23783	19683
B01	18	10	46	10	243	20020	367	21521	24	12	10	102	329	27735	591	25106
	10	41	11	10	220	19885	235	23243	13	67	22	58	244	21263	384	22015
	20	10	10	10	141	16517	237	29310	44	87	24	57	429	26843	323	26845
B02	25	35	12	10	10	10	12	37	14086	17840	16852	18069	120374	231782	174008	214533
	37	24	10	10	43	88	10	51	14088	16228	14736	13960	151450	151602	147931	144269
	42	10	40	24	33	11	41	48	15133	13308	15937	13849	156088	194917	150325	187064
B03	17	14	10	14	21	31	12	40	14311	18799	17549	20163	122375	240070	169020	224566
	10	10	10	10	21	82	10	10	15886	17125	15423	16784	136534	166995	147748	156456
	18	26	10	10	30	16	10	32	12999	17566	14863	16072	172879	204294	164764	204147
B05	10	10	20	40	1819	3124	1569	3044	40	30	30	40	2179	3824	2609	3465
	10	50	10	10	1779	2934	1939	3384	40	30	40	30	2289	4105	2179	3455
	20	30	20	10	1669	3014	1739	3164	40	20	50	40	2589	4454	2849	4243
B06	10	30	10	10	7353	33321	5962	35895	70	60	40	50	6251	37686	5520	33431
	60	30	40	10	6342	33783	6252	34527	20	120	10	20	5421	34527	6021	35433
	10	40	20	10	4830	32948	5961	35654	40	50	50	60	5621	41773	6501	47789
B91	1223	1494	1204	1153	1504	1297	1516	1467	710238	741043	659576	677057	685982	802074	696730	734656
	1023	1185	822	984	1206	1336	1335	1386	648332	621928	598927	576389	527443	619753	515162	599440
	766	1215	744	1015	1150	1168	1117	1317	542969	639468	506603	582843	592413	738689	523765	765685
B92	10	10	10	10	10	41	39	28	10	10	10	55	10	10	73	10
	11	35	10	10	10	16	27	23	10	10	34	27	10	29	42	10
	51	10	31	10	10	18	79	22	10	10	10	10	45	34	10	11
B93	55729	92645	61678	84373	178806	380672	154150	400042	59596	107929	64012	104916	195672	411392	240134	408419
	48900	80510	41393	80287	196635	350882	215978	369362	70535	84647	72509	80125	265666	300176	262834	303919
	47307	63122	47851	71405	200242	289563	216455	309497	66054	88391	65943	85956	285754	373333	261111	369601
B95	210	50	190	40	72083	101040	69270	98193	226	85	185	64	73570	136364	95881	129917
	160	70	160	40	73501	95540	69240	103929	156	86	146	46	126994	112195	120203	104856
	100	50	90	90	86037	90482	82000	100318	86	66	76	56	127465	102370	130318	122394
B96	60	30	30	10	10	13	10	13	8946	10369	9407	10108	52768	88891	74250	84631
	10	10	10	10	37	14	10	22	7254	6483	7404	6553	65171	65446	64221	62758
	10	10	10	30	25	25	86	33	6573	6763	6733	7614	77638	82365	73860	82249
C01	40	40	30	100	60	60	40	140	392773	334953	472410	350106	393336	369388	391452	346090
	130	60	60	31	30	130	80	90	419649	354708	436675	349722	361572	371431	425192	394314
	10	90	130	100	90	110	50	170	463745	376691	475738	353566	434317	341640	422569	281764
C03	65	95	106	66	116	66	76	56	974720	788248	972251	831049	946944	864680	872338	811638
	36	76	26	46	46	56	26	96	1021409	894320	1070064	877472	954432	913014	889117	918281
	34	26	76	76	56	76	56	66	1101966	882765	1035690	915434	991001	951096	1003269	951354
C04	140	240	240	260	80	200	320	310	1821	1810	2123	2076	23066	13621	19928	14367
	200	100	220	210	190	180	260	310	2285	2025	2117	2234	22460	14341	21260	14495
	160	110	210	130	200	210	320	340	2049	1983	2093	1967	21159	12429	26410	12764
C05	101	131	826	727	231	131	780	729	26094	22288	30478	24306	169753	175342	171807	168558

inducers	None	Lara	VAI	Lara VAI	IPTG	Lara IPTG	VAI ITPG	Lara VAI ITPG	aTc	Lara aTc	VAI aTc	Lara VAI aTc	IPTG aTc	Lara IPTG aTc	VAI ITPG aTc	Lara VAI ITPG aTc
	171	151	721	692	141	171	701	638	29218	22639	32502	22729	173714	176969	168301	193592
	111	131	761	803	251	261	773	808	26827	22157	28895	22862	157195	154333	161742	147627
C06	171	251	149191	142901	191	281	124649	131602	717	558	150303	119388	1120	1450	143391	129099
	241	301	121321	113957	201	211	116844	134375	816	527	133897	115880	1200	1189	132822	138585
	241	131	121545	108710	431	271	110107	137309	667	598	127879	109144	1153	1104	126997	113482
C09	1337	1767	1637	1477	29510	31691	29370	31701	22338	21935	22246	21896	96512	127640	97975	107847
	1408	1577	1488	1127	29018	33721	27722	29198	23391	24424	21654	23882	101178	131917	97650	112978
	2396	1507	1368	1347	28464	32675	29611	33128	24554	24182	24343	23582	107964	136734	107862	116836
C11	693302	630551	651194	552732	637635	591905	601858	561011	711617	667151	652787	557976	700890	614331	642673	546463
	721211	624628	614291	573249	714039	618778	609017	560293	761459	689347	729671	578904	829732	688393	694405	580303
	789651	661947	704477	603931	729464	667639	647666	594784	789063	690003	711012	625691	781629	604324	670184	474240
C12	720065	686573	645911	624969	655207	658772	617084	600992	732130	691521	747813	617296	776655	689444	657445	587919
	754726	683093	689158	619927	841702	650016	702175	616187	795730	707012	858444	627489	903344	689589	826660	665547
	877937	759955	880845	679319	879736	749973	794973	705440	946731	781035	818592	658500	887332	658754	835167	515268
C13	9551	8712	9301	6749	9674	8539	8751	6922	13853	12095	13246	9906	15929	13495	13587	11522
	8345	7794	8362	6272	10024	7672	7347	6869	14989	12031	13237	10290	15844	14764	14303	12735
	9442	8201	8677	7339	11675	9651	10421	8358	16817	13480	14368	11151	17455	14400	15778	10157
C14	186	197	156	177	227	377	207	247	377550	351552	433864	361996	412410	350716	396467	334984
	157	187	196	88	196	207	267	227	436359	354388	448194	340668	394921	359626	426423	348117
	86	127	136	227	306	276	306	206	425493	337915	415802	330689	404329	311896	369357	242261
C15	8094	7364	9237	7434	6904	7504	7955	7064	210069	248409	248651	258784	230441	246710	210166	234216
	9466	7755	9937	7945	8985	7845	7944	8205	252069	266202	255692	259067	224343	250898	213857	263244
	8565	7274	9087	7234	7745	7044	8496	8356	216162	236899	211068	227476	195397	211268	185927	187742
C16	7590	7821	8120	7620	7281	8131	7712	7861	203198	243471	247096	253726	212836	246237	203026	225417
	9154	7580	9725	7911	8813	8031	8983	8181	243491	253339	251858	254007	221089	239885	201359	251268
	8962	7742	8932	8071	7731	7881	8612	8851	204034	230876	207386	218161	185840	212054	185716	189405
C18	147	166	104	125	124	79	201	205	202	235	187	227	243	256	345	193
	105	60	201	133	73	149	176	105	211	256	213	324	262	354	172	252
	122	140	161	201	183	118	202	170	192	245	181	198	301	244	315	383
C19	171	221	719	571	271	301	742	808	531	540	1067	901	27171	19088	28716	18796
	321	171	632	593	241	351	563	899	521	640	938	911	28899	20664	26629	19699
	91	141	461	833	251	381	773	868	381	640	1100	832	26508	17505	27914	15470
C20	147	257	147	136	1317	1935	1107	1855	154	132	134	103	4231	4129	3709	4433
	118	126	148	166	1277	1815	1207	1635	153	261	153	142	3519	4805	3479	4291
	256	86	167	116	1107	1745	1127	2185	202	220	262	103	3427	5004	3557	4293
C21	7844	5821	6692	5201	7231	6429	6560	5920	79794	72540	88790	67332	78672	89958	74712	77840
	7303	5631	6293	5081	7081	6599	6841	5648	92638	71690	98727	65360	90710	90739	77710	92132
	7674	6152	7203	4851	7281	7240	7280	6239	89887	75233	87827	66492	83437	88871	77448	81621
C22	4880	4795	4857	4917	4576	4918	5020	4318	9125	7567	9941	7140	10121	7954	9425	7812
	4951	5001	4592	4540	5562	5053	4882	4854	10321	7804	10789	7896	11291	9056	10116	8258
	5598	5099	5954	4967	5212	5124	5288	5140	9444	8333	9650	7539	10806	7885	8822	7619

inducers	None	Lara	VAI	Lara VAI	IPTG	Lara IPTG	VAI ITPG	Lara VAI ITPG	aTc	Lara aTc	VAI aTc	Lara VAI aTc	IPTG aTc	Lara IPTG aTc	VAI ITPG aTc	Lara VAI ITPG aTc
C23	691907	612104	594535	548969	592680	592935	581688	534753	676154	641037	627537	580477	730241	614837	600488	533855
	678043	632466	623601	585406	729438	620763	625283	563528	799585	658896	758528	609853	771020	664931	657376	607350
	690198	637953	671223	601005	667952	668169	588134	583479	718677	666752	640195	589811	702428	622172	614868	520173
C24	5877	5747	5375	4929	35832	29446	36192	31527	7189	5943	6554	5527	49293	36247	43190	33694
	5971	5734	6224	4893	42227	31120	41172	31188	7264	5536	7871	4970	55123	39175	51870	37440
	6747	6119	7241	5259	43238	33206	45453	38551	7794	6113	7049	5387	56361	38214	48791	32707
C26	423	327	407	257	451	391	442	608	22389	18859	21917	18798	179518	225669	171985	206791
	267	372	303	491	673	532	472	360	21075	20171	22171	18140	176262	238954	164290	203742
	349	379	363	360	441	421	649	479	23042	19491	21486	18467	182403	233998	168453	213581
C27	4782	2480	3270	2060	3303	2702	2832	2172	5754	4567	4926	3625	6155	5198	4617	4601
	3270	2130	2169	1400	3094	2252	2722	1893	5595	4504	5244	3625	6505	5407	5036	4464
	3080	1990	2560	1610	3444	2902	2562	2482	5582	4730	4602	3682	5870	5126	4691	3992
C29	209818	211259	242964	226268	166315	193759	162111	200550	172163	215150	187464	218086	168273	234292	174386	234295
	153087	191137	145899	195837	152307	194875	157644	207999	169517	213169	161293	221466	176033	228160	194764	239006
	181819	173373	183634	213831	179232	209676	173660	218454	194257	227329	202184	228339	204875	245344	230560	250911
C31	290	231	450	291	291	520	390	571	310	651	510	590	941	1105	1092	1056
	131	151	291	231	449	470	389	760	550	431	580	460	1060	944	1162	1005
	121	221	411	411	550	509	490	700	580	430	580	540	1021	805	1133	957
C32	5197	2977	4618	2799	4239	3858	3769	3868	4534	3411	6005	3563	5212	5259	5203	4430
	3971	2529	4161	2749	4159	3688	4099	3908	4885	3732	5456	3553	6093	5269	5323	4660
	4411	3269	4081	2939	4419	4068	4359	4218	5225	3901	5076	3693	5552	5019	5854	4350
C33	5985	9515	5135	7474	6277	8002	5166	7542	32938	47478	36826	36957	35270	49369	30727	43970
	5855	7864	4505	6743	5236	8002	4475	7042	31779	45977	28450	37138	33930	49832	29913	40917
	5705	8654	5245	6933	5486	8412	5356	7932	33478	46913	30253	38738	34515	54474	29260	44777
C34	190	235	172	166	293	229	225	149	170121	132270	160226	125204	175695	130265	161551	115431
	177	250	252	266	270	223	217	193	167149	137531	173442	126132	156640	133889	154409	121326
	290	216	340	225	382	225	232	284	181646	137401	166854	125720	170429	140747	160238	121866
C35	1057	1521	1094	978	1152	1050	1024	1087	1103994	858360	1047307	879005	1034645	912711	932317	842940
	1016	1153	1022	948	1044	1179	784	1212	1084387	880962	1025243	867436	923040	912265	879575	884368
	1283	1104	1004	1060	884	940	745	1265	1105619	903631	1049696	885367	956416	1003197	976917	981484
C37	4630	27113	3389	15681	3883	29632	2191	15553	4041	23648	2598	17136	5538	28713	3177	18848
	3796	22908	2011	14993	3893	27042	2248	17897	5180	26982	2934	19493	5556	35064	4014	24550
	4491	25777	2776	18217	5401	30636	3418	21455	6009	28716	3346	20675	6867	35419	3885	23122
C38	302	360	322	283	328	501	337	575	843	860	902	812	1396	1468	1277	1382
	136	303	324	284	419	523	536	565	982	750	882	712	1494	1505	1286	1507
	264	381	343	292	457	671	566	553	1021	1009	953	871	1583	1636	1265	1141
C39	50	10	40	50	2930	2140	450	12463	20	30	20	10	4491	8325	2530	2890
	30	10	50	30	10128	17505	12473	3741	10	50	10	20	3420	9337	11010	8696
	20	10	10	20	6693	3751	9747	4882	50	10	10	10	3270	13104	3020	10869
C40	574	816	664	878	589	930	788	920	41649	19966	42424	17823	42456	21061	37080	19618
	446	747	456	739	650	1080	568	861	43240	20247	40430	18173	43321	20619	37563	18223

