Title
Designing Genetic Circuits for Memory and Communication

Author
Chen, David

Publication Date
2014

Peer reviewed|Thesis/dissertation

Designing Genetic Circuits for Memory and Communication

By

David Chen

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Joint Doctor of Philosophy
with University of California San Francisco

in

Bioengineering

in the

Graduate Division

of the

University of California, Berkeley

Committee in Charge:

Professor Adam P. Arkin, Chair
Professor Hana El-Samad
Professor David F. Savage

Fall 2014

Designing Genetic Circuits for Memory and Communication

by

David Chen

Abstract

Designing Genetic Circuits for Memory and Communication

by

David Chen

Joint Doctor of Philosophy with University of California San Francisco

in Bioengineering

University of California, Berkeley

Professor Adam P. Arkin, Chair

The goal of synthetic biology is to allow the rapid design of organisms that can find diverse uses in environmental remediation, chemical production, or human health. Genetic engineering has traditionally been done by trial and error, but synthetic biology seeks to apply engineering principles and build complex circuits by rationally composing genetic parts together. We envision engineering microbes to form spatial communities for applications such as programmable tissues. We present novel circuit designs for memory and communication, which are basic building blocks for programming these behaviors. Our memory device uses molecular sequestration instead of cooperativity that was used in almost all previously built synthetic switches. In addition, our design allows predictable tuning of the switching boundaries and enables the rapid design of custom bistable switches that can function as a set-reset latch.

We also present designs for contact-based communication by utilizing the recently discovered contact-dependent inhibition (CDI) system. Such a communication channel could allow programmed spatial features with micron-scale resolution, which can be advantageous compared to existing communication methods that rely on diffusible molecules. We present two strategies for harnessing the CDI system. In the first method, we fuse a small transcriptional activator to the protein that is delivered during CDI. In the second method, we exploit the known biology that the delivered domain can co-localize two other proteins. We use this co-localization effect to trigger an increase in activity from a split enzyme, and design an ultrasensitive response to the small number of molecules delivered during the CDI process. While we were not able to show control of gene expression in touching *E. coli* cells, we believe that our circuit designs can guide future engineering efforts.

Table of contents

Index of Figures

Acknowledgements

I want to start by thanking my family. My parents have supported me since I was born and have encouraged my academic pursuits. I also want to thank my younger brother William, who I overlapped with during my first four years in Berkeley while he did his undergraduate studies.

I also want to thank the many members of the Arkin Lab and others on campus who have helped me during my time here. When I did my first rotation here, I learned the basics of synthetic biology from Julius Lucks and Lei Qi. Gavin Price and Josh Hug were extremely helpful when discussing the sequestration-based bistability project. Weston Whitaker provided a number of plasmids I used for cloning, and I want to thank Tim Hsiau for advice with cloning. I also want to thank Esteban Toro for being available to discuss ideas for the contact-based communication project. Will Holtz was also very helpful when thinking about the CDI project. Kelly Wetmore served as the *de facto* lab manager during my time here and helped ensure that the lab ran smoothly, and I also want to thank Gwyneth Terry for taking care of any administrative issues.

Finally, I want to thank Adam Arkin for supporting me the past six years as I learned the art and craft of doing science. Genetic engineering is a young field with tremendous opportunities and challenges, and Adam has been invaluable in understanding this landscape.

Preface

Biology is not usually viewed as a programmable platform. Biological systems are complicated with many moving parts. Many of the important pieces are still not identified today and of the parts we do know the identities of, we often do not fully understand their function and mechanism. Despite these difficulties, I believe we will eventually treat living cells as computable platforms, with which we design circuits and have them execute predictably inside a living cell.

For inspiration, I look to one of the first people to use the title of programmer. As Edsger W. Dijkstra graduated from his doctoral program, no one knew what it meant to be a programmer; when Dijkstra was married in 1957 and forced to list his occupation, the government rejected his use of "programmer" and forced him to list himself as "theoretical physicist". In that era of computing, the importance of programming was not appreciated. As Dijkstra noted in his 1972 ACM Turing Award lecture, "the preoccupation with the physical aspects of automatic computing is still reflected in the names of the older scientific societies in the field, such as the Association for Computing Machinery." When the users were too busy just keeping the machines running, it was difficult for most users of that time period to imagine our ability to design and execute the amazingly complex software programs we have today.

The design of biological circuits will likely follow the same trajectory. We are still hamstrung today by worrying about the physical aspects of construction (ie, cloning), operation (keeping the cells happy), and debugging (probing what's going on inside a cell requires a tremendous investment). Despite the challenges inherent in biology, we sought to push forward our ability to design complex circuits in living cells.

1. Introduction

In the 1970's, we discovered enzymes that allowed us to cut-and-paste DNA and manipulate the building blocks of life in a directed fashion [1]. This gave birth to the era of recombinant DNA, allowing breakthroughs in diverse areas from human health to agriculture. These tentative steps towards genetic engineering shared an empirical approach that was driven by trial and error, with an emphasis on accomplishing the current task and not on learning design rules that could guide future engineering efforts [2].

By the end of the 20th century, there was a recognition that genetic engineering was too laborious and slow. Enabled by the plummeting costs of DNA synthesis, the new practitioners in genetic engineering called their field synthetic biology and sought to bring traditional engineering concepts into the design of genetic circuits [3]. The key idea of synthetic biology is that we should have genetic parts that can be used as building blocks for different applications. Along with the concept of reusable parts are the two closely related ideas of abstraction and standardization [4,5]. Abstraction means that we specify what functionality the parts provide, but we don't specify the underlying details about the implementation. Standardization allows us to swap in different parts as long as they provide the same functionality. These ideas assume that biology contains modules, pieces whose functionality can be separated from the context [6,7]. If biology can be broken down into reusable modules, biological circuits can be designed in a rational manner, sharing techniques such as used in designing electrical circuits [8].

The first *in vivo* demonstrations for the idea of composing genetic parts together were published in early 2000. Gardner and Collins used a simple ordinary differential equation (ODE) model to design an architecture with existing genetic parts that could result in a bistable system [9]. This device, also referred to as the toggle switch, resembles the electronic SR latch and can serve a similar purpose of storing memory. Simultaneously, Elowitz published his implementation of an oscillator (usually referred to as the repressilator), which also composed genetic elements to build a dynamical device that was consistent with the expected behavior [10].

Despite these successes, we are still far from deploying engineered microbes in the full spectrum of applications that were originally envisioned. Part of the problem is that we simply do not have enough parts that we trust will work as desired. The parts should be reliably independent so that they do not interfere with other parts or the host cell, even in different environments [11]. The parts must also have the property of orthogonality, which means that similar parts do not cross-talk with each other. Orthogonality is easily accomplished in electrical circuits by laying out the wires and elements so that parts do not physically interact; in the context of a genetic circuit where molecules are freely diffusing, orthogonality can be much more difficult to engineer but can be accomplished by carefully considering how the molecules interact [12].

A core capability of engineered microbes is the ability to differentiate into different specialized roles. Programs with complex logic rules will require many tunable switches that are independent and orthogonal. Much of the effort in synthetic biology has been the characterization of combinatorial and sequential logic gates that can be used to control bio-molecular activity in response to external inputs such as environmental changes [13,14]. Previously demonstrated memory-based logic gates relied on proteins with specific mechanistic requirements that are difficult to tune. By utilizing a different method for generating the ultrasensitivity required in a bistable memory device, we present a design that is tunable in a predictable manner. Our memory device should be scalable and enable the rapid design of many switches with specified properties.

Based on the decisions enabled by the memory elements, we can control which cells specialize into their different roles by using cellular communication. Programmed communication in *Escherichia coli* cells has historically been done with quorum sensing molecules, which are diffusible molecules that allow the control of gene expression in target cells as a function of the density of the sender cells. While quorum sensing has been used for many purposes such as controlling cell density and sensing and killing pathogenic bacteria [15,16], greater spatial control can be achieved if communication is based on contact. Using a recently discovered contact-dependent mechanism for *E. coli* cells to inhibit growth in their neighbors, we propose circuit designs that enable contact-based communication.

In this dissertation, we present detailed investigations into novel memory and communication devices. We provide an overview of the known biology and previous work in the field. As appropriate, we model the designed circuits with ordinary differential equations, using techniques from nonlinear dynamics [17]. We show the experimental results, and we discuss if they are consistent with the designed behavior and their broader ramifications. Finally, we conclude the dissertation by reviewing how our results help make the engineering of biology easier.

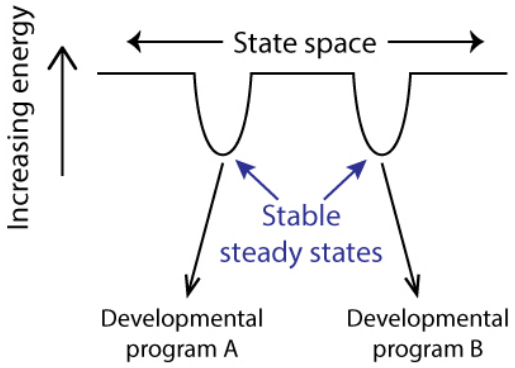
2. Sequestration-based bistability

2.1. Introduction to bistability in biological systems

2.1.1. Key properties of bistability

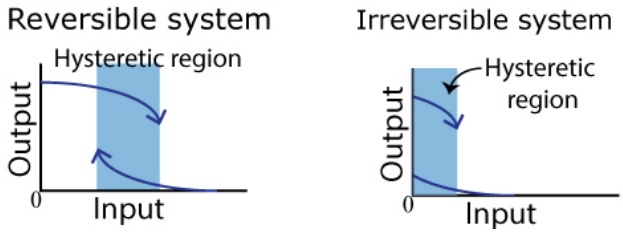
A system is called bistable if there are two stable steady states that a system can permanently reside in. If the cell is forced into either state, then the cell remains in its new state. Figure 1 shows a bistable system conceptually on an energy landscape; we can envision a ball that rolls on this energy landscape, and the ball is trapped in either of the two wells that represent stable steady states. Bistability allows a cell to lock-on to a developmental program because it remains trapped in the specified state.

Figure 1. Conceptual overview of bistability. Shown is a system with two stable steady states and each leads to a different developmental program.



The other property that immediately arises from bistability is hysteresis. This is the region where the system exhibits a memory-dependence, i.e. the system can be in two different states for the same input level and which one is chosen depends on the history of how the system entered the region. The hysteretic region provides a buffering region so that slight variations from noise do not result in drastic changes in output [14,18]. As long as the noise is not so strong that it exceeds the hysteretic region, then the system should only switch state when expected to by a large change in input. In addition, the bistable system can be designed to be irreversible if we move the hysteretic region to the left so that the left side occurs at negative concentrations. Since negative concentrations are impossible, the system can only do one transition (illustrated in Figure 2).

Figure 2. Conceptual example of an irreversible bistable system. If one branch of the hysteretic region occurs in a negative concentration regime, the system becomes irreversible. In the plot on the right, the system can only transition to the lower output value.



Bistable circuits are critical in natural developmental pathways, both for their ability to reduce noise-driven chatter and for their ability to force irreversible transitions [19–22]. Recently, there has also been interest in building synthetic switches for use in engineered circuits. By combining logic gates with memory devices, engineers can design novel circuits for applications such as tumor-killing bacteria by integrating sensors for cell density and oxygen concentration [23,24].

2.1.2. Requirements for building a bistable switch

Bistability requires a positive feedback loop, either directly or indirectly, as shown in Figure 3. This requirement was recently proven by Soule [25]. The second requirement is that the system contains a source of ultrasensitivity; ultrasensitivity can be interpreted as a sigmoidal response curve. We can gain some intuition for this requirement by looking at a simple system where a protein x produces more of itself with a production function $f(x)$ and is degraded with a standard first-order rate $\gamma * x$. Figure 4 shows an example on the left where the production rate is not ultrasensitive and we observe that there are two steady states, and only one of them is stable (indicated with filled circle). On the right, we show an example of an ultrasensitive function using the commonly used Hill function with Hill coefficient of 2 [26]. The key qualitative difference is that the production curve is buffered at low values of x, which allows another stable steady state to emerge. Ultrasensitive functions generally form a sigmoidal S-shape, which allows a third intersection of the production and degradation functions.

Figure 3. Bistability requires a positive feedback loop. Shown is a direct positive feedback loop in the form of autoregulation and an indirect positive feedback loop via mutual inhibition.

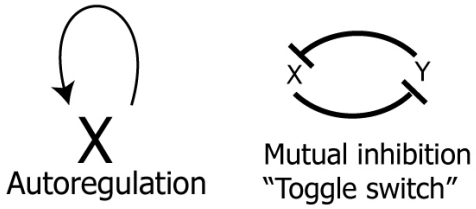


Figure 4. Intuition for why ultrasensitivity is required for bistability. The production curves are shown in blue with their corresponding production rate ($f(x)$ where $K_m = 1$), and the degradation curve is shown in purple ($\gamma = 1$). The intersection of the two curves indicate steady states, i.e. where $\frac{dx}{dt} = 0$. Filled circles represent stable steady states and unfilled circles correspond to unstable steady states.

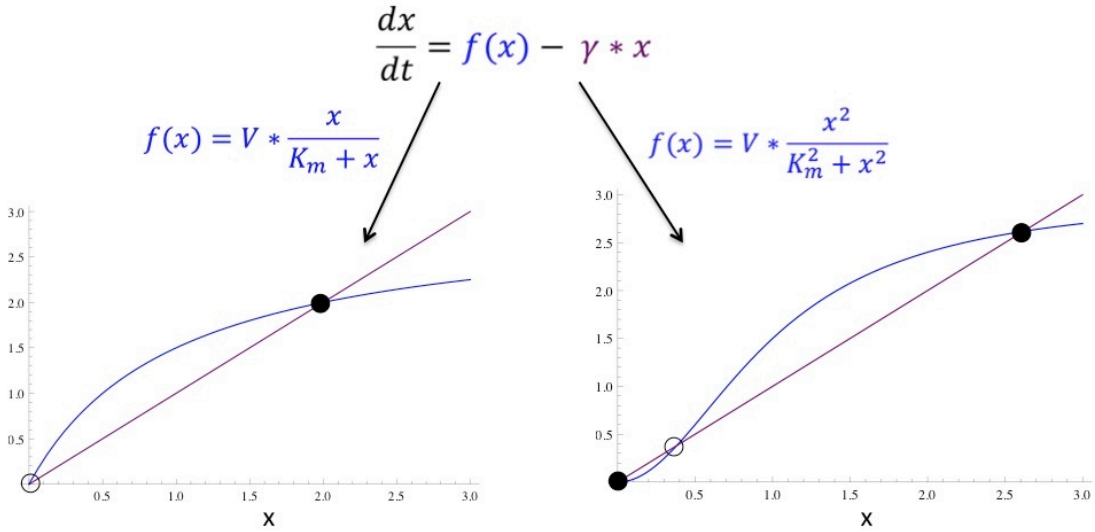
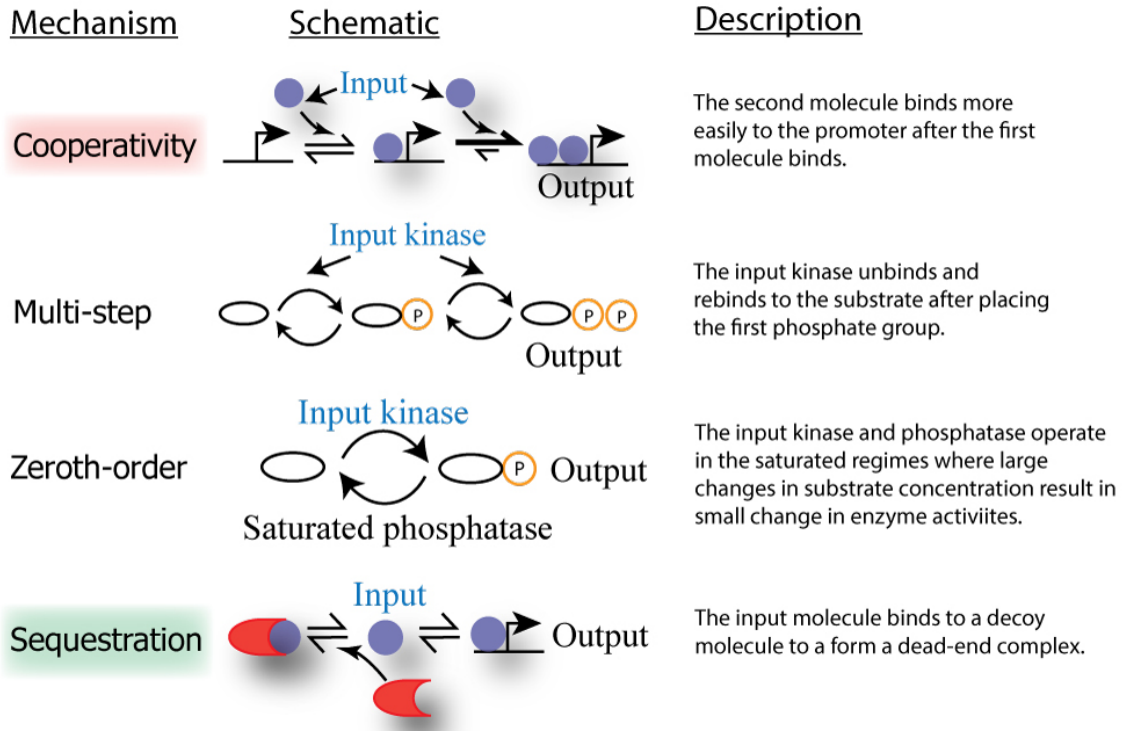


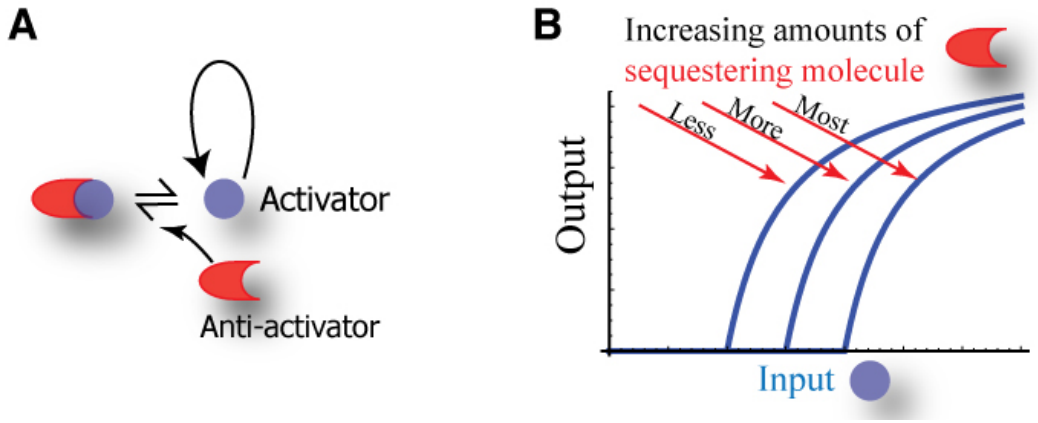
Figure 5 provides graphical descriptions for a number of proposed sources of ultrasensitivity [27–29]. Despite the diverse possible methods for generating ultrasensitivity, almost all *in vivo* synthetic systems built to date rely on molecular cooperativity [9,30–33]. Some possible exceptions are discussed in the Discussion section.

Figure 5. Examples of sources of ultrasensitivity. Cooperativity, which is the mechanism most commonly used in synthetic circuits, is highlighted in red. In our bistable circuit, we use sequestration, which is highlighted in green.



Recent work has shown sequestration's importance in natural systems, and sequestration-based ultrasensitivity has been demonstrated synthetically in eukaryotic systems [34–38]. Nonetheless, no one has yet demonstrated *in vivo* that sequestration with positive feedback is sufficient to build a bistable switch. Figure 6 provides an overview of how we can utilize the properties of molecular sequestration. Panel A shows a schematic of the bistable circuit; the key architectural feature is that we have an activator molecule that produces more of itself and that there is an anti-activator molecule that binds tightly to the activator to form a dead-end complex. Panel B illustrates that the response curve from sequestration can be predictably adjusted by varying the expression level of the anti-activator, and we demonstrate how this property can be utilized to predictably tune the switching boundaries.

Figure 6. Exploiting molecular sequestration **a,** Schematic of our bistable circuit where the blue circle represents a sigma factor that activates itself and red indicates an anti-sigma factor that binds to the activator to form a dead-end complex. **b,** Intuition for why sequestration allows predictable tuning of the response curve.



2.2. Results

2.2.1. Model for sequestration-based bistable circuit

Before we show the experimental results, we first model the circuit shown conceptually in Figure 6. We make the following assumptions that simplify the modeling process:

- Transcription is much slower than protein binding (enables the pseudo-steady-state approximation)
- Gene expression is controlled only at the transcriptional level or by sequestration of the activator
- All molecules degrade at the same rate
- Well-mixed interior of the cell
- Molecule numbers are continuous and system is deterministic

We denote ***x* as the concentration of total sigma factor (activator)** and ***y* as the concentration of total anti-sigma factor (anti-activator)**. We describe the rate of change of total sigma factor concentration with the following ordinary differential equation:

$$\frac{dx}{dt} = basal + V\frac{free\_sigma\_factor}{K_m + free\_sigma\_factor} - \gamma * x$$

The first term represents basal production of the sigma factor that is independent of sigma factor concentration (i.e., from the inducible Ptet promoter in Figure 14). In the second term, we use a hyperbolic function for gene expression consistent with non-cooperative binding of the sigma factor to the promoter; V is a lumped term representing the rate of maximal protein production from the feedback loop and K_m is the concentration of free sigma factor for half-maximal production. Finally, we use a first-order loss term (γ includes both dilution and degradation).

