Synthetic Circuits for Feedback and Detection in Bacteria

Thesis by
Victoria Hsiao

In Partial Fulfillment of the Requirements for the
degree of
Doctor of Philosophy, Bioengineering

CALIFORNIA INSTITUTE OF TECHNOLOGY
Pasadena, California

2016
Defended April 25th, 2016

Victoria Hsiao
ORCID: 0000-0001-9297-1522

ACKNOWLEDGEMENTS

I am so grateful to have had the opportunity to do my graduate studies at Caltech, and to have gotten to know so many thoughtful and generous people.

First, I would like to thank my advisor, Richard Murray, for being a great advisor and mentor by providing just the right amounts of guidance and independence. Richard has given me so many opportunities to present my work, to develop my ideas, and to collaborate with others. I would also like to thank Paul Rothemund for his feedback on data, ideas, and life. While I did not start to know Paul until the later half of my grad school career, he has been an invaluable collaborator and mentor. I would also like to thank the other members of my committee – Michael Elowitz, Niles Pierce, and Lea Goentoro – for being extremely supportive and contributing their thoughts on my projects.

Thank you to the Murray Lab. The Murray Lab can only be described as a loosely associated group of fun and brilliant people all doing their own thing; it is a group with whom I have grown as a scientist and as a person, and that I will miss being a part of day-to-day. Furthermore, I need to thank all past and current members of the Ladies Who Lift – especially Marcella Gomez, Anu Thubagere, Melissa Tanner, and Camille McAvoy. Exercise is 100% better when it includes new and amazing friends. Thank you also to all the amazing people who helped us celebrate Galentine's Day and Friendsgiving. In particular, I need to thank Clare Hayes, not only for paying half the rent, but also for being a smart and hilarious friend. Finally, I would like to thank my family and Tommy for their unwavering encouragement and support.

Finding happiness and balance during graduate school can be hard. I could not have done it without the help of this incredible community, and I could not have asked for a better experience.

ABSTRACT

Synthetic biology, by co-opting molecular machinery from existing organisms, can be used as a tool for building new genetic systems from scratch, for understanding natural networks through perturbation, or for hybrid circuits that piggy-back on existing cellular infrastructure. Although the toolbox for genetic circuits has greatly expanded in recent years, it is still difficult to separate the circuit function from its specific molecular implementation. In this thesis, we discuss the function-driven design of two synthetic circuit modules, and use mathematical models to understand the fundamental limits of circuit topology versus operating regimes as determined by the specific molecular implementation. First, we describe a protein concentration tracker circuit that sets the concentration of an output protein relative to the concentration of a reference protein. The functionality of this circuit relies on a single negative feedback loop that is implemented via small programmable protein scaffold domains. We build a mass-action model to understand the relevant timescales of the tracking behavior and how the input/output ratios and circuit gain might be tuned with circuit components. Second, we design an event detector circuit with permanent genetic memory that can record order and timing between two chemical events. This circuit was implemented using bacteriophage integrases that recombine specific segments of DNA in response to chemical inputs. We simulate expected population-level outcomes using a stochastic Markov-chain model, and investigate how inferences on past events can be made from differences between single-cell and population-level responses. Additionally, we present some preliminary investigations on spatial patterning using the event detector circuit as well as the design of stationary phase promoters for growth-phase dependent activation. These results advance our understanding of synthetic gene circuits, and contribute towards the use of circuit modules as building blocks for larger and more complex synthetic networks.

PUBLISHED CONTENT AND CONTRIBUTIONS

Hsiao, Victoria, Yutaka Hori, et al. (In press). "A population-based temporal logic gate for timing and recording chemical events". In: *Molecular Systems Biology*.
V.H. conceived of the circuit design, constructed the necessary experimental strains, performed experimental work and data analysis, ran model simulations, and wrote the manuscript. Y.H. developed the stochastic model and derived the mathematical results.

Hsiao, Victoria, Emmanuel L C de los Santos, et al. (2015). "Design and implementation of a biomolecular concentration tracker." In: *ACS Synthetic Biology* 4.2, pp. 150–161. DOI: 10.1021/sb500024b.
V.H. participated in the conception of the project, cloned all the plasmids, designed and executed all experimental and microscopy work, prepared the data, and wrote the manuscript. V.H. and E.L.C.S designed the mathematical model together and V.H. ran simulations.

Santos, Emmanuel LC de los, Victoria Hsiao, and Richard M Murray (2013). "Design and implementation of a biomolecular circuit for tracking protein concentration". In: *Proceedings of the American Control Conference*, pp. 1–5.
V.H. and E.L.C.S are co-first authors on this conference paper. V.H. and E.L.C.S. designed the mathematical model together and V.H. ran simulations. The manuscript was co-written.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

LIST OF TABLES

Chapter 1

INTRODUCTION

Synthetic biology is the study of re-purposing existing molecular machinery for engineered applications. Synthetic gene circuits can be used in three possible ways: to create a completely new module that is intended to be independent from host operations, to better understand existing networks by adding on an inducible module for perturbation and measurement, or to implement a hybrid system that diverts an existing pathway or signal towards a synthetic module. In order to create and apply synthetic circuits, we need a basic understanding of the underlying biology, but the implementation and testing of synthetic systems often also leads to new biological insights and characterizations. Although the toolbox for genetic circuits has greatly expanded in recent years, it is still difficult, from an engineering perspective, to separate the circuit function from its specific molecular implementation. As a result, it is critical to build mathematical models that capture the overall circuit topology and intended function. Simulation results can then be used to determine which aspects of experimental outcomes are a result of the circuit topology, and which are implementation-specific idiosyncrasies.

In this thesis, we describe the design and implementation of two novel synthetic circuit modules: a biomolecular tracker that regulates protein stoichiometry via negative feedback (Chapter 2), and a temporal logic gate that both senses and records the order and timing of chemical events (Chapter 3). We use modeling and simulation to show that underlying circuit topology governs overall circuit behavior, and that performance limitations and time constants for each system were set by the particular sets of genetic and molecular parts used. This distinction, in which the model does not depend on the specific implementation but is tuned to have realistic parameters, enabled model-driven experimental design and also hypothesis testing for unexpected experimental results. In Chapter 4, we extend the temporal logic gate by adding a spatial element for diffusion-induced patterning and differentiation. Finally, in Chapter 5, we design stationary-phase combinatorial promoters, and discuss possible applications for understanding natural biofilm structures or for implementing growth-phase dependent delays in synthetic circuits.

First, in Chapter 2, we describe a protein concentration regulation circuit that sets

the concentration of an output protein relative to the concentration of a reference protein. Though positive and negative feedback systems are an essential feature of endogenous signaling networks, synthetic circuits more commonly rely on library-based screening to find optimal expression levels. This type of open-loop approach is sensitive to downstream synthetic modules, environmental factors, host organism, and growth phase (Del Vecchio, Ninfa, and Sontag, 2008; Klumpp, Zhang, and Hwa, 2009; Cardinale, Joachimiak, and Adam Paul Arkin, 2013). In this chapter, we design a module that regulates relative stoichiometry between two proteins. Constant relative ratios between proteins has been measured in a variety of complex signaling pathways, including apoptosis (Federspiel et al., 2016), the MAP kinase pathway (Kolch, 2000), and G-protein coupled receptors (Ostrom, Post, and Insel, 2000), but it is still difficult to implement more than one regulatory loop in synthetic circuits. Regulation of stoichiometry ensures constant activity despite cell-to-cell variation and intramolecular noise. The functionality of this circuit relies on a single negative feedback loop that is implemented via small programmable protein scaffold domains. In addition to building and demonstrating dynamic signal tracking of the experimental circuit, we build a mass-action model to understand the relevant timescales of the tracking behavior as well as how the input/output ratios and circuit gain might be tuned with circuit components.

In Chapter 3, we discuss the design and implementation of another novel module, a temporal logic gate with permanent genetic memory. More importantly, we show that population-level analysis of stochastic single cell responses can provide additional information not encoded in single cells. Previously, much work had been done in creating Boolean logic gates for static chemical signals using molecular-concentration based systems (Gardner, Cantor, and Collins, 2000; Anderson, Voigt, and Adam P Arkin, 2007; Moon et al., 2013) or bacteriophage integrases (Bonnet, Subsoontorn, and Endy, 2012; Bonnet, Yin, et al., 2013; Siuti, Yazbek, and Lu, 2013). Temporal logic gates, which process time-varying chemical signals, have been much less explored. Pioneering work by Friedland *et al.* used serine integrase-based recombination for the counting and detection of sequential pulses of inducers (Friedland et al., 2009). But thus far, no work has studied the potential for temporal logic gates to provide information about the duration of a signal, or the time between two chemical events. Specifically, we design a circuit with four different genetic states for *no inducer*, *a only*, *b only*, and *a then b*. Two of these genetic states, *a only* and *a then b*, express RFP and GFP fluorescence, respectively. Although the circuit design does have a unique genetic state for *b then a*, we find that analysis of GFP

population fractions creates a distinct population-level output for a *b then a* event. The most compelling advantage of engineered biological systems over manmade sensors lies in their inherent capabilities for replication and parallel sensing with minimal energy and resource requirements. Thus, deployment of synthetic bacterial devices would almost certainly involve populations of cells, never just a single cell. Using a combination of Markov-chain stochastic simulations and experimental data, we find that stochasticity in digital single-cell responses to chemical inputs translates into analog population-level differentiation, and that this additional information can provide nuanced details on past events. This system enables us to begin thinking about using population distributions as circuit readouts, and to design feedback for maintaining distributions of differentiated cells.

In Chapter 4, we use the same temporal logic gate *E. coli* strain, but use 2D diffusion to implement the timing between cell exposure to inducers. This creates spatial patterning in which cells genetically differentiate based on their distance from the inducer point sources. While it is well known that steady-state morphogen gradients are a critical part of embryonic development, more recent studies have shown that the high levels of precision and robustness in cell differentiation are actually derived from early decoding of transient, pre-steady-state, morphogen signals (Bergmann et al., 2007; Tamari and Barkai, 2012). In this chapter, we discuss experimental setup for diffusion-based experiments, present some preliminary data on spatial patterns generated from the temporal logic gate strain, and characterize the effect of growth rate and media on patterning. We show that the temporal logic gate works when the timing between inducers is regulated by distance between inducer point sources, and that distinct populations of differentiated cells can be visualized via fluorescence. We also design a modified version of the temporal logic gate in which fluorescence only occurs in the overlap region of the two diffusing inducers. When implemented experimentally, this design was met with limited success, and the reasons for that are also discussed in detail.

In Chapter 5, we present a small study on combinatorial stationary phase *E. coli* promoters and some of their possible applications. Current testing of synthetic circuits is done almost exclusively in exponential growth phase, but this is neither a realistic environment for natural bacteria nor a reasonable expectation for engineered strains outside of the laboratory. Stationary phase gene expression in *E. coli* has been widely studied (Miksch and Dobrowolski, 1995; Lee and Gralla, 2001; Lacour and Landini, 2004; Shimada et al., 2004) and development of inducible stationary-

phase active promoters would have a number of significant advantages. In particular, circuit functionality could be partitioned into exponential phase tasks and stationary phase tasks, or cells could be programmed to conserve resources until some quorum had been reached. Here, we have identified 13 native σ38 promoters from the literature, as well as a σ38 consensus sequence, and created a test construct with a GFP reporter. We then measured their activity and growth rate over time in DH5α-Z1 *E. coli* cells, and found that all the promoters turned on in late-log phase but with varying activity levels. The three promoters with the highest expression profiles were then combined with tetO operator sites to create $\sigma 38 - tetO$ combinatorial promoters. We show, via time-course traces, that 7 out of 9 engineered promoters exhibit the correct behavior where they are only active during late-log phase and if aTc is present to relieve tetR repression.

Additionally, macroscale fluorescence microscopy was used to visualize the spatial structure of biofilms comprised of σ38-GFP reporter strains (Chapter 5). *E. coli* naturally form biofilms that are comprised of multiple layers of cells in different growth phases (Serra et al., 2013; Hobley et al., 2015). The availability of stationary phase active promoters in the context of biofilms means that cells could be programmed to express different functions based on their location within the naturally occurring biofilm. Natural biofilms already take advantage of spatial differences in growth phase – cells on the biofilm periphery are in exponential phase and expand quickly, but are also more susceptible to attack and stress, while cells on the interior in stationary phase are protected but also receive less nutrients (Liu et al., 2015). Combinatorial stationary phase promoters could be used to couple synthetic model activity to existing cellular infrastructure processes, such as the complex mechanisms that determine growth phase and σ38 production. Rather than design a synthetic timer, we could use growth-phase dependent promoters to implement delays or oscillatory behaviors in synthetic circuit activity.

The work presented in this thesis advances our understanding of synthetic gene circuits, and contributes towards the use of circuit modules as building blocks for larger and more complex synthetic networks. In future work, we discuss possibilities of using the temporal logic gate not as an endpoint readout of past events, but as a population generator that can take a homogenous population of "stem cell" bacteria, and through timed inputs, reliably create distributions of genetically distinct daughter cells. A controlled population distribution generator would create new possibilities for testing distribution-based feedback and building multi-strain consortia.

References

Anderson, J Christopher, Christopher A Voigt, and Adam P Arkin (2007). "Environmental signal integration by a modular AND gate." In: *Molecular Systems Biology* 3, p. 133.

Bergmann, Sven et al. (2007). "Pre-steady-state decoding of the Bicoid morphogen gradient." In: *PLoS Biology* 5.2, e46.

Bonnet, Jerome, Pakpoom Subsoontorn, and Drew Endy (2012). "Rewritable digital data storage in live cells via engineered control of recombination directionality." In: *Proceedings of the National Academy of Sciences of the United States of America* 109.23, pp. 8884–8889.

Bonnet, Jerome, Peter Yin, et al. (2013). "Amplifying genetic logic gates." In: *Science* 340.6132, pp. 599–603.

Cardinale, Stefano, Marcin Pawel Joachimiak, and Adam Paul Arkin (2013). "Effects of genetic variation on the E. coli host-circuit interface." In: *Cell Reports* 4.2, pp. 231–237.

Del Vecchio, Domitilla, Alexander J Ninfa, and Eduardo D Sontag (2008). "Modular cell biology: retroactivity and insulation." In: *Molecular Systems Biology* 4, p. 161.

Federspiel, Joel D et al. (2016). "Assembly dynamics and stoichiometry of the apoptosis signal-regulating kinase (ASK) signalosome in response to electrophile stress." In: *Molecular & cellular proteomics : MCP*, mcp.M115.057364.

Friedland, Ari E et al. (2009). "Synthetic gene networks that count." In: *Science* 324.5931, pp. 1199–1202.

Gardner, T S, C R Cantor, and J J Collins (2000). "Construction of a genetic toggle switch in Escherichia coli." In: *Nature* 403.6767, pp. 339–342.

Hobley, Laura et al. (2015). "Giving structure to the biofilm matrix: an overview of individual strategies and emerging common themes". In: *FEMS microbiology reviews* 39.5, pp. 649–669.

Klumpp, Stefan, Zhongge Zhang, and Terence Hwa (2009). "Growth Rate-Dependent Global Effects on Gene Expression in Bacteria". In: *Cell* 139.7, pp. 1366–1375.

Kolch, W (2000). "Meaningful relationships: the regulation of the Ras/Raf/MEK/ERK pathway by protein interactions." In: *The Biochemical journal* 351 Pt 2.Pt 2, pp. 289–305.

Lacour, Stephan and Paolo Landini (2004). "SigmaS-dependent gene expression at the onset of stationary phase in Escherichia coli: function of sigmaS-dependent genes and identification of their promoter sequences." In: *Journal of Bacteriology* 186.21, pp. 7186–7195.

Lee, S J and J D Gralla (2001). “Sigma38 (rpoS) RNA polymerase promoter engagement via -10 region nucleotides.” In: *The Journal of Biological Chemistry* 276.32, pp. 30064–30071.

Liu, Jintao et al. (2015). “Metabolic co-dependence gives rise to collective oscillations within biofilms.” In: *Nature* 523.7562, pp. 550–554.

Miksch, G and P Dobrowolski (1995). “Growth phase-dependent induction of stationary-phase promoters of Escherichia coli in different gram-negative bacteria.” In: *Journal of Bacteriology* 177.18, pp. 5374–5378.

Moon, Tae Seok et al. (2013). “Genetic programs constructed from layered logic gates in single cells”. In: *Nature* 491.7423, pp. 249–253.

Ostrom, R S, S R Post, and P A Insel (2000). “Stoichiometry and compartmentation in G protein-coupled receptor signaling: implications for therapeutic interventions involving G(s).” In: *The Journal of pharmacology and experimental therapeutics* 294.2, pp. 407–412.

Serra, Diego O et al. (2013). “Microanatomy at cellular resolution and spatial order of physiological differentiation in a bacterial biofilm.” In: *mBio* 4.2, e00103–13.

Shimada, Tomohiro et al. (2004). “Classification and strength measurement of stationary-phase promoters by use of a newly developed promoter cloning vector.” In: *Journal of Bacteriology* 186.21, pp. 7112–7122.

Siuti, Piro, John Yazbek, and Timothy K Lu (2013). “Synthetic circuits integrating logic and memory in living cells.” In: *Nature Biotechnology* 31.5, pp. 448–452.

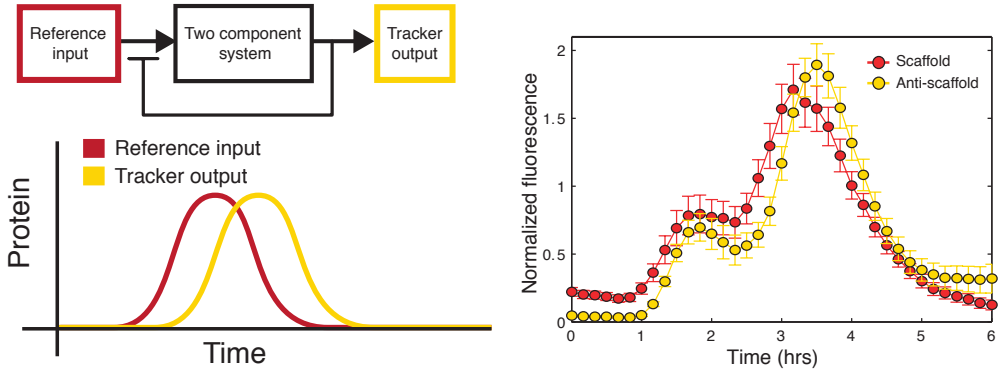
Tamari, Zvi and Naama Barkai (2012). “Improved readout precision of the Bicoid morphogen gradient by early decoding.” In: *Journal of Biological Physics* 38.2, pp. 317–329.

Chapter 2

DESIGN AND IMPLEMENTATION OF A BIOMOLECULAR CONCENTRATION TRACKER

A version of this chapter has been published as a conference paper (mathematical model, discussed in Santos, Hsiao, and Richard M Murray, 2013), and as a journal article (experimental results, presented in Hsiao et al., 2015).

2.1 Abstract



As a field, synthetic biology strives to engineer increasingly complex artificial systems in living cells. Active feedback in closed loop systems offers a dynamic and adaptive way to ensure constant relative activity independent of intrinsic and extrinsic noise. In this work, we use synthetic protein scaffolds as a modular and tunable mechanism for concentration tracking through negative feedback. Input to the circuit initiates scaffold production, leading to colocalization of a two-component system and resulting in the production of an inhibitory anti-scaffold protein. Using a combination of modeling and experimental work, we show that the biomolecular concentration tracker circuit achieves dynamic protein concentration tracking in *Escherichia coli* and that steady state outputs can be tuned.

2.2 Introduction

Implementation of reliable feedback and control in engineered circuits is a continuing challenge in synthetic biology. Though positive and negative feedback systems are an essential feature of natural biological networks, synthetic circuits more commonly rely on library-based screening to find optimal expression levels. Not only

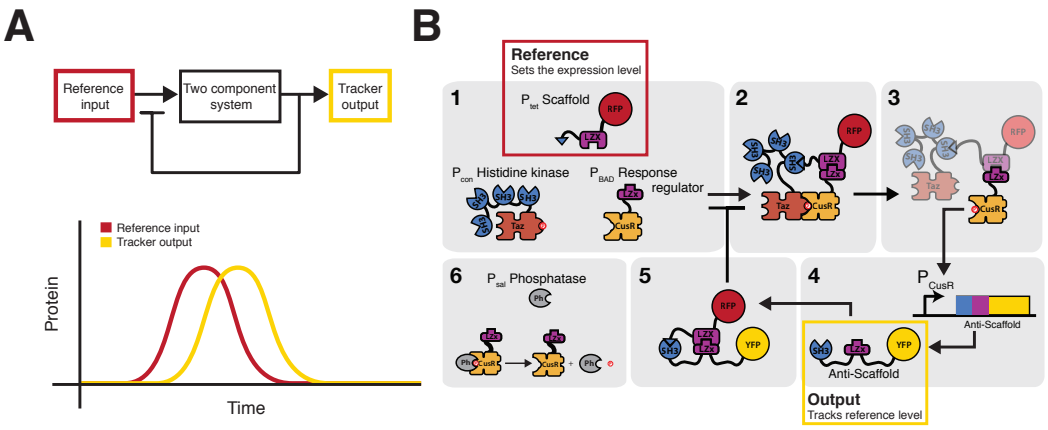


Figure 2.1: Overview of circuit design. A) The circuit takes an input that sets the reference value. The input proportionally modulates activity of a two component signaling system that then produces an output. The output triggers a negative feedback response. The negative feedback is the mechanism that generate real-time tracking behavior. B) The specific implementation of the circuit is shown. The circuit regulates the production of the amount of target protein (anti-scaffold-YFP) with respect to the amount of reference protein (scaffold-RFP). Expression of the target is dependent on the amount of free scaffold. The target contains domains which sequester free scaffold creating a negative feedback loop. Scaffold, response regulator, and phosphatase concentrations are induced via P_{tet}, P_{BAD}, and P_{sal}, respectively.

are the resulting systems sensitive to relative concentrations between components, but each time the circuit is expanded, the network of regulatory sequences must be re-optimized to account for increased load on cell machinery (Klumpp, Zhang, and Hwa, 2009). More importantly, this type of open loop approach only optimizes for a single set of environmental parameters, and inherently does not accommodate for stochastic cell-to-cell variation, changes due to cell growth cycles, or changes in cell loading from other circuit modules (Cardinale, Joachimiak, and Arkin, 2013).

Closed loop systems provide regulation of individual components that is robust with respect to environmental disturbances. Negative feedback is a common feature of natural pathways, and has been shown to decrease transcriptional response time (Rosenfeld, Elowitz, and Alon, 2002), to provide stability and reduce fluctuations (Becskei and Serrano, 2000), and to be necessary for oscillatory behavior (Ferrell, 2013).

Active feedback in biological systems has been previously considered at various levels. Recent studies have designed and studied an RNA-based rate regulating

circuit with two opposing negative feedback loops (Franco, Forsberg, and Richard M. Murray, 2008), a system utilizing an RNA binding protein to repress translation of its own mRNA (Stapleton et al., 2012), and analysis of noise in transcriptional negative feedback (Dublanche et al., 2006). There have also been demonstrations of an *in silico* closed loop system, in which a computer measured fluorescence output and automatically modulated the activity of a photosensitive transcription factor (Milias-Argeitis et al., 2011). In that study, the negative feedback occurred in the software control system outside of the cell.

In this work, we present an *in vivo* protein concentration tracker circuit. To our best knowledge, this is the first demonstration of dynamic molecular tracking entirely within the cell environment. This circuit contains a single negative feedback loop implemented with scaffold proteins and operates on the timescale of one cell cycle. We show that negative feedback implemented through sequestration results in "tracking" behavior: the proportional modulation of one protein concentration (the *anti-scaffold*) relative to that of the reference protein (the *scaffold*) over a range of reference induction levels.

2.3 Results and discussion

Scaffold-based circuit design and implementation

Previously, Weston R Whitaker et al., (2012) designed a scaffold-dependent two-component system in which the phosphotransfer was mediated by a synthetic scaffold protein consisting of small protein-protein binding domains. They demonstrated that weak natural cross-talk between a noncognate histidine kinase and response regulator pair could be artificially amplified via colocalization onto the scaffold. By fusing the kinase to the Crk SH3 domain and the response regulator to half of a leucine zipper, both would be recruited in the presence of a scaffold protein consisting of the SH3 ligand and the other half of the leucine zipper. Forcing the kinase and response regulator into close proximity greatly enhances the level of phosphotransfer and thus the level of downstream expression. The kinase-regulator pair of Taz and CusR was chosen because of measured low levels of cross-talk upon long incubations of purified proteins (Skerker et al., 2005).

Building upon this scaffold-dependent two-component system, we designed a negative feedback circuit by introducing an anti-scaffold molecule that competitively inhibits scaffold function. The *scaffold* molecule consists of a leucine zipper domain (LZX) linked to the SH3 ligand via flexible glycine-serine repeats (Figure 1). The

two component system is comprised of the chimeric kinase Taz linked to four SH3 domains and the response regulator CusR linked to a single leucine zipper (LZx) domain (Figure 1A). The presence of the scaffold recruits the HK Taz and RR CusR into close proximity by forming a ternary complex, resulting in the phosphorylation of CusR. The phosphorylated CusR becomes an active transcription factor, binding to its natural promoter (P_{CusR}) and activating expression of the *anti-scaffold* protein (Figure 1B). The anti-scaffold consists of the complementary LZx and SH3 ligand domains, which allow it to competitively bind to and consequently sequester the scaffold protein (K_d = 6 nM for the leucine zipper and K_d = 100 nM for the SH3 domain (Acharya et al., 2002; Weston R. Whitaker and John E. Dueber, 2011)). This prevents further phosphorylation of the response regulator, and halts further production of the anti-scaffold. In the absence of any scaffold protein, no activated response regulator activity is observed (Figure BT-S2).

We implemented the circuit in a ΔCusS ΔCusR *E. coli* knockout strain (Weston R Whitaker et al., 2012). In the absence of CusS, the native bifunctional histidine kinase/phosphatase partner for CusR, activated CusR proteins remain phosphorylated. Accordingly, we re-introduced a CusS(G448A) mutant behind an inducible promoter to tune response regulator deactivation. The G448A mutation disrupts the ATP binding site, eliminating kinase autophosphorylation without affecting phosphatase activity (Weston R Whitaker, 2012; Zhu and Inouye, 2002). This created a tunable phosphate sink in our circuit and ensures tight coupling between present scaffold and activated response regulator concentrations. The negative feedback circuit with the anti-scaffold is referred to as the *closed loop circuit*. As a control, we also built an *open loop circuit*, which instead of P_{CusR}-driven expression of the anti-scaffold, only the anti-scaffold reporter is expressed.

We constructed the circuit as a three plasmid system, in which the kinase is constitutively expressed and the scaffold, response regulator, and phosphatase were cloned behind the inducible promoters P_{tet}, P_{BAD}, and P_{sal}, respectively. Dynamic tracking behavior was visualized by adding medium strength ssrA degradation tags (C-terminal, RPAANDENYAAAV) to the scaffold-RFP and anti-scaffold-YFP fusion proteins (Andersen and Molin, 1998). The fluorescent reporters mCherry RFP and Venus YFP were chosen on account of their similar maturation times (∼5 and 15 min, respectively) (Nagai et al., 2002; Shaner et al., 2004).

Parameter		Value	Units	Description
HK_{tot}		100	nM	Histidine kinase
RR_{tot}		0 - 3000	nM	Response regulator
Sc_{tot}		0 - 5000	nM	Scaffold
Ph_{tot}		0 - 5000	nM	Phosphatase
β_{AS}		10	nM/s	Transcription + Translation
β_0		0.01		Leaky promoter activity (1% of total induction)
γ		3.84×10^{-4}	s^{-1}	Deg/dilution [Groban et al., 2009, Munsky et al., 2009]
γ_{ssrA}		1.5γ	s^{-1}	Degradation for ssrA-tagged proteins
n		2		Hill coefficient for anti-scaffold activation
K_D		1	nM	K_D for AS activation
k_{dephos}		0.003	s^{-1}	Phosphatase mediated dephosphorylation [cf. Groban et al., 2009]
r		2.8×10^{-4}	s^{-1}	Decay constant for diffusion of inducer [cf. Munsky et al., 2009]
Forward and reverse reaction rates				
k_{HKp}	k_f	0.003	s^{-1}	HK autophosphorylation [cf. Groban et al., 2009]
	k_r	0.0001	s^{-1}	[cf. Pazy et al., 2009]
k_{cogp}	k_f	102.1	s^{-1}	Cognate HK-RR phosphorylation [cf. Groban et al., 2009]
	k_r	0.00294	s^{-1}	[cf. Groban et al., 2009]
k_{noncog}	k_f	0.0031	$s^{-1} M^{-1}$	Non-cognate HK-RR phosphorylation [cf. Groban et al., 2009]
	k_r	0.0002	$s^{-1} M^{-1}$	[cf. Groban et al., 2009]
k_{SH3}	k_f	1×10^5	$s^{-1} M^{-1}$	SH3 domain/ligand binding [Solomaha et al., 2005]
	k_r	$k_{f:SH3}(0.1 \times 10^{-6})$	s^{-1}	$K_D = 0.1$uM [Whitaker and Dueber, 2011]
k_{LZX}	k_f	1×10^5	$s^{-1} M^{-1}$	Leucine zipper binding
	k_r	$k_{f:LZX}$ (0.01×10^{-6})	s^{-1}	$K_D = 0.01$uM [Whitaker and Dueber, 2011]
$k_{Sc:HK}$	k_f	$4\ k_{f:SH3}$	$s^{-1} M^{-1}$	Scaffold binding to HK with 4 SH3 domains
	k_r	$k_{r:SH3}$	s^{-1}	
$k_{Sc:RR}$	k_f	$k_{f:LZX}$	$s^{-1} M^{-1}$	Scaffold binding to RR with 1 LZX domain
	k_r	$k_{r:LZX}$	s^{-1}	
$k_{Ph:RR}$	k_f	1×10^5	$s^{-1} M^{-1}$	Phosphatase binding to RR_p
	k_r	1×10^{-3}	s^{-1}	
Closed Loop Anti-scaffold interactions				
$k_{Sc:AS}$	k_f	$k_{f:LZX} + k_{f:SH3}$	$s^{-1} M^{-1}$	Scaffold binding to anti-scaffold
	k_r	$0.001\ k_{r:LZX}$	s^{-1}	
$k_{AS\text{-}SH3}$	k_f	$k_{f:SH3}$	$s^{-1} M^{-1}$	Anti-scaffold binding to Sc:RR complex
	k_r	$0.001\ k_{r:SH3}$	s^{-1}	
$k_{AS\text{-}LZX}$	k_f	$k_{f:LZX}$	$s^{-1} M^{-1}$	Anti-scaffold binding to Sc:HK complex
	k_r	$0.001\ k_{r:LZX}$	s^{-1}	

Note: In the open loop circuit there is no anti-scaffold, so the closed loop anti-scaffold interaction rates are all zero.

Table 2.1: Table of model parameters. Parameters estimated from the literature are cited.

Modeling dynamics and steady state circuit behavior

The circuit was modeled using differential equations with all chemical reactions between species explicitly defined. The model omits transcriptional activity and accounts only for protein level behavior. With the exception of the anti-scaffold production term, all other terms are derived from mass action kinetics. A basic model of the circuit was previously published (Santos, Hsiao, and Richard M Murray, 2013). Here, we have expanded the model by adding the phosphatase species and all accompanying reactions. The 25 species arise from combinations of scaffold (Sc),

response regulator (RR), histidine kinase (HK), anti-scaffold (AS), and phosphatase (Ph) binding complexes. In total, the model consists of 80 reactions, 25 differential equations, and 26 parameters (See SI for complete list of chemical reactions). Many parameters (Table 2.1) were selected from experimental values found in the literature (Pazy et al., 2009; Groban et al., 2009; Solomaha et al., 2005), and others were estimated within a physiologically reasonable range.

Model reactions can be classified into five categories: production and degradation, phosphorylation, scaffold complex formation, activation, and irreversible sequestration. Phosphorylated species are denoted with a subscript p (e.g. RR_p), and complexes are denoted with a colon separating the participating species (e.g. Sc : AS). Though the possibility of modeling the scaffold as an enzyme-like species was considered, we could not assume that either the kinase or response regulator would always be in excess, a requirement of the substrate in a Michaelis-Menten reaction. Therefore, Michaelis-Menten kinetics were deliberately avoided.

The production rate, β, of the scaffold, histidine kinase, response regulator, and phosphatase are determined by user input of the total steady state value (in nM) multiplied by the degradation/dilution rate γ. This ensures constant concentration of these species in solution. The degradation rate γ is applied universally for all species and is estimated based on a cell division time of 30 minutes (Groban et al., 2009).

The phosphorylation reactions describe the autophosphorylation of HK and dephosphorylation of RR_p. Key reactions that describe this process are:

$$\mathrm{HK} \underset{k^r_{\mathrm{HK_p}}}{\overset{k^f_{\mathrm{HK_p}}}{\rightleftharpoons}} \mathrm{HK_p}, \tag{2.1}$$

$$\mathrm{RR_p} + \mathrm{Ph} \underset{k^r_{\mathrm{Ph}}}{\overset{k^f_{\mathrm{Ph}}}{\rightleftharpoons}} \mathrm{RR_p{:}Ph} \xrightarrow{k_{\mathrm{dephos}}} \mathrm{RR} + \mathrm{Ph}.$$

The phosphatase forms a complex with the RR_p prior to dephosphorylation. We model both phosphorylation and dephosphorylation with a two-step reaction model, an approach consistent with previous models (Huang and Ferrell, 1996). Rate constants for kinase phosphorylation and dephosphorylation of the response regulator were chosen based on cognate and noncognate phosphorylation rates measured for natural two-component systems, and occur on the order of seconds (Groban et al.,

2009). The following equations show phosphorylation in the absence and presence of scaffold:

$$\mathrm{HK_p} + \mathrm{RR} \underset{k^r_{\mathrm{noncog}}}{\overset{k^f_{\mathrm{noncog}}}{\rightleftharpoons}} \mathrm{HK} + \mathrm{RR_p}, \tag{2.2}$$

$$\mathrm{Sc{:}HK_p{:}RR} \underset{k^r_{\mathrm{cogp}}}{\overset{k^f_{\mathrm{cogp}}}{\rightleftharpoons}} \mathrm{Sc{:}HK{:}RR_p}.$$

Reaction rates for scaffold complex formation were based on the kinetics of the protein-protein interaction domains SH3 domain/ligand and LZX/LZx. SH3 domain/ligand binding has an estimated association affinity K_d of 0.1 uM while leucine zippers have a K_d of approximately 0.01 uM (Posern et al., 1998; Acharya et al., 2002; Grunberg et al., 2010; Weston R. Whitaker and John E. Dueber, 2011). Here we have examples of histidine kinase and response regulator binding to scaffold via SH3 and LZX binding, respectively:

$$\mathrm{Sc} + \mathrm{HK_p} \underset{k^r_{\mathrm{SH3}}}{\overset{4k^f_{\mathrm{SH3}}}{\rightleftharpoons}} \mathrm{Sc{:}HK_p}, \tag{2.3}$$

$$\mathrm{Sc{:}HK_p} + \mathrm{RR} \underset{k^r_{\mathrm{LZX}}}{\overset{k^f_{\mathrm{LZX}}}{\rightleftharpoons}} \mathrm{Sc{:}HK_p{:}RR}.$$

A phosphorylated response regulator becomes an active transcription factor. We considered all possible complexes with $\mathrm{RR_p}$ as possible activators (shown as $\mathrm{RR_{active}}$ Since the response regulator, CusR, dimerizes upon phosphorylation, the total rate of AS production, k_{fAS}, is modeled as a second-order Hill function:

$$\emptyset \xrightarrow{k_{fAS}} \mathrm{AS}, \tag{2.4}$$

$$k_{fAS} = \beta_{AS}\left[\beta_0 + \left(\frac{\mathrm{RR^2_{active}}}{K_D^2 + \mathrm{RR^2_{active}}}\right)\right], \tag{2.5}$$

where $\mathrm{RR_{active}} = \mathrm{RR_p} + \mathrm{Sc{:}RR_p} + \mathrm{Sc{:}HK{:}RR_p} + \mathrm{Sc{:}HK_p{:}RR_p} + \mathrm{Sc{:}RR_p{:}AS}$.

The negative feedback component comes about through the irreversible sequestration of the scaffold once it has bound to the anti-scaffold. We made the assumption that the individual SH3 and LZX domains on the anti-scaffold bind independently, at the same rates as HK and RR binding. However, once either the SH3 or LZX component of the AS has bound to the Sc, this results in a local concentration of the free domain that is substantially higher than the K_D. Therefore, we assume the other domain quickly displaces any competing species and sequesters the entire Sc.

The effective irreversibility comes about through steric hinderance of competing HK and RR species, both of which only have one compatible binding domain to the Sc:

$$\begin{aligned} \text{Sc} + \text{AS} &\xrightarrow{k^f_{\text{Sc:AS}}} \text{Sc:AS}, \\ \text{Sc:HK} + \text{AS} &\underset{k^r_{\text{LZX}}}{\overset{k^f_{\text{LZX}}}{\rightleftharpoons}} \text{Sc:HK:AS} \xrightarrow{k^r_{\text{SH3}}} \text{Sc:AS} + \text{HK}, \\ \text{Sc:RR} + \text{AS} &\underset{k^r_{\text{SH3}}}{\overset{k^f_{\text{SH3}}}{\rightleftharpoons}} \text{Sc:RR:AS} \xrightarrow{k^r_{\text{LZX}}} \text{Sc:AS} + \text{RR}. \end{aligned} \tag{2.6}$$

The validity of the model was tested by comparing the open and closed loop circuits. In the open loop circuit, the negative feedback binding reactions are set to zero (Table 1). Experimentally, this was done by replacing the anti-scaffold with a fluorescent reporter alone. Figure 2.2A shows simulated steady state values for anti-scaffold (or fluorescent reporter) output over a range of scaffold concentrations (0 - 1000 nM), with either 0 nM or 100 nM of response regulator. In the cases with no response regulator, the circuit does not function and production of output is solely due to simulated leaky anti-scaffold production (β_0). When response regulator molecules are present, the open loop circuit output decreases significantly with increasing scaffold. Though it is not intuitive, this can be explained as the scaffold single occupancy effect (Weston R. Whitaker and John E. Dueber, 2011; Good, Zalatan, and Lim, 2011), where an overabundance of free scaffold leads to binding of only kinase or response regulator but not both. When we examine the prevalence of these intermediate species (Sc:HK, Sc:RR) in simulation, we can see that the total concentration of singly-bound scaffold increases, decrease in output is indeed observed (Figure BT-S1A). The same effect also occurs in the closed loop circuit, but much higher concentrations of scaffold are needed, since the anti-scaffold sequestration lowers the effective number of free scaffold molecules in solution (Figure BT-S1B).

Experimental data for the circuit closely recapitulated the model predictions (Figure 2.2B). First, without induction of RR for both open and closed loop circuits, there is no output YFP. Secondly, the open loop circuit shows the single scaffold occupancy effect at lower concentrations of scaffold. In the case of no scaffold induction, the open loop circuit has about four times more background than the closed loop circuit. This is due to leakiness in scaffold production in the absence of aTc. In the closed loop circuit, leaky production of scaffold is subdued by the negative feedback, while

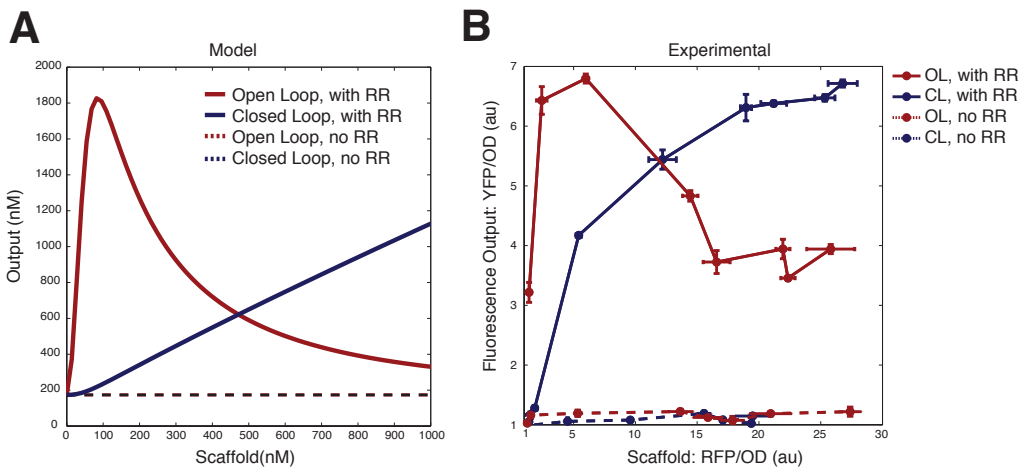


Figure 2.2: Open loop versus closed loop. A) Model predictions of scaffold circuit with and without negative feedback. Solid lines show anti-scaffold output over a range of scaffold concentrations (0 - 1000 nM) for open and closed loop circuits with constant response regulator (100 nM). Dotted lines show lack of output in the absence of response regulator. Open loop circuit shows scaffold single occupancy effect at lower levels of scaffold. B) Steady state experimental data of open and closed loop circuits with and without response regulator matches model predictions. Both sets of experimental data were normalized by the autofluorescence of a control *E. coli* strain (Figure BT-S2).

in the non-regulated open loop, we see significant production of YFP. All data was normalized to the autofluorescence of a control *E. coli* strain (Figure BT-S2).

We compared protein expression to fluorescence output to verify the use of fluorescence traces as a proxy for protein concentration. Western blot quantification was done with an analogous circuit containing a bicistronic scaffold(3xFLAG)/RFP and anti-scaffold-GFP(3xFLAG) (Figure BT-S3). mCherry is expressed from its own RBS instead of tethering directly to the scaffold (12 kDal) to provide a substantial size difference from the anti-scaffold (44 kDal). Quantification of band intensities show good agreement between anti-scaffold expression and measured fluorescence output (Figure BT-S4). These results served to validate both the model and the use of synthetic scaffolds as a tunable mechanism for negative feedback.

Characterization of step response

We characterized circuit response time by testing the closed loop response to step inputs. Using a programmable microfluidic plate (CellAsic) under a microscope, step induction of the scaffold protein was achieved by flowing in 0, 37.5, or 75

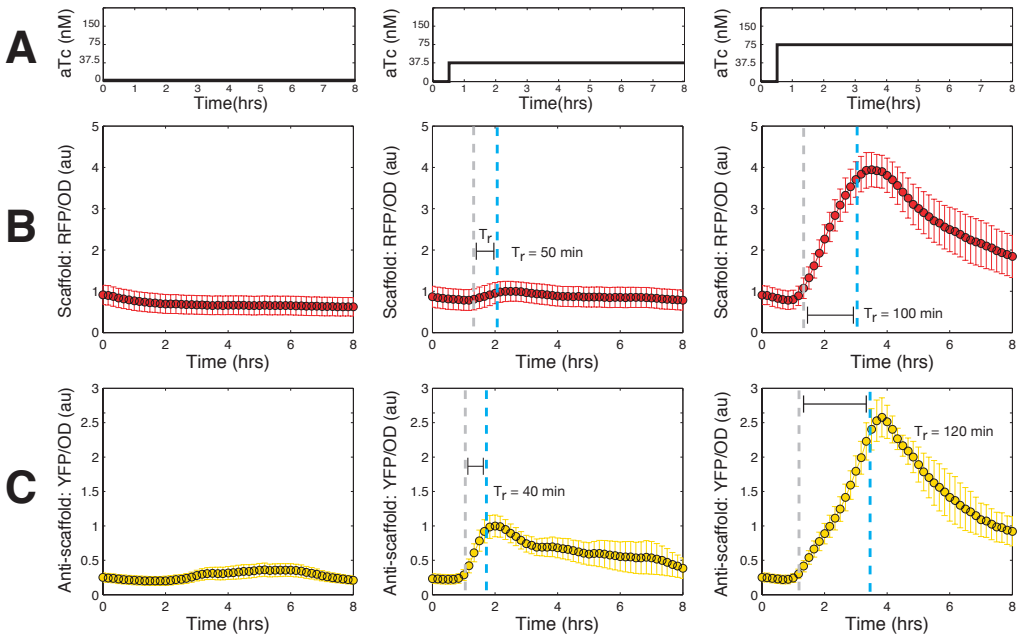


Figure 2.3: Step induction of closed loop circuit. A) aTc induction of Sc-RFP began 30 minutes after start of experiment and continued for the rest of the experiment. B) Scaffold-RFP/OD measurements for no induction (left), 37.5 nM induction (middle), and 75 nM induction (right). Response time (T_r) is quantified by finding the time needed for fluorescence to increase from 10% (gray dotted line) to 90% of the maximum value (blue dotted line). A two-fold increase in aTc results in a four-fold increase in scaffold expression and a two-fold increase in response time. The insets show growth curves for each condition. C) AS-YFP/OD measurements show 2.5 fold increase between the two inputs and a three-fold increase in response time. Fluorescent measurements are normalized such that the maximum of the middle column (37.5 nM aTc) is 1 a.u. to better visualize fold change.

nM of aTc (Figure 2.3A) after 30 minutes of growth in normal media. Cellular production of response regulator and phosphatase was pre-induced by incubating cells with arabinose and salicylate. Microscopy analysis methods are described in Figure BT-S5. Growth curves for all the conditions are in Figure BT-S7. In all conditions, expression of scaffold-RFP (Figure 2.3B) began about 30 minutes after induction, and occurred almost simultaneously with that of anti-scaffold-YFP (Figure 2.3C). Although we had selected mCherry and Venus-YFP on account of their similar maturation times, Venus still matures faster than mCherry (5 and 15 min, respectively). We believe that although there is a delay in mCherry maturation, the scaffold is immediately functional, leading to the near overlap of RFP and YFP expression. In order to better visualize the fold change, fluorescence output is

normalized by the maximum value of the lowest step input.

Response times (T_r) for fluorescent detection of scaffold (RFP) and anti-scaffold (YFP) were quantified. In control theory, response time is the amount of time needed for an output signal to increase from 10% to 90% of its final steady state. As cells reach stationary phase, circuit expression gradually turns off, and no steady state in fluorescence output is maintained. The 0 nM aTc case shows basal expression of the fluorescent proteins. We observed that scaffold induction, regulated by a P_{tet} promoter, has a 4-fold expression increase between 37.5 nM (Figure 2.3C) and 75 nM (Figure 2.3D) induction, but only a 2-fold increase in response time (50 min to 100 min). anti-scaffold output, regulated by the scaffold concentration, shows a 2.5 fold increase in maximum expression and a 3-fold increase in response time (40 min to 120 min).

This step input characterization revealed that scaffold and anti-scaffold fluorescence could be observed almost simultaneously about one cell cycle (30 min) after aTc induction of scaffold transcription. Following induction of the circuit, the response time to maximum expression increases in a linear-like fashion with increasing scaffold induction.

Circuit closely follows three step induction

Following step input characterization, we investigated circuit response to multiple step-up inputs. Figure 2.4 shows the results of a three step scaffold induction experiment with one hour steps corresponding to 50 nM increases of aTc inducer. Growth curves are shown in Figure BT-S8. The single negative feedback loop in the circuit represses overproduction of anti-scaffold but there is no mechanism for feedback in the case of an excess of scaffold or anti-scaffold. As such, the model predicts that increases in inducer will lead to immediate increases of scaffold followed closely by the anti-scaffold but once induction is turned off, degradation of proteins depends on the endogenous ClpXP degradation machinery (Figure 2.4A). Additionally, the upward slope of each curve should overlap until induction ceases.

Step-up induction was performed on cells pre-incubated in arabinose and salicylate, activating expression of response regulator and phosphatase, respectively. As shown in Figure 2.4B, experimental results for a three step induction are consistent with model predictions, and show overlapping curves during the ascent, with each individual curve dropping off slowly as induction ceases. The chemical induction of the scaffold produces a much smoother output curve compared to the response

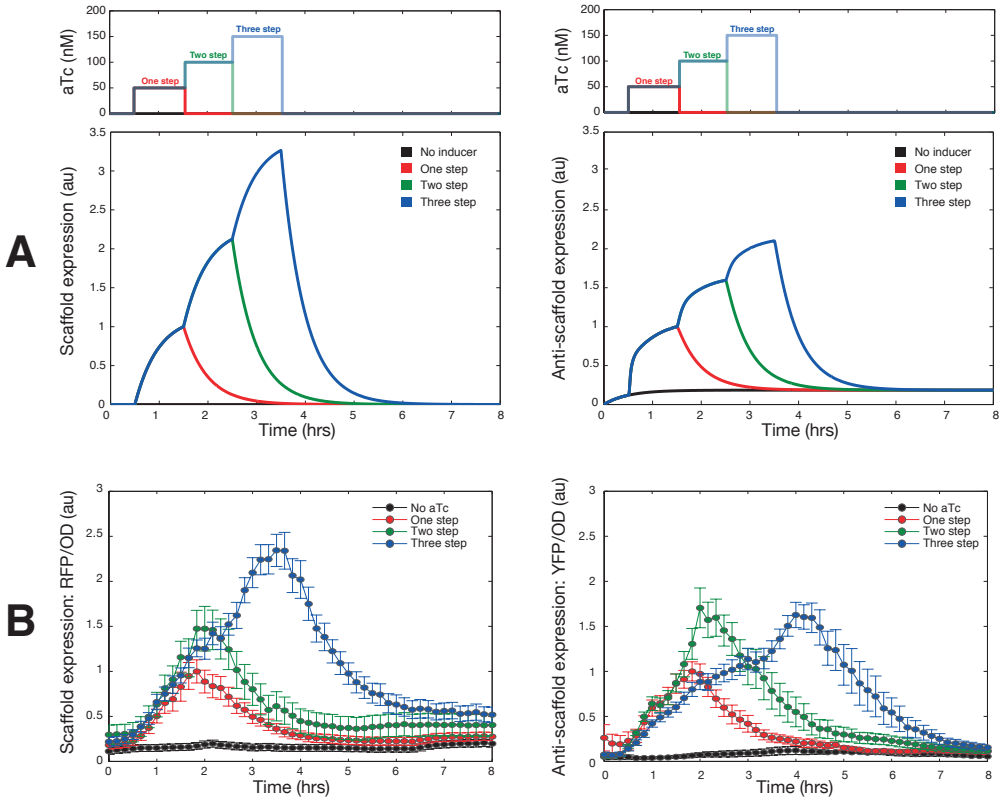


Figure 2.4: Multi-step induction of tracker circuit. A) Simulation results for a three step induction show overlapping response times with each curve decreasing based on degradation rate after induction ceases. Upper panel shows aTc induction pattern with one hour steps increasing in 50 nM increments starting 30 minutes after start of experiment. B) Experimental time traces for Sc-RFP show overlapping fluorescence output, with each curve decreasing at a time proportional to the number of steps. Corresponding anti-scaffold-YFP data show similar overlaps and proportional decreases. Fluorescent measurements are normalized such that the maximum value of the one step curve is 1 a.u. to better visualize fold change. Growth curves are shown in Figure BT-S8.

regulator-modulated anti-scaffold. Due to high levels of leaky expression, the open loop circuit did not respond to multi-step inductions (Figure BT-S8).