inducers	None	Lara	VAI	Lara VAI	IPTG	Lara IPTG	VAI ITPG	Lara VAI ITPG	aTc	Lara aTc	VAI aTc	Lara VAI aTc	IPTG aTc	Lara IPTG aTc	VAI ITPG aTc	Lara VAI ITPG aTc
	455	797	575	839	919	949	738	970	43280	19985	41275	18243	43562	19233	38397	16819
C41	1885	53495	2040	38507	1602	45129	1503	39720	1784	57650	1822	45414	2195	55374	2305	45874
	1335	48001	1306	37798	1405	44121	1544	37815	1935	57933	1916	46119	2633	59869	2391	47134
	1470	51393	1620	39234	1780	48220	1681	41931	2040	60668	2179	47166	2718	62230	2215	49511
C42	325	327	235	629	1747	1568	1306	920	586	646	586	628	2142	2003	1783	1626
	357	368	297	470	1795	1288	1416	1259	626	716	516	608	2090	1982	2022	1626
	276	388	276	329	1575	1636	1586	1549	565	775	636	648	2241	1992	1993	1716
C44	25730	7676	19157	2473	2032	1454	1691	1827	1816	1340	1665	1682	2064	2297	2405	2192
	1287	932	917	923	1478	1594	1389	1395	1715	1240	1365	1191	2118	1895	2254	2151
	1316	922	977	1033	1341	1482	1459	1575	1205	1199	1476	1151	2009	2167	2265	1932
C45	28077	55498	25505	46785	19994	58321	18219	51984	23491	58077	27270	48997	25692	66109	21708	57402
	22254	50202	19596	43127	19193	57048	18831	52939	25146	58451	27301	51215	27631	70601	23787	59260
	23401	53320	21995	47238	20096	62755	19815	55401	25749	61228	24635	51720	27513	67324	23135	58456
C46	88261	108151	98704	107013	92303	107318	94500	106638	167169	164938	213762	170186	186215	181039	181484	172854
	99975	108589	98552	109131	89313	114604	91326	123453	184518	168578	216406	172603	182320	176456	177680	182946
	89541	102126	97769	108580	86945	112176	89301	124440	180664	165338	179304	164530	169577	166432	168676	160072
C47	22722	23292	21969	20925	382646	541243	372725	423181	26507	28810	30746	25124	479704	538580	422978	469518
	24792	25893	25425	21277	412487	472317	397027	455831	32007	28316	34763	25172	516771	525758	440818	499536
	23920	22531	25615	21526	370196	476696	361780	463055	29930	28869	28768	23309	492828	475020	421442	435148
C48	62	78	121	60	897	1326	1073	1537	196468	215060	184156	172519	203077	205861	214400	190538
	91	258	110	148	1242	1897	1227	1920	235609	210821	216526	214523	261426	205977	257717	201600
	131	157	110	149	1027	1654	1287	1955	216834	211120	217873	187665	258662	215452	246366	220990
C49	24414	28632	26655	26534	22770	29173	23132	25630	26605	29647	28825	28272	27557	31102	24666	27870
	25653	26243	25150	25139	24729	27426	24959	26745	31619	28039	32605	28704	29031	31562	26753	31244
	26908	27949	29631	28071	25029	30238	24316	30381	31045	29133	28876	28171	28227	28254	26149	26592
C51	7527	4096	6827	2666	6592	4769	5740	3928	13487	7188	13055	5846	14011	8519	11696	6648
	6480	3508	4917	2767	6613	4259	5819	3460	13196	7719	10050	6068	15473	10371	13309	7227
	6598	3556	5557	3166	7173	4487	6631	4178	14097	7808	12465	6697	17669	9930	13558	8421
C52	14377	13554	13084	10369	11209	13063	11400	12772	20577	20209	23358	17592	24391	26021	22103	22477
	11922	10459	11240	9507	11119	13213	11570	12422	21350	19396	22696	18164	24771	26734	23127	22809
	11681	11831	12302	10409	11530	13794	11820	13133	22052	21203	22504	19066	24219	26181	23889	23632
C53	216	468	317	362	457	530	248	621	635	676	644	629	634	925	575	828
	187	400	178	392	337	580	328	691	634	586	664	579	814	1014	904	987
	267	399	277	342	387	839	418	660	724	845	755	788	953	964	774	837
C54	9578	7204	8306	6063	6963	7512	6482	7553	7736	6602	7916	5742	8894	8973	7033	7292
	7325	5862	6444	4651	7143	6982	6352	6181	8946	6983	8667	5372	8374	8953	7383	7713
	7125	5742	7075	5442	7093	7843	6082	7352	9157	7143	7636	5852	9786	8593	8425	7302
C55	2808	2339	2038	1920	2624	3364	2333	2804	2719	2799	2308	2079	3207	4409	2818	3121
	2169	2110	1749	1470	2820	3253	2341	2843	2828	2489	2488	2189	3583	4488	2997	3631
	2579	2310	2169	2070	2863	3261	2392	2933	3248	2799	2339	2059	3893	3919	2848	3261
C56	1136	548	1167	330	87121	91240	91699	90123	1130	889	1570	600	100810	105110	104413	96761

inducers	None	Lara	VAI	Lara VAI	IPTG	Lara IPTG	VAI ITPG	Lara VAI ITPG	aTc	Lara aTc	VAI aTc	Lara VAI aTc	IPTG aTc	Lara IPTG aTc	VAI ITPG aTc	Lara VAI ITPG aTc
	1000	460	840	330	99198	90975	101100	90083	1410	769	1240	760	123839	111057	106457	97331
	760	540	910	420	90186	96552	101446	93840	1320	919	1390	740	125177	107941	106764	98704
C59	94805	94636	86928	84954	67174	98470	70019	90922	74322	91268	90198	83589	81375	108982	80131	103654
	76050	84577	73043	75357	63610	102771	63250	104018	83906	88842	98957	87647	87089	115057	82286	108341
	75989	90047	80591	85492	67611	113035	71347	128313	77086	97068	81242	89809	85024	115296	77859	98812
C61	317	266	246	237	257	346	187	287	3466	2435	4146	2505	79250	63613	68629	54504
	177	187	137	147	227	266	327	366	3284	2425	3535	2255	89006	73769	83964	62542
	146	206	186	216	237	336	237	336	4315	3055	4006	3085	91686	75288	80365	61401
C63	563	417	405	420	409	653	598	358	834	675	901	597	909	789	872	905
	429	359	437	301	438	365	537	467	600	715	1010	728	1201	889	834	873
	250	368	366	321	359	604	398	486	622	644	867	698	899	980	814	765
C64	345	356	295	237	322	471	412	491	424	574	633	495	898	767	799	569
	306	286	236	287	382	511	342	361	633	444	663	465	878	677	669	789
	246	216	296	267	342	380	302	470	523	434	624	505	638	828	649	649
C65	5133	3171	4172	2751	4761	3641	3891	3200	4742	3822	5142	2811	5022	4211	4341	3461
	5003	3632	4032	2711	4091	3761	4071	3140	5092	3742	4681	2901	5522	3971	5092	3401
	4773	3432	4342	2541	5072	3551	3881	3441	5332	3722	4061	2941	5622	3761	4211	3060
C66	307	298	258	189	396	456	477	527	427	438	437	359	825	865	696	676
	258	198	298	239	366	436	367	447	527	428	647	359	735	885	606	836
	168	198	168	299	456	436	447	466	597	478	418	358	934	845	686	566
C67	7675	5693	6404	5013	6419	6586	6300	6548	7184	6103	7765	5353	8756	8644	7186	7125
	6204	4923	5763	4362	6970	7327	6440	6839	7244	6284	8015	5433	8854	8602	7936	7374
	6284	5363	6714	4662	6069	7226	5980	6847	7695	6703	6643	5733	9573	8404	8186	7434
C68	30	10	10	20	180	290	110	510	20	30	30	10	29593	45332	26890	30467
	10	50	30	30	350	770	320	360	10	20	10	10	22080	33061	43538	44052
	90	20	10	10	670	630	390	420	10	40	10	10	47962	73217	23737	830
C69	100	110	120	140	276	195	156	167	150	191	142	242	267	308	259	193
	140	140	180	90	197	215	187	207	111	160	321	181	307	317	299	293
	149	70	120	80	157	185	196	277	230	300	132	142	196	335	157	249
C70	837	1052	780	985	800	1520	791	1062	58160	30130	69438	25441	66331	27903	56746	25507
	640	1033	671	895	950	1418	1041	1180	61815	29768	61747	25743	66593	29792	60613	26853
	580	752	611	864	870	1387	910	1468	58481	32210	59684	26565	65005	33818	59571	27209
C71	65190	98030	52186	83787	49044	106098	45476	96221	59940	111459	64503	97641	72043	136786	59343	117470
	52674	94151	44216	79663	53141	116835	47583	101243	62787	111072	55868	98890	71109	142280	64191	118908
	52520	94169	43649	83218	51392	116526	46864	110521	63728	117165	59674	103791	73630	169189	72699	150019
C75	153363	133371	141883	119725	118255	151516	122632	126113	143459	139897	160055	131939	163423	171289	150099	138420
	121611	125257	134934	115890	125031	141892	129372	131355	166107	140747	165954	128873	152590	170755	137193	150592
	115267	128158	139252	113667	116999	148030	120999	137429	152696	149896	124796	125032	164061	165654	136171	134821
C76	425	469	557	291	22402	26684	25805	28387	497	427	604	579	29013	30790	29658	31429
	341	350	338	312	23636	29399	24408	29551	523	337	463	399	27849	31605	27650	29658
	222	480	407	342	22994	29548	26629	31459	645	556	510	380	26792	28731	25180	27038

inducers	None	Lara	VAI	Lara VAI	IPTG	Lara IPTG	VAI ITPG	Lara VAI ITPG	aTc	Lara aTc	VAI aTc	Lara VAI aTc	IPTG aTc	Lara IPTG aTc	VAI ITPG aTc	Lara VAI ITPG aTc
C77	188	189	228	189	441	399	390	499	428	469	407	339	788	717	618	738
	218	229	198	269	570	389	480	488	457	439	587	469	788	588	628	688
	118	139	238	149	480	629	550	388	437	288	537	449	858	749	610	620
C78	3924	3519	4183	3088	4007	4273	3396	3706	4263	3990	4802	3538	5169	4750	4420	3852
	3483	3560	3455	2879	4001	4604	3493	3895	4438	4099	4807	3388	4615	4909	4155	3790
	3546	3750	3501	2889	4052	4533	3683	4471	4223	3957	4190	3196	5262	4109	3940	3821
C79	1067	538	927	358	62901	79630	54932	67663	1747	1020	1557	760	80344	93367	69076	78618
	907	488	757	289	63164	78739	54781	68502	1844	1080	1365	881	87089	100684	72638	84234
	1067	598	727	309	65116	84019	58051	76647	1687	960	1477	851	98627	92831	74603	86883
C80	4149	1958	4069	1219	106392	106765	112382	103216	5523	3141	6650	2227	126977	115781	121677	113650
	4199	1979	4040	1359	101751	103967	105040	111820	6438	2881	6334	1997	124741	117940	121248	114615
	4289	1729	4090	1359	98029	108353	106392	133764	6301	2543	6027	1889	122250	104364	121490	108182
C81	10	30	20	10	60	10	20	10	5232	4261	5122	6003	5692	7554	8035	11340
	20	10	20	20	20	40	10	10	4451	6493	4952	7935	5112	9016	7895	9147
	50	30	30	20	40	20	10	20	6713	23837	15169	13956	23395	38927	13976	25092
C82	927	660	649	511	830	930	751	819	69566	31111	75281	26983	77703	31924	65124	28074
	581	542	522	462	661	849	671	820	71650	31835	67269	27926	77885	32224	65608	27934
	589	650	540	630	670	1028	739	979	75801	31954	67813	26880	82323	30158	66405	25953
C83	60182	36551	46373	37127	39321	29778	39805	38306	41503	36189	42739	37159	36766	39349	43575	47784
	38832	33151	36660	35287	35849	37720	38747	36585	40487	39168	34431	42010	40379	45420	50931	46302
	52453	43799	45781	40578	43570	39198	46854	39582	45832	43446	52429	47489	43489	56843	57519	56364
C84	64672	32195	49248	35595	44715	31311	40897	39176	49775	35946	50459	37003	39286	37874	46415	45316
	40172	32408	38402	33172	37827	36367	42106	35523	40836	38160	37162	40042	42647	47150	53220	47079
	56008	40151	47847	39771	43816	38127	51134	35872	49236	40372	52968	43806	45113	52268	58721	47865
C85	26750	20284	27654	20666	27512	22873	26608	19772	28035	18989	28647	19661	30103	22489	29389	21014
	26218	19321	24751	18599	27332	20885	25674	22140	28818	20845	28335	21056	30776	24578	31772	24457
	25404	17646	26720	17987	28928	22140	32296	21687	29663	21367	29411	20223	32888	21656	26857	21286
C86	31941	22736	30148	21200	39495	49381	42941	46755	27020	27901	29513	26187	47500	58568	48901	53304
	24433	21319	27015	22303	42687	47891	43555	46845	29601	28113	31009	27391	46259	55792	47974	52940
	36198	22453	27195	24151	40644	49422	44118	46252	30204	27530	31623	24861	48031	62455	49998	56239
C87	7278	3579	6469	2510	33009	38195	32204	35966	8951	4027	8658	2958	42814	43175	39624	41851
	7413	3219	6710	2801	35503	37542	35492	37937	8577	4047	9058	2828	39875	39742	42647	41990
	7214	3699	7711	2781	34196	36384	31591	41136	7979	3556	7382	2879	38988	39471	34958	36286
C90	107219	75670	113086	79672	84675	90681	95020	81460	111038	72996	122479	85744	117137	95860	118740	94906
	109989	72622	102090	74272	116069	83831	117119	89189	140306	75223	151989	83604	125709	95301	142502	100517
	92100	69920	125645	72420	107642	86580	109740	105716	108766	87236	129229	88515	138172	97975	111990	94530
C91	883	748	662	458	1024	1290	854	1064	3098	2923	2603	2567	4762	3633	3630	3372
	872	529	404	449	1049	1202	1070	1052	3849	2783	3202	2507	4151	3349	3537	3118
	565	719	560	339	1270	1200	1011	1328	3307	2943	2738	2449	4368	3394	3748	2894
C92	914	668	655	631	1170	1367	1142	1270	434279	386200	395062	351583	503112	381494	458876	342754
	856	692	560	514	1029	1227	1331	1248	527570	384783	492835	355240	488638	398702	452256	354781

inducers	None	Lara	VAI	Lara VAI	IPTG	Lara IPTG	VAI ITPG	Lara VAI ITPG	aTc	Lara aTc	VAI aTc	Lara VAI aTc	IPTG aTc	Lara IPTG aTc	VAI ITPG aTc	Lara VAI ITPG aTc
	708	581	589	543	1232	1336	1130	1443	451652	373347	419428	338240	479389	374973	414138	321498
C93	159687	136889	151271	117909	129515	147661	115256	129688	142783	130866	152895	120364	164355	172568	151682	149800
	145858	111650	121102	102538	139436	146943	127758	134944	160604	138394	163476	122037	158357	155464	156762	153081
	130907	117787	128209	110285	125622	150059	129463	141062	148415	139890	142089	124898	168522	162151	131988	144116
C94	543	447	345	431	655	601	637	674	37734	20809	42211	18133	44713	22062	42491	19758
	406	302	360	353	603	691	605	643	40076	21702	43097	19938	42034	20951	41956	19687
	368	330	469	472	626	690	515	892	38074	23889	39062	20139	47563	27808	40158	18185
C96	47256	30634	44145	26263	36804	30218	34701	30980	38979	30181	45375	29397	44808	42712	48647	34769
	38981	25359	36858	23742	38888	35707	37297	35677	44278	29065	50801	31055	49383	41644	50239	42429
	40056	34444	49003	31236	44214	39451	39700	45050	53189	40713	46905	34583	59405	49654	47032	32537
D03	20	10	30	20	20	40	30	10	80	20	60	70	50	60	40	70
	10	40	20	30	40	20	10	30	10	40	50	40	60	40	60	90
	30	20	10	60	10	20	30	20	60	80	30	40	30	70	70	60
D04	10	40	40	30	18478	58465	21458	56113	30	120	70	190	42924	112167	40769	96892
	20	20	10	40	15500	55901	19090	47932	40	80	50	40	31312	70039	29372	66661
	30	30	40	40	13876	42974	14327	41625	60	70	30	80	31010	79852	31070	73409
D05	20	10	10	10	8876	5342	11441	6964	90	20	70	110	5062	6233	8095	7885
	20	10	10	20	10359	8556	11771	9026	30	80	70	40	2230	2330	3440	2070
	20	60	40	30	1920	2360	2380	2420	40	50	60	30	2500	3340	3110	4021
D07	10	20	30	20	10	30	10	40	210	60	160	80	105598	133491	108520	122946
	20	10	10	10	20	60	10	20	90	100	100	50	72215	79193	87325	74726
	20	20	50	10	30	20	20	30	60	90	70	90	90076	84364	87549	83795
D08	10	20	10	70	200	90	240	80	760	1260	850	1000	238483	204764	219784	168255
	10	10	20	40	150	30	210	60	430	500	500	520	230363	265983	229033	230508
	10	10	10	30	160	110	140	120	430	770	430	520	237131	254998	217772	198795
D09	1070	480	640	340	8866	6083	8656	6073	8626	3370	8435	2680	117671	98031	134646	98498
	970	320	740	300	10088	8065	11691	8115	3741	1670	1910	1560	61586	73936	61910	69989
	150	180	190	190	3170	3140	3110	4061	1580	1870	1680	1660	64275	78747	64437	74918
D10	10	10	20	30	10	30	30	40	50	40	40	80	30377	45594	29201	36219
	10	20	10	80	10	20	40	50	60	30	70	50	22361	25775	27252	23596
	20	30	50	50	50	20	10	10	50	60	110	60	27131	29462	26428	27101
D11	30	10	40	10	20	30	10	50	212403	287248	243063	251722	265888	339400	260456	309987
	10	10	10	40	70	50	10	60	204617	190262	191110	185279	180155	226177	208985	169828
	40	10	60	60	20	20	40	30	189964	186723	158759	180536	208281	244803	205756	207246
D13	2710	460	2590	510	141205	115439	133644	120017	3370	510	3180	470	179815	174275	189850	159570
	3791	310	4541	270	135402	128941	149935	121079	580	330	340	220	98294	121303	110353	112971
	600	230	380	160	95926	105262	93712	102159	790	400	390	290	125305	145323	129207	137100
D14	43065	388168	38161	408748	31553	415257	29171	385299	44284	408246	41906	383434	58576	447204	42682	385310
	36028	357249	32739	358097	38192	401040	32166	413800	60889	280360	43579	248323	56315	290689	43478	245521
	38292	232306	32297	207608	42319	262075	34892	237901	50846	278611	42591	249282	55426	341378	46581	320744
D16	20	30	40	50	10	50	40	50	70	90	80	100	60	40	50	100