Using the pseudo-steady-state approximation, we can use the relationship previously derived in the literature where K_d represents the binding disassociation constant between sigma factor and anti-sigma factor [27,35,37]:

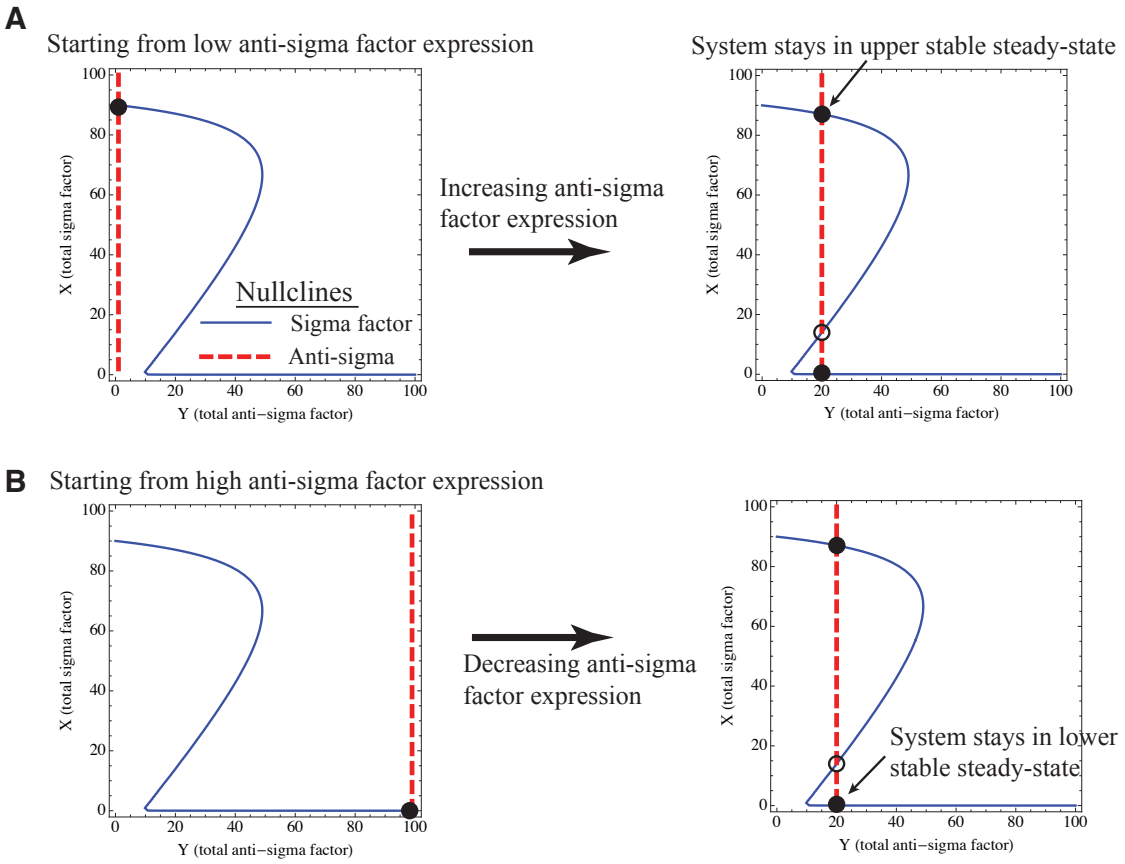
$$free\_sigma\_factor = x - K_d - y + \sqrt{(x + K_d + y)^2 - 4xy}$$

Note that $(x + K_d + y)^2 - 4xy$ can be rewritten as $(x - y - K_d)^2 + 4 * x * K_d$. Plugging in this expression into the ODE and non-dimensionalizing by K_d and γ and using "hat" to denote non-dimensionalized variables, we obtain:

$$\frac{d\hat{x}}{d\hat{t}} = \alpha + \beta * \frac{\hat{x} - 1 - \hat{y} + \sqrt{(\hat{x} + \hat{y} + 1)^2 - 4\hat{x}\hat{y}}}{\kappa + \hat{x} - 1 - \hat{y} + \sqrt{(\hat{x} + \hat{y} + 1)^2 - 4\hat{x}\hat{y}}} - \hat{x} \qquad \text{Equation 1}$$

In the final non-dimensionalized equation, $\alpha = \frac{basal}{\gamma * K_d}$ represents the production rate from the aTc-inducible promoter, $\beta = \frac{V}{\gamma * K_d}$ is the maximal rate of sigma factor production from the positive feedback loop, and $\kappa = \frac{2K_m}{K_d}$ is the half-maximal concentration for production from the feedback loop. With the non-dimensionalized system, we can gain some intuition for why this system provides bistability from the nullclines shown in Figure 7.

Figure 7. Nullclines illustrate why hysteresis is expected as anti-sigma factor expression is varied. The solid blue line corresponds to the nullcline for total sigma factor concentration ($\frac{d\hat{x}}{d\hat{t}} = 0$) and the dashed red line is a vertical line corresponding to the total concentration of anti-sigma factor. **a,** The system starts from a monostable ON state when there is low anti-sigma factor expression and shows that the system stays at the nearest stable steady state (filled circles represent stable steady states and open circles represent the unstable steady state). As anti-sigma factor expression is increased, the anti-sigma factor nullcline (shown with a dashed red line) moves to the right. The system stays in the nearest stable steady state. **b,** The system starts from a monostable OFF state at high anti-sigma factor expression. We decrease anti-sigma factor expression to the same final value as in Panel A, but now the system rests at a different stable steady state. Parameters: $\beta = 100, \kappa = 20, \alpha = 0.01$.



The equation describing our bistable system can be solved for the three roots analytically, but the solutions are unwieldy in the general case. In the limit of $\alpha = 0$, the solutions are substantially simpler, and we are able to gain some intuition over the roles of β and κ. In this limit of no extra sigma factor induction ($\alpha = 0$), we compute the three roots to be:

$$x = 0 \qquad \text{Equation 2}$$

$$x = \frac{2\beta(-4+\kappa) + \kappa(2 + 2y - \kappa - \sqrt{4y^2 + (-2-2\beta+\kappa)^2 - 4y(-2+2\beta+\kappa)})}{4(-2+\kappa)} \qquad \text{Equation 3}$$

$$x = \frac{2\beta(-4+\kappa) + \kappa(2 + 2y - \kappa + \sqrt{4y^2 + (-2-2\beta+\kappa)^2 - 4y(-2+2\beta+\kappa)})}{4(-2+\kappa)} \qquad \text{Equation 4}$$

The first solution represents the OFF state of the switch, the second solution represents the unstable branch, and the final equation is the ON value of the switch.

We can use these equations to find the values of y that cause the saddle node bifurcations that mark the boundaries of the memory region. The left boundary of the hysteretic region occurs when Equation 3 (the unstable branch) crosses from negative values to positive values. We find that the left bifurcation occurs at

$$y = \frac{2\beta - \kappa}{\kappa} \qquad \text{Equation 5}$$

We immediately see that this is positive as long as $2\beta > \kappa$ (note that $\beta > 0$ and $\kappa > 0$). We can locate the right boundary of the memory region since this is when the term contained under the square-root of Equation 3 or Equation 4 becomes negative. This occurs when $4y^2 + (-2 - 2\beta + \kappa)^2 - 4y(-2 + 2\beta + \kappa) = 0$, which is a quadratic equation for y. We find the right bifurcation to occur at

$$y = \frac{1}{2}(-2 + 2\beta + \kappa - 2\sqrt{2}\sqrt{-2\beta + \beta\kappa}) \qquad \text{Equation 6}$$

This implies the restriction that $\kappa > 2$ for $\sqrt{-2\beta + \beta\kappa}$ to be real. Since $\kappa = 2\frac{K_m}{K_d}$, the $\kappa > 2$ restriction sets a limit on the binding affinity versus the affinity of the sigma factor to the promoter (ie, if the binding affinity is too weak, K_d increases and κ no longer satisfies our condition). When $\kappa < 2$, this circuit design can not result in bistability. Before we investigate the implications of Equation 5 and Equation 6, we also need to check that the values for x are physically realizable at the saddle node bifurcations.

The left bifurcation always occurs at $x = 0$, so there is no restriction here. On the other hand, we need to check that the right saddle node meets at a positive value for x. We substitute Equation 6 into the expression outside the square-root of the numerator of Equation 3 and Equation 4, $2\beta(-4 + \kappa) + \kappa(2 + 2y - \kappa) > 0$, and find that $\beta > \frac{\kappa^2}{2(-2+\kappa)}$ is our new restriction that is always stricter than the previously found $\beta > \frac{\kappa}{2}$. Thus, the maximal rate of production from the feedback loop must be strong enough to have a positive ON value. At low κ that is just slightly greater than 2, it becomes extremely difficult to satisfy $\beta > \frac{\kappa^2}{2(-2+\kappa)}$.

Now that we can restrict the valid domain for β and κ, we can investigate how the left and right boundaries are shifted as we change β and κ. We take partial derivatives of Equation 6 to analyze the right boundary and find that $\frac{\partial y}{\partial \kappa} < 0$ and $\frac{\partial y}{\partial \beta} > 0$ in the restricted domain. Since Equation 5 shows that the left boundary scales as $y \sim \frac{2\beta}{\kappa}$, we see that both the left and right boundaries move to increasing y as β increases and both boundaries move to decreasing y as κ increases. In the limit that $\beta \gg \kappa$, both boundaries are linearly proportional to β (ie, for every doubling of β, both boundaries also double).

In addition, we can investigate the dynamic range of our switch, which is the difference of free sigma factor in the ON state versus the OFF state. In the limit of infinite amounts of anti-sigma factor, there is no free sigma factor, so the concentration is 0 in the OFF state. If we take the limit of the ON value (Equation 4) as y approaches 0, we find that $x = \frac{2\beta(-4+\kappa)+\kappa(2-\kappa+\sqrt{(-2-2\beta+\kappa)^2})}{4(\kappa-2)}$ (note that total sigma factor is the same as free sigma factor in this limit of no anti-sigma factor).

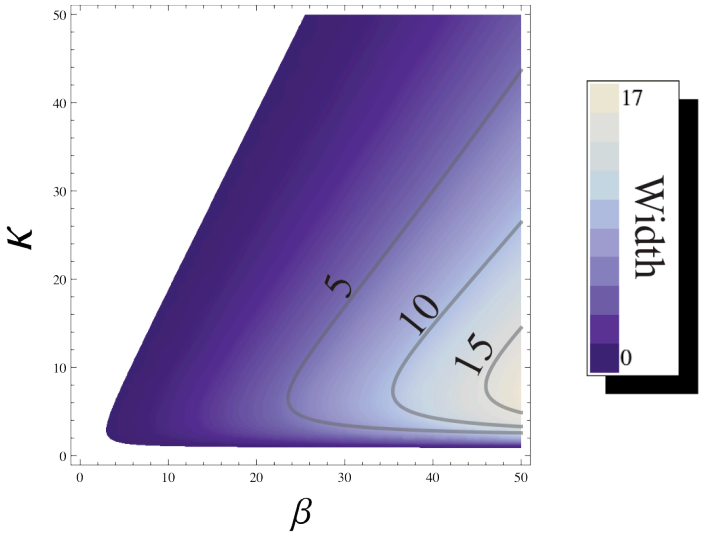
We case on the term within the quadratic $(-2-2\beta+\kappa)$. When this term is positive, we find that $x = -\frac{2\beta}{\kappa-2}$, which is never positive if we satisfy the condition that $\kappa > 2$. Hence, this case is not possible and we check the other case when $-2-2\beta+\kappa < 0$ and find that

$$x = \frac{2\beta - k}{2} \qquad \text{Equation 7}$$

This is the ON value of the switch in the limit of no anti-sigma factor and results in the same restriction as previously found ($2\beta > \kappa$). Thus, we maximize the dynamic range of our switch by maximizing β or reducing κ to its limit of 2.

We define the width of the hysteretic region to be the difference of the left and right boundaries and plot this width as a function of the two parameters. Figure 8 shows that the width continues to increase as β increases. Increasing β, the maximal rate of production from the feedback loop, increases the size of the memory region and the dynamic range of the switch. Both effects were desirable for us to demonstrate the bistable nature of our synthetic switch, which is why we maximized β in our system (cloning details in the Methods section).

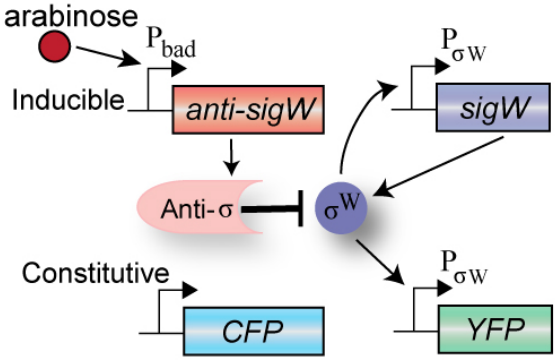
Figure 8. Width of hysteretic region as a function of two parameters. The plot shows the difference between the right and left boundaries, which we define as the width of the memory (hysteretic) region. A few selected contours are also indicated.



2.2.2. Experimental demonstration of hysteresis

We chose to demonstrate our sequestration-based switch in *Escherichia coli* (*Ec*), using the SigW sigma factor and its cognate anti-sigma factor (RsiW) from *Bacillus subtilis* as the activator and anti-activator (see Methods for cloning details and Figure 9 for schematic). To build the positive feedback loop, we placed the sigma factor under control of a SigW-dependent promoter. Expression of the anti-sigma factor was controlled by an arabinose-inducible promoter. Finally, system output was monitored by placing YFP expression under control of the SigW-responsive promoter.

Figure 9. Implementation of sequestration-based circuit *in vivo*. $P_{\sigma W}$ denotes the promoter that the SigW sigma factor recognizes and P_{bad} is the arabinose-inducible promoter that allows us to induce anti-sigma factor.



We call the cells ON when the positive feedback loop is active and YFP is expressed, and cells without YFP expression are considered OFF. The defining feature of bistability is a hysteretic region where the steady-state response of the system depends on the previous state. To experimentally locate this region, we grew cells in the absence of arabinose and at full induction (0.1%) overnight and then stored both sets of aliquots at -80°C. The cells from the aliquots were then inoculated into intermediate arabinose concentrations and allowed to grow for 24 doublings before being imaged by fluorescence microscopy. Figure 10 shows the memory region where the system responds differently depending on whether it was previously ON or OFF. Cells that were previously in full arabinose induction remain OFF in the memory region and then begin to turn ON at 10^{-5} % arabinose (orange arrowheads facing left). Cells that were previously grown in the absence of arabinose are fully ON in the memory region and begin to switch OFF at $8*10^{-4}$ % arabinose (maroon-colored arrowheads facing right). Constitutive CFP expression was used to normalize the YFP expression and eliminate dead cells from the analysis. Figure 11 shows that the distribution of cells in the ON and OFF states shifts at the switching boundaries as expected.

Figure 10. *In vivo* demonstration of hysteresis. Cells contain the circuit shown in Figure 9 using the SigW sigma factor and its cognate RsiW anti-sigma factor. Starting with master cultures that were forced to be fully OFF or ON, we grew the cells for more than 24 doublings at the indicated arabinose concentrations. The cells were imaged on a Zeiss fluorescent microscope. The shaded hysteretic region shows inducer concentrations where the system can be at different states depending on where it was previously. The solid black line shows the stable steady states from the model and the dashed line shows the unstable steady state from the model.

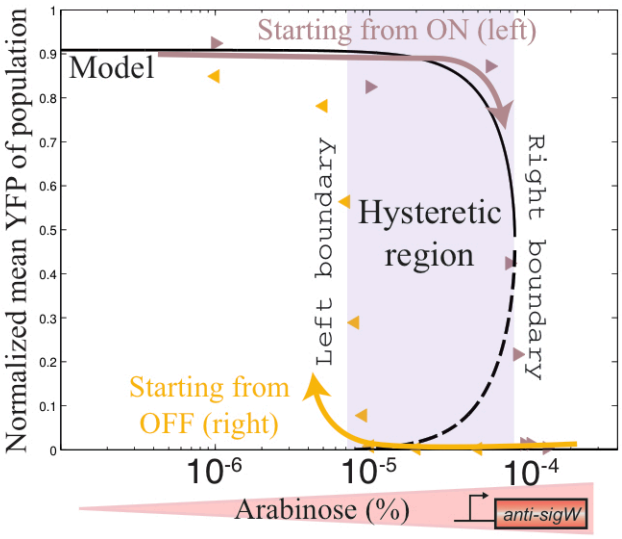
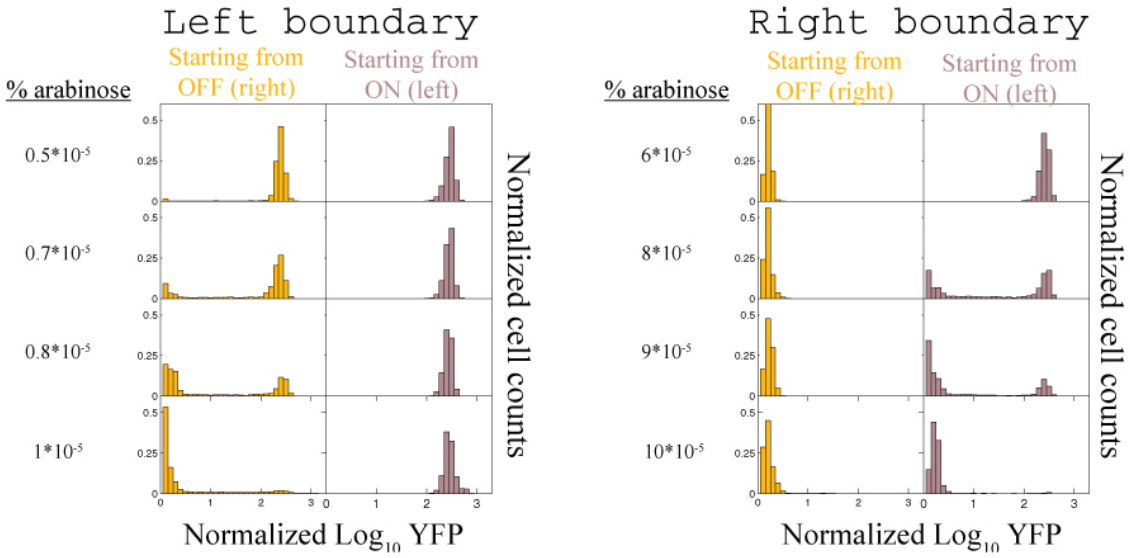


Figure 11. System behaves as expected at the boundaries of the hysteretic region. We show the distributions of the fluorescent proteins at the indicated arabinose concentrations, which correspond to the points shown in Figure 10. At the left boundary, we see that cells that were already fully ON stay ON while the cells that were previously OFF switch ON as we move to lower arabinose concentrations. At the right boundary, the cells that were previously OFF stay OFF while cells that were previously ON switch to OFF as we move to higher arabinose concentrations.



Since we built our device from a pair of *B. subtilis* proteins that bind in a 1:1 stoichiometry, we claim that our device is a demonstration of bistability where molecular sequestration is sufficient to generate the required ultrasensitivity. To further support this claim, we verified that we do not observe other possible sources of nonlinear behavior. Figure 12 shows the response of the PsigW promoter to a linear cascade; as expected, we not observe an ultrasensitive response curve suggesting that the promoter responds to the SigW protein with a Michaelis-Menten-like (hyperbolic) response curve. We also measured the growth rates of the ON and OFF states by measuring OD growth curves and fitting to an exponential growth function (Table 1). We found that ON cells grew approximately 4.5% slower than OFF cells (mean doubling time of 36.7 minutes versus 35.0 minutes). We believe that this small effect on growth rate is not sufficient to permit bistability; for comparison, the circuit based on bistability from growth rate differences by Tan *et al* had ON cells that grew 161% slower than the OFF cells [39].

Figure 12. Absence of anti-sigma factor results in hyperbolic response curve. **a,** Schematic of circuit implemented in APA4311 cells in absence of positive feedback loop. **b,** Cells were inoculated from -80°C stocks stored at an OD600 of 0.1 and grown in M9 with different levels of aTc. Cells were diluted approximately 10-million-fold and examined on a fluorescence microscope after 12 hours of growth. Values represent the mean of each fluorescence reporter over all cells in the population. The points were overlaid with the fit $7 * 10^{-7} + 549 * \frac{cfp^{.8838}}{402.8+\mathrm{CFP}^{.8838}}$. Cells were grown in the absence of arabinose (no anti-sigma factor expressed). As expected, the cells do not exhibit a sigmoidal response in this linear cascade.

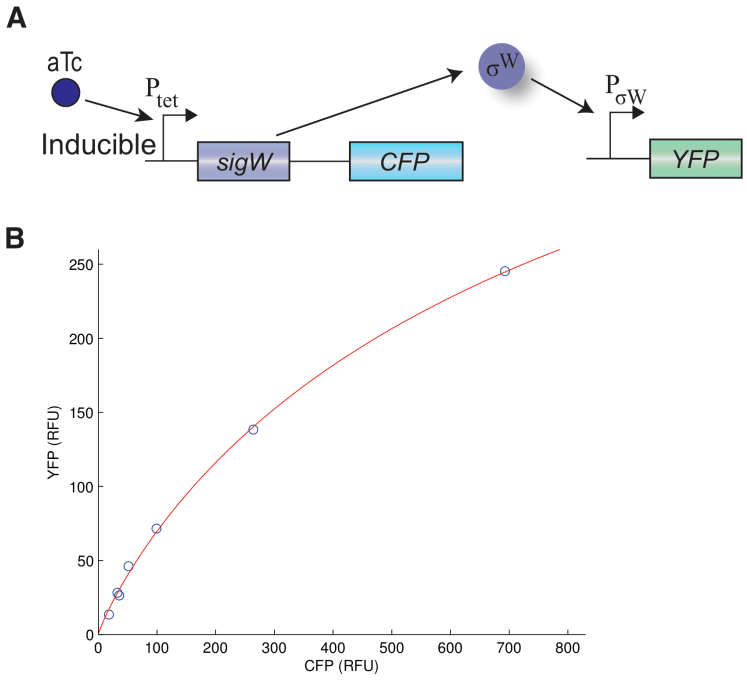


Table 1. Comparison of growth rates of ON and OFF cells for SigW-based switch. Aliquots of APA4309 cells that were previously forced ON and OFF for use during the hysteresis experiments were inoculated into cultures tubes containing no arabinose and $15*10^{-5}$% arabinose. The experiment was performed in triplicate, and the OD600 was measured every 30 minutes. The OD600 values between 0.01 and 0.1 were used to calculate the growth rate using Matlab's curve-fitting function to an exponential form.

	OD measured at time (minutes since first measurement)					**Doubling time (min)**
	0	**30**	**60**	**90**	**120**	
ON, repl. 1	0.0059	0.0117	0.0206	0.037	0.0634	37.3
ON, repl. 2	0.0066	0.0126	0.0223	0.0394	0.0691	36.6
ON, repl. 3	0.0059	0.0114	0.0204	0.0365	0.0645	36.2
OFF, repl. 1	0.0149	0.031	0.0511	0.0941	0.1335	35.4
OFF, repl. 2	0.0083	0.0154	0.0283	0.0503	0.0917	35.0
OFF, repl. 3	0.0138	0.0262	0.0476	0.0865	0.1272	34.6

2.2.3. Characterization of switching rates between states

In our analyses so far, we have only investigated the steady-state behavior of the dynamical system. Ideally, the system holds onto the designated state indefinitely, but in reality, noise can allow the system to switch to the other stable steady-state. We use the model introduced by Nevozhay *et al* to measure the rates of switching between the two states, with the two states allowed to have different growth rates [40]:

$$\dot{N_L} = -r * N_L + f * N_H + g_L * N_L$$
$$\dot{N_H} = r * N_L - f * N_H + g_H * N_H \qquad \text{Equation 8}$$

In our system, the ON cells are also considered the High cells and have the experimentally measured growth rate $g_H = 1.134/hr$. The OFF (Low) cells have growth rate $g_L = 1.188/hr$.