Inducer diffusion rates contribute to cumulative effect of sequential pulses

We observed in our model that variations in inducer diffusion rate would greatly affect the outcome of sequential pulses (Figure 2.5A). The removal of aTc from the cytoplasm and surrounding media is not instantaneous, and induction does not

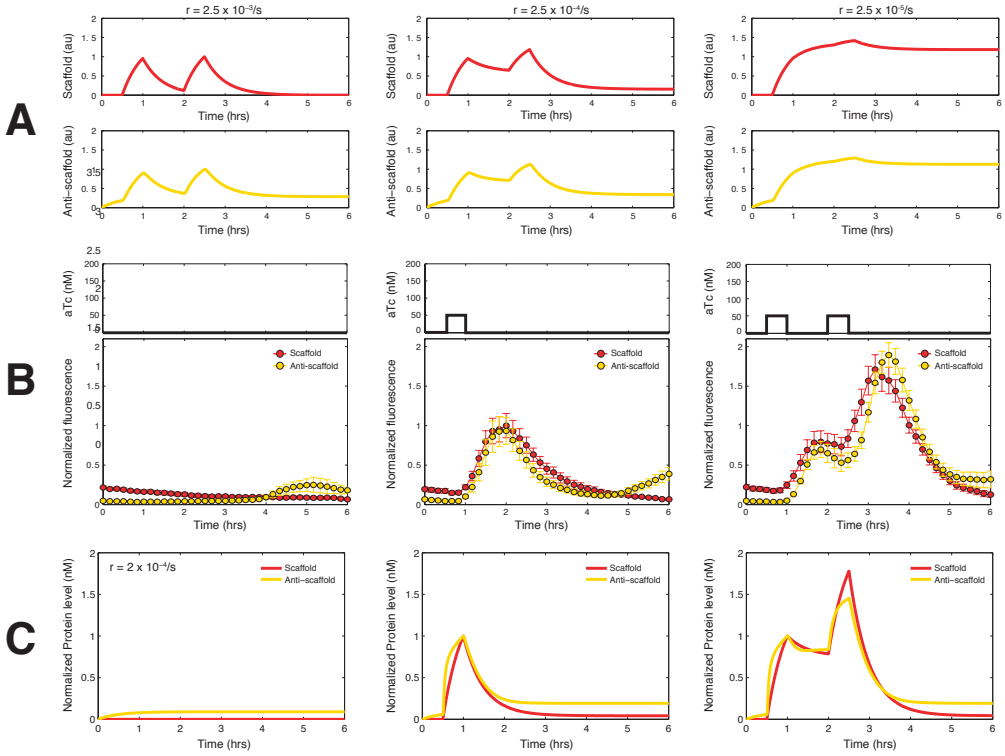


Figure 2.5: Two pulse induction of circuit. A) Model results for a range of inducer decay constants from 2.8×10^{-3} to 10^{-5}/s. Fast diffusion (left) shows two independent pulses, intermediate diffusion (middle) results show some overlapping protein from first and second pulses, and slow diffusion (right) shows large amounts of overlapping protein from the first to the second pulse. B) Experimental data for zero, one, and two pulses of 50 nM aTc. Data are normalized by maximum of single pulse induction (middle column). C) Simulations with improved inducer diffusion rates ($r = 2 \times 10^{-4}$/s).

go to zero. Growth curves are shown in Figure BT-S9. Given two sequential 30 minutes pulses spaced one hour apart, the diffusion constant determined whether two independent, identical outputs occurred, or if an additive effect would take place. Essentially, if the first pulse of inducer is not given sufficient time to diffusion out of environment, aTc molecules from the first pulse are still present when the second pulse occurs. We modeled inducer diffusion following a pulse with an exponential decay term, $\beta_{Sc} = \beta_{ind} \exp(-rt)$ (Munsky, Trinh, and Khammash, 2009). Figure 2.5A shows two pulse simulation results when the default decay constant ($r = 2.8 \times 10^{-4}$ /s, middle column) is increased or decreased by 10-fold.

When we tested two pulse induction *in vivo* (Figure 2.5B), we ran simultaneous

experiments with zero, one, and two 30 minute pulses of aTc(50 nM). The single pulse fluorescence maximum (Figure 2.5B, middle column) was normalized to 1 a.u. It is clear from the two pulse fluorescence output data that the diffusion rate of aTc after a pulse *in vivo* was actually much slower than expected *in silico*. In fact, so much of the scaffold from the first pulse remained that there was almost a two fold increase in maximal expression during the second pulse. This was an effect which had not been apparent previously during the multi-step inductions, where we showed sequential increases in inducer concentration. These data show that modulation of pulse frequency, but not concentration, can result in the same additive effect as increasing inducer concentration.

We then sought to improve our model by tuning the inducer decay constant (Figure 2.5C), generating outputs which demonstrated the nearly two-fold increase observed *in vivo*. Although the optimized decay rate ($r = 2 \times 10^{-4}$/s) better captured gene expression during log phase, we consistently observed a rapid decrease in fluorescence as cells approached stationary phase. We believe this is due to upregulation of ClpX and other ssrA machinery in stationary phase (Farrell, Grossman, and Sauer, 2005). This resulted in improved model performance when simulating dynamic circuit behavior.

Model-informed exploration of parameter space

Circuit limitations were explored *in silico*. Specifically, we investigated the effects of tuning response regulator and phosphatase concentrations on the ability of the anti-scaffold output to track the scaffold reference. Response regulator and phosphatase concentrations are easily accessible parameters via inducible promoters in our experimental system. In Figure 2.6A, a scan of input-output response curves is shown over a range of response regulator and phosphatase concentrations (See Figure BT-S10 for explicit values). For each curve in the grid, the scaffold concentration in which the single occupancy drop-off occurs was found, and the slope of the curve up to that concentration was found with a linear fit. The maximum scaffold occupancy limit is the concentration of scaffold molecules at which each scaffold molecule only has either a response regulator or histidine kinase. The slope of the curve up to that point represents the anti-scaffold to scaffold ratio which can be achieved by the circuit. In the case where the single occupancy limit does not appear, the last concentration is used. Data shown in Figure 2.6B indicates that increasing response regulator values result in a greater AS/Sc ratio (up to 1.5 fold increase), while increasing phosphatase serves to bring down that ratio. The

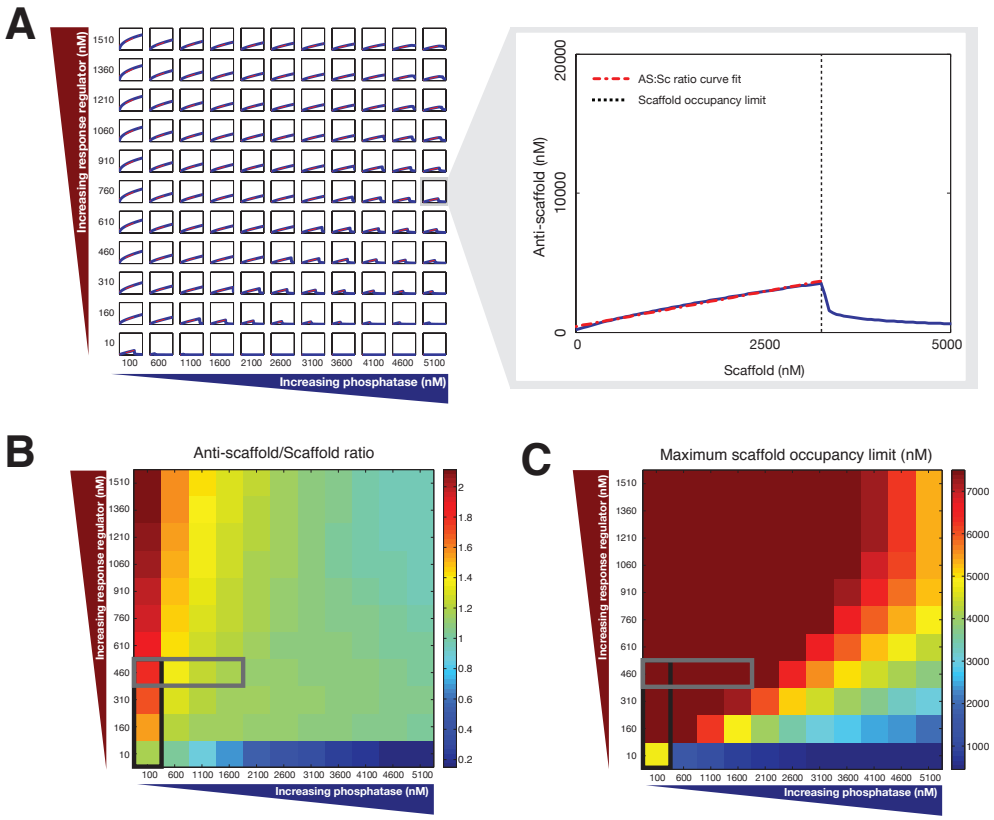


Figure 2.6: Model-based exploration of parameter space. A) Simulations of scaffold to anti-scaffold inputs and outputs over a range of phosphatase (100 - 5100 nM, 500 nM increments) and response regulator (10 - 1510 nM, 150 nM increments) concentrations. Enlargement shows the scaffold single occupancy limit concentration and curve fitting for each curve. Red dotted lines show curve fits - the slope represents the anti-scaffold to scaffold ratio. B) Heat map showing anti-scaffold to scaffold ratio for each curve shown in part A. Increasing response regulator results in greater AS/Sc ratios. Gray box represents estimated experimental phosphatase induction range. Black box estimates experimental response regulator induction range. C) Heat map of maximum scaffold occupancy limit. Higher concentrations of phosphatase result in decreased maximum scaffold occupancy limit.

effect of increasing phosphatase is apparent when the maximum scaffold occupancy limit is examined (Figure 2.6C). Furthermore, the simulations show that some minimal amount of phosphatase is necessary for a sufficiently high response regulator turnover rate so as to approach a 1:1 ratio. As phosphatase concentration increases, active response regulators are quickly dephosphorylated, decreasing the efficacy of the scaffolds, lowering the maximum occupancy concentration, and making the drop-off more steep. Based on experimental outputs of our circuit, however, we

believe the actual achievable dynamic range of the circuit is limited to the lower left corner of the parameter space. The qualitatively estimated induction range is shown with the black and gray rectangles in Figure 2.6B and 2.6C.

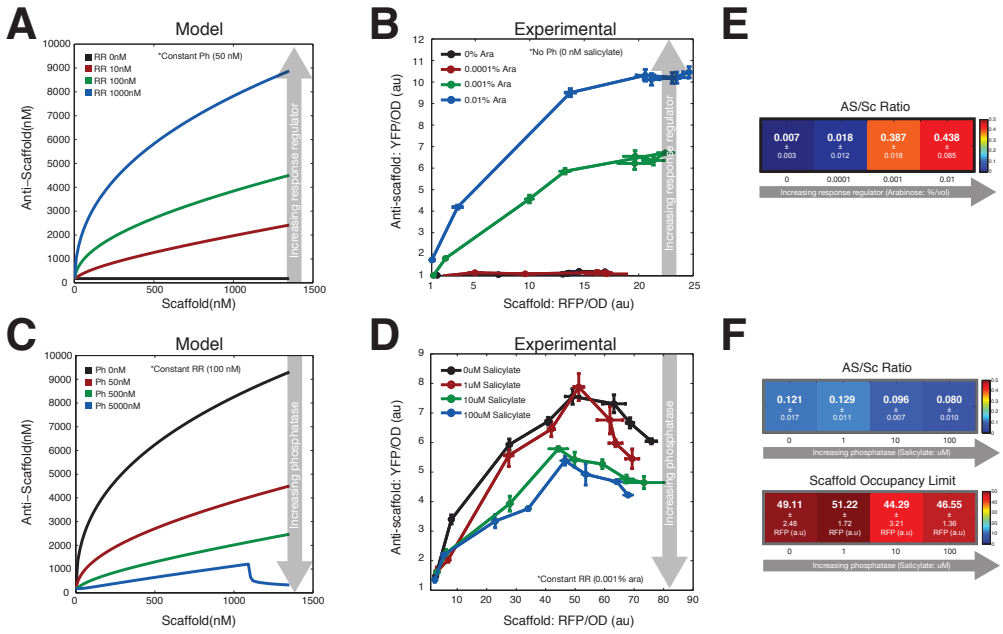


Figure 2.7: Steady state experimental tuning of response regulator and phosphatase. A) Simulation data of input-output curves with increasing response regulator concentrations (0 - 1000 nM). Increasing response regulator increases the scaffold occupancy limit as well as overall AS/Sc ratio. B) Experimental data of steady state scaffold to anti-scaffold curves with ten-fold increases in response regulator induction (0 - 0.01 % arabinose). There was no additional induction of phosphatase (0 nM salicylate). C) Simulation data of input-output curves with increasing phosphatase concentrations (0 - 5000 nM) with constant response regulator concentration of 100 nM. Increasing phosphatase decreases the scaffold occupancy limit and overall AS/Sc ratio. D) Experimental data of steady state circuit behavior with ten-fold increases in salicylate. Response regulator concentration is constant (0 - 100 uM salicylate). E) Ratios of YFP/RFP ratios from (B) as a proxy for As/Sc ratios with increasing response regulator. Scaffold occupancy limit was not observed in response regulator experiments. F) Ratios of YFP/RFP ratios and scaffold occupancy limit values from (D) with increasing phosphatase. All experimental data was normalized by baseline auto-fluorescence values.

By modulating response regulator and phosphatase concentrations, a range of maximal expression levels for scaffold and anti-scaffold can be achieved. Figure 2.7A and B shows steady state circuit response to varying levels of response regulator induction in both the model and experimental circuit. Increasing RR concentra-

tions increases the gain of the system by increasing the number of available active transcription factors for the AS promoter. In simulation data (Figure 2.7A), we see that the scaffold occupancy effect is mitigated by higher levels of response regulator. This is consistent with our previous explanation, since more regulator means almost all free scaffold molecules will exist as Sc:RR. Experimental data for tuning response regulator concentration via ten-fold increases of arabinose (Figure 2.7B) do not extend the scaffold levels far enough to show the occupancy effect, but the increasing output gain is evident.

The presence of phosphatase in the circuit modulates the amount of time that phosphorylated response regulator is active. As such, tuning phosphatase concentrations changes RR $\longleftrightarrow$ RR_p cycling time. Early versions of the circuit did not include the phosphatase species (Santos, Hsiao, and Richard M Murray, 2013), and we were unable to observe dynamic behavior due to buildup of RR_p. Using the model, we explored the effects of adding a phosphatase prior to testing *in vivo*. Figures 2.7C and D shows steady state responses across a range of phosphatase concentrations. Simulation results show that increasing phosphatase decreases overall circuit output (Figure 2.7C) by decreasing the average time RR_p is active. Experimental results (Figure 2.7D) support model predictions and show this suppression of output with increased induction via salicylate.

In Figure 2.7E and F, these experimental steady state data are analyzed using the same techniques shown in Figure 2.6. Figure 2.7E shows anti-scaffold to scaffold ratio and scaffold occupancy limit as calculated based on RFP/YFP fluorescence data with ten-fold increases in response regulator induction with no phosphatase present. Similar to the analysis used in the model, if the single occupancy drop is not observed, the highest scaffold concentration is taken. Figure 2.7F shows the same metrics with ten-fold increases in phosphatase induction with constant response regulator (0.001% arabinose induction). Experimental data is presented as a function of fold change from background fluorescence, and so cannot be compared directly with model data (presented in nM). However, the overall trends are in agreement. As response regulator increases, we see a significant increase in anti-scaffold to scaffold ratio, and little change in the occupancy limit. With increasing phosphatase, we see a slight decrease in AS/Sc ratio and scaffold occupancy limit. We believe these data show us that our experimental range occupies only a small fraction of that shown by our model (Figure 2.7B,C), and that these limitations are due to the limited dynamic range of the inducible promoters (P_{BAD}-RR, P_{sal}-Phos,

Figure BT-S11).

Scaffold-based circuits for rapid feedback

We have designed a novel negative feedback tracker circuit using modular synthetic scaffold proteins and a two-component system with scaffold-dependent phosphorylation. The use of scaffold proteins for negative feedback could potentially be a robust way of linking modules and ensuring constant performance despite intrinsic and extrinsic noise. scaffold proteins have been shown to be powerful hubs for organization of regulatory feedback in natural networks, usually by colocalization of phosphorylation machinery (Good, Zalatan, and Lim, 2011). Previous studies have rewired the naturally occurring Ste5 scaffold in the yeast MAPK cascade to redirect signals, to modify delays in signaling time, and to introduce ultra sensitivity (Park, Zarrinpar, and Lim, 2003; Bashor et al., 2008). The modular scaffold proteins used in this study were previously used to control phosphotransfer to non-cognate response regulators, building a synthetic signaling pathway (John E Dueber et al., 2009; Moon et al., 2010; Weston R. Whitaker and John E. Dueber, 2011). Here we have taken those same scaffolding modules and built an entirely synthetic feedback circuit. The system allows for tunable control of output gain and cycling time. Most importantly, the proportional anti-scaffold tracking of the scaffold is maintained over a range of component concentrations.

After we designed the circuit framework, we constructed and then experimentally validated an ODE-based mathematical model. Through selection of parameters and reaction rates based on the literature, we obtained a model able to reasonably predict circuit behavior. Comparisons between simulation and experimental data confirmed the presence of scaffold-mediated negative feedback, and we used the model to scan the parameter space in a way that would have been time and resource intensive to explore *in vivo*. We found that steady state circuit gain can be tuned by changing response regulator concentrations and cycling time is controlled by varying phosphatase levels, observations which were supported by experimental data. Following initial step induction system characterization of step input response time, expression of both the reference (Sc-RFP) and output (AS-YFP) protein was shown to be fast and responsive to multi-step inputs. Finally, we found that pulse-modulated induction could result in additive circuit response, leading to improvement of the model through more accurate inducer diffusion parameter values.

A scaffold-based biomolecular tracking circuit has potential applications in active

regulation of component expression in synthetic circuits. The relatively small size (approx. 60 AA) of the scaffold and anti-scaffold proteins facilitates attachment to larger proteins, represented in this work by mCherry-RFP and Venus-YFP. Rather than open loop tuning of regulatory sequences and large-scale screening, scaffold-based negative feedback could be utilized. By attaching the scaffold to a native protein, it may also be possible to tie synthetic circuit inputs to naturally occurring cycles *in vivo*. It is well known that many natural cell processes such as developmental segmentation, circadian clocks, and stem cell multipotency involve oscillatory gene expression (Bessho, 2003; Imayoshi et al., 2013).

Furthermore, response to signal transduction may be modulated not by amplitude, but by frequency (Cai, Dalal, and Elowitz, 2008). We have shown that the scaffold-modulated protein tracker follows changes in both amplitude and frequency, and exhibits good agreement with a mass-action model. Future iterations of this design may improve tracking fidelity by including reverse feedback loop to compensate for over-expression.

2.4 Materials and Methods

Cell strain, media

The circuit was implemented in the *E.coli* cell strain WW62, a variant of BW27783 (CGSC 12119) with knockouts of EnvZ, OmpR, CusS, CusR, CpxA, and CpxR. All cell culture was done in optically clear MOPS EZ Rich defined medium (Teknova, M2105), with 0.4% glycerol instead of 0.2% glucose. The use of glycerol as a carbon source was done to prevent interference with the arabinose induction of the P_{BAD} promoter.

Tested arabinose induction levels were 0, 0.0001%, 0.001%, 0.01%, and 0.1% (20% stock solution). Anhydrotetracycline (aTc) was diluted in media at concentrations of 0, 5, 15, 30, 60, 90, 120, 150 nM. Sodium salicylate was resuspended at a stock concentration of 100 mM and diluted 1:1000 in media for experiments.

Plasmids

Plasmids used in this study were derived from those used in Whitaker *et al.*(Weston R Whitaker et al., 2012).The plasmid encoding the SH3-ligand-LZX-mCherry scaffold (pVH001) has a high copy backbone(ColE1) with ampicillin resistance. The CusR-LZx response regulator and SH3-domain-LZx-VenusYFP anti-scaffold plasmids (pVH003 for closed loop, pVH009 for open loop) are on a medium copy backbone (pBBR1) with kanamycin resistance. The 4SH3-domain-Taz histidine kinase and

CusS-G448A phosphatase are on a low copy plasmid (p15A) with chloramphenicol resistance. Detailed plasmid maps are shown in Figure BT-S11, and a complete list of plasmids and strains can also be found in the Supplementary Materials.

Plate reader experiments

Plate reader data were collected on a Biotek H1MF machine using the kinetic read feature. Cell were grown in two consecutive overnight cultures in MOPS EZ rich media. On the day of the experiment, overnight cultures were diluted 1:40, and grown to OD $\sim$ 0.1 prior to start of experiment. Cells were incubated in the plate reader at 37C and shaken at 800rpm between reads. Measurements were taken every five minutes. Cells were grown in clear bottomed 96-well microplates (PerkinElmer, ViewPlate, 6005182) and sealed with breathable clear membranes (Sigma Aldrich, Breath-Easy, Z380059). mCherry was read at excitation/emission of 580/610 with gain 140, Venus was read at 500/540 with gain 100, GFP was read at 488/525 with gain 75.

Analysis of the data was done by taking fluorescence readings at late log phase for each independent well. Experimental conditions were done in triplicate and repeats were averaged. Fluorescence per OD was normalized by the fluorescence of a control strain (lacking mCherry/YFP/GFP) such that the cell autofluorescence equals 1 a.u. (Figure BT-S2). Error bars shown are standard error of the mean.

Western blots

Cultures were grown for five hours in a deep-well microplate at 37C with a range of aTc from 0 - 120 nM. Arabinose was kept constant at 0.001 %. No salicylate was used. After five hours, OD600, RFP, and GFP were measured in a plate-reader. Western blot samples were collected and spun down. Because aTc concentration can affect growth rates, the volume spun down was calculated based on OD to ensure consistent cell mass. Pellets were resuspended in lysis buffer and boiled for 10 minutes. Samples were run on 4-20% tris-glycine gels (Novex, 150V for 1 hr) and a semi-dry transfer apparatus was used (Bio-Rad, 15V for 20min) to transfer onto a PVDF membrane. Monoclonal anti-FLAG M2-peroxidase (HRP) antibody was diluted 1:88,000 in 5% milk. Blot imaging was done using the Chemi Hi Resolution setting on a BioRad ChemiDoc MP imager. Quantification of band intensity was done using Image Lab 5.0 (BioRad).

Microscopy

Step induction data were taken using the CellAsic ONIX microfluidic perfusion system for bacteria (B04A). The microscope is an Olympus IX81-ZDC enclosed in a custom heater box. Images were taken using a 100x oil immersion phase objective. Fluorescence filters are 580/630 for mCherry (Chroma 41027) and 510/560 for YFP (Chroma 31040 JP2). Microscope media was augmented with oxidative scavengers Trolox (200 nM) and sodium ascorbate (2 mM).

Overnight cultures were then diluted 1:500 in media containing arabinose (0.01%) and/or salicylate (100uM) four hours prior to loading in the CellAsic plate. This is to ensure steady state concentrations of response regulator and phosphatase prior to aTc induction of the scaffold. Cells were diluted 1:10 again before loading. Microscopy movies were taken inside a temperature controlled environment set at 37C, and images were taken at 10 minute intervals. Exposure time was 10 ms for brightfield and 500 ms for mCherry and YFP fluorescent channels.

Analysis of microscope movies was done using custom algorithms in ImageJ and MATLAB. For each frame, the phase image is converted to a binary mask of the cell colony. The mask is then used to find total mCherry and YFP fluorescence in the frame. After subtraction of background fluorescence, the total fluorescence is normalized by the total cell area (fluorescence intensity per pixel). For step induction experiments, fluorescence is normalized such that the maximum fluorescence of the lowest concentration induction is equal to 1 a.u. Figure BT-S5 shows the microscopy analysis workflow. Error bars shown in microscopy time trace data are the standard error of the mean between analysis of different positions (n = 7 to 10) on the same experimental plate.

Model implementation

The model was implemented using the Simbiology toolbox in MATLAB and the ode15 solver (See Supplementary files for MATLAB code).

2.5 Acknowledgements

VH, ELCS, and WRW designed the research. VH and ELCS created the mathematical model and performed analysis. VH performed cell culture experiments and analysis. VH wrote the manuscript. WRW provided initial plasmids and critical technical insight.

The authors thank Jongmin Kim for insightful discussion and suggestions on the

model. This research was conducted with government support under and awarded by DoD, Air Force Office of Scientific Research, National Defense Science and Engineering Graduate (NDSEG) Fellowship, 32 CFR 168a. Additional support was granted in part by the Benjamin M. Rosen Bioengineering Center, the NIH/NRSA training grant 5 T32 GM07616, the Gordon and Betty Moore Foundation through Grant GBMF2809 to the Caltech Programmable Molecular Technology Initiative, and the Institute for Collaborative Biotechnologies through grant W911NF-09-0001 from the U.S. Army Research Office.

The content of the information does not necessarily reflect the position or the policy of the Government, and no official endorsement should be inferred.

2.6 Supplementary information

Supplementary Figures

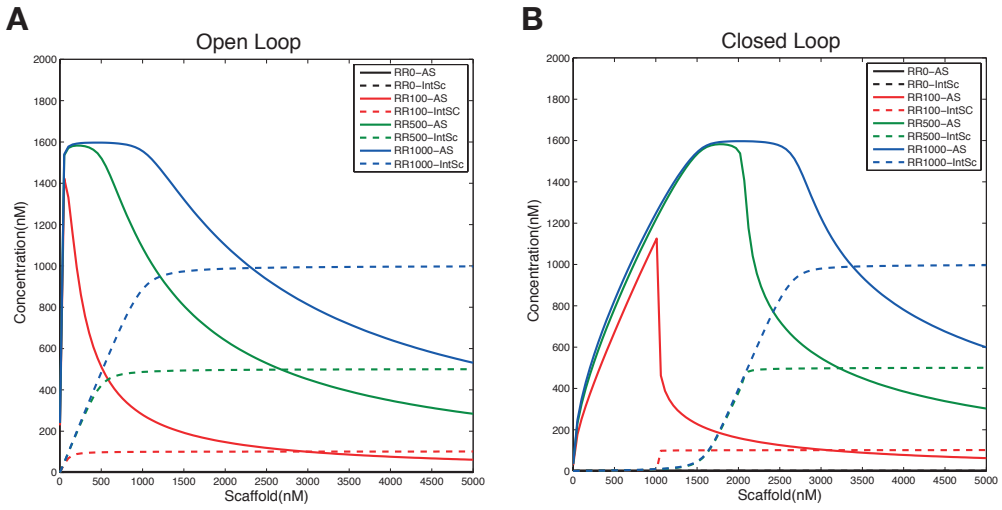


Figure BT-S1: Modeling of single occupancy effect. Dotted lines show intermediate scaffold complexes (labeled as IntSc). IntSc is the sum of Sc:RR, Sc:RR_p,Sc:HK, Sc:HK_p. Solid lines represent antiscaffold concentrations. It is apparent from the model that the drop off in steady stand antiscaffold concentrations corresponds with the increase in intermediate scaffold complexes. This is true for both the open and closed loop circuits. However, the closed loop circuit exhibits this behavior at much higher concentrations of scaffold relative to the open loop.

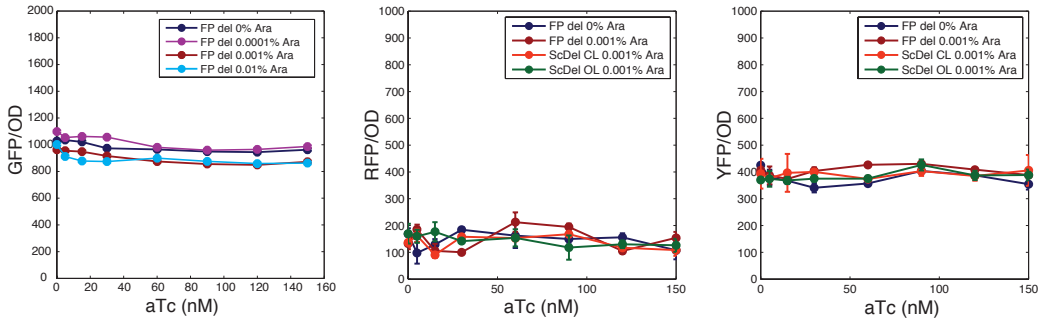


Figure BT-S2: GFP, RFP, and YFP autofluorescence of control *E. coli* strains. *FP del* shows a control strain transformed with the three plasmid system without fluorescence reporters. FP del provides the native autofluorescence of our particular *E. coli* strain, and these values are used for normalization of plate reader data. *Sc Del* is a control strain without the scaffold plasmid, and shows the level cross talk between the HK and RR in the absence of Sc. Auto-fluorescence of the *E. coli* is shown to be independent of arabinose and aTc concentrations.

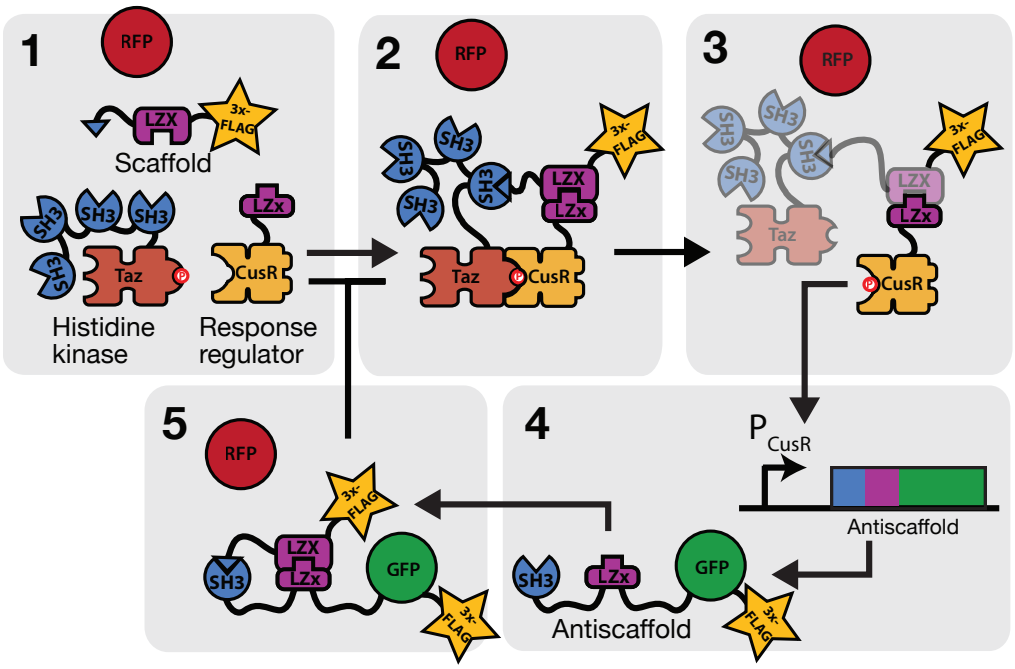


Figure BT-S3: Circuit used for Western blot analysis. The circuit used for Western blot analysis has the same architecture as the circuit in Figure 1 of the main text. The main differences are the bicistronic scaffold/mCherry and the use of GFP instead of YFP. mCherry is expressed from its own RBS instead of tethering directly to the scaffold (12 kDa) to provide a substantial size difference from the antiscaffold (44 kDa). The scaffold and antiscaffold are tagged with the 3xFLAG epitope. We expect the addition of the tag not to significantly affect circuit function. None of the molecules have degradation tags and there is no additional phosphatase in the system.

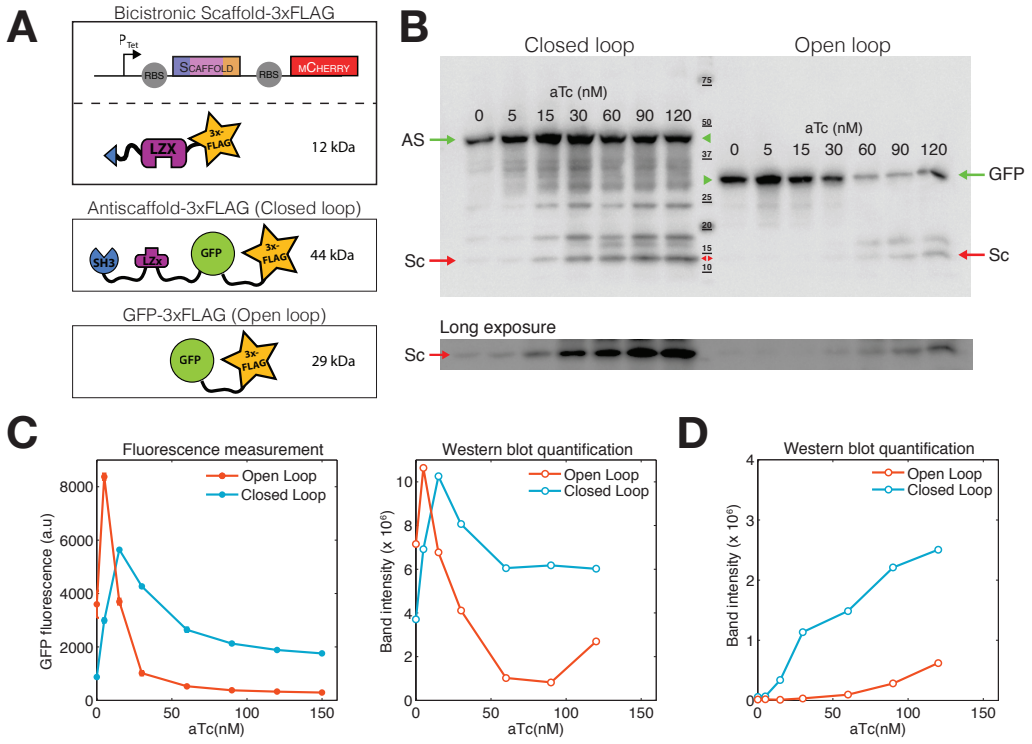


Figure BT-S4: Comparison of fluorescence data to Western blots. A) Both the scaffold and antiscaffold are tagged with the 3xFLAG-tag epitope. The scaffold is expressed on a separate RBS from the mCherry to create a size difference between the scaffold-3xFLAG (12 kDa) and the antiscaffold-GFP-3xFLAG (44 kDa) bands. B) Western blot (15s exposure) of closed loop versus open loop over a range of aTc induction values (0-120 nM). AS denotes antiscaffold band at 44 kDa, Sc denotes scaffold band at 12 kDa, and GFP denotes open loop GFP expression at 29 kDa. Longer exposure (42s) of the blot reveals the scaffold bands more clearly. C) GFP fluorescence versus Western blot quantification of antiscaffold-GFP-3xFLAG. Western blot results for the closed loop versus open loop circuits (Figure 3B) show bands at the expected molecular weights, with significantly darker bands for the antiscaffold (CL) / GFP (OL) than the scaffold (15s exposure). The fluorescence measurements of the experimental cultures were taken immediately prior to lysis for protein blotting. Comparison of GFP/OD and antiscaffold band intensity as a function of aTc induction shows very good agreement in output curves between measured fluorescence and antiscaffold concentrations in both the open and closed loop circuits. D) Western blot quantification of the scaffold-3xFLAG. Measured RFP/OD and band intensity of scaffold cannot be compared since RFP expression is independent of scaffold expression. The discrepancy between open loop and closed loop scaffold expression is due to sequestration of the scaffold in the Sc:AS complex, which we believe protects the scaffold from degradation in the closed loop circuit.

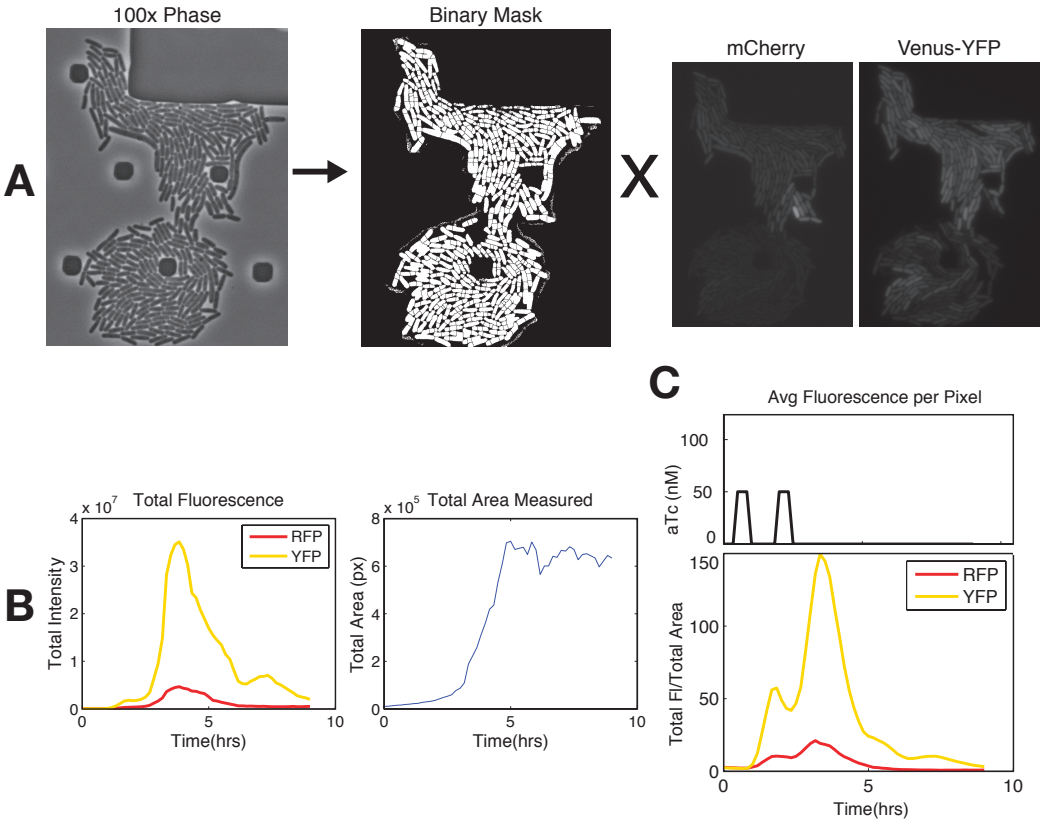


Figure BT-S5: Microscopy analysis overview. A) 100x phase images are converted into binary masks. CellAsic plate features such as the small and large square pillars shown in the figure, are filtered out in ImageJ. The remaining cell segmentation is saved as a binary .PNG file. The binary image is converted into a matrix in MATLAB, and used to extract only cell fluorescence from mCherry and Venus-YFP channels. The fluorescence background average is calculated from the area outside the mask and subtracted from every subsequent frame. B) Total RFP and YFP fluorescence and total cell area are extracted from each frame. C) The total fluorescence is divided by total area to find average fluorescence per pixel. The fluorescence traces shown in Figures 3, 4, and 5 are derived by averaging the average fluorescence per pixel of each frame, with an average of 5-7 frames per experimental condition. Microscopy analysis protocol based on a protocol established by the Phillips group at Caltech (`http://www.rpgroup.caltech.edu/courses/PBL/bootcamp2011/protocols_and_references/Matlab_Tutorial_2010.pdf`).

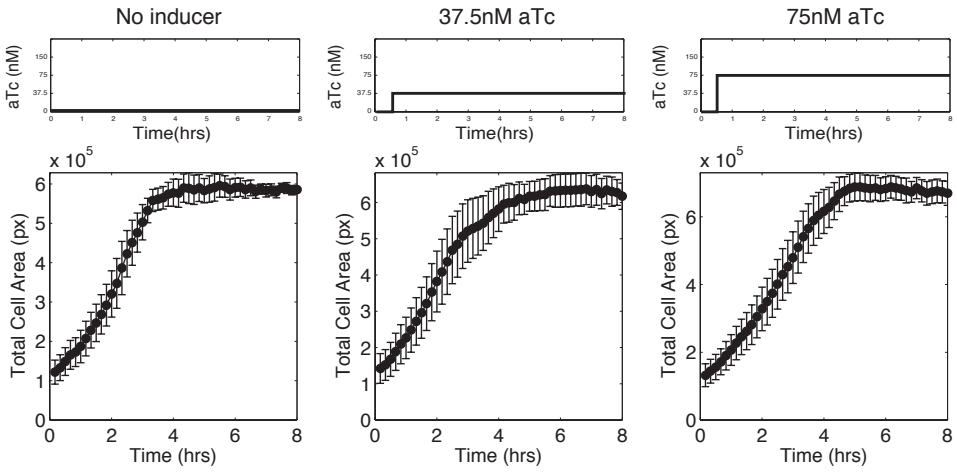


Figure BT-S6: Growth curves for step induction experiment shown in Figure 3 of the main text. Shown is the quantified total cell area extracted from phase contrast microscopy images.

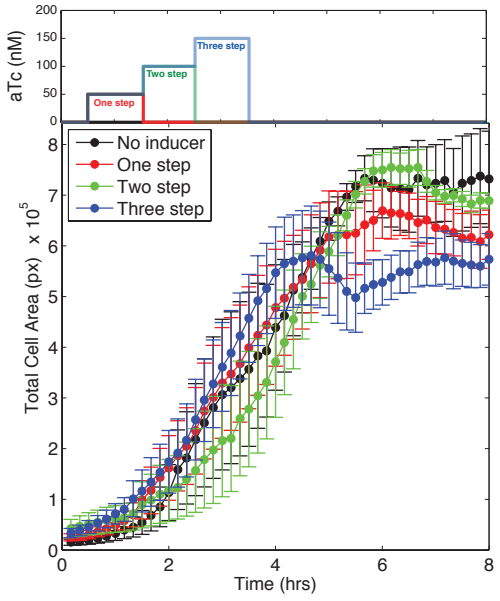


Figure BT-S7: Growth curves for multi-step induction experiment shown in Figure 4 of the main text. Shown is the quantified total cell area extracted from phase contrast microscopy images.

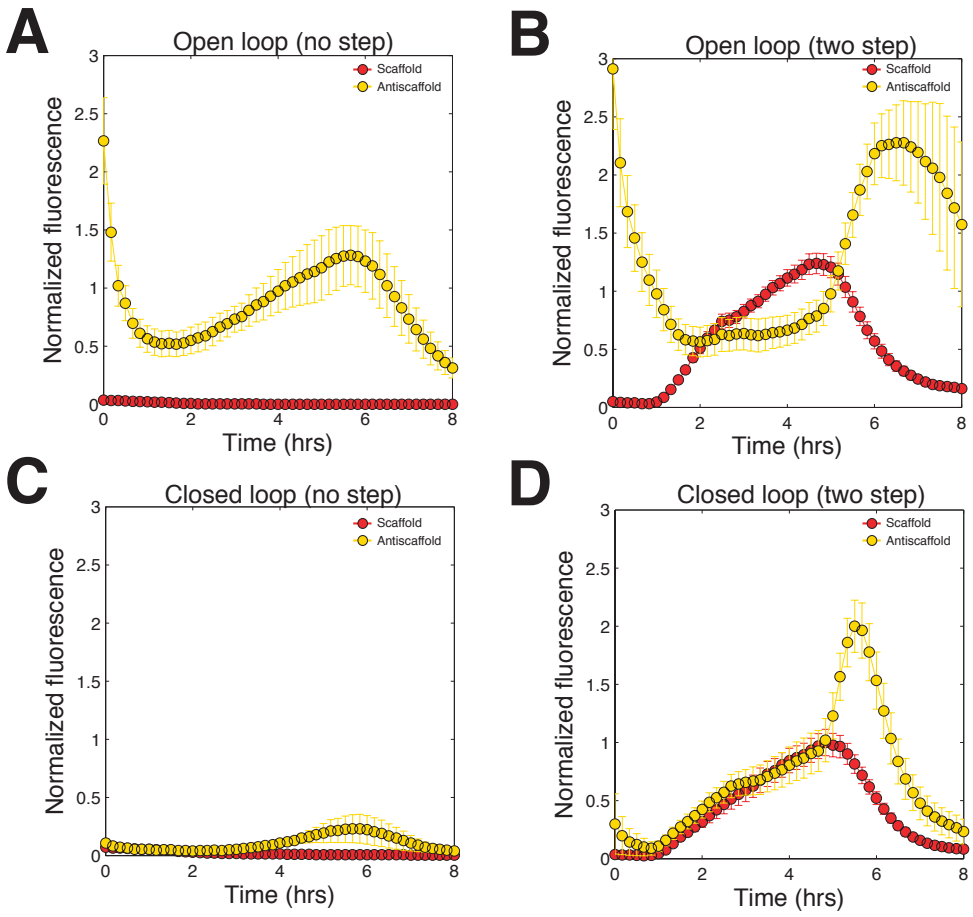


Figure BT-S8: A) Open loop circuit with no aTc induction. Red trace is scaffold-mCherry, yellow traces is antiscaffold-Venus. B) Open loop circuit with 30 minutes of media only, followed by 2 hours of media with 50 nM aTc, followed by 2 hours of 150 nM aTc, then back to media only. C) Closed loop circuit with no aTc induction. D) Closed loop circuit with two step induction described in (B). Fluorescence traces for all four panels are normalized by the closed loop maximum values. RFP is normalized by $\text{clRFP}_{\text{max}}$ and YFP is normalized by $\text{clYFP}_{\text{max}}/2$.

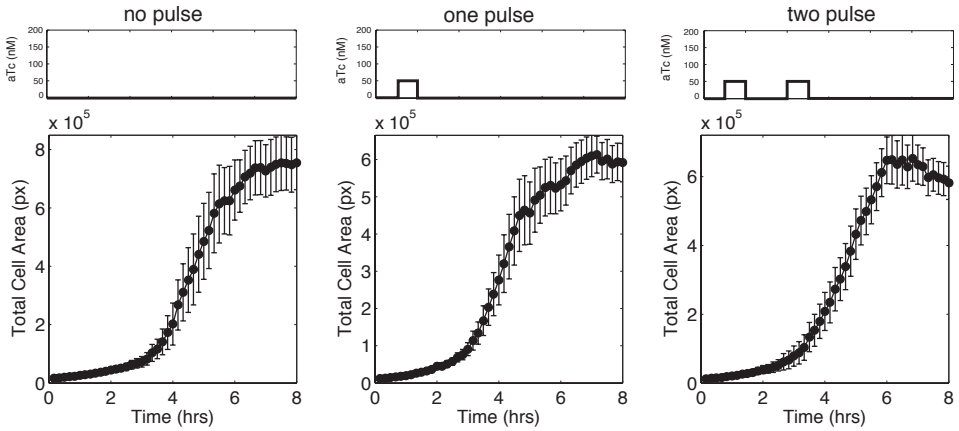


Figure BT-S9: Growth curves for oscillating induction experiment shown in Figure 7 of the main text. Shown is the quantified total cell area extracted from phase contrast microscopy images.

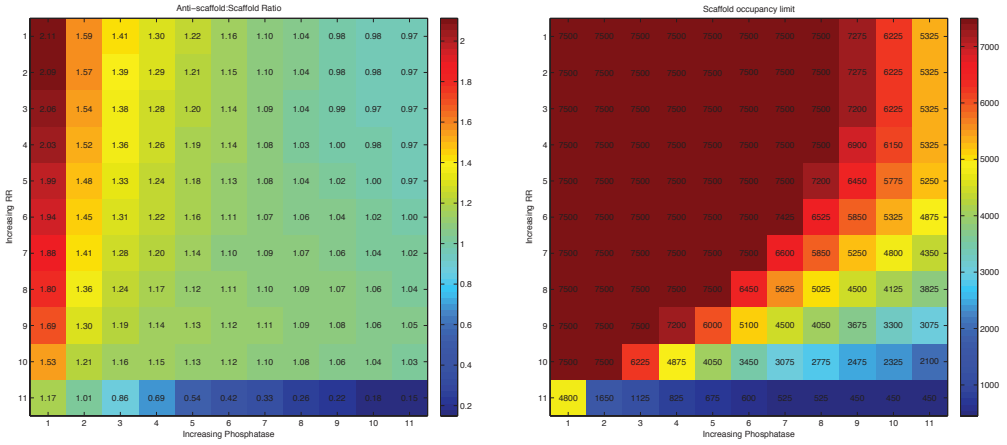


Figure BT-S10: Heat maps from parameter space exploration with values shown. The model was used to find the antiscaffold to scaffold ratio as well as the scaffold occupancy limit over an 11 x 11 matrix of response regulator and phosphatase values. Both RR and Ph values cover the range of 0 to 100 nM in 10 nM increments.

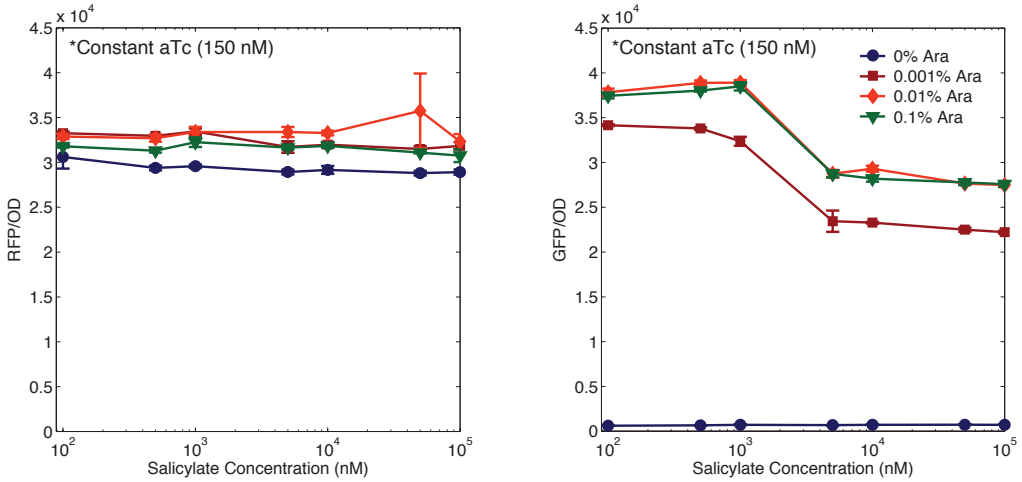


Figure BT-S11: P_{sal} response curve is shown over a range of salicylate levels (0, 1, 10, 100uM). The circuit used was an open loop circuit with P_{sal}-CusS(G448A) and P_{CusR}-GFP. RFP represents scaffold-mCherry, and GFP is the open loop output (no antiscaffold). As salicylate concentration increases, induction of CusS(G448A) phosphatase concentration increases. Increased CusS(G448A) activity decreases concentrations of activated CusR response regulator, resulting in lowering levels of GFP.