inducers	None	Lara	VAI	Lara VAI	IPTG	Lara IPTG	VAI ITPG	Lara VAI ITPG	aTc	Lara aTc	VAI aTc	Lara VAI aTc	IPTG aTc	Lara IPTG aTc	VAI ITPG aTc	Lara VAI ITPG aTc
	10	40	10	30	10	40	10	40	100	70	60	110	60	40	100	140
	50	40	60	10	10	30	10	40	90	100	60	110	70	80	170	110
D17	55214	69797	48043	46319	38947	96313	50059	71092	59465	45966	53337	47528	46642	83086	47841	51774
	43629	67915	44596	56244	57445	54669	54094	61940	60344	49131	60586	47236	66499	57658	64740	51360
	49182	43357	51713	39561	52561	49000	53105	51441	52450	48567	54922	47801	62304	74959	58142	66034
D18	40	10	10	20	40	40	10	40	1986291	1915247	2209744	2399427	1917144	1773607	1872940	1912606
	10	10	10	10	40	30	20	20	2052013	2376710	2325011	2477130	2375257	2442023	2271526	2468000
	70	40	20	40	20	40	20	30	2319984	2372351	2317496	2522016	2315013	2459564	2256303	2343586
D19	62092	53398	42994	24550	28457	18528	27704	21579	26267	27724	16944	33212	14637	13655	10909	14607
	20515	11982	28236	9798	16061	23064	12513	26227	5662	4021	4231	4331	3430	4311	4311	2940
	4511	3991	4371	3561	4321	5442	3290	3661	4251	4932	4031	4041	5472	6383	4922	6323
D20	40	40	40	20	20	10	20	30	70	20	120	40	80	100	110	40
	20	10	30	30	20	80	20	80	50	70	40	40	30	60	80	140
	10	70	30	30	20	30	40	30	70	60	60	60	120	50	90	70
D24	5332	6643	4882	7704	4801	6443	5892	7084	7104	7074	7875	6613	5952	7014	5742	7715
	7124	5362	7644	5632	6703	8756	7254	8726	2790	2390	2520	2600	1900	2810	2270	2510
	1810	2080	1870	2150	1620	2510	1840	2560	1930	2480	1990	2970	3140	3861	2920	3631
D25	10	10	40	40	780	470	780	480	90	60	50	100	600	750	630	850
	40	30	10	80	360	720	580	890	50	70	50	90	190	260	370	340
	40	20	50	30	240	330	310	300	100	60	70	80	300	320	380	380
D27	37447	61829	39450	57607	29724	73582	41101	69776	80865	79193	72964	77237	74483	106322	82021	92696
	36551	56567	42360	60758	40275	51330	42098	56759	91619	75323	94778	80034	88482	91264	98701	89619
	55911	48033	57859	51612	56143	57738	62061	60031	88716	78818	90624	81737	89112	96862	85936	94971
D28	97797	91447	94768	88848	116723	166692	156449	164779	113216	95266	121089	102983	161542	171167	162673	155710
	94484	91031	117814	96729	120783	148694	120293	171680	147823	109823	145917	109304	200821	177395	200873	178795
	121507	88097	117763	94534	162077	149586	170898	157517	143755	107960	142270	112676	190882	183001	197348	171094
D29	10	10	30	10	20	20	10	40	6173	8245	5892	10349	9327	17947	9056	13214
	40	10	30	20	40	60	40	10	6373	9958	4511	8486	3100	19010	11481	25022
	30	30	10	20	10	10	10	20	5382	24912	13896	10649	15831	98366	10669	840
D31	20	30	40	10	40	30	10	40	70	70	100	190	100	50	110	90
	20	10	10	30	30	30	10	70	80	110	50	80	130	110	80	110
	30	50	10	60	70	20	30	40	80	100	130	140	100	80	170	110
D32	370	370	260	430	280	400	570	360	430	390	400	370	230	440	300	400
	280	270	470	310	180	270	280	310	150	200	180	110	220	190	180	180
	80	60	140	90	80	110	150	70	230	180	230	210	200	180	240	250
D34	197059	331576	157650	323688	141973	400068	217451	392518	226125	318046	245458	348099	236049	396556	252202	314789
	164800	338637	201514	362225	172545	258671	204607	327603	328438	303747	353166	315242	362937	348132	337250	338837
	303904	240273	323445	274747	267122	289629	316834	336278	329695	285936	365285	305051	307713	379580	286367	365072
D35	50594	50765	51441	50735	41866	54609	41856	48395	91995	95784	96343	88848	110088	96598	97166	76630
	45110	52783	45483	46027	46964	36109	45896	30136	105903	128532	100552	117182	113084	128155	110964	112859
	66206	60960	68097	59071	63658	61061	63972	57092	117468	120283	109864	112126	105781	122243	93214	102922

inducers	None	Lara	VAI	Lara VAI	IPTG	Lara IPTG	VAI ITPG	Lara VAI ITPG	aTc	Lara aTc	VAI aTc	Lara VAI aTc	IPTG aTc	Lara IPTG aTc	VAI ITPG aTc	Lara VAI ITPG aTc
D36	57819	48053	53035	52419	53287	55477	65973	56093	105008	86311	90350	86696	81179	78656	86220	81342
	51158	40225	58718	43276	51673	62708	60768	67166	43810	45291	43397	48809	37910	51683	42309	47408
	25363	29985	24078	30809	27804	33825	27131	35666	37980	51764	37799	52147	50523	67632	52510	64052
D37	60	20	40	40	20	70	50	60	60	60	60	80	100	80	80	120
	20	10	20	30	10	50	10	70	90	110	80	110	80	100	70	160
	20	10	40	40	20	30	50	50	60	50	80	160	70	90	130	140
D38	28668	44697	26910	36219	26750	53106	28879	44495	46742	83968	43528	69321	62031	100929	61829	79933
	19170	46601	24440	39752	33413	38574	33735	32639	58294	61435	53307	50916	54659	68461	53216	58405
	30286	28638	28940	27784	29643	36028	29040	30668	49273	56759	48305	52914	52944	77359	51713	69098
D40	144738	189252	130045	154550	136292	319702	183104	295507	158276	169592	158964	156192	272487	312611	255603	240190
	104550	205870	133633	184918	161255	222097	173060	234614	262514	234915	237496	214631	345483	333499	327741	319785
	224121	201990	219939	189086	287342	278842	275533	295403	231517	220541	240211	209565	321578	365201	313041	325228
D41	33111	10008	18207	11901	19030	10599	29884	11681	16743	12974	20896	16512	7334	11210	8235	15901
	11430	7985	20364	8075	12823	14768	13294	19040	3871	3741	3641	4771	3410	4611	3721	3771
	2620	3420	3000	3300	2930	4441	2710	4191	2740	4781	3170	4841	4511	7925	5402	8786
D42	6813	8486	6493	7004	6233	8626	8405	7995	8165	7634	8145	7244	6703	7855	7334	7084
	6493	6463	8626	6613	6163	7274	5832	7704	6493	4701	5682	4411	5412	5722	5862	4641
	4771	3651	5112	3481	4872	4791	4952	4071	4751	4982	4942	4211	6173	6203	6403	5182
D43	81402	129984	58182	112666	50251	127695	65832	122477	77622	120967	81899	118069	87721	128299	91934	108846
	59162	119140	70110	115632	59758	99352	62607	111779	123835	110149	119110	96963	104753	113288	114704	101782
	85317	82497	91914	81859	89721	93844	92655	89477	106820	105272	108754	99077	120334	120498	116254	116600
D44	9016	8446	7164	7905	9217	9267	12733	8696	10228	10739	10609	9006	6994	9507	9026	8976
	7885	6353	9036	6003	10539	10098	10559	10248	4251	3851	2930	3571	2890	3831	3631	3390
	2730	2410	2300	2520	2490	3030	2360	2850	2590	3491	2430	3360	4071	5332	4211	5322
D45	40	10	30	50	60	30	40	30	137909	285612	143550	302969	108031	225234	137408	251681
	30	10	20	10	70	50	30	80	63628	71446	59496	74402	50130	82335	66974	68603
	30	10	20	40	20	30	10	30	49656	73359	49363	69685	64467	101345	68087	106616
D46	40	60	80	20	40	80	40	80	415556	397440	447297	380068	573376	472058	521910	357439
	60	10	20	80	70	60	50	90	608589	426887	580425	402097	516039	463004	590572	420410
	30	50	40	50	40	40	110	60	475915	417323	490999	391654	577301	472845	581205	448428
D49	190	80	147382	132835	280	130	173904	150109	190	170	145814	100196	220	270	235883	178260
	120	70	131026	114123	130	130	158256	119426	150	200	145558	146060	140	180	166074	164954
	70	30	135259	130402	40	90	137899	149237	70	120	146285	135402	110	150	159981	163844
D50	50	50	50	120	50	100	120	90	155987	201245	157814	180536	204482	221777	212206	202342
	60	60	60	80	30	100	40	50	164451	147884	174810	148940	163351	166168	177219	158974
	30	30	50	80	30	40	60	90	148582	145824	158040	144390	175314	179486	188302	186981
D51	40487	39369	35304	35062	37377	39651	40134	38403	42632	42239	38252	36411	34268	47408	40114	42138
	32508	31362	35928	30970	32407	41675	34348	40396	32759	29623	29000	27563	26197	28096	24711	25544
	20916	21960	20133	22151	21779	25554	19822	22251	25514	28467	23978	25846	26097	35093	26920	35254
D52	330	230	230	210	260	250	420	290	280	290	360	370	190	260	230	260
	160	170	320	130	160	290	110	310	140	90	130	150	130	110	170	210

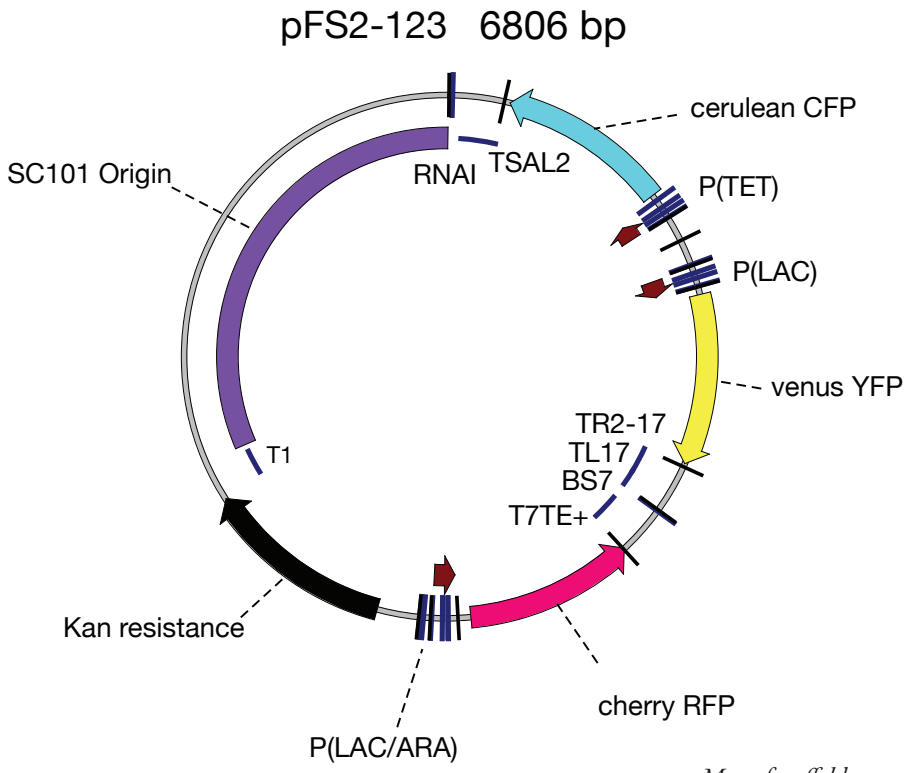
inducers	None	Lara	VAI	Lara VAI	IPTG	Lara IPTG	VAI ITPG	Lara VAI ITPG	aTc	Lara aTc	VAI aTc	Lara VAI aTc	IPTG aTc	Lara IPTG aTc	VAI ITPG aTc	Lara VAI ITPG aTc
	70	100	130	90	90	180	120	140	160	200	130	170	280	220	170	310
D53	60	90	60	50	32830	152612	43448	140529	180	110	160	80	56385	146194	63224	112625
	100	90	60	20	39410	79396	42179	89995	140	90	100	90	92482	179507	92614	168285
	50	30	80	40	75404	147567	77967	157876	60	80	50	100	86524	194012	86767	177291
D54	20555	36179	18930	31814	15339	36853	18889	34670	25946	34338	25926	30276	29713	37739	27141	29030
	18338	34459	21227	31261	19040	26157	19481	28437	38353	31583	34882	28538	37135	34157	34992	29171
	29734	24480	28488	23014	28508	27573	28689	24098	33021	29442	33172	27905	35777	38665	34781	31613
D56	10	10	40	40	2330	4992	2930	4751	90	70	120	60	144461	205105	147936	158266
	10	20	10	20	3010	3491	3100	3501	110	120	70	80	222170	259799	232233	232555
	10	20	40	30	4561	4241	4751	4531	140	80	60	70	223208	235925	218736	224079
D57	60	50	50	100	510	800	590	660	70	110	60	140	580	840	660	820
	40	20	50	50	480	640	570	880	50	80	50	70	350	310	400	350
	40	40	70	30	180	190	220	270	120	110	40	80	310	350	270	380
D60	252108	501102	292273	458784	224817	506469	271787	497528	294961	491996	328374	445532	430472	540552	357026	483024
	268952	445706	250451	444854	250273	432341	266558	425117	505740	437702	485538	421449	405400	486166	454482	434274
	338117	372342	365104	359211	356941	412451	354141	410140	409038	440777	418299	434757	463975	557393	462334	525920
D61	30	120	10	120	200	73491	60	87275	30	100	10	40	160	102210	100	70110
	50	50	30	110	160	105425	180	72934	10	140	30	70	70	99769	240	92706
	40	130	10	60	220	92097	190	61748	40	110	30	60	130	120089	150	106382
D62	60	30	70	60	80	110	60	70	119313	139598	141194	124365	125734	143551	133316	148684
	30	60	20	80	70	80	20	70	74635	74391	64396	70353	63800	81230	67288	73663
	50	70	70	50	70	80	60	40	57809	76913	64406	73227	72205	103665	77926	104295
D64	133459	89457	116254	90309	130219	95530	132703	92289	108591	108479	130147	103593	77217	105741	75131	108612
	95235	70889	102322	69857	102047	101203	100338	115662	42813	49787	35173	51915	34399	54902	38071	45886
	29914	41665	31221	39440	31804	53196	31191	44193	32478	59556	35264	50382	44848	78413	47549	79760
D65	30	40	70	80	1620	1170	2300	1230	70	100	70	110	1550	1150	1630	1290
	30	30	40	30	1340	1330	1300	1670	60	60	100	60	1170	390	1050	460
	20	40	20	20	870	370	680	400	80	60	80	50	1140	570	1150	540
D66	4922	3831	3891	4431	5062	4081	6813	5232	4351	5072	5572	5182	3721	4821	3891	4511
	4141	3020	5542	3350	4091	4721	3801	5732	2040	2000	1980	2100	1690	1850	1990	1910
	1090	1570	1340	1450	1310	1440	1340	1480	1290	1620	1540	1980	1960	2460	1840	2920
D68	100	10	50	70	183928	240586	209824	223540	140	80	130	60	274957	271401	287132	245219
	140	10	110	40	212962	226696	211430	236850	70	130	110	100	158153	182980	167000	157321
	60	40	50	20	123273	141471	111678	123324	70	120	150	90	159827	209566	163207	189953
D69	91853	143632	63264	138594	60132	166887	75931	158471	98996	135290	94748	113929	166373	166744	130249	140140
	74888	152673	75222	152642	73146	121313	73774	121252	162632	161727	162868	162600	146490	172370	172349	160844
	131506	139976	143663	140437	138677	149760	138155	150201	150058	158030	152119	149996	160546	161985	159036	162374
D70	30	40	30	60	40	30	60	10	40	50	60	110	70	50	50	80
	10	40	10	90	70	80	10	40	40	40	40	80	20	50	90	90
	30	50	60	50	20	30	70	80	60	70	80	60	30	50	70	100
D71	22462	29884	19782	26719	16643	29312	19261	30990	24460	25655	24249	23275	33483	34308	29422	30749