Following the nomenclature used previously by Nevozhay *et al*, we define $R(t) = \frac{N_L}{N_H}$. Using Mathematica to analytically solve this system and simplify the expression for R(t), we obtain

$$R(t) = \frac{f * (e^{a1*t} - 1)(2 * N_{H0} + N_{L0}) + N_{L0}(e^{a1*t}(a1 - g_H + g_L - r) + a1 + g_H - g_L + r)}{a1 * N_{H0}(e^{a1*t} + 1) - (e^{a1*t} - 1)(N_{H0} * (f - g_H + g_L) - r(N_{H0} + 2N_{L0}))}$$

where

$$a1 = \sqrt{f^2 + (g_H - g_L + r)^2 + 2f(g_L + r - g_H)}$$

Note that our solution for $R(t)$ can be shown to be the same as the expression shown in the paper by Nevozhay *et al.*

We substitute $R_0 = \frac{N_{L0}}{N_{H0}}$ into the above equation and obtain

$$R(t) = \frac{(-1 + e^{\mathrm{a1}t})f(2 + \mathrm{R_0}) + (\mathrm{a1} + \mathrm{g_H} - \mathrm{g_L} + e^{\mathrm{a1}*t}(\mathrm{a1} - \mathrm{g_H} + \mathrm{g_L} - r) + r)\mathrm{R_0}}{\mathrm{a1}(1 + e^{\mathrm{a1}t}) - (-1 + e^{\mathrm{a1}t})(f - \mathrm{g_H} + \mathrm{g_L} - r(1 + 2\mathrm{R_0}))}$$

To experimentally measure *r* and *f*, we force the cells fully ON or OFF to begin. Thus, R_0 is initially 0 or infinity. We can simplify the previous equation for these two limits.

In the limit $R_0 \to \infty$ that all cells are in the Low state to start, we simplify $R(t)$ to:

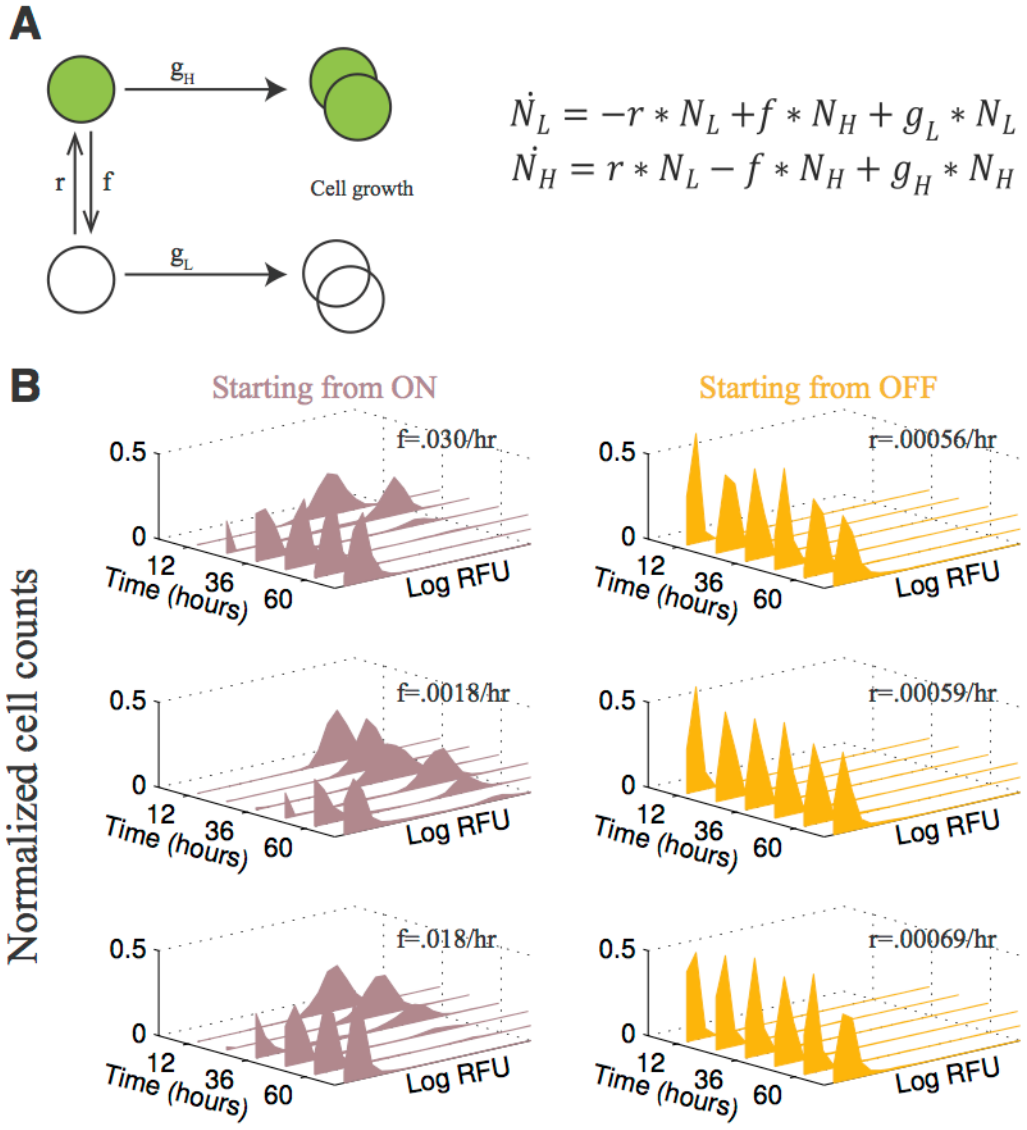
$$\frac{\mathrm{a1}(1 + e^{\mathrm{a1}t}) + (-1 + e^{\mathrm{a1}t})(f - \mathrm{g_H} + \mathrm{g_L} - r)}{2(-1 + e^{\mathrm{a1}t})r} \qquad \text{Equation 9}$$

In the second limit $R_0 \to 0$ that all cells are in the High state to start,

$$\frac{2(-1 + e^{\mathrm{a1}t})f}{\mathrm{a1}(1 + e^{\mathrm{a1}t}) + (-1 + e^{\mathrm{a1}t})(f - \mathrm{g_H} + \mathrm{g_L} - r)} \qquad \text{Equation 10}$$

The experiment is performed by assaying the cells and diluting them back every 12 hours to keep them in exponential growth phase. The parameters *r* and *f* are fit for each time course using Matlab's Nelder-smead fminsearch function starting from 100 different initial values for *r* and *f*. Both *r* and *f* are initialized from a uniform distribution ranging from 0.01 to 1. Within the Nelder-smead routine, the error is computed as the sum of the differences from the computed $R(t)$ using Equation 9 or Equation 10 and the experimentally measured ratio of OFF to ON cells. Finally, *f* is calculated as the median of the 100 optimizations for the time courses starting from ON cells, and *r* is calculated similarly for time courses starting from OFF cells. We find that ON cells switched OFF in the memory region at a rate of .0165 +/- .014 per hour while OFF cells switched ON at a much slower rate of $6.1*10^{-4}$ +/- $7*10^{-5}$ per hour. At the chosen concentration of anti-sigma factor production rate, we observe that the system has an easier time jumping from the ON to the OFF state. This could be disadvantageous in applications that require the cell to hold the ON state for long periods of time. On the other hand, we can make it harder to jump from ON to OFF if we use a lower anti-sigma factor production rate (i.e., a lower arabinose concentration). This would simultaneously make it easier to jump from the OFF to ON state, however, so would be less useful if we wanted to hold the OFF state indefinitely.

Figure 13. Measurement of stochastic switching rates. We quantified the rate of switching between the two states using the model previously described by Nevozhay *et al* [40]. **a,** Cells can switch between the low (OFF) and high (ON) states at two different rates r and f, and the two states can grow at different growth rates g_L and g_H. **b,** We measured the rates *r* and *f* from three replicates where we start at time 0 with cells fully low (OFF) or fully high (ON) and observe the change in the distributions. APA4309 cells were grown in M9 with 4*10^{-5}% arabinose.

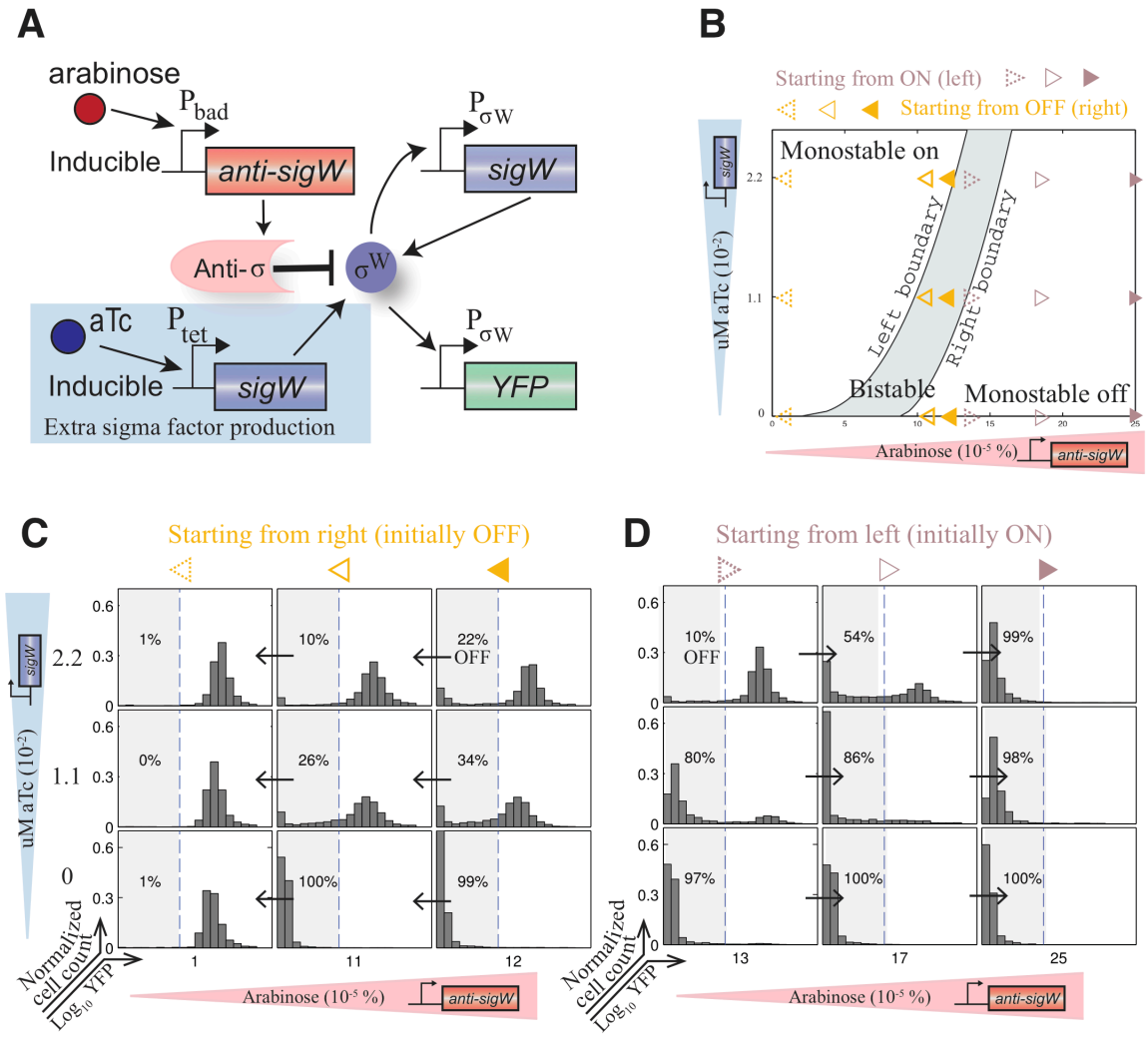


2.2.4. Demonstration of tunable switching boundaries

To tune the switching boundaries of our sequestration-based switch, we added another transcriptional unit with the sigma factor under control of a promoter inducible with anhydrotetracycline (aTc) (schematic shown in Panel A of Figure 14). In this system, the two induced proteins compete with each other. In the limit of high expression of anti-sigma factor and no sigma factor induction, all the cells are turned OFF. In the other limit where the sigma factor is strongly induced and no anti-sigma factor is present, all the cells are forced ON. In between the two extremes, we expect a bistable region whose

boundaries vary with the total concentrations of the proteins. The numerically computed bifurcation diagram for this system is plotted in Panel B. We predict that cells at higher levels of sigma factor (aTc) induction require higher levels of anti-sigma factor (arabinose) induction to cross the boundaries. We experimentally verified this relationship by preparing aliquots of master cultures that were forced to be OFF or ON. Initially OFF cells were grown at the inducer concentrations indicated in Panel B with left arrowheads, and we found that increasing aTc concentration results in switching ON at higher arabinose concentrations, shown in Panel C. Similarly, initially ON cells were grown at the concentrations indicated with right arrowheads, and we found that increasing aTc concentration requires higher levels of arabinose to switch OFF, shown in Panel D of Figure 14.

Figure 14. Demonstration of tunable switching boundaries. **a,** Schematic of the system that is implemented in APA4310 cells that allows aTc induction to control extra sigma factor production. **b,** Bifurcation diagram overlaid with arrowheads denoting inducer concentrations where the system was investigated. Cells initially OFF were induced at the concentrations indicated with left arrowheads, and cells initially ON were induced at the concentrations marked with right arrowheads. **c,** Histograms show that cells initially OFF switch ON at higher arabinose levels with increasing aTc induction. Cells were grown at the indicated arabinose and aTc concentrations (corresponding to left arrowheads from panel b) for 12 hours, and at least 1000 cells were imaged per histogram. The percentage indicates the proportion of cells considered OFF (sum of bars in shaded area). **d,** Cells initially ON require more arabinose to switch OFF with increasing aTc. Arabinose and aTc concentrations correspond to right arrowheads from panel b.

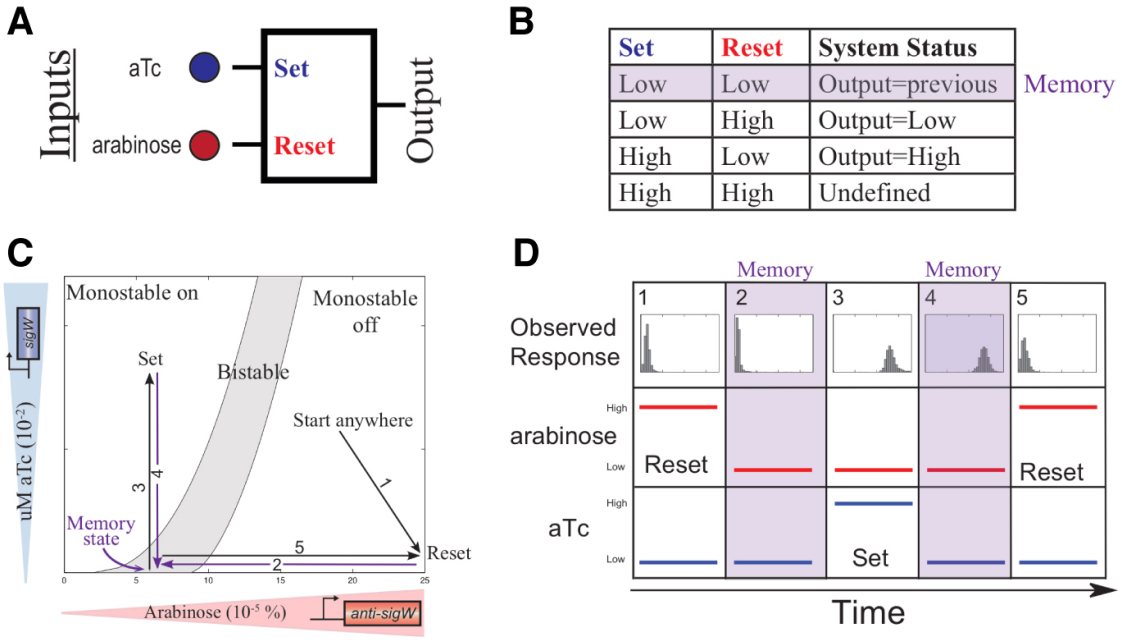


2.2.5. Demonstration of irreversible latch

From the bifurcation diagram, we also predicted that we could design our switch to function as a set-reset latch. The latch is a true memory element rather than a hysteretic switch. It holds the state of the last input: set or reset. Panel A of Figure 15 illustrates that the two inputs to the latch are arabinose and aTc, and Panel B shows the logic table the set-reset latch implements. To demonstrate the latch with APA4310 cells, we sequentially performed the steps illustrated in the bifurcation diagram of Panel C. We first initialized the latch into the reset state by forcing the cells OFF with full induction of anti-sigma factor (step 1 of Panel C and response shown in Panel D). We then inoculated the cells with both inputs set low, so the cells are in the memory state and remain OFF (step 2). While maintaining the same concentration of arabinose, we pulsed the cells into the ON state by growing them in aTc and observed that upon inoculating them into the memory state with both inputs set low, the cells remained ON (steps 3 and 4). Notably, the cells at step 4 are at the same inducer concentrations as used at step 2, but the cells after step 4 are ON compared to being OFF after step 2. After 12 hours in the memory state, we observed that less than 2% of the cells switched to the incorrect state. Finally, we could

reset the latch by pulsing the cells with excess arabinose to force them OFF again (step 5).

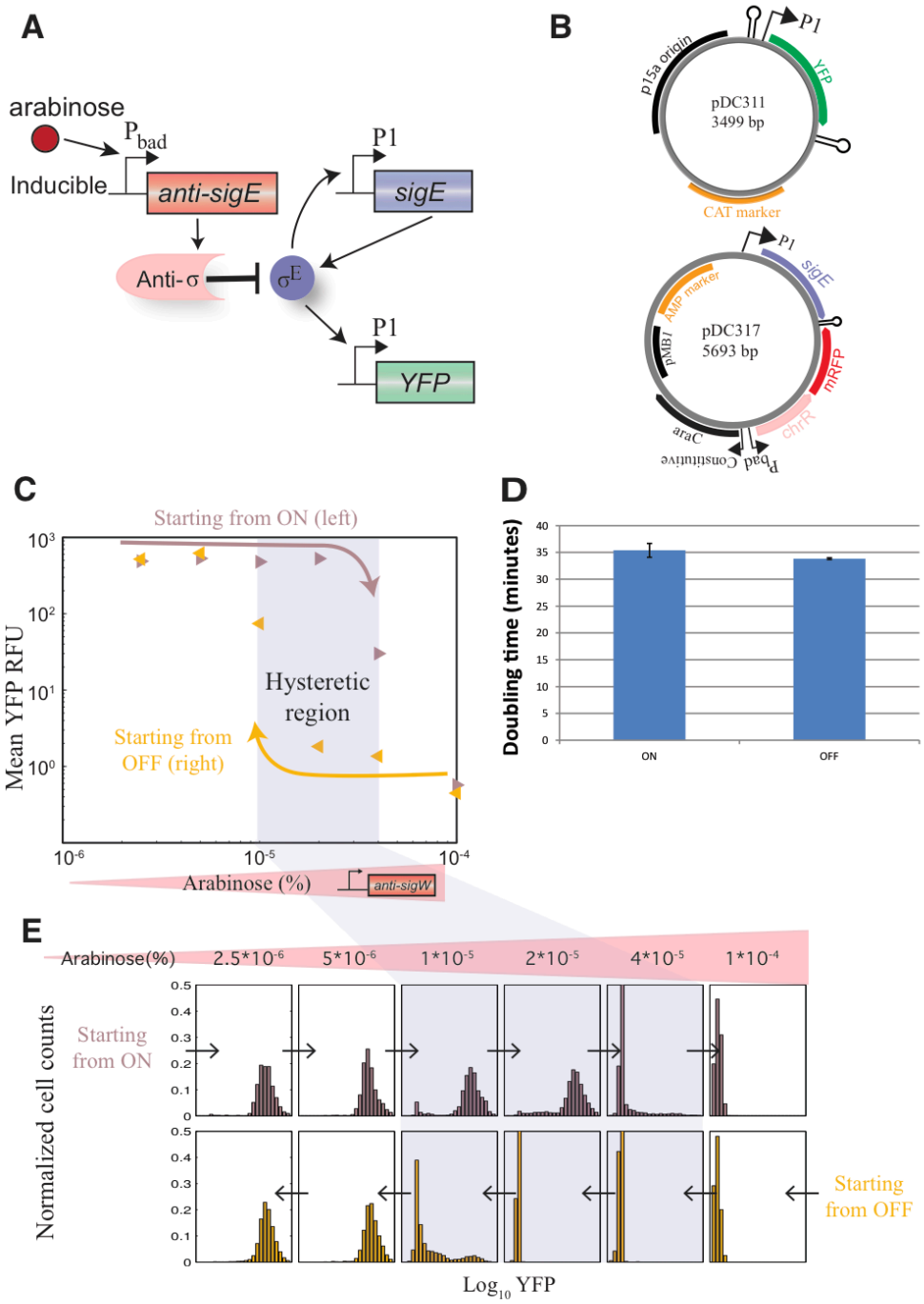
Figure 15. Demonstration of irreversible latch. **a**, Schematic of set-reset latch **b**, Logic table corresponding to set-reset latch **c**, Using the previously determined switching boundaries, a set-reset latch can quickly be designed. Path numbers correspond to observed points in the next panel. **d**, The system retains a memory of the previous set or reset pulse. Steps 2 and 4 were analyzed after allowing the system to grow for 12 hours, and cells were grown in the Set and Reset stages for 9 hours.



2.2.6. Demonstration using *R. sphaeroides* sigma and anti-sigma

A key claim of our design is that it should be easy to design new memory devices since the requirements are quite minimal. To demonstrate that our design is not specific to the properties of SigW, we also demonstrated hysteresis in a bistable system based on SigE and its cognate anti-sigma factor ChrR from *Rhodobacter sphaeroides*. Figure 16 shows the circuit design and implementation and the experimental results. Panel A shows a schematic of the circuit; one other difference is that P1 (which is naturally found in front of *sigE* in *R. sphaeroides* and is recognized by SigE) replaces PsigW. Panel B shows the plasmid maps for the constructed plasmids that correspond to the design of the previous panel. In panel C, we followed the same protocol used for the SigW-based device and prepared master cultures that were forced ON (0% arabinose) or OFF (0.1% arabinose), and then the master cultures were inoculated into intermediate arabinose concentrations. The corresponding histograms are shown in panel E. In panel D, we show the doubling times as calculated from measuring the OD600 at half-hour intervals in triplicate experiments. The ON cells have a mean doubling time of 35.4 minutes versus 33.8 minutes for the OFF cells. The results are consistent with what we observed for our SigW/RsiW-based switch, indicating that our design should be generalizable to different pairs of sigma factor and anti-sigma factor.