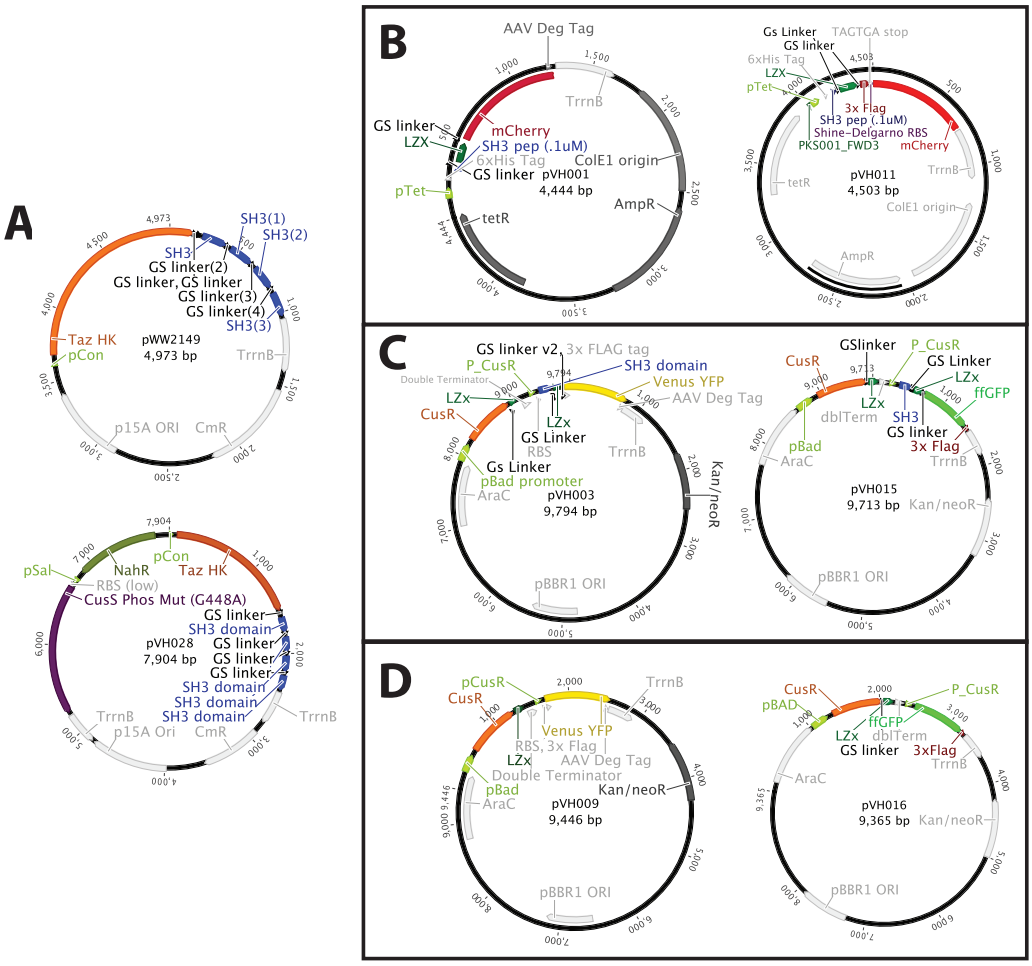


Figure BT-S12: Plasmids used in experimental work. A) Histidine kinase plasmids. pWW2149 has the Taz chimeric histidine kinase with four SH3 binding domains behind a constitutive promoter. pVH028 has the mutated phosphatase, CusS mut G448A, behind an inducible pSal promoter in addition to the Taz HK. Backbone is p15A ori with CmR. B) Scaffold plasmids. pVH001 contains the scaffold-mCherry-AAV fusion behind an inducible pTet promoter. pVH011 contains the bicistronic scaffold/mCherry. Backbone is ColE1 with AmpR. C) Closed loop antiscaffold plasmids. pVH003 contains the response regulator, CusR, behind a pBAD promoter, as well as the antiscaffold-YFP behind the pCusR promoter. pVH015 has a fast-folding GFP reporter(ffGFP) (instead of YFP) with no degradation tag. D) Open loop antiscaffold plasmids. pVH009 is the open loop version of pVH003, in which the sequence for the antiscaffold has been deleted, and only pCusR-YFP remains. pVH016 has ffGFP reporter (instead of YFP) with no degradation tag. Plasmid maps generated with the software Geneious (version 6.1 created by Biomatters).

Total list of chemical equations used in model

Mass Action Reactions

Production and Degradation Reactions (other species are degraded with rate γ and not produced)

$$\phi \underset{\gamma}{\overset{HK_{tot}\gamma}{\rightleftharpoons}} HK \tag{2.1}$$

$$\phi \underset{\gamma}{\overset{RR_{tot}\gamma}{\rightleftharpoons}} RR \tag{2.2}$$

$$\phi \underset{\gamma}{\overset{Sc_{tot}\gamma}{\rightleftharpoons}} Sc \tag{2.3}$$

$$\phi \underset{\gamma}{\overset{Ph_{tot}\gamma}{\rightleftharpoons}} Ph \tag{2.4}$$

Autophosphorylation of histidine kinase

$$HK \underset{k_{rHK_p}}{\overset{k_{fHK_p}}{\rightleftharpoons}} HK_p \tag{2.5}$$

$$Sc:HK \underset{k_{rHK_p}}{\overset{k_{fHK_p}}{\rightleftharpoons}} Sc:HK_p \tag{2.6}$$

$$Sc:HK:RR \underset{k_{rHK_p}}{\overset{k_{fHK_p}}{\rightleftharpoons}} Sc:HK_p:RR \tag{2.7}$$

Non-scaffold mediated background phosphorylation

$$HK_p + RR \underset{k_{rNonCog}}{\overset{k_{fNonCog}}{\rightleftharpoons}} HK + RR_p \tag{2.8}$$

Scaffold binding to HK-SH3 or RR-LZX

$$Sc + HK \underset{k_{rSc:HK}}{\overset{k_{fSc:HK}}{\rightleftharpoons}} Sc:HK \tag{2.9}$$

$$Sc + HK_p \underset{k_{rSc:HK}}{\overset{k_{fSc:HK}}{\rightleftharpoons}} Sc:HK_p \tag{2.10}$$

$$Sc + RR \underset{k_{rSc:RR}}{\overset{k_{fSc:RR}}{\rightleftharpoons}} Sc:RR \tag{2.11}$$

$$Sc + RR_p \underset{k_{rSc:HK}}{\overset{k_{fSc:HK}}{\rightleftharpoons}} Sc:RR_p \tag{2.12}$$

Trimer formation

$$Sc : HK + RR \underset{k_{rSc:RR}}{\overset{k_{fSc:RR}}{\rightleftharpoons}} Sc : HK : RR \tag{2.13}$$

$$Sc : HK_p + RR \underset{k_{rSc:RR}}{\overset{k_{fSc:RR}}{\rightleftharpoons}} Sc : HK_p : RR \tag{2.14}$$

$$Sc : HK + RR_p \underset{k_{rSc:RR}}{\overset{k_{fSc:RR}}{\rightleftharpoons}} Sc : HK : RR_p \tag{2.15}$$

$$Sc : HK_p + RR + p \underset{k_{rSc:RR}}{\overset{k_{fSc:RR}}{\rightleftharpoons}} Sc : HK_p : RR_p \tag{2.16}$$

$$Sc : RR + HK \underset{k_{rSc:HK}}{\overset{k_{fSc:HK}}{\rightleftharpoons}} Sc : HK : RR \tag{2.17}$$

$$Sc : RR + HK_p \underset{k_{rSc:HK}}{\overset{k_{fSc:HK}}{\rightleftharpoons}} Sc : HK_p : RR \tag{2.18}$$

$$Sc : RR_p + HK \underset{k_{rSc:HK}}{\overset{k_{fSc:HK}}{\rightleftharpoons}} Sc : HK : RR_p \tag{2.19}$$

$$Sc : RR_p + HK_p \underset{k_{rSc:HK}}{\overset{k_{fSc:HK}}{\rightleftharpoons}} Sc : HK_p : RR_p \tag{2.20}$$

Phosphorylation of Response Regulator

$$ScHK_pRR \underset{k_{rCogP}}{\overset{k_{fCogP}}{\rightleftharpoons}} Sc : HK : RR_p \tag{2.21}$$

Dephosphorylation of Response Regulator

$$Ph + RR_p \underset{k_{r_{Ph:RR}}}{\overset{k_{f_{Ph:RR}}}{\rightleftharpoons}} Ph : RR_p \xrightarrow{k_{f_{dephos}}} Ph + RR \tag{2.22}$$

$$Ph + Sc : RR_p \underset{k_{r_{Ph:RR}}}{\overset{k_{f_{Ph:RR}}}{\rightleftharpoons}} Ph : Sc : RR_p \xrightarrow{k_{f_{dephos}}} Ph + ScRR \tag{2.23}$$

$$Ph + Sc : HK : RR_p \underset{k_{r_{Ph:RR}}}{\overset{k_{f_{Ph:RR}}}{\rightleftharpoons}} Ph : Sc : HK : RR_p \xrightarrow{k_{f_{dephos}}} Ph + ScHKRR \tag{2.24}$$

$$Ph + Sc : HK_p : RR_p \underset{k_{r_{Ph:RR}}}{\overset{k_{f_{Ph:RR}}}{\rightleftharpoons}} Ph : Sc : HK_p : RR_p \xrightarrow{k_{f_{dephos}}} Ph + Sc : HK_p : RR \tag{2.25}$$

$$Ph + Sc : RR_p : AS \underset{k_{r_{Ph:RR}}}{\overset{k_{f_{Ph:RR}}}{\rightleftharpoons}} Ph : Sc : RR_p : AS \xrightarrow{k_{f_{dephos}}} Ph + Sc : RR : AS \tag{2.26}$$

(2.27)

Sequestration of Scaffold by Antiscaffold

$$Sc + AS \xrightarrow{k_{f_{ScAS}}} ScAS \tag{2.28}$$

$$Sc : HK + AS \underset{k_{rAS_{LZX}}}{\overset{k_{fAS_{LZX}}}{\rightleftharpoons}} ScHKAS \xrightarrow{k_{rAS_{SH3}}} Sc : AS + HK \tag{2.29}$$

$$Sc : HK_p + AS \underset{k_{rAS_{LZX}}}{\overset{k_{fAS_{LZX}}}{\rightleftharpoons}} ScHK_pAS \xrightarrow{k_{rAS_{SH3}}} Sc : AS + HK_p \tag{2.30}$$

$$Sc : RR + AS \underset{k_{rAS_{SH3}}}{\overset{k_{fAS_{SH3}}}{\rightleftharpoons}} ScRRAS \xrightarrow{k_{rAS_{LZX}}} Sc : AS + RR \tag{2.31}$$

$$Sc : RR_p + AS \underset{k_{rAS_{SH3}}}{\overset{k_{fAS_{SH3}}}{\rightleftharpoons}} ScRR_pAS \xrightarrow{k_{rAS_{LZX}}} Sc : AS + RR_p \tag{2.32}$$

Additional Reactions with Phosphatase-bound Complexes

$$Sc + Ph : RR_p \underset{k_{rSc:HK}}{\overset{k_{fSc:HK}}{\rightleftharpoons}} Ph : Sc : RR_p \tag{2.33}$$

$$Sc : HK + Ph : RR_p \underset{k_{rSc:RR}}{\overset{k_{fSc:RR}}{\rightleftharpoons}} Ph : Sc : HK : RR_p \tag{2.34}$$

$$Sc : HK_p + Ph : RR_p \underset{k_{rSc:RR}}{\overset{k_{fSc:RR}}{\rightleftharpoons}} Ph : Sc : HK_p : RR_p \tag{2.35}$$

$$Ph : Sc : RR_p + HK \underset{k_{rSc:HK}}{\overset{k_{fSc:HK}}{\rightleftharpoons}} Ph : Sc : HK : RR_p \tag{2.36}$$

$$Ph : Sc : RR_p + HK_p \underset{k_{rSc:HK}}{\overset{k_{fSc:HK}}{\rightleftharpoons}} Ph : Sc : HK_p : RR_p \tag{2.37}$$

$$Ph : Sc : RR_p + AS \underset{k_{rAS_{SH3}}}{\overset{k_{fAS_{SH3}}}{\rightleftharpoons}} Ph : Sc : RR_p : AS \xrightarrow{k_{rAS_{LZX}}} Sc : AS + Ph : RR_p \tag{2.38}$$

Non-Mass Action - Activation of Antiscaffold Expression

$$\phi \underset{\gamma}{\overset{k_f}{\rightleftharpoons}} As \tag{2.39}$$

$$k_{fAS} = \beta_{AS}\left[\beta_0 + \left(\frac{\mathrm{RR}^2_{\mathrm{active}}}{\mathrm{K}^2_{\mathrm{D}} + \mathrm{RR}^2_{\mathrm{active}}}\right)\right] \tag{2.40}$$

where $\mathrm{RR}_{\mathrm{active}} = \mathrm{RR_p} + \mathrm{Sc} : \mathrm{RR_p} + \mathrm{Sc} : \mathrm{HK} : \mathrm{RR_p} + \mathrm{Sc} : \mathrm{HK_p} : \mathrm{RR_p} + \mathrm{Sc} : \mathrm{RR_p} : \mathrm{AS}$.

List of plasmids and cell strains used

Plasmids		
Name	Resistance	Description
pWW2149	Chloramphenicol	Contains pCon-Taz (histidine kinase). p15A origin of replication, low copy.
pVH028	Chloramphenicol	Contains pCon-Taz (histidine kinase) and pSal-CusSmut (phosphatase). Also has NahR for the pSal promoter. p15A origin of replication, low copy.
pVH001	Carbenicillin/Amp	Contains pTet-Scaffold-mCherry-AAV and TetR genes. ColE1 origin of replication, high copy.
pVH011	Carbenicillin/Amp	Contains bicistronic pTet-Scaffold(3xFLAG)-ShineDelgarno-mCherry and TetR genes. ColE1 origin of replication, high copy.
pVH003	Kanamycin	Contains pBAD-CusR (response regulator) and pCusR-antiscaffold-YFP-AAV. pBBR1 origin of replication, medium copy.
pVH015	Kanamycin	Contains pBAD-CusR (response regulator) and pCusR-antiscaffold-GFP(3xFLAG). pBBR1 origin of replication, medium copy.
pVH009	Kanamycin	Contains pBAD-CusR (response regulator) and pCusR-YFP-AAV. pBBR1 origin of replication, medium copy.
pVH016	Kanamycin	Contains pBAD-CusR (response regulator) and pCusR-GFP(3xFLAG). pBBR1 origin of replication, medium copy.
pVH002	Carbenicillin/Amp	Contains pTet-Scaffold with no fluorescent protein. Used as a control for autofluorescence.
pVH004	Kanamycin	Contains pBAD-CusR and pCusR-antiscaffold with no fluorescent protein. Used as a control for autofluorescence.

Strains (All in WW62 *E.coli*)		
Name	Plasmids Contained	Description
eVH15	pWW2149/pVH011/pVH015	Closed loop Western blot circuit. Cm/Carb/Kan resistance.
eVH16	pWW2149/pVH011/pVH016	Open loop Western blot circuit. Cm/Carb/Kan resistance.
eVH38	pVH028/pVH001/pVH003	Closed loop tracker circuit. Cm/Carb/Kan resistance.
eVH39	pVH028/pVH001/pVH009	Open loop tracker circuit. Cm/Carb/Kan resistance.
eVH05	pww2149/pVH002/pVH004	Autofluorescence control strain. Cm/Carb/Kan
eVH42	pVH028/pVH003	Scaffold-less closed loop control. Cm/Kan
eVH43	pVH028/pVH009	Scaffold-less open loop control. Cm/Kan

References

Acharya, Asha et al. (2002). "A Heterodimerizing Leucine Zipper Coiled Coil System for Examining the Specificity of aPosition Interactions: Amino Acids I, V, L, N, A, and K". In: *Biochemistry* 41.48, pp. 14122–14131.

Andersen, Jens Bo and Soren Molin (1998). "New Unstable Variants of Green Fluorescent Protein for Studies of Transient Gene Expression in Bacteria". In: *Appl Environ Microbiol* 64.6, pp. 2240–2246.

Bashor, C J et al. (2008). "Using Engineered Scaffold Interactions to Reshape MAP Kinase Pathway Signaling Dynamics". In: *Science* 319.5869, pp. 1539–1543.

Becskei, Attila and Luis Serrano (2000). "Engineering stability in gene networks by autoregulation ". In: *Nature* 405, pp. 590–593.

Bessho, Y (2003). "Oscillations, clocks and segmentation". In: *Current Opinion in Genetics & Development* 13.4, pp. 379–384.

Cai, Long, Chiraj K Dalal, and Michael B Elowitz (2008). "Frequency-modulated nuclear localization bursts coordinate gene regulation". In: *Nature* 455.7212, pp. 485–490.

Cardinale, Stefano, Marcin Pawel Joachimiak, and Adam Paul Arkin (2013). "Effects of genetic variation on the E. coli host-circuit interface." In: *Cell Reports* 4.2, pp. 231–237.

Dublanche, Yann et al. (2006). “Noise in transcription negative feedback loops: simulation and experimental analysis”. In: *Mol Syst Biol* 2.

Dueber, John E et al. (2009). “Synthetic protein scaffolds provide modular control over metabolic flux”. In: *Nature Biotechnol* 27.8, pp. 753–759.

Farrell, Christopher M, Alan D Grossman, and Robert T Sauer (2005). “Cytoplasmic degradation of ssrA-tagged proteins”. In: *Mol Microbiol* 57.6, pp. 1750–1761.

Ferrell Jr, James E (2013). “Feedback loops and reciprocal regulation: recurring motifs in the systems biology of the cell cycle”. In: *Curr Opin Cell Biol*, pp. 1–11.

Franco, Elisa, Per-Ola Forsberg, and Richard M. Murray (2008). “Design, modeling and synthesis of an *in vitro* transcription rate regulatory circuit.” In: *Proceedings of the American Control Conference*.

Good, Matthew C, Jesse G Zalatan, and Wendell A Lim (2011). “Scaffold Proteins: Hubs for Controlling the Flow of Cellular Information”. In: *Science* 332, pp. 680–686.

Groban, Eli S et al. (2009). “Kinetic Buffering of Cross Talk between Bacterial Two-Component Sensors”. In: *J Mol Biol* 390.3, pp. 380–393.

Grunberg, R et al. (2010). “Building blocks for protein interaction devices”. In: *Nucleic Acids Research* 38.8, pp. 2645–2662.

Hsiao, Victoria et al. (2015). “Design and implementation of a biomolecular concentration tracker.” In: *ACS Synthetic Biology* 4.2, pp. 150–161. DOI: 10.1021/sb500024b.

Huang, Chi-Ying F and James E Ferrell Jr (1996). “Ultrasensitivity in the mitogen-activated protein kinase cascade”. In: *Proceedings of the National Acadamy of Sciences USA* 93, pp. 10078–10083.

Imayoshi, Itaru et al. (2013). “Oscillatory control of factors determining multipotency and fate in mouse neural progenitors.” In: *Science* 342.6163, pp. 1203–1208.

Klumpp, Stefan, Zhongge Zhang, and Terence Hwa (2009). “Growth Rate-Dependent Global Effects on Gene Expression in Bacteria”. In: *Cell* 139.7, pp. 1366–1375.

Milias-Argeitis, Andreas et al. (2011). “brief communications”. In: *Nature Biotechnol* 29.12, pp. 1114–1116.

Moon, Tae Seok et al. (2010). “Use of modular, synthetic scaffolds for improved production of glucaric acid in engineered E. coli”. In: *Metab Eng* 12.3, pp. 298–305.

Munsky, Brian, Brooke Trinh, and Mustafa Khammash (2009). “Listening to the noise: random fluctuations reveal gene network parameters”. In: *Molecular Systems Biology* 5.

Nagai, Takeharu et al. (2002). "A variant of yellow fluorescent protein with fast and efficient maturation for cell-biological applications." In: *Nature Biotechnology* 20.1, pp. 87–90.

Park, Sang-Hyun, Ali Zarrinpar, and Wendell A Lim (2003). "Rewiring MAP Kinase Pathways Using Alternative Scaffold Assembly Mechanisms". In: *Science* 299.5609, pp. 1061–1064.

Pazy, Yael et al. (2009). "Matching Biochemical Reaction Kinetics to the Timescales of Life: Structural Determinants That Influence the Autodephosphorylation Rate of Response Regulator Proteins". In: *J Mol Biol* 392.5, pp. 1205–1220.

Posern, Guido et al. (1998). "Development of highly selective SH3 binding peptides for Crk and CRKL which disrupt Crk-complexes with DOCK180, SoS and C3G ". In: 16.15.

Rosenfeld, Nitzan, Michael B Elowitz, and Uri Alon (2002). "Negative Autoregulation Speeds the Response Times of Transcription Networks". In: *J Mol Biol* 323.5, pp. 785–793.

Santos, Emmanuel LC de los, Victoria Hsiao, and Richard M Murray (2013). "Design and implementation of a biomolecular circuit for tracking protein concentration". In: *Proceedings of the American Control Conference*, pp. 1–5.

Shaner, Nathan C et al. (2004). "Improved monomeric red, orange and yellow fluorescent proteins derived from Discosoma sp. red fluorescent protein." In: *Nature Biotechnology* 22.12, pp. 1567–1572.

Skerker, Jeffrey M et al. (2005). "Two-Component Signal Transduction Pathways Regulating Growth and Cell Cycle Progression in a Bacterium: A System-Level Analysis". In: *PLoS Biology* 3.10, e334.

Solomaha, Elena et al. (2005). "Kinetics of Src homology 3 domain association with the proline-rich domain of dynamins: specificity, occlusion, and the effects of phosphorylation." In: *The Journal of biological chemistry* 280.24, pp. 23147–23156.

Stapleton, James A et al. (2012). "Feedback Control of Protein Expression in Mammalian Cells by Tunable Synthetic Translational Inhibition". In: *ACS Synth Biol* 1.3, pp. 83–88.

Whitaker, Weston R (2012). "Engineering Modular Post-Translational Control Strategies in Prokaryotes". In: *PhD Thesis, UC Berkeley, USA*, pp. 1–151.

Whitaker, Weston R. and John E. Dueber (2011). "Chapter nineteen - Metabolic Pathway Flux Enhancement by Synthetic Protein Scaffolding". In: *Synthetic Biology, Part A*. Ed. by Chris Voigt. Vol. 497. Methods in Enzymology. Academic Press, pp. 447–468. DOI: http://dx.doi.org/10.1016/B978-0-12-385075-1.00019-6. URL: http://www.sciencedirect.com/science/article/pii/B9780123850751000196.

Whitaker, Weston R et al. (2012). “Engineering Robust Control of two-component system phosphotransfer using modular scaffolds”. In: *PNAS*, pp. 18090–18095.

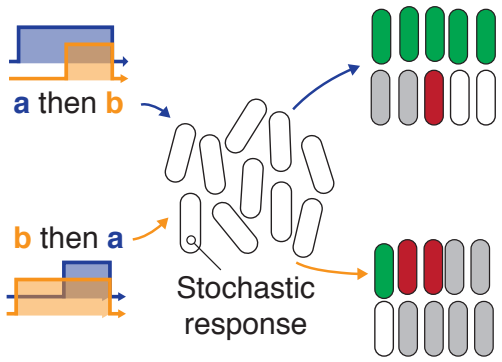
Zhu, Yan and Masayori Inouye (2002). “The role of the G2 box, a conserved motif in the histidine kinase superfamily, in modulating the function of EnvZ”. In: *Mol Microbiol* 45.3, pp. 653–663.

Chapter 3

A POPULATION-BASED TEMPORAL LOGIC GATE FOR TIMING AND RECORDING OF CHEMICAL EVENTS

A version of this chapter has been published as a journal article (Hsiao et al., In press).

3.1 Abstract



Engineered bacterial sensors have potential applications in human health monitoring, environmental chemical detection, and materials biosynthesis. While such bacterial devices have long been engineered to differentiate between combinations of inputs, their potential to process signal timing and duration has been overlooked. In this work, we present a two-input temporal logic gate that can sense and record the order of the inputs, the timing between inputs, and the duration of input pulses. Our temporal logic gate design relies on unidirectional DNA recombination mediated by bacteriophage integrases which detect and encode sequences of input events. For an *E. coli* strain engineered to contain our temporal logic gate, we compare predictions of Markov model simulations with laboratory measurements of final population distributions for both step and pulse inputs. Although single cells were engineered to have digital outputs, stochastic noise created heterogeneous single cell responses that translated into analog population responses. Furthermore, when single cell genetic states were aggregated into population-level distributions, these distributions contained unique information not encoded in individual cells. Thus, final differentiated sub-populations could be used to deduce order, timing, and duration of transient chemical events.

3.2 Introduction

Engineered bacteria could one day be powerful self-replicating biosensors with environmental, health, and industrial applications. Synthetic biology has made important strides in identifying and optimizing genetic components for building such devices. In particular, much work has focused on Boolean logic gates which detect the presence or absence of static chemical signals (Gardner, Cantor, and Collins, 2000; Anderson, Voigt, and Arkin, 2007; Wang et al., 2011; Moon et al., 2013; Shis et al., 2014) and compute a digital response.

Temporal logic gates, which process time-varying chemical signals, have been much less explored. Pioneering work by Friedland *et al.* used serine integrase-based recombination for the counting and detection of sequential pulses of inducers (Friedland et al., 2009). But thus far, no work has studied the potential for temporal logic gates to provide information about the duration of a signal, or the time between two chemical events. Here, we present a temporal logic gate that allows us to infer analog signal timing and duration information about the sequential application of two inducer molecules to a population of bacterial cells.

Similar to previous temporal logic gates, our design takes advantage of the irreversibility of serine integrase recombination. While bistable switches have been successfully deployed as memory modules in genetic circuits (Kotula et al., 2014), such switches require constant protein production to maintain state, and are sensitive to cell division rates and growth phase. The large serine integrases, on the other hand, reliably and irreversibly flip or excise unique fragments of DNA (Yuan, Gupta, and Van Duyne, 2008). Thus logic circuits built from integrases intrinsically include DNA-level memory that requires virtually no cellular resources to maintain state, thus enabling permanent and low-cost genetic differentiation of individual bacterial cells based on transient integrase induction. Further advantages of the serine integrates include the short length (40-50 bp) and directionality of their attachment sites. Serine integrases recognize flanking DNA binding domains (attB, attP) and subsequently digest, flip or excise, and re-ligate the DNA between the attachment sites. Flipping or excision activity is determined by the relative orientation of the sites, which allows complex orientation-dependent behavior to be programmed into integrase circuits. Well-known serine integrases include Bxb1, TP901-1, and ϕC31, all of which have been used to demonstrate static-input logic gates (Siuti, Yazbek, and Lu, 2013; Bonnet, Yin, et al., 2013), and some have cofactors that can reverse directionality (Bonnet, Subsoontorn, and Endy, 2012; Khaleel et al., 2011). Re-

cently, an entirely new set of 11 orthogonal integrases was characterized, greatly expanding the set of circuits that can be built (Yang et al., 2014).

In contrast to previous studies of temporal logic gates, our work leverages the stochastic nature of single-cell switching to create a robust population level response to a time-varying chemical signal. The fundamental nature of living cells that makes them so attractive for engineering – their extremely low energy operation in the limit of using small numbers of molecules to represent information – is also inextricably linked to stochasticity and noise. By traditional engineering standards, synthetic circuits would ideally perform identically in every cell in a population. When this ideal is applied to biology, the stochastic nature of molecular processes, particularly at low copy numbers, presents a significant barrier to reliable outputs from engineered cells. Thus, while natural cellular dynamics and differentiation take advantage of noisy gene expression (Elowitz et al., 2002; Süel et al., 2007) synthetic circuits often require noise reduction for proper function (Dunlop et al., 2008). Recent work has taken a different direction, towards understanding of population-level dynamics. This includes analysis of both stochastic cellular responses to inputs (Uhlendorf et al., 2012; Ruess et al., 2015), and changes in collective population-level memory in response to stress (Mathis and Ackermann, 2016). Such efforts suggest that a deeper understanding of the inherent heterogeneity in biological systems might eventually lead to circuit design which operates on distributions of cellular responses, rather than depending on homogeneous responses from all cells.

It is with this vision in mind that we designed a two-input temporal logic gate using strategically interleaved and oriented integrase (Bxb1, TP901-1) DNA recombination sites and used this gate to engineer an *E. coli* strain with four possible genetically-differentiated end states. This strain contains single genomic copies of the temporal logic gate, ensuring digital-yet-stochastic responses from individual cells. We then utilized the heterogeneity of individual cellular responses to encode sequences of chemical inputs into the overall population response, and used a stochastic model of single cell trajectories to predict the population response. By analyzing the distributions of final cell states, we can deduce the timing and pulse duration of transient chemical pulses, and show that cumulative population level distributions contain additional event information not encoded in any single cell. Furthermore, because the states are genetically encoded, we can recover details of a chemical event long after its occurrence.

3.3 Results

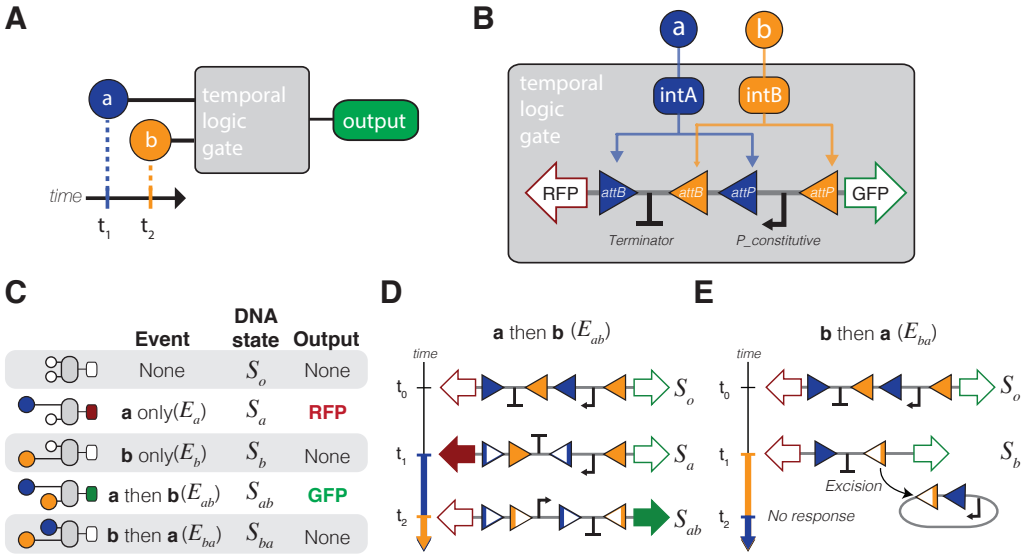


Figure 3.1: Design overview of a temporal logic gate. A) A temporal logic gate distinguishes between two chemical inputs (**a**,**b**) with different start times. B) Implementation of the temporal logic gate using a set of two integrases with overlapping attachment sites. Chemical inputs **a** and **b** activate production of integrases intA and intB, which act upon a chromosomal DNA cassette. C) Table with all possible inputs and outcomes to the event detector. D) Sequence of DNA flipping following inputs with inducer **a** before inducer **b** (event E_{ab}). E) Sequence of DNA flipping following inducer inputs with **b** first (event E_{ba}). In any events in which **b** precedes **a**, the uni-directionality of the intB attachment sites results in excision.

Design of a two-integrase temporal logic gate

We have designed a two-input temporal logic gate that differentiates between the start times of two chemical inputs and produces unique outputs accordingly (Figure 3.1A). The design relies on a system of two-integrases with nested integrase attachments sites (Figure 3.1B). The use of integrases irreversibly invert segments of DNA, resulting in a memory feature that can be maintained for multiple generations (Bonnet, Subsoontorn, and Endy, 2012).

The design of the integrase temporal logic gate hinges on interleaving the attB attachment site of integrase B (intB) with the attP site of integrase A (intA), thus ensuring that the possible DNA flipping outcomes are mutually exclusive (Figure 3.1B). The serine integrases used in this design are TP901-1 (intA) and Bxb1 (intB). The fluorescent proteins mKate2-RFP (RFP) and superfolder-GFP (GFP) are used as placeholders for future downstream gene activation as well as real-time readouts

of the logic gate. The design also features a terminator (Bba-B0015) and a strong constitutive promoter (P7). In the case where there are no inputs, the terminator prevents expression of RFP from the constitutive promoter.

There are five possible basic events that could occur in a two-input system (Figure 3.1C): no input, inducer **a** only (E_a), inducer **b** only (E_b), inducer **a** followed by **b** at a later time (E_{ab}), and inducer **b** followed by **a** at a later time (E_{ba}). Consequently, in a perfectly resolved temporal logic gate there should be five unique DNA states corresponding to the five types of events: S_o (the initial state), S_a, S_b, S_{ab}, and S_{ba}. This design is limited to only four DNA states due to excision when E_b occurs ($S_b = S_{ba}$). The two fluorescent outputs correspond to the two states that occur when inducer **a** is detected first – RFP is produced when the cell is in state S_a, and GFP is produced when the cell is in state S_{ab}.

Figure 3.1D illustrates the sequence of recombination that occurs during an event E_{ab} that results in DNA state S_{ab} and the production of GFP. Upon addition of inducer **a** at time t_1, TP901-1 flips the DNA between its attachment sites, reversing the directionality of the terminator and the Bxb1 attB recognition site (state S_a). Then, when inducer **b** is added at some time t_2 that is greater than t_1, the directionality of the Bxb1 sites is such that the DNA is flipped to reverse the directionality of the P7 constitutive promoter (state S_{ab}). If inducer **b** is added first (Figure 3.1E), the Bxb1 attachment sites are uni-directional, a configuration that results not in recombination, but in excision of the DNA between the sites (state S_b).

Once DNA recombination has occurred, it is irreversible. The unique attB and attP attachment sites are recombined into attL and attR sites, respectively, to which the integrases cannot bind without additional excisionase cofactors (Ghosh, Pannunzio, and Hatfull, 2005). The nesting of the integrase attachment sites is the key design feature that produces the temporal ***a** then **b*** logic, and the irreversibility of the recombination records the event in DNA memory. The result is a genetic record that can both be sequenced later and immediately read via constitutive production of fluorescent outputs.

A Markov model for integrase recombination

The most compelling advantage of engineered biological systems over manmade sensors lies in their inherent capabilities for replication and parallel sensing with minimal energy and resource requirements. Thus, deployment of synthetic bacterial devices would almost certainly involve populations of cells, never just a single cell.

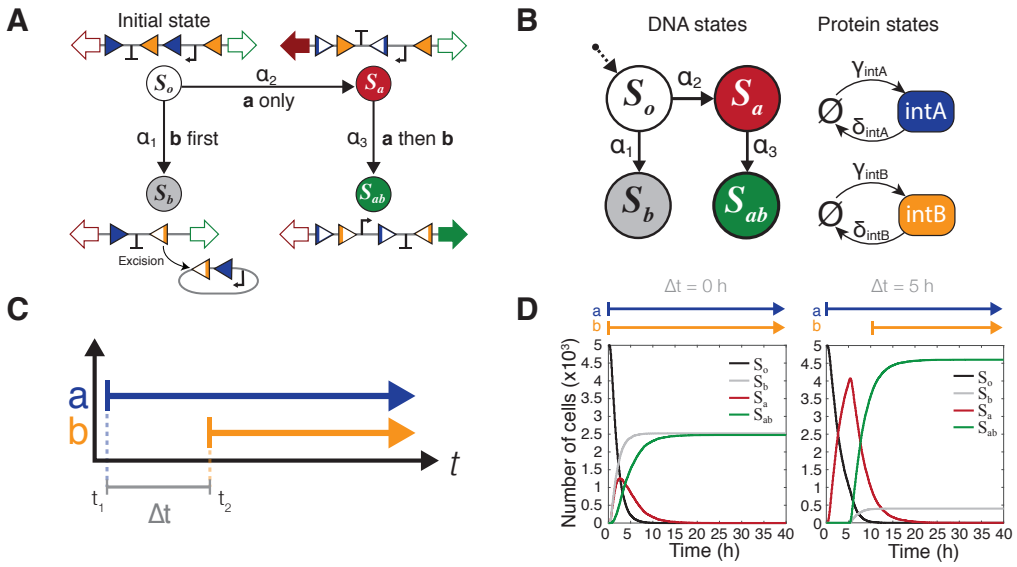


Figure 3.2: A Markov model of integrase-mediated DNA flipping. A) The four possible DNA states, illustrated with DNA state diagrams. All DNA begins in the initial state S_o and there are no reverse processes. The propensity functions α_1,α_2, and α_3 are dependent on the concentration of the two integrases and correspond to the events ***b*** *first* (E_b), ***a*** *only* (E_a), and ***a*** *then* ***b*** (E_{ab}), respectively. B) Representation of the same model as a Markov chain. Integrases are represented simply as protein states with production (γ_A, γ_B) and degradation (δ_A, δ_B) rates. C) Graphical representation of inducer step functions. Δt is defined as difference between the start time of the first inducer and start time of the second. D) Simulation results for inducer separation times of 0 and 5 hours. There are four possible DNA states, but all cells end up in either the S_b or S_{ab} final states. Individual trajectories are simulated for 5000 cells and the number of cells in each DNA state are summed for each time point (Appendix Fig. ED-S2).

It is therefore important to understand how stochastic single cell responses affect overall population-level distributions and outcomes. We created a Markov model of integrase-mediated DNA flipping and then used a stochastic simulation algorithm (Gillespie, 1977) to simulate individual cell trajectories (Figure 3.2A). All of the four possible DNA states are represented in the model: the original state (S_o), the intB excision state (S_b), the intA single flip state (S_a), and the ***a*** *then* ***b*** double flip state (S_{ab}). We have implemented the system experimentally by chromosomally integrating the target DNA into the genome of the *E. coli* cell. This allows us to assume that each cell only has one copy of the temporal logic gate (Haldimann and Wanner, 2001), and that each cell can be characterized by the tuple (DNA, IntA, IntB) (Figure 3.2B). The DNA term is S_o, S_a, S_b, or S_{ab}, and IntA and IntB are non-negative

integers representing the molecular copy number of each integrase. Once a DNA cassette has flipped into any of the states other than the original state S_o, there is no reverse process. The logic gate is designed such that if integrase B is expressed prior to integrase A, the DNA cassette is excised and the chain reaches the dead-end S_b state. In order for a cell to successfully detect E_{ab}, it first needs to switch into state S_a then transition into state S_{ab} upon addition of inducer **b**.

Since each cell contains only a single copy of the temporal logic gate DNA, we can expect each cell to behave differently, and to be highly susceptible to internal and external noise. This stochastic behavior will create a heterogeneous population response that can be analyzed for a more complex profile of event than if all the cells behaved uniformly. In order to capture the heterogeneity of cell population, we model the temporal logic gate using a stochastic model. Specifically, the stochastic transitions between the DNA states and the production/degradation of integrases are mathematically modeled by a continuous-time Markov chain over the state space $(\text{DNA}, \text{IntA}, \text{IntB})$ as illustrated in Figure 3.2B. Definitions of transition rates can be found in Appendix Table S1.

In silico, the dynamics of a single cell translates to each stochastic simulation of the Markov model starting with $(\text{DNA} = S_o, \text{IntA} = 0, \text{IntB} = 0)$ state. We define $\mathbb{P}_t(S_o), \mathbb{P}_t(S_a), \mathbb{P}_t(S_b)$ and $\mathbb{P}_t(S_{ab})$ as the probability that the DNA state of a single cell is S_o, S_a, S_b and S_{ab} at time t, respectively.

The temporal dynamics of the probability can be modeled by the following ordinary differential equation (ODE)

$$\frac{d}{dt}\begin{bmatrix}\mathbb{P}_t(S_o)\\ \mathbb{P}_t(S_a)\\ \mathbb{P}_t(S_b)\\ \mathbb{P}_t(S_{ab})\end{bmatrix} = \begin{bmatrix}-\mathbb{E}_t[\alpha_1(\text{IntB})|S_o] - \mathbb{E}_t[\alpha_2(\text{IntA})|S_o] & 0 & 0 & 0\\ \mathbb{E}_t[\alpha_2(\text{IntA})|S_o] & -\mathbb{E}_t[\alpha_3(\text{IntB})|S_a] & 0 & 0\\ \mathbb{E}_t[\alpha_1(\text{IntB})|S_o] & 0 & 0 & 0\\ 0 & \mathbb{E}_t[\alpha_3(\text{IntB})|S_a] & 0 & 0\end{bmatrix}\begin{bmatrix}\mathbb{P}_t(S_o)\\ \mathbb{P}_t(S_a)\\ \mathbb{P}_t(S_b)\\ \mathbb{P}_t(S_{ab})\end{bmatrix}, \tag{3.1}$$

where the notation $\mathbb{E}_t[\cdot|\cdot]$ stands for the conditional expected value at time t (Full derivation, Appendix Section 12.1).

Serine integrases are produced as monomers that form dimers, search for specific attB and attP sequences, and, once both attB and attP sites are occupied, form a tetramer (dimer of dimers) that digests, flips, and re-ligates the DNA (Yuan, Gupta, and Van Duyne, 2008; Rutherford et al., 2013). Though some cooperativity in $\Phi C31$ binding to attB has been found (McEwan, Rowley, and Smith, 2009), cooperativity in Bxb1 or TP901-1 integrase binding to attB and attP not been observed (Ghosh, Pannunzio, and Hatfull, 2005; Singh, Rockenbach, et al., 2014).

Rather than account for all individual DNA-integrase interactions, we have created a minimal model of stochastic transitions where only the final DNA states (S_o, S_a, S_b, S_{ab}) and the number of integrase monomer molecules (intA, intB) are tracked and all integrase activity is encompassed in the $k_{\text{flip}*}$ term. Since no cooperativity has been observed in Bxb1 or TP901-1 DNA binding (i.e. occupation of attB does not increase the probability of attP binding), we represent the required tetramerization as a fraction where flipping efficiency is zero unless at least four molecules are present. Thus, the propensity functions for state transitions as a function of integrase concentration, $\alpha_i(\text{Int}_*)$, are defined as

$$\alpha_i(\text{Int}_*) := k_{\text{flip}*}\left(\frac{\text{Int}_*(\text{Int}_*-1)(\text{Int}_*-2)(\text{Int}_*-3)}{K_{\text{d}*}^4+K_{\text{d}*}^3\,\text{Int}_*+K_{\text{d}*}^2\,\text{Int}_*(\text{Int}_*-1)+K_{\text{d}*}\,\text{Int}_*(\text{Int}_*-1)(\text{Int}_*-2)+\text{Int}_*(\text{Int}_*-1)(\text{Int}_*-2)(\text{Int}_*-3)}\right), \quad (3.2)$$

where Int_* is integrase concentration, $K_{\text{d}*}$ is the dissociation constant, $k_{\text{flip}*}$ is the rate of flipping if the tetramer is formed, $i = 1, 2, 3$, and $* = A, B$ (See Appendix ED-S1 for visualization of $\alpha_i(\text{I}nt_*)$, Appendix Section 12.2 for full derivation).

We also define the time between the introduction of the first inducer (t_1) and the arrival of the second inducer (t_2) as the *inducer separation time* (Δt), such that

$$\Delta t = t_2 - t_1, \quad (3.3)$$

as shown in Figure 3.2C.

In the following set of simulations and experiments, we will consider cases with step inputs (Figure 3.2C), where the inducers are either present or not present. Concentrations of the inducers when they are "on" will be held constant. Also, it is important to note that inducer **a** is still present during and after time Δt when inducer **b** is introduced.

Simulations of the Markov model were done with biologically plausible parameters in order to predict qualitative circuit behavior (Appendix Table S1). We limited the parameters to only the basic processes (integrase production, degradation, and DNA flipping), and parameter values were chosen to be within biological orders of magnitude. The single production rate constants, $k_{\text{prod}A}$ and $k_{\text{prod}B}$, combine the transcription and translation rates of each integrase. When an integrase in the model is induced, its production rate, γ_*, is the sum of $k_{\text{prod}*}$ and any leaky transcriptional expression, $k_{\text{leak}*}$ ($*$ = intA or intB). The integrase monomer disassociation constant, $K_{\text{d}*}$, was estimated from measured Bxb1 binding constants (Singh, Ghosh, and

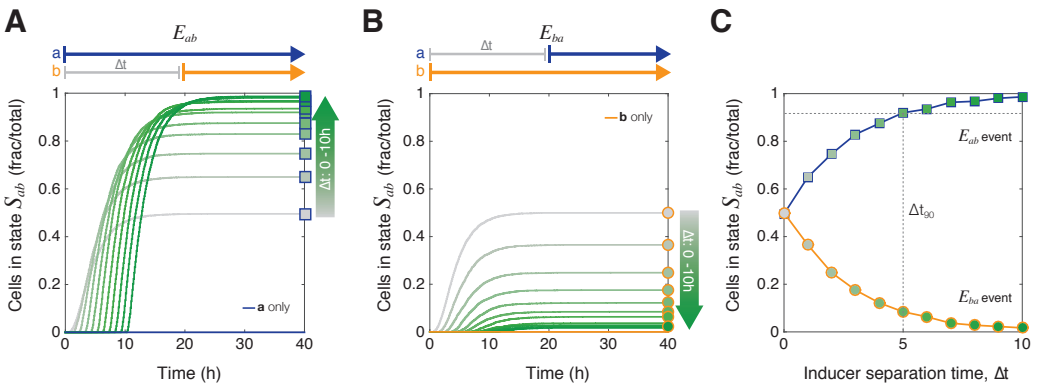
Hatfull, 2013). Parameter values for preliminary simulations were $k_{\mathrm{prod}A} = k_{\mathrm{prod}B} = 50(\mu\mathrm{m}^3 \cdot \mathrm{hr})^{-1}$, $k_{\mathrm{deg}} = 0.3\mathrm{hr}^{-1}$ (2.3 hr half-life), $k_{\mathrm{flip}A} = k_{\mathrm{flip}B} = 0.4\mathrm{hr}^{-1}$, $k_{\mathrm{leak}A} = k_{\mathrm{leak}B} = 0(\mu\mathrm{m}^3 \cdot \mathrm{hr})^{-1}$, and $K_{\mathrm{d}A} = K_{\mathrm{d}B} = 10$ molecules.

Our analysis of initial numerical simulation results highlight the significant role that the inducer separation time, Δt, plays in setting the final population distributions (Figure 3.2D). For each Δt, individual cell trajectories were generated with the assumption that each cell only has one copy of the target DNA (N = 5000 trajectories). Then, at every time point, the total number of cells in each DNA state is counted (Appendix ED-S2). Figure 3.2D shows the contrast between adding both inducers simultaneously (Δt = 0h) and adding inducer **b** after a 5 hour delay (Δt = 5h). Since both inducers are present by the end of simulation, all of the cells must have a final state that is either the S_{ab} state or the S_b state. No cells remain in the original S_o configuration. S_a is a transient state that builds up prior to the addition of inducer **b** and begins to convert to S_{ab} immediately after the introduction of **b**. These initial simulation results suggest that Δt may be a way to reliably tune the final population fractions of S_{ab} versus S_b state cells.

Population distributions reflect inducer order and separation time

We used the model to further investigate the effects of varying both inducer order and separation time on population distributions in our experimental system, and to understand the possible outcomes. In Figure 3.3, we simulate *in silico* cell populations that have been exposed to a sequence of overlapping step functions (N = 5000 trajectories).

In the case of an E_{ab} event, the proportion of cells that successfully detect ***a** then **b*** and switch to state S_{ab} is a function of the inducer separation time, Δt (Figure 3.3A). High Δt means increasing the time that cells spend in only inducer **a**, allowing for most of the population to transition from $S_o \rightarrow S_a$ before the addition of any inducer **b**. Exposing cells to the inverse sequence of events, E_{ba}, results in a decrease of S_{ab} cells proportional to increasing Δt (Figure 3.3B). High Δt in an E_{ba} event means that $S_o \rightarrow S_b$ is the dominant reaction and cells that get partitioned into S_b will not respond to **a**. If we plot the final number of S_{ab} cells from both E_{ab} and E_{ba} as a function of Δt (Figure 3.3C), we see that the two curves do not overlap. S_{ab} fractions exposed to E_{ab} increase monotonically with Δt, while those exposed to E_{ba} decrease monotonically with Δt. Thus measuring the fraction of S_{ab} cells is sufficient to determine both the order of events and the timing, Δt, between them.



D

Event	Single-cell level: DNA state	Single-cell level: Color	Population level: Population fraction (%) S_o	S_a	S_b	S_{ab}
None	S_o	None	100			
a only	S_a	RFP		100		
b only	S_b	None			100	
a then b	S_{ab}	GFP				>50
b then a	S_b	None				<50

Figure 3.3: Simulation results for inducer separation time for $\Delta t = 0 - 10$h. A) The population fraction (N/5000 cells) that switches into state S_{ab} following an E_{ab} event is dependent on the inducer separation time, Δt. The gray to dark green color gradient represents increasing Δt values. Square markers indicate final population fractions for specific values of Δt. B) In the case of the inverse E_{ba} event, the fraction of cells in state S_{ab} decreases proportionally to Δt. Circular markers indicate final population fractions for specific values of Δt. C) Final S_{ab} cell fractions from Figure 3.3A, B are plotted as a function of Δt. Blue line with square markers are endpoint population fractions from an E_{ab} event. Yellow line with circular markers are final endpoint population fractions from an E_{ba} event. The gradient inside the markers corresponds to increasing Δt value. The dotted gray line corresponds to the Δt_{90}, the value of Δt at which $\geq 90\%$ of the cells are in state S_{ab}. All simulations were done with a population of N = 5000 cells. D) Chart showing differences in information that can be recorded at the single-cell versus the population level. In particular, E_{ba} does not have a unique single-cell genetic state, but has a clear distinct population-level phenotype.

Additionally, we can define a detection limit, Δt_{90}, for which the inducer separation time results in $\geq 90\%$ of population switching into the S_{ab} state (Figure 3.3C). This Δt_{90} limit provides a way to capture the two response regimes of the population. If the inducer separation time is less than the detection limit ($\Delta t < \Delta t_{90}$), then the rate of population switching is fast enough such that the number of S_{ab} cells will correspond uniquely to some Δt value. If $\Delta t > \Delta t_{90}$, then most cells have already

switched to a final state, and the differences in S_{ab} cell count are too small to uniquely determine Δt.

The single-cell limitations of the temporal logic gate circuit can be overcome by measuring the number of S_{ab} cells as a fraction of total cells. Though the logic gate itself does not have a unique genetic S_{ba} state and cannot distinguish between a ***b*** *only* event versus a ***b*** *then* ***a*** event, these simulation results suggest that population-level fractional phenotypes can provide this additional information (Figure 3.3D). In the case of E_{ab}, fractions of S_{ab} will always be above 50%, while S_{ab} fractions less than 50% indicate E_{ba}. Additional figures showing how populations of S_a, S_b, and S_o cells change with Δt can be found in Appendix ED-S3.