inducers	None	Lara	VAI	Lara VAI	IPTG	Lara IPTG	VAI ITPG	Lara VAI ITPG	aTc	Lara aTc	VAI aTc	Lara VAI aTc	IPTG aTc	Lara IPTG aTc	VAI ITPG aTc	Lara VAI ITPG aTc
	18930	24550	19301	24801	16181	24108	19090	23988	22843	20565	20615	19602	18819	21127	20675	19471
	16232	15620	16563	15299	16492	16111	15841	18478	17505	18709	17355	18639	21679	21107	20083	20746
D72	232629	326422	245927	290426	219898	339568	252265	337673	267906	335792	332335	268115	330349	335794	338657	304462
	258034	226593	230748	252849	251899	323372	330255	305198	197761	222191	182000	211036	159262	220106	175840	210124
	147966	185764	140939	176560	141297	187342	132283	183000	151729	221900	150468	221423	187734	262724	186692	272916
D73	23636	31583	18879	29925	29714	21007	24159	18729	19231	33282	18538	31493	16462	27111	15901	26217
	21127	16462	17004	12402	16883	31855	14026	30256	9147	9317	5352	6403	7634	11861	6273	6343
	5672	7975	4391	5382	7044	10128	5462	6223	6894	11962	5832	6483	8916	15229	8115	10759
D77	570	80	480	140	1160	180	1440	180	132263	142311	143785	145886	101314	136825	119579	145630
	520	140	630	210	440	200	350	140	59435	66519	46410	63961	50301	69948	52752	62526
	110	90	110	140	140	150	60	100	43669	70130	48285	62223	60536	95581	63173	98650
D78	2060	2570	2000	2110	3230	1480	4091	1990	2310	2380	2630	2710	1520	1610	1230	1820
	1560	970	2730	760	1480	1780	1760	2410	680	530	630	780	440	480	740	830
	300	230	290	270	240	340	270	360	490	650	500	570	810	890	730	1020
D79	20	40	40	50	20	30	70	30	1849937	1658333	2057674	1889256	2077786	1751911	2093520	1737058
	10	10	20	30	30	60	10	30	2117706	2178466	2211959	2243646	1923685	2217220	2054456	2120724
	10	10	10	30	10	20	10	20	1828671	2233413	1984362	2267458	2161418	2596919	2374815	2685795
D80	710	990	97075	77632	920	940	88422	72245	870	1080	116539	51542	750	880	96994	69260
	420	350	71223	46027	420	990	96191	65771	400	370	52298	53650	260	350	66115	51693
	90	40	50261	41544	170	140	49444	47095	200	410	48456	46752	410	400	62809	64032
D81	160	140	130	210	250	180	370	260	150	100	190	230	120	140	180	210
	90	180	180	130	100	200	80	240	60	60	90	110	100	140	120	190
	50	60	70	80	60	40	70	90	60	70	90	90	60	60	80	140
D83	14186	13846	14658	12402	13184	16813	16823	16643	15279	16161	16863	14257	17445	15580	17204	14748
	13525	11070	13966	10459	14447	15740	16974	14768	9076	9317	9257	8225	7014	9868	9417	8496
	6473	7564	6803	6964	6493	7795	5942	7694	7865	9127	7034	8926	8736	11080	8556	10960
D85	123089	97065	96414	93285	127940	104560	126817	102484	151739	90868	103949	77248	137264	140469	124263	131783
	95479	73592	91802	74695	100572	94768	99596	87691	114959	101355	86169	90076	115244	128206	109467	116539
	94727	75050	76944	75850	109355	99586	98376	98966	113033	92960	96089	87031	139096	119967	122732	99616
D86	22803	22964	21498	25384	28709	29543	28809	35455	20756	21067	20164	22793	30879	29804	32508	30397
	18759	14387	17886	16613	28960	30317	28920	35052	18057	11110	20485	12322	25655	18217	26247	18639
	15289	10218	19150	11370	22843	15810	25795	18047	16563	10869	18779	11661	25645	17275	28789	18107
D87	1210	1110	810	790	36773	26448	38152	26960	760	940	900	980	30196	28719	30427	28719
	910	330	1230	360	27141	25785	28357	28498	370	200	210	220	17224	18769	19822	17565
	230	140	180	150	20826	16161	18679	17054	300	270	390	220	26036	21910	26026	17736
D89	60	100	70	100	110	100	100	70	160	190	270	180	160	150	110	160
	30	70	20	60	70	100	70	140	120	110	140	130	130	90	110	110
	30	10	30	40	30	30	40	60	80	170	120	80	160	140	150	170
D90	40	10	40	30	87863	78220	95520	87833	150	70	150	150	78706	81859	75951	91376
	30	10	20	30	60516	74017	67875	91122	110	110	100	100	40517	63082	51431	61778
	10	10	20	20	43025	57163	45442	61981	80	140	90	160	57658	75060	61041	75698

inducers	None	Lara	VAI	Lara VAI	IPTG	Lara IPTG	VAI ITPG	Lara VAI ITPG	aTc	Lara aTc	VAI aTc	Lara VAI aTc	IPTG aTc	Lara IPTG aTc	VAI ITPG aTc	Lara VAI ITPG aTc
D91	40	20	30	20	930	41886	880	40376	430	470	590	510	1680	39973	1620	38876
	10	10	10	30	1230	38262	1220	37517	390	490	470	500	1280	24801	1100	23476
	50	20	20	10	710	21569	730	23024	520	480	430	500	1350	27935	1440	22763
D92	1870	2460	1640	2310	2300	1440	2440	1610	2110	1860	2090	1850	930	1290	1220	1160
	1310	740	2080	860	1300	1420	1170	1580	610	540	530	620	350	640	660	670
	380	340	420	670	490	500	430	660	520	610	510	650	650	800	740	960
D93	1410	2020	1660	2010	1690	1680	2200	1880	2200	2240	1970	2520	790	1420	1010	1440
	1130	1010	1500	1160	970	1500	1210	2190	730	430	680	390	610	350	660	440
	630	300	630	290	480	420	600	370	620	420	670	380	590	480	680	530
D94	12954	18879	16422	20404	125785	159540	149545	142730	19030	18749	20033	24982	124641	128687	131047	141205
	12132	11982	15770	12583	88797	140857	102485	146859	6113	6813	7815	8946	45856	74250	78200	93549
	6673	6563	6924	9387	62435	78717	68239	86169	6213	8235	7715	10799	81169	107125	94849	119181
D95	28136	311030	27644	369609	23165	282771	28277	326442	32920	373481	35716	389883	18869	248428	17746	250263
	24389	200181	28528	204979	19090	287174	20836	333044	17666	180588	15309	189592	14447	187683	13525	178136
	14497	125060	13385	140836	12693	140365	13515	156326	16482	173306	16583	183443	14186	204070	14818	181443
D96	37407	30568	33664	34751	39903	32166	44908	40537	57284	39832	49817	45331	30578	27312	32277	27965
	35928	18147	41474	22371	29724	29804	35676	35928	21067	15480	18488	17736	10288	17375	19592	20174
	15480	13896	17595	15871	18308	15069	17555	19221	15981	18167	18930	19511	25564	21990	25153	19120

Appendix B: The Reporter Scaffold

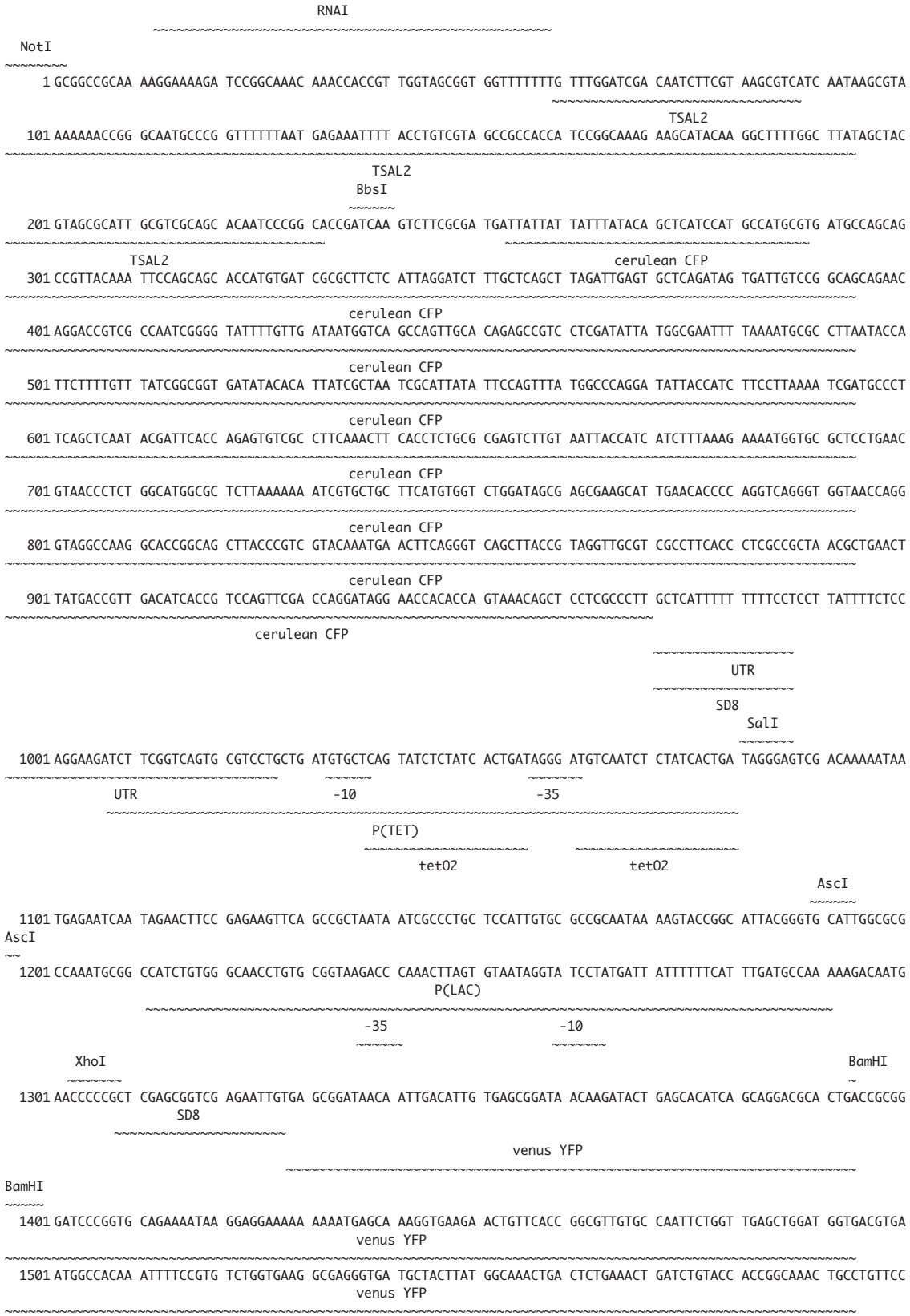
This appendix contains the complete sequence of the 3-color reporter plasmid pFS2-123 described in Chapter 3. Here the beginning of the scaffold is numbered as the first base pair, corresponding to the description of the genetic features in the Chapter 3 Methods and Materials. A map of the plasmid shows the layout of the relevant genetic features. The replacement promoter sequence used to remove LacI repression of yfp is shown separately.