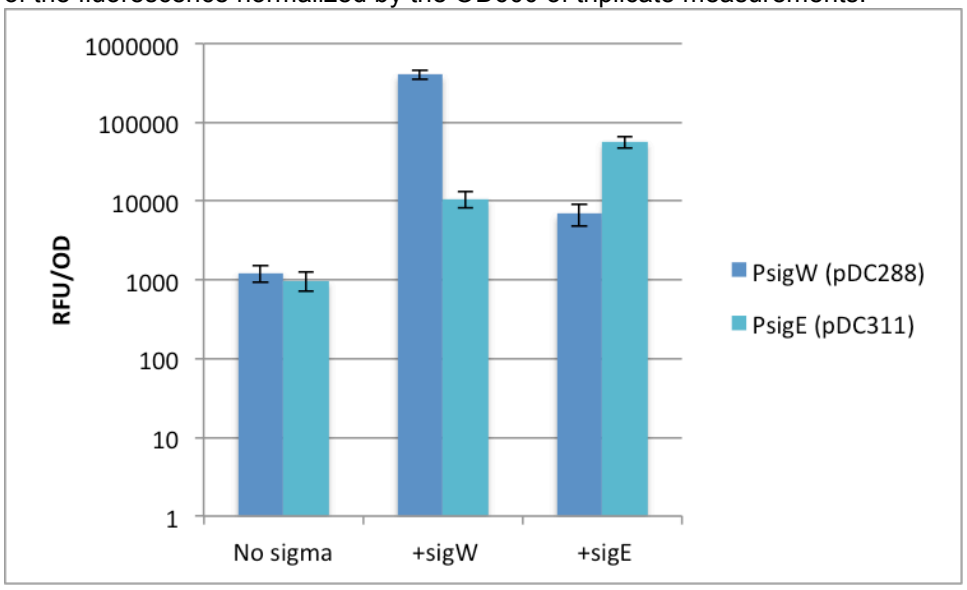
Figure 16. Demonstration of hysteresis using a different pair of proteins. The circuit was built from a pair of sigma factor and anti-sigma factor from *Rhodobacter sphaeroides* and implemented in *Escherichia coli* cells. At least 1000 cells were observed by fluorescence microscopy after 12 hours of growth for the data shown in panels C and E. Detailed descriptions in the main text.



We also did a quick check to determine if we could incorporate both the SigW and SigE-based switches in the same cell. Figure 17 shows fluorescence measurements from a plate reader. BW27783 cells were co-transformed with a medium-copy plasmid containing the reporter for SigW or the reporter for SigE and a high-copy plasmid expressing no sigma

factor, the SigW switch, or the SigE switch. The PsigW promoter produces the most fluorescent protein in the presence of SigW as expected, but the promoter is also somewhat activated by the presence of the SigE sigma factor. The PsigE promoter is also most active in the presence of the SigE sigma factor (although not as strong as PsigW's response to SigW), and is also slightly activated by the SigW sigma factor. The results show that the *B. subtilis* SigW and *R. sphaeroides* SigE sigma factors exhibit some cross-talk in *E. coli* with the promoters we chose to use. Recent work demonstrated other pairs of sigma factors with their cognate promoters that exhibit less cross-talk [41].

Figure 17. Test of orthogonality between SigW and SigE and their promoters. Cells were inoculated from colonies and allowed to grow for 4 hours in M9 in the absence of arabinose. Cells were then assayed on a Tecan Safire2 plate fluorometer. Shown are the mean and standard error of the fluorescence normalized by the OD600 of triplicate measurements.



2.3. Discussion

While bistability from sequestration has been shown *in vitro* [42], we have verified *in vivo* that bistability can be generated without relying on cooperative transcription factors. We note that since the topological requirements for our device are so minimal, the same design was previously found as part of an *in silico* evolutionary search for genetic circuit designs [43]. Recent papers have suggested other methods for building bistability in the absence of cooperativity. Modeling of the MprA/MprB system in mycobacteria illustrated that sequestration of a sigma factor by its anti-sigma factor could play a critical role in allowing bistability in the natural system, and they propose experiments to demonstrate their predictions *in vivo* [44]. Alternatively, nonlinearity in growth rate combined with positive feedback could result in bistability [39], but the *in vivo* synthetic system did not exhibit steady-state hysteresis on the population level, indicating that building memory units based on this design could be problematic. As another possibility, two positive feedback loops without cooperativity or other sources of ultrasensitivity was

theoretically shown to be sufficient to generate bistability [45]. While this is an elegant topology since the requirements are minimal and the design facilitates tuning, the *in vivo* demonstration contained basal ultrasensitivity in one of the feedback loops, so this design remains to be built using pathways without inherent ultrasensitivity.

Biological devices should be evaluated in terms of operating performance and scalability. The key operational feature of our design is that it allows predictable tuning of the switching boundaries by adjusting expression of the sequestering protein. Another advantage is that the simple topology based on positive autoregulation lets us describe the key characteristics with analytic expressions, such as shown in Equation 5 and Equation 6. Thus, if we desire a certain dynamic range or width of hysteretic region for a future application, the analytical results guide us in choosing optimal parameter values. In comparison, the toggle switch is not analytically tractable so the parameters that allow bistability must be determined numerically [9]. One caveat with our system, however, is that the width of the hysteretic region is coupled with the tuning of the switching boundaries (shown in Figure 8).

The scalability of any biological device is limited both by the number of orthogonal parts and by host interaction. We chose to use an extracytoplasmic function (ECF) sigma factor since the ECF sigma factors and non-ECF sigma factors recognize different consensus promoters via distinct mechanisms of promoter binding [46]. There are 43 groups of ECF sigma factors that recognize different consensus promoters, and *Escherichia coli* contains two ECF sigma factors that belong to different groups than SigW's group [47]. Since 35 of the ECF sigma factor groups have cognate anti-sigma factors that should be orthogonal, the strategy presented here should be scalable. We observe a small amount of cross-talk between SigW and SigE and their cognate promoters (Figure 17), so using these systems in the same cell could require understanding the determinants of orthogonality. Recently, Rhodius *et al* characterized at least 20 sigma factors and their cognate promoters that are highly orthogonal inside *E. coli* cells [41]. In addition, these sigma factor-promoter pairs are highly repressed by their cognate anti-sigma factors, but they observe some cross-talk between anti-sigma factors to non-cognate sigma factors. The scalability of our bistable design based on sigma factors and anti-sigma factors may also be limited by host interaction since we observe a mild effect on the growth rate when the *B. subtilis* SigW sigma factor is expressed. These could be resolved in the future by identifying and fixing the parasitic interactions with the host.

Memory modules will be key components of synthetic biology applications in the future, so genetic engineers will need a large palette of different types to choose from. While the genetic toggle switch based on mutually inhibitory transcription factors has been the standard design for a set-reset latch [9], each instance requires two cooperative transcription factors while our design requires only one activator/anti-activator pair per switch. Future work could further compare different properties of the competing designs; for example, the toggle switch was shown to be more robust to growth rate changes than a positive autoregulatory loop [48]. An *in vivo* demonstration was also shown for a T-latch, which features a push-on, push-off behavior [49].

Although we have focused on bistability, memory units can also be built on recombinases [50–52], which may be especially attractive for long-term storage, but these devices lack a hysteretic region that provides buffering and the scalability of these parts is still unproven. A recent design based on recombinases especially demonstrated the stability of the device over 120 doublings, although switching was only 85% complete after a RESET pulse [53]. In contrast, our device is 99% complete in switching after either a SET or RESET pulse of 9 hours, but we find that the half-life of the ON state in the memory region is only 24 hours. Our current device may not be suitable for applications that require long-term stability, but this could be improved in the future; for example, if the mild growth defect were fixed and both states grew at the same rate, the half-life of the ON state should be about 43 hours. The sequestration-based switch should be advantageous for applications requiring scalability and predictable tuning.

2.4. Methods

2.4.1. Cloning Details

The *sigW* and *rsiW* genes were cloned from *B. subtilis* 168 (*Bs*) genomic template using primers o330/o331 for *sigW* and o332/o333 for *rsiW* (oligonucleotide sequences are in Table 3). When we express the anti-sigma factor *rsiW* in *E. coli*, we only use the N-terminal 87 amino acids since this domain normally remains in the cytoplasm, does not contain residues that target RsiW for proteolytic degradation, and was previously found to be sufficient to sequester the sigma factor in *Bs* [54,55]. For the SigW-dependent promoter, we used primers o328/o329 to clone the promoter that drives the natural *sigW*/*rsiW* operon in *Bs* [56] . We observed some leakage from this promoter in the absence of *sigW* expression, but this leakage was removed when we changed the UP element of the promoter to the *E. coli* consensus (using oDC383/oDC384 and the Phusion Site-Directed Mutagenesis protocol). This modified promoter exhibits a 100-fold induction ratio and was used to control transcription of both the positive feedback loop and the YFP reporter.

As explained in the analytical modeling section, increasing maximal sigma factor production rate from the feedback loop (β in the model) increases the width of the hysteretic region and increases the difference between ON and OFF states. We selected for high expression by constructing an RBS library in the 5' UTR region between PsigW and *sigW* (using primers o362/o363) and screened for candidates that exhibited strong fluorescence when co-transformed with a reporter plasmid containing PsigW-GFP. We also created an RBS library for *rsiW* expression (using primers o364/o365) and selected for an RBS that produced enough anti-sigma factor to turn off fluorescence from a strain containing PsigW-GFP and *sigW* being expressed from the Ptet promoter.

The PsigW-sigW and Pbad-rsiW transcriptional units were placed on a high copy plasmid (pMB1 origin) with a strong transcriptional terminator (rrnB terminator) between them. The PsigW-Venus and Ptet-sigW transcriptional units were placed on a p15a plasmid,

again with a strong transcriptional terminator between them. To facilitate debugging of the inducible promoters, RFP was added to the operon controlled by Pbad (using primers o404/o405 and o391/o406) and CFP was added to the operon controlled by Ptet (using o395/o396 and o390/o394) using SLIC [57]. The Venus (YFP) and Cerulean (CFP) reporters were obtained from pZS2-123 [58]. RFP is a codon-optimized mRFP obtained by Prof. J. Christopher Anderson, U.C. Berkeley, from the Parts Registry (part E1010) and was modified at nucleotide positions 6-11 to fix an RBS issue.

The sigE switch based on *Rhodobacter sphaeroides* was cloned using o416/o417 off pBS16 [59]. The sigE/chrR system was previously shown to function in *E coli* [60]. The reporter plasmid pDC311 was created by using o418/o419 off pDC288 to swap the sigW-responsive promoter with the sigE-responsive P1 promoter. Following the same procedure described above for the sigW-based switch, an RBS library was created between the P1-sigE 5' UTR region using o420/o421 and the candidates were screened using pDC311. The *chrR* gene was also cloned from pBS16 using o422/o423 and used to replace *rsiW* in pDC298 using SLIC with o424/o425. The candidate from the RBS library (using o428/o429) and the Pbad-chrR (using o426/o427) operons were combined using SLIC to make pDC317.

Plasmid maps are shown in Figure 18, and Table 2 lists the plasmids and strains described previously. Plasmids were transformed into BW27783, a strain designed to allow homogeneous expression from the pBad promoter [61]. Oligos were ordered from IDT with standard desalting.

Table 2. Plasmids and strains used to demonstrate sequestration-based bistability

Plasmids

Name	Resistance	Origin	Description
pDC296	ampR	pMB1	Contains SigW positive feedback loop and Pbad controlling RsiW and mRFP expression
pDC297	cmR	p15a	Reporter plasmid for SigW with constitutive CFP expression
pDC298	ampR	pMB1	Pbad controlling RsiW and mRFP expression
pDC304	cmR	p15a	Reporter plasmid for SigW and contains Ptet controlling SigW and CFP expression
pDC311	cmR	p15a	Reporter plasmid for SigE
pDC317	ampR	pMB1	Contains SigE positive feedback loop and Pbad controlling ChrR and mRFP expression

Strains

Name	Background	Plasmids
APA4309	BW27783	pDC296+pDC297
APA4310	BW27783	pDC296+pDC304
APA4311	BW27783	pDC298+pDC304
APA4327	BW27783	pDC311+pDC317

Figure 18. Plasmid maps for plasmids constructed for the SigW-based devices. Stem-loop structures denote transcriptional terminators (rrnB terminator).

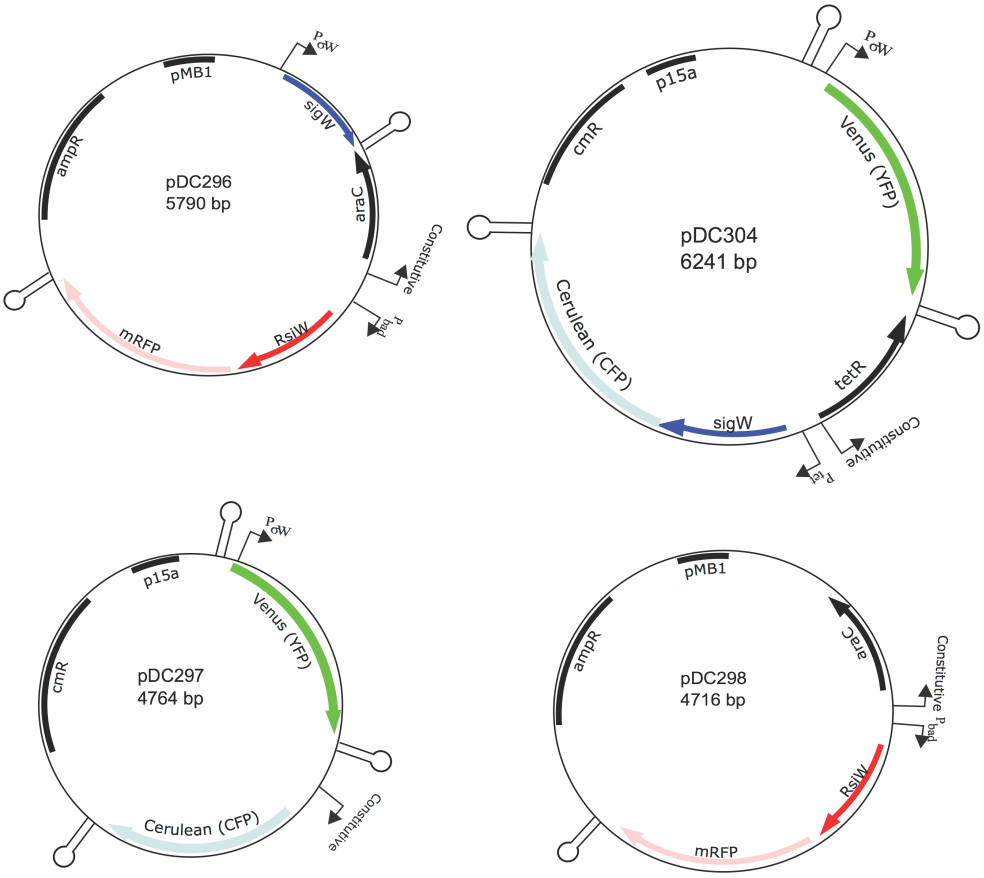


Table 3. Oligonucleotides used for cloning of the sequestration-based bistable device. Oligos that end in lowercase on the 3' side denote annealing region on the 3' side in lowercase, restriction sites in uppercase, and a 5' buffer region in lowercase that is removed upon cutting with the restriction enzymes.

Name	Description	Sequence
o328	PsigW for	actgaGAATTCATGAGATCTTCTACACCCTGCCAAAAATT
o329	PsigW rev	catgtCTCGAGTTAGGATCCTTTATCTAACCTCTGCCTTC
o330	sigW forward	actgaGAATTCATGAGATCTatggaaatgatgattaaaaaaagaattaaacaagtc
o331	sigW rev	catgtCTCGAGTTAGGATCCttaaagatcccttaattgttttctaagagcc
o332	rsiW for	actgaGAATTCATGAGATCTATGAGCTGTCCTGAACAAAT
o333	rsiW rev	catgtCTCGAGTTAGGATCCATGGGTTCTGAACCATCTTT
o362	PsigW Rbs lib rev	NccttccttNNNNaccggTCtgtatgtatacgagc
o363	PsigW Rbs lib for	DDDDDDatggaaatgatgattaaaaaaagaattaaacaagtc

o364	Strong Pbad-rsiW for	AGGNDDDDDDATGAGCTGTCCTGAACAAATTGtgc
o365	Strong Pbad-rsiW rev	ccttNNNNgtcgggagaggtaccgctag
o383	iPCR PsigW rev	ATTTTTGGCAGGGTGTAGAaga
o384	iPCR PsigW for	ggaaaatttttttcaaaagtaAAATTGAAACCTTTTGAAACGAAGC
o390	For End of BB into TrrnB	GGATCCtaaCTCGAGtaatgaga
o391	Rev RsiW	ATGGGTTCTGAACCATCTTTTTACAG
o394	Rev sigW	ttaaagatcccttaattgttttctaagag
o395	For Cerulean	ctcttagaaaacaattaagggatctttaatcttcctggagaaaataaggagg
o396	Rev Cerulean	aagcttctcattaCTCGAGttaGGATCCTTATTTATACAGCTCATCCATGCC
o404	Rev Trrnb with XbaIPstI	ttatgtCTGCAGTCTAGAaattagttcaccgacaaacaacag
o405	mRFP for	CTGTAAAAAGATGGTTCAGAACCCATTAATAAgaaagaggagaaaggatctatggca
o406	mRFP rev	aGGATCCTTACTTGTACAGCTCGttattaagcaccggtggagt
o416	P1-sigE for	atacagaattcatgAGATCTcctgatccagactggc
o417	P1-sigE rev	cattaCTCGAGttaGGATCCTtAACTCATATGCTGGCGCAGA
o418	P1 reporter for	taacgtccTTCTTAcggcggccgggccagtctGGATCAggTATAAACGCAGAAAGGCCCA
o419	P1 reporter rev	agatcacggcaggagtTAAGGAGGAaaaaaaaATGAGCAAAGGTG
o420	RBS Lib for	DDDDDDaTGACGGACAAGAGTGACAGGA
o421	RBS Lib rev	NccttccttNNNNACTCCTGCCGTGATCTTAACGT
o422	chrR for	ACCCGAAAGGAGGTAATGATATGACGATCCGGCACCATGT
o423	chrR rev	ttctcctctttcTTATTATtAGATGCGGAAGAAGGGCT
o424	Vec for	TAATAAgaaagaggagaaaggatctatggca
o425	Vec rev	ATCATTACCTCCTTTCGGGTc
o426	Ins for	ctgttgtttgtcggtgaactTTATTAAGCACCGGTGGAGTG
o427	Ins rev	gaggaagcCTGCAGTCTAGActatgctactccatcgagccgt
o428	Vec for	TCTAGACTGCAGgcttcct
o429	Vec rev	agttcaccgacaaacaacaga

2.4.2. Growth conditions

Cells were grown at 37°C in M9 minimal medium supplemented with glycerol and casamino acids; media contains per liter: 11.28g 5xM9 salt (Sigma M6030), 300 mg thiamine hydrochloride, 8 mL 50% glycerol, 2g casamino acids (BD Bacto), 20 mL of 0.1 M magnesium sulfate, 200 uL of 0.5M calcium chloride, filling to 1L with distilled water and filtering before storage at 4°C. Antibiotics were added as appropriate with carbenicillin at 50 ug/ml and chloramphenicol at 25 ug/ml. To prepare master cultures, cells were grown overnight in 0.1% arabinose (Sigma A3256) or the absence of arabinose to an OD600 of 0.1 and then stored in 10% glycerol aliquots of 50-100 uL at -80°C (8-strip PCR tubes, T-3135-1, from Bioexpress were found to be the most convenient for

storing aliquots). Master cultures for Fig 3b-d were grown in the absence of aTc and 0 or 0.1% arabinose. The experiments were performed by thawing a frozen aliquot at 37°C for a minute and then diluting the cells 16-million-fold into 2 mL of the m9 medium. Cells were grown overnight at 37°C at 200 rpm until an OD600 of approximately 0.1 and then cells were concentrated via a microcentrifuge for microscopic analysis. For the tunable boundary experiments, cells were grown at the indicated concentrations of arabinose and aTc (Clontech 631310). For the set-reset latch experiments, cells were pulsed OFF by growing in 0.1% arabinose and in the absence of aTc. Cells were grown in the memory state in the absence of aTc and $6.3*10^{-5}$ % arabinose. Cells were pulsed ON by growing at the same arabinose concentration and approximately 10^{-3} uM aTc.

2.4.3. Microscopy and image analysis

Cells were imaged on a Zeiss Axio Observer D1 using a 63x plan-apochromat Ph3 oil-immersion objective. Cells were imaged with phase contrast and with the appropriate fluorescence filter set (38 for GFP/YFP, 45 for RFP, and 47 for CFP) with the following exposure times (150 ms for RFP, 50 ms for GFP/YFP, and 150 ms for CFP). Fluorescence excitation was provided with an X-CITE 120XL. Images were captured with an AxioCam MRm using the provided Zeiss Axiovision software.

Images were analyzed using Matlab software, starting with the routines provided with CellTracer [62]. Phase contrast images were first thresholded and then segmented using the CTiterativeSelectiveSegmentation function, which recursively separates the cells in an area of the image. Cells were then discarded if they failed checks on eccentricity or size (removes clumps of cells that could not be segmented as individuals). After segmenting the phase contrast images, the fluorescence images were first flat-field normalized. Then the fluorescence for each cell in the image was calculated. For each cell, the background fluorescence in a 20-by-20 pixel box surrounding the cell was calculated and then subtracted from the mean fluorescence in the area within the cell. Finally, dead cells or debris mis-classified as cells were removed by excluding cells that failed to exhibit any CFP when expected (ie, cells containing pDC304 and grown in aTc induction or cells containing pDC297 with CFP under constitutive expression). For cells not expected to express CFP, cells were excluded if they failed to exhibit any RFP fluorescence. For most images, no more than 1% of the cells after segmentation were excluded as dead.

2.4.4. Other instruments

Bulk fluorescence measurements were performed on a Tecan Infinite M1000 (RFP settings were 584/607 nm for excitation and emission with 5nm bandwidths, 0 μs lag, 20 μs integration, 50 reads, calculated z-position, and gain set to 170) or a Tecan Safire 2 (YFP measured with 510/528 excitation and emissions wavelengths and 5nm bandwidths, 130 gain, 10 reads, 20 μs integration). Optical density measurements at 600 nm for growth curves were performed on an UVmini-1240 (Shimadzu).

2.4.5. Numerical computations

To numerically compute the bifurcation diagrams, we need to determine values for β, κ, α and $\hat{y}$ that correspond to Equation 1. The first two parameters are independent of the two inducers while α is a function of aTc and $\hat{y}$ is a function of arabinose. We can measure the relative amount of $\hat{y}$ by measuring the induction curve of the Pbad promoter as a function of arabinose (shown in Figure 19). We can also measure the relative amounts of α by measuring the Ptet induction curve as a function of aTc (shown in Figure 20).