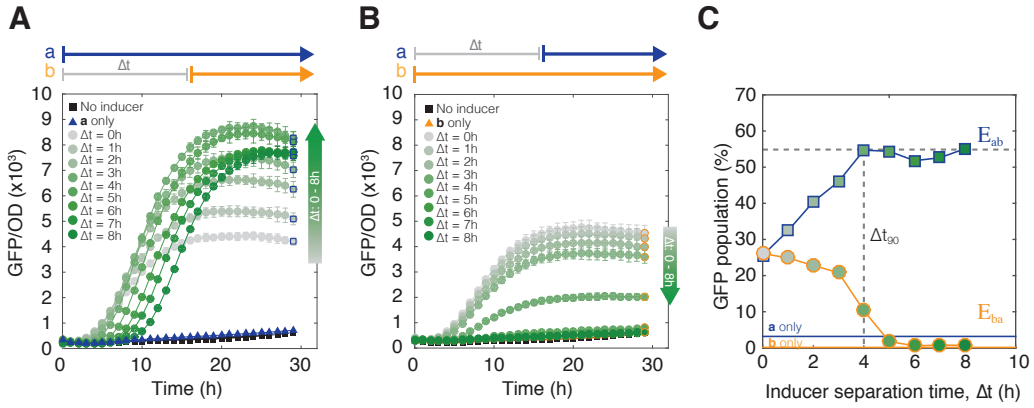


Figure 3.4: *In vivo* results for varying inducer separation time from $\Delta t = 0 - 8$h. A) Populations of cells exposed to an E_{ab} event sequence. Cell-switching to state S_{ab} (indicated by GFP fluorescence) begins when inducer **b** (aTc) is added. Maximum normalized GFP fluorescence increases as a function of the inducer separation time Δt. Gray to dark green gradient represents increasing Δt values. Square markers are final endpoint measurements. Error bars represent standard error of the mean. B) Cells exposed to the inverse E_{ba} sequence of events. GFP fluorescence is inversely proportional to the inducer separation time between **b** and **a**. Circular markers are final endpoint measurements. C) Final population distributions from Figure 3.4A, B at 30 hours are plotted as a function of Δt. Cells were gated by GFP fluorescence to identify percentage of S_{ab} cells. Dotted line marks Δt_{90} detection limit.

In vivo step induction data supported model predictions and showed that population fractions of S_{ab} cells could be tuned using Δt (Figure 3.4). DH5α-Z1 cells were chromosomally integrated with one copy of the integrase target DNA and then transformed with a high copy plasmid containing Ptet-Bxb1 and PBAD-TP901-1. When Δt was varied from 0 – 8 hours, we observed results qualitatively similar to model predictions. In Figure 3.4A, the cells have been exposed to an E_{ab}

event, where inducer **a** is present from time $t = 0$ h to t_{end}, and **b** is present from $t = \Delta t$ h to t_{end}. GFP expression during time course measurements is used as a proxy for S_{ab} state cells and flow cytometry was used to measure final populations. Comparisons of bulk fluorescence versus cytometry cell counts suggest that in single copy integrants, overall GFP is a good approximation of population S_{ab} levels (Appendix Fig. ED-S12).

In Figure 3.4A, the number of cells in the GFP-expressing S_{ab} state increases proportionally with increasing Δt, and continue to be responsive even when the two inducers are separated by 8 hours. There is some expression of GFP in the presence of only inducer **a** (E_a), indicating some basal levels of intB. RFP expression, a proxy for the number of cells in state S_a, begins to increase at $t = 0$ h and drops at time $t = \Delta t$ when inducer **b** is added (Appendix Fig. ED-S4A). Aligning all of the GFP expression curves by Δt (Appendix Fig. ED-S5) shows that lower values of Δt not only have lower final GFP expression values, but also have slower rates of GFP production. This is consistent with modeling results because if we assume inducer **b** has an equal probability of entering any one cell, then in case of small Δt ($\Delta t \leq 4$ hours) there is a much larger number of S_o cells and so the rate of $S_a \rightarrow S_{ab}$ state conversion will be lower. In the case of $\Delta t > 4$ hours, the majority of cells in the population are already in the S_a state configuration, and so the rate of cell state conversion to S_{ab} will be much higher. When cells are exposed to E_{ba}, the number of S_{ab} cells decreases proportionally to Δt (Figure 3.4B), and there is no RFP expression above background (Appendix Fig. ED-S4B). In both types of events, the cells maintained their state for up to 30 hours in liquid culture and when re-streaked as single colonies. (Additional data with a more distinct color scheme and OD curves for this set of experiments can be found in Appendix Figures S6, S7. Single colony analysis in Appendix Fig. ED-S11.)

Final S_{ab} (GFP) population fractions are sufficient to differentiate between populations that have been exposed to E_{ab} versus E_{ba} within one hour of separation time between inducers (Figure 3.4C). Final populations after 30 hours of growth were measured via flow cytometry and plotted against Δt. As Δt increases, so does the S_{ab} sub-population. The cells that encountered E_{ba} have lower S_{ab} fractions with high Δt, and at $\Delta t = 6$ h the final S_{ab} sub-population is equal to the baseline expression of a ***b*** *only* population, indicating that the addition of inducer **a** after a 6 hour exposure to only inducer **b** has no effect at all. Based on where the GFP fraction exceeds 90% of the maximum S_{ab} population fraction, the Δt_{90} detection limit for

the experimental system is ~ 4 hours. These experimental results show that the S_{ab} population fraction clearly diverges for E_{ab} and E_{ba} when $\Delta t \neq 0$h, indicating that S_{ab} fractions alone can be used to determine both event order and separation time.

Further analysis of population-level data for all of the measurable fluorescent cell states can provide additional insights into differences in partition growth rates and leaky integrase expression (Figure EV3.5, Appendix Fig. ED-S8–ED-S10). In Figure EV1, experimental populations from the step input experiments have been gated into quadrants such that S_{ab}, S_a, and $S_o + S_b$ populations can be counted. Even with maximum induction at highest Δt, the maximum population fraction that can be switched appears to be approximately 60% of the total population. We believe this is due to the non-fluorescent cells (S_o,S_b) having a slight growth advantage over differentiated cells. Studies have shown that unnecessary protein production has inverse effects on cell growth (Tan, Marguet, and You, 2009; Scott et al., 2010), and even with single copy integrants, this would result in some overrepresentation of non-fluorescent cells in the population. Single colony analysis of the final populations shows that S_o cells persist in the population even with 30 – 40 hours of inducer exposure (Appendix Fig. ED-S11E).

Leaky expression of intA and intB can also be inferred from the *no inducer*, ***a*** *only*, and ***b*** *only* populations (Figure EV1A, EV1B) and we can conclude that leaky expression is quite low, not exceeding ~ 0.5 – 3 %. Even accounting for the overrepresentation of non-fluorescent cells, the baseline population split when both **a** and **b** are added simultaneously (Δt = 0 h) is just under 50% of the total GFP population fraction. This suggests that the integrase flipping rates, $k_{\text{flip}A}$ and $k_{\text{flip}B}$, may not be equal and that the basal expression rates, $k_{\text{leak}A,B}$ should be non-zero.

Varying model parameters for integrase activity and basal expression

Prior to proceeding with additional model-driven experimental designs, model parameters were modified to better represent asymmetrical integrase activity. The parameters for integrase flipping and leaky basal expression were tuned to account for the asymmetrical population responses to E_{ab} versus E_{ba} events (Figure 3.4C). We hypothesized that this asymmetry arises from a combination of unequal integrase activity when searching for and flipping the DNA, as well as leaky background expression of the integrases (Figure 3.6).

To understand overall trends in model behavior, we varied $k_{\text{flip}A}$ and $k_{\text{leak}B}$ while holding the other parameters constant. When the relative flipping efficiency of intA

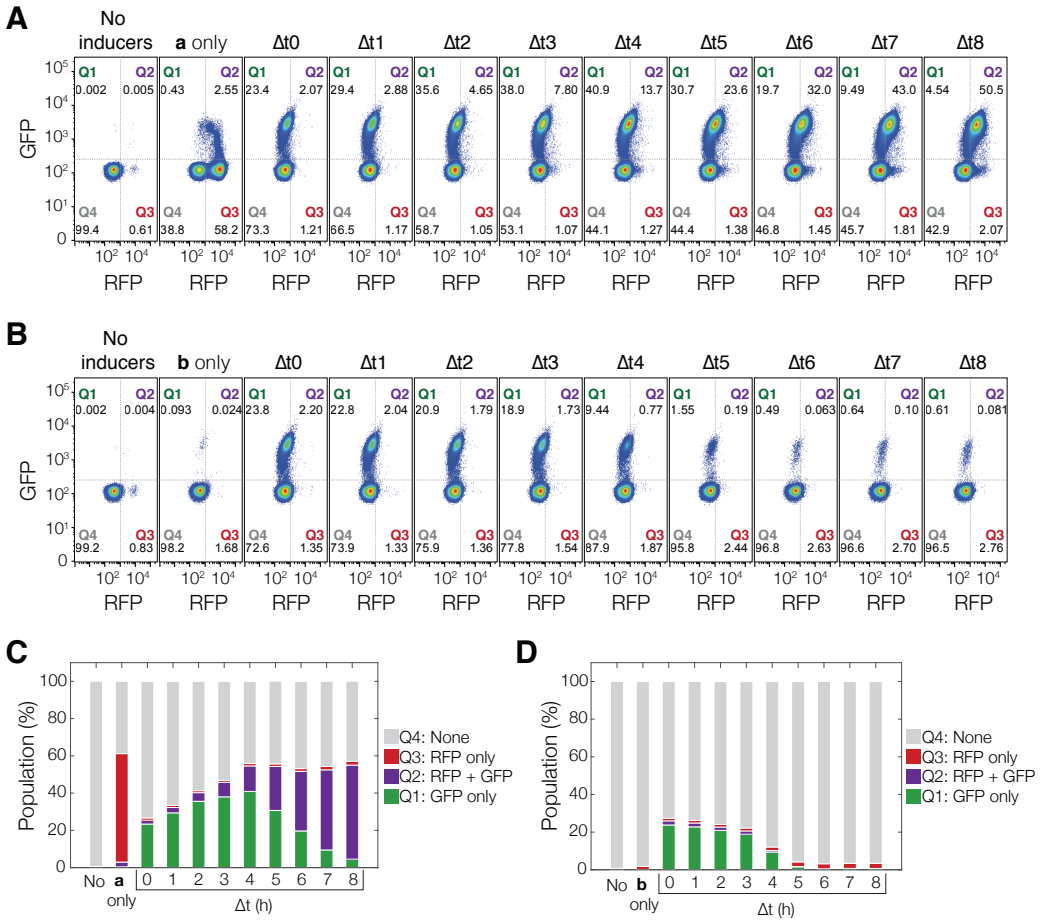


Figure 3.5: Flow cytometry data for varying inducer separation time from Δt = 0 – 8h ($\sim 10^6$ cells per population). The populations are gated by fluorescence into quadrants Q1 (GFP only), Q2 (GFP and RFP), Q3 (RFP only), and Q4 (non-fluorescent). There is a transitory phase (Q2) in which cells contain both GFP and RFP. This is due to slow dilution of RFP through cell division even after cells have switched to S_{ab} and begun production of GFP. Cells in Q1 + Q2 are used for the final GFP population fractions in Figure 3.4C. A) E_{ab} cell populations plotted by their RFP and GFP expression with increasing Δt. Leaky expression of PBAD-intA can be estimated by looking at Q3 of the *No inducers,* ***b*** *only* populations (~0.5 – 2%). Leaky expression of Ptet-intB can be estimated with Q1 + Q2 fractions of the ***a*** *only* population (~ 2-3%). B) E_{ba} populations with increasing Δt. C) Population fractions by quadrant for E_{ab}. D) Population fractions by quadrant for E_{ba}. Individual flow cytometry histograms can be found in Appendix Fig. ED-S8–ED-S10.

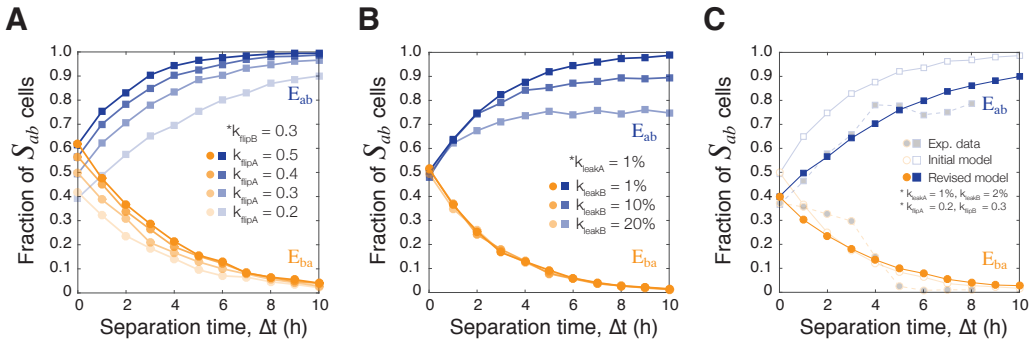


Figure 3.6: Varying model parameters for integrase flipping and leaky expression. A) As DNA flipping rates of intA ($k_{\text{flip}A}$) are decreased relative to $k_{\text{flip}B}$, the population of S_{ab} cells at Δt = 0 h has a downward shift. Simulations are done with N = 3000 trajectories/marker. B) Increasing the leaky expression of intB ($k_{\text{leak}B}$) changes the maximum threshold of cells that correctly identify S_{ab} even at high Δt. Leakiness is defined as a percentage of the induced integrase production rate (k_{prod*}). C) The model was revised to more closely match the experimental data by constraining parameters for leaky expression and varying integrase flipping (N = 5000). Mean squared error was calculated between the experimental data and the initial and revised models to find an optimized pair of $k_{\text{flip}A,B}$ values (Appendix Fig. ED-S13). The final parameters were set to be $k_{\text{flip}A} = 0.2\text{hr}^{-1}$, $k_{\text{flip}B} = 0.3\text{hr}^{-1}$, $k_{\text{leak}A} = 0.01 \cdot k_{\text{prod}A}(\mu m^3 \cdot \text{hr})^{-1}$, and $k_{\text{leak}B} = 0.02 \cdot k_{\text{prod}B}(\mu m^3 \cdot \text{hr})^{-1}$ (Appendix Table S2).

($k_{\text{flip}A}$) was varied from 0.2 to 0.5 hr^{-1}($k_{\text{flip}B}$ = 0.3 hr^{-1}), we observed a bias in the baseline population split when both inducers are introduced simultaneously, Δt = 0 h (Figure 3.6A, N = 3000). Previously in the preliminary model (Figure 3.3C), the two integrases were assigned equal flipping rates, and the population split was expected to be 50/50 for S_a/S_b. As the flipping rate of intA decreases relative to that of intB, that baseline shifts downwards to favor the more active integrase, intB. Varying the basal expression of intB ($k_{\text{leak}B}$) from 1% to 20% of the intB production rate ($k_{\text{prod}B}$) monotonically decreases the maximum S_{ab} population fraction that can be reached in an E_{ab} event (Figure 3.6B, N = 3000). If there is a constant level of un-induced intB, then there will always be a minimum population of S_b cells inhibiting the maximum fraction of S_{ab} cells.

These simulation results showed that by varying $k_{\text{flip}A}$ and $k_{\text{leak}B}$, we could tune the baseline shift at Δt = 0 and the maximum S_{ab} ceiling at high Δt to better approximate our experimental system. However, experimental measurements of leaky integrase expression showed that leaky expression was actually quite low (1% for intA, 1-3%

for intB)(Figure 3.4C, Figure EV1, ***b** only*, ***a** only*). Given actual measurements for $k_{\text{leak}A,B}$, we constrained those parameters and fit the model by varying $k_{\text{flip}A,B}$.

In order to find the best pair of values for $k_{\text{flip}A}$ and $k_{\text{flip}B}$, the flipping efficiency parameters for both integrases were varied from 0.1 to 0.6 $\text{h}r^{-1}$ *in silico* (N = 500 cell trajectories), creating a matrix of simulated S_{ab} population fractions for each combination (Appendix Fig. ED-S13). Leaky basal expression of the integrases was held constant based on experimentally measured values (k_{leak_a} = 1% of $k_{\text{prod}A}$, k_{leak_b} = 2% of $k_{\text{prod}B}$), and experimental data was normalized to a 70% population maximum for fitting purposes. Mean squared error was found by comparing model fits with experimental data (Appendix Fig. ED-S13A), and the combination with the minimum MSE was chosen (Appendix Fig. ED-S13B).

Figure 3.6C shows Δt versus S_{ab} simulation results for final revised parameters. The final parameters were set to be $k_{\text{flip}A} = 0.2\text{h}r^{-1}$, $k_{\text{flip}B} = 0.3\text{h}r^{-1}$, $k_{\text{leak}A} = 0.01 \cdot k_{\text{prod}A}(\mu m^3 \cdot \text{h}r)^{-1}$, and $k_{\text{leak}B} = 0.02 \cdot k_{\text{prod}B}(\mu m^3 \cdot \text{h}r)^{-1}$ (Appendix Table S2). The introduction of leaky integrase expression into the model suggests that the high leaky expression of intB, around 2% of the population will "detect" E_{ba} and be in state S_{ba} even when no inducer **a** has been introduced. Additionally, preliminary simulation results suggest that the Δt_{90} detection limit can be tuned by increasing or decreasing the overall production rate $k_{\text{p}rod*}$ (∗ = A or B) (Appendix Fig. ED-S14), though this remains to be experimentally verified in future work.

In silico parameter space exploration shows that varying $k_{\text{flip}*}$ and $k_{\text{leak}*}$ parameters enables tuning of baseline Δt = 0h split for E_{ab} / E_{ba} and the maximum ceiling for S_{ab} population fraction. Fold-change variations in relative rates allowed us to understand overall trends in the final populations, and we adjusted the model to account for inequalities in integrase flipping and leaky basal expression. Since leaky expression was measured to be small, we primarily tuned flipping rates. This process led us to more relevant model-informed predictions of experimental outcomes. With the refined model, we were interested to see whether distributions of the RFP-expressing S_a state could provide information that measuring S_{ab} fractions alone could not.

Deducing pulse width from S_a population fractions

Using the fraction of S_{ab} (GFP) cells alone, we can determine Δt values up to a Δt_{90} limit for any given sequence of two step inputs. Now consider a pulse type of event, in which inducer **a** begins at time t = 0 h, remains constant throughout, and

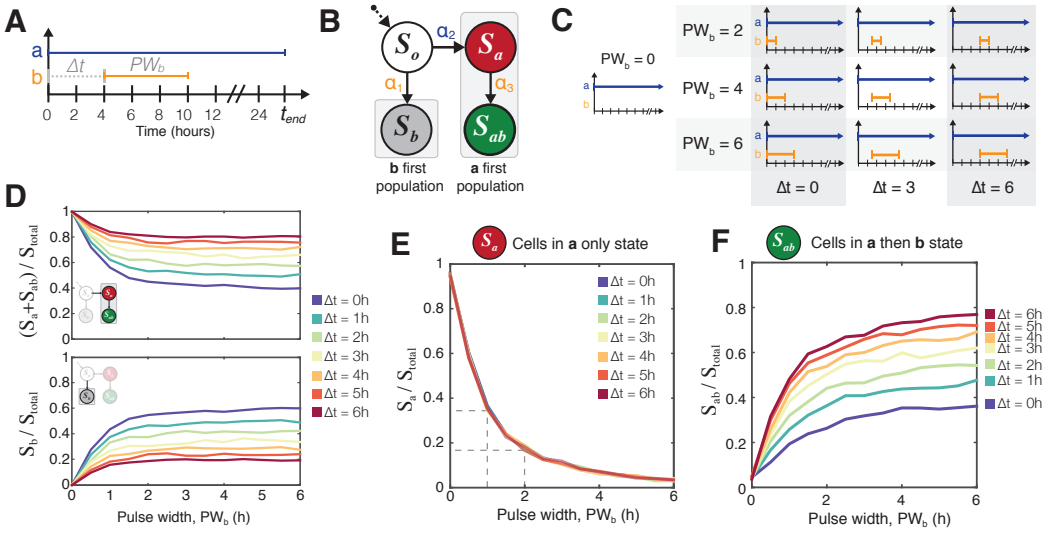


Figure 3.7: Simulation results for pulse width modulation. Simulations were done with revised parameters found in Figure 3.6C. A) Inducer **a** can be used as a reference signal against which to measure the time and duration of the inducer **b** pulse. B) The population eventually divides into one of two partitions: those that see inducer **a** first and those that see inducer **b** first. Only if a cell has entered the ***a** first* pathway does it have the possibility to express RFP or GFP. Furthermore, S_a can be thought of as a necessary precursor to S_{ab}. C) A matrix illustrating a subset of the Δt and PW_b values to be tested. D) Simulation results show that for any given Δt, the number of cells in S_b = total number of cells – (S_a + S_{ab}) E) The fraction of the population in the S_a state is totally independent of Δt and depends only on the pulse duration of inducer **b**. F) Once PW_b is known, then the fraction of the population in S_{ab} state can be used to find the time at which the pulse of inducer **b** began. N = 3000 cell trajectories for each value of Δt, PW_b.

inducer **b** is introduced as a finite pulse at time $t = \Delta t$ h (Figure 3.7A). The start time of inducer **a** then becomes a reference for when the entire system is activated and ready to detect inducer **b**. Cell states are measured via flow cytometry at time t_{end}, where $t_{end} > 24$ hours. Modeling results presented in this section are using the refined set of parameters defined in Figure 3.6C and Appendix Table S2.

If either of the two inducers is present in the media to some limit t_{end}, we would expect all of the S_o cells will end up in one of two populations (Figure 3.7B). Cells that encounter inducer **b** first will be in the S_b state, while cells that encounter **a** first will either be in the S_a or S_{ab} states. In the previous sections, once an inducer was added to the population, it was not removed, and the assumption was made that at times greater than 24 hours, only a negligible number of S_o cells remained. This type of step function induction also meant that only the number of S_{ab} cells (GFP)

was needed to uniquely determine the separation time Δt because *any and all* cells that had switched to S_a would eventually become S_{ab}.

However, in the case of a transient pulse, some cells that are in the S_a state (RFP) will not ever encounter inducer **b**. Assuming the $k_{\text{leak}B}$ is small, these cells will remain in the S_a state. Therefore, the population of ***a** first* cells equals $S_a + S_{ab}$. We simulated a matrix of populations exposed to varying inducer separation times (Δt) and inducer **b** pulse widths (PW_b) to measure the resolution of detectable events (Figure 3.7C). In simulation (Figure 3.7D), we can see that the two populations mirror each other to add up to 100% of the total cells (N = 3000 cells, additional simulations in Appendix Fig. ED-S16).

Given that the step induction of **b** is equivalent as a pulse of infinite length ($PW_b = \infty$) and our prior experimental evidence showing that virtually no cells remain in state S_a when $PW_b = \infty$, we reasoned that the number of S_a cells could be used to deduce information about the pulse width of **b**. This hypothesis was tested *in silico* by running a matrix of simulations with varying Δt and PW_b. In Figure 3.7E, we see that the fraction of S_a cells over the total number of cells decreases monotonically with increasing PW_b, and the curves overlap regardless of Δt. The overlap occurs despite non-zero leakiness of intA and intB. The maximum number of S_a cells does not go to 1 at $PW_b = 0$h because of leaky intB expression ($k_{\text{leak}B} = 0.02 \cdot k_{\text{prod}B}$).

Analytically, we solved Eq. 3.1 for $\mathbb{P}_t(S_a)$ to ensure that the S_a population fraction is only dependent on PW_b. If inducer **a** is used as a constant reference signal, all cells transition into either of S_a, S_b or S_{ab} state, thus $\mathbb{P}_\infty(S_a) = 1 - (\mathbb{P}_\infty(S_b) + \mathbb{P}_\infty(S_{ab}))$. If we assume that the basal leaky expression of intB is zero ($k_{\text{leak}B} = 0$), $\mathbb{P}_t(S_b) + \mathbb{P}_t(S_{ab}) = 0$ holds for $t \leq \Delta t$, since there is no intB that turns DNA state into S_b or S_{ab}. Then, we can show that $\mathbb{P}_t(S_b) + \mathbb{P}_t(S_{ab})$ is dependent only on PW_b, the duration of the pulse width of inducer B, for $t > \Delta t$. This conclusion holds as long as $k_{\text{leak}B}$ is negligibly small compared to other kinetic constants (k_{flipA}, k_{flipB}, k_{deg}, γ_A and $k_{\text{prod}B}$) (See Appendix Section 12.3 for full derivation).

If S_a population fractions can be modulated by changing PW_b, then conversely, we should be able to use measured experimental RFP population fractions as a way to determine PW_b. Once PW_b is known, then the S_{ab} fraction can be used to uniquely determine the time between inducers, Δt (Figure 3.7F). Furthermore, the genetically encoded state means that these population fractions should be maintained and measurable at a time, t_{end}, that is much later than the time of the events.

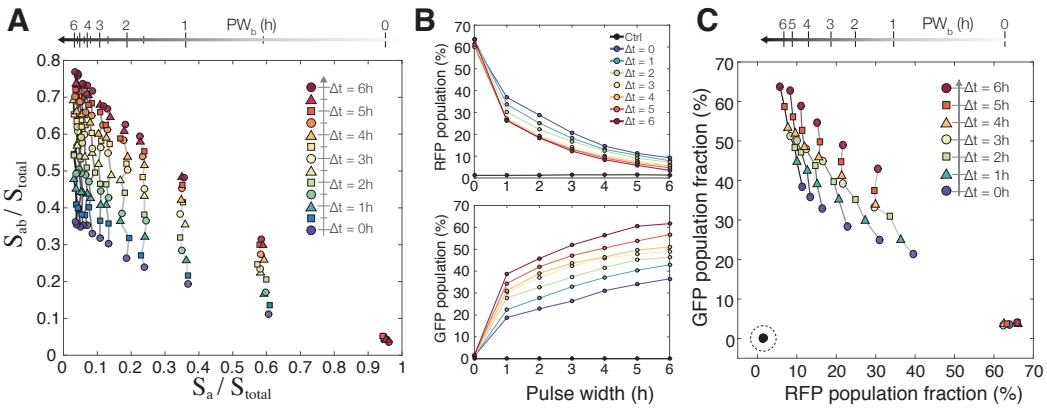


Figure 3.8: Determining arrival time and pulse duration of inducer **b** with population fractions. A) Simulation results from testing an 11×11 matrix of parameters with Δt and PW_b varying from 0 – 6 hours in increments of 0.5 hours. Each point represents a population of 3000 cells. Increasing PW_b goes from right to left, and increasing Δt goes from bottom to top. B) Experimental results showing RFP and GFP expression as a function of increasing Δt and PW_b. Fluorescence values have been normalized to the highest GFP and RFP fluorescence in the sample set. Experimental results from exposing temporal logic gate *E. coli* populations to varying PW_b and Δt values (0 – 6 hours, 0.01%/vol L-ara, 200 ng/ml aTc, measurements taken at 48 hours). C) A scatterplot of each population using their RFP and GFP fractions as coordinates ($\sim 10^6$ cells per population). The non-induced control samples are indicated with a dotted circle on the bottom left, and the samples with PW_b = 0h are on the bottom right. Samples with the same PW_b are connected with a solid line, and line darkness represents increasing PW_b duration. Samples with the same Δt are shown with the same colored shape marker and increasing Δt goes from bottom to top.

These conclusions can be extended in simulation to create a scatterplot of S_a cells versus S_{ab} cells in a population (Figure 3.8A) over an 11×11 parameter matrix varying Δt and PW_b from 0 – 6 hours in increments of 0.5 hours (Additional plots in Appendix Fig. ED-S17). Each point on the chart in Figure 3.8A represents a simulated population (N = 3000) exposed to a unique combination of Δt and PW_b values. Vertical lines represent the same PW_b value, and points with the same shape and color have the same Δt value. The simulation results suggest sufficient resolution of events as long as PW_b and Δt values are between 0 – 4 hours. For any single value of PW_b, we can follow the increasing Δt values vertically and see that the population response saturates after 4.5 hours resulting in overlapping between populations with $4.5 < \Delta t < 6$ hours. We can trace any individual Δt value horizontally from right to left, and observe that the points begin to cluster and overlap when $4.5 < PW_b < 6$ hours. These simulation data suggest that there should

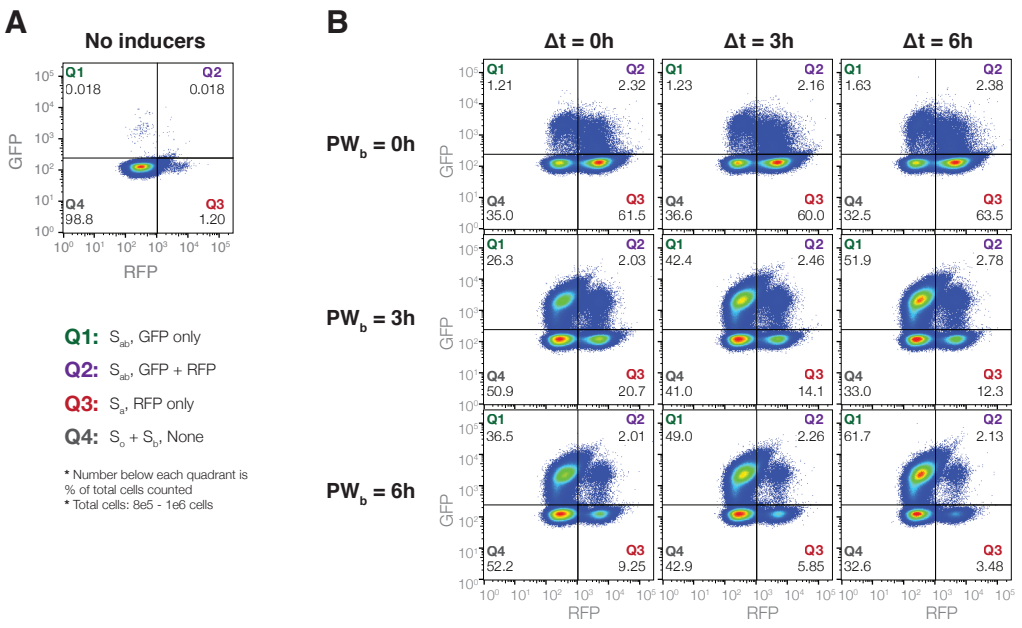


Figure 3.9: Selected flow cytometry panels for Figure 3.8C populations. Cell populations are gated by RFP and GFP fluorescence into quadrants Q1 (S_{ab}, GFP only), Q2 (S_{ab}, RFP and GFP), Q3 (S_a, RFP only), and Q4 (S_o,S_b, non-fluorescent). Percentage of total cells in each quadrant is shown under the quadrant label in each panel. A) Control population not exposed to any inducers. There is minimal leaky expression (1.2%) into Q3 after 36 hours of growth. B) Populations for PW_b and Δt values of 0, 3, and 6 hours. For PW_b = 0h populations, $\sim$ 60% of the cells switch to S_a (Q3), with $\sim$ 3% intB leaky expression going into Q1 and Q2. As PW_b increases, the S_a fraction drops from 60%(PW_b=0h) to 10-20%(PW_b=3h) to < 10%(PW_b=6h). As Δt increases, the percentage of cells in S_{ab} (Q1) increases from 20-40% (Δt = 0h) to 40-50% (Δt = 3h) to 50-60% (Δt = 6h). S_a populations (Q3) also drift downwards with increasing Δt, rather than staying constant as predicted in simulation. Lower Δt results in higher S_b populations, which, combined with S_o cells, make up Q4. Critically, the percentage of the population expressing both RFP and GFP simultaneously (Q2) is always < 3%. This ensures that RFP is a reliable determinant of S_a state cells, and subsequently, of PW_b.

be some defined detection range of Δt and PW_b where each possible combination of the two is uniquely identifiable.

Experimentally, we tested a 7×7 matrix of varying Δt and PW_b (0 – 6 hours, 1 hour increments) on independent populations of the temporal logic gate *E. coli* strain (Figure 3.8B). All populations, except for the control, were exposed to inducer **a** (L-ara 0.01%/vol) at time t_0 to t_{end}. Pulses of inducer **b** (aTc, 200ng/ml) were achieved by sampling 5μl of the population and diluting 1:100 into fresh media with only

inducer **a** (M9CA + 0.01%/vol L-ara). Populations were sampled in triplicate and pooled for flow cytometry analysis after 24 additional hours of growth in inducer **a** ($\sim$ 36 hours after start of experiment)(Figure EV3.5). Variability between triplicates measured via bulk fluorescence can be found in Appendix Figs. ED-S18, ED-S19. For all values of Δt, the number of S_a cells (RFP) is highest when there is no exposure to inducer **b** (PW_b = 0h) and decreases monotonically as a function of PW_b (Figure 3.8B, top). We see a more pronounced separation of the Δt curves when we look at S_{ab} (GFP) cell fractions (Figure 3.8B, bottom). The number of S_{ab} cells is dependent on both Δt and PW_b and increases proportionally with both increasing **b** pulse duration and inducer separation time.

By counting population fractions of RFP versus GFP expressing cells, we can resolve the different populations that result from varying Δt and PW_b values (Figure 3.8C). As with Figure 3.8A, each point on the graph represents an independent population of cells (OD $\sim$ 0.7, $\sim 10^6$ cells counted per population). All of the populations exposed to either or both of the inducers occupy fractional coordinates that are unique from that of the *no inducer* controls (indicated by dotted circle). We see that if Δt is constant and PW_b increases (Figure 3.8C, right to left) then the S_a fraction decreases as S_{ab} fractions increase. For constant PW_b with increasing Δt(Figure 3.8C, bottom to top), the S_a cell fraction remains mostly constant relative to increasing S_{ab}. In the case where there is no **b** pulse (PW_b = 0h), we see maximum S_a (RFP) cell fractions of about 60% with minimal S_{ab} populations that are about the same as *no inducer* S_{ab} levels. Overall, populations with different PW_b exposures are well separated by S_a (RFP) fraction up to 4 hours. Even for PW_b at 5 and 6 hours the populations have unique S_a/S_{ab} coordinates, just not unique S_a fractional values.

This method of profiling is only valid if the fraction of S_a state cells can be used as a measure of PW_b that is independent of Δt. In previous experiments with step inputs (Figure EV3.5), there would be a significant population of cells with both GFP and RFP fluorescence, since they had transitioned to S_{ab} but had not yet fully diluted out built up RFP protein levels from being in S_a for extended periods. If a significant percentage of the population remained in this transition state (Q2), that would make RFP an unreliable measure of S_a state cells. However, flow cytometry analysis of the pulse-modulated populations (Figure EV3.5) showed that although there were some cells expressing both RFP and GFP (Q2), these cells were always less than 3% of the total population. (Additional flow cytometry analysis can be found in Appendix

Fig. ED-S20–ED-S25.) Thus, RFP was measured to be a reliable determinant of S_a state cells, and subsequently, of PW_b.

For any given PW_b, we observed higher experimental S_a (RFP) population fractions with lower Δt (Figure 3.8A top), resulting in a diagonal slant for each value of PW_b (Figure 3.8C). Upon further investigation, we believe this is due to a slower $\mathrm{S_o} \xrightarrow{\alpha_1} \mathrm{S_b}$ transition than we anticipated. In our model, we assume $\mathrm{S_o} \xrightarrow{\alpha_1} \mathrm{S_b}$ equal to $\mathrm{S_a} \xrightarrow{\alpha_3} \mathrm{S_{ab}}$, since both transitions are mediated by intB. However the diagonal S_a fractions with increasing Δt for each value of PW_b suggest that the α_1 transition rate may be actually be slower than α_2 or α_3. Simulation results with adjusted transition rates ($\alpha_1 < \alpha_2 = \alpha_3$) recapitulated the slanting S_a population fractions (Appendix Fig. ED-S27). This inequality in transition rates could have arisen from differences in DNA sequence length or from differences in the DNA excision required for $S_o \rightarrow S_a$ instead of the recombination that occurs in the other transitions. Differences in DNA excision or recombination for a single integrase are important experimental parameters, but do not ultimately affect our conclusions about the overall system. Despite unequal intB transition rates, experimental implementation of the temporal logic gate still produces unique (S_a, S_{ab}) fractional coordinates for each combination of Δt, PW_b, even though S_a values are not unique for higher PW_b.

Model-informed predictions on population fractions in response to pulses of inducer **b** led to experiments that could produce unique S_a and S_{ab} coordinates for different combinations of Δt and PW_b. However, experimental data also revealed areas in which the model had been oversimplified. While it is important to have a model to understand overall properties and limitations of the experimental system, it is also impractical to design simulations that can account for all possible variations that might occur in the implementation of biological devices. Therefore, we believe that future workflows should also involve calibration protocols for specific applications of engineered biological populations.

Practical use and calibration of populations for event detection

Curve-fitting methods were used to automatically convert experimentally measured RFP and GFP population fractions into PW_b and Δt values and to evaluate the resolution with which population ratios can be used to determine inducer separation time and pulse duration. Using the experimental data from Figure 3.8B,C, we generated fitting curves for PW_b as a function of RFP population percentage(R), and for Δt as a function of both GFP population percentage (G) and PW_b (Appendix

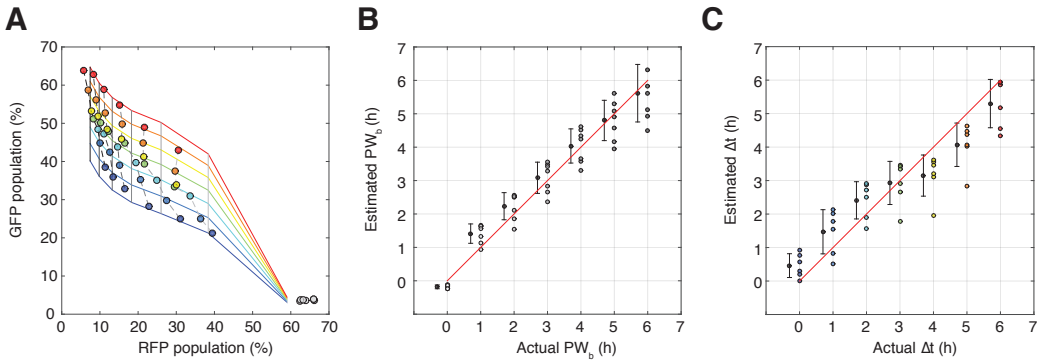


Figure 3.10: Determining prediction resolution for PW_b and Δt from population data. A) A mesh generated from fitted curves for PW_b as a function of RFP population percentage(R) and Δt as a function of pulse width and GFP population percentage(G). Experimental data is overlaid. B) Comparison of actual versus estimated PW_b values generated by fitted function $PW_b(R)$. For each actual PW_b value, the average of the estimated PW_b values with ± 1 standard deviation (Slightly offset on the x-axis for better comparison). C) Comparison of actual versus estimated Δt generated by the fitted function $\Delta t(G, PW_b)$. For each actual Δt values, the average of the estimated Δt with ± 1 standard deviation (Slightly offset on the x-axis for better comparison).

Fig. ED-S28, ED-S29,Table S3). We will denote these functions with $PW_b(R)$ and $\Delta t(G, PW_b)$, respectively. The functions $PW_b(R)$ and $\Delta t(G, PW_b)$ can then be used to generate a mesh of estimated PW_b and Δt values for any given normalized fluorescence values (Figure 3.10A, Appendix Eq. 8-11).

The estimated values were compared against the actual values to determine the approximate time window with which a specific PW_b or Δt can be resolved. For each actual value of PW_b and Δt, we calculated the average and standard deviation for the set of estimated values. The standard deviation allows us to visualize the range for which the majority of predictions will fall for any given actual value. For instance, a PW_b of 1 hour can be detected ±0.25 hours, but as PW_b increases, this prediction window widens and for $PW_b \geq 3$ hours, the resolution of detection is closer to ±1 hour (Figure 3.10B). Similarly, predicted values of Δt fall within ±0.5 hours for $0 < \Delta t < 3$ hours and increase to ±1 hours when $\Delta t \geq 3$ hours (Figure 3.10C). Using these fitting functions, we can also pre-generate a reference table that converts RFP and GFP population fractions into predicted PW_b and Δt values (Appendix Table S4).

3.4 Discussion

Engineered biological systems have inherent capabilities for replication, parallel processing, and energy efficiency. These advantages rely on the existence of bacteria not as single cells, but as populations. As the field moves forward with synthetic gene circuits, it is important to understand outcomes not just as single-cell outputs but as overall population-level distributions.

We have designed and implemented a temporal logic gate that takes advantage of the population dynamics to collectively sense and record sequences of transient chemical inputs. We show both that single cells independently sense and record events, and that aggregate population fractions create unique outcomes that provide information not encoded in single cells. As with all engineered systems, proper calibration of these temporal logic gate populations will be required prior to deployment in the "field." We envision a process similar to the one described in this report. First, experimental populations are exposed to a matrix of PW_b and Δt values. This will set the maximum and minimum RFP and GFP population fractions and provide necessary data for determining the Δt_{90} limit and producing the fitting functions $PW_b(R)$ and $\Delta t(G, PW_b)$. Once the fitting functions have been determined, values for PW_b and Δt for experimental samples can be estimated within ± 0.25 to 1 hour of the actual values. A calibrated table could also be generated and used for as a reference for samples that have been exposed to unknown conditions.

The stochastic nature of molecular processes often presents a significant barrier to homogenous outputs from an engineered population of cells. This implementation of event detection via population fractions takes advantage of stochastic and heterogenous individual responses to environmental conditions in order to map final population fractions back to unique sequences and durations of chemical events. The sensitivity of the system and the Δt_{90} detection limit could potentially be modulated by increasing or decreasing protein production rates via tuning of plasmid copy numbers, signal concentration, or transcription/translation sequences. The use of digital cellular outputs combined with the analog population response creates event detection systems that are more robust to stochasticity and can be tuned more easily. We plan to explore these possibilities in future work.

As a proof-of-concept, we have used the common laboratory inducers L-arabinose and aTc as inputs, but we anticipate that our temporal logic gate system could be used modularly with any biosensors of choice. In particular, we believe there are possibilities for detection of miRNAs and biofilm formation. Stable populations

of microRNAs (miRNAs) circulating in the blood have generated a lot of interest as biomarkers for human health (Cortez et al., 2011). These short (~ 20-30nt) regulatory RNAs have been shown to have sequential tissue-specific expression signatures that correlate with pregnancy, tumor formation, and other diseases (Gilad et al., 2008; Mitchell et al., 2008), and synthetic biology has developed many customizable RNA sensors (Friedland et al., 2009; Green et al., 2014). Detection of miRNAs would require implementation of the temporal logic gate in mammalian cells. Though recombinase-based synthetic circuits have not been shown in mammalian cells, serine integrases have been used quite effectively in a wide variety of mammalian cell types, primarily for genome editing and integration (Keravala et al., 2006; Xu et al., 2013).

Another possible application of this would be detection of harmful biofilms. Biofilms are self-assembling, highly structured, multi-species consortia that develop in stages and have sophisticated networks of interaction and function (Stoodley et al., 2002; Flemming and Wingender, 2010; Elias and Banin, 2012). Unnatural biofilm development in environments such as industrial water sources or waste streams can be both harmful for both the natural environment and the industrial mechanisms. Detection of biomarkers for known strains of biofilm colonizers would provide early warning of changing ecosystems, and although we do not yet fully understand these networks, it is known that quorum-sensing plays a critical role in the process. Quorum-sensing molecules and receptors are available in the synthetic biology toolbox and so may provide an accessible way of detecting the sequential colonization of different microbes. Field deployment of engineered bacteria will likely involve transient signals, low-nutrient environments, and possibly even other microbial competitors (i.e. soil, flowing rivers, the digestive tract). We used minimal media in this study to better approximate low-nutrient environments, and anticipate further characterization in more customized 'local' environments (i.e. gut model or air model or soil model) and with hardier microbial chassis.

Finally, this study focused on the population outputs as indicators of past events, but we believe that this temporal logic gate could be used to reliably differentiate a single strain into controlled sub-populations via input pulse order, duration, and frequency. In recent years, it has been recognized that many natural systems modulate cellular behavior not only by changing the concentration of signaling molecules but also by regulating signal pulse frequency (Cai, Dalal, and Elowitz, 2008; Lin et al., 2015). If we consider the fluorescent proteins GFP and RFP in this circuit as

simply placeholders for downstream genes, then this system could easily be applied as a top-down population differentiator. By modulating the sequence of inputs, one could systematically predict and create mixed populations of genetically differentiated cells. This greatly expands our capability to design synthetic systems that have controllable distributions as outcomes, not just digital on/off phenotypes. Furthermore, we can then begin to develop frameworks for understanding the role of feedback and control theory in modulating these sub-populations given different starting distributions or uneven growth rates due to resource limitations. As the scientific community turns towards further understanding of microbiomes and multi-cellular consortia, engineered bacteria populations could be used not only as a tool for investigating the activities of natural communities but also as a way to build synthetic communities from the ground up.

3.5 Materials and methods

Cell strains and plasmids

All plasmids used in this study were designed in Geneious 7.1 (Biomatters, Ltd.) and made using standard Gibson isothermal cloning techniques. Integrases Bxb1 and TP901-1 are on a high-copy plasmid (pVHed05, plasmid map in Figure ED-S9) with a ColE1 origin of replication (original template from the Dual Recombinase controller (Bonnet, Yin, et al., 2013), Addgene Plasmid 44456). Integrase A (Bxb1) is behind a Ptet promoter and integrase B (TP901-1) is behind a PBAD promoter. The plasmid has been modified with an additional TetR gene. The temporal logic gate was integrated into the Phi80 site on the *E. coli* chromosome using CRIM integration (Haldimann and Wanner, 2001) and screened for single integrant colonies. The integration plasmid template and DH5α-Z1 strain were generously provided by J. Bonnet and D. Endy and modified to contain the temporal logic gate (pVHed07, plasmid map in Figure ED-S9).

Additional DNA and oligonucleotides primers were ordered from Integrated DNA Technologies (IDT, Coralville, Iowa).

A custom formulation of M9CA media was used for all experiments. The media contained 1x M9 salts (Teknova, M1906) augmented with 100mM NH_4CL, 2mM $MGSO_4$, 0.01% casamino acids, 0.15 μg/mL biotin, 1.5 μM thiamine, and 0.2% glycerol, and then sterile filtered (0.2 μm).

Simulations of the model

The stochastic simulation algorithm by Gillespie (Gillespie, 1977) was implemented to generate the sample paths of individual cells using the Markov model (see Table S6 for the definitions of Markov transitions and transition rates). All simulation runs and their analyses were done with MATLAB (R2014b,The MathWorks, Inc.). Simulated populations were done with 3000 – 5000 individual cell trajectories.

Experimental methods

Prior to all experiments, cells were grown overnight from plate cultures in M9CA for two days, then diluted to OD 0.1 and recovered for 4-6 hours at 37^oC. L-arabinose and anhydrous tetracycline (aTc) were used as inducers **a** and **b**, respectively. L-ara was used a concentration of 0.01% by volume, and aTc was used a concentration of 200 ng/ml (450nM). All media contained the antibiotics chloramphenicol (Sigma Aldrich, Inc (C0378); 50μg/ml) and kanamycin (Sigma Aldrich, Inc (K1876); 30μg/ml). All experiments were performed with the aid of timed liquid handling by a Hamilton STARlet Liquid Handling Robot (Hamilton Company).

For step function experiments, the cells were diluted to OD 0.06–0.1 into a 96-well matriplate (Brooks Automation, Inc., MGB096-1-2-LG-L) with 500μl total volume in M9CA. Cultures were incubated at 37^oC in a BioTek Synergy H1F plate reader with linear shaking (1096 cycles per minute) (BioTek Instruments, Inc.) and inducers were added at appropriate time by the Hamilton robot. OD and fluorescence measurements (superfolder-GFP ex488/em520, mKate2-RFP ex580/em610) were taken by the BioTek every 10 minutes. Each experimental condition was done was triplicate on the plate.

For the pulse experiments, single 500μl cultures were grown at 37^oC in the BioTek plate reader (linear shaking, 1096 cycles per minute) and inducers added at time Δt by the Hamilton liquid handler. Pulses were achieved through dilution of the culture into fresh M9CA media containing 0.01% L-arabinose. The Hamilton was programmed to sample 5 μl of the culture and dilute it into 500 μl of fresh M9CA + 0.01% L-ara to achieve pulsatile exposure to aTc. This was done in three independent triplicates for each experimental condition. 96-well deep-well plates containing the diluted cultures were then incubated at 37^oC incubated for an additional 24 hours ($\sim$ 36 – 40 hours from start of experiment). Final endpoint populations were measured using the plate reader and also stored and further analyzed using flow cytometry.

Analysis of experimental data was done using custom MATLAB scripts. All de-

picted error bars are standard error of the mean. Fitting of curves was done in MATLAB.

Flow cytometry

Experimental cultures were spun down, washed, and resuspended in sterile PBS with 15% glycerol and stored at -80C (Jahn et al., 2013). Cultures were then thawed on ice, and diluted to 10^6 cells/ml in sterile PBS prior to running on the flow cytometer. Flow cytometry was done using a MACSQuant(r) VYB (Miltenyi Biotec, Germany) at the Caltech flow cytometry core facility. Flow data analysis and gating was done with FlowJo Version 10.0.8r1 (Flowjo, LLC, Ashland, OR). For inducer separation time experiments shown in Figure 3.4, $\sim 10^5$ cells were measured per population. For pulse induction experiments shown in Figure 3.8, $\sim 10^6$ cells were measured per population.

3.6 Acknowledgements

The authors would like to thank J. Bonnet and D. Endy for the initial plasmids used in this work, S. Sanchez for critical assistance in automation and liquid handling, D. Perez for flow cytometry assistance, and C. Hayes for discussions.

V.H. is supported by the U.S. Department of Defense (DoD) through the National Defense Science & Engineering Graduate Fellowship (NDSEG) Program. Y.H. is supported by JSPS Fellowship for Research Abroad. Research supported in part by the Institute for Collaborative Biotechnologies through grant W911NF-09-0001 from the U.S. Army Research Office. The content of the information does not necessarily reflect the position or the policy of the Government, and no official endorsement should be inferred.

3.7 Author Contributions

V.H. conceived of the circuit design, constructed the necessary experimental strains, performed experimental work and data analysis, ran model simulations, and wrote the manuscript. Y.H. developed the stochastic model and derived the mathematical results. P.W.K.R provided feedback and guidance on data analysis and interpretation. R.M.M provided feedback and guidance on overall project vision, circuit design, and interpretation of results.

3.8 Conflict of Interest

The authors declare that they have no conflict of interest.

3.9 Appendix (Supplementary Information)

A Markov Model for integrase recombination

Parameters used for simulations

The parameters were chosen to be in biological orders of magnitude. Tetrameterization of the integrase is represented with the expression

$$\alpha_i(\mathrm{Int}_*) := k_{\mathrm{flip}*}\left(\frac{\mathrm{Int}_*(\mathrm{Int}_*-1)(\mathrm{Int}_*-2)(\mathrm{Int}_*-3)}{K_{\mathrm{d}*}^4+K_{\mathrm{d}*}^3\,\mathrm{Int}_*+K_{\mathrm{d}*}^2\,\mathrm{Int}_*(\mathrm{Int}_*-1)+K_{\mathrm{d}*}\mathrm{Int}_*(\mathrm{Int}_*-1)(\mathrm{Int}_*-2)+\mathrm{Int}_*(\mathrm{Int}_*-1)(\mathrm{Int}_*-2)(\mathrm{Int}_*-3)}\right) \quad (3.4)$$

for $i = 1, 2, 3$ and $* = A, B$ (see **Tetramerization of integrase** in Appendix Section 12.2 for the derivation, and Fig. S1).