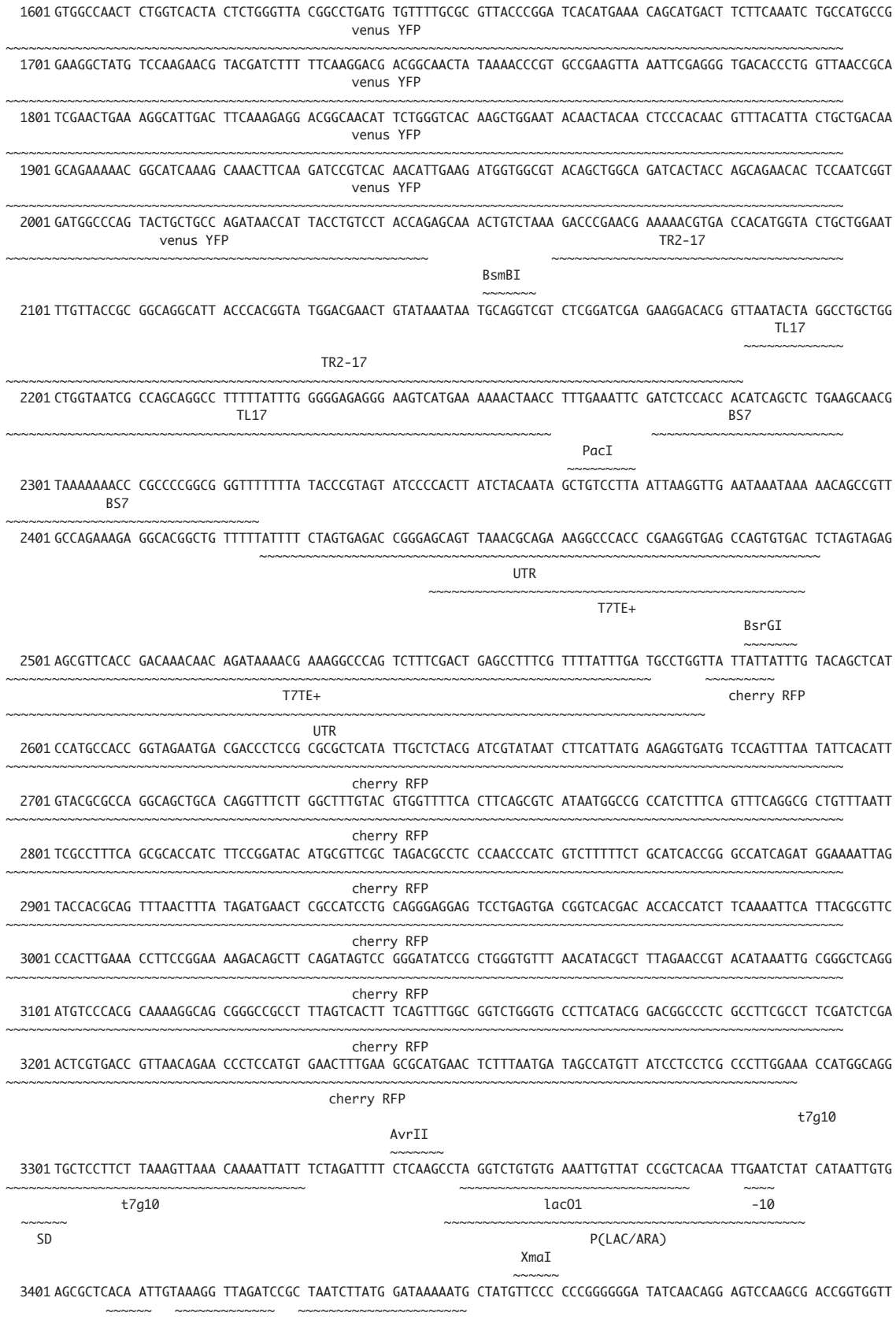
Map of scaffold sequence

```
cccgggcatacgttaaatctatcaccgcaagggataaatatctaacaccgtgcgtgTTGACTattttacctctggcggtGATAATggttgcatgcctagg
XmaI site                                              -35 box              -10 box        AvrII site
~~~~~~                                                 ~~~~~~               ~~~~~~         ~~~~~~
```

Lambda P(R) promoter inserted between XmaI and AvrII sites in plasmid pFS2-12R



The pFS2-123 Sequence



The pFS2-123 Sequence