However, we require additional parameters (i.e., conversion factors and leakage in the absence of induction) to convert from the Pbad and Ptet induction curves to $\hat{y}$ and α. First, we forced the Pbad promoter to be non-leaky since this promoter is known to be fairly tight in the absence of arabinose induction [63]. In contrast, we observed slight leakage from the Ptet promoter since the switching boundaries were shifted to higher levels of arabinose in Figure 14 in the absence of aTc induction compared to the boundaries in Figure 10, which shows cells without the transcriptional unit producing extra sigma factor. We calculated a value of α at 0 aTc that is consistent with the shifted boundaries of Figure 14. We fit the remaining parameters to match the switching boundaries from Figure 10 (when α must be 0) and the boundaries at 0.5*10-2 uM aTc from Figure 14: we numerically optimized for the values for β, κ, the conversion ratio from arabinose induction to $\hat{y}$, and α that corresponds to 0.5*10-2 uM aTc. The values of α at 0 aTc and at 0.5*10-2 uM aTc were then used to convert the aTc dose-response to values of α.

Summary of calculated parameters and dose-response functions:

- $\beta = 1812.9$
- $\kappa = 3106.7$
- $\hat{y} = 0.3161 * 407 * \frac{\text{ara}^{1.746}}{510+\text{ara}^{1.746}}$ where ara is arabinose in units of 10^{-4} %
- $\alpha = 0.13 * (2.024 + 1152 * \frac{\text{aTc}^{.6834}}{1+\text{aTc}^{.6834}})$ where aTc is in units of uM

For Figure 10, we compute the bifurcation diagram by starting at the lowest value of $\hat{y}$ and moving to higher values of $\hat{y}$ while solving for the steady-state value of $\hat{x}$; if we observe a large decrease in the new steady-state value of $\hat{x}$, we identify the current value of $\hat{y}$ as a saddle node that marks the right boundary of the hysteretic region. Similarly, we can identify the left boundary of the hysteretic region by starting at a high value of $\hat{y}$ and identifying a large jump at the new steady state of $\hat{x}$ while decreasing $\hat{y}$. Finally, we can identify the unstable branch by searching for a third steady state between the saddle nodes. For the bifurcation plots shown in Figure 14 and Figure 15, we set β, κ and α values and compute the values of $\hat{x}$ and $\hat{y}$ that correspond to the switching boundaries (ie, saddle-node bifurcations). At the saddle nodes, the rate equation is at steady state, and the production curve is tangent with the loss curve (i.e. Equation 1 is zero and the derivative of Equation 1 with respect to $\hat{x}$ is one at values of $\hat{x}$ and $\hat{y}$ that correspond to a saddle-node bifurcation).

Figure 19. Arabinose induction curve controlling anti-sigma factor production. Points represent mean and standard deviation of four replicates of cells with pDC298 using bulk fluorescent measurements on a plate reader. Points were fit using nonlinear least squares to a Hill function, 32.36+407*ara^1.746/(510+ara^1.746).

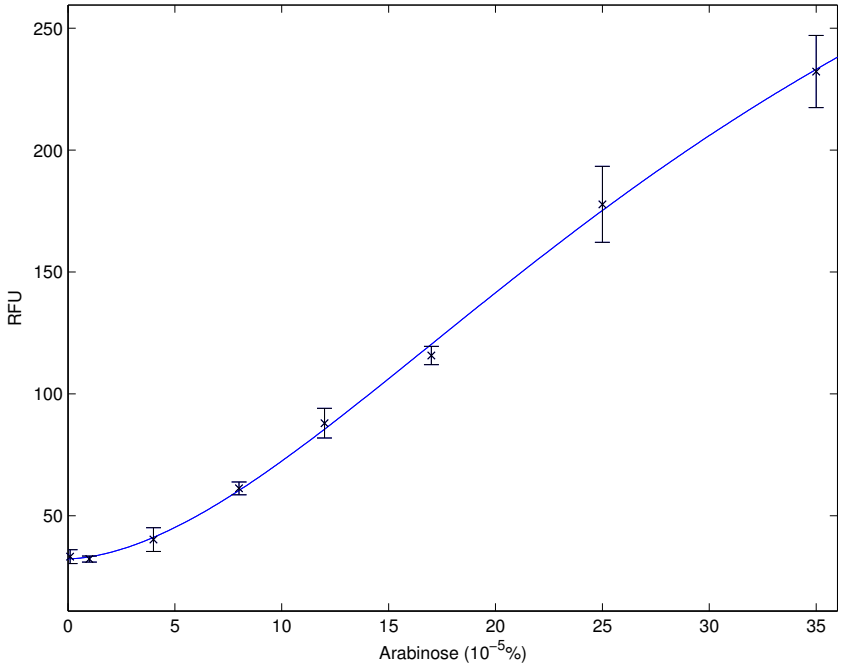
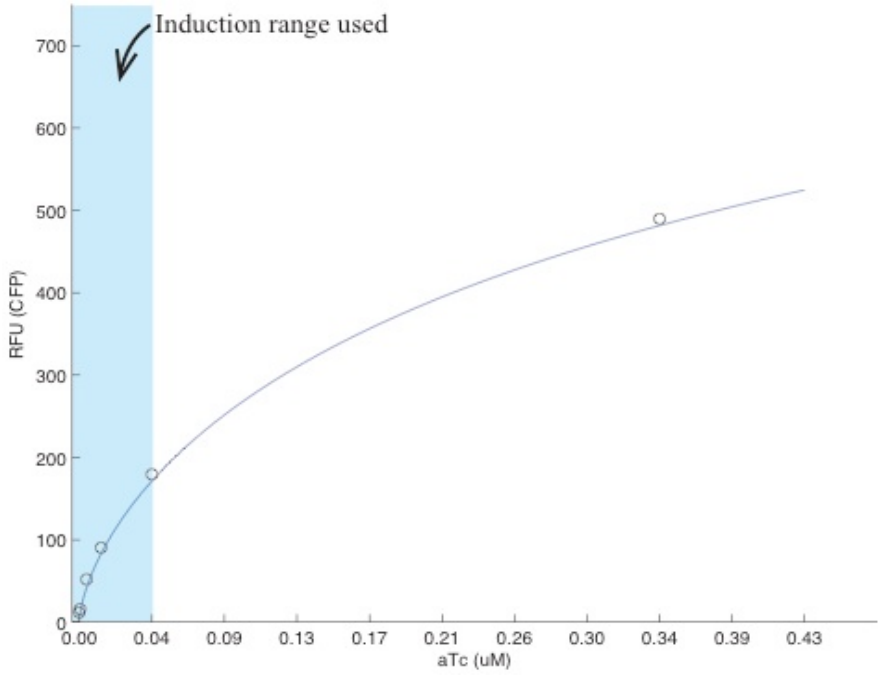


Figure 20. ATC induction curve controlling extra sigma factor production. Fluorescence values were used from the CFP fluorescence values used to plot the x-axis in Figure 12. The dose-response points were fit to the Hill function $10.23 + 1152 * \frac{\text{aTc}^{.6834}}{1+\text{aTc}^{.6834}}$ via nonlinear least squares.



3. Contact-based cellular communication

3.1. Introduction

3.1.1. Why engineer contact-based communication?

Synthetic biologists are interested in cellular communication both for its possibilities in enabling new applications and as a model to study development in natural systems. Inspired by how effective natural microbial consortia are at producing the chemicals they need to stay alive in harsh and diverse environments, researchers are trying to use synthetic consortia for producing metabolic chemicals [64,65]. The key idea is that different cells can specialize at different tasks and then pass their products along to another type of cell for further processing.

Besides metabolic applications, spatial communication can be used for cells to compute logic in a distributed manner [66,67]. This facilitates parts re-use since the same genetic elements can be used in cells that are spatially separated. This type of distributed computing also allows users to modify the computation by placing the cells in different locations. Thus, spatial patterns can be used as inputs to create desired logic computations. Cellular communication can also enable programmed spatial patterns; for example, Basu *et al* created bulls-eye patterns by incorporating a band-pass function [68].

While there are many potential applications that spatial communication can enable, we are also extremely interested in using synthetic communication to understand development in natural settings. Higher organisms utilize long-range signaling for patterning of large fields of cells, but utilize contact-dependent signaling for fine patterning [69]. In contrast, microbes generally perform communication only with diffusible molecules, except for a few exceptions such as in *Myxococcus xanthus* [70]. By engineering contact-based communication in bacteria, we could study what design rules govern different modes of communication. Recent work by Ferreira and Arcak theoretically explored what types of patterns can be formed with contact-based lateral inhibition [71].

We were especially interested in investigating Turing patterns that Alan Turing proposed almost half a century ago [72]. Observing patterns such as stripes and spots on the skins of animals, Turing proposed a possible chemical mechanism that required a slow-diffusing activator and a fast-diffusing inhibitor. While there have been a few natural processes that are consistent with a Turing model [73], systems found from nature are always complicated by unknown interactions. A synthetic Turing model could enable us to truly understand how well Turing patterns can work in forming developmental patterns. Past efforts at implementing Turing patterns have been hamstrung by the limited range in diffusivity of quorum sensing molecules. A contact-based communication

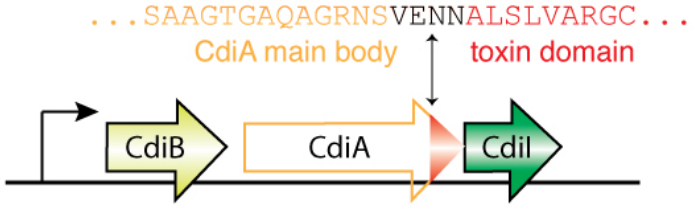
channel provides the equivalent of an extremely short-range (slow-diffusing) activator, which should enable us to build a Turing pattern in combination with a quorum-sensing molecule that serves as the fast-diffusing inhibitor.

3.1.2. Overview of the contact-dependent inhibition system

The contact-dependent inhibition (CDI) system was discovered by David Low's laboratory when they co-cultured a strain of *E. coli* from a rat feces with a standard laboratory strain of *E. coli* [74]. They found that the laboratory strain stopped growing in the co-culture, but this growth inhibitory effect only occurred if the two strains could physically contact each other. Separating the two strains with a mesh size that prevents the *E. coli* cells from passing but permits the media to freely diffuse does not result in growth inhibition. The evidence strongly indicated that this growth inhibition required contact between the cells.

The original growth inhibitory strain was called EC93 and likely works by puncturing a hole in the inner membrane [75]. The members of Low's laboratory found that the CDI system is found in a diverse range of bacteria and characterized the growth inhibitory activity from three more homologues: another strain of *E. coli* called UPEC536 that causes urinary tract infection, a strain of *Yersinia pestis* (the microbe that causes the black death), and a strain of *Dickeya dadantii* [76]. Figure 21 illustrates the modules of the CdiA protein. Aoki *et al* found that the killing domain downstream from the VENN motif can be considered as a module that can be fused to a different CdiA main body (the CdiA protein upstream of the VENN motif). For example, one can fuse the *Y. pestis* toxin domain to the EC93 main body to deliver this toxin into target *E. coli* cells. The toxins that have been shown to work in a fusion to the UPEC536 CdiA main body range in size from 205 to 264 amino acids. Recent work indicates that each main body can only target specific types of species [77]. In addition, the CdiI proteins only recognize and provide immunity to their cognate killing domains; for example, the CdiI from EC93 will not provide immunity to the killing domain of the CdiA from UPEC536 [76].

Figure 21. Overview of the natural CDI system. The contact-dependent inhibition system is naturally found as a 3-gene operon. CdiB serves as a transporter on the outer membrane of the sender cell. CdiA is a large protein with more than 3,000 amino acids with a C-terminus that encodes for a toxin domain (shown as a red triangle) that is delivered to the target receiver cells. CdiI is a small immunity protein of 129 amino acids that prevents growth inhibition of cells containing the CDI operon. The CdiA protein consists of a main body to the left of the VENN motif and a toxin domain that is downstream of the VENN motif.



Members of Low's lab and Christopher Hayes' lab further characterized the UPEC536 CDI system and found that it functioned as a tRNAse [78]. The tRNAse activity of UPEC536 requires expression of the CysK protein, so target cells with *cysK* deleted are

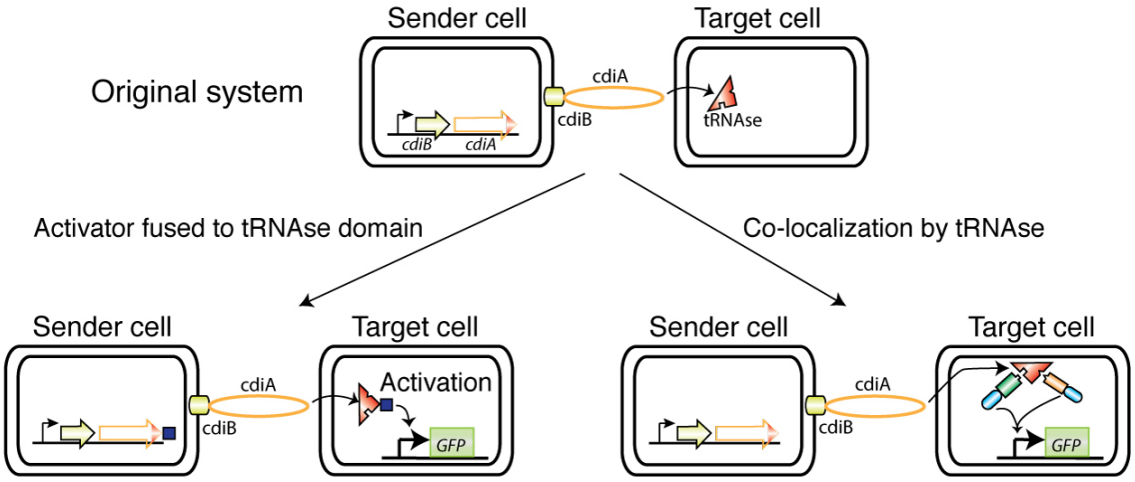
resistant to the growth inhibitory activity. While the exact mechanism of how CysK activates the tRNAse activity is unknown, Diner *et al* found that CysK binds to the C-terminal end of the tRNAse. In addition, they found that the tRNAse can simultaneously bind the UPEC536 CdiI immunity protein. When both CysK and the CdiI protein are bound, the tRNAse activity is prevented. CysK is a protein used in cysteine biosynthesis and is normally present in the cytoplasm. Webb *et al* showed with immunofluorescence microscopy that the tRNAse domain is found in the cytoplasm of target cells [79].

3.2. Results

3.2.1. Overview of strategies for harnessing the CDI system

Our goal was to utilize the CDI system as a mechanism to control gene expression in target cells that are in contact with sender cells. In the characterized CDI systems, the toxin domain leads to growth inhibition and does not trigger a specific signal that we can use to modulate gene expression. We chose to start with the tRNAse toxin domain since it is known to reach the cytoplasm of target cells. We had two approaches to utilize the tRNAse-based CDI system (shown conceptually in Figure 22). First, we could directly fuse a protein that can trigger a change in the expression level of a target gene. The simplest example is a small protein that can activate expression of a promoter that controls our gene of interest. The advantage of this approach is that it is conceptually straightforward since we are attempting to simply deliver another protein with the existing CDI system. The disadvantage with this method, however, is that we are perturbing the natural delivery process. At this point, we do not know the details for how proteins are delivered into the cytoplasm of target cells and what perturbations are tolerated by the CDI system.

Figure 22. High-level overview of two approaches for using CDI. The original system delivers a tRNAse domain (shown as a red triangle) that is at the C-terminus of the CdiA protein. On the left, we show the first approach of fusing a protein fragment, indicated as a blue square, to the tRNAse domain. On the right, we show an alternative scheme where we exploit the known property of the tRNAse that it can simultaneously bind to two other proteins (shown in green and orange) and can thus co-localize these two proteins.



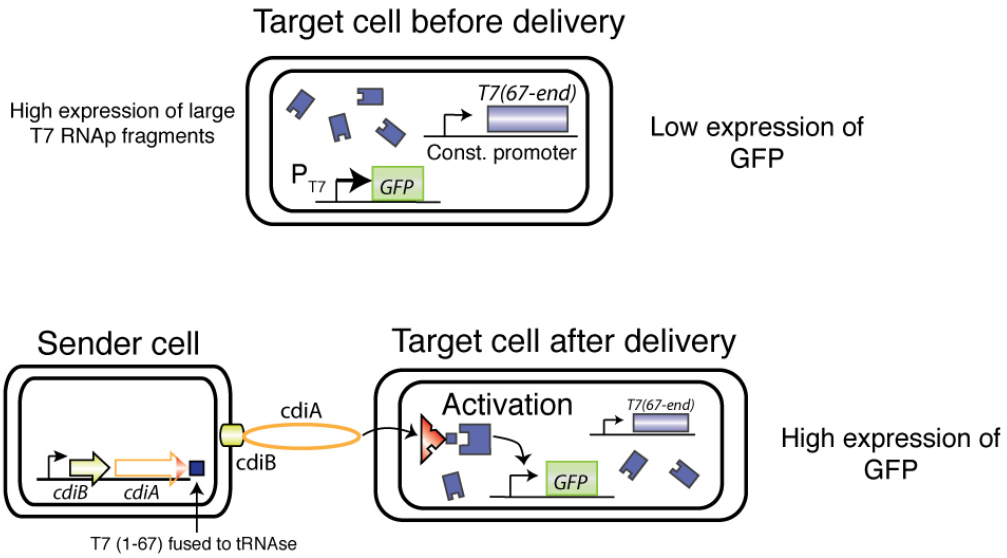
In order to mitigate the biological uncertainty that the first approach requires, we also devised a second approach that leaves the existing delivery process intact (shown schematically on the bottom-right of Figure 22). We utilize the fact that the tRNAse killing domain can bind simultaneously to both the Cysk and CdiI proteins. Thus, delivery of the tRNAse domain brings the CysK and CdiI proteins spatially close together, and we can exploit the spatial co-localization to trigger a change of gene expression. This scheme resembles two-hybrid systems that are commonly used for detecting protein-protein interactions [80–82]. This approach does not require us to perturb the transfer mechanism so there is no uncertainty about whether the CDI process can still deliver the tRNAse domain; instead, the difficulty with this approach is detecting the possibly small change in signal that results from spatial co-localization.

3.2.2. Method of protein fusion to the CdiA protein

In the first scheme where we fuse a protein or peptide fragment to the CdiA protein, there are two characteristics we think are ideal. First, the peptide fragment should be a very potent activator so that even if only a smaller number of molecules are successfully delivered into the cytoplasm of the receiver cells, the delivered molecules can trigger a large change in gene expression. We preferred a molecule that can directly increase gene expression, but one can also envision indirect schemes such as a catalyst that increases production of a metabolite whose concentration can be sensed by a regulator. Second, this peptide fragment should be as short as possible since we expect that a shorter fragment is less likely to perturb the ability of the CDI system to transport this modified tRNAse domain.

Our primary candidate for this approach is a small fragment from a split T7 RNA polymerase. T7 RNA polymerase (RNAp) is a highly potent transcriptional activator that tends to severely inhibit the host *E. coli* cell's ability to grow since so many resources are diverted by the T7 RNAp. The T7 RNAp system is routinely used for over-expressing (heterologous) proteins in *E. coli* since the polymerase is highly selective for its promoters and up to 50% of the host cell's protein fraction can be diverted to the heterologous protein [83]. Members of the Voigt lab recently demonstrated that they could split this protein into at least two pieces (unpublished work), and one of the pieces only consists of 67 amino acids (which we occasionally refer to as the small half). Figure 23 provides an overview of this scheme where we fuse the small fragment to the CdiA protein and over-express the large fragment of T7 RNAp in the receiver cells. We attempted to fuse the small T7 RNAp fragment at two locations in the tRNAse: after 103 amino acids of the tRNAse and at the very end of the tRNAse (223 amino acids). The motivation for fusing the T7 RNAp in the middle is that there could be a size limit the CDI system permits, so we wanted to attempt a "swap" in addition to the "append" design.

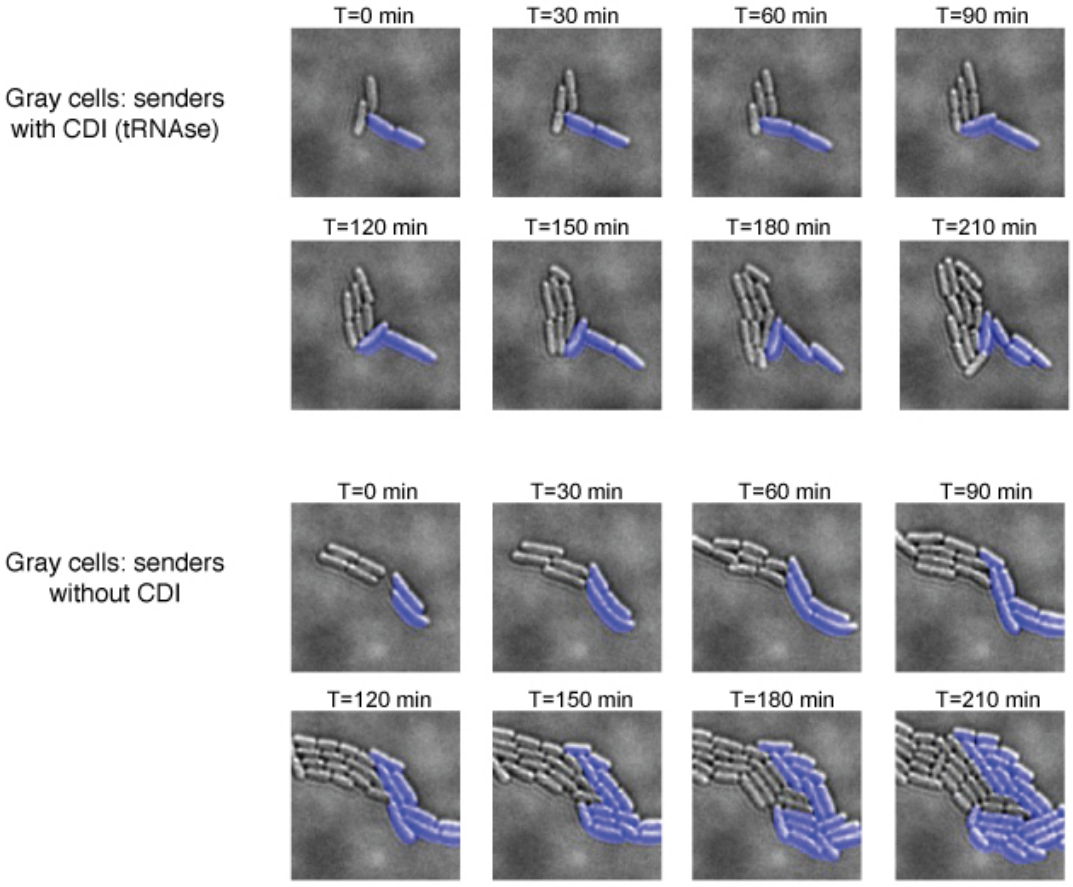
Figure 23. Overview for scheme with fused T7 RNAp fragment. Sender cells contain the small T7 RNAp fragment (1-67aa) fused to the tRNAse domain. Target (receiver) cells constitutively express a high concentration of the large T7 RNAp fragment (67 aa to the end). Only after both halves are expressed in the same cell is the downstream promoter able to fire.