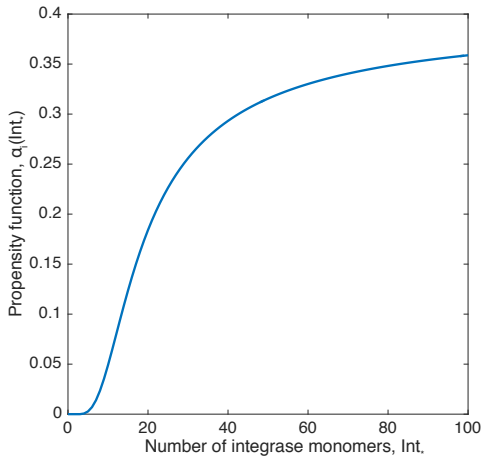


Figure ED-S1: Visualizing the nonlinear term for integrase tetramerization (Eq. 3.4). The propensity function, $\alpha_i(\mathrm{Int}_*)$, as a function of integrase monomers, Int_*, is zero until at least four monomers are present. Parameters for flipping and dissociation constant are $k_{\mathrm{flip}*} = 0.4\ \mathrm{h}r^{-1}$, and $K_{\mathrm{d}*} = 10$ molecules.

Parameter	Value		Units
α_1	See equation (3.4)		
α_2	See equation (3.4)		
α_3	See equation (3.4)		
δ_A	k_{deg}(IntA)		
δ_B	k_{deg}(IntB)		
γ_A	$k_{\text{prod}A} + k_{\text{leak}A}$,	if inducer **a** exists	
	$k_{\text{leak}A}$,	if no inducer present	
γ_B	$k_{\text{prod}B} + k_{\text{leak}B}$,	if inducer **b** exists	
	$k_{\text{leak}B}$,	if no inducer present	
$k_{\text{prod}A,B}$	50		$(\mu m^3 \cdot hr)^{-1}$
k_{deg}	0.3		hr^{-1}
$k_{\text{flip}A}$	0.4		hr^{-1}
$k_{\text{flip}B}$	0.4		hr^{-1}
$k_{\text{leak}A}$	0		$(\mu m^3 \cdot hr)^{-1}$
$k_{\text{leak}B}$	0		$(\mu m^3 \cdot hr)^{-1}$
$K_{\text{d}A}$	10		molecules
$K_{\text{d}B}$	10		molecules

Table 3.1: Initial Markov transition rates and parameters. We define the rate of DNA state transitions as α_1, α_2 and α_3 using the rate of DNA flipping for a unit concentration of IntA($k_{\text{flip}A}$) and IntB($k_{\text{flip}B}$). The notations IntA and IntB denote the copy number of each integrase, and $[S_o] = 1([S_a] = 1)$ if the DNA state is $S_o(S_a)$ and $[S_o] = 0([S_a] = 0)$ otherwise. The production and degradation rates of the integrases are defined by γ and δ, respectively. $k_{\text{prod}A}$ and $k_{\text{prod}B}$ are the protein production rate constants, and $k_{\text{leak}A}$ and $k_{\text{leak}B}$ are the basal leaky expression rate constants. We assume the plasmid copy number is proportional to the volume of a cell. In this paper, we use 1 μm^3 (1 femtoliter) as the estimated volume of a single *E. coli* cell. The integrase degradation/dilution rate constant $k_{\text{deg}} = 0.3\text{hr}^{-1}$ sets the protein half-life to approximately 2.3 hours. The binding constant, $K_{\text{d}*}$ was estimated based on Bxb1 K_d binding constants 70 nM (Singh et al, 2013). When converted into molecules in 1 femtoliter volume, this translated to 7 molecules ($70\text{nM} \times \frac{10^{23}\text{molecules}}{\text{L}} \times 10^{-9} \times \frac{1L}{10^{15}\mu m^3} = 7$ copies.)

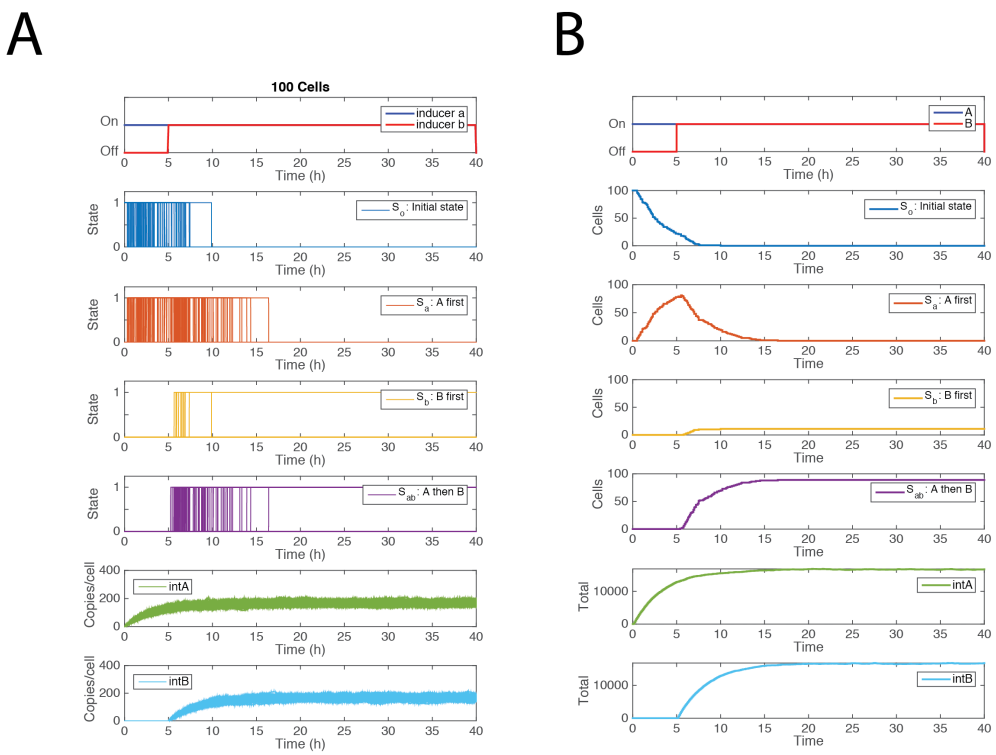


Figure ED-S2: Example of individual cell trajectories and total summed population from stochastic simulations (Δt = 5h). A) Individual simulated cell trajectories for the possible cell states. A sample of 100 cells out of the population of 5000 has been shown here for clarity. The panels, from top to bottom, show time and duration of induction, cells in state S_o (blue), cells in state S_a (red), cells in state S_b (yellow), cells in state S_{ab} (purple), copies per cell of integrase A (green), and copies per cell of integrase B (sky blue). B) Summed totals of all possible DNA and protein states for all 5000 cells.

Characterization of inducer separation time

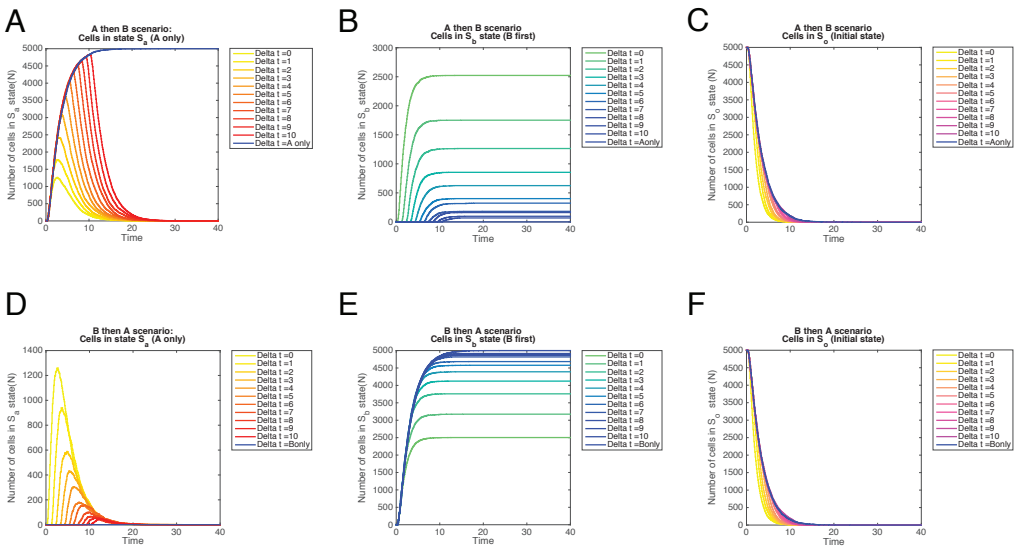


Figure ED-S3: Effect of varying Δt on number of S_a, S_b, and S_o state cells in simulation (Supplementary to Figure 3 in the main text). For each value of Δt, a population of 5000 individual cell trajectories was generated and summed. A) S_a cell counts for E_{ab} event. In the case of *a only*, 100% of the cells become S_a. For the other states, the cell count drops off at time Δt as S_a transition into S_{ab}. B) S_b state cells for E_{ab}. The number of S_b cells that transition is a function of available S_o cells left at time Δt. With high Δt, the most cells are already in S_a. C) S_o state cells for E_{ab} decrease exponentially with time as they convert into either S_a or S_o. D) S_a state cell count with an E_{ba} event are inversely proportional to Δt. E) S_b state cells gain fractional dominance with increasing Δt during with an E_{ba} event. In the case of *b only*, 100% of the cells become S_b. F) S_o state cells decrease exponentially with E_{ba} event as well.

Experimental results for varying inducer separation time

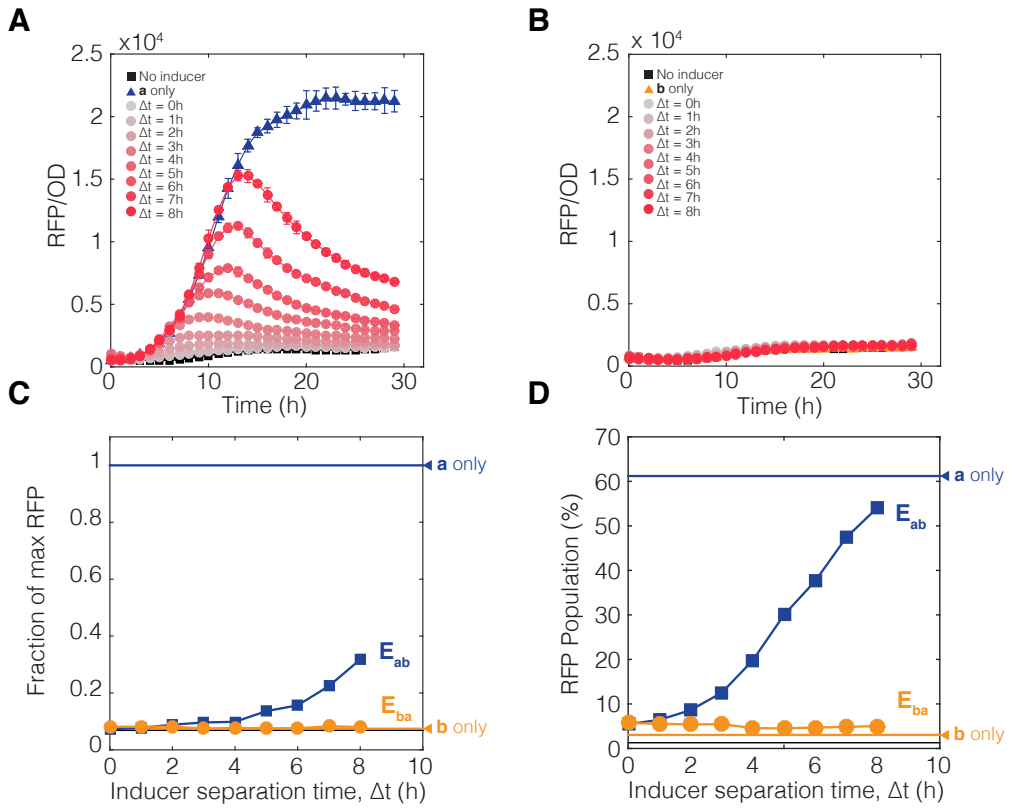


Figure ED-S4: RFP expression for *in vivo* experiments with increasing Δt (Supplementary to Figure 4 in the main text). A) RFP expression as a proxy for S_a state cells when population is exposed to E_{ab}. B) RFP expression when the inverse E_{ba} event occurs. C) Endpoint RFP bulk fluorescence measurement of cultures as a function of Δt. In an infinite step induction experiment, we expect no cells to be expressing RFP since all S_a cells become S_{ab}. However, cultures with later Δt values spend up to 8 hours in S_a and build up a lot of RFP that does not completely dilute even upon switching to S_{ab}. D) Flow cytometry counts of RFP population. The flow cytometry shows that a high percentage of cells are expressing a low amount of RFP. Quadrant analysis of RFP vs GFP populations shows that these RFP-expressing cells are all in Q2, the transitory quadrant in which cells have switched to S_{ab} but still have undiluted RFP molecules (Figure EV1). See also Appendix Figure ED-S12 for conversion between flow populations and bulk intensity measurements.

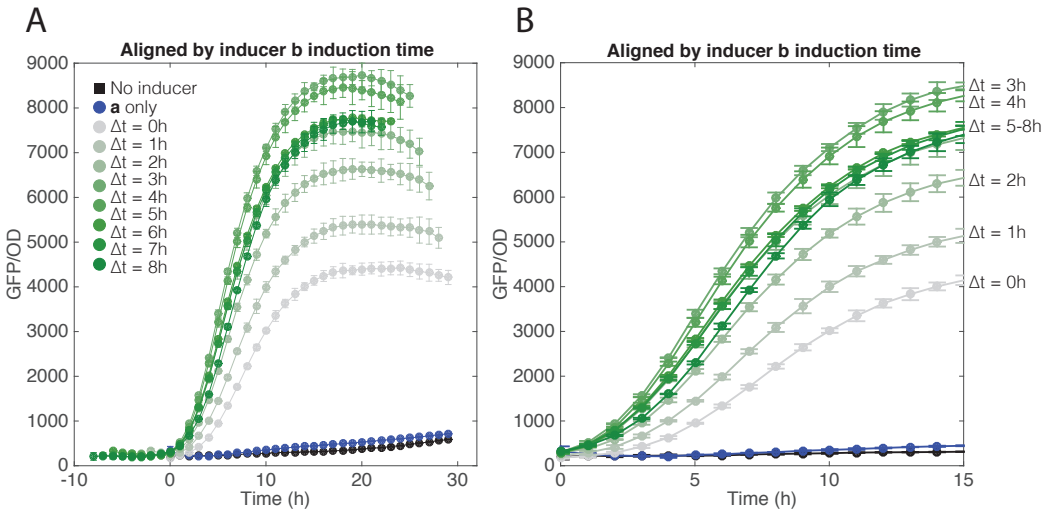


Figure ED-S5: *In vivo* GFP expression curves aligned by Δt (Supplementary to Figure 4 in the main text). A) Curves have been aligned by Δt such that cell switching to S_{ab} and GFP expression all starts at time 0. B) Zoomed in panel shows different slopes of the various Δt curves.

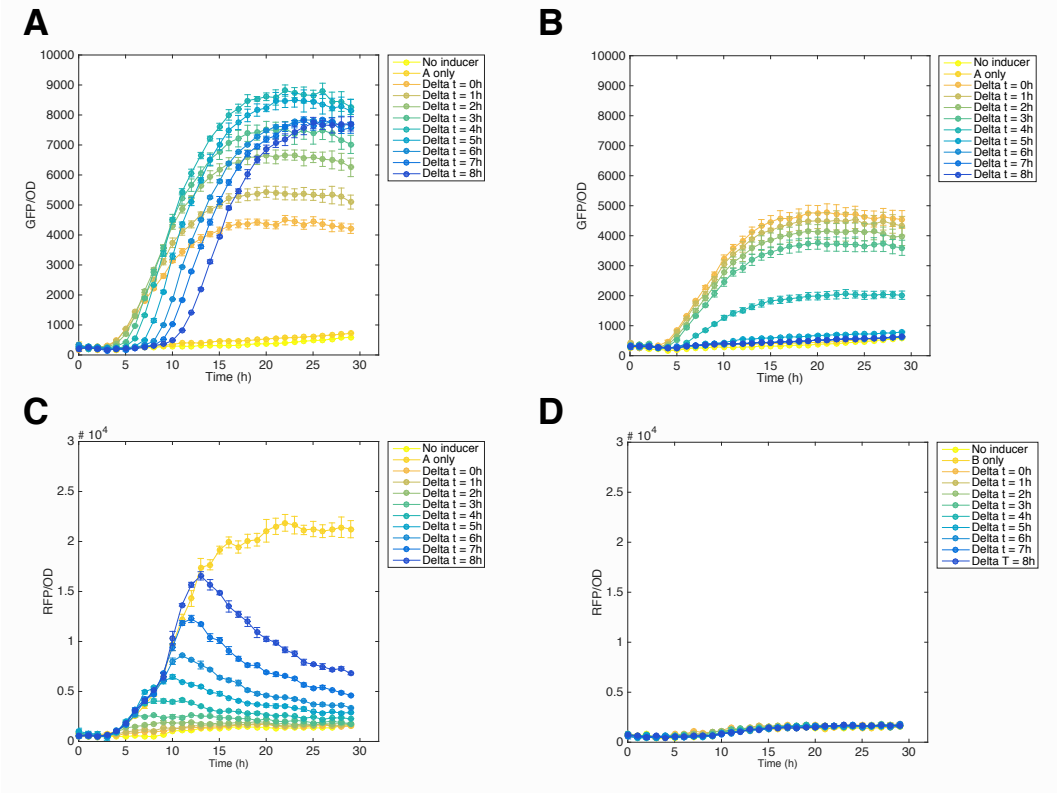


Figure ED-S6: Time-course data for Figure 4 with more separated color scheme. The color gradient used in Figure 4 can make it difficult to distinguish individual curves, and so here we have more color-separated plots. A) GFP fluorescence with event E_{ab}. B) GFP fluorescence with event E_{ba}. C) RFP fluorescence with E_{ab}. D) RFP fluorescence with E_{ba}.

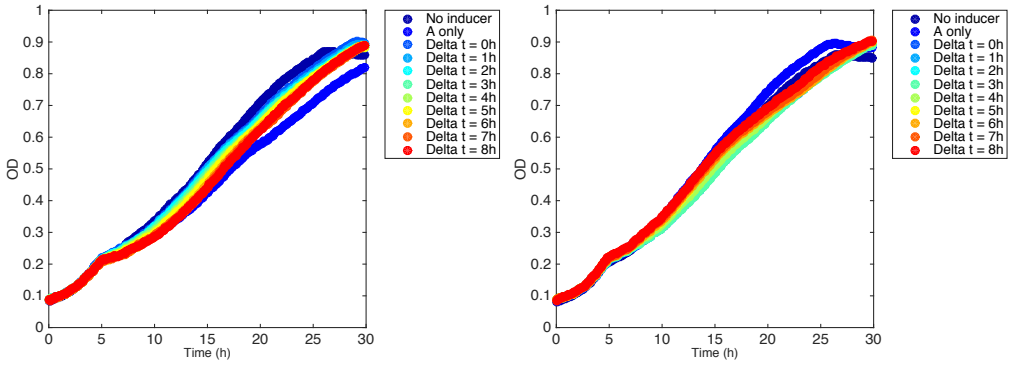


Figure ED-S7: OD growth curves for *in vivo* experiments with increasing Δt (Supplementary to Figure 4 in the main text). Growth curves are fairly linear due to growth in M9CA minimal media at 37C. A) OD growth curves for cells subjected to E_{ab} event. B) OD growth curves for cells subjected to E_{ba} event

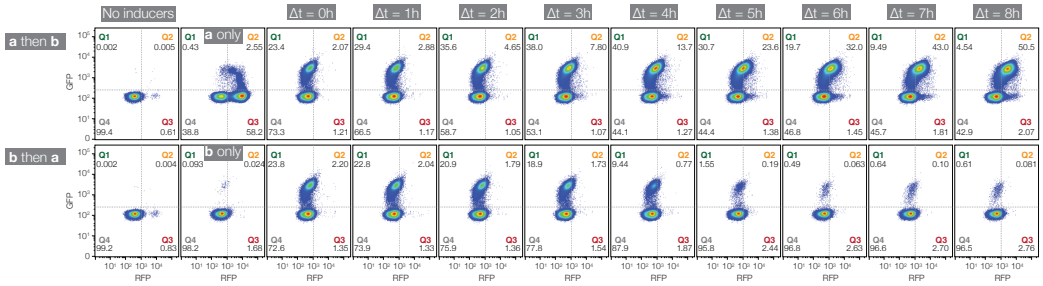


Figure ED-S8: Flow cytometry populations, RFP vs GFP (Fig. 4). Populations are split into quadrants Q1 (GFP only, S_{ab}), Q2 (GFP + RFP, S_{ab}), Q3 (RFP only, S_a), and Q4 (non-fluorescent, S_o + S_b.) ~ 100,000 cells were analyzed for each population.

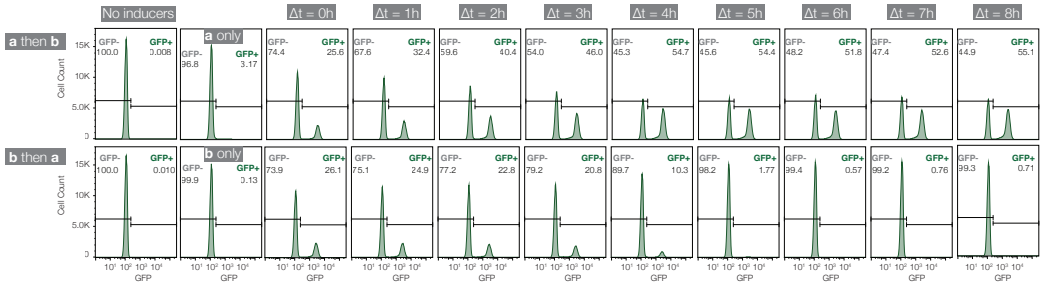


Figure ED-S9: Flow cytometry GFP histograms (Fig. 4). ~ 100,000 cells were analyzed for each population.

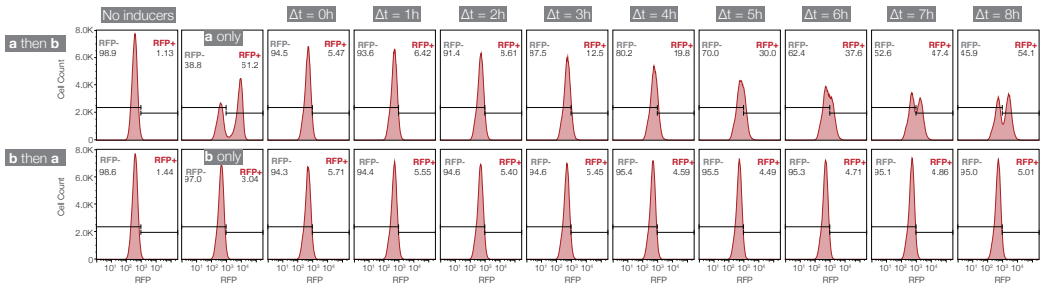


Figure ED-S10: Flow cytometry RFP histograms (Fig. 4). ~ 100,000 cells were analyzed for each population.

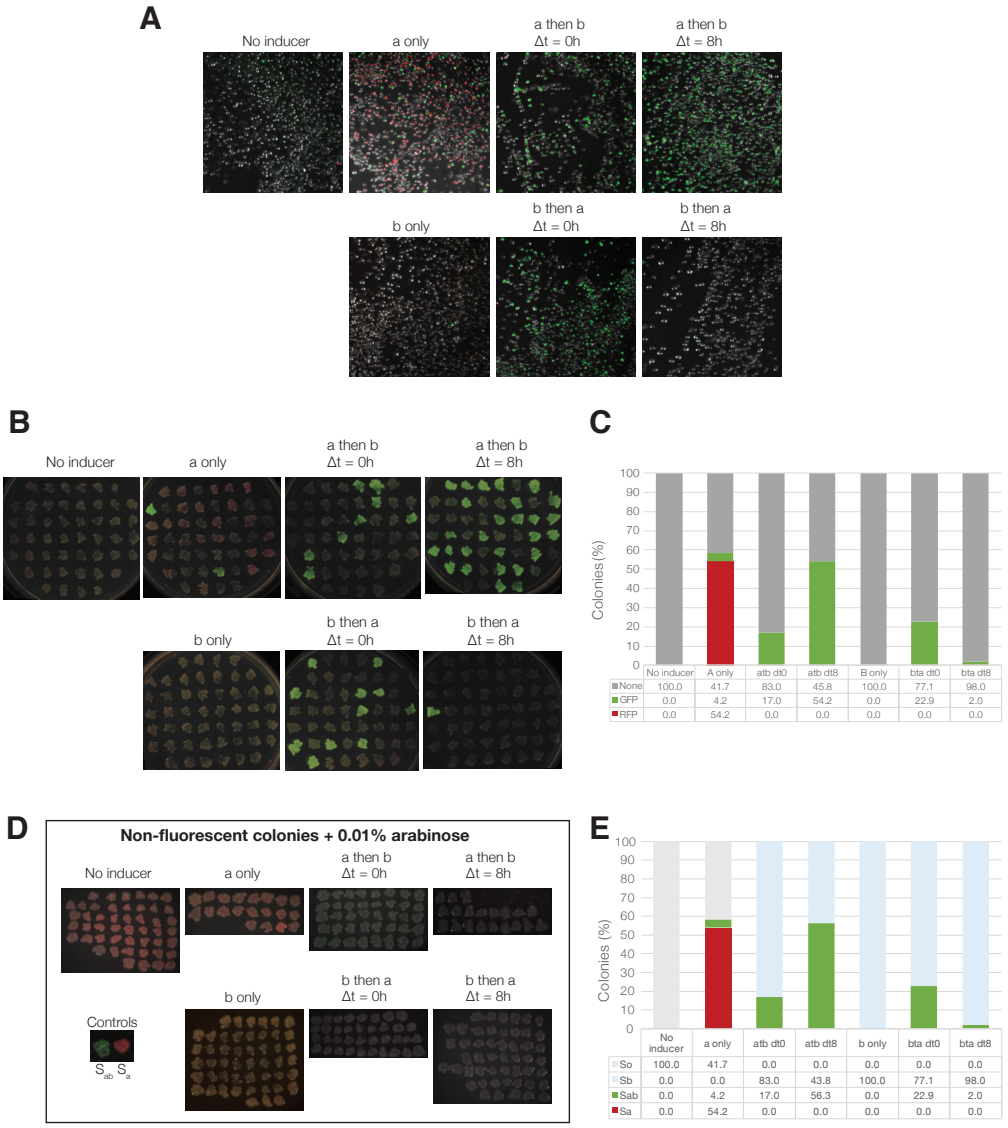


Figure ED-S11: Single colony analysis of non-fluorescent colonies in Δt experiment (Fig. 4) was done to determine genetic state (S_o or S_b). A) Experimental cultures were diluted 1:10,000 after experiment and plated on LB agar plates with no inducer. B) 48 ± 2 single colonies were randomly picked from each plate and re-streaked on a new agar plate. C) Single colonies were counted based on fluorescence and the resulting distributions are similar to those measured via flow cytometry. D) The non-fluorescent colonies from each condition were then re-streaked again onto plates with 0.01% arabinose. Only S_o cells would turn red (S_a), while S_b cells would remain non-fluorescent. We determined that 100% of the *no inducer* and *a only* non-fluorescent colonies were S_o, while the 100% of the non-fluorescent colonies in the other experimental conditions were S_b. E) Revised genetic state distributions based on single colony analysis of non-fluorescent colonies.

Comparing plate reader fluorescence readings with flow cytometry

We were interested to know how bulk culture florescence compared with actual single cell expression profiles. With bulk fluorescence, it is possible that bimodal expression of fluorescent molecules result in a few bright cells dominating the overall fluorescence measurement, and so we wanted to ensure that this was not the case with our time-course measurements. Endpoint bulk fluorescence was measured via BioTek Synergy H1F plate reader (BioTek Instruments, Inc, VT, USA) and normalized by the maximum fluorescence. Flow cytometry was done with a MACSQuant VYB flow cytometer (Miltenyi Biotec, Germany), and for both RFP and GFP, cells were counted and their relative fluorescence intensity was measured. Flow cytometry data was gated using FlowJo Version 10.0.8r1 (Flowjo, LLC, Ashland, OR).

In Figure ED-S12, we compare bulk fluorescence (Fig. ED-S12A) with flow cytometry populations (Fig. ED-S12B), then reconstruct the bulk fluorescence measurements by multiplying the cell counts with average measured intensity (Fig. ED-S12C). We find that GFP fluorescence is not disproportionately skewed by bulk fluorescence, indicating that cells that are "on" in state S_{ab} have a relatively tight distribution and are not overly dominated by a minority of bright cells. This can also be seen in the GFP histograms (Fig. ED-S9). For this experiment, in which both inducers are present long after Δt induction, we would expect no cells to remain in state S_a. However, we still measure RFP at the endpoint. We find that this RFP is leftover RFP from cultures that spent more time in S_a prior to transitioning to S_{ab} (Fig. EV1). These cells have stopped production of RFP, but existing RFP concentrations have not yet diluted completely. Flow cytometry analysis of cell counts find a high number of cells with low RFP fluorescence. This can also be seen in the RFP histograms (Fig. ED-S10).

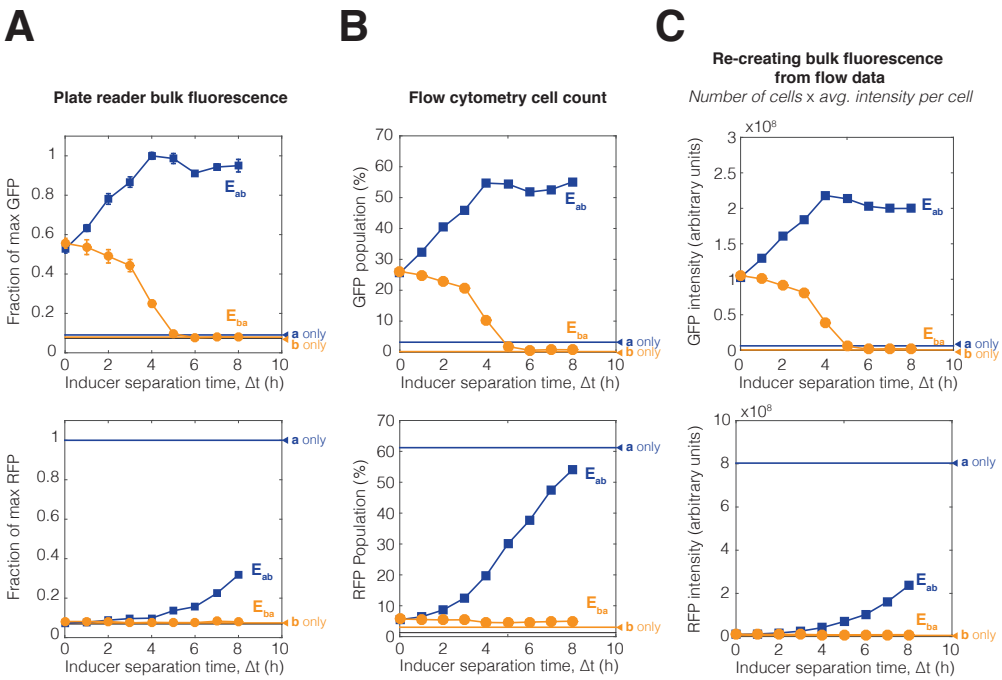


Figure ED-S12: Comparison of plate reader bulk fluorescence readings with flow cytometry cell counts. A) Bulk fluorescence GFP and RFP readings normalized by max GFP and max RFP. B) Flow cytometry counts of cell percentages about GFP and RFP gated thresholds. C) Re-creating bulk fluorescence data from average intensity per cell multiplied by number of cells.

Varying model parameters for integrase activity and leaky basal expression

Parameter	Value	Units
$k_{\text{prod}A}$	50	$(\mu m^3 \cdot hr)^{-1}$
$k_{\text{prod}B}$	50	$(\mu m^3 \cdot hr)^{-1}$
k_{deg}	0.3	hr^{-1}
$k_{\text{flip}A}$	0.2	hr^{-1}
$k_{\text{flip}B}$	0.3	hr^{-1}
$k_{\text{leak}A}$	$0.01*k_{\text{prod}A}$	$(\mu m^3 \cdot hr)^{-1}$
$k_{\text{leak}B}$	$0.02*k_{\text{prod}B}$	$(\mu m^3 \cdot hr)^{-1}$

Table 3.2: Table of revised parameters for uneven flipping to better match experimental data. IntA was set to be less efficient in flipping, and leakiness was added.

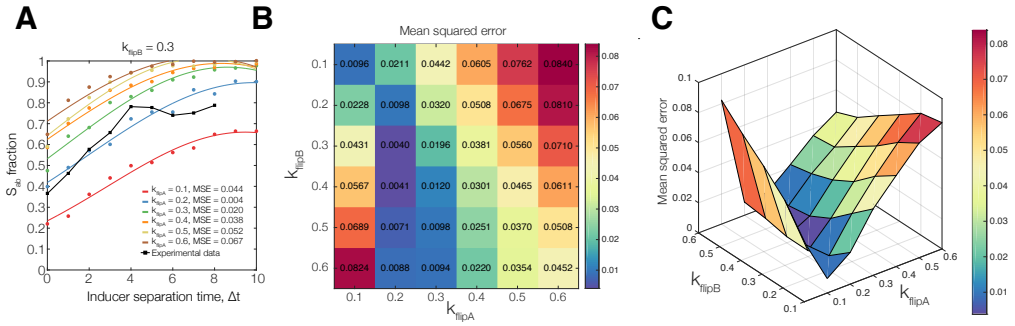


Figure ED-S13: Fitting model parameters for Figure 5C. In stochastic simulations, the flipping efficiency parameters for both integrases, $k_{\text{flip}A,B}$, were varied from 0.1 to 0.6 hr^{-1} for E_{ab} (N = 500 cell trajectories). Leaky basal expression of the integrases were held constant based on experimentally measured values (k_{leak_a} = 1% of $k_{\text{prod}A}$, k_{leak_b} = 2% of $k_{\text{prod}B}$). A) Simulation results for each set of $k_{\text{flip}A,B}$ parameters were fit to a one-term Gaussian function (MATLAB, *fit(x,y,'gauss1')*). Mean squared error (MSE) was calculated by comparing the fitted curves to experimental data from Figure 4C (MATLAB, *goodnessoffit(reference,model)*). This graph shows fits from varying $k_{\text{flip}A}$ for constant $k_{\text{flip}B} = 0.3\text{hr}^{-1}$. Experimental data is shown in black. B) Heatmap showing MSE values for combinations of $k_{\text{flip}A,B}$ parameters. Lower MSE values indicate a better fit. Best fit is for $k_{\text{flip}B} = 0.3\text{hr}^{-1}$, $k_{\text{flip}A} = 0.2\text{hr}^{-1}$. C) Surface plot showing MSE values for combinations of $k_{\text{flip}A,B}$. Lower MSE values indicate a better fit.

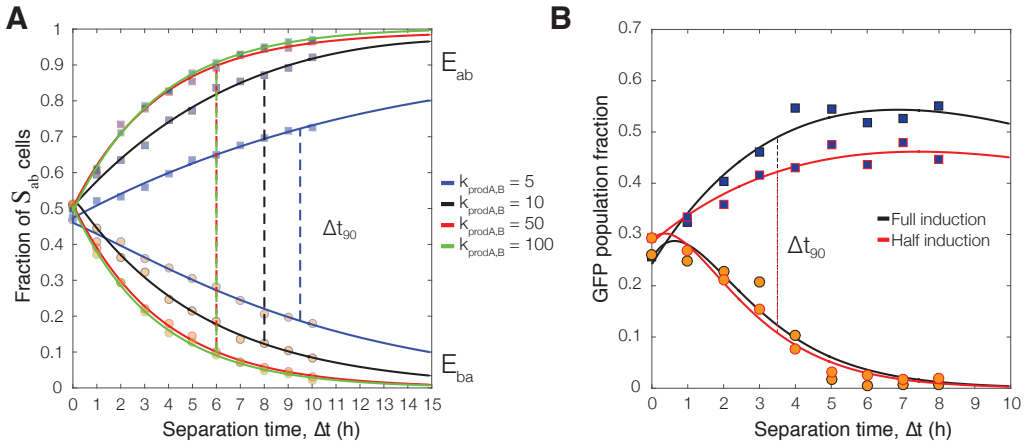


Figure ED-S14: Varying protein production rates to tune the Δt_{90} limit. A) Simulation results for varying $k_{prodA,B}$ from 5 to 100 $(\mu m^3 \cdot hr)^{-1}$(N = 2500 per population). Δt_{90} is the limit with which S_{ab} population fractions can be used to resolve unique Δt values, and therefore determines overall system sensitivity to inputs. Based on simulation results, Δt_{90} is inversely proportional to protein production rate. Curve fits were generated for each set of simulated populations (MATLAB, 2-term exponential fit) in order to find Δt_{90}. B) Experimental results show lower Δt_{90} at half induction of integrases. Protein production rate was modulated by reducing the concentrations of the inducers. We compared population-level responses with full inducer concentrations (ara: 0.01%/vol, aTc: 200ng/ml) and half inducer concentrations (ara: 0.005%/vol, aTc: 100ng/ml). The data was fit to a 2-term exponential function (MATLAB, 2-term exponential fit) and the Δt_{90} limit was estimated based on the fitted curve. The Δt values are consistent with being in the saturation regime of integrase production.

Simulation results suggest that the Δt_{90} detection limit can be tuned by increasing or decreasing the overall production rate k_{prod*} ($*$ = A or B) (Appendix Fig. ED-S14). In Figure 4C, the Δt_{90} limit was $\sim$ 4 hours, meaning that within the 0 – 4 hour window, S_{ab} population fraction can be used to uniquely determine Δt. Outside of this window, the only assertion that can be made is that $\Delta t > 5$ hours. *In silico*, we see that the rate of protein production is inversely proportional to the Δt_{90} detection limit (Appendix Fig. ED-S14A). When $k_{\mathrm{prodA,B}}$ is high, integrase molecules accumulate faster, increasing the probability of DNA flipping, and thus causing the S_{ab} population fraction to saturate at lower Δt values. However, within that smaller time window, S_{ab} fractions would also be measurably different at much smaller intervals, and so Δt could be resolved with much higher resolution. When protein production is slow, the stochastic DNA recombination events happen less

frequently, resulting in a population that is more sensitive to inputs for a longer period of time (high Δt_{90}), but has lower resolution overall since the population fractions are not changing as quickly. These simulation results were compared to some preliminary experimental data in which lower production rates for intA and intB were approximated by halving the inducer concentrations for both **a** and **b** (Appendix Fig. ED-S14B, ED-S15). Δt_{90} was estimated by fitting curves to the experimental data to determine maximum S_{ab} (MATLAB, 2-term exponential fit). When inducer concentrations were halved (ara: 0.005%/vol, aTc: 100ng/ml), we see that the Δt_{90} is the same as before, so even with half induction, we are still in the saturation regime of integrase production.

Varying protein production rates more accurately is something we would like to pursue further. We limited the scope of this study to a single concentration of inducer and Δt_{90} such that we could fully understand the information that can be gained from other states in the system.

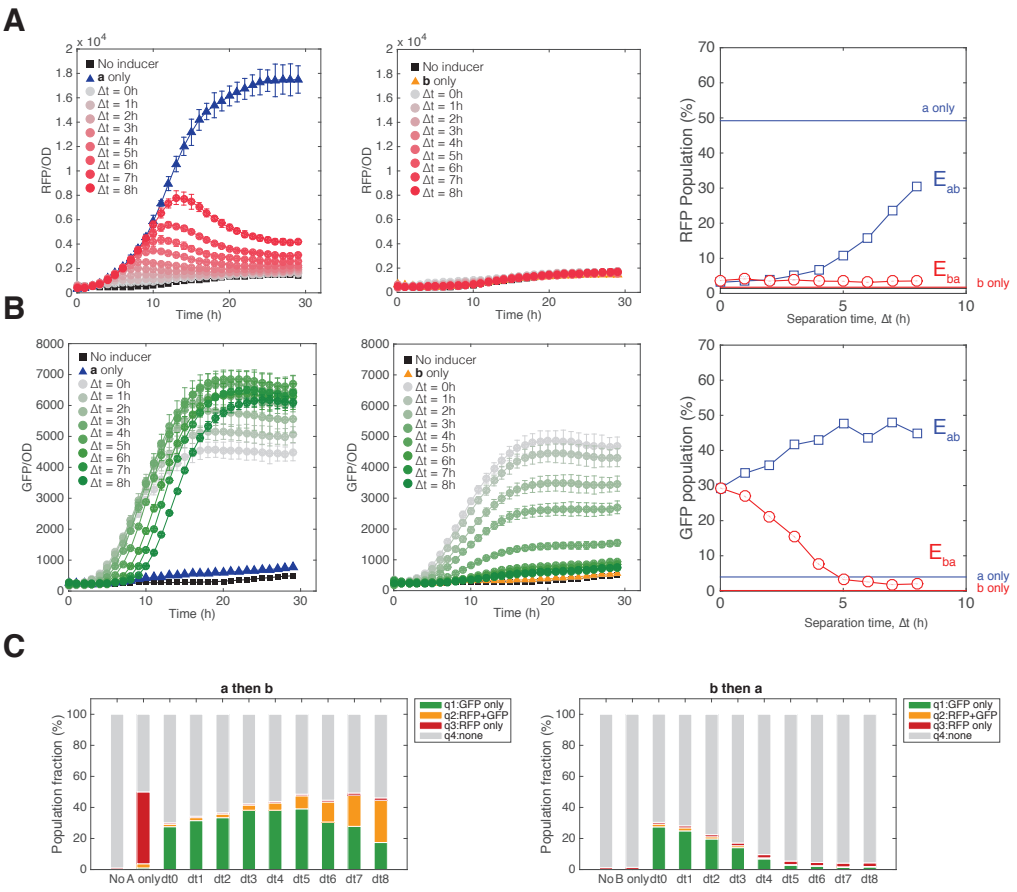


Figure ED-S15: Varying protein production rates, timecourse. Here, we have used half the normal inducer concentrations to test the effects of lower protein production rates. Concentrations of **a** and **b** are 0.005%/vol arabinose and 100ng/ml aTc. A) RFP fluorescence over time for E_{ab}(left), E_{ba}(center), and endpoint population fractions as measured by flow cytometry (right). B) GFP fluorescence over time for E_{ab}(left), E_{ba}(center), and endpoint population fractions as measured by flow cytometry (right). C) Population distributions gated by quadrants. Overall population behavior was the same as full inducers, but response was more graded with increasing Δt.

Deducing inducer pulse width: simulations

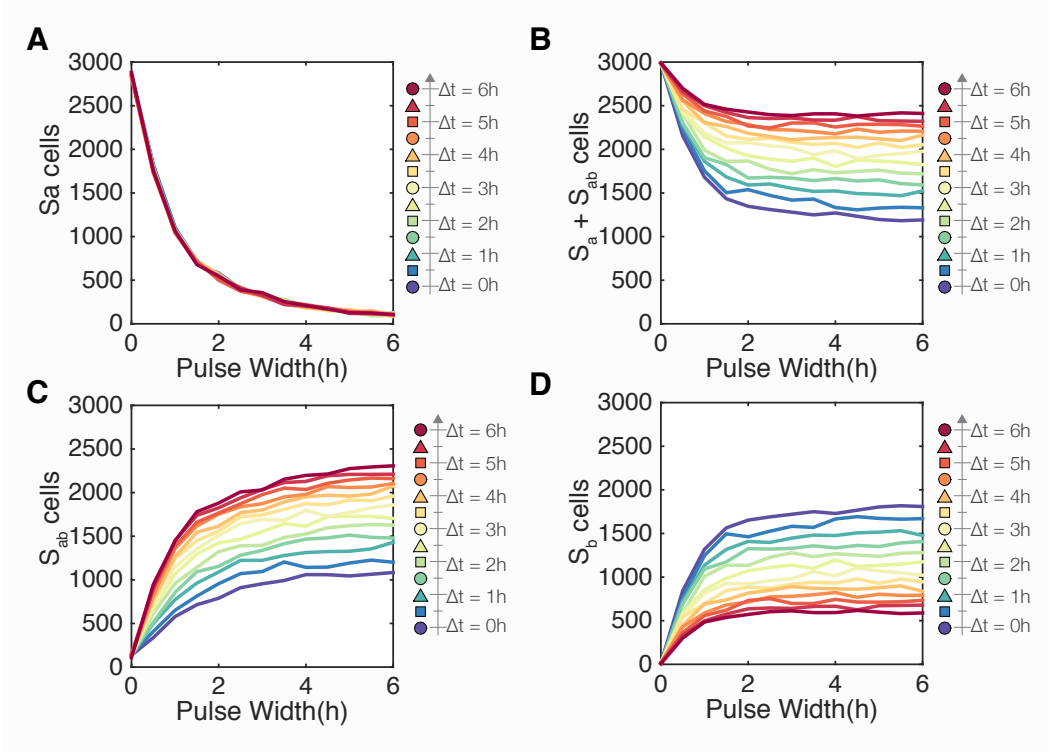


Figure ED-S16: Deducing pulse width, additional states with varying Δt, PW_b, N = 3000 cells. A) S_a cells decrease as function of PW_b. S_a fraction is independent of Δt. B) The sum of S_a + S_{ab} is the fraction of cells that see ***a*** *first*, and this increases with Δt and PW_b. C) The number of S_{ab} cells increases with Δt and PW_b. D) The number of S_b cells decreases with Δt but increases with PW_b.

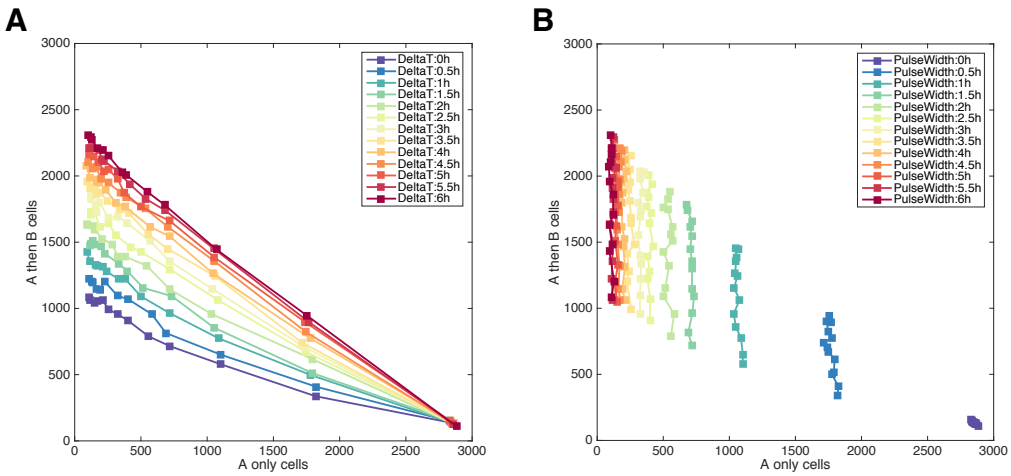


Figure ED-S17: Unique populations for different combinations of Δt and PW_b (Fig.7). Each point represents a simulation with 3000 cells. A) Lines represent increasing Δt values. B) Lines represent increasing PW_b values.

Deducing inducer pulse width: experimental

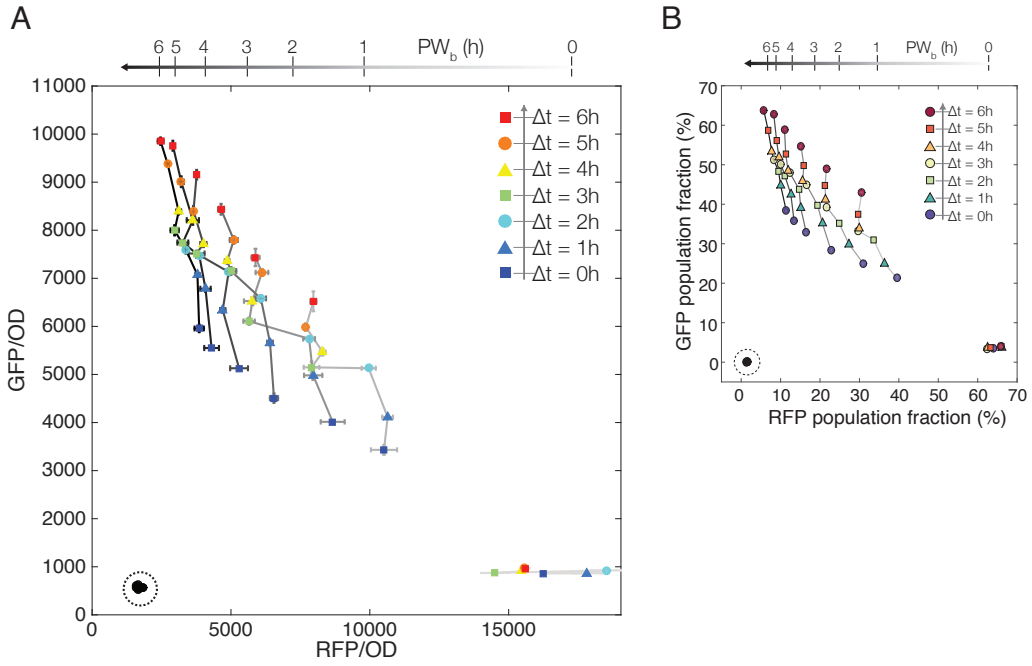


Figure ED-S18: Fluorescence measurements vs flow cytometry for Figure 7C. Experimental pulse modulated populations were sampled and grown in triplicate in 96-well microplates. Δt and PW_b were varied from 0 – 6 hours. Dotted circle shows control populations with no inducer exposure. A) Fluorescence measurements show standard error between triplicates for RFP and GFP fluorescence. Differences in single cell fluorescence expression results in skewing of population coordinates. B) Triplicates were pooled for flow cytometry measurements (Same data as Figure 7C).