```
              -35         ara2                 ara1
~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
                        P(LAC/ARA)
          NheI
         ~~~~~~~
 3501 GCATGTCTAG CTAGCTAGAA CAGGACTAGC TAATGGTTTC TTAGACGTCG GAATTGCCAG CTGGGGCGCC CTCTGGTAAG GTTGGGAAGC CCTGCAAAGT
                                                                                        Kan resistance
                                                                                        ~~~~~~~~~~~~~~~~
 3601 AAACTGGATG GCTTTCTTGC CGCCAAGGAT CTGATGGCGC AGGGGATCAA GATCTGATCA AGAGACAGGA TGAGGATCGT TTCGCATGAT TGAACAAGAT
                                          Kan resistance
~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
 3701 GGATTGCACG CAGGTTCTCC GGCCGCTTGG GTGGAGAGGC TATTCGGCTA TGACTGGGCA CAACAGACAA TCGGCTGCTC TGATGCCGCC GTGTTCCGGC
                                          Kan resistance
~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
 3801 TGTCAGCGCA GGGGCGCCCG GTTCTTTTTG TCAAGACCGA CCTGTCCGGT GCCCTGAATG AACTGCAGGA CGAGGCAGCG CGGCTATCGT GGCTGGCCAC
                                          Kan resistance
~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
 3901 GACGGGCGTT CCTTGCGCAG CTGTGCTCGA CGTTGTCACT GAAGCGGGAA GGGACTGGCT GCTATTGGGC GAAGTGCCGG GGCAGGATCT CCTGTCATCT
                                          Kan resistance
~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
 4001 CACCTTGCTC CTGCCGAGAA AGTATCCATC ATGGCTGATG CAATGCGGCG GCTGCATACG CTTGATCCGG CTACCTGCCC ATTCGACCAC CAAGCGAAAC
                                          Kan resistance
~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
 4101 ATCGCATCGA GCGAGCACGT ACTCGGATGG AAGCCGGTCT TGTCGATCAG GATGATCTGG ACGAAGAGCA TCAGGGGCTC GCGCCAGCCG AACTGTTCGC
                                          Kan resistance
~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
 4201 CAGGCTCAAG GCGCGCATGC CCGACGGCGA GGATCTCGTC GTGACCCATG GCGATGCCTG CTTGCCGAAT ATCATGGTGG AAAATGGCCG CTTTTCTGGA
                                          Kan resistance
~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
 4301 TTCATCGACT GTGGCCGGCT GGGTGTGGCG GACCGCTATC AGGACATAGC GTTGGCTACC CGTGATATTG CTGAAGAGCT TGGCGGCGAA TGGGCTGACC
                                Kan resistance
~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
 4401 GCTTCCTCGT GCTTTACGGT ATCGCCGCTC CCGATTCGCA GCGCATCGCC TTCTATCGCC TTCTTGACGA GTTCTTCTGA GCGGGACTCT GGGGTTCGAG
 4501 AGCTCGCTTG GACTCCTGTT GATAGATCCA GTAATGACCT CAGAACTCCA TCTGGATTTG TTCAGAACGC TCGGTTGCCG CCGGGCGTTT TTTATTGGTG
~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
                                                T1
 4601 AGAATCCAAG CACTAGGGAC AGTAAGACGG GTAAGCCTGT TGATGATACC GCTGCCTTAC TGGGTGCATT AGCCAGTCTG AATGACCTGT CACGGGATAA
~~~~~~~~~~~~~~~~~                                          ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
      T1                                                                   SC101 Origin
 4701 TCCGAAGTGG TCAGACTGGA AAATCAGAGG GCAGGAACTG CTGAACAGCA AAAAGTCAGA TAGCACCACA TAGCAGACCC GCCATAAAAC GCCCTGAGAA
~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
                                          SC101 Origin
 4801 GCCCGTGACG GGCTTTTCTT GTATTATGGG TAGTTTCCTT GCATGAATCC ATAAAAGGCG CCTGTAGTGC CATTTACCCC CATTCACTGC CAGAGCCGTG
~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
                                          SC101 Origin
 4901 AGCGCAGCGA ACTGAATGTC ACGAAAAAGA CAGCGACTCA GGTGCCTGAT GGTCGGAGAC AAAAGGAATA TTCAGCGATT TGCCCGAGCT TGCGAGGGTG
~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
                                          SC101 Origin
 5001 CTACTTAAGC CTTTAGGGTT TTAAGGTCTG TTTTGTAGAG GAGCAAACAG CGTTTGCGAC ATCCTTTTGT AATACTGCGG AACTGACTAA AGTAGTGAGT
~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
                                          SC101 Origin
 5101 TATACACAGG GCTGGGATCT ATTCTTTTTA TCTTTTTTTA TTCTTTCTTT ATTCTATAAA TTATAACCAC TTGAATATAA ACAAAAAAAA CACACAAAGG
~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
                                          SC101 Origin
 5201 TCTAGCGGAA TTTACAGAGG GTCTAGCAGA ATTTACAAGT TTTCCAGCAA AGGTCTAGCA GAATTTACAG ATACCCACAA CTCAAAGGAA AAGGACTAGT
~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
                                          SC101 Origin
 5301 AATTATCATT GACTAGCCCA TCTCAATTGG TATAGTGATT AAAATCACCT AGACCAATTG AGATGTATGT CTGAATTAGT TGTTTTCAAA GCAAATGAAC
~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
                                          SC101 Origin
 5401 TAGCGATTAG TCGCTATGAC TTAACGGAGC ATGAAACCAA GCTAATTTTA TGCTGTGTGG CACTACTCAA CCCCACGATT GAAAACCCTA CAAGGAAAGA
~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
                                          SC101 Origin
 5501 ACGGACGGTA TCGTTCACTT ATAACCAATA CGCTCAGATG ATGAACATCA GTAGGGAAAA TGCTTATGGT GTATTAGCTA AAGCAACCAG AGAGCTGATG
~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
                                          SC101 Origin
 5601 ACGAGAACTG TGGAAATCAG GAATCCTTTG GTTAAAGGCT TTGAGATTTT CCAGTGGACA AACTATGCCA AGTTCTCAAG CGAAAAATTA GAATTAGTTT
~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
                                          SC101 Origin
 5701 TTAGTGAAGA GATATTGCCT TATCTTTTCC AGTTAAAAAA ATTCATAAAA TATAATCTGG AACATGTTAA GTCTTTTGAA AACAAATACT CTATGAGGAT
~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
                                          SC101 Origin
 5801 TTATGAGTGG TTATTAAAAG AACTAACACA AAAGAAAACT CACAAGGCAA ATATAGAGAT TAGCCTTGAT GAATTTAAGT TCATGTTAAT GCTTGAAAAT
~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
                                          SC101 Origin
```

```
  5901 AACTACCATG AGTTTAAAAG GCTTAACCAA TGGGTTTTGA AACCAATAAG TAAAGATTTA AACACTTACA GCAATATGAA ATTGGTGGTT GATAAGCGAG
~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
                                            SC101 Origin
  6001 GCCGCCCGAC TGATACGTTG ATTTTCCAAG TTGAACTAGA TAGACAAATG GATCTCGTAA CCGAACTTGA GAACAACCAG ATAAAAATGA ATGGTGACAA
~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
                                            SC101 Origin
  6101 AATACCAACA ACCATTACAT CAGATTCCTA CCTACGTAAC GGACTAAGAA AAACACTACA CGATGCTTTA ACTGCAAAAA TTCAGCTCAC CAGTTTTGAG
~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
                                            SC101 Origin
  6201 GCAAAATTTT TGAGTGACAT GCAAAGTAAG CATGATCTCA ATGGTTCGTT CTCATGGCTC ACGCAAAAAC AACGAACCAC ACTAGAGAAC ATACTGGCTA
~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
                                            SC101 Origin
  6301 AATACGGAAG GATCTGAGGT TCTTATGGCT CTTGTATCTA TCAGTGAAGC ATCAAGACTA ACAAACAAAA GTAGAACAAC TGTTCACCGT TAGATATCAA
~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
                                            SC101 Origin
  6401 AGGGAAAACT GTCCATATGC ACAGATGAAA ACGGTGTAAA AAAGATAGAT ACATCAGAGC TTTTACGAGT TTTTGGTGCA TTTAAAGCTG TTCACCATGA
~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
                                            SC101 Origin
  6501 ACAGATCGAC AATGTAACAG ATGAACAGCA TGTAACACCT AATAGAACAG GTGAAACCAG TAAAACAAAG CAACTAGAAC ATGAAATTGA ACACCTGAGA
~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
                                            SC101 Origin
  6601 CAACTTGTTA CAGCTCAACA GTCACACATA GACAGCCTGA AACAGGCGAT GCTGCTTATC GAATCAAAGC TGCCGACAAC ACGGGAGCCA GTGACGCCTC
~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
                                            SC101 Origin
  6701 CCGTGGGGAA AAAATCATGG CAATTCTGGA AGAAATAGCG CTTTCAGCCG GCAAACCTGA AGCCGGATCT GCGATTCTGA TAACAAACTA GCAACACCAG
~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
                                            SC101 Origin
  6801 AACAGC
~~~~~~
......
```

The pFS2-123 Sequence

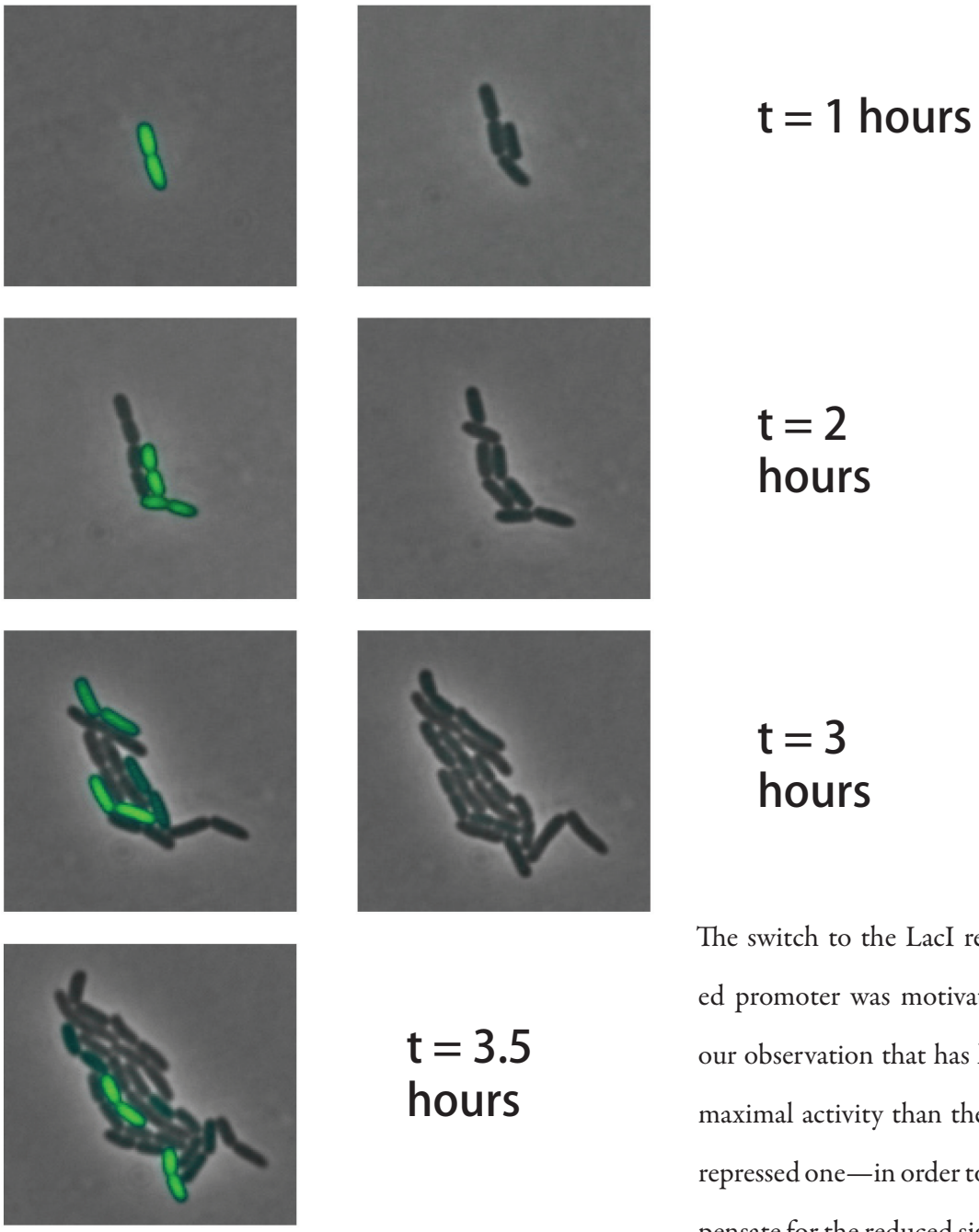
Appendix C: Repressilator Experiments

This appendix describes two simple experiments with the repressilator, an oscillatory network composed of three repressor TFs: LacI represses TetR represses λ cI represses LacI. In the original experiment (Elowitz and Leibler, 2000), TetR also repressed the expression of GFP, where GFP was tagged with a moderate ssrA degradation signal (ASV, the repressors were tagged with a stronger signal LVA). The behavior of the 'classic' repressilator is noisy and unsynchronized: genetically identical daughter cells from the same parent rapidly become uncorrelated. The first experiment below describes a somewhat more appealing "christmas tree" version of the repressilator circuit. The second experiment confirms some underlying assumptions of the original design.

Making the repressilator "blink"

One small mystery of the repressilator circuit was that the magnitude of the fluorescent GFP signal increased over time. This must be due to a changing environment: in the experimental protocol, the cells are first grown in liquid media before they are sandwiched between a slab of (the same) media with 1.5% low melting-point agarose and a coverslip, sealed, and grown within a temperature controlled chamber at 32°C. Since the cell culture is very dilute, some difference between the liquid- and solid-phase environments creates a steady increase in the average cellular fluorescence during microcolony growth. The resulting oscillations look more like the waves of an incoming tide than the steady pulse of a sine wave.

To make the oscillations more regular, I simply strengthened the GFP degradation tag. For this I used a different reporter plasmid pZE12-*cfplva* containing: the same origin of replication (ColE1), an ampicillin resistance marker (instead of kanamycin), a LacI controlled promoter (instead of a TetR controlled one) (Lutz and Bujard, 1997), CFP instead of GFP, and a strong ssrA degradation tag (LVA) instead of a moderate one (ASV). I used CFP for this experiment because it is very photostable—I didn't want to accidentally bleach the reporter to get the desired result.



The switch to the LacI regulated promoter was motivated by our observation that has higher maximal activity than the TetR repressed one—in order to compensate for the reduced signal of the destabilized CFP. To accommodate this new reporter plasmid, I swapped the antibiotic resistance marker in the original repressilator (ampicillin) for a spectinomycin resistance gene. Except for the antibiotics, the repressilator experiment was repeated as described (Elowitz and Leibler, 2000). With the new reporter plasmid, the oscillations become more pulse-like. Some daughter cells appeared to completely repress CFP expression, even after several hours of growth (~50 min

/ division). These cells appeared to be "stuck" in a non-fluorescent state. No cells were found to persist in a fluorescent state for more than 2-3 cell cycles. This experiment revealed that sister cells carrying the repressilator can exhibit different oscillation periods, and that the noisy phenotype is exacerbated by asymmetry in the circuit components.

Three color oscillations

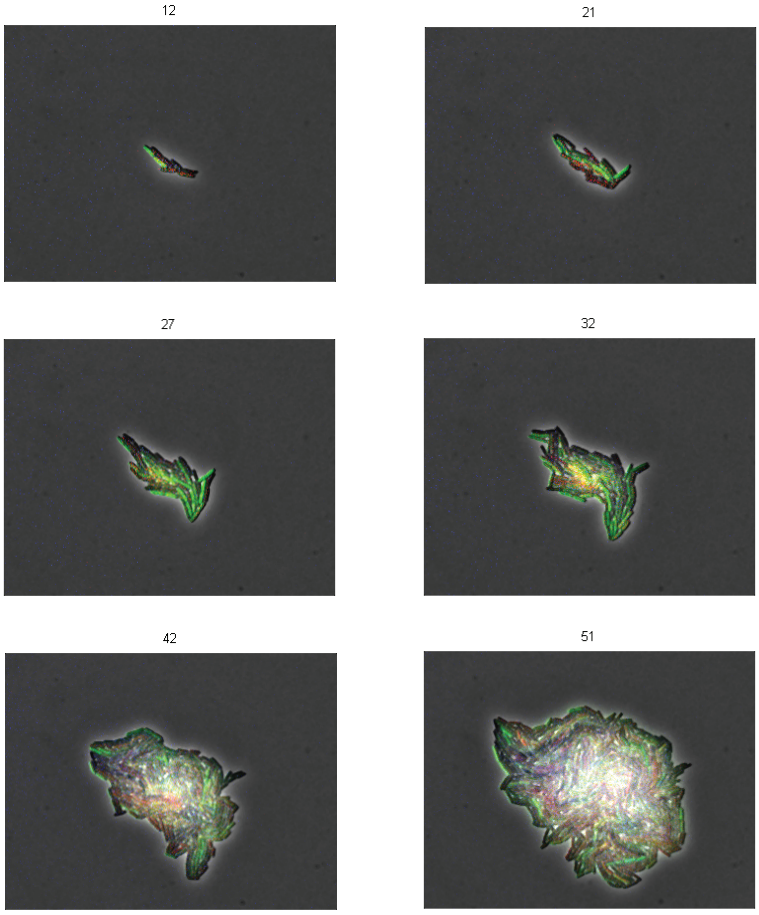
The repressilator is a remarkably robust circuit. It is also genetically stable: after multiple generations, almost every cell containing the repressilator plasmids will oscillate when grown in a microcolony. The above experiments make use of promoter fusions controlling GFP variants to observe the repressor oscillations. Ideally, one would like to construct a repressilator circuit in which each repressor is tagged (protein fusion) with a different, distinguishable fluorescent protein. With only one fluorescent protein, we must infer that the reporter oscillations are representative of the underlying circuit behavior. With this simple three-repressor circuit, there are very few ways in which oscillations could be generated—we assume that LacI represses TetR, TetR represses cI, and cI represses LacI.

To verify this assumption, I constructed a repressilator containing three (stable) reporter genes with three promoter fusions. The reporter plasmid, pFE2-12R, is a variant of one of the constructs characterized in Chapter 3 (it only differs in the origin of replication, in this case it is ColE1). To avoid saturation of protein expression (Introduction, Figure 4), no degradation tags were used on the reporters. Here, TetR represses the *cfp* gene, LacI represses the *yfp* gene, and cI represses the *rfp* gene. To represent these colors, I use blue (*cfp*), green (*yfp*), and red (*rfp*).

The period of the repressilator is about 2 cell cycles, so it is possible to observe oscillations with these stable reporters: after a promoter is shut off, its reporter concentration will dilute to about a fourth of its maximal value before production starts anew. If we begin with a cell where cI concentration is high and LacI concentration is low, *yfp* will be high and *rfp* will be low (and the cell is green). In this state the TetR concentration will increase, so both *cfp* and cI will be on the decline.

As the cI concentration drops, *rfp* and LacI production will increase until the cell becomes red. This increase in LacI results in a decrease in TetR (along with *yfp*). When TetR has diluted sufficiently, the cell becomes blue.

The repressilator with reporter plasmid pFE2-12R reveals this directionality. Frame numbers correspond to ten-minute increments; the movie shows microcolony growth from 120 to 510 minutes. Inspection the time-lapse images confirms this trend: green → red → blue → green. A repressilator with the inverse "wiring diagram" (LacI → cI → TetR → LacI) would oscillate in the opposite direction: green → blue → red → green.



References

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

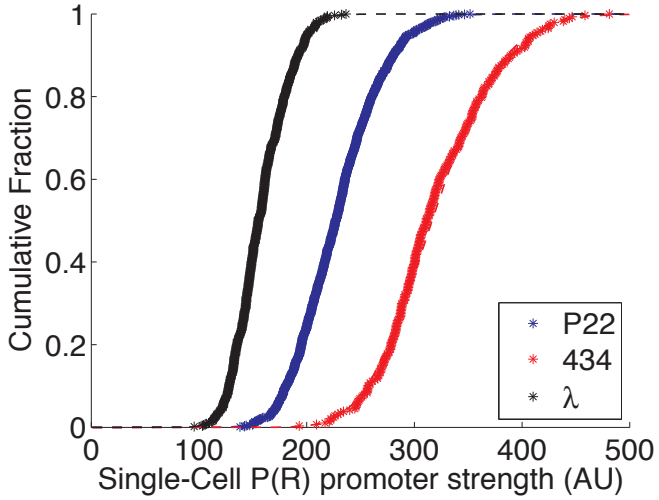
Lutz R, Bujard H (1997) Independent and tight regulation of transcriptional units in Escherichia coli via the LacR/O, the TetR/O and AraC/I1-I2 regulatory elements. *Nucleic Acids Res* **25:** 1203-1210.

Appendix D: Phage Circuits

In this appendix, I discuss several regulatory circuits I designed and built from the "lambdoid phage" repressors. I initially examined four of these repressors: cI from λ, cI from 434, c2 from Salmonella phage P22, and the HK022 repressor protein. I used the sequences for the repressors and promoters available in the Biobricks parts registry (parts.mit.edu). The HK022 repressor has been reported to have an extremely high cooperativity, but its target promoters are extremely weak compared to the other three phages—I was not able to detect unregulated promoter activity with *rfp*. The remaining three phage promoters showed very strong activity: 434 was the strongest, followed by P22 and λ. I was able to control the expression of target promoters with the p22 and λ repressor genes—unfortunately, the 434 repressor did not effectively repress its target.

Noise distributions of lambdoid promoters

Studies of stochastic noise in gene expression (Cai et al, 2006; Elowitz et al, 2002) have revealed transcription to be a bursty process. Transcripts are created in bursts, which occur during a transient active state of the promoter. Each transcriptional burst produces multiple mRNAs, and each mRNA can then produce several proteins. For an unregulated promoter, bursty behavior can be characterized with two noise parameters: the burst frequency a and the burst size b. These parameters describe a gamma distribution, a general stochastic form derived from the chemical master equation (Friedman et al, 2006). When the variance of autofluorescence is significant, the distributions are no longer gamma. For this reason, *rfp* is the optimal reporter for measuring burst parameters, since autofluorescence is barely detectable in the Crimson channel (Chapter 3). Experimentally, when a strong unregulated promoter drives the expression of a stable reporter protein such as RFP, these two parameters represent the number of transcriptional bursts per cell cycle (a) and the amount of fluorescence (in arbitrary units) generated by each burst (b). The burst size could be further calibrated in terms of proteins per cell (Rosenfeld et al, 2006)—here we use arbitrary units simply to compare the different promoters.