Our end goal with this approach is to demonstrate control of gene control by transferring the small T7 RNAp fragment during co-culture on agar pads. We will first verify that the CDI transfer process works in our experimental setup. After confirming that the toxin can be delivered via the CDI system, we will then investigate that the split T7 fragments behave properly. To eliminate uncertainty in the delivery process when fusing a T7 RNAp fragment to the toxin domain, we will initially express all the fragments in the same cell. We will then fuse the small T7 RNAp fragment to the entire CdiA protein and test if we can detect transfer of the small T7 RNAp fragment from one cell to another.

Figure 24 shows that sender cells containing the CDI system with the wild-type tRNAse domain can prevent growth in target cells that they are touching. This confirms that the transfer process can work on the agar pads that we use in our experiments and in our strains of *E. coli*.

Figure 24. Verification that CDI transfer can occur on agar pads. Receiver cells constitutively express a fluorescent protein and are colored blue in this figure; receiver cells contain pJT26 in BW27783 cells so express *cysK* from the natural chromosomal location. Sender cells are grey and express the CDI system (via pCDI233) or are decoy cells without the CDI system; sender cells are in CDI299 cells and do not contain a *cysK* gene.



We then verified that the small T7 fragment and the fusion of the large T7 fragment work as expected when expressing everything in a single cell (i.e., simulating the final state after the transfer process). In this experiment, we have fused 67 amino acids of T7 RNAp with a linker after the first 103 amino acids of the tRNAse domain (pCDI663). Figure 25 shows the constructs that were built to perform this preliminary experiment. We constitutively express two versions of the large fragment from T7 RNAp: alone and fused to CdiI. We included the version with a CdiI fusion since we hypothesized that CdiI must bind tightly to the tRNAse (so that cells expressing CdiI don't get cleaved by the tRNAse), so CdiI's binding to the tRNAse could help stabilize the interaction of the two T7 RNAp fragments. The results from a plate fluorometer (Figure 25) indicate that the split T7 RNA polymerase system is working when everything is expressed within a single cell. We observe a slight increase in fluorescence already when the high-copy reporter plasmid is present from some leaky expression of the GFP. When just the big half of T7 RNA polymerase is present (either pCDI679 or pCDI680), there is a negligible increase in fluorescence although we observe that the pCDI679 version has a slight

increase in GFP production. Nonetheless, when the small half is included via pCDI663, we observe a dramatic increase in GFP expression.

Figure 25. Split T7 RNAp fusions work as desired. Cells are fully induced with arabinose during the experiments. Shown are mean of triplicate fluorescence values as measured on a Tecan plate fluorometer normalized by OD600.

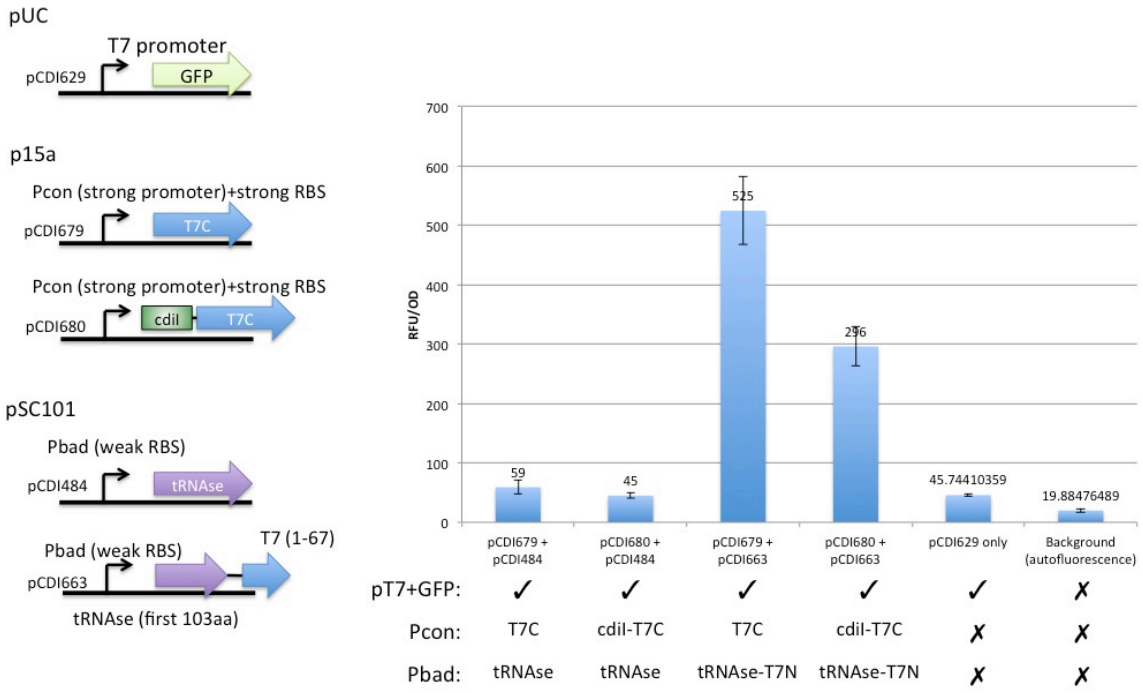
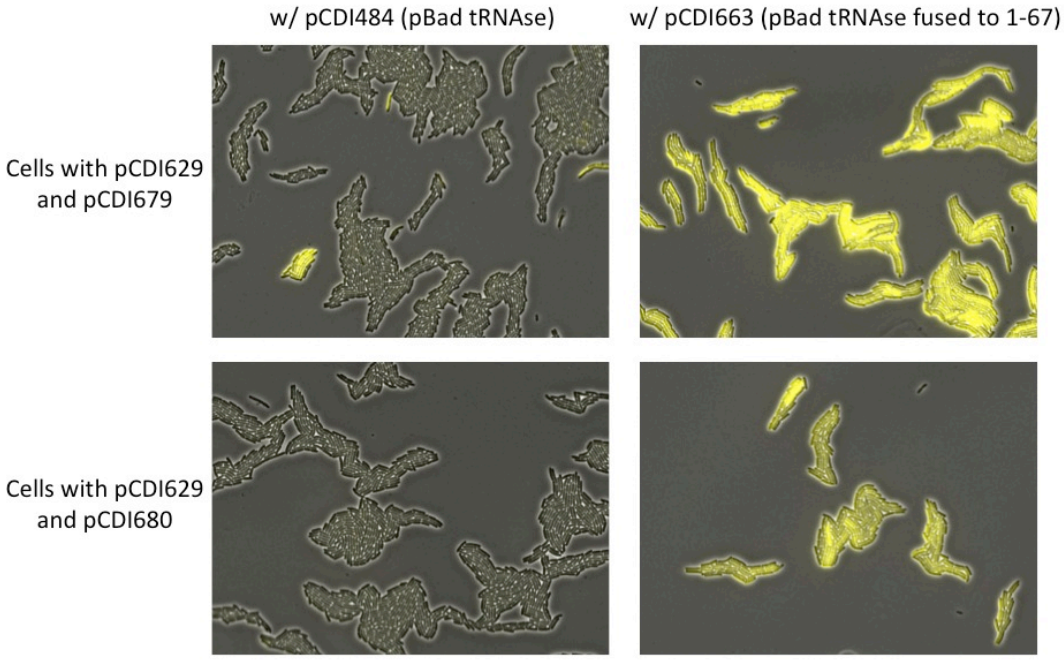


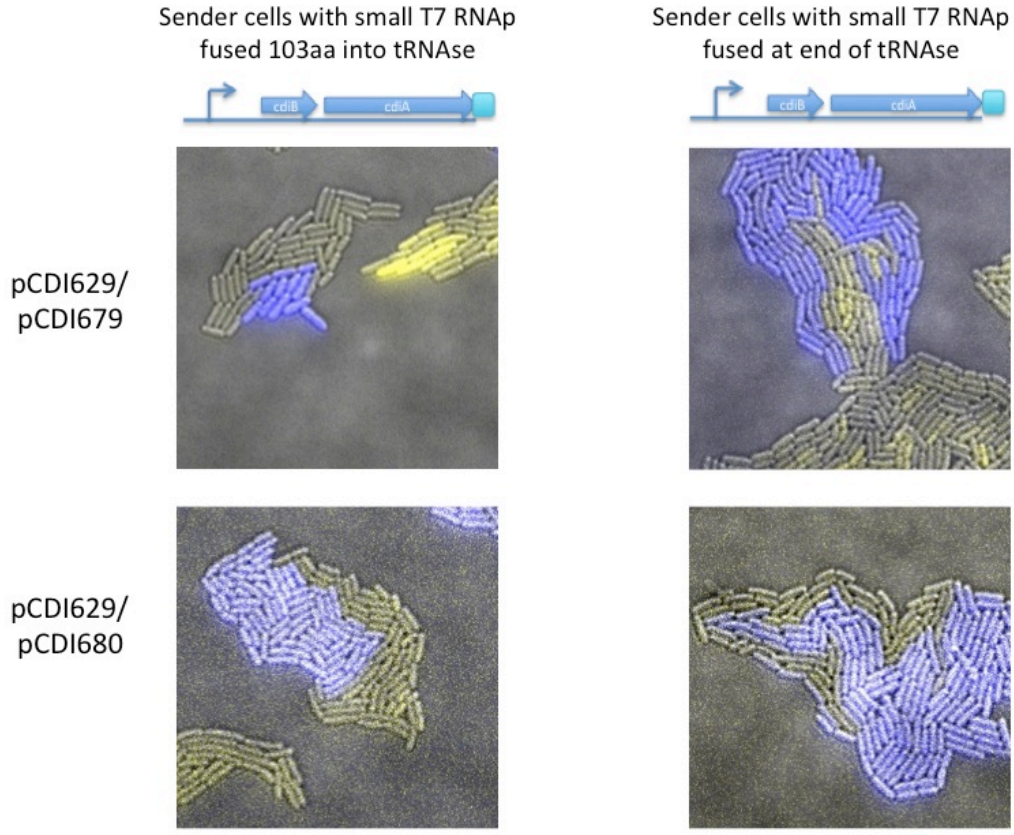
Figure 26 verifies these results on a single-cell basis by observing individual cells with fluorescence microscopy. The results are consistent with the bulk measurements from the plate fluorometer, indicating that pCDI679 has higher background GFP expression than pCDI680, caused by a small fraction of cells that are able to highly express GFP, but we can still observe a substantial increase in the expression of GFP when the small T7 RNAp fragment is also present.

Figure 26. Verification that split T7 RNA polymerase works with single-cell measurements. CDI299 cells containing the indicated three plasmids were grown on an agar pad with arabinose induction and then imaged on a Zeiss fluorescent microscope. In the absence of the small tRNAse fragment (left column), the cells tend to not express GFP (colored as yellow). In the presence of the small tRNAse fragment (right column), the cells express much higher amounts of GFP. Cells with pCDI679 are occasionally able to express GFP in the absence of the small tRNAse fragment.



Finally, we tested if we can deliver the small T7 RNAp fragment. For this test, we fused the small T7 RNAp fragment both at the same location as used previously and at the very end of the tRNAse domain. We expect that protein fusions at the end of the tRNAse domain can function properly because the final four residues of the tRNAse are exposed for binding to the CysK proteins, indicating that the end of the tRNAse is not buried [78]. Figure 27 shows an experiment where we grow the sender cells containing the small T7 RNAp fragment near the receiver cells, using agar pads that we previously verified can allow delivery of the tRNAse domain. We fail to observe an increase in GFP intensity in the receiver cells (colored yellow) that are in contact with sender cells (colored blue). In the case where the receiver cells express the big T7 RNAp fragment from pCDI679, we observe some receiver cells that are not in contact with sender cells that strongly express GFP (consistent with results shown in Figure 26).

Figure 27. Testing split T7 RNAp system with CDI transfer process. Sender cells (pCDI596 and pCDI657/pCDI658) are colored blue and receiver cells (pCDI629 and pCDI679/pCDI680) are colored yellow. Cells were grown on an agar pad and imaged on a Deltavision fluorescence microscope.



Despite our best efforts at identifying a protein that is small and highly potent at activating gene expression, our results so far are not promising for this strategy. There are two potential problems: either the CDI system can not tolerate the changes we made to the tRNAse or we are delivering too few molecules to observe a change in GFP production even if we are successfully delivering the small T7 RNAp molecules that are fused to the tRNAse. Unfortunately, there is no easy way to debug this system. Because the tRNAse does not exhibit toxic activity when fused to the small T7 RNAp fragment, we cannot assay for growth inhibition in the receiver cells. At this point, we do not understand enough of the CDI biology to predict which modifications are tolerated by the transfer process. The ideal way to diagnose the system is to purify antibodies that bind to the small T7 RNAp fragment, and then we could image receiver cells with immunostaining after co-culturing with sender cells, similar to what has previously been used to monitor CdiA transfer [79]. This would definitively answer the question if the CDI system can still deliver the tRNAse with the small T7 RNAp fusion and could be quantitative enough to indicate how many molecules are transferred [84]. In order to predict how many molecules need to be transferred for us to observe a shift in GFP expression, we would use the cells that express both halves in the same cell (ie, without

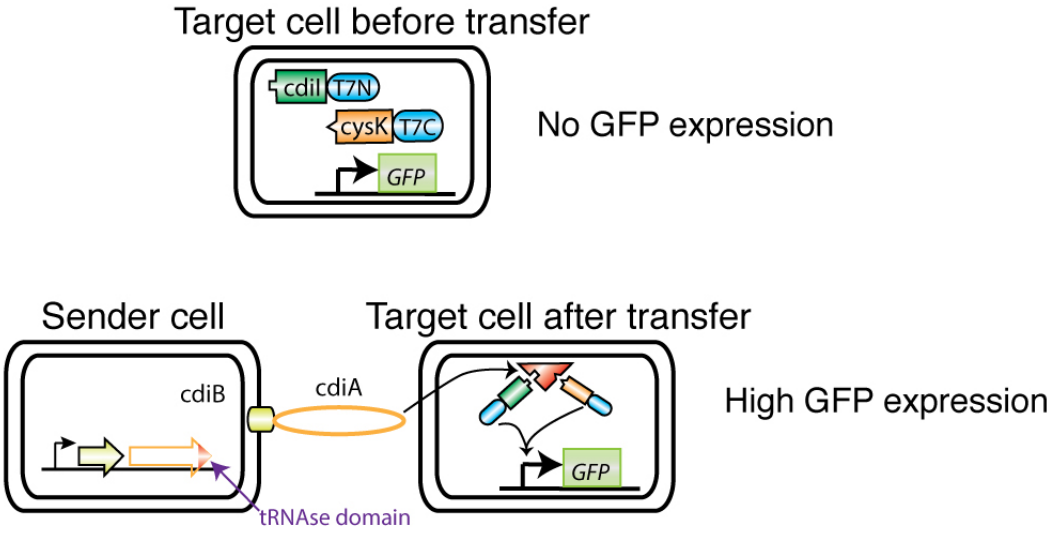
the transfer process) and perform quantitative Western blotting along with fluorescence measurements.

Even if we could figure out why the system is failing to deliver adequate numbers of the small T7 RNAp fragment, we are not sure what changes to make to improve the transfer process. Instead of attempting to debug this strategy of a small-peptide fusion to the tRNAse domain, we pursued an alternative scheme that does not require modifying the delivery process at all.

3.2.3. Method of spatial co-localization

Diner *et al* showed that CdiI (the immunity protein) and CysK (the protein whose binding is required for tRNAse activity) can bind simultaneously to the tRNAse [78]. We sought to combine this known property of the tRNAse with ideas developed in the past decade for detecting protein-protein binding via two-hybrid or three-hybrid assays [80–82]. The common idea behind these techniques is that we start with two protein fragments that do not have activity when not interacting with the other fragment. When the two fragments are brought in close proximity, they form a complete enzyme or transcriptional activator. Our hypothesis is that since the CdiI and CysK proteins can simultaneously bind the tRNAse, we should be able to fuse two different proteins to the CdiI and CysK so that when they bind the tRNAse simultaneously, both proteins can interact to form the complete protein.

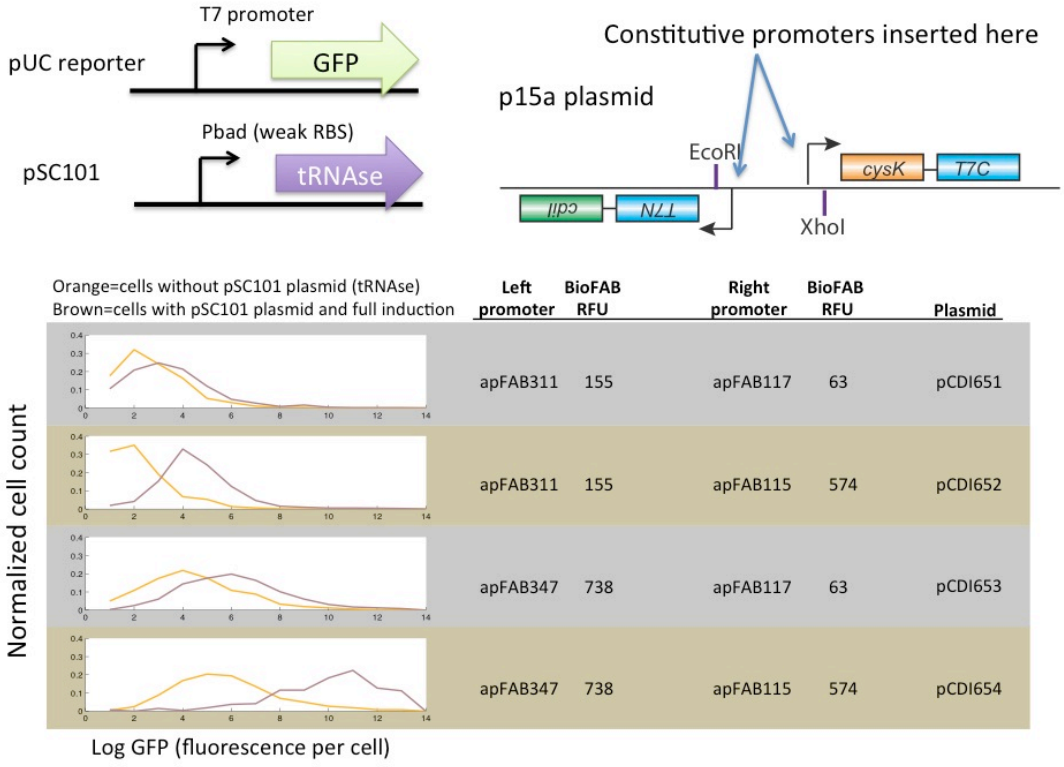
Figure 28. Overview of strategy for using tRNAse as a scaffold for co-localization. Target (receiver) cells constitutively express split T7 RNAp (denoted as T7N and T7C) molecules that are fused to CdiI or CysK. The tRNAse domain of CdiA has a single mutation that prevents tRNAse activity but does not otherwise disrupt transfer or the ability of CysK or CdiI to bind.



We again chose to use the split T7 RNA polymerase, although now that we do not have a constraint of needing one of the halves to be as small as possible, we used the split that was recently published by Shis and Bennett [85]. This split was shown to provide an extremely high fold-change in expression level only when both split fragments are

present whereas the previous split occasionally showed some GFP expression if only the large T7 RNAp fragment is present (Figure 26). Figure 29 shows the constructs that were used to test that our co-localization scheme can work when we express both split proteins and the tRNAse scaffold in the same cell. We designed a p15a plasmid that contains a cloning site into which we can insert constitutive promoters of different strengths. We chose a few promoters from the BioFAB parts registry and built four combinations [86]. We prevent killing activity by mutating one conserved residue that removes the tRNAse's catalytic activity and should not disrupt binding activity or the transfer process [78,79]. We observe that cells with the tRNAse show increased levels of GFP expression, consistent with the co-localization effect working. We also find that with increased expression of the split fusions, we observe higher levels of basal fluorescence, indicating that the split proteins are capable of interacting with each other in the absence of the tRNAse scaffold.

Figure 29. Testing that co-localization scheme works when everything is expressed in a single cell. Each cell in the experiment contains a pUC reporter plasmid (pCDI629) and a p15a plasmid (pCDI651-pCDI654) that constitutively expresses both split fusions. The left promoter drives the N-terminal half of T7 RNA polymerase fused to CdiI and the right promoter drives CysK fused to the C-terminal half of T7 RNA polymerase. Cells were assayed by fluorescence microscopy.

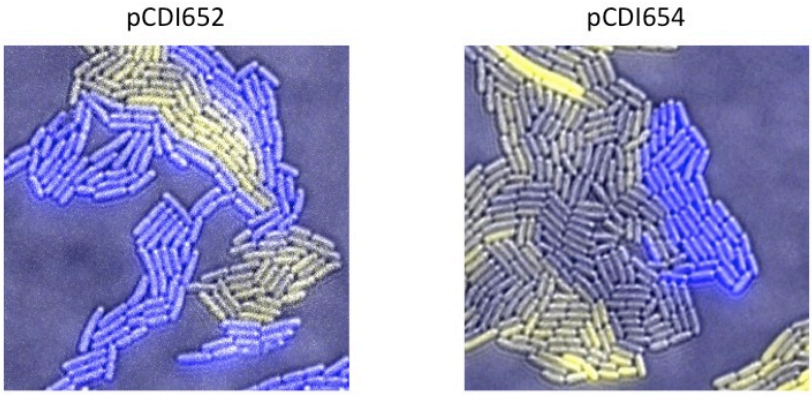


Left promoter	BioFAB RFU	Right promoter	BioFAB RFU	Plasmid
apFAB311	155	apFAB117	63	pCDI651
apFAB311	155	apFAB115	574	pCDI652
apFAB347	738	apFAB117	63	pCDI653
apFAB347	738	apFAB115	574	pCDI654

Upon observing this success, we also attempted to test if we could observe a change in fluorescence if we delivered the tRNAse molecules via the CDI process. We chose our best two plasmids, pCDI652 and pCDI654, and used them as receiver cells with sender cells that contain the CDI system. We allowed the sender and receiver cells to grow near

each other on an agar pad, but we fail to observe the desired behavior that receiver cells that are touching sender cells exhibit higher levels of GFP expression.

Figure 30. Testing the spatial co-localization scheme with CDI transfer. Sender cells contain pCDI596 and pCDI616 in CDI299 cells and are colored blue. Receiver cells contain the indicated plasmid and pCDI629 in CDI299 cells and are colored yellow. Note that we do not use the same mapping of intensity to brightness in the two images since pCDI654 has much higher levels of GFP expression. Cells were imaged on a Deltavision fluorescence microscope.



3.2.3.1. Improving sensitivity of spatial co-localization

We suspect that we did not observe the desired behavior with the transfer process because we are only transferring a small number of tRNAse molecules. Thus, we sought to modify the design to enhance the response of the system so that there is a better separation of the GFP distribution even if only a small number of tRNAse molecules are delivered. Another way to view the problem is that the presence of the tRNAse scaffold increases the effective concentration of T7 RNAp molecules; thus, our goal was to design an ultrasensitive response so that slight increases in T7 RNAp concentration can lead to strong changes in output.