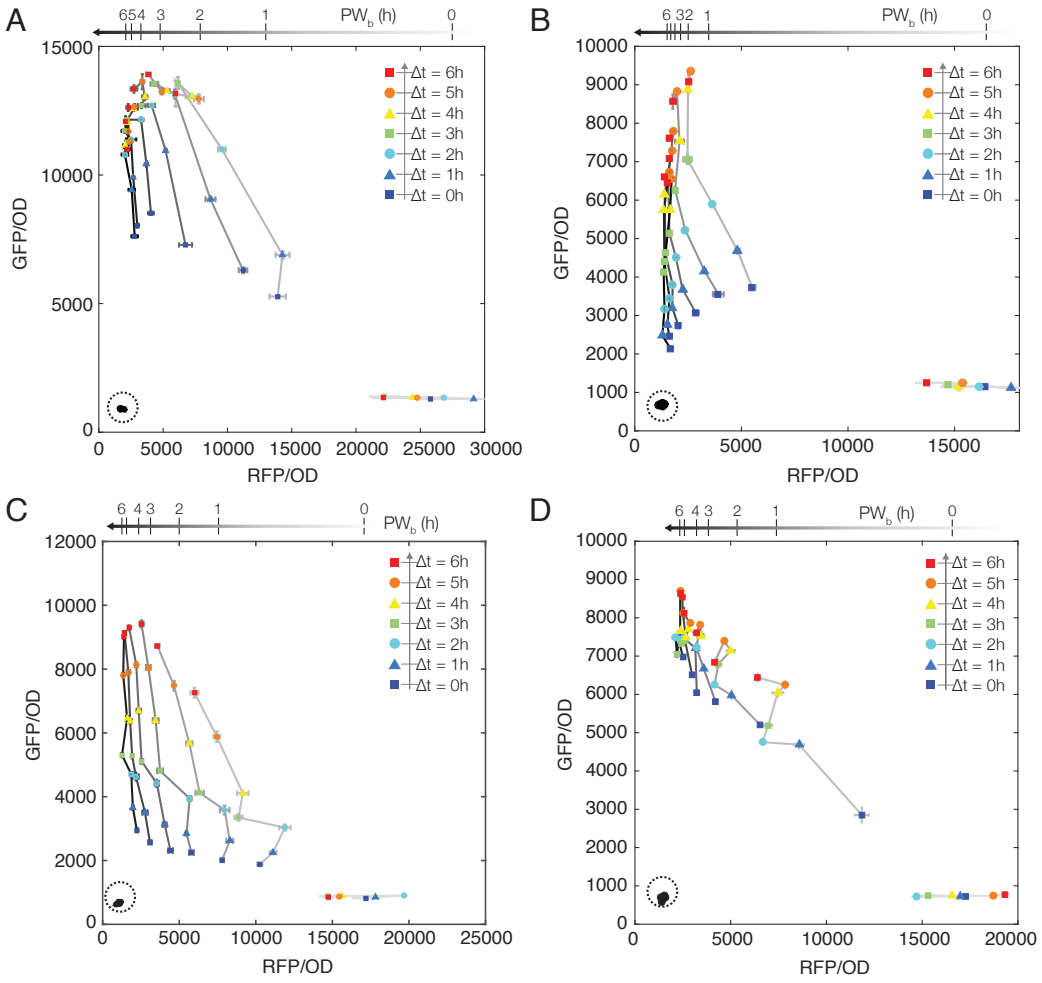


Figure ED-S19: Additional replicates of pulse experiments show final populations are sensitive to starting cell density. Δt and PW_b were varied from 0 – 6 hours. Dotted circle shows control populations with no inducer exposure. These data were not used for the main manuscript. While overall trends were consistent, bulk fluorescence of final populations showed skewing, which we hypothesize is a function of starting OD combined with differences in growth rates and variation in single cell fluorescence. While starting OD is quite low, these differences can become magnified over the 12 hour time frame with which the cultures are being sampled (e.g. $\Delta t = 6\,\text{h} + PW_b = 6$ h results in the last population getting sampled 12 hours after start of experiment). All populations were then grown for an additional 24 h after sampling. Optimal starting OD was determined to be 0.06 – 0.08 (data used for Fig. 7 had starting OD of 0.072–0.074, Fig. ED-S18). A) Bulk fluorescence (plate reader) of pulse modulated populations sampled from batch cultures that started at OD 0.083–0.1. B) Bulk fluorescence of populations sampled from cultures started at OD 0.12–0.14. C) Fluorescence of populations sampled from cultures started at OD 0.066 – 0.073. D) Fluorescence of populations sampled from cultures started at OD 0.060 – 0.065.

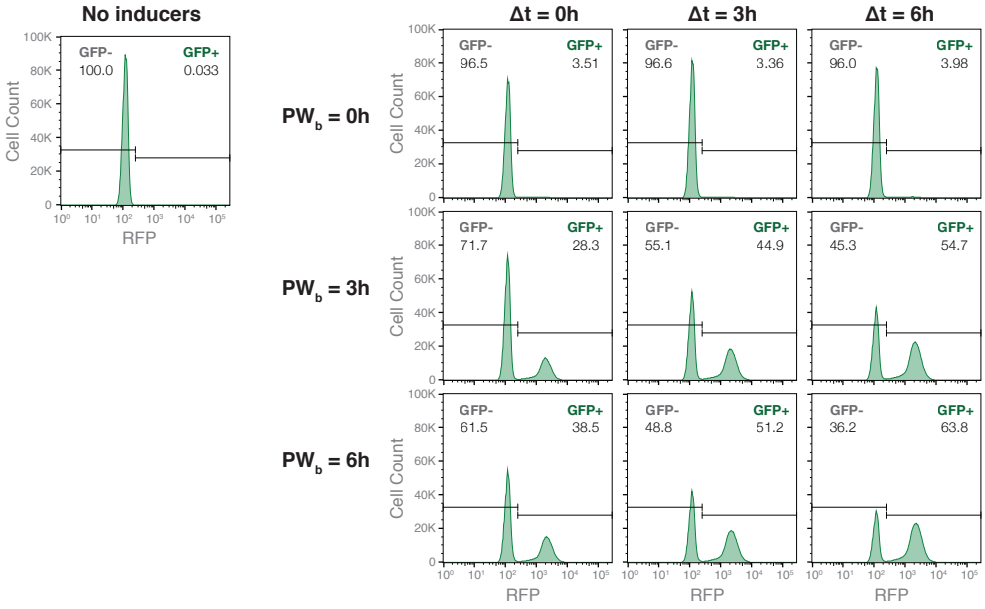


Figure ED-S20: Flow cytometry data for Figure 7BC. GFP histograms, selected panels. ~ 1 million cells were measured for each population.

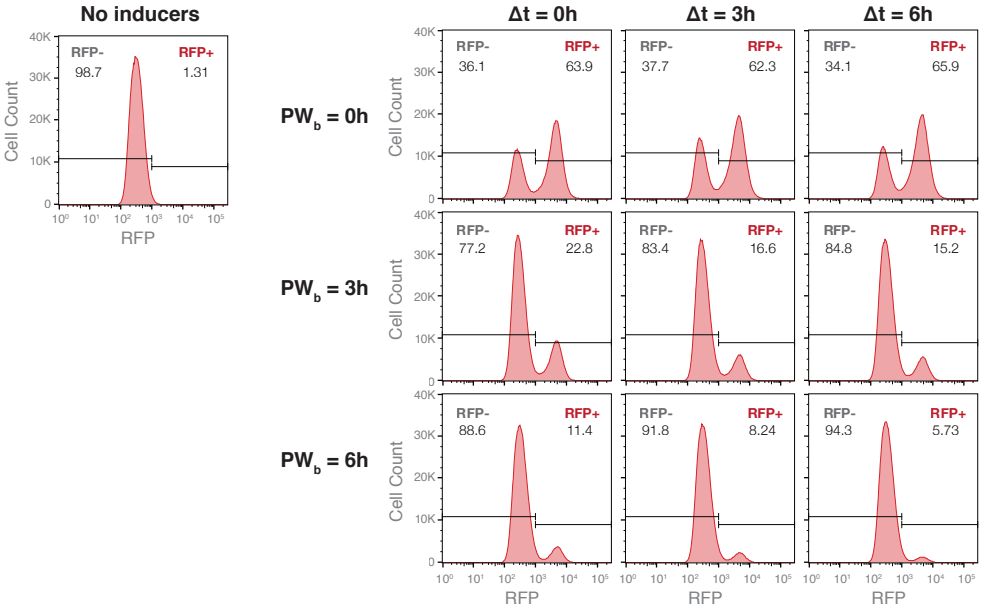


Figure ED-S21: Flow cytometry data for Figure 7BC. RFP histograms, selected panels. ~ 1 million cells were measured for each population.

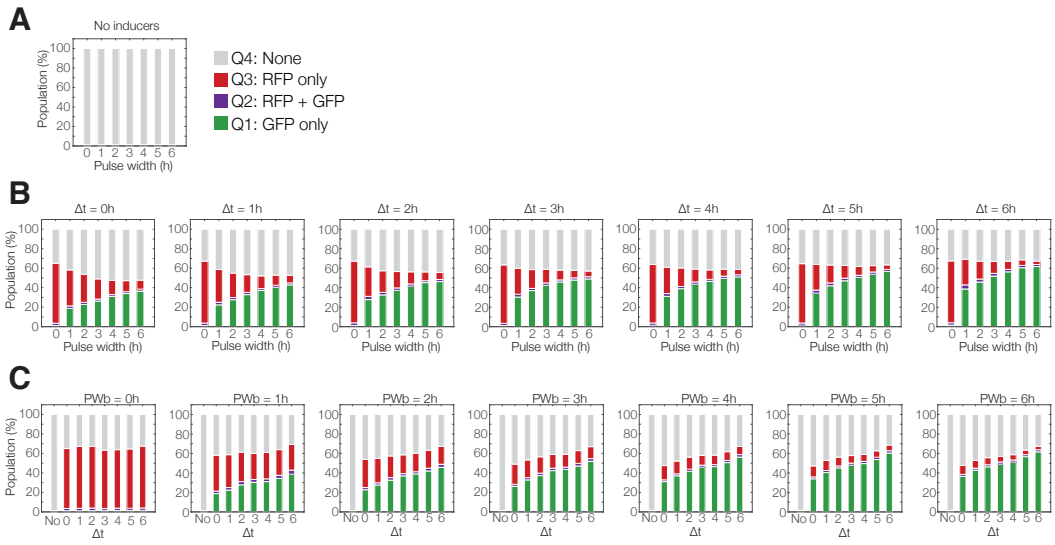


Figure ED-S22: Population quadrants for Figure 7BC. $\sim 10^6$ cells were measured for each population. Populations are split into quadrants Q1 (GFP only, S_{ab}), Q2 (GFP + RFP, S_{ab}), Q3 (RFP only, S_a), and Q4 (non-fluorescent, S_o + S_b.) The Q2 population is $< 3\%$ for all conditions. A) Cultures that were incubated without any inducer exposure remained non-fluorescent. B) Population distributions as they changed with increasing PW_b. Individual subplots (left to right) are increasing Δt. C) Population distributions as they changed with increasing Δt. Individual subplots (left to right) are increasing PW_b.

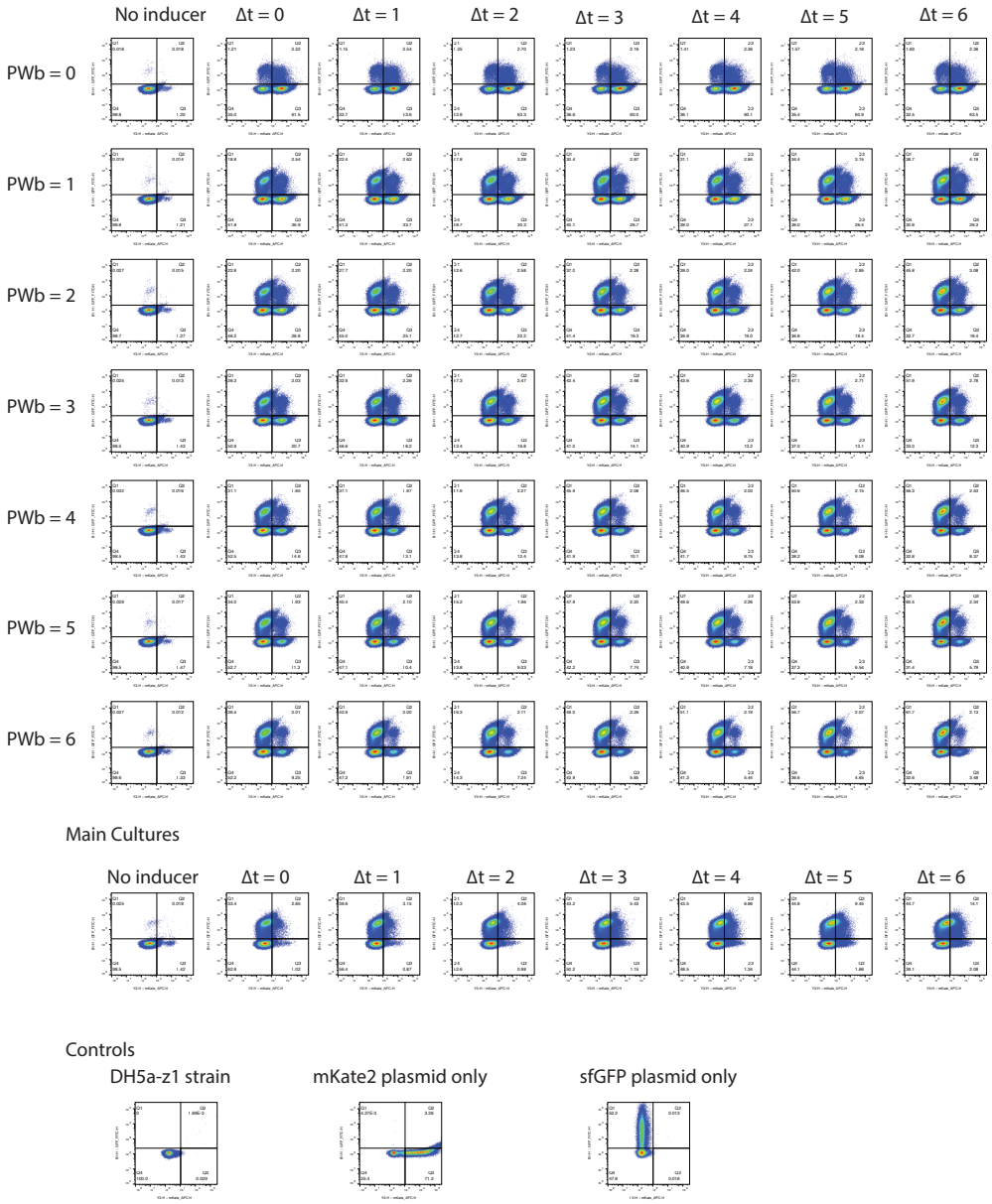


Figure ED-S23: Complete flow cytometry data for Figure 7BC, RFP vs GFP. ~ 1 million cells per population.

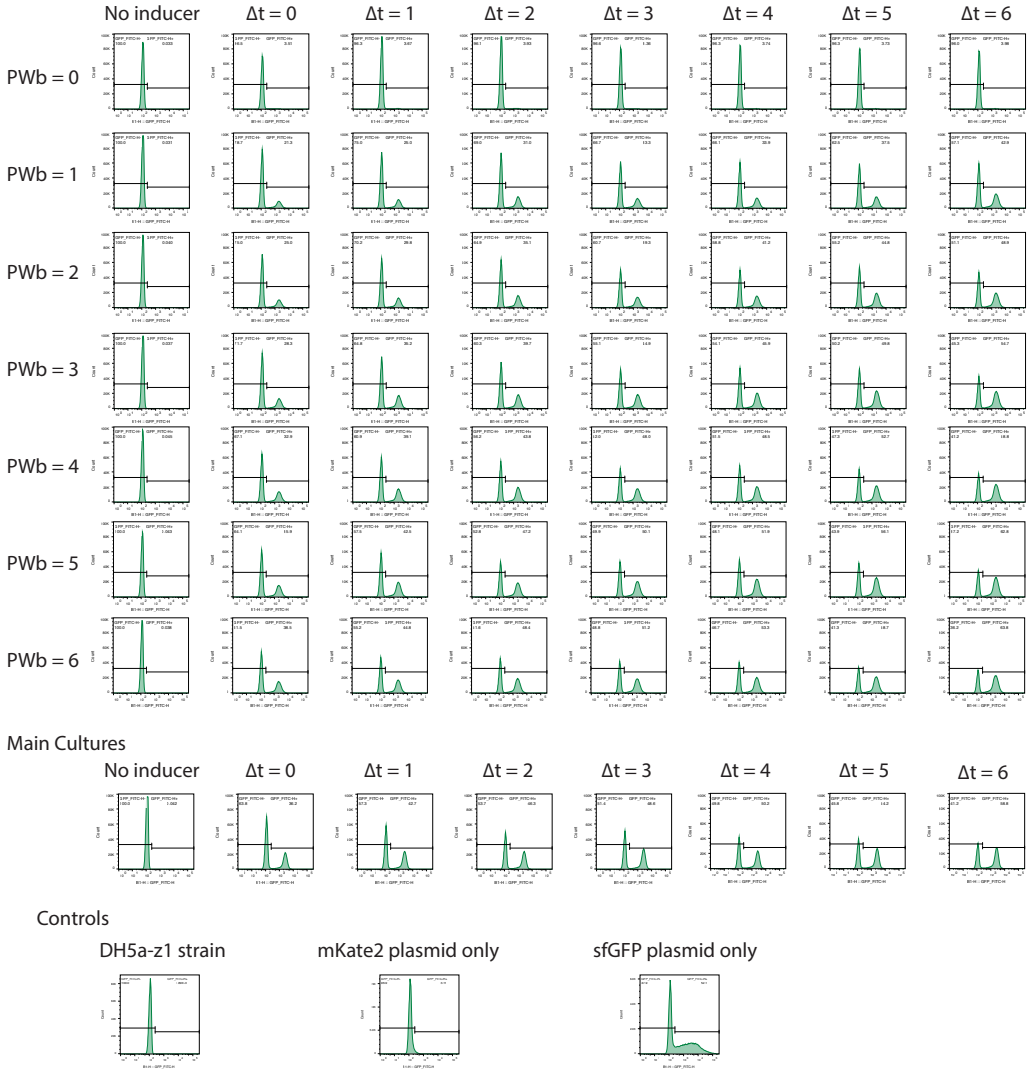


Figure ED-S24: Complete flow cytometry data for Figure 7BC, GFP histograms. ~ 1 million cells per population.

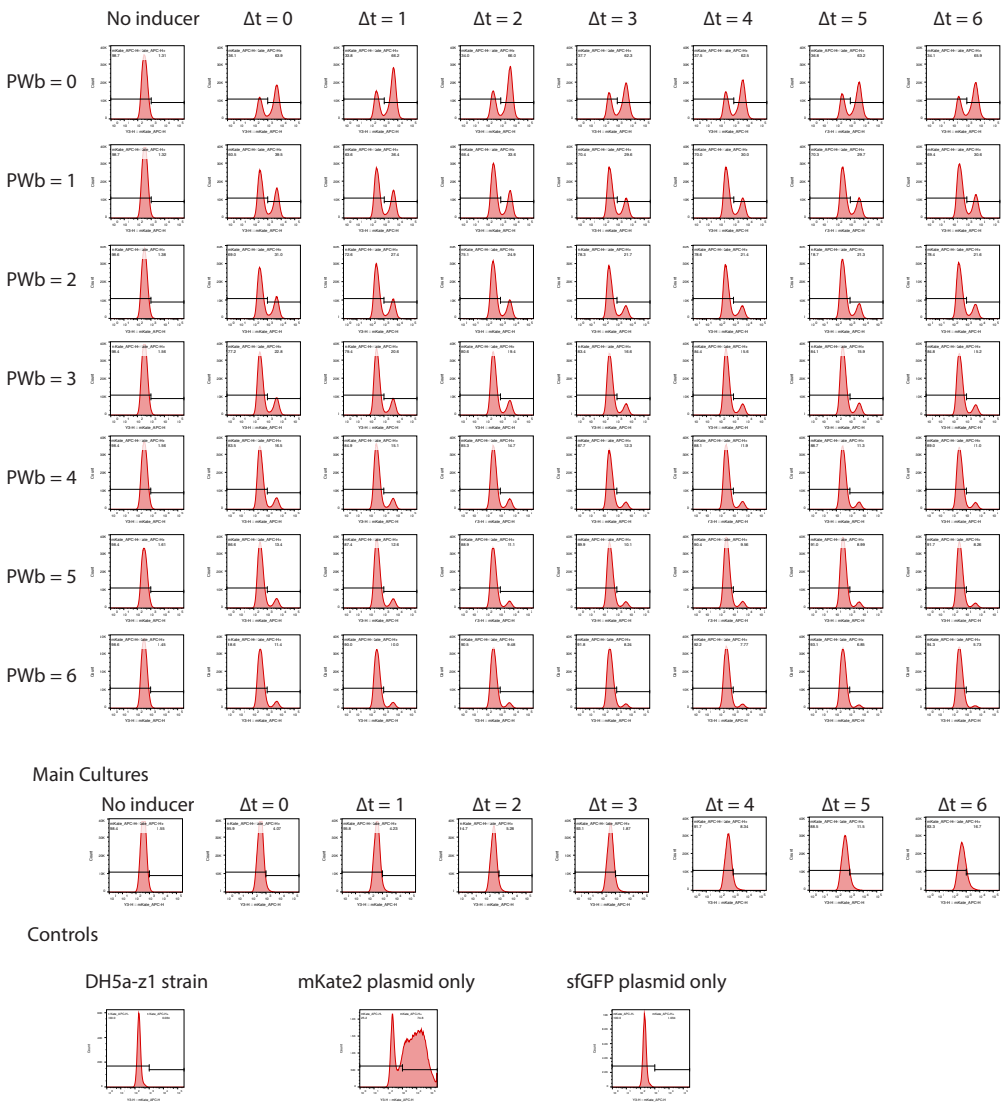


Figure ED-S25: Complete flow cytometry data for Figure 7BC, RFP histograms. ~ 1 million cells per population.

Single colony analysis of pulse modulated populations

Since a significant fraction of all experimental populations from pulse experiments consisted of at least 30% non-fluorescent cells, we wished to determine whether these colonies were S_o or S_b state cells. Five experimental populations from the same experimental cultures as Fig. 7BC were diluted 1:10,000 and plated onto LB agar plates with no inducers (Figure ED-S26A). We selected populations from the corners of the experimental matrix to get the widest range of results (Δt = 0, 6, PW_b = 0, 6). 60±10 colonies were re-streaked onto a new agar plate with no inducers (Figure ED-S26B). The number of RFP (S_a), green(S_{ab}), and non-fluorescent (S_o, S_b) colonies were counted. In Figure ED-S26C, we see that population distributions from single cell counts closely matched overall flow cytometry data.

We chose the first 8 non-fluorescent colonies from each population for detailed analysis (Figure ED-S26D). We colony-PCR amplified the genomically-integrated DNA memory cassette for each colony (S_o,S_a,S_{ab} = 404bp, S_b = 220bp). We also included controls from the original strain (S_o), a RFP fluorescent colony (S_a) and a GFP fluorescent colony (S_{ab}). We then purified each PCR-amplified product and sequence confirmed all products (Sequencing primers, ED_seq_1F/ED_seq_1R). The 8 non-fluorescent colonies were also re-streaked on LB agar + 0.01% arabinose plate to separate S_o versus S_b cells (Figure ED-S26E). When exposed to fresh arabinose, only S_o state cells should turn red. The results from re-streaking onto inducer **a** matched PCR and sequencing results exactly.

Using S_o:S_b ratios derived from colony counts (Figure ED-S26F), the original non-fluorescent distributions shown in Figure ED-S26C were revised. Our random sample for the *no inducer* population revealed no leaky expression, though flow analysis revealed about 1–2% leaky fluorescent expression. For the PW_b = 0h populations, these populations never encountered inducer **b**, and so have similar S_a population fractions. Of the remaining cells for the two PW_b = 0h cases, we see some intB leaky expression, resulting in non-zero S_{ab} and S_b fractions for both. While it is not surprising that higher exposure to inducer **b** (PW_b = 6h) would result in mostly S_b cells, it was surprising that some fraction of S_o persisted over the entire 40 h experiment. We conclude that the integrase controller plasmid has minimal leaky expression, and that over-representation of non-fluorescent states is likely due to a growth advantage over fluorescent states. Furthermore, these data show that overall integrase flipping (S_a + S_{ab} + S_b) is about 90% efficient with about 10% persistent S_o population which can be utilized for future responses.

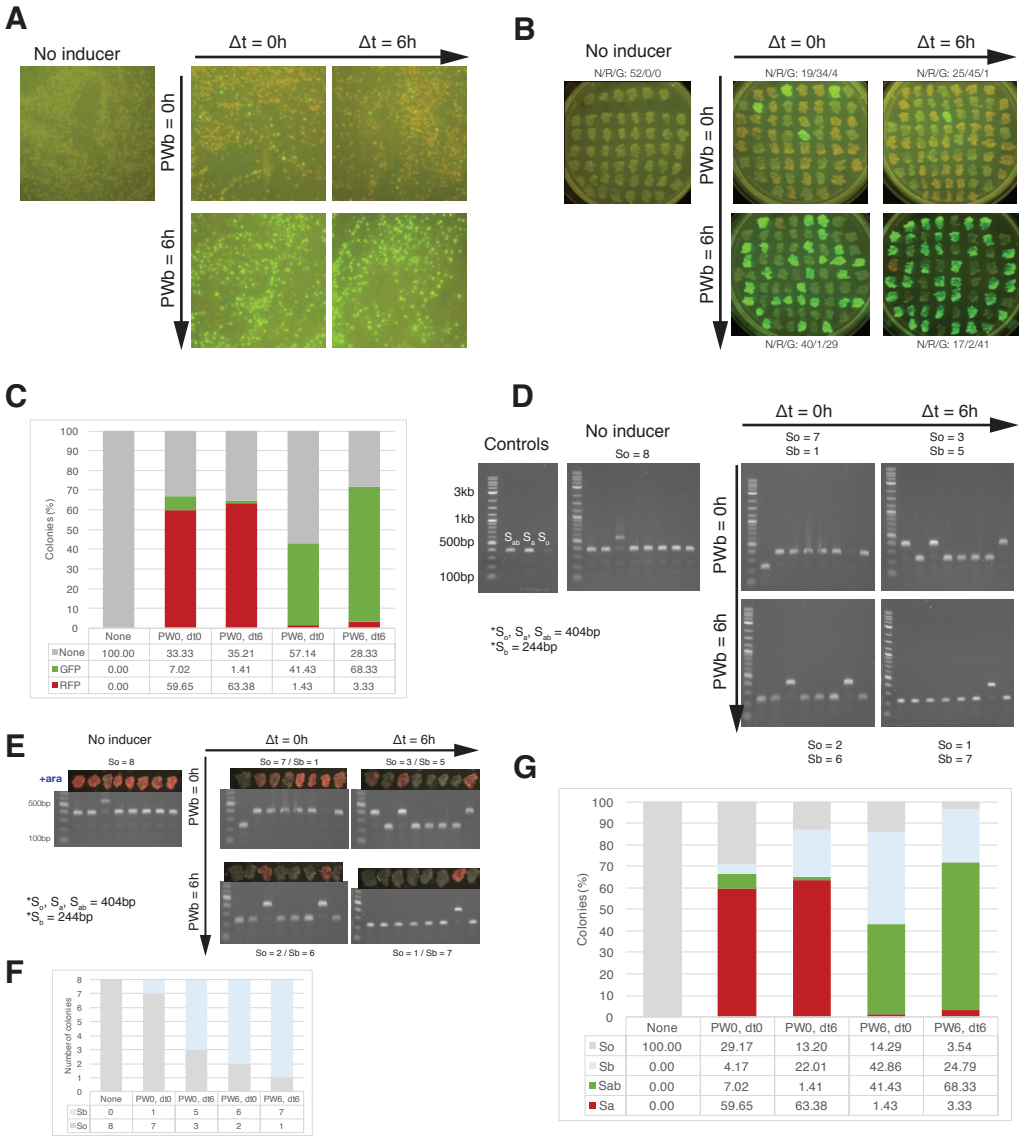


Figure ED-S26: Single colony analysis of pulse modulated populations to determine genetic state. A) Five experimental populations from the same experimental cultures as Fig. 7BC were diluted 1:10,000 and plated onto LB agar plates with no inducers. B) 60 ± 10 individual colonies were re-streaked onto a new agar plate with no inducers. The number of RFP (S_a), green(S_{ab}), and non-fluorescent (S_o, S_b) colonies were counted. C) Population distributions from single cell counts closely matched overall flow cytometry data for entire population. D) We used colony PCR to amplify the genomically-integrated DNA memory cassette from 8 non-fluorescent colonies for each population (S_o,S_a,S_{ab} = 404bp, S_b = 220bp). Controls are from the original strain (S_o), a highly RFP fluorescent colony (S_a) and a highly GFP fluorescent colony (S_{ab}). E) The 8 non-fluorescent colonies were also re-streaked on LB agar + 0.01% ara plate to test whether only S_o state cells would turn red. Colonies matched PCR and sequencing results exactly. F) Colony counts of S_o versus S_b cells for the non-fluorescent fraction of each population. G) Revised distributions based on S_o versus S_b population ratios derived from panel F.

Model exploration of S_a dependence of Δt

Though our model predicted complete independence of S_a state from Δt separation times (Appendix Fig. ED-S27A), our experimental outcome showed a small linear dependence (Figure 7B,top), where lower Δt values resulted in higher S_a population fractions.

This dependence on Δt resulted in a right-to-left slant in RFP population fractions for any given PW_b value that was not predicted by our model (Figure 7A versus 7C). Upon examination of our model, we believe this is the result of unequal reaction rates during the $\mathrm{S_o} \xrightarrow{\alpha_1} \mathrm{S_b}$ transition compared to $\mathrm{S_a} \xrightarrow{\alpha_3} \mathrm{S_{ab}}$. In our model we had assumed that these rates were equal, since both are mediated by intB:

$$\alpha_1 = k_{\mathrm{flip}B} f(\mathrm{Int}_B), \tag{3.5}$$

$$\alpha_3 = k_{\mathrm{flip}B} f(\mathrm{Int}_B), \tag{3.6}$$

where $f(\mathrm{Int}_B)$ is the tetramerization term:

$$f(\mathrm{Int}_B) := k_{\mathrm{flip}B}\left(\frac{\mathrm{Int}_B(\mathrm{Int}_B-1)(\mathrm{Int}_B-2)(\mathrm{Int}_B-3)}{K_{\mathrm{d}B}^4+K_{\mathrm{d}B}^3\mathrm{Int}_B+K_{\mathrm{d}B}^2\mathrm{Int}_B(\mathrm{Int}_B-1)+K_{\mathrm{d}B}\mathrm{Int}_B(\mathrm{Int}_B-1)(\mathrm{Int}_B-2)+\mathrm{Int}_B(\mathrm{Int}_B-1)(\mathrm{Int}_B-2)(\mathrm{Int}_B-3)}\right), \tag{3.7}$$

where Int_B is integrase B concentration, $K_{\mathrm{d}B}$ is the dissociation constant, and $k_{\mathrm{flip}B}$ is the rate of flipping if the tetramer is formed.

We had made this assumption because the DNA attachment sites attB and attP are the same for both transitions, and so binding kinetics should be the same. Changing other parameters such as $K_{\mathrm{d}B}$ or $K_{\mathrm{flip}B}$ did not regenerate slanting behavior since these parameters were universal for both α_1 and α_3. However, $\mathrm{S_o} \longrightarrow \mathrm{S_b}$ is an excision reaction rather than recombination, and so the physical looping of the DNA could have different kinetics (Appendix Fig. ED-S27D).

Only when we consider the excision reaction to be slower than the other two recombination reactions,

$$\alpha_1 < \alpha_2 = \alpha_3,$$

were we able to see this effect of RFP drifting with increasing Δt.

In Appendix Figure ED-S27E and F, simulation results for $\alpha_1 = 0.6\alpha_2 = 0.6\alpha_3$ show the separation of S_a curve by Δt, and nonlinear RFP with increasing Δt in the RFP

vs GFP plot. When the transition rate is decreased even more ($\alpha_1 = 0.4\alpha_2 = 0.6\alpha_3$, Appendix Figure ED-S27G, H), the slant increases even more.

Intuitively, the reason slower $S_o \rightarrow S_b$ transition rates would cause this effect is because at lower separation times the dominating cell state is S_o, and so the predominant reactions are $\mathrm{S_o} \xrightarrow{\alpha_1} \mathrm{S_b}$ versus $\mathrm{S_o} \xrightarrow{\alpha_2} \mathrm{S_a}$. In the case of equal reaction rates, 50% goes to S_b, and 50% goes to S_a. $\mathrm{S_a} \xrightarrow{\alpha_3} \mathrm{S_{ab}}$ can only occur after S_a cells appear, and so cannot occur until after some delay. If $\alpha_1 < \alpha_2$, however, then the population split will be unequal as S_o cells are more likely to transition to S_a over S_b.

For large Δt, the dominating cell state is S_a, and so the predominant reactions are $\mathrm{S_a} \xrightarrow{\alpha_3} \mathrm{S_{ab}}$. In this case, few S_o remain, so α_1 and α_2 become less relevant as α_3 converts S_a cells into S_{ab} in a pulse width dependent manner.

So, if we consider the $\mathrm{S_a} \xrightarrow{\alpha_3} \mathrm{S_{ab}}$ conversion rate to be the baseline, then $\mathrm{S_o} \xrightarrow{\alpha_2} \mathrm{S_a}$ is generating a *higher* proportion of S_a cells than predicted at low Δt because $\alpha_2 > \alpha 1$.

Uneven transition rates are not unsurprising for experimental systems, however, changing PW_b is still the dominating determinant of cell fractions. When designing future systems it may be relevant to characterize switching rates. Despite unequal intB transition rates, each combination of PW_b and Δt still maps to unique (S_a, S_{ab}) fractional coordinates, even though S_a values are not unique for higher PW_b.

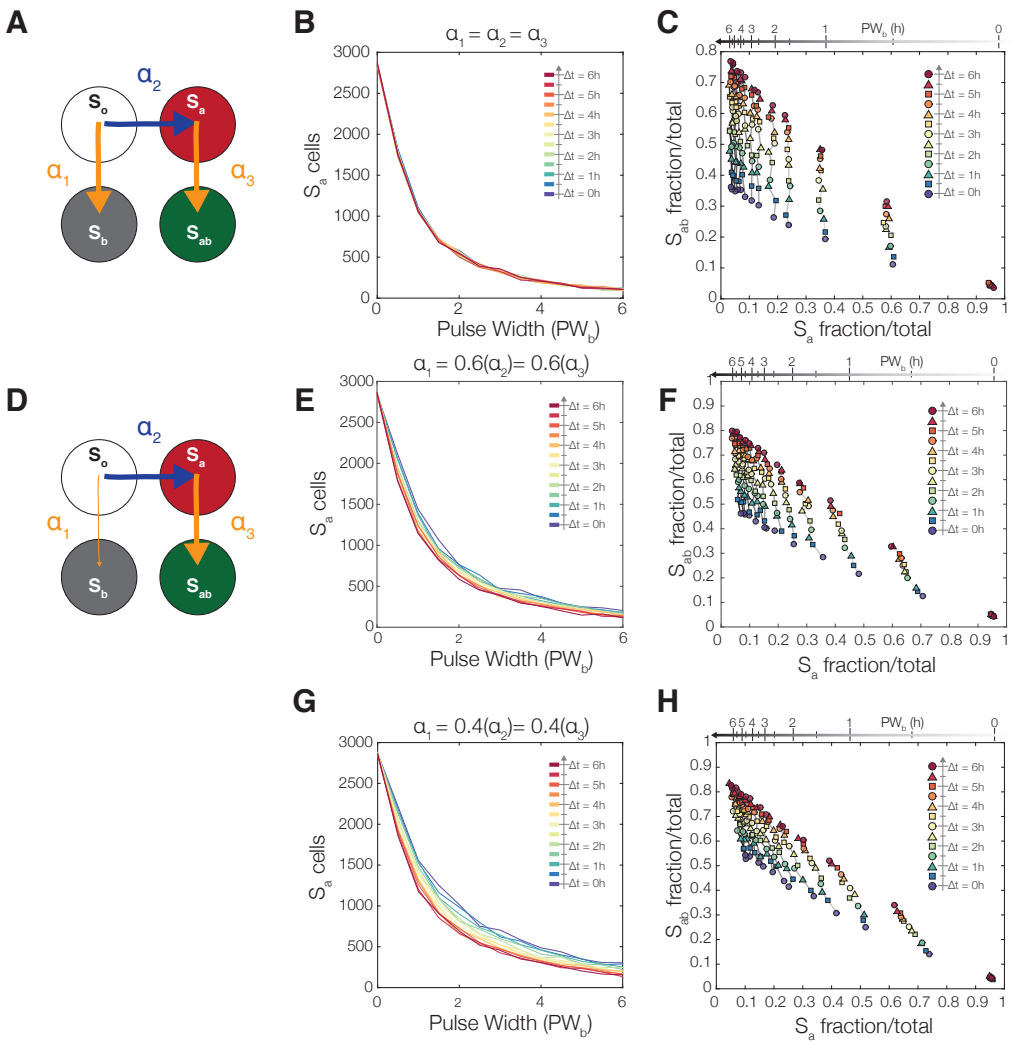
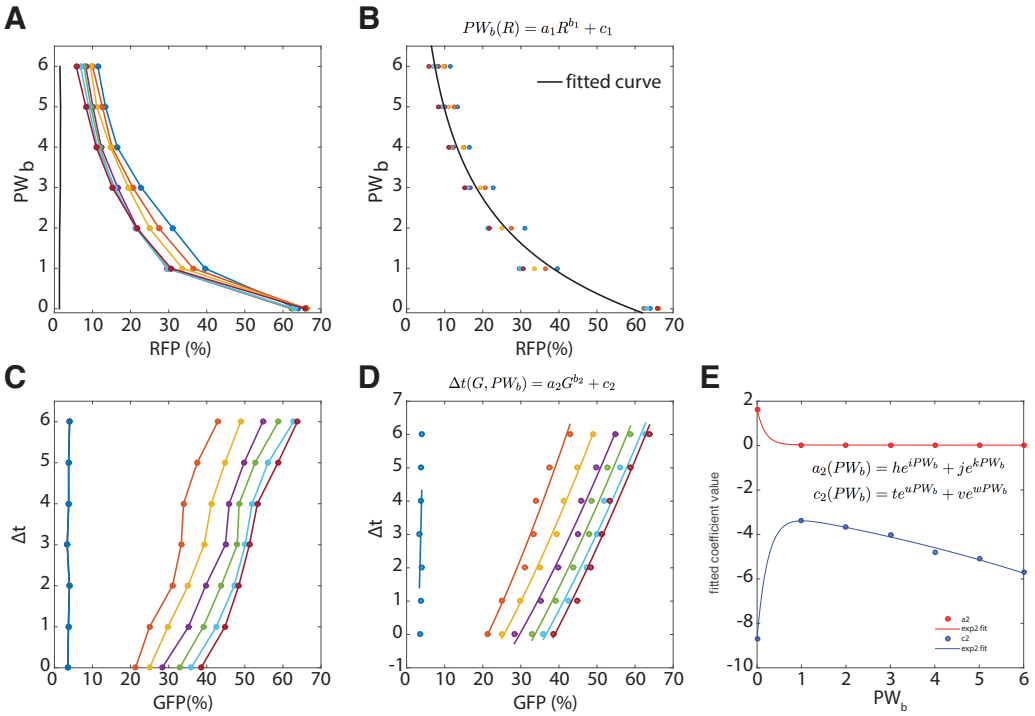


Figure ED-S27: Simulations with unequal intB transition rates. A) Initial model parameters assume equal transition probabilities for α_1, α_2, and α_3. B) Model simulations showing total independence of S_a from Δt values. C) S_a versus S_b populations with constant S_a fractions for any given PW_b. D) We hypothesized that the excision reaction from $S_o \rightarrow S_b$ maybe be slower than $S_a \rightarrow S_{ab}$. E) S_a cell count as a function of pulse width (PW_b) with $\alpha_1 = 0.6\alpha_2 = 0.6\alpha_3$. Δt curves no longer completely overlap and low Δt values result in higher S_a fractions. F) S_a fraction versus S_{ab} fraction shows right to left slanting behavior observed in experimental results (Figure 7B,7C). G) S_a cell count as a function of pulse width (PW_b) with even slower $S_o \rightarrow S_b$ transition rate ($\alpha_1 = 0.4\alpha_2 = 0.6\alpha_3$). H) S_a fraction versus S_{ab} fraction shows right to left slanting behavior.

Practical use and calibration



F

Parameter	Fitted Value
a_1	24.3
b_1	-0.258
c_1	-8.483
b_2	1.5
h	1.608
i	-5.587
j	0.029
k	-0.046
t	-5.742
u	-4.115
v	-2.944
w	0.1113

Figure ED-S28: Fitting experimental data for PW_b, Δt. A) RFP population fractions from experimental data plotted with known pulse widths on the y-axis. B) Curve fit to determine PW_b dependence on RFP. A power fit (general form $PW_b(R) = a_1 R^{b_1} + c_1$) to the data generates parameters $a_1 = 24.3, b_1 = -0.25, c_1 = -8.4$. C) GFP measurements from experimental data plotted with known Δt values. D) Curve fits to determine Δt dependence on GFP and RFP population fractions. Data from each value of PW_b is fitted to a different curve of general form $\Delta t(G, PW_b) = a_2 G^{b_2} + c_2$. Parameter b_2 was separately fit to be 1.5. Parameters a_2 and c_2 are functions of PW_b. E) Parameters a_2 and c_2 are then fitted to their own exponential curves (general forms $a_2(PW_b) = he^{iPW_b} + je^{kPW_b}$ and $c_2(PW_b) = te^{uPW_b} + ve^{wPW_b}$) to determine dependence on PW_b. Fitted parameters are: h = 1.6, i = -5.5, j = 0.029, k = -0.046, t = -5.7, u = -4.1, v = -2.9, w = 0.1. F) A table of all the fitted parameters.

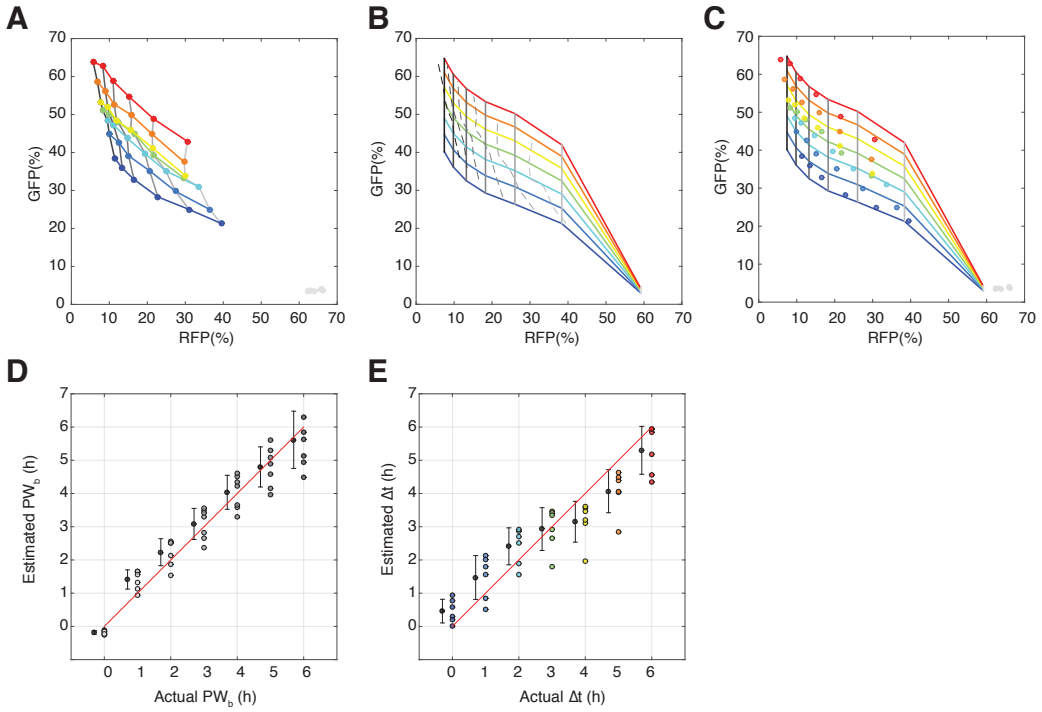


Figure ED-S29: Resolution for determining PW_b, Δt from population distributions. A) Connected experimental values create a mesh. Area between meshlines represent the accuracy with which values of PW_b and Δt can be determined from experimentally derived RFP and GFP population fractions. RFP fractions for pulse widths from 0 – 6 hours are well separated but with decreasing resolution as PW_b increases. GFP fraction is dependent on both RFP and GFP and is also well separated, with the exception of $\Delta t = 2, 3$ hours. B) Mesh generated from the curve fits for $PW_b(R), \Delta t(G)$ discussed in Figure ED-S28. Experimental values are colored by Δt value to show fit. C) Mesh generated from the curve fits for $PW_b(R), \Delta t(G)$ discussed in Figure ED-S28. Experimental values are colored by PW_b value to show fit. D) Estimated PW_b values were generated from experimental RFP population fraction (%) using the fitted equation for $PW_b(R)$. The estimated values were plotted against the actual PW_b values of the experiment. The vertical gray bars show approximate spread in estimated values, the numbers above the bars indicate length of the bars (in hours). The variance in estimated values increases with higher PW_b. If there is no pulse, the resolution with which we can deduce that based on fluorescence is ±0.25 hours. If the pulse width is 3 hours or greater, our prediction capabilities decrease to a window of ±1 hour. For each actual PW_b values, estimated PW_b averages with ± 1 standard deviation are slightly offset on the x-axis for better comparison. E) Estimated versus actual values for Δt. Estimated Δt values are generated using the fitted curve for $\Delta t(G, PW_b)$. Variance in Δt predictions is more consistent than that for PW_b, with a resolution of ±0.25 hours for actual Δt from 0 – 3 hours, and an estimation window of ±0.5 hours for actual values between 4 – 6 hours. For each actual Δt values, predicted Δt averages with ± 1 standard deviation are slightly offset on the x-axis for better comparison.

Fitting equations and reference tables

Fitting of curves was done with experimental data from Figure 7C using the MATLAB curve fitting toolbox. Curves for $PW_b(R)$ and $\Delta t(G, PW_b)$ were fit to two-term power functions. Curves for the Δt coefficients $a_2(PW_b)$ and $c_2(PW_b)$ were fit to two-term exponential functions.

$$R = \text{RFP population (\%)} \tag{3.8}$$

$$G = \text{GFP population (\%)} \tag{3.9}$$

$$PW_b(R) = a_1 R^{b_1} + c_1 \tag{3.10}$$

$$\Delta t(G, PW_b) = a_2 G^{b_2} + c_2 \tag{3.11}$$

$$a_2(PW_b) = he^{iPW_b} + je^{kPW_b} \tag{3.12}$$

$$c_2(PW_b) = te^{uPW_b} + ve^{wPW_b} \tag{3.13}$$

Parameter	Fitted value
a_1	24.3
b_1	-0.258
c_1	-8.483
b_2	1.5
h	1.608
i	-5.587
j	0.029
k	-0.046
t	-5.742
u	-4.115
v	-2.944
w	0.1113

Table 3.3: Fitted parameters for $PW_b(R)$, $\Delta t(G, PW_b)$

Δt Prediction table

	GFP						
RFP	**0%**	**10%**	**20%**	**30%**	**40%**	**50%**	**60%**
0%	–	–	–	–	–	–	–
1%	0	0	0	0	0	0	0
10%	0	0	0	0	0.7	3.0	5.6
20%	0	0	0	0.2	2.5	5.0	7.9
30%	0	0	0	0.9	3.3	6.0	9.1
40%	0	0	0	2.9	6.3	10.2	14.5
50%	0	2.9	16.1	33.1	53.3	76.2	101.6
60%	0	52.6	∞	∞	∞	∞	∞

PW_b Prediction table

	GFP						
RFP	**0%**	**10%**	**20%**	**30%**	**40%**	**50%**	**60%**
0%	∞	∞	∞	∞	∞	∞	∞
1%	15.8	15.8	15.8	15.8	15.8	15.8	15.8
10%	4.9	4.9	4.9	4.9	4.9	4.9	4.9
20%	2.7	2.7	2.7	2.7	2.7	2.7	2.7
30%	1.6	1.6	1.6	1.6	1.6	1.6	1.6
40%	0.9	0.9	0.9	0.9	0.9	0.9	0.9
50%	0.4	0.4	0.4	0.4	0.4	0.4	0.4
60%	0	0	0	0	0	0	0

Table 3.4: Generated table of PW_b and Δt based on fitted curves. RFP and GFP are population fractions. Use of this system for event detection requires calibration of the system first in the lab by running experimental conditions for PW_b and Δt from 0 to 6 hours, fitting for the appropriate parameters, and generation of a similar table prior to deployment in the "field".

Derivation: A Markov model for integrase-based temporal logic gates

NOTE: Full derivations were done by Yutaka Hori and can be found in the appendix of Hsiao et al., In press.

List of plasmids and cell strains used

Plasmids		
Name	Resistance	Description
pVHed05	Cm	Controller plasmid (slightly modified Dual Recombinase Controller)
pVHed07	Kan	Integration plasmid for temporal logic gate in Phi80 site
pAH123 (Addgene 66077)	Amp (30C)	Helper plasmid needed for chromosomal integration in Phi80 site

Cell strains		
Name	Resistance	Description
DH5α-Z1		received from Endy lab
E. coli *pir 2+*		Necessary for cloning integration plasmids (contains the *pir* protein needed for replication of R6K origin of replication)
eVHed07	Kan/Cm	Chromosomally integrated temporal logic gate strain with integrase controller plasmid

Sequencing primers		
Name	Sequence	TM
ED_seq_F1	AAGCTTATGCCAACACAATT	59C (with Phusion Hotstart Flex 2x Mastermix)
ED_seq_R1	AGCTTCGTGGTTTGTCTG	59C

Plasmid maps

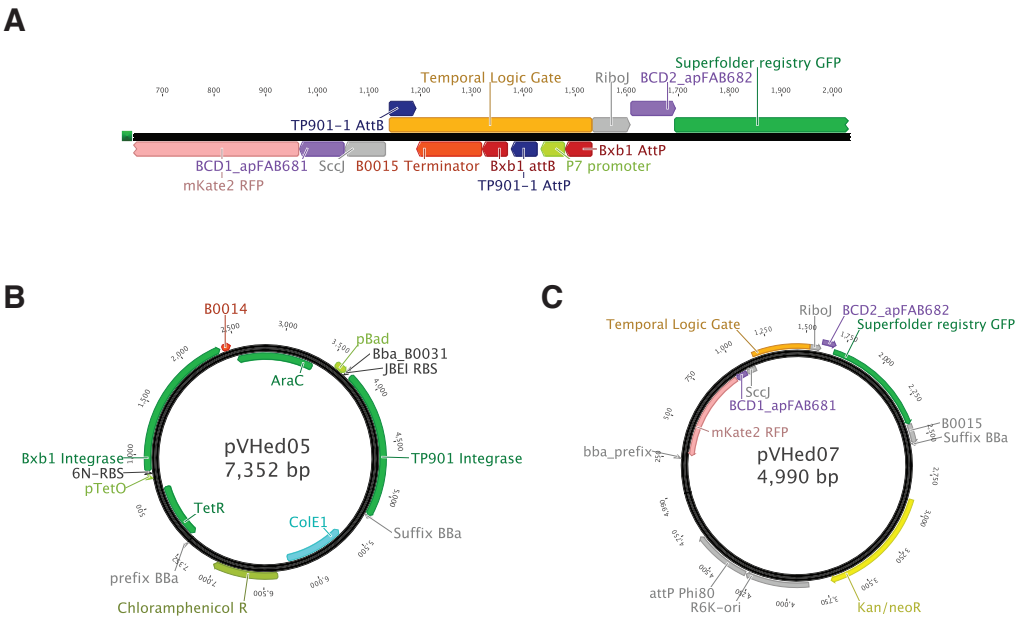


Figure ED-S30: Plasmid maps of temporal logic gate system. A) Design of the temporal logic gate. B) Controller plasmid for integrase A (Ptet-Bxb1) and integrase B (PBAD-TP901-1)

References

Anderson, J Christopher, Christopher A Voigt, and Adam P Arkin (2007). "Environmental signal integration by a modular AND gate." In: *Molecular Systems Biology* 3, p. 133.