Measurements of ~500 cells for each of these three promoters revealed a very good fit to a gamma distribution. Interestingly, the burst size *b* increased with promoter strength while the burst frequency *a* remained roughly constant at about 40 bursts per cell cycle. Since the mRNA transcripts from all three promoters are identical, this difference corresponds simply to the number of transcripts generated per burst. Biochemical studies have shown that strong promoters are rate-limited at the late stages of transcription initiation (Kammerer et al, 1986). We do not know why transcripts are produced in bursts, though these results indicate a functional link between the late stages of transcription and the number of transcripts produced during a burst.

Promoter	Reporter	Relative strength[1]	**Burst size *b***	**Burst frequency *a***
λ P(R)	*rfp*	1.7	3.5-4.1	38-45
P22 P(R)	*rfp*	2.4	6.0-6.9	33-38
434 P(R)	*rfp*	3.3	7.3-9.3	35-44

Noise inference of cooperativity

We next asked whether genetic noise could be exploited to infer properties of gene regulation. A critical characteristic of any gene regulation system is the sharpness, or cooperativity, of the regulation. Circuit behaviors such as oscillation or bistability are sensitive to this property, which is frequently parameterized by an approximate Hill coefficient. It was recently shown that it is possible to make *in vivo* measurement of effective cooperativity using time-lapse movies (Rosenfeld et al, 2006). In that work, the repressor level was varied by dilution during growth. Here, we asked whether cooperativity

1 Median fluorescence relative to the $P_{lac/ara1}$ promoter under the fully inducing conditions, measured in Chapter 3.

could be inferred instantaneously, by using natural fluctuations to vary a transcription factor level, and observing correlations between its concentration and that of its target protein.

We built two variants of a synthetic repressor cascade. The first, on plasmid pFS1-1/2cI*/R, expressed a destabilized form of the λ cI repressor gene as an operon fusion with *yfp*. The cI-repressed promoter P(R) controls *rfp*. The second variant, pFS1-1/2cI*/R2, contained the same cascade circuit with the affinity-lowering OR2* point mutation in the P(R) promoter (Ptashne, 2005). Wild type (MG1655) cells containing each circuit were grown and measured in minimal media. Here, the "leaky" expression of cI* (i.e., without IPTG induction) was sufficient to partially repress *rfp*.

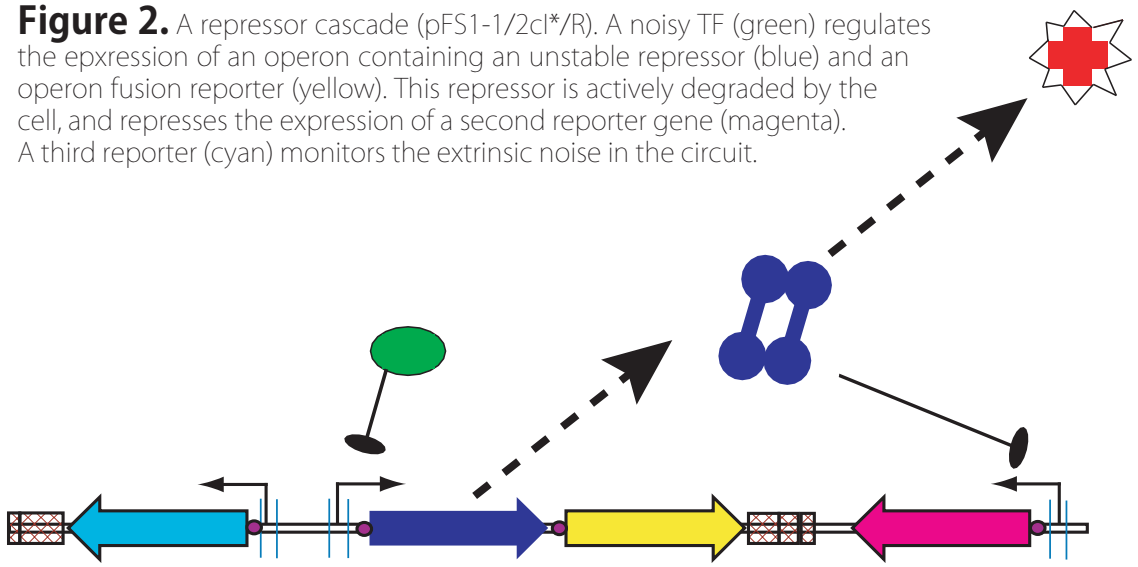
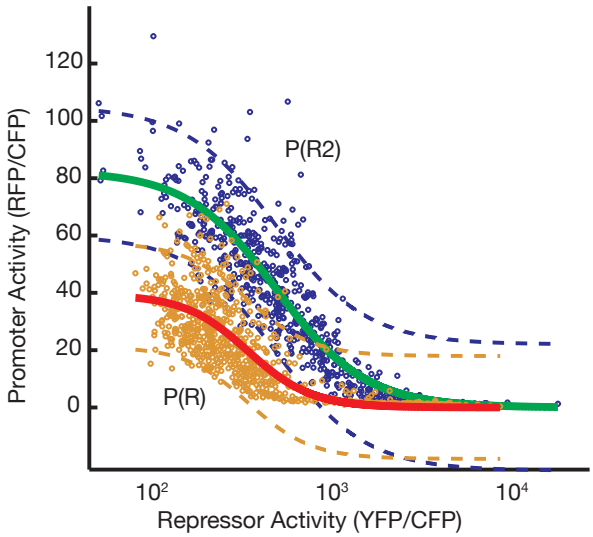


Figure 2. A repressor cascade (pFS1-1/2cI*/R). A noisy TF (green) regulates the epxression of an operon containing an unstable repressor (blue) and an operon fusion reporter (yellow). This repressor is actively degraded by the cell, and represses the expression of a second reporter gene (magenta). A third reporter (cyan) monitors the extrinsic noise in the circuit.

There was a distinct anti-correlation between the repressor operon (*yfp*) and its target genes (*rfp*). This indicated that the level of cI* expression was close to the threshold level required to repress the target promoters. The Hill equation captures this threshold response:

$$P(x) = \frac{\alpha}{1 + \frac{x^{\eta}}{\kappa^{\eta}}}$$

When the repressor concentration (x) is near the threshold level (κ) the output level is most sensitive to input fluctuations. Several factors contribute to output level fluctuations. Intrinsic (uncorrelated) noise contributes a purely stochastic component to the output level. Extrinsic noise between cells results in a positive correlation between the input and output levels. The intrinsic component of noise in the cI* level is propagated through the cascade, as is the extrinsic component of cI* noise. Both of these contributions result in anti-correlation between the input and output. The ideal measurement system would have large intrinsic noise in cI*, and zero noise (both intrinsic and extrinsic) for the other contributions. Since the MG1655 strain does not contain the TetR repressor, the *cfp* level reflects the extrinsic noise contribution. To account for extrinsic fluctuations, we normalized both the input (*yfp*) and output (*rfp*) levels by the constitutive (*cfp*) level. This correction mitigates the extrinsic noise contributions to cI* and the output, respectively. A plot of input versus output shows that the cell populations sample a large range of the promoter response functions.



We used these input-output distributions to fit the Hill function model. In each case, we measured the maximal promoter strength (α, arbitrary units) by deleting the cI* repressor gene from the circuit. We then fit the promoter threshold (κ, arbitrary units) and cooperativity (η, dimensionless) to the distributions by a simple non-linear fitting algorithm (Matlab function nlinfit). We used a bootstrap (with replacement) sampling procedure to determine 95% confidence intervals for each parameter. The change in the cooperativity of the OR2* mutation is detectable, and goes from about 2.4 to about 1.7. These values agree with previous *in vivo* measurements of the stable cI repressor (Rosenfeld et al, 2006).

Promoter	Strength (α)	Threshold (κ)	Cooperativity (η)
λ R	36.5-42.2	350-408	2.05-2.88
λ R2	78.1-86.9	502-588	1.59-1.91

This experiment shows two important results. First, we can exploit genetic noise to infer quantitative parameters of a genetic circuit. In this case, we were able to determine the cooperativity parameter for two different promoters. Though we did not measure the circuit threshold and promoter strength directly, we were able to obtain consistent relative values by comparison of the two circuits. This shows that the (LVA) degradation tag on the cI repressor does not significantly change the response parameters: the effect of protein degradation is merely to lower the concentration (activity) of the repressor. Importantly, this result could not be determined with a repressor dilution method (Rosenfeld et al, 2006) which relies on the systematic lowering of repressor concentration during cell growth. The second result is that we were able to measure a promoter response function instantaneously, without any external circuit perturbation. Since the cI* activity was poised near its threshold, we did not need to induce any aspect of the circuit to infer its behavior. Since the experiment was instantaneous (i.e., snapshots), this method could be extended to flow cytometry measurements. The advantage here is throughput: it is conceivable to measure *every* single-input promoter response function in an organism such as *E. coli* with this method. All that is needed is to identify the TF and promoter of interest, use a noise source (such as LacI) to control the TF expression level, and tune the system so that the TF activity fluctuates about its threshold value (κ). Coupling this analysis with the rules for combinatorial promoter function (Chapter 2), we can hope to build a quantitative model of promoter function that takes into account regulation by multiple—noisy—interacting inputs.

References

Cai L, Friedman N, Xie XS (2006) Stochastic protein expression in individual cells at the single molecule level. *Nature* **440:** 358-362.

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

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

Friedman N, Cai L, Xie XS (2006) Linking stochastic dynamics to population distribution: an analytical framework of gene expression. *Phys Rev Lett* **97:** 168302.

Kammerer W, Deuschle U, Gentz R, Bujard H (1986) Functional dissection of Escherichia coli promoters: information in the transcribed region is involved in late steps of the overall process. *The EMBO journal* **5:** 2995-3000.

Ptashne M (2005) Regulation of transcription: from lambda to eukaryotes. *Trends in biochemical sciences* **30:** 275-279.

Rosenfeld N, Perkins TJ, Alon U, Elowitz MB, Swain PS (2006) A fluctuation method to quantify in vivo fluorescence data. *Biophys J* **91:** 759-766.

Appendix E: A Synthetic *Escherichia coli* Predator-Prey Ecosystem

Lingchong You, Robert Sidney Cox III, and Frances H. Arnold

Abstract

We describe synthetic ecosystems of interacting *E. coli* populations that communicate and influence each other's behavior via *de novo* engineered genetic regulatory circuits. To establish cell-cell communication, we exploit components from "quorum sensing" systems that bacteria use to sense population density and coordinate behavior for diverse physiological functions. Mimicking mechanisms of programmed cell death, we use bacteriophage lysis genes to regulate population densities by controlling rates of cell death. Here we describe an ecosystem consisting of two types of cells that mutually regulate their gene expression and therefore survival; it resembles well-studied predator-prey systems in terms of basic logic and dynamics. Numerical analysis shows that the circuit generates stable oscillations in population densities and intracellular gene expression for a wide range of biologically feasible parameter values. In addition, the analysis highlights the key design features required to achieve the target circuit function in an experimental system. Systems such as this, which couple genetic regulation and cell-cell communication, will allow us to explore the dynamics of interacting populations in a well-defined experimental framework.

Introduction

We wish to construct interacting *Escherichia coli* populations—in essence, synthetic ecosystems—using genetic regulatory networks and intercellular communications systems to control and coordinate the behavior. One goal is to explore how single-celled bacteria can be programmed to exhibit complex behaviors, reminiscent of multicellular organisms. A second goal is to build simple, artificial ecosystems, where we can observe how well-controlled exchanges of information, in the form of signaling compounds that regulate gene expression, manifest themselves in the dynamics of interacting populations. Here we describe a simple circuit design based on naturally-occuring intercellular communications and genetic regulatory components that can convert *E. coli* into two populations whose dynamics mimic a predator-prey system.

We use the components of "quorum-sensing" systems to establish intercellular communication and coordinate behavior across a bacterial population. Many bacteria can sense their population density and coordinate a response via a small-molecule signal, called an autoinducer, that is produced intracellularly and can diffuse across the cell membrane into the medium(1, 2). The extracellular and, because they are coupled by diffusion, the intracellular concentrations of the signal reflect the density of the cells that are producing it. As the cell density increases, so does the total concentration of the autoinducer. When the intracellular concentration reaches a sufficient level, the autoinducer activates a cognate transcriptional regulator which further stimulates the production of autoinducer. This positive feedback results in a bistability: after a certain cell density all cells in the population express the autoinducer maximally. Other genes are also expressed in the same operon, including diverse physiological functions such as bioluminescence, pathogenicity, and biofilm formation. The use of quorum sensing in engineering applications has been demonstrated experimentally in recent work (3,20).

To construct populations that can interact in a programmable fashion, we want to connect

intracellular communication to cell growth and survival. Here we chose to mimic mechanisms of programmed cell death (4–6) and use bacteriophage lysis genes to control death rates. The lysis genes encode proteins that effect cell death and/or cell wall disruption at the late stage of infection by lytic bacterial viruses (7). Previous studies have demonstrated that rates of cell death can be regulated by expressing lysis genes with inducible promoters (8–10). Here, we use lysis genes and quorum sensing components to engineer a feedback circuit in which two populations of cells communicate to control the death rate of the other.

Many different circuits, programming different fundamental ecological interactions, can be built from these basic components. Here, we present a prototype ecosystem consisting of two *E. coli* populations that mutually regulate their gene expression and survival in a manner similar to the classic relationship of predator and prey. Based on results from mathematical modeling and numerical analysis, we point out potential experimental strategies that will facilitate the desired system function.

Circuit Design

The two *E. coli* population types, detailed in Figure 1, communicate and control each other's population density by producing small-molecule signals (acyl-homoserine lactones, or AHLs) that can freely diffuse across cell membranes into the medium (11) and regulate gene expression. The dynamics are similar to a predator-prey system: without the 'prey', the 'predator' population decays at a high rate due to expression of a lysis gene it carries. As the prey grows, it produces an AHL that diffuses through the medium into the predator, where it rescues the predator by inhibiting lysis gene expression. The predator produces a second AHL that diffuses into the prey and initiates synthesis of the lysis gene, effecting "predation". This circuit differs from a canonical predator-prey system in that, instead of acting as a food source, the prey provides an 'antidote'. It satisfies the broader definition of predation for a two-species ecosystem, where one species suffers from the growth of the second and the second benefits from the growth of the first (12).

We employ ordinary differential equations (ODEs) to model the major kinetic events during the functioning of this circuit, including cell growth and death, production, diffusion, and degradation of AHLs, production and degradation of transcriptional regulators, activation of the transcriptional regulators by AHLs, and regulation of lysis protein expression by activated transcriptional regulators. The resulting model contains 20 coupled ODEs and 40 kinetic parameters, many of which are as yet poorly characterized. Because the objective of this model is to capture overall dynamics of the underlying system, we followed conventional model reduction techniques to simplify it (13). Model reduction can also highlight parameters that have similar effects on system dynamics, which is useful for guiding experiments when "fine-tuning" system dynamics. By assuming that 1) concentrations of all interacting species other than the lysis proteins and the cells are at a quasi-steady state, 2) cell densities (in volume fraction) are much smaller than unity, and 3) diffusion of AHLs is much faster than their degradation, we reduce the model to four differential equations:

$$\frac{dc_1}{d\tau} = c_1(c_{max\,1} - c_1)/c_{max\,1} - e_1 c_1 \qquad 1a$$

$$\frac{dc_2}{d\tau} = \mu c_2(c_{max\,2} - c_2)/c_{max\,2} - e_2 c_2 \qquad 1b$$

$$\frac{de_1}{d\tau} = \frac{\kappa_{E\,1}}{1+\alpha_1 c_2^{\beta}} - \delta_{E\,1} e_1 \qquad 1c$$

$$\frac{de_2}{d\tau} = \frac{\kappa_{E\,2}\alpha_2 c_1^{\beta}}{1+\alpha_2 c_1^{\beta}} - \delta_{E\,2} e_2 \qquad 1d$$