3.2.3.2. Intuition for why positive feedback can improve I/O response

We attempted to design an ultrasensitive response by introducing a positive feedback loop, and in order to motivate the benefits of positive feedback, we first present an analytic investigation of a simplified system. Suppose that we have a protein species called x that drives itself in a positive feedback loop. We assume no cooperativity or other source of ultrasensitivity except for the positive feedback loop, so we model the production rate with a Michaelis-Menten-like term. We also assume that degradation and dilution is a standard first-order term ($\gamma * x$):

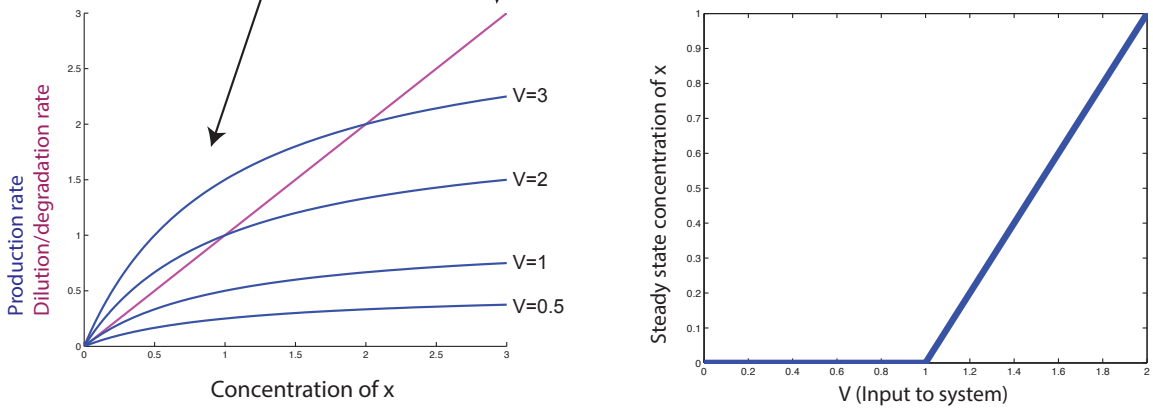
$$\frac{dx}{dt} = V\frac{x}{K_m + x} - \gamma * x \qquad \text{Equation 11}$$

We solve for the steady states for x and find that $x = 0$ or $x = \frac{V - K_m * \gamma}{\gamma}$. Because there are at most two steady states, only one of them is stable. There is a threshold for which steady state is stable: when $V < K_m * \gamma$, the system is stable at $x = 0$. Otherwise, when V

is greather than the threshold $K_m * \gamma$, the system is stable at $\frac{V - K_m * \gamma}{\gamma}$. If we view V as the input to the system, we observe a steady-state response for x that is highly reminiscent of the buffered responses that we obtained from molecular sequestration. Figure 31 provides some sample plots and the overall system I/O response when $\gamma = 1$ and $K_m = 1$. The intuitive explanation is that when the production rate (V) is too weak, the system remains stably in the off state. As we increase the production rate above the threshold $K_m * \gamma$, the system moves up to a higher stable steady state.

Figure 31. Simple demonstration for how positive feedback can give ultrasensitivity. The system described in Equation 11 leads to an ultrasensitive response when V is viewed as the input to the system and we observe the stable steady state of x as the output. $K_m = 1$ and $\gamma = 1$ in the plots below.

$$\frac{dx}{dt} = V\frac{x}{K_m + x} - \gamma * x$$



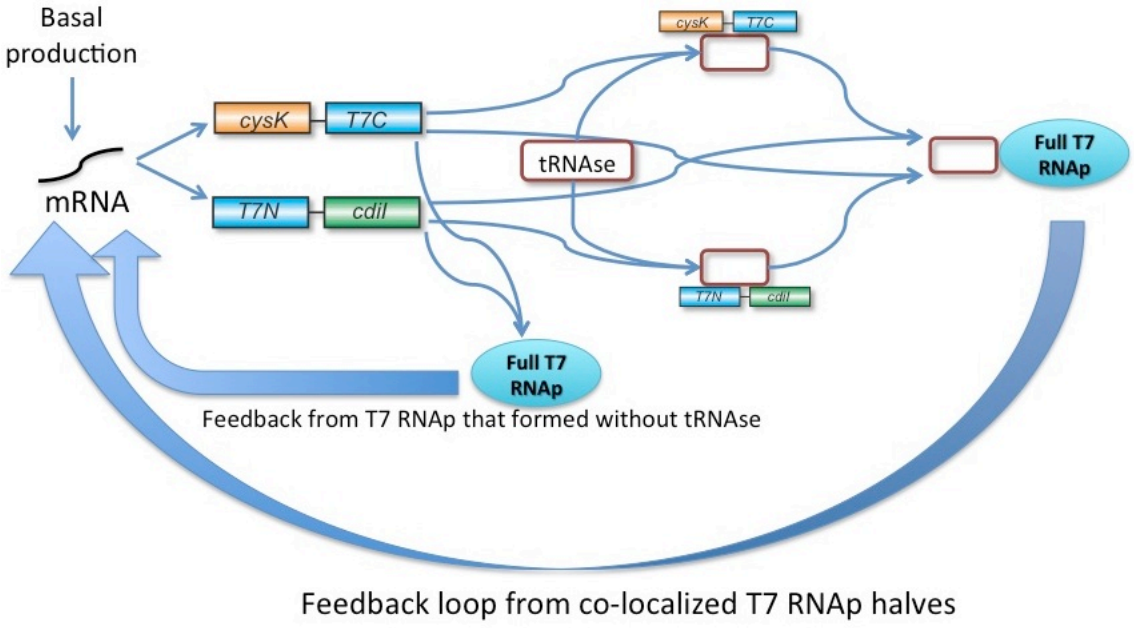
Returning to our system with tRNAse recruiting the two halves, X represents the concentration of T7 RNA polymerase molecules and V represents the effectiveness of the T7 RNA polymerase molecules. When more tRNAse molecules are present in the cell, the split T7 RNAp halves are more likely to bind to the tRNAse scaffold and then bind to each other, equivalent to increasing V.

3.2.3.3. Schematic for co-localization scheme with positive feedback

With this intuition in mind, Figure 32 illustrates how we incorporate a positive feedback loop with our system that utilizes the tRNAse as a scaffold. In the schematic, we account for complete T7 RNAp molecules that form even without the aid of tRNAse and we also have the co-localization of the split proteins via the tRNAse scaffold. We use the same mRNA to produce both split proteins and we allow a constitutive production source of

the mRNA. This allows us to independently tune the basal amount of mRNA before the tRNAse enters the cell.

Figure 32. Reaction diagram for co-localization scheme with positive feedback. The mRNA transcripts lead to translation of both split proteins. The mRNA can be produced by positive feedback of T7 RNAp molecules or by a basal (constitutive) production source.



3.2.3.4. Analytical treatment for simplified system

The system shown in Figure 32 requires 7 ordinary differential equations, so it is difficult to gain intuition for the system. We define the species used in the complete reaction network in Table 4. We present an analytical investigation of a simplified system where we make a number of assumptions to eventually reduce it to a 2-ODE system.

Table 4. Species used to model the co-localization scheme

Species name	Description
a	CysK fused to half of T7 RNA polymerase
b	CdiI fused to half of T7 RNA polymerase
c	Complex with *a* and *b* bound together (complete T7 RNA polymerase)
T	tRNAse that is not bound to anything else
aT	tRNAse bound to *a*
bT	tRNAse bound to *b*
cT	tRNAse bound to both *a* and *b* (complete T7 RNA polymerase)
mRNA	mRNA from which *a* and *b* are both produced
A_{tot}	Total concentration of *a* molecules ($A_{tot} = a + aT + c + cT$)
B_{tot}	Total concentration of *b* molecules ($B_{tot} = b + bT + c + cT$)
T_{tot}	Total concentration of tRNAse molecules ($T_{tot} = aT + bT + cT + T$)

First, we assume no un-scaffolded binding of a and b (i.e., no c is formed), and we assume rapid equilibrium of the protein-protein binding reactions. We let $K_{d,a}$ denote the disassociation constant of a molecules to the tRNAse and $K_{d,b}$ denote the disassociation constant of b molecules to the tRNAse. We make the critical assumption that binding of a and b to the tRNAse occurs independently, i.e. the disassociation constants $K_{d,a} = \frac{a*T}{aT} = \frac{a*bT}{cT}$ and $K_{d,b} = \frac{b*T}{bT} = \frac{b*aT}{cT}$. This allows us to write the current concentrations of $aT, bT,$ and cT as:

$$aT = \frac{T * a}{K_{d,a}}$$

$$bT = \frac{T * b}{K_{d,b}}$$

$$cT = \frac{a * bT}{K_{d,a}} = \frac{b * aT}{K_{d,b}} = \frac{aT * bT}{T} = \frac{a * b * T}{K_{d,a} * K_{d,b}}$$

We then apply mass conservation, i.e. $A_{tot} = a + aT + cT = a + \frac{T*a}{K_{d,a}} + \frac{a*b*T}{K_{d,a}*K_{d,b}}$, $B_{tot} = b + bT + cT = b + \frac{T*b}{K_{d,b}} + \frac{a*b*T}{K_{d,a}*K_{d,b}}$, and $T_{tot} = aT + bT + cT + T = T * \left(\frac{a}{K_{d,a}} + \frac{b}{K_{d,b}} + \frac{a*b}{K_{d,a}*K_{d,b}} + 1\right)$ to derive current concentrations for $a, b,$ and T:

$$a(A_{tot}, B_{tot}, T_{tot}) = \frac{1}{2}(A_{tot} - K_{d,a} - T_{tot} + \sqrt{4A_{tot} * K_{d,a} + \left(T_{tot} + K_{d,a} - A_{tot}\right)^2})$$

$$b(A_{tot}, B_{tot}, T_{tot}) = \frac{1}{2}(B_{tot} - K_{d,b} - T_{tot} + \sqrt{4B_{tot} * K_{d,b} + \left(T_{tot} + K_{d,b} - B_{tot}\right)^2})$$

$$T(A_{tot}, B_{tot}, T_{tot}) = \frac{T_{tot}}{1 + \frac{a}{K_{d,a}} + \frac{b}{K_{d,b}} + \frac{a * b}{K_{d,a} * K_{d,b}}}$$

$$= \frac{4 * K_{d,a} * K_{d,b} * T_{tot}}{\left(A_{tot} + K_{d,a} - T_{tot} + \sqrt{4A_{tot}K_{d,a} + \left(K_{d,a} + T_{tot} - A_{tot}\right)^2}\right)\left(B_{tot} + K_{d,b} - T_{tot} + \sqrt{4B_{tot}K_{d,b} + \left(K_{d,b} + T_{tot} - B_{tot}\right)^2}\right)}$$

We treat the total number of tRNAse molecules as an input parameter that does not change over time. This allows us to reduce the system to a 3-ODE system where we only need to keep track of the concentrations for the total amount of a molecules, the total amount of b molecules, and the concentration of mRNA. Figure 33 provides an overview of the simplified system that we use in our analytic treatment. Note that $cT(A_{tot}, B_{tot}, T_{tot})$ can now be found by substituting into $\frac{a*b*T}{K_{d,a}*K_{d,b}}$ and is the function that represents the number of molecules of tRNAse bound to both halves (ie, both a and b molecules are bound). Because we assume no c molecules are formed, production of the mRNA only comes from positive feedback caused by cT molecules. We also assume that

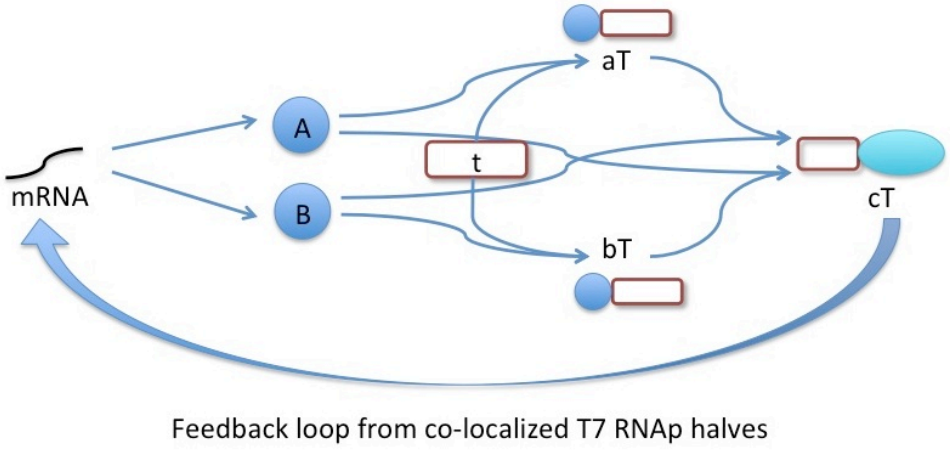
all protein molecules are degraded with the same rate constant γ, although we allow the mRNA molecules to degrade at a different rate γ_{mRNA}.

$$\frac{dA_{tot}}{dt} = K_a * mRNA - \gamma * A_{tot}$$

$$\frac{dB_{tot}}{dt} = K_b * mRNA - \gamma * B_{tot}$$

$$\frac{dmRNA}{dt} = V_m \frac{cT(A_{tot}, B_{tot}, T_{tot})}{K_m + cT(A_{tot}, B_{tot}, T_{tot})} - \gamma_{mRNA} * mRNA$$

Figure 33. Schematic of simplified 3-ODE system. We further assume that a and b molecules have identical properties in our analytic investigation to reduce the system to two ordinary differential equations.



We can reduce the system from three ODE's to two ODE's if we make a final assumption that a and b have identical properties, i.e. $K_{d,a} = K_{d,b}$, $K_a = K_b$, and that the concentrations of the two are always the same, $A_{tot} = B_{tot}$. Equations 12 show the final model incorporating all assumptions used in the analytic investigation:

$$cT(A_{tot}, T_{tot}) = \frac{T_{tot}\left(K_{d,a} + T_{tot} - A_{tot} - \sqrt{4 * A_{tot} * K_{d,a} + \left(K_{d,a} + T_{tot} - A_{tot}\right)^2}\right)^2}{\left(K_{d,a} + A_{tot} - T_{tot} + \sqrt{4 * A_{tot} * K_{d,a} + \left(K_{d,a} + T_{tot} - A_{tot}\right)^2}\right)^2}$$

Equations 12

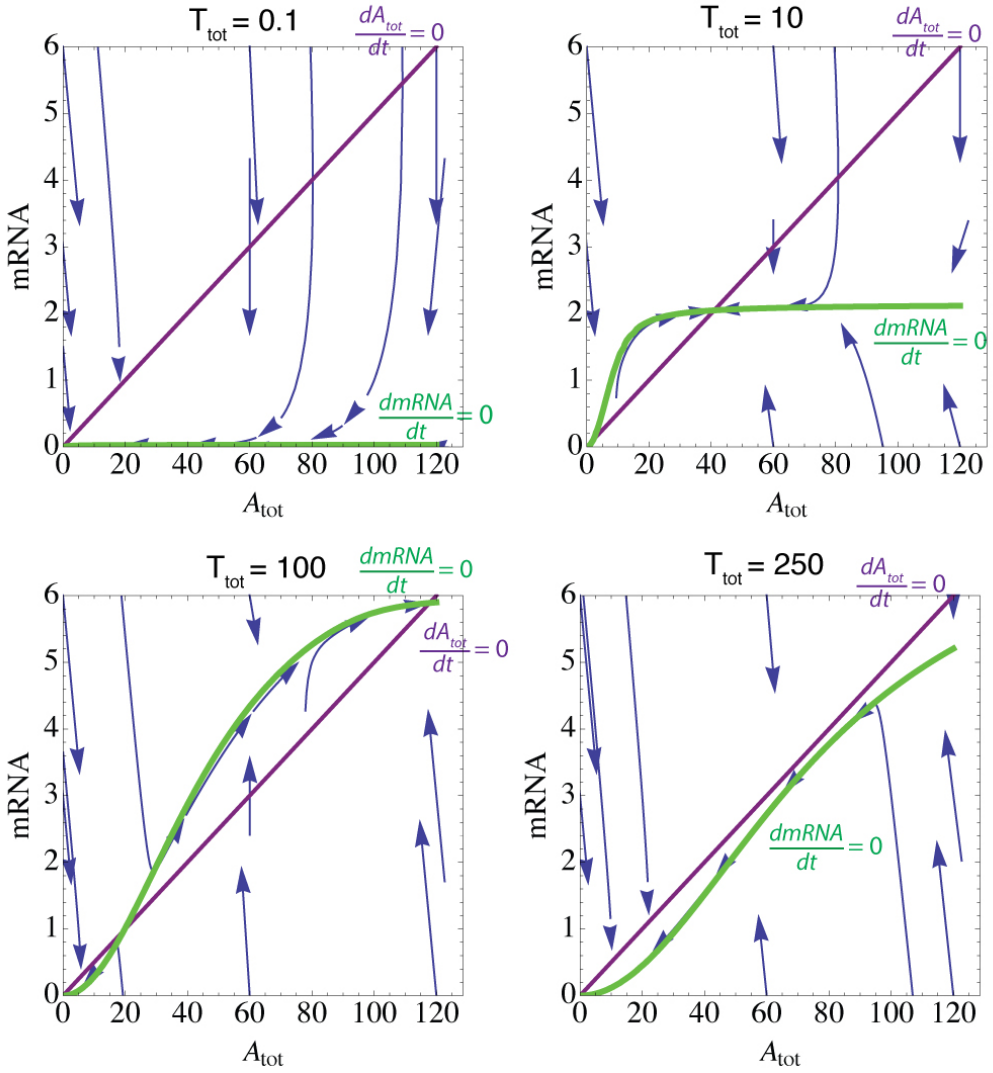
$$\frac{dA_{tot}}{dt} = K_a * mRNA - A_{tot}$$

$$\frac{dmRNA}{dt} = V_m * \frac{cT}{K_m + cT} - 10 * mRNA$$

Notice that we have non-dimensionalized away the dilution rate for the A_{tot} molecules, and we assume that the mRNA is degraded at a rate that is 10 times the dilution rate for the protein molecules. This is consistent with proteins that are extremely stable so the loss comes from dilution every time the cell divides (we assume the cells divide every 30 and a half-life of 3 minutes for the mRNA).

Figure 34 provides sample nullclines for the system. At low T_{tot}, the system is at a monostable low state as expected, since no complete T7 RNAp complexes are able to form in the absence of tRNAse molecules that serve as a scaffold for co-localization. As T_{tot} is increased, we observe that the nullcline for mRNA production is shifted upwards, allowing a stable steady state that is ON. In fact, the system becomes bistable since there are three intersections of the nullclines, and two of the intersections are stable steady states. Finally, when T_{tot} is increased too far, the mRNA nullcline is shifted too far to the right (even as the mRNA nullcline is shifted upwards), and the system turns off again to a monostable low state. At a high T_{tot}, we can compensate and recover the high stable steady state by increasing K_a, which reduces the slope for the A_{tot} nullcline. Before investigating the behavior of the nullclines, we might have expected that we could optimize the system to work at a low concentration of tRNAse (T_{tot}) and expect the response to be better at higher concentrations of tRNAse. Instead, we find that we need to carefully optimize the system parameters for the actual tRNAse concentration used in the transfer process.

Figure 34. Sample nullclines for the simplfied 2-ODE system. As T_{tot} is varied, the nullcline for mRNA production changes. The parameters for the plots in the figure: $V_m = 75, K_m = 25, K_{d,a} = 1, K_a = 20$.