Bonnet, Jerome, Pakpoom Subsoontorn, and Drew Endy (2012). "Rewritable digital data storage in live cells via engineered control of recombination directionality." In: *Proceedings of the National Academy of Sciences of the United States of America* 109.23, pp. 8884–8889.

Bonnet, Jerome, Peter Yin, et al. (2013). "Amplifying genetic logic gates." In: *Science* 340.6132, pp. 599–603.

Cai, Long, Chiraj K Dalal, and Michael B Elowitz (2008). "Frequency-modulated nuclear localization bursts coordinate gene regulation". In: *Nature* 455.7212, pp. 485–490.

Cortez, Maria Angelica et al. (2011). "MicroRNAs in body fluids—the mix of hormones and biomarkers". In: *Nature Reviews Clinical Oncology* 8.8, pp. 467–477.

Dunlop, Mary J et al. (2008). "Regulatory activity revealed by dynamic correlations in gene expression noise." In: *Nature Genetics* 40.12, pp. 1493–1498.

Elias, Sivan and Ehud Banin (2012). "Multi-species biofilms: living with friendly neighbors." In: *FEMS microbiology reviews* 36.5, pp. 990–1004.

Elowitz, Michael B et al. (2002). "Stochastic gene expression in a single cell." In: *Science* 297.5584, pp. 1183–1186.

Flemming, Hans-Curt and Jost Wingender (2010). "The biofilm matrix". In: *Nature Reviews Microbiology*.

Friedland, Ari E et al. (2009). "Synthetic gene networks that count." In: *Science* 324.5931, pp. 1199–1202.

Gardner, T S, C R Cantor, and J J Collins (2000). "Construction of a genetic toggle switch in Escherichia coli." In: *Nature* 403.6767, pp. 339–342.

Ghosh, Pallavi, Nicholas R Pannunzio, and Graham F Hatfull (2005). "Synapsis in phage Bxb1 integration: selection mechanism for the correct pair of recombination sites." In: *Journal of Molecular Biology* 349.2, pp. 331–348.

Gilad, Shlomit et al. (2008). "Serum microRNAs are promising novel biomarkers." In: *PLoS ONE* 3.9, e3148.

Gillespie, Daniel T (1977). "Exact stochastic simulation of coupled chemical reactions". In: *The Journal of Physical Chemistry* 81.25, pp. 2340–2361.

Green, Alexander A et al. (2014). "Toehold Switches: De-Novo-Designed Regulators of Gene Expression". In: *Cell* 159.4, pp. 925–939.

Haldimann, A and B L Wanner (2001). "Conditional-Replication, Integration, Excision, and Retrieval Plasmid-Host Systems for Gene Structure-Function Studies of Bacteria". In: *Journal of Bacteriology* 183.21, pp. 6384–6393.

Hsiao, Victoria et al. (In press). "A population-based temporal logic gate for timing and recording chemical events". In: *Molecular Systems Biology*.

Jahn, Michael et al. (2013). "Comparison of preservation methods for bacterial cells in cytomics and proteomics". In: *Journal of Integrated OMICS* 3.1.

Keravala, Annahita et al. (2006). "A diversity of serine phage integrases mediate site-specific recombination in mammalian cells." In: *Molecular genetics and genomics : MGG* 276.2, pp. 135–146.

Khaleel, Thanafez et al. (2011). "A phage protein that binds φC31 integrase to switch its directionality". In: *Molecular Microbiology* 80.6, pp. 1450–1463.

Kotula, Jonathan W et al. (2014). "Programmable bacteria detect and record an environmental signal in the mammalian gut." In: *PNAS* 111.13, pp. 4838–4843.

Lin, Yihan et al. (2015). "Combinatorial gene regulation by modulation of relative pulse timing." In: *Nature* 527.7576, pp. 54–58.

Mathis, Roland and Martin Ackermann (2016). “Response of single bacterial cells to stress gives rise to complex history dependence at the population level.” In: *Proceedings of the National Academy of Sciences of the United States of America*, p. 201511509.

McEwan, Andrew R, Paul A Rowley, and Margaret C M Smith (2009). “DNA binding and synapsis by the large C-terminal domain of phiC31 integrase.” In: *Nucleic Acids Research* 37.14, pp. 4764–4773.

Mitchell, Patrick S et al. (2008). “Circulating microRNAs as stable blood-based markers for cancer detection.” In: *Proceedings of the National Academy of Sciences of the United States of America* 105.30, pp. 10513–10518.

Moon, Tae Seok et al. (2013). “Genetic programs constructed from layered logic gates in single cells”. In: *Nature* 491.7423, pp. 249–253.

Ruess, Jakob et al. (2015). “Iterative experiment design guides the characterization of a light-inducible gene expression circuit.” In: *Proceedings of the National Academy of Sciences of the United States of America* 112.26, pp. 8148–8153.

Rutherford, Karen et al. (2013). “Attachment site recognition and regulation of directionality by the serine integrases.” In: *Nucleic Acids Research* 41.17, pp. 8341–8356.

Scott, Matthew et al. (2010). “Interdependence of Cell Growth and Gene Expression: Origins and Consequences”. In: *Science* 330.6007, pp. 1099–1102.

Shis, David L et al. (2014). “Modular, Multi-Input Transcriptional Logic Gating with Orthogonal LacI/GalR Family Chimeras”. In: *ACS Synthetic Biology*, p. 140728104557004.

Singh, Shweta, Pallavi Ghosh, and Graham F Hatfull (2013). “Attachment site selection and identity in bxb1 serine integrase-mediated site-specific recombination.” In: *PLoS Genetics* 9.5, e1003490–e1003490.

Singh, Shweta, Kate Rockenbach, et al. (2014). “Cross-talk between Diverse Serine Integrases”. In: *Journal of Molecular Biology* 426.2, pp. 318–331.

Siuti, Piro, John Yazbek, and Timothy K Lu (2013). “Synthetic circuits integrating logic and memory in living cells.” In: *Nature Biotechnology* 31.5, pp. 448–452.

Stoodley, P et al. (2002). “Biofilms as complex differentiated communities.” In: *Annual Review of Microbiology* 56, pp. 187–209.

Süel, Gürol M et al. (2007). “Tunability and noise dependence in differentiation dynamics.” In: *Science* 315.5819, pp. 1716–1719.

Tan, Cheemeng, Philippe Marguet, and Lingchong You (2009). “Emergent bistability by a growth-modulating positive feedback circuit”. In: *Nature Methods* 5.11, pp. 842–848.

Uhlendorf, Jannis et al. (2012). “Long-term model predictive control of gene expression at the population and single-cell levels.” In: *Proceedings of the National Academy of Sciences of the United States of America* 109.35, pp. 14271–14276.

Wang, Baojun et al. (2011). “Engineering modular and orthogonal genetic logic gates for robust digital-like synthetic biology”. In: *Nature Communications* 2, p. 508.

Xu, Zhengyao et al. (2013). “Accuracy and efficiency define Bxb1 integrase as the best of fifteen candidate serine recombinases for the integration of DNA into the human genome”. In: *BMC Biotechnology* 13.1, p. 1.

Yang, Lei et al. (2014). “Permanent genetic memory with >1-byte capacity.” In: *Nature Methods* 11.12, pp. 1261–1266.

Yuan, Peng, Kushol Gupta, and Gregory D Van Duyne (2008). “Tetrameric structure of a serine integrase catalytic domain.” In: *Structure* 16.8, pp. 1275–1286.

Chapter 4

APPLICATION OF TEMPORAL LOGIC GATE FOR DIFFUSION-ACTIVATED SPATIAL PATTERNING

4.1 Abstract

In previous work (Chapter 3), we demonstrated a two-input temporal logic gate circuit in which order and timing of the two-inputs could be sensed and recorded. This work was done in liquid culture, and inducers were introduced to the cultures at varying times. Here, we show that differential exposure of temporal logic gate cells to inducer order and timing can be achieved via point source diffusion on a 2D surface. We demonstrate that final cell state is then a function of position and proximity to inducer point sources, where only cells in the overlap region between the two inducers are exposed to an ***a*** *then* ***b*** event. This resulted in spatially differentiated sub-populations that could be imaged with fluorescence microscopy. We characterize the effects of growth rate and inducer distance on patterning. Additionally, we design a modified version of the temporal logic gate, called the Dual Function Cassette (DFC), that has unique outputs (though not unique genetic states) for ***a*** *then* ***b*** and ***b*** *then* ***a***, and report on those spatial patterning results.

4.2 Introduction

Spatial differentiation with diffusing inducers has been a long-standing model of multi-cellular development. Many studies have been done elucidating the effect of morphogen gradients on embryonic development, in which cell fates are dependent on steady state signal concentration (Porcher and Dostatni, 2010). In particular, the passive diffusion of the Bicoid (Bcd) transcription factor in *Drosophila* along the anterior-posterior axis is a classic example. Bcd induces transcription of downstream genes, called gap genes, in a concentration-dependent manner. However, modeling the steady state concentration gradient of Bcd alone does not account for the observed robustness in precision of patterning nor the rapid speed with which downstream gap genes are expressed (< 90 minutes) (Bergmann et al., 2007). Recent evidence shows that gap gene expression is defined, not by the steady state gradient, but by transient pre-steady-state Bcd profiles (Bergmann et al., 2007). Further investigations into Bcd mediated gap gene activation showed that this "early decoding" during the first 14 cell divisions is what enables patterning precision and robustness to noise (Tamari and Barkai, 2012). Morphogen induced differentiation triggered by geometric confinement has been shown to be sufficient for replicating early embryonic spatial patterning in human stem cells (Warmflash et al., 2014).

Synthetic bacterial systems could potentially be useful as simplified models of multi-cellular development. While much work has been done investigating spatial patterning with reaction-diffusion systems (Kondo and Miura, 2010; Hori et al., 2015), our temporal logic gate circuit uniquely interprets transient signals and encodes these signals into DNA states with differentiated transcriptional activity. Synthetic feedback circuits with quorum sensing molecules have been created to generate bullseye patterns (Basu et al., 2005) and macroscale colony structures (Blanchard and Lu, 2015) but these systems have no memory and cells do not retain differentiated genetic states.

The two-integrase temporal logic gate design featured in Chapter 3 utilizes stochastic responses in single cell behavior to generate population fractions that reflect inducer order and timing between the two inducers, and retains memory of events through DNA recombination. In previous experiments, inducers were added to liquid cultures at different times and population fractions were measured. We hypothesized that another way to implement delayed exposure to inducers was to use spatial diffusion of molecules. In the following chapter, we show preliminary experimental results in which we subject a uniform lawn of event detector *E. coli* to diffusing

inducers and observe the subsequent fluorescence patterning. This resulted in spatially differentiated sub-populations that were dependent on inducer diffusion rates and concentrations. We also designed a new event detector circuit, called the Dual Function Cassette (DFC), that has unique outputs (though not unique genetic states) for ***a*** *then* ***b*** and ***b*** *then* ***a***, and report on spatial patterning results for those new circuits.

4.3 Spatial patterning with temporal logic gate

Chapter 3 introduced an integrase-based temporal logic gate (Figure 3.1) that used two integrases (Bxb1 and TP901-1) to implement logic with four unique genetic states (S_o, S_a, S_b, S_{ab}). The final population fractions were characterized by manual addition of inducers into liquid culture. However, we hypothesized that slow diffusion on a 2D surface might also be sufficient for triggering event detection in temporal logic gate cells. Here, we show that diffusion-based delays are sufficient to create the timing differences necessary to trigger the temporal logic gate in a reliable way. Spatial patterns are characterized by testing plasmid-based and chromosomally integrated temporal logic gate strains on different types of media.

Figure 4.1A show the experimental set-up for 2D diffusion experiments. Standard petri dishes with 1% agar + M9CA minimal media were made and a biopsy punch (2mm diameter) was used to create point sources for liquid inducers, which were added at high concentration (arabinose 2%/vol, aTc 20 μg/ml). We expect these inducers (both small molecules) diffuse at a rate proportional to the density of the agar and the concentration of inducer, and to have some radius of induction. Since the inducers do not diffuse instantaneously, cells spread uniformly over the surface would encounter different sequences of events (Figure 4.1B). In the non-overlap regions, cells closer to the inducer **a** source will see ***a*** *only* and switch to state S_a (RFP fluorescence), while cells closer to the **b** source switch to state S_b (no fluorescence). Within the overlapping region between the two inducers, there will be two populations: Cell closer to source **a** will encounter ***a*** *then* ***b*** and switch to S_{ab} (GFP fluorescence), while cells closer to source **b** will encounter ***b*** *then* ***a*** and switch to state S_b (no fluorescence). We tested both a plasmid-based version of the temporal logic gate (multiple copies of DNA targets, much higher fluorescence), and the chromosomally integrated version used in the previous chapter (single chromosomal copy of DNA target, lower fluorescence output). Additional information about plasmids and strains constructed are described in Materials and Methods (Table 4.1).

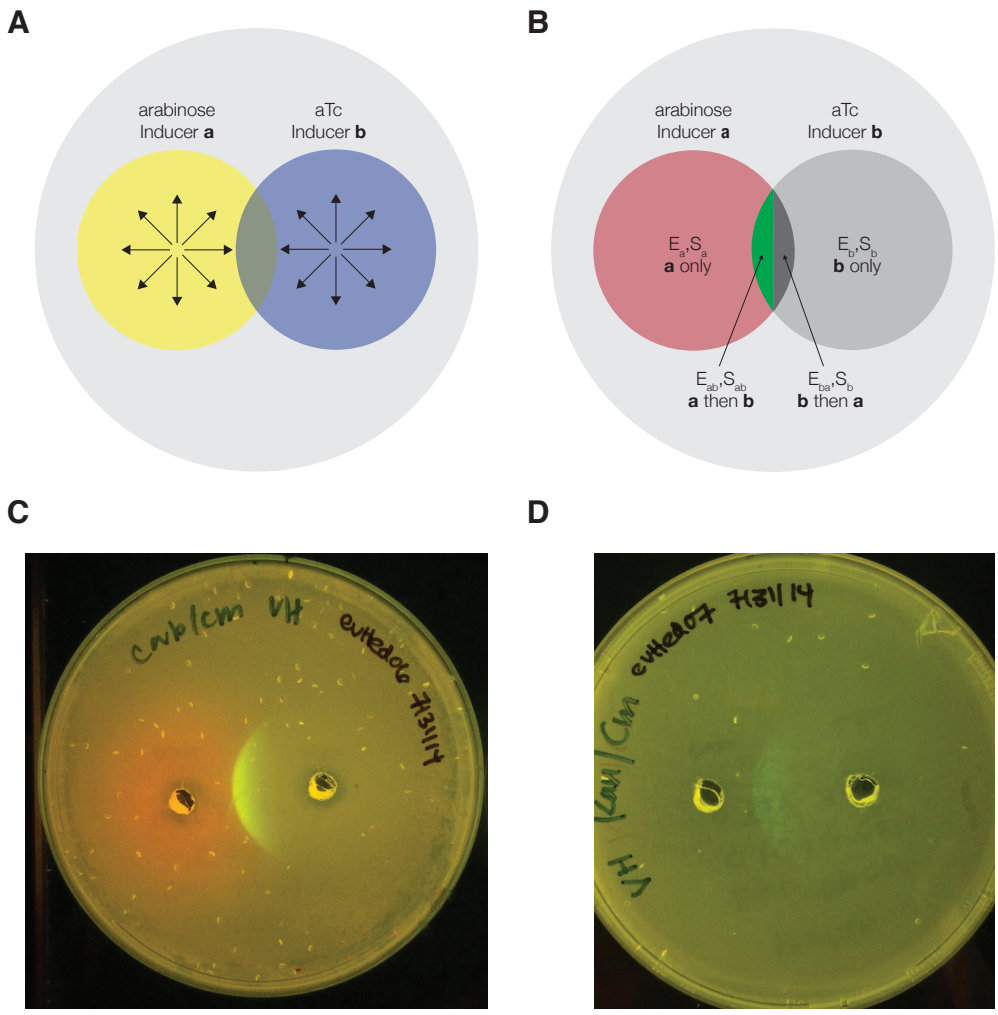


Figure 4.1: 2D diffusion experiment with temporal logic gate. A) Diagram of inducer diffusion. Point sources of arabinose (inducer **a**) and aTc (inducer **b**) will induce cells within a certain radius. Only a small region in the center will be exposed to both inducers. B) Diagram of expected cell states based on overlapping inducer diffusion. C) Spatial patterning from a uniform lawn of plasmid-based temporal logic gate strain. RFP expression occurs where the cells encounter ***a*** *only* and switch to state S_a, while GFP expression occurs only where cell encounter ***a*** *then* ***b*** and switch to S_{ab}. D) Diffusion pattern with chromosomally integrated temporal logic gate strain.

Diffusion experiments with a uniform lawn of the plasmid-based temporal logic gate created a bright red circle around the inducer **a** point source intersecting a bright green crescent centered around the inducer **b** point source (Figure 4.1C). The RFP expression corresponded to where cells encountered ***a*** *only* and switched to state S_a, while the GFP expression corresponded to where cells encountered ***a*** *then*

b and switched to S_{ab}. The chromosomally integrated strain resulted in the same pattern but with lower fluorescence overall (Figure 4.1D). Higher resolution images of Figure 4.1C,D are shown in Figure 4.3BE.

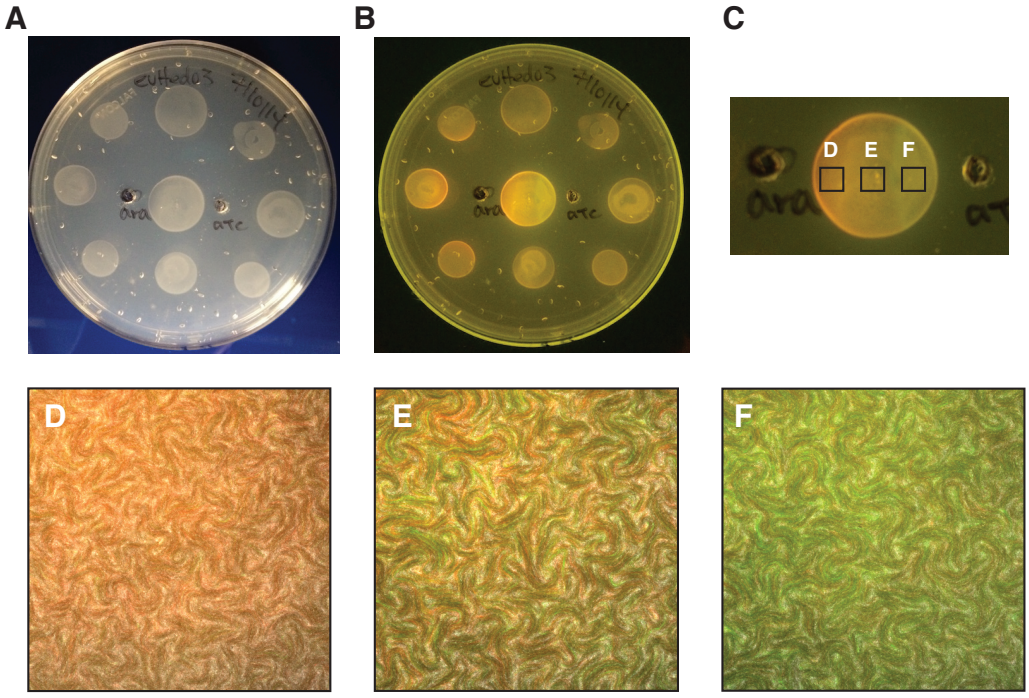


Figure 4.2: Microscopy of spatially differentiated colonies. Cell cultures of plasmid-based temporal logic gate were mixed with liquid media and spotted around inducer point sources. A) Plate imaged in ambient light shows overnight cell growth is independent of location on the plate. B) Blue light imaging shows RFP/GFP segmentation only between the two point sources. C) Cells in the overlap region (middle circle) were extracted and imaged at 10x magnification (Olympus IX81). Images were taken of cells in the red, middle, and green quadrants (labeled D,E,F). D) Within the overlap region, cells closer to the arabinose point source encounter inducer ***a*** *only* (S_a) and so express RFP. The densely grown cells show curly growth patterns. E) Cells that are equidistant from the two point sources are a mix of S_a (RFP) and S_{ab} (GFP). F) Cells in the overlap region that are closer to the aTc point source can sense ***a*** *then* ***b***, and switch to state S_{ab} (GFP).

Diffusion-differentiated cells were imaged at 10x magnification on a fluorescence microscopy (Figure 4.2). Cultures of plasmid-target strain were spotted onto an agar plate with inducer point sources and incubated overnight (Figure 4.2A). RFP fluorescence (S_a state, ***a*** *only*) was observed for regions encircling the arabinose point source, GFP fluorescence (S_{ab}, ***a*** *then* ***b***) was only observed in the region between the the two point sources, and no fluorescence was observed for cells encircling the

aTc source or that were too far away from either source (Figure 4.2B). Fluorescence patterns were confirmed by further imaging cells in the overlapping region at 10x magnification (Figure 4.2C). Cells in the region between the two point sources show segmentation based on distance from each point source (Figure 4.2C) – cells closer to the arabinose source are predominantly red (S_a)(Figure 4.2D), those in the middle are a mixture of red and green (Figure 4.2E), and those closer to the aTc source are green (S_{ab})(Figure 4.2F). A sharp border between RFP and GFP populations was not observed; instead, there was a gradual transition where the population became predominantly red or green.

4.4 Effect of growth rate on spatial patterning

Spatial patterning results were compared for plasmid and chromosomal event detector strains growing on three different types of media: LB, M9CA, and MOPS. In previous liquid culture experiments (Chapter 3), cells were grown in minimal media M9CA with 0.2% glycerol – the slow cell division time ensured integrase buildup in each cell, and prolonged the time before stationary phase, thus producing more consistent outcomes. Here, we hypothesized that rich media would also result in both less efficient DNA flipping and less time for the cells to respond to inducer, since the population saturates faster. Cell doubling times in LB, MOPS, and M9CA are roughly 20 minutes, 45 minutes, and 2 hours, respectively. MOPS and M9CA, unlike LB, are defined medias and so have less background autofluorescence in the GFP channel.

Spatial patterning outcomes are dependent on the relative ratio between the rate of inducer diffusion and the rate at which the population reaches stationary phase. While inducer diffusion rates are a function of agar density, the type of media sets the growth rate. This relationship between population growth and inducer diffusion becomes clear if we focus only on the diameter of RFP expression surrounding the arabinose (left) point source in Figure 4.3. Since the starting concentration of arabinose was the same for all six conditions, the radius of RFP expression reflects how far arabinose had diffused when the population reached stationary phase and stopped responding. The RFP radius is greatest for patterns on M9CA (Figure 4.3B,E) and notably smaller for those on LB (Figure 4.3A,C) and MOPS (Figure 4.3C,F) – this is true for both plasmid and chromosomal strains. Cell response to aTc also follows this pattern, but it is more difficult to visualize since GFP expression only occurs if population saturation is *slow* relative to inducer diffusion. This only occurs on M9CA (Figure 4.3C,F) and perhaps in some cases of MOPS (Figure

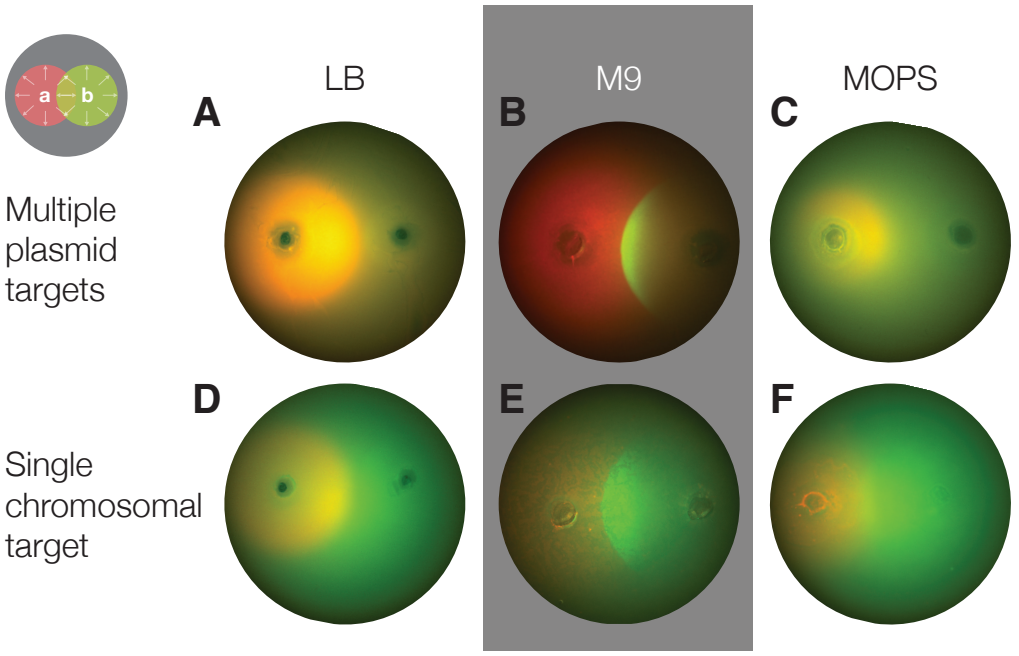


Figure 4.3: Comparison of spatial patterning with LB, M9CA, and MOPS media. Fluorescence images of diffusion experiments on agar plates were taken using an Olympus MVX10 macroscope (0.63x). The plasmid-based temporal logic gate strain was compared to the chromosomally-integrated strain. Arabinose (inducer **a**) point source was placed on the left, and aTc (inducer **b**) point source was placed on the right. A) Plasmid-based strain grown on LB shows RFP expression around arabinose point source but no GFP. B) Plasmid-based strain grown on M9 shows clear segmentation between the circle of S_a around the arabinose point source and a crescent of GFP (S_{ab}) where the arabinose ring intersects with the aTc ring. C) Plasmid-based strain grown on MOPS shows a smaller diameter ring of RFP expression compared to LB. D) Chromosomal strain grown on LB small RFP circle around arabinose point source. E) Chromosomal strain grown on M9CA shows sharply delineated region of GFP cells (S_{ab}) where arabinose and aTc overlap. F) A slight GFP overlap region can be seen for chromosomal cell grown on MOPS.

4.3F). Additionally, note the sharp crescent of GFP expressing cells where arabinose intersects with aTc, in contrast with the largely diffuse ring of RFP expressing cells surrounding the arabinose point source. We believe this is because the induction curve for aTc induction of Ptet is much sharper than the induction curve for arabinose/ PBAD. The ultrasensitivity in de-repression of tetR results in a remarkably crisp GFP circle where aTc and arabinose meet.

Overall, just as with liquid culture experiments, we concluded that M9CA was the best medium with which to grow the event detector circuits, and the plasmid version was possibly even preferable due to brighter fluorescence.

4.5 Dual Function Cassette Designs

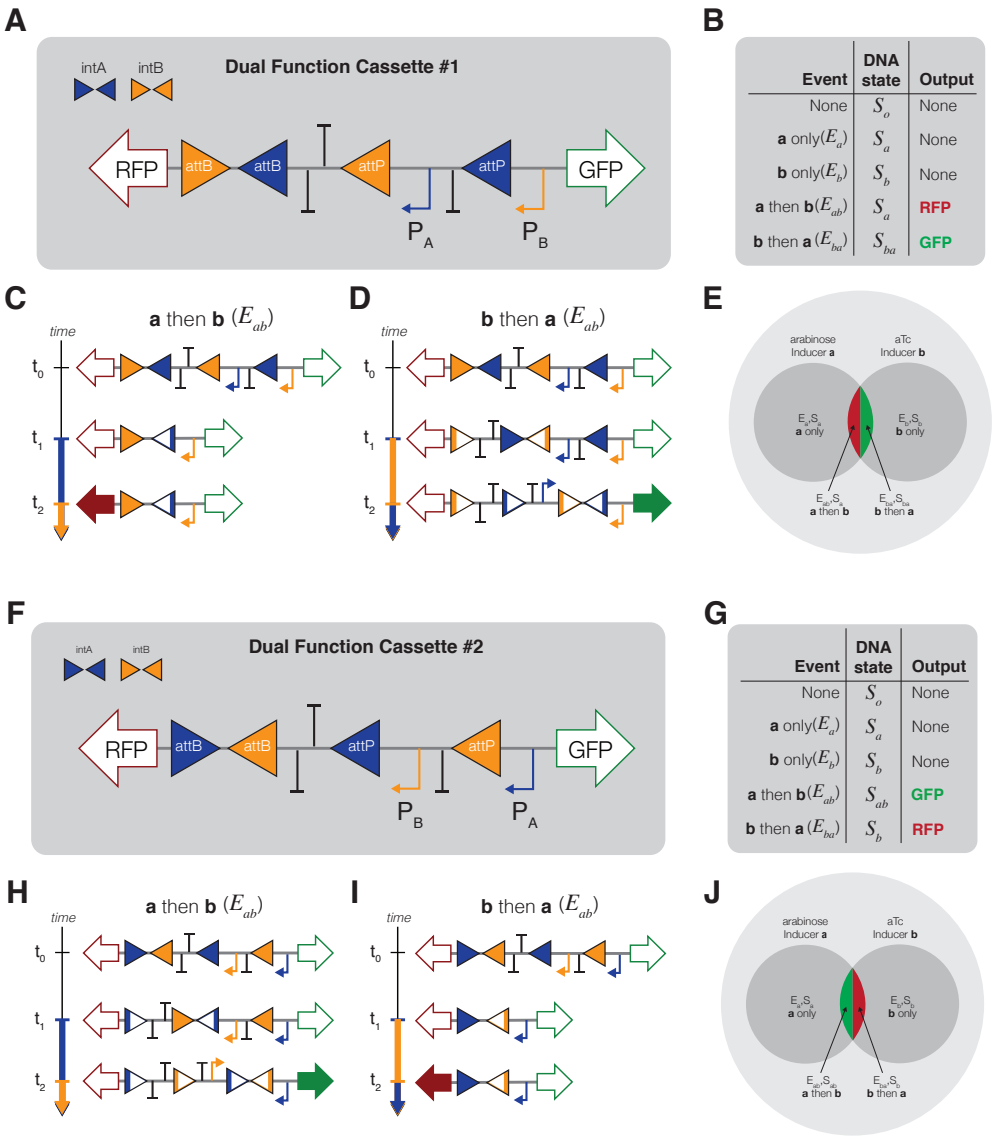


Figure 4.4: Designs for dual function cassette (DFC) event detector circuits. A) Layout for DFC design # 1. B) Truth table for DFC #1: RFP is expressed for ***a*** *then* ***b*** and GFP is expressed for ***b*** *then* ***a***. C) Mechanism of action for DFC #1 response to E_{ab} shows excision when exposed to inducer **a**, and then activation of the Ptet promoter when exposed to inducer **b**. D) Mechanism of action for DFC #1 response to E_{ba} shows two step flipping and activation of PBAD. E) Expected spatial patterning and genetic state of cells when exposed to diffusing point sources for DFC #1. F) DFC #2 design is the same as DFC #1 except the integrases have been switched. G) Truth table for DFC #2. H) DFC #2 response to E_{ab}. I) DFC #2 response to E_{ba}. J) Expected fluorescence and genomic state patterning for DFC #2.

We designed a modified version of the temporal logic gate to change the fluorescence outcomes for E_{ab} and E_{ba}. In particular, the original circuit expressed RFP with ***a*** *only* and GFP for ***a*** *then* ***b***, with no fluorescent output for ***b*** *then* ***a***. In order to fully visualize the overlap region between the two inducers, we designed a circuit that would only express fluorescence for ***a*** *then* ***b*** and ***b*** *then* ***a***.

The Dual Function Cassettes (DFCs) fundamentally utilize the same circuit topology as the original design, but the DNA target designs now have inducible promoters rather than a constitutive promoter (Figure 4.4). Furthermore, these inducible promoter are the same promoters used to induced the integrases. The two designs have the same topology – the positions of the two integrase binding sites in DFC #2 is simply the inverse of their positions in DFC #2. mKate2, the RFP previously used in the temporal logic gate, was switched out for mCherry because mCherry has faster maturation times.

In Figure 4.4A, the design for DFC #1 consists of the same interleaved integrase attachment sites as in the previous temporal logic gate circuit, however the constitutive promoter has been replaced with the same promoters that activate production of the integrases (P_A = PBAD, P_B = Ptet). Additional terminators have also been added to prevent read-through. The truth table for the DFC #1 has RFP expression for E_{ab} and GFP expression for E_{ba} (Figure 4.4B); however, this does not arise from a new unique genetic state. Rather, the addition of the first promoter removes obstruction by the terminators and the second inducer then activates the promoter (Figure 4.4C,D). If the circuit works as intended, the final fluorescence patterning would occur only where the two inducer diffusion circles overlap (Figure 4.4E). The design and outputs for DFC #2 are the inverse of DFC #1 (Figure 4.4F–J). Additional information about plasmids and strains are described in Materials and Methods (Table 4.2).

4.6 2D diffusion results for Dual Function Cassette circuits

The Dual Function Cassette (DFC) circuits were implemented as plasmid targets in DH5α-Z1 *E. coli* (Table 4.2). Two different distances between inducer point sources (1.5cm and 2.5cm) were tested, as well as low (Figure 4.5) and high (Figure 4.6) concentrations of inducer source. All patterning was done on M9CA agar plates. Inducer concentrations were 2% arabinose w/vol (inducer **a**), and 20ug/ml aTc (inducer **b**)) for high induction and 1/10th of that for low induction.

Diffusion-induced patterning results for the DFC strains at low inducer concen-

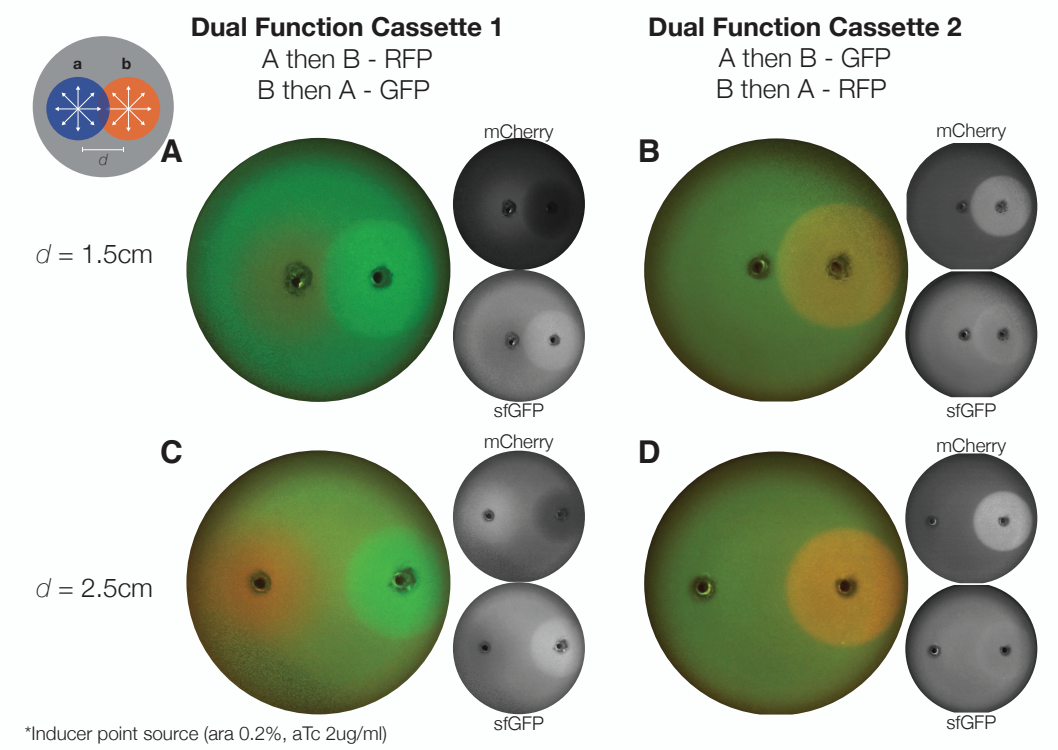


Figure 4.5: 2D diffusion patterning results for DFCs, low inducer source concentrations. Experiments were done with M9CA agar with *E. coli* lawn of DFC#1 and DFC#2 strains. Inducer point sources were 1/10th the usual concentration (inducer **a**(left): arabinose 0.2%, inducer **b**(right): aTc 2ug/ml). A) DFC #1 strain with 1.5cm distance between inducer point sources. Color image is composite of RFP and GFP channels (shown in smaller circles). B) DFC #2 strain with d = 1.5cm between inducers. C) DFC #1 strain with d = 2.5cm between point sources. D) DFC #2 strain with d = 2.5cm between point sources.

trations suggested inducer concentrations were too low for DNA recombination to occur. Rather than seeing fluorescence only in the overlap region, we observed uniform circles around each of the inducer point sources, which give an indication of background fluorescence expression. In Figure 4.5A, the DFC #1 strain has been exposed to inducer point sources that are 1.5cm apart, with low concentrations. Looking back at the circuit diagram (Figure 4.4A), we hypothesize that this RFP expression is background fluorescence from terminator run-through by P_A. Similarly, the GFP fluorescence observed in these results may be the result of backwards activity by P_B. In Figure 4.5C, where the inducers clearly do not overlap, this leaky background expression is overtly clear. Furthermore, when P_A is the promoter immediately upstream of GFP (DFC #2), we see no leaky fluorescence expression around the arabinose point source. Figure 4.4B,D also shows that the crisp circular

pattern is characteristic of Ptet and not of the fluorescent reporter. From these patterning results, we observed significant background fluorescence around each point source. We believe this may be due to insufficient terminator insulation and bidirectional activity by P_B (Ptet), though more experiments need to be done to confirm this.

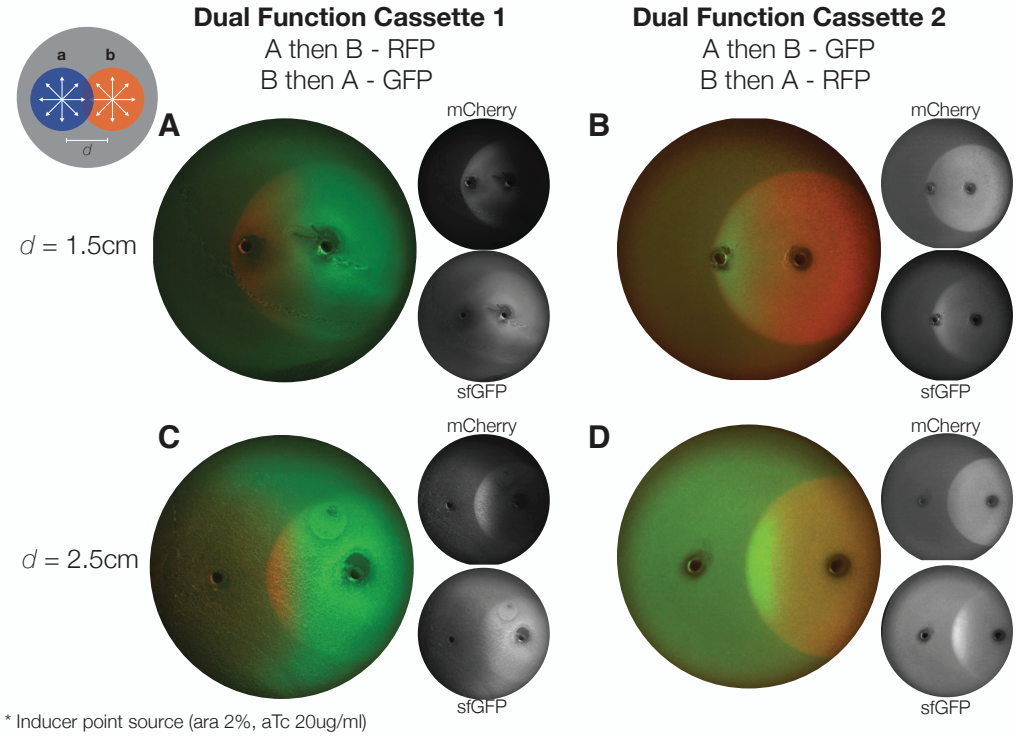


Figure 4.6: 2D diffusion patterning results for DFCs, high inducer source concentrations. Experiments were done on M9CA agar with *E. coli* lawn of DFC#1 and DFC#2 strains. Inducer point source concentrations were inducer **a** (left): arabinose 2%, inducer **b** (right): aTc 20ug/ml. A) DFC #1 strain with 1.5cm distance between inducer point sources. Color image is composite of RFP and GFP channels (shown in smaller circles). B) DFC #2 strain with d = 1.5cm between inducers. C) DFC #1 strain with d = 2.5cm between point sources. D) DFC #2 strain with d = 2.5cm between point sources.

Patterning results for high concentration inducer point sources show ***a*** *then* ***b*** dependent fluorescence but not ***b*** *then* ***a*** dependent fluorescence for both circuit types (Figure 4.6). For DFC #1 (Figure 4.6A,C), we expect cells in the overlap region to express RFP for ***a*** *then* ***b*** and GFP for ***b*** *then* ***a***. Despite some background RFP expression, significant RFP fluorescence over background was observed in the overlap region between the point sources (Figure 4.6A, mCherry inset), indicating ***a*** *then* ***b*** differentiation. However, we do not observe GFP fluorescence above background,

possibly because background GFP expression is already too high. For DFC #2 patterns (Figure 4.6B,D), we see the inverse effect, where regions of ***a*** *then* ***b*** overlap show high fluorescence, but RFP background is too high to observe any additional fluorescence for ***b*** *then* ***a*** regions. Overall, inducer concentrations are high enough such that both distances between inducers (1.5cm and 2.5cm) have overlap regions, with the 1.5cm distance having a larger overlap region that extends beyond the other inducer point source.

4.7 Conclusions and Future Work

We have shown that two-dimensional diffusion of inducers can be used to induce DNA recombination and differentiation in temporal logic gate *E. coli* strains. When cells are spread uniformly, their distance relative to each of the point sources determines the order and timing with which they encounter each inducer. Since two of the final genetic states express fluorescence, we were able to use macro- and microscale fluorescence imaging to produce spatial patterning and to investigate the relationship between inducer diffusion rate and cellular growth rate.

We also discussed modified designs of the temporal logic gate, called the Dual Function Cassette #1 and #2 (DFCs). These new circuits were designed to produce fluorescence only in the overlap region between the two diffusing inducer. Testing of spatial patterning with the DFCs showed high background fluorescence and uneven promoter responses for ***a*** *then* ***b*** versus ***b*** *then* ***a***. Preliminary analysis of these experiments suggests that terminator run-through and bidirectional promoter expression may be the reason, though more experiments need to be done to confirm this.

In future work, we would like to test other designs for patterning – particularly, new designs that would generate unique genetic states as well as unique fluorescence outputs. Previous studies have shown that a single integrase can have multiple pairs of orthogonal attachment sites by mutating the central dinucleotide at the heart of the DNA recombination reaction (Ghosh, Bibb, and Hatfull, 2008; Colloms et al., 2014). Results from Roquet et al., n.d. (unpublished) show that a finite state machine with unique genetic states for all possible inducer sequence combinations can be made using these orthogonal sites. This greatly expands the design space for genetic memory and DNA recombination. Additionally, we would like to build a 2D model to simulate communities of cells and their responses to diffusing inducers. Using these models, we would like to show that the unique aspect of transient

signal encoding in DNA memory can lead to spatial patterning regimes that are inaccessible by Turing reactions alone.

4.8 Materials and Methods

Media, agar plates, and inducers

Media used in these experiments were LB (Sigma, L3022), MOPS EZ Rich (Teknova M2105, 0.4% glycerol instead of glucose), and a custom formulation of M9CA. The media contained 1x M9 salts (Teknova, M1906) augmented with 100mM NH4CL, 2mM MGSO4, 0.01 % casamino acids, 0.15 ?g/mL biotin, 1.5 ?M thiamine, and 0.2% glycerol.

M9 and MOPS agar plates for diffusion experiments were made by autoclaving bactoagar under normal plate-making protocols at twice the normal concentration (15g / 500mL), and making the media at also twice the normal concentration (500mL). When the agar was cooled to 50C, the two solutions were mixed. These plates were cooled to room temperature and stored at 4C indefinitely.

Prior to experiment, plates were warmed at 37C. Low-melting point top agar with cell cultures was prepared as described in *Appendix D* (Final top agar 0.33%, cell OD 0.15). Once top agar had solidified, inducer point sources were create using a 2mm biopsy punch and inducers were added (2% arabinose wt/vol, 20ug/ml aTc).

Plasmids and strains

pVHed02	Temporal logic gate on low-copy plasmid (CarbR)
DRC	High copy plasmid with Ptet-Bxb1 and PBAD-TP901-1 (Dual Recombinase Controller, Addgene #44456)(CmR)
pVHed05	Modified DRC plasmid with additional copy of tetR repressor (CmR)
eVHed03	Plasmid-based temporal logic gate with DRC controller plasmid
eVHed06	Plasmid-based temporal logic gate with pVHed05 controller plasmid
eVHed07	Chromosomally integrated temporal logic gate strain (Φ80 site) with pVHed05 controller plasmid

Table 4.1: Table of relevant strains for temporal logic gate experiments

Microscopy

All patterning images in this section were taken with a Olympus MVX10 macroscope (0.63x), with the exception of Figure 4.2 – those 10x images were taken with an inverted Olympus IX81 scope. Agar plates were prepared as detailed in Appendix D, *Protocol for creating an even lawn of E. coli using top agar*, and incubated at

pVHed05	Dual recombinase controller plasmid + extra tetR. Has PBAD-TP901-1 and Ptet-Bxb1 (CmR)
pVHed22	Plasmid with DFC #1 – Dual response with RFP for A then B and GFP for B then A (CarbR)
pVHed23	Plasmid with DFC #2 – Dual response with GFP for A then B and RFP for B then A (CarbR)
eVHed13	DFC #1 experimental strain – DH5α-Z1 strain with pVHed22 + pVHed05 (CarbR/CmR)
eVHed14	DFC #2 experimental strain – DH5α-Z1 strain with pVHed23 + pVHed05 (CarbR/CmR)

Table 4.2: Table of relevant plasmids and strains for dual function cassette (DFC) designs

37C overnight.

References

Basu, Subhayu et al. (2005). "A synthetic multicellular system for programmed pattern formation". In: *Nature* 434.7037, pp. 1130–1134.

Bergmann, Sven et al. (2007). "Pre-steady-state decoding of the Bicoid morphogen gradient." In: *PLoS Biology* 5.2, e46.

Blanchard, Andrew E and Ting Lu (2015). "Bacterial social interactions drive the emergence of differential spatial colony structures." In: *BMC systems biology* 9.1, p. 59.

Colloms, Sean D et al. (2014). "Rapid metabolic pathway assembly and modification using serine integrase site-specific recombination." In: *Nucleic Acids Research* 42.4, e23–e23.

Ghosh, Pallavi, Lori A Bibb, and Graham F Hatfull (2008). "Two-step site selection for serine-integrase-mediated excision: DNA-directed integrase conformation and central dinucleotide proofreading". In: *PNAS* 105.9, pp. 3238–3243.

Hori, Yutaka et al. (2015). "Coordinated Spatial Pattern Formation in Biomolecular Communication Networks". In: *IEEE Transactions on Molecular, Biological and Multi-Scale Communications* 1.2, pp. 111–121.

Kondo, Shigeru and Takashi Miura (2010). "Reaction-diffusion model as a framework for understanding biological pattern formation." In: *Science* 329.5999, pp. 1616–1620.

Porcher, Aude and Nathalie Dostatni (2010). "The bicoid morphogen system." In: *Current biology : CB* 20.5, R249–54.

Roquet, Nathaniel et al. *Synthetic State Machines in Living Cells*. Tech. rep.

Tamari, Zvi and Naama Barkai (2012). “Improved readout precision of the Bicoid morphogen gradient by early decoding.” In: *Journal of Biological Physics* 38.2, pp. 317–329.

Warmflash, Aryeh et al. (2014). “A method to recapitulate early embryonic spatial patterning in human embryonic stem cells.” In: *Nature Methods* 11.8, pp. 847–854.

Chapter 5

STATIONARY PHASE PROMOTERS FOR GROWTH-PHASE INDUCED DELAYS AND SPATIAL PATTERNING

5.1 Abstract

Current bacterial synthetic circuits rely on the fast dilution and high protein expression that occurs during exponential phase. However, constant exponential phase is both difficult to ensure in a lab environment and almost certainly impractical in any natural setting. Here, we characterize the performance of 13 *E. coli* native σ38 promoters, as well as a previously identified σ38 consensus promoter. We then make tetO combinatorial versions of the three strongest promoters to allow for inducible delayed expression. The design of these combinatorial promoters allows for design of circuits with inducible stationary phase activity that can be used for phase-dependent delays in dynamic circuits or spatial partitioning of biofilms.

5.2 Introduction

We envision the future of synthetic biology to involve the deployment of engineered bacteria into potentially harsh and minimal nutrient environments for long periods of time. Although current synthetic circuit testing is done almost exclusively in exponential growth phase, this is neither a realistic environment for natural bacteria nor a reasonable expectation for engineered strains outside of the laboratory. The common constitutive and inducible promoters reduce their gene expression activity considerably when the population reaches stationary phase, and there are not yet well-defined parts libraries for stationary phase active promoters. Stationary phase gene expression in E. coli has been widely studied (Miksch and Dobrowolski, 1995; Lee and Gralla, 2001; Lacour and Landini, 2004; Shimada et al., 2004; Gorochowski et al., 2014) but we have not yet been able to take advantage of stationary phase stage promoters for synthetic circuits.

Here, we present a small step towards finding, characterizing, and engineering a stationary-phase active promoter library by mining the existing literature for known stationary phase sigma-factor promoters, testing their activity, and engineering them into inducible combinatorial promoters. The creation of this library would have a number of significant advantages. Most importantly, cell functionality could be partitioned into exponential phase tasks and stationary phase tasks. Cells could

be programmed to conserve resources until some quorum had been reached, and experimental strains would not have to be constantly diluted to maintain exponential growth. Furthermore, dynamic circuits that rely on delayed stationary phase activity would be repressed until the start of the experiment and presence of the inducer. We also examine the use of these promoters for visualization and spatial partitioning of bacterial biofilms.