All state variables and parameters in these equations are scaled with respect to the predator growth rate and written in a dimensionless form. Briefly, τ denotes time; c_1 and c_2 are the volume fraction of the predator and the prey, respectively; e_i (i = 1 for predator or 2 for prey) lumps effects of the concentration and lethality of the lysis protein in the corresponding cell. We assume that cell growth follows logistic kinetics with a carrying capacity of c_{maxi}. We further assume that regulation by AHLs (via binding to transcriptional regulators) follows Hill kinetics, with a cooperativity of β. The parameter μ represents the growth rate constant of the prey; κ_{Ei} and δ_{Ei} are the full synthesis

rate (uninhibited for the predator or fully induced for the prey) and the degradation rate constants of the corresponding lysis proteins; α_i lumps effects of production, diffusion, and degradation of the corresponding AHL with the sensitivity of its target regulatory protein to AHL concentration.

Results and Discussion

In addition to its resemblance to a predator-prey system in overall population dynamics, the system also demonstrates rich temporal patterns in intracellular gene expression. For example, a typical set of simulation results demonstrates stable oscillations in the cell densities and the concentrations of lysis proteins (Figure 2). Further analysis shows that the qualitative behavior of the system depends on parameters that directly affect accumulation of lysis proteins in either cell type, in particular, κ_{E1}, κ_{E2}, δ_{E1}, and δ_{E2}. Because δ_{Ei} and κ_{Ei} have opposite effects on the system dynamics, we focus our discussion on κ_{Ei}. As shown in Figure 3a, the system has a stable steady state when κ_{E1} is small. As κ_{E1} increases, the steady state becomes unstable by passing a Hopf bifurcation point(14); it becomes stable again upon passing another Hopf bifurcation point. Between these two points the system generates stable oscillations, with amplitude that first increases and then decreases with increasing κ_{E1}. The system does not oscillate when κ_{E1} is too small or too large. The system will fail to oscillate if $\left.\frac{\kappa_{E1}}{1+\alpha_1 c_2^{\beta}}\right|_{c_2=0} = \kappa_{E1} < \delta_{E1}$; under these conditions the predator will survive even without prey. Likewise, oscillation will fail when $\frac{\kappa_{E1}}{1+\alpha_1 c_{max\,2}^{\beta}} > \delta_{E1}$; here, the predator will die even if the prey is in abundance. These inequalities highlight how the system dynamics depends on the model parameters.

The system will also fail to oscillate if κ_{E2} is small, due to insufficient regulation by the predator on the prey (Figure 3b). Unlike the case of κ_{E1}, however, the system continues to oscillate for large κ_{E2} (Figure 3b). In this case, the maximum death rate of the prey is high. However, as the prey density drops to a low level, the predator density will eventually decrease because it needs the AHL from

the prey to repress its lysis gene. As predator density drops, expression of the lysis gene in the prey will slow down, and the prey will resume growth. Mathematically, for any finite value of κ_{E2}, the *effective* production rate of the lysis protein in the prey (represented by the term $\frac{\kappa_{E2}\alpha_2 c_1^{\beta}}{1+\alpha_2 c_1^{\beta}}$) will approach zero with decreasing predator density, permitting the prey to recover.

The other parameters are less critical. For example, our results show that cooperativity in AHL action is unimportant: the system can oscillate for all β values equal to or greater than 1, which encompasses essentially all biologically feasible regimes. In addition, when the other parameters are held constant at their base-case values, either α_1 or α_2 can be changed over orders of magnitude with only minor effects on the qualitative behavior of the circuit (not shown). However, these variations can result in significant changes in the oscillatory region defined by κ_{E1} and κ_{E2}, particularly in the case of α_1. In general, the oscillatory region is larger for larger α_1, because the regulatory range ($\frac{\kappa_{E1}}{1+\alpha_1 c_{max2}^{\beta}} \rightarrow \kappa_{E1}$) for the lysis protein synthesis rate in the predator will be broader. Interestingly, the same changes in α_2 have no significant effect on the oscillatory region (not shown). These changes will have very minor effect on the regulatory range of the lysis protein in the prey ($\frac{\kappa_{E2}\alpha_2 c_{max1}^{\beta}}{1+\alpha_2 c_{max1}^{\beta}}$), whose upper limit is essentially constant ($\approx \kappa_{E2}$) when $\alpha_2 > 100$.

These numerical results provide detailed guidance for experimental implementation. In particular, they highlight the importance of controlling the expression, lethality, and stability of the lysis proteins. This should be feasible, considering that a wide spectrum of phage λ holin gene mutants of varying lethality have been characterized(10, 15). In addition, production of the lysis proteins can be controlled by manipulating the strength of the corresponding promoters and ribosome binding sites. Our results indicate that oscillation is more likely when the accumulation rate and signaling sensitivity of AHL1 are large. This can be engineered by using strong promoters and ribosome binding sites for *luxI*, along with high copy number plasmids (if necessary). These design features should also reduce interference from intracellular noise. Our laboratory has demonstrated

that the function of a genetic circuit can be optimized by directed evolution (16), a well-established technology for improving protein functions *in vitro* and *in vivo* (17–19). In particular, directed evolution was used to achieve proper matching of the different components in the circuit, and allowed rapid optimization of a nonfunctional initial design. We anticipate that a similar strategy of mutagenesis and screening will facilitate tuning of kinetic parameters to achieve oscillatory behavior and will allow us to efficiently explore circuit function in different regions of the parameter space. Finally, we note that the idea of building synthetic ecosystems that incorporate population sensing and regulation goes well beyond the particular implementation described here. Addiction modules(6) that bacteria use to program death can be used in place of a lysis gene to regulate population density; the "killing" circuitry can also be reversed—to create a mutualistic relationship—by regulating a complementing gene to allow growth of an auxotroph.

Recent studies have demonstrated the feasibility of constructing xenobiotic input-output modules(20) and autonomous circuits that lead to stable(21), bi-stable(22), or oscillatory(23) gene expression. Yet these studies also indicate several engineering hurdles that may interfere with circuit function: noise in intracellular processes (24, 25), cell-to-cell variation across a population(23), and uncontrolled cell growth. These hurdles are likely to be overcome by the circuit design strategy described here. In particular, synchronization of intra-cellular behavior across a population, achieved by inter-cellular communication (26, 27), may render the circuit more resistant to noise in individual cells. And, by imposing limits on population densities, regulated cell death will allow a substantially wider time window for experimentation and characterization than offered by cell populations without density control.

In addition to addressing particular engineering challenges in genetic circuit design, these synthetic ecosystems will serve as well-defined systems for exploring evolutionary and ecological questions regarding, for example, the generation and maintenance of biodiversity (28-30) and the role of programmed cell death in bacteria (5). In these systems, there is a clear and explicit mapping

between the genetic construct and population dynamics. This mapping illuminates one of the central questions in ecology(31): how interactions at the molecular level are manifested in the temporal patterns of population dynamics. Although the current circuit demonstrates predation, the basic design strategy can be applied in a straightforward manner to program other ecological interactions,includingmutualism(orsymbiosis),competition,commensalismandamensalism(12). For instance, mutualism can be established when two types of cells produce AHLs that repress cell death in each other. Similarly, cells can be programmed to synthesize AHLs that mutually activate cell death, thereby creating competing populations. The essentially unlimited configurations that are possible with these basic elements will allow us to study the interplay between environment, gene regulation and population dynamics. With the addition of control over population mixing or segregation, it will be possible to program bacterial populations to mimic development and differentiation in multicellular organisms.

Computational Methods

The design and numerical analysis were carried out using modeling software Dynetica (32) (also see http://www.its.caltech.edu/~you). Stability and bifurcation analysis were carried out using XPP version 5.53 (http://www.math.pitt.edu/~bard/xpp/xpp.html) implementing a subset of a bifurcation analysis software AUTO.

Acknowledgments

We thank Cynthia Collins, Jared Leadbetter, Lianhong Sun, Yohei Yokobayashi, Ron Weiss, and Erik Winfree for discussions and comments. This material is based upon work supported by the Defense Advanced Research Projects Agency (DARPA) under Award No. N66001-02-1-8929. *Disclaimer:* Any opinions, findings, and conclusions or recommendations expressed in this publication are those of the author(s) and do not necessarily reflect the views of the DARPA.

Figure 1. *A synthetic predator-prey ecosystem.* We use two orthogonal quorum sensing modules to effect two-way communications between cell populations: the LuxI/LuxR system from the marine luminous bacterium *Vibrio fischeri* (33) and the RhlI/RhlR system from bacterium *Pseudomonas aeruginosa*(34). A lysis gene, such as the phage λ holin gene(35) or phage ϕX174 *E* gene(36), is used to regulate cell death. The gene is ON in the predator but OFF in the prey in the absence of regulation. The LuxI gene in the prey leads to the synthesis of an acyl-homoserine lactone (AHL1), which diffuses across the cell membrane into the medium and then into the predator, where it binds and activates LuxR. Active LuxR represses transcription of the lysis gene (E1) carried by the predator by binding to an engineered promoter containing a binding site for LuxR(37). The predator produces another AHL (AHL2) that diffuses into the prey, where it activates RhlR. Active RhlR activates transcription of the lysis gene (E2) by binding to its native promoter. AHL turnover, necessary for the proper functioning of the circuit, can be modulated by varying the medium pH(38) or with enzymes (acylase or lactonase) that can degrade AHL(39, 40). The outer boxes represent cells; solid arrows represent activation or production; dotted arrows denote diffusion; bars represent inhibition or degradation. The cognate receptors (LuxR and RhlR) for the AHLs are expressed constitutively and are omitted from the figure for clarity.

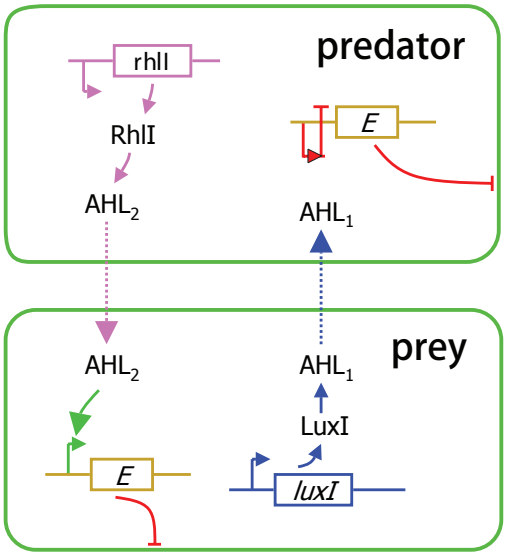


Figure 2. Simulated time courses of cell densities and lysis protein concentrations (E1 and E2 correspond to the predator and the prey, respectively) for the base parameter setting: $\mu = 1$; $c_{max1} = c_{max2} = 0.05$; $\kappa_{E1} = \kappa_{E2} = 2$; $\alpha_1 = \alpha_2 = 500$; $\delta_{E1} = \delta_{E2} = 1$; $\beta = 1.2$. See Supplementary information for sources and justification of parameter values.

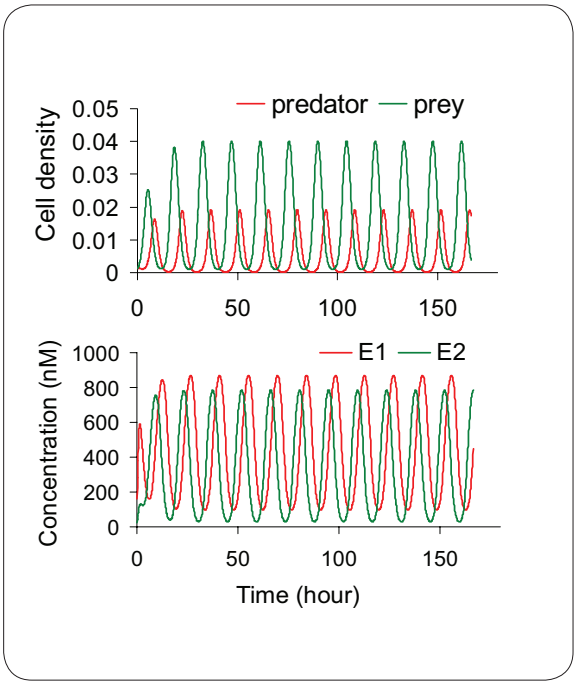
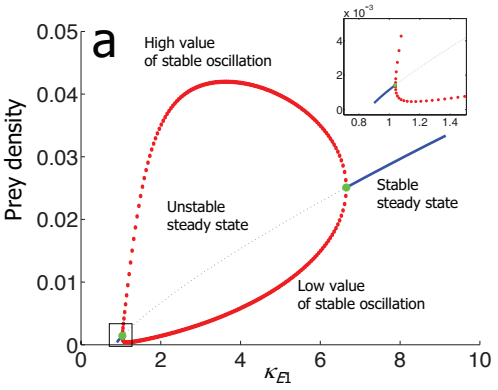
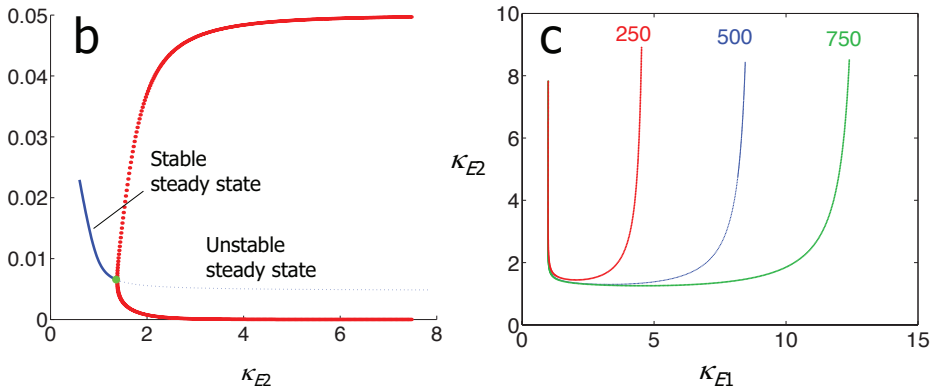


Figure 3. *Dependence of the steady-state prey density and/or its oscillation amplitudes on (a) κ_{E1} and (b) κ_{E2}.* The system will always have a non-trivial steady state that may be stable (thick solid lines) or unstable (thin dotted lines), depending on parameter values. Oscillations (thick dotted lines) result when the steady state is unstable. Inset in (a), corresponding to the small box at the left corner, shows in detail the first Hopf bifurcation with increasing κ_{E1}.





References

1. Miller, M. B. & Bassler, B. L. (2001) *Annu Rev Microbiol* **55,** 165-99.
2. Fuqua, C., Parsek, M. R. & Greenberg, E. P. (2001) *Annu Rev Genet* **35,** 439-68.
3. Weiss, R. & Knight, T. (2000) in *6th International Workshop on DNA-Based Computers, DNA 2000*, eds. Condon, A. & Rozenberg, G., Leiden, The Netherlands), pp. 1-16.
4. Kerr, J. F., Wyllie, A. H. & Currie, A. R. (1972) *Br J Cancer* **26,** 239-57.
5. Lewis, K. (2000) *Microbiol Mol Biol Rev* **64,** 503-14.
6. Engelberg-Kulka, H. & Glaser, G. (1999) *Annu Rev Microbiol* **53,** 43-70.
7. Young, I., Wang, I. & Roof, W. D. (2000) *Trends Microbiol* **8,** 120-8.
8. Morita, M., Asami, K., Tanji, Y. & Unno, H. (2001) *Biotechnol Prog* **17,** 573-6.
9. Hori, K., Kaneko, M., Tanji, Y., Xing, X. H. & Unno, H. (2002) *Appl Microbiol Biotechnol* **59,** 211-6.
10. Smith, D. L. & Young, R. (1998) *J Bacteriol* **180,** 4199-211.
11. Kaplan, H. B. & Greenberg, E. P. (1985) *J Bacteriol* **163,** 1210-4.
12. May, R. M. (1974) *Stability and complexity in model ecosystems* (Princeton University Press, Princeton, NJ, USA).
13. Edelstein-Keshet, L. (1988) *Mathematical models in biology* (McGrow-Hill, Inc., New York).
14. Seydel, R. (1994) *Practical bifurcation and stability analysis: from equilibrium to chaos* (Springer-Verlag, New York).
15. Chang, C. Y., Nam, K. & Young, R. (1995) *J Bacteriol* **177,** 3283-94.
16. Yokobayashi, Y., Weiss, R., and Arnold, F. H., Directed evolution of a genetic circuit. Proc Natl Acad Sci U S A 99 (26), 16587 (2002).
17. Arnold, F. H. & Volkov, A. A. (1999) *Curr Opin Chem Biol* **3,** 54-9.
18. Giver, L. & Arnold, F. H. (1998) *Curr Opin Chem Biol* **2,** 335-8.
19. Arnold, F. H. (2001) *Nature* **409,** 253-7.
20. Weiss, R. and Knight Jr., T., presented at the 6th International Workshop on DNA-Based

Computers, Leiden, The Netherlands, 2000 (unpublished).

21. Becskei, A. & Serrano, L. (2000) *Nature* **405,** 590-3.
22. Gardner, T. S., Cantor, C. R. & Collins, J. J. (2000) *Nature* **403,** 339-42.
23. Elowitz, M. B. & Leibler, S. (2000) *Nature* **403,** 335-8.
24. Elowitz, M. B., Levine, A. J., Siggia, E. D. & Swain, P. S. (2002) *Science* **297,** 1183-6.
25. Ozbudak, E. M., Thattai, M., Kurtser, I., Grossman, A. D. & van Oudenaarden, A. (2002) *Nat Genet* **31,** 69-73.
26. McMillen, D., Kopell, N., Hasty, J. & Collins, J. J. (2002) *Proc Natl Acad Sci U S A* **99,** 679-84.
27. Glass, L. (2001) *Nature* **410,** 277-84.
28. Kerr, B., Riley, M. A., Feldman, M. W. & Bohannan, B. J. (2002) *Nature* **418,** 171-4.
29. Lenski, R. E. & Hattingh, S. E. (1986) *J Theor Biol* **122,** 83-93.
30. Czaran, T. L., Hoekstra, R. F. & Pagie, L. (2002) *PNAS* **99,** 786-790.
31. Bohannan, B. & Lenski, R. E. (2000) *Ecology Letters* **3,** 362-377.
32. You, L., Hoonlor, A. & Yin, J. (2002) *Bioinformatics* **In press**.
33. Fuqua, W. C., Winans, S. C. & Greenberg, E. P. (1994) *J Bacteriol* **176,** 269-75.
34. Pesci, E. C., Pearson, J. P., Seed, P. C. & Iglewski, B. H. (1997) *J Bacteriol* **179,** 3127-32.
35. Wang, I. N., Smith, D. L. & Young, R. (2000) *Annu Rev Microbiol* **54,** 799-825.
36. Young, K. D. & Young, R. (1982) *J Virol* **44,** 993-1002.
37. Egland, K. A. & Greenberg, E. P. (2000) *J Bacteriol* **182,** 805-11.
38. Schaefer, A. L., Hanzelka, B. L., Parsek, M. R. & Greenberg, E. P. (2000) *Methods Enzymol* **305,** 288-301.
39. Dong, Y. H., Wang, L. H., Xu, J. L., Zhang, H. B., Zhang, X. F. & Zhang, L. H. (2001) *Nature* **411,** 813-7.
40. Leadbetter, J. R. & Greenberg, E. P. (2000) *J Bacteriol* **182,** 6921-6.