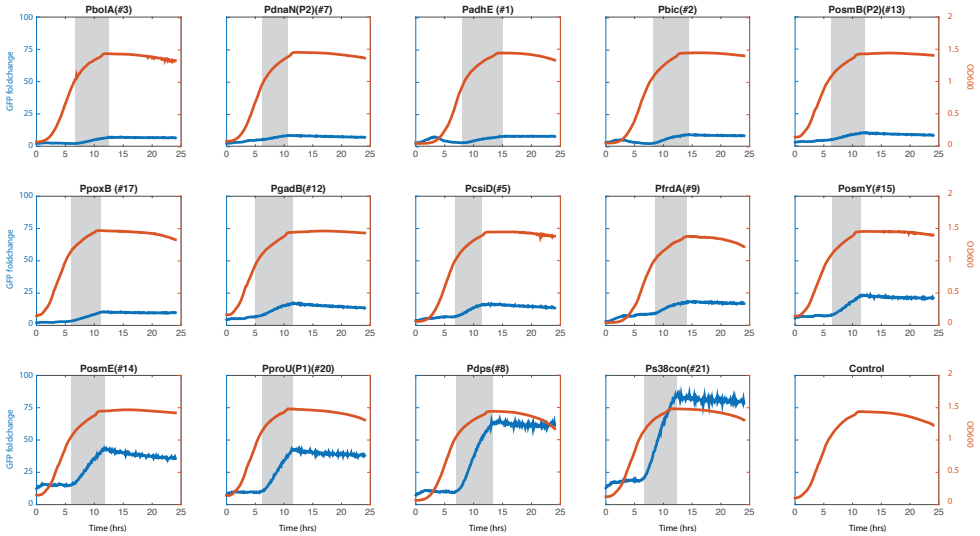


Figure 5.1: Panel of stationary phase promoters. Native promoter sequences in DH5a-z1 E. coli as previously identified by Lee *et. al* (Lee and Gralla, 2001), ordered by increasing strength of GFP expression. Orange traces represent OD600 and blue traces represent GFP fold change as normalized by the DH5a-Z1 control strain. Time of GFP increase is highlighted with gray boxes. All of the promoters turned on in late-log phase. Cells were grown at 37C in LB media.

5.3 Native promoter panel

We began with a set of 21 sigma38 promoters previously identified in the literature (Lee and Gralla, 2001). σ38, also known as σS or RpoS, is a RNA polymerase co-factor that is selectively expressed during stationary phase from the *rpoS* gene. First identified by Lange and Hengge-Aronis in 1991(Lange and Hengge-Aronis, 1991), σ38 is the main stationary phase and stress response promoter and controls more than 500 genes (Serra et al., 2013). After identifying the section of putative promoter sequences, we designed a test plasmid with the promoter sequence, a standardized bicistronic ribosomal binding sequence (BCD2-RBS), and a fast folding green fluorescent protein (sfGFP). Figure 5.1 shows the full panel of σ38 promoters

tested, in which all the promoters have minimal expression until late-log phase and then express GFP with a range of characteristic strengths. The full panel includes 13 native promoters and one sigma38 consensus promoter created from the consensus motifs of the other promoters (Lee and Gralla, 2001). The control was a non-fluorescent DH5a-Z1 strain. All of the promoters turned on in late-log phase – we are currently investigating whether that is due to late-log expression of σ38 or individual cells reaching stationary phase at different times.

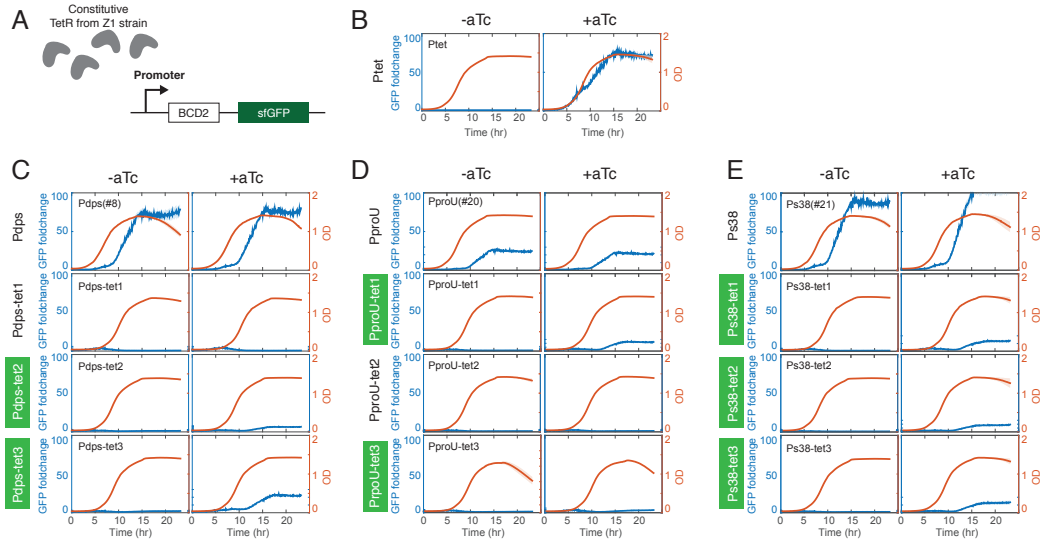


Figure 5.2: Panel of combinatorial σ38 promoters. Combinatorial versions of the strong three native σ38 promoters were designed and tested. Orange lines represent OD600 traces, blue lines show GFP fold change over control. A) Design of the test system consisted of cloning each promoter variation in front of BCD2-sfGFP and then transforming the plasmid into a Z1 strain that constitutively makes TetR repressor. B) Control plasmid with Ptet-GFP that only expresses GFP when aTc is present in the media (Added at time t=0h). C) Pdps native promoter and variations. Working combinatorial promoters are highlighted in green. D) PproU promoter and variations. E) Ps38 consensus and variations.

5.4 Engineered combinatorial promoters

From this panel of working σ38 late-log phase promoters, we chose the strongest three promoters, PproU(#20), Pdps(#8), and Ps38consensus (#21) to create combinatorial promoters. For each of the three s38 promoters, we designed three variations of a tetR combinatorial promoter. In the tet1 variant, a single tetO DNA binding motif was placed directly upstream of the promoter sequence 3' of the -10 region. In the tet2 variant, the tetO motif was placed between the -10 and the -35 regions,

with all of the original spacing bases deleted. In the tet3 variant, the tetO motif was again placed between the -10 and -35 regions with 2 original spacing bases on either side.

All nine combinatorial promoter variants, plus the original promoters, and a control Ptet promoter were tested in media with and without aTc in a DH5a-Z1 strain which constitutively produces tetR (Figure 5.2A). The control Ptet-GFP plasmid turns on only when aTc is present in the media (added at time = 0h) (Figure 5.2B). The original stationary phase promoters turn on in late log phase independently of aTc, and the successful combinatorial promoters only turn on in late log phase when aTc is present (Figure 5.2CDE). The working combinatorial promoters are highlighted with green rectangles, with seven out of nine working on the first attempt.

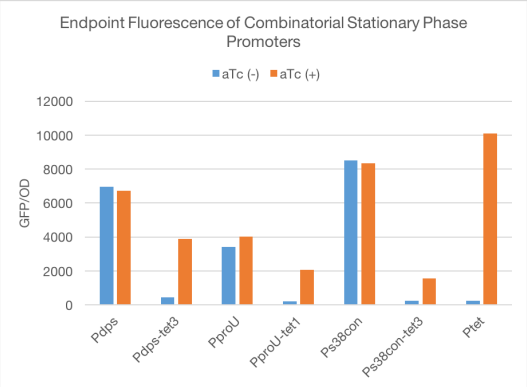


Figure 5.3: Endpoint fluorescence of combinatorial versus native stationary phase promoters.

We can compare the performance of the top three combinatorial promoters (dps-tet3, proU-tet1, s38con-tet3) with the native promoters, and we see at least a 50% reduction in overall expression strength (Figure 5.3). This indicates that replacing the bases between the -10 and -35 regions either decreases overall σ38 affinity, or the aTc induction was not high enough to full relieve tetR repression.

5.5 Spatial patterning of biofilms

E. coli naturally form biofilms that are comprised of multiple layers of cells in different growth phases (Serra et al., 2013; Hobley et al., 2015). The availability of stationary phase active promoters in the context of biofilms means that cells could be programmed to express different functions based on their location within the naturally occurring biofilm. Within a natural biofilm, the cells on the edges are in exponential phase and dividing to expand the biofilm, the cells in the bottom

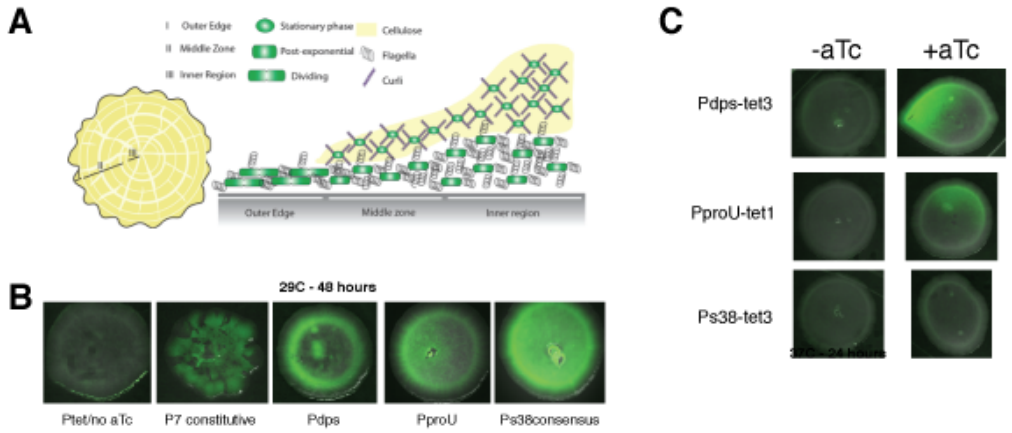


Figure 5.4: Use of stationary phase promoters for understanding biofilms. A) Diagram copied from Hobley *et. al* (Hobley et al., 2015). *E. coli* biofilms are comprised of multiple layers of cells in different growth phases. B) Imaging of native σ38 promoters within an *E. coli* biofilm. Ptet-GFP with no aTc and the constitutive P7-GFP strains are presented as controls. C) Biofims of colonies with combinatorial σ38 promoters with and without aTc.

central layer are post-exponential and no longer dividing, and the cells in the upper central layer, furthest from the nutrients in the agar, are in stationary phase (Figure 5.4A) (Hobley et al., 2015). Natural biofilms already take advantage of spatial differences in growth phase – Cells on the biofilm periphery are in exponential phase and expand quickly, but are also more susceptible to attack and stress, while cells on the interior in stationary phase are protected but also receive less nutrients (Liu et al., 2015). Additionally, we could potentially combinatorial stationary phase promoters to purposely couple synthetic model activity to the existing cellular infrastructure, i.e. the complex mechanisms that determine growth phase and σ38 production. Rather than design a synthetic timer, we could use growth-phase dependent promoters to implement delays or oscillatory behaviors in synthetic circuit activity.

In preliminary experiments, when we spot strains with the native σ38 promoters onto LB agar, we see spatially dependent expression of GFP compared to a constitutive strain (Figure 5.4B). Notably, we see differential expression of the each of the three different stationary phase promoters, with Pdps expression primarily around the outer ring, PproU expression going out in tendrils from the center, and s38consensus expression evenly spread out (Figure 5.4B).

When the combinatorial promoters Pdps-tet3, PproU-tet1, and Ps38-tet3 are tested on plates with and without aTc, we can observe GFP expression turning on based on aTc availability. Because the overall output of these promoters is much weaker than the endogenous promoters, it is more difficult to see the same phenotypic differences observed in Figure 5.4B, thus far.

5.6 Conclusion

In future work, we would like to further understand the biological mechanisms behind these promoters – are they active in late log phase, or are different cells reaching stationary phase at different times? What is the protein production rate after the population reaches full stationary phase? We would also like to explore interesting new genetic circuit designs that can be achieved with delays implemented using these promoters. Directed evolution and error-prone PCR libraries of these promoters could also be used to select for higher expression or tighter expression windows.

In order to fully take advantage of spatially-dependent expression, we need to first understand promoter expression in the context of the *E. coli* biofilm. In addition to taking top down images, we are planning to take cross-sections of the biofilm to see if layers can be visualized.

σ38 is not only the main stationary phase promoter, it's also the general stress response promoter. It would be advantageous to find the promoter sequences of less promiscuous growth-phase dependent transcription factors.

5.7 Materials and Methods

All cells were grown in LB media at 37C in DH5a-z1 E. coli. Biofilms were spotted from 5-10ul of liquid cultures and grown at 29C or 37C for 24-48 hours.

5.8 Acknowledgments

VH is supported by the Department of Defense (DoD) through the National Defense Science & Engineering Graduate Fellowship (NDSEG) Program. Research supported in part by the Institute for Collaborative Biotechnologies through grant W911NF-09-0001 from the U.S. Army Research Office. The content of the information does not necessarily reflect the position or the policy of the Government, and no official endorsement should be inferred.

References

Gorochowski, Thomas E et al. (2014). "Using synthetic biological parts and microbioreactors to explore the protein expression characteristics of Escherichia coli." In: *ACS Synthetic Biology* 3.3, pp. 129–139.

Hobley, Laura et al. (2015). "Giving structure to the biofilm matrix: an overview of individual strategies and emerging common themes". In: *FEMS microbiology reviews* 39.5, pp. 649–669.

Lacour, Stephan and Paolo Landini (2004). "SigmaS-dependent gene expression at the onset of stationary phase in Escherichia coli: function of sigmaS-dependent genes and identification of their promoter sequences." In: *Journal of Bacteriology* 186.21, pp. 7186–7195.

Lange, R and R Hengge-Aronis (1991). "Identification of a central regulator of stationary-phase gene expression in Escherichia coli." In: *Molecular Microbiology* 5.1, pp. 49–59.

Lee, S J and J D Gralla (2001). "Sigma38 (rpoS) RNA polymerase promoter engagement via -10 region nucleotides." In: *The Journal of Biological Chemistry* 276.32, pp. 30064–30071.

Liu, Jintao et al. (2015). "Metabolic co-dependence gives rise to collective oscillations within biofilms." In: *Nature* 523.7562, pp. 550–554.

Miksch, G and P Dobrowolski (1995). "Growth phase-dependent induction of stationary-phase promoters of Escherichia coli in different gram-negative bacteria." In: *Journal of Bacteriology* 177.18, pp. 5374–5378.

Serra, Diego O et al. (2013). "Microanatomy at cellular resolution and spatial order of physiological differentiation in a bacterial biofilm." In: *mBio* 4.2, e00103–13.

Shimada, Tomohiro et al. (2004). "Classification and strength measurement of stationary-phase promoters by use of a newly developed promoter cloning vector." In: *Journal of Bacteriology* 186.21, pp. 7112–7122.

Chapter 6

CONCLUSIONS AND FUTURE DIRECTIONS

6.1 Conclusions

Synthetic biology is a continuous process of identifying molecular components, testing them in a non-native environment, and re-factoring them to improve their performance against specified metrics. While it is not yet obvious what bioengineering paradigms should be, it is clear that mixing-and-matching biological components does not follow the rules of traditional engineering standards and practices. Recent work, including this one, has shown that the field is pivoting away from precise uniformity and minimizing noise. Instead we have begun to design for, and even take advantage of, the stochasticity and delay inherent in biological systems.

Although a great deal of progress has been made in creating libraries of characterized genetic and molecular tools and components, it is still difficult to abstract away from the specific limitations of the way that circuits are implemented. We have taken a function-based design approach to two projects: a negative feedback loop that works as a concentration regulator, and a temporal logic gate that can detect order and timing between two inputs. In both circuits, the molecular implementations determined the limitations and time constants of the experimental system, the overall behavior of the system was determined by the circuit topology, and could be modeled without precise fitting of biological parameters. Furthermore, we used both models for forward predictions on experimental outcomes as well as insight into unexpected experimental results.

In Chapter 2, we demonstrated a novel negative feedback tracker circuit that maintains the ratio between an input and an output protein. The underlying topology of the circuit was a single negative feedback loop, a common motif for regulating steady state concentrations. Our implementation with modular synthetic scaffold proteins and a scaffold-dependent two-component system resulted in extremely rapid responses to input. This is because two of the components, the kinase and response regulator, could be constitutively produced with minimal cross-talk. We used simulation data to investigate how output gain, steady-state ratios, and cycling time could be tuned by changing how quickly the response-regulator is de-activated, and verified these predictions experimentally.

In Chapter 3, we designed and implemented a temporal logic gate that takes advantage of population dynamics to collectively sense and record sequences of transient chemical inputs. We not only demonstrated that single cells independently sense and record events but also that aggregate population fractions create unique outcomes, providing information not encoded in single cells. The use of bacteriophage integrases provided a way to simultaneously achieve temporal logic and memory within a relatively small region of DNA; however, the ability of the circuit to generate of population fractions proportional to past events relied only there being delay and stochasticity in single cell responses to inputs. This was proven with a stochastic Markov-chain model that represented integrase activity only as probabilities for state transitions, but still produced useful predictions on overall population-level outcomes. Therefore, the conclusions of this study could be generalized for other implementation of temporal logic gates, provided the network topology consisted of delayed, stochastic digital switches.

In Chapter 4, we demonstrated that the temporal logic gate also be used for spatial patterning and differentiation. This was achieved by spreading a uniform lawn of cells across an agar plate with a point source for each inducer. The inducer order and timing was then determined by a cell's proximity and placement relative to the two diffusing point sources. Using fluorescence images, we characterized the effects of inducer concentration, cellular growth rate, and diffusion rate on patterning.

In Chapter 5, we identified 13 σ38 stationary phase promoters from the literature and designed 7 combinatorial promoters that activate only in late-log phase and in the presence of inducer. We tested applications of these promoters for visualizing the growth phase of cells across a biofilm.

6.2 Future work

In Chapter 2, modular scaffold domains were used to implement a negative feedback loop for ratiometric control of two proteins. This would be an alternative to open-loop tuning of each component in a synthetic circuit, where absolute concentration is not important as long as ratios are maintained. scaffold proteins in natural networks are hubs for signal cascades, usually by colocalizing phosphorylation machinery (Good, Zalatan, and Lim, 2011), and previous studies have rewired the MAPK cascade in yeast to redirect signals, modify delays, and introduce ultra sensitivity (Park, Zarrinpar, and Lim, 2003; Bashor et al., 2008). Although endogenous signaling cascades use protein ratios for feedback and activation, there has not yet

been a synthetic circuit with more than one layer of feedback. Future iterations of our design could improve tracking fidelity by including a second reverse feedback loop to compensate for over-expression, and orthogonal circuits could be made with other pairs of two-component systems. This will facilitate building of synthetic signal cascades by regulating ratios between parts of the cascade.

In Chapter 3, we designed a integrase-based temporal logic gate for detecting sequences inputs and genetically modifying the cells accordingly. Furthermore, we showed that counting population fractions are a reliable method of determining past events. Future directions for this circuit are primarily focused on applications for generating distributions and spatial patterning for multi-cellular consortia. The temporal logic gate could be used not as a endpoint readout of past events, but as a population generator that can take a homogenous population of "stem cell" bacteria, and through timed inputs, reliably create distributions of genetically distinct daughter cells. A controlled way to generate population distributions opens up possibilities for testing distribution-based feedback. We are collaborating with another graduate student to apply a finite state projection model to our integrase circuit, and to design additional layers of feedback once the cells have differentiated.

Preliminary work for spatial patterning using the temporal logic gate was presented in Chapter 4. Although much work has been done on diffusion-reaction patterns, our system has unique properties that result from genetic memory of transient events. We are collaborating with another graduate student on a 2D model to explore patterns and addition feedback that could be implemented. The goals of this model would be to show how our system creates a platform for downstream communication and feedback that is not accessible with Turing patterns alone. Eventually, this may create a model system that could provide insights on embryonic development or be a starting platform for a synthetic multi-strain bacterial consortia.

In Chapter 5, we designed combinatorial stationary phase promoters. We plan to use the natural un-modified promoters to understand location-dependent growth within natural biofilms, and then to use to the combinatorial promoters for position-dependent activation with a synthetic biofilm. Furthermore, inducible stationary phase promoters are "AND" gates with built in delay, since one of the inputs would be late-log phase. This opens up possibilities for circuit design, in which a time-delay threshold is implemented. For instance, one could design a circle in which cells wait to commit to a differentiated state until all they are saturated, allowing conservation of resources up until that time. Additionally, it is difficult to use the

same parts for different circuits if they are to be implemented in the same cell. Stationary phase promoters would allow for partitioning of phase-dependent tasks, and provide temporal orthogonality for repeated parts.

The work presented in this thesis contributes towards a framework where circuit design can be abstracted away from the specific molecular implementation. More directly, this work has led to future applications for building synthetic multi-cellular communities with multiple layers of feedback within single cells, between cells, and between populations of differentiated cells. These robust biological devices would modulate their activity in response to growth phase, competition, and specified inputs.

Appendix A

CRIM CHROMOSOMAL INTEGRATION PROTOCOL

A.1 Overview

The CRIM (Conditional-replication, integration, and modular) plasmid system is a series of plasmid vectors which use the R6K replicon for maintenance as an autonomous DNA element in pir+ hosts and contain an attP site for site-specific integration into the chromosome of non-pir+ hosts. Integration into non-pir+ hosts requires the expression of the corresponding lambdoid phage Int protein (i.e. integration into attBλ requires attPλ and Intλ). The procedure outlined here is directly adapted from the 2001 Haldimann and Wanner paper.

Sources:

- Haldimann, A., & Wanner, B. L. (2001). Conditional-Replication, Integration, Excision, and Retrieval Plasmid-Host Systems for Gene Structure-Function Studies of Bacteria. Journal of Bacteriology, 183(21), 6384?6393. http://doi.org/10.1128/JB.183.21.6384-6393.2001
- http://openwetware.org/wiki/Rao:CRIM

A.2 Materials

Desired host containing the Int expressing helper plasmid at 30^{o}C. All of these helper plasmid have the Ts SC101 replicon and the int gene under λPRM regulated by the Ts λcI857 repressor. The idea is to cure the plasmid while expressing the corresponding Int.

Plasmid	integration site
pINTts	Int(λ)
pAH69	Int(HK022)
pAH123	Int(ϕ80)
pAH121	Int(P21)
pAH130	Int(P22) - Not recommended for e14+ strains (e.g. MG1655).

A.3 Creating your CRIM plasmid

- You will need a pIT3 backbone that has an attP Phi80 site and an R6k-origin of replication (i.e. pIT3_KP_280 from Jerome Bonnet).
- The Phi80 integrase will integrate the entire plasmid into the chromosome at the Phi80 site. Use your favorite cloning method (Gibson, golden gate, etc), but at the end, make sure you transform into a competent E. coli pir2+ strain - Only a pir2+ strain will replicate your plasmid (since it has an R6k-ori).
- Recover and grow at 37C on a plate as normal.

A.4 CRIM integration

1. Grow an overnight of your Int expressing strain in LB (Amp 100 ug/mL) at 30^oC.
2. The following morning, subculture your Int expressing strain 1:100 in fresh LB(Amp 100 μg/mL).
    - Concentrate cells for electroporation 100X. 5 mL of culture will make 50 μL of electrocompetent cells.
3. Grow with vigorous shaking at 30^oC, until your culture has reached an OD600 of around 0.6.
    - Do not let the cells enter late-log/stationary phase. You need them to continue growing during recovery so they can express Int effectively.
4. Prepare the cells for electroporation at 100X concentration (modified from Emzo's quick and dirty electrocomp protocol)
    - Take 1 ml of the culture and spin down in 5000rpm for 2 min
    - Resuspend in 10% cold glycerol, keep on ice, spin down again.
    - Repeat 3x, then resuspend in 50ul (since there's always some liquid left, that will give you 100ul)
    - Electroporate with 100ul + 1ul of CRIM plasmid.
    - Recover in 1 mL LB or SOC for 1 hr at 37^oC for efficient recovery and Int expression.
5. After one hour, move the cultures to 42^oC for 30 minutes before plating.
    - This step helps cure the plasmid and allows for high level expression of the Int protein. This step increases the efficiency of integration and should not be omitted. A high level of false positives have been observed when recovery is only done at 37^oC.

6. Plate on selective agar for your CRIM's marker at 37C.
    - It is advisable to plate 1:10 and 1:100 dilutions as the integration process is usually efficient and this helps reduce false positives.
7. The next day, restreak some colonies on a fresh plate and grow at 42^{o}C to clear the helper plasmid.
8. The third day, PCR check the integration using the primers suggested in the paper.

Appendix B

M9CA MINIMAL DEFINED MEDIA RECIPE

Source: Adapted from protocol found on Rob Philips's lab website

M9 salts are used for making minimal media; the salts provide basic ionic buffering for the cells, and also provide an environment with comfortable osmotic properties. However, there are no nutrients.

B.1 5x M9 salts for bacteria

Per Liter:

- 30g Na_2HPO_4
- 15g KH_2PO_4
- 5g NH_4Cl
- 2.5g NaCl
- 15mg $CaCl_2$ (optional)

Add ingredients to water in 2L flask and heat with stirring until dissolved. Pour into bottles with loosened caps and autoclave. Cool to < 50 deg C before adding nutritional supplements and antibiotics. Tighten caps and store concentrated media indefinitely at room temperature.

B.2 M9CA recipe

Before use, dilute concentrated media to 1x w/ sterile water and add the following sterile solutions, per liter to make minimal (MGC) media:

M9CA recipe:

- 1X M9 minimal media (1X salts + 1mL 1M $MgSO_4$)
- 0.2% glycerol
- 0.01% casamino acids
- 0.15 ug/mL biotin
- 1.5 uM thiamine

For 1L total volume, mix together:

- 200 mL sterile 5X M9 salts
- 2mL 1M $MgSO_4$
- 4mL 50% glycerol
- 2mL 5% casamino acids
- 1mL 150 ug/mL biotin
- 250 uL 6mM thiamine
- 800 mL water to 1000 mL total
- mix together, filter sterilize

To make 1L plates:

- Autoclave 15g agar in 800mL water for 15 min.
- Add everything else listed above (5x M9, $MgSO_4$, glycerol, CA, biotin, thiamine)
- Cool to 50 deg C. and add antibiotics.
- Pour 32-40 mL medium/plate for 25-30 plates/liter.

Appendix C

SEMI-DRY WESTERN BLOT PROTOCOL

Western blot protocol used for Hsiao, Santos, et al., 2015 (Used for proteins 30-50 kDal) (Based primarily on David Chan lab's Western protocol / also C. Smolke lab protocol found online)

C.1 Lysing cells and running SDS-PAGE gel

1. Resuspended 50ul of cell pellet in 100ul of lysis buffer.
2. Added 12.5 ul of resuspended cells + 12.5ul of protein loading dye (save the rest of the resuspended cells in -20C)
3. Boil cells by heating to 95C for 10 minutes
4. Load 20ul of sample into 4-12% tris gel
5. Prepare the ladder by mixing 5ul of Western C Protein standards (ladder for westerns) and 5ul of Kaleidoscope prestained ladder (multi colors to help look at how well the gel runs/transfers). Don't heat this mixture and load the 10ul into the ladder well of your gel.
6. Run gels at 150V for 1hr (or 200V for 30 min).

C.2 Pre-Equilibration and Transfer

1. Pre-equilibrate gels for 5 - 10 minutes in transfer buffer. (5 min for smaller MW proteins, 10 minutes for other)
2. Soak Whatman paper (2 / gel) in 1x semi-dry transfer buffer
3. Cut PVDF membranes to the size of gel and also trim off the top left corner (to keep track of orientation),
   - Soak in 100% MeOH until opaque / transparent ~1 min
   - Transfer to MilliQ water until it stops streaking ~2 min
   - Soak in 1x semi-dry transfer buffer ~5 min
4. Remove Whatman paper from 1x semi-dry transfer buffer, and place on semi-dry transfer apparatus
5. Make the transfer sandwich by aligning PVDF with the bottom layer of Whatman paper, placing the trimmed off corned in the top left.
6. Using two hands, carefully move the gel onto the PVDF.

7. Using hands, quickly re-wet the top Whatman paper and place on top of gel.
8. Cover the apparatus and run at 15V for 20 minutes.

C.3 Blocking and antibody staining

1. Immerse the PVDF in 5% Milk in TBST (2.5g in 50ml 1x TBST). Place on a nutator for 1 hr at room temp.
2. Prepare the 3xFLAG-HRP antibody 1:88,000 in 5% milk in TBST (0.5ul of antibody in 4.4ml milk/tbst) also add 0.44 ul of Strep-HRP antibody for the Western C ladder. Vortex.
3. Prepare the pouches, add the PVDF, seal two sides and add 2ml of antibody solution to the front of the PVDF.
4. Leave on the nutator for 1 hr at room temperature (or overnight in the cold room).
5. Wash 3x in TBST (15 min each)
6. Add 2ml of immobilon hrt mix to the top of the pvdf. Pipette it over the pvdf for about a minute.

C.4 Developing

Developing on a Bio-Rad ChemiDoc gel imager

1. Place the PVDF on the imaging stage
2. Open the Image Lab software and select “Blot” then “chemi high sensitivity”
3. You can have it auto-detect exposure time or set your own

Developing on an X-ray developer

1. Put the PVDF on the Kodak developing cassette.
2. Make sure you turn the machine on 15 minutes before use to warm up. (Put the lid on, turn on the water, turn on the button on the left.)
3. When developing, turn off all the lights and turn on the red lamp.
4. Take out a sheet of film and place it opposite the PVDF on the cassette.
5. Do 1s, 5s, 10s, 30s exposures. Press the button on the right of the developer to get it to roll the film in. (For a <1s exposure, use your hand to press film directly onto the PVDF.)

C.5 Recipes

Protein Loading Dye Recipe

- For 50 mL
- 12.5 mL glycerol
- 3.75 g SDS
- 15.5 mL 0.5 M, Tris pH 6.8 (or 7.75 mL dH2O + 7.75 mL 1.0 M Tris)
- Pinch of bromophenol blue (10 mg)
- Bring volume up to 50 mL
- May need to add some water before adding SDS to get the SDS soluble. Watch out! SDS takes up lots of volume!
- After you make the main solution, add 5 uL of BME for every 95 uL of dye before use (do this in the hood, it's an organic solvent).

Lysis Buffer

- 50 mM Tris (0.302 g in 50mL)
- 150 mM NaCl (0.420 g in 50mL)
- 0.1% Triton X-100 (50ul in 50mL)
- pH 8.0

10x SDS - PAGE Running Buffer

- 30.3g Tris-HCl
- 71.3g Glycine
- 5g SDS
- Or, you can just buy the 10x and dilute down to 1x.

Semi-dry Transfer Buffer

- 10x Semi-dry transfer buffer stock (1 L)
- - 58.1g Tris HCl
- - 29.3g Glycine
- - 10g SDS

1x Semi-dry Transfer in MeOH (20%)

- - 50ml 10x buffer
- - 100ml MeOH
- - 350ml ddH20

10x TBS

- For 1 L:
- - 80g NaCl
- - 2g KCl
- - 30g Tris-HCl
- - 700ml H20
- * Adjust pH to 8.0 with HCl
- * Adjust volume to 1L

1X TBST

- for 1L:
- - 100ml 10x TBS
- - 900ml ddH20
- - 1ml Tween-20

Appendix D

PROTOCOL FOR CREATING AN EVEN LAWN OF E.COLI USING TOP AGAR

Used for diffusion experiments in Chapter 4. Adapted from the methods of Basu, S., Gerchman, Y., Collins, C. H., Arnold, F. H., & Weiss, R. (2005). A synthetic multicellular system for programmed pattern formation. Nature, 434(7037), 1130-1134. doi:10.1038/nature03461

Materials

- Use agarose and media to create a 1% stock solution of agar in media
- Ultrapure low melting temp agarose from Life (16520-050) (It may not be necessarily to use the expensive agarose, I just used it because it was already in the lab).
- The low melting temp agarose melts at 65C and doesn't solidify until 25C
- Cell growth media (M9, MOPS, LB etc) + antibiotics
- Overnight culture of cells (Measure the OD first)
- Pre-warm agar plate that you're going to add the top agar to (pre-warm in 37C so that the plate isn't cold and the top agar doesn't solidify instantly).

Protocol

- Re-melt the 1% agar stock solution in the microwave (1min, mix every 10 sec).
- Combine the agar, fresh media, and cell culture in a falcon tube such that
- The final agar concentration is 0.33%
- The final cell OD600 is 0.15
- The total volume is 3mL (Make up the difference with fresh media + antibiotics)
- Pipette 1 - 2 ml of the mixture on top of the pre-warmed agar plate so that the surface is evenly covered but there isn't extra liquid sloshing around. The low agar concentration means that you have time before it solidifies so no need to rush.
- Let cool for a little bit at room temp until agar solidifies.

- Making plates with bacterial lawns for spatial patterning
    1. Make M9 agar plates (1% agar). These can be made anytime ahead of the experiment and kept at 4C.
    2. Incubate culture of cells overnight.
    3. Dilute overnight culture 1:100, allow to outgrow for 2-4 hours.
    4. Make 1% top agar in M9 (no antibiotics) with ultra pure low melting point agar (ThermoFischer, #16520050) by microwaving it.
    5. Mix top agar with liquid culture to obtain final cell OD 0.15 with 0.35% agar.
    6. Pipette 2ml of the top agar + cell mix on a pre-warmed M9 agar plate and then remove 1ml once evenly spread over the plate.
    7. Let cool at room temperature for 1 hr.
    8. Punch holes for inducers using a biopsy punch (or just a cut off pipette tip in a pinch)
    9. Add inducers at maximum concentration (arabinose 20%, aTc 20mg/ul) into the holes until filled – the volume will be about 50ul.
    10. Place the plate in 37C incubator with the lid facing up for about 2 hours (just so the inducer won't drip out, but may not be a huge problem).
    11. Flip plate over to prevent condensation, and incubate at 37C overnight.
    12. After 24 or 48 hours, image the place on blue light source or with the stereoscope.

Appendix E

MAKING CHEMICALLY COMPETENT CELLS

Adapted from protocol developed by Wes Whitaker.

Preparation:

- (Two days before) Streak cells onto LB agar plate and grow overnight
- (One day before) Inoculate cells from plate into a 5mL LB culture and grow to saturation overnight.

Pre-chill the following (on ice or in -20C):

- Rotor (Set the centrifuge to 4C)
- 25mL pipet for resuspension
- Two 50-ml Falcon tubes instead
- 50-100 .6mL Eppendorf tubes
- TSS solution (on ice)
- 200ul tips

Protocol:

- Dilute the overnight culture 1:100 in 100mL LB
- Outgrow to OD 0.5–0.7 at 37C (~2–3 h)
- Divide all 100mL into the two pre-chilled 50ml Falcon tubes.
- Spin Falcon tubes at 3500 rpm for 5 min in pre-chilled centrifuge.
- Do the rest of the protocol on ice or at 4C (I fill an autoclave tray with ice and stick everything in it)
- Resuspend in 5 mL ice-cold TSS
- Add 1mL KCM
- Aliquot into 60uL/tube (Just stick the little aliquots into the ice)?
- Freeze in liquid N2 or directly place in -80C for storage. (I keep aliquots on ice until I'm all done and then dump them in a box and put them in the -80C.)

TSS Recipe

- 10% PEG-3350
- 5% DMSO
- 20mM $MgCl_2$

For 500mL TSS:

- 50g PEG
- 25mL DMSO
- 10mL $MgCl_2$ (190mg)
- 500mL LB
- Sterile filter

KCM Recipe:

- 0.5M KCl (74.55g/mol)
- 0.15M $CaCl_2$ (110.98g/mol)
- 0.25M $MgCl_2$ (95.211g/mol)

For 10ml KCM:

- 372mg KCl
- 166mg $CaCl_2$
- 238mg $MgCl_2$
- 10ml MilliQ water
- Autoclave

Transformations:

1. Thaw 60ul pre-aliquoted chem. competent cells on ice.
2. Add 1ul plasmid (1–10ng/ul) (or 10ul of ligation mix, or 2-3ul of Gibson reaction)
3. Gently flick to mix
4. Incubate on ice for 10 minutes
5. Heat shock at 42C for 90 s
6. Add 100uL 2YT or SOC (or even LB) and incubate for 30-90min depending on needed efficiency (may not be necessary for Carb/Amp plasmids) ?
7. Plate on LB/Agar with antibiotics, grow overnight at 37C

Appendix F

LIQUID HANDLING ROBOT PROTOCOLS

These protocols were used for Hsiao, Hori, et al., In press, and written by Sean Sanchez, our robotics lab technician. The liquid handling robot used is the STARlet by Hamilton Robotics interfaced with a BioTek H1MF plate reader for incubation and measurement.

F.1 Step induction for event detector populations

Protocol name: *Timed_induction.med*

Materials

- 96 well Matriplate (600ul volume)(Brooks Automation, Inc., MGB096-1-2-LG-L)
- 40ml of M9CA media in 50ml Falcon tube
- 16ml of eVHed07 cells at OD0.3 in 50ml Falcon tube (measure the OD using in a matriplate with BioTek)
- (Dilute overnight culture 4 hrs ahead of experiment to allow recovery)
- 600ul Eppendorf of Arabinose inducer at concentration of 1% (0.01% final)
- 600ul Eppendorf of aTc inducer at concentration 20ug/ml (200ng/ml final)

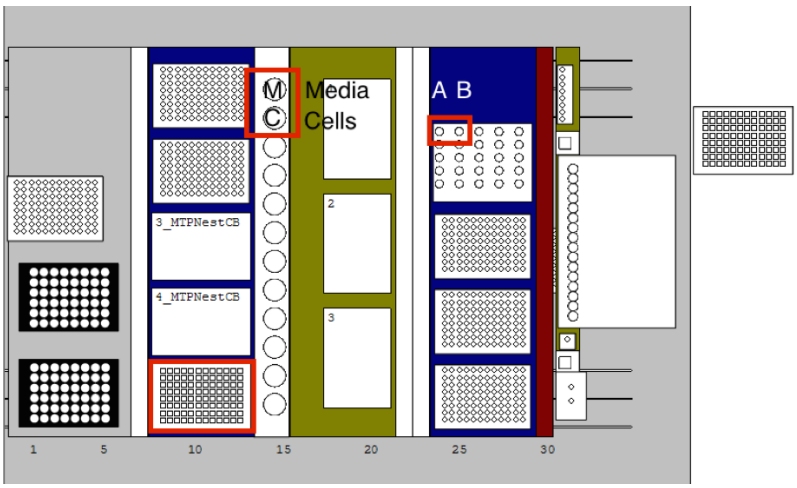


Figure F.1: Hamilton robot deck layout.

Setup (Figure F.1)

- Media and Cells go in the falcon tube holders
- Inducers A (arabinose) and B (atc) go in the chiller rack
- Empty Matriplate goes in the row 10, facing forward
- Now you are ready to run the protocol (green play button)

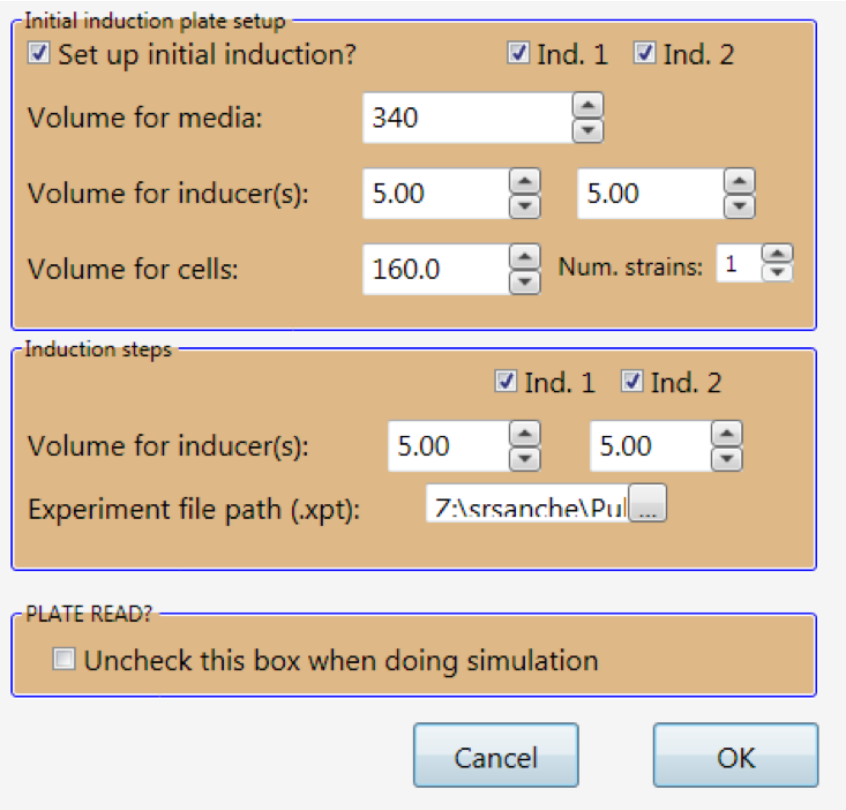


Figure F.2: Hamilton robot step input volumes.

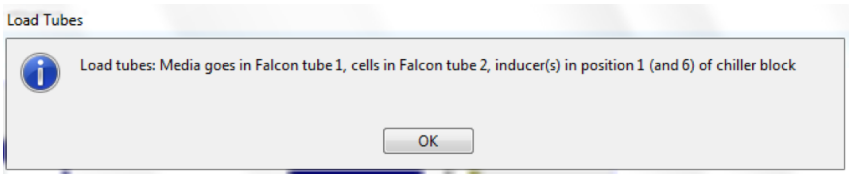


Figure F.3: Hamilton robot step input prompt.

Running the protocol

- The protocol will first add 340 ul of M9 media to each well (will take about 20 minutes)

- It will then add inducer A to the top four rows and inducer B to the bottom four rows
- Then it will add 80ul of cells to each well by sucking up 80ul x 12 per channel and dispensing down the row
- Then it will add *another* 80ul of cells to each well (this is because it can?t do 160ul for all the wells at a time)
- Then it will add the opposite inducer to Column 3 (the Δt = 0 h row)
- Then it will put the plate in the BioTek plate reader (prewarmed to 37C) and the plate reader will shake for 10 minutes before reading.
- The plate reader will read to the Gen5 experiment file called *VH_Robot test - 2 inducer.xpt*

F.2 Pulse induction for event detector populations

Protocol name: *Pulse Induction.med*

- This robot protocol will induce cultures in a 96 well matriplate that are growing in the biotek at 37C
- At 0h, 30s, 1/2/3/4/5/6 hours it will take 5ul samples from the main culture and dilute it into a 96 well deep well plate that contains fresh media with inducer A. This will simulate a pulse of inducer B
- Starting OD 0.06 (OD IS KEY) – use 500ul in a matriplate to test the OD before you begin
- Final arabinose concentration is 0.01%
- Final aTc concentration is 200ng/ml

Materials

- 96 well matriplate (600ul volume)
- 100ml M9CA media
  - 14ml of M9CA with no inducer
  - 85ml of M9CA + 0.01% arabinose (42.5ul of 20% stock in 85ml)
- 50ml 0.2um sterile filtered MilliQ water
- Starting culture of eVHed07 (diluted overnight in the morning to 0.3 and allowed to recover for 4-6hrs)
- Two 1ml deep well 96 well plates
- 600ul Eppendorf of arabinose inducer at concentration of 1% (0.01% final)

  - Dilute stock 1:20 (30ul in 600ul)
- 600ul Eppendorf of aTc inducer at concentration 20ug/ml (200ng/ml final)
  - Dilute stock 1:100 (6ul in 600ul)

Setup for dilution plates B1 and B2

- See Figure F.4
- Add 500ul of M9CA with NO INDUCERS to the first three columns of plate B1 (use multi-channel pipette with media reservoir)
- Then add 500ul of M9CA + 0.01% arabinose to all the other wells in B1 and B2

Setup for matriplate with main cultures

- See Figure F.5
- Dilute starting eVHed07 culture such that the starting concentration is 0.066 and total volume is 500ul (use BioTek to check – this is *critical* because measurements on the nanodrop will be slightly different).
- Fill all other wells with 500ul of filtered water (this maintains humidity and decreases evaporation of cultures).

Robot protocol

- See Figure F.6 to find where all the things go on the robot deck
  - Inducers in the chill rack
  - Matriplate in the back facing the plate reader
  - Plate 1 in the middle facing the plate reader
  - Plate 2 on the outside facing the plate reader
- Set the inducer volumes for the robot to 5ul of each so that they?re diluted 1:100
- Check the box that says *No for simulation*
- Figure F.7 shows time table with which cultures get sampled and diluted.

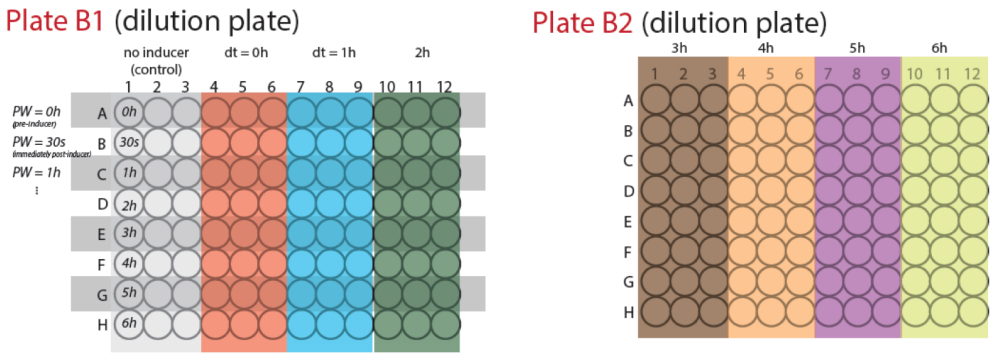


Figure F.4: Pulse exp. dilution plates. These are 96-well deep well plates with 500ul of fresh M9CA + 0.01% arabinose that 5ul of the cultures get diluted into. These plates incubate for an additional 24 hrs at 37C after all cultures have been sampled. Populations in these plates are what get measured as the final populations.

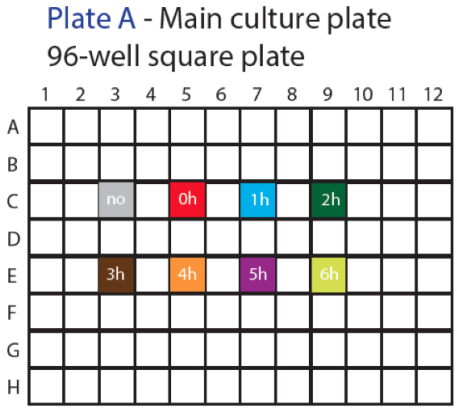


Figure F.5: Pulse experiment culture plate setup on matriplate.

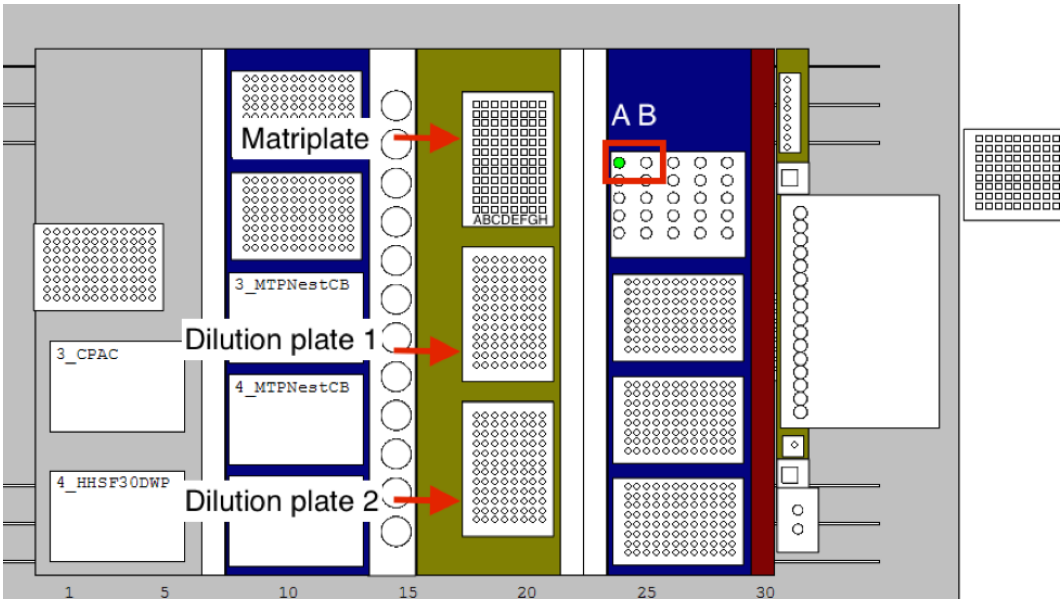


Figure F.6: Pulse exp. Hamilton robot layout.

Table 1 (Time table for adding inducer B)

	no inducer	dt = 0 h	dt = 1h	dt = 2h	dt = 3h	dt = 4h	dt = 5h	dt = 6h
t = 0 h		add inducer B						
t = 1 h			add inducer B					
t = 2 h				add inducer B				
t = 3 h					add inducer B			
t = 4 h						add inducer B		
t = 5 h							add inducer B	
t = 6 h								add inducer B
t = 7 h								
t = 8 h								
t = 9 h								
t = 10 h								
t = 11 h								
t = 12 h								
t = 13 h								

Table 2 (Time table for taking out 5ul samples)

	no inducer	dt = 0 h	dt = 1h	dt = 2h	dt = 3h	dt = 4h	dt = 5h	dt = 6h
t = 0 h	sample pw = 0	sample pw = 0						
t = 30s	sample pw = 30s	sample pw = 30s	sample pw = 0					
t = 1 h	sample pw = 1	sample pw = 1	sample pw = 30s	sample pw = 0				
t = 2 h	sample pw = 2	sample pw = 2	sample pw = 1	sample pw = 30s	sample pw = 0			
t = 3 h	sample pw = 3	sample pw = 3	sample pw = 2	sample pw = 1	sample pw = 30s	sample pw = 0		
t = 4 h	sample pw = 4	sample pw = 4	sample pw = 3	sample pw = 2	sample pw = 1	sample pw = 30s	sample pw = 0	
t = 5 h	sample pw = 5	sample pw = 5	sample pw = 4	sample pw = 3	sample pw = 2	sample pw = 1	sample pw = 30s	sample pw = 0
t = 6 h	sample pw = 6	sample pw = 6	sample pw = 5	sample pw = 4	sample pw = 3	sample pw = 2	sample pw = 1	sample pw = 30s
t = 7 h			sample pw = 6	sample pw = 5	sample pw = 4	sample pw = 3	sample pw = 2	sample pw = 1
t = 8 h				sample pw = 6	sample pw = 5	sample pw = 4	sample pw = 3	sample pw = 2
t = 9 h					sample pw = 6	sample pw = 5	sample pw = 4	sample pw = 3
t = 10 h						sample pw = 6	sample pw = 5	sample pw = 4
t = 11 h							sample pw = 6	sample pw = 5
t = 12 h								sample pw = 6

Figure F.7: Pulse exp. sampling time table.

INDEX