3.2.3.5. Numerical investigation of the complete system

We return to a more complete model for the scheme that is shown in Figure 32. Equations 13 show the seven ODE's that describe this system (molecular species described previously in Table 4). Compared to the simplified system analyzed analytically, we now allow un-scaffolded binding of the two split proteins. We treat the complexes of T7 RNAp formed from un-scaffolded binding and co-localization via tRNAse as equivalent in their ability to transcribe mRNA; thus, we use the sum of c and cT for the production rate of mRNA in the feedback loop. We continue to assume independent binding of both types of proteins to the tRNAse, i.e. $K_{d,a} = \frac{a*T}{aT} = \frac{a*bT}{cT}$ and $K_{d,b} = \frac{b*T}{bT} = \frac{b*aT}{cT}$. We also assume as before that all proteins dilute at the same rate except for the mRNA, which is allowed to degrade at a faster rate. We use a doubling

time of 30 minutes and assume mRNA degrades with a half-life of 3 minutes, so we set γ=0.0231/minute and γ_{mRNA}=0.231/minute. We set $K_m = 10$ nM since this is consistent with what was previously measured for T7 RNA polymerase although it is unknown if our split T7 RNAp molecules have the same binding properties as what was measured for the full-length T7 RNAp [87].

$$\frac{da}{dt} = K_a * mRNA - K_{on,un} * a * b + K_{off,un} * c - K_{on,a} * a * tRNAse + K_{off,a} * aT - K_{on,a} * a * bT + K_{off,a} * cT - \gamma * a$$

$$\frac{db}{dt} = K_b * mRNA - K_{on,un} * a * b + K_{off,un} * c - K_{on,b} * b * tRNAse + K_{off,b} * bT - K_{on,b} * b * aT + K_{off,b} * cT - \gamma * b$$

$$\frac{dc}{dt} = K_{on,un} * a * b - K_{off,un} * c - \gamma * c$$

$$\frac{daT}{dt} = K_{on,a} * a * tRNAse - K_{off,a} * aT - K_{on,b} * b * aT + K_{off,b} * cT - \gamma * aT$$

$$\frac{dbT}{dt} = K_{on,b} * b * tRNAse - K_{off,b} * bT - K_{on,a} * a * bT + K_{off,a} * cT - \gamma * bT$$

$$\frac{dcT}{dt} = K_{on,a} * a * bT + K_{on,b} * b * aT - K_{off,a} * cT - K_{off,b} * cT - \gamma * cT$$

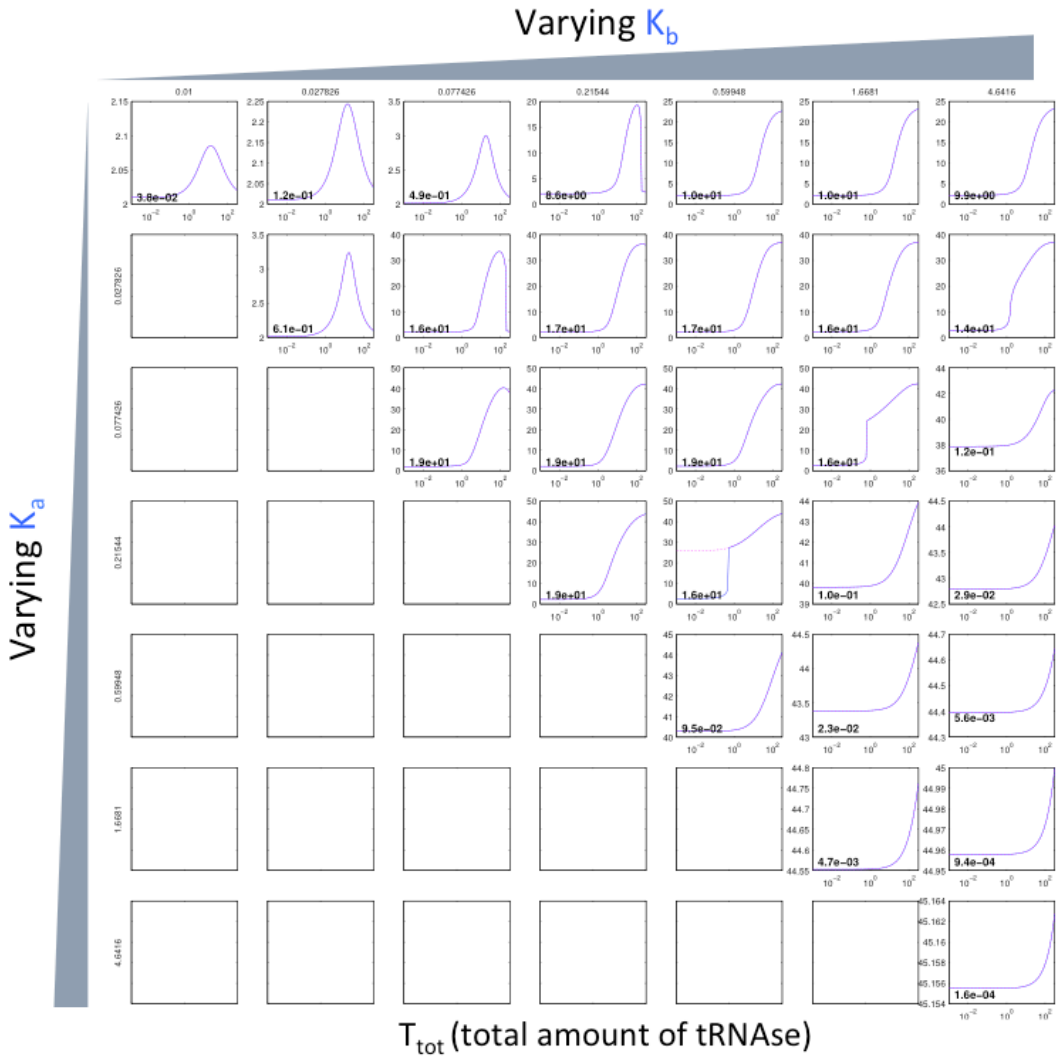
$$\frac{dmRNA}{dt} = basal_{mRNA} + V_m \frac{c + cT}{K_m + c + cT} - \gamma_{mRNA} * mRNA$$

Equations 13

We note that the parameter V_m encapsulates the copy number of the plasmid that contains the T7 promoter, so this parameter is tunable over a couple orders of magnitude. We choose $V_m = 10$ nM/min, which corresponds to a steady-state concentration of 43.3 mRNA molecules given our chosen degradation rate for mRNA. Given the previously found transcription rate for T7 RNAp of 40 bp/sec [88] and that our two split proteins are about 4000 bp in total, it takes approximately 1.67 minutes to transcribe the mRNA for both proteins. This corresponds to a copy number of approximately 72, which can be easily achieved with a high-copy plasmid in *E. coli*.

We begin by investigating an ideal case where un-scaffolded binding is extremely rare, so we let both $K_{d,a}$ and $K_{d,b}$ be 10 nM and set $K_{d,un} = \frac{K_{off,un}}{K_{on,un}} = 10$ uM. Figure 35 shows sample bifurcation plots of steady-state mRNA concentration as we vary total tRNAse concentration along the x-axis of each plot. We observe some of the behaviors that we already found from the analytic investigation. Generally, as we increase the concentration of tRNAse, the steady-state mRNA concentration increases, and for some parameters, we get an ultrasensitive response as desired. When the parameters are just right, the system becomes bistable in an irreversible manner. Finally, when the total concentration of tRNAse is increased too high, the steady-state level of mRNA drops back down.

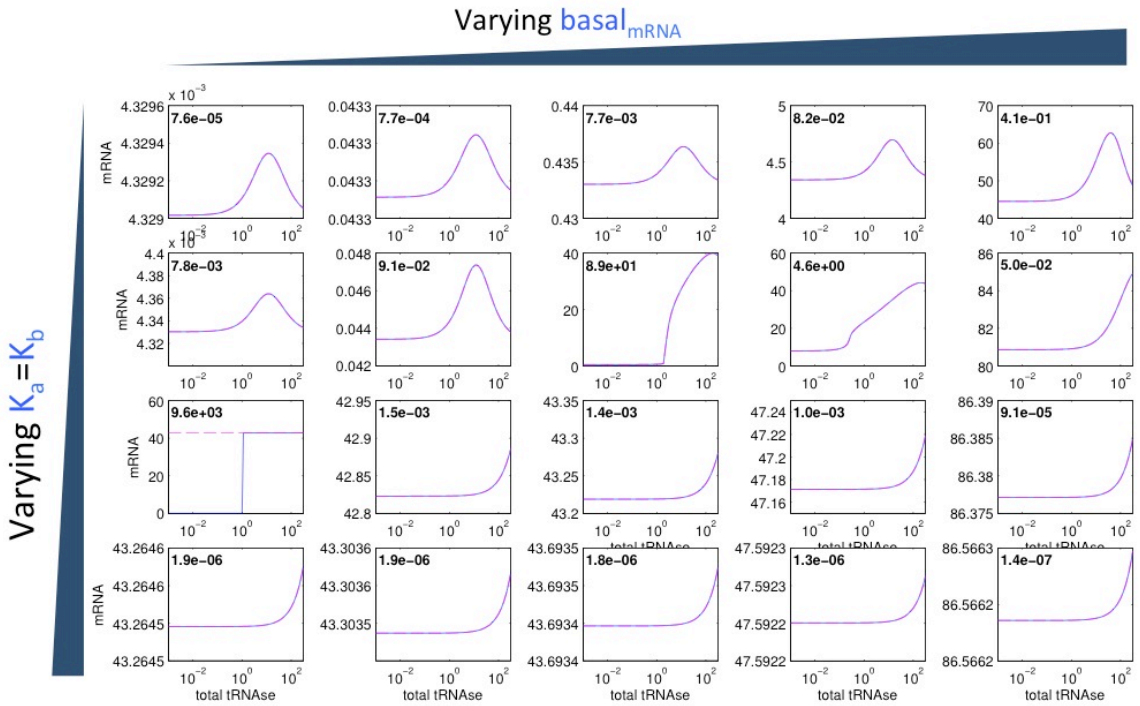
Figure 35. Bifurcation plots are consistent with results from analytic investigation of simplified model. Shown are the stable steady-states for mRNA concentration as T_{tot} is varied from 10^{-3} to 315 nM. The solid blue line represents the low stable steady-state and the pink dotted line shows the high stable steady-state. The lower-left corner displays the maximal fold-response change (maximal steady-state mRNA concentration divided by the mRNA concentration at $T_{tot} = 10^{-3}$). The parameters were set to $K_{on,a} = K_{on,b} = .01$, $K_{off,a} = K_{off,b} = .1, V_m = 10, K_m = 10, K_{on,un} = .00001$, $K_{off,un} = .1$, $basal_{mRNA} = .464$; since the grid is symmetric with this choice of parameters, we only display the upper half and the diagonal results.



One of the key ways our system can fail is if the two split halves bind to each other and turn on the feedback loop in the absence of the tRNAse. We investigated how much worse the binding affinity of the two split halves must be compared to the binding affinity to each split protein to the tRNAse scaffold. Recall that we define the binding affinities as $K_{d,un} = \frac{K_{off,un}}{K_{on,un}}$, $K_{d,a} = \frac{K_{off,a}}{K_{on,a}}$, and $K_{d,b} = \frac{K_{off,b}}{K_{on,b}}$. For the following analysis, we hold the on-rate for all reactions as 0.01 (per nM-hour). In Figure 36 we present a grid search as K_a, K_b, and $basal_{mRNA}$ are varied and $K_{d,un} = 1$ uM and both $K_{d,a}$ and $K_{d,b}$ are set to 10 nM. The results suggest that there exist parameters that provide the desired

behavior if $K_{d,un}$ (binding affinity without the tRNAse scaffold) is at least two orders of magnitude worse than binding of the CysK and CdiI proteins to the tRNAse.

Figure 36. Bifurcation analysis when binding affinity differs by 100-fold. Each plot shows the steady-state mRNA concentration as the total tRNAse concentration is varied from 10^{-3} to 315 nM. The upper-left corner displays the maximal fold-response change. The other parameters were held at: $K_{on,a} = K_{on,b} = .01$, $K_{off,a} = K_{off,b} = .1$, $V_m = 10, K_m = 10$, $K_{on,un} = .0001$, $K_{off,un} = .1$. We perform a grid search on the parameters: $basal_{mRNA}$ varies from 10^{-3} to 10, and $K_a = K_b$ vary from 10^{-2} to 10, with steps 10-fold greater than the previous value.

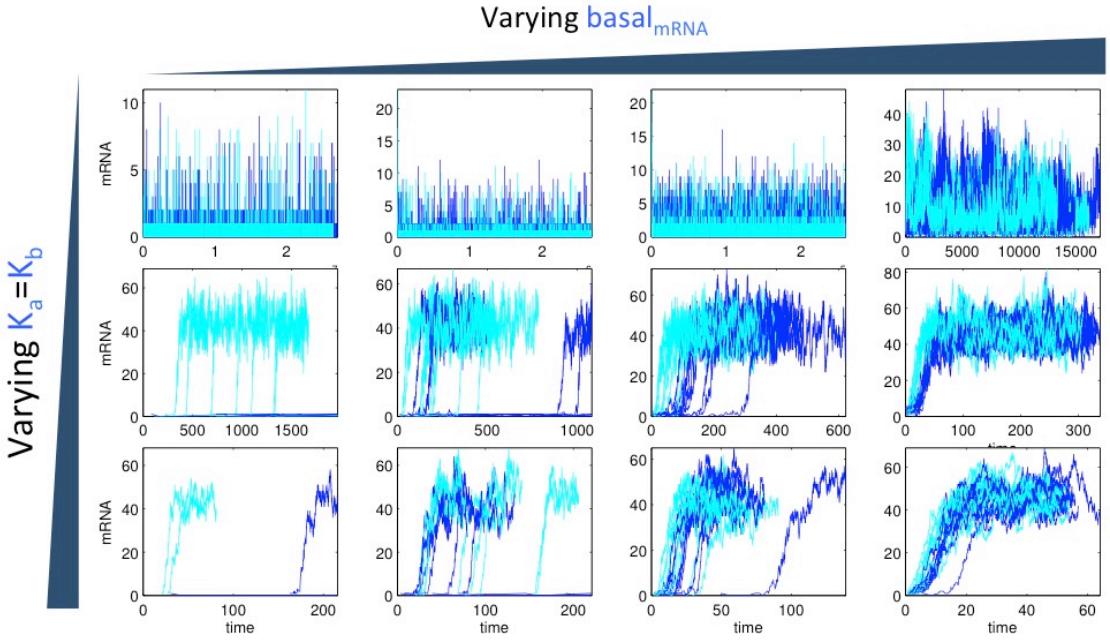


The ODE treatment just presented assumes that the system dynamics evolve in a continuous and deterministic manner. As much recent work has emphasized, stochastics in biological circuits can drastically alter the actual behavior [89–91]. We check that the previous results are also valid when we simulate the circuit behavior using an exact simulation via Gillespie's algorithm [92]. Note that when converting the system for stochastic simulation, we assume each nM in concentration corresponds to one molecule, and we assume that cells with the tRNAse contain 20 molecules of tRNAse since an email from David Low suggests that 20-200 molecules of the toxin are delivered, although these results are still unpublished.

Figure 37 shows the results of stochastic simulations for the same parameters as shown in Figure 36. The stochastic results suggest we can still find parameter sets that provide the desired behavior of ON cells in the presence of a low number of molecules of tRNAse while remaining OFF in the absence of tRNAse. The results are not perfect, however, as the best case probed in our grid search has only six of the ten cells turning ON in the first 2000 minutes of the simulation. The best case shown in Figure 37 occurs when $K_a = K_b = 1/\text{min}$, which is compatible with the experimentally measured translation

rate of 15 amino acids per second in *E. coli* [93]. Nonetheless, these results suggest that as long as the K_d for binding of the T7 RNAp split proteins without tRNAse is two orders of magnitude worse than the binding of CysK and CdiI to the tRNAse, we can expect parameters that qualitatively provide the desired behavior.

Figure 37. Stochastic simulations when binding affinity differs by 100-fold. Dark-blue traces show trajectories for cells that don't contain tRNAse and light-blue traces correspond to cells containing 20 molecules of tRNAse. In each plot, we simulate 10 cells for each tRNAse concentration until 100,000 reactions have fired. The x-axis is time in minutes while the y-axis shows number of molecules. The other parameters were held at: $K_{on,a} = K_{on,b} = .01$, $K_{off,a} = K_{off,b} = .1$, $V_m = 10, K_m = 10$, $K_{on,un} = .0001$, $K_{off,un} = .1$. We perform a grid search on the parameters: $basal_{mRNA}$ varies from 10^{-3} to 1, and $K_a = K_b$ vary from 10^{-1} to 10, with each step 10-fold greater than the previous value.



For comparison, Figure 38 shows the simulated trajectories for the cells when $K_{d,un} = 100$ nM and both $K_{d,a}$ and $K_{d,b}$ are set to 10 nM. When the binding affinity of the two halves is only 10-fold worse than the binding affinity for both proteins to the scaffold, we have a much harder time obtaining the desired behavior. Of the parameters tested, the best combination results in two of the ten cells with 20 molecules of tRNAse switching to the ON state. As a final comparison, Figure 39 shows the trajectories when $K_{d,un} = 10$ uM and both $K_{d,a}$ and $K_{d,b}$ are kept at 10 nM. When the binding affinity of the two split proteins is much worse than the binding of the CysK and CdiI to the tRNAse, we much more easily obtain the desired behavior. We observe a parameter set that gives all ten out of ten cells with tRNAse turning ON while all the cells without tRNAse remain OFF.

Figure 38. Stochastic simulations when binding affinity differs by 10-fold. Similar to Figure 37 with the same parameters, except $K_{on,un} = .001$.

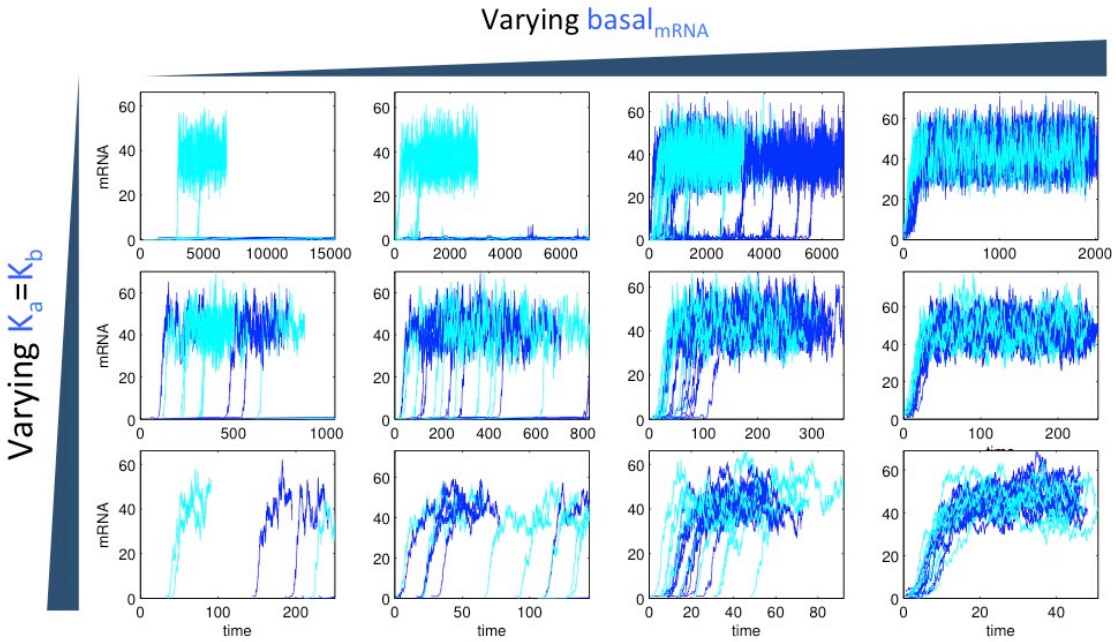
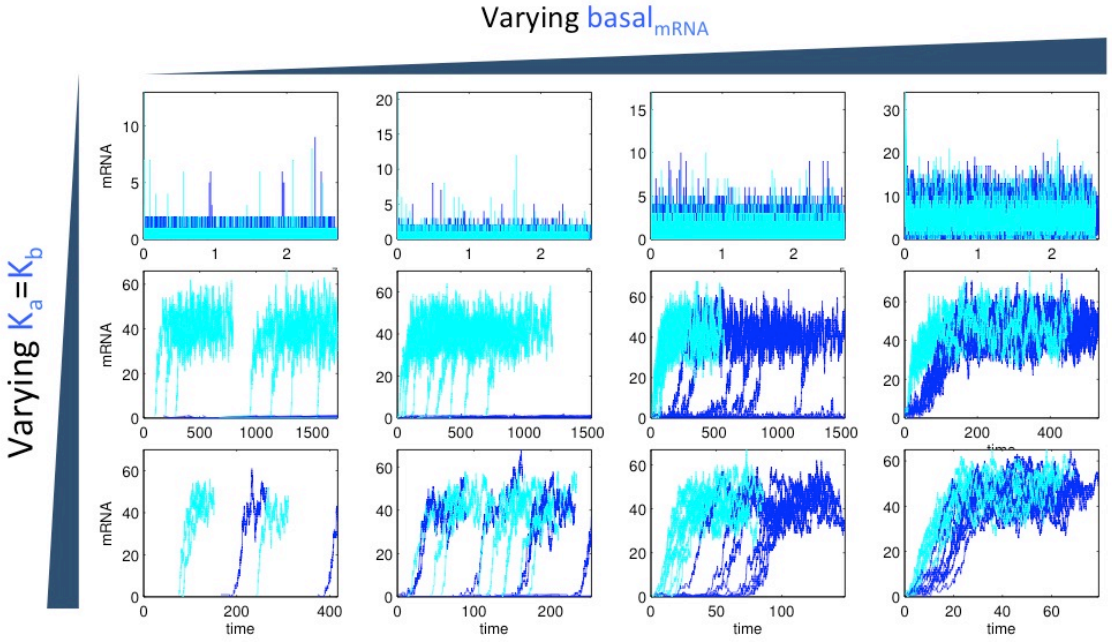


Figure 39. Stochastic simulations when binding affinity differs by 1,000-fold. Similar to Figure 37 with the same parameters, except $K_{on,un} = .00001$.

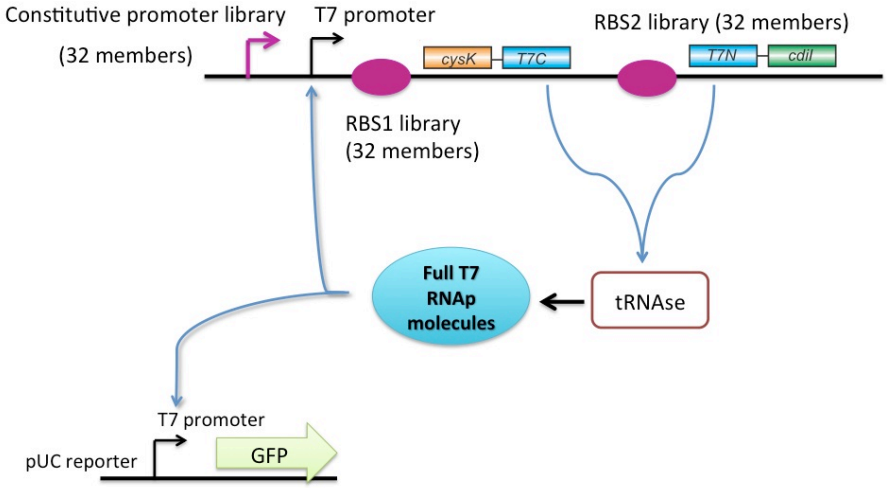


3.2.3.6. Attempt at experimental demonstration

While we do not know the actual biochemical parameters for our system, the simulation results presented above indicate that this strategy could potentially be achievable *in vivo*.

We chose to build a combinatorial library that allows us to simultaneously vary K_a, K_b, and $basal_{mRNA}$. Both genes are expressed from the same transcriptional unit, and we vary the ribosome binding site (RBS) for each of the proteins (overview of library design shown in Figure 40). We assume each RBS allows us to independently vary K_a and K_b although there may be some translational coupling. This two-gene operon is driven by a constitutive promoter and a T7 promoter. The constitutive promoter library allows us to vary the basal expression level of the mRNA transcript while the T7 promoter provides positive feedback from T7 RNA polymerase molecules.

Figure 40. Scheme for implementation of library to search parameter space. We construct a library containing 32^3 members by varying both ribosome binding sites and the strength of the constitutive promoter (library elements indicated with magenta color). The same mRNA transcript encodes for both the CysK-T7C fusion and the T7N-CdiI fusion; we assume production of the mRNA transcript is the summation of the activities of constitutive promoter and the T7 promoter.



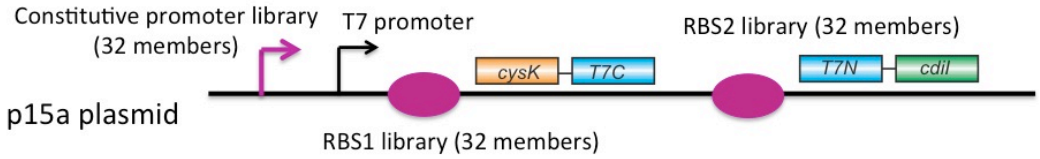
The library for the constitutive promoter is based off the strong J23100 promoter that was designed by Chris Anderson [94]. Anderson constructed a family with 19 members and characterized their strength by measuring production of a fluorescent protein; the 19 members span from 1 relative fluorescent unit (RFU) to 2547 RFU's. We designed our library to cover at least 1 to 396 RFU's (i.e., we wanted to ensure that some members of our library were as weak as possible). Both RBS libraries were constructed using the RBS calculator with the program set to generate as diverse a library as possible [95]. Since we can not guarantee that the library members actually span the intended parameter ranges, we allowed the libraries to be as large as possible to maximize our chances of getting a parameter set that results in the desired behavior. We decided that allowing each library to contain 32 members results in a final library size of $\mathbf{32^3} = 32{,}768$, which is a size that we can expect to fully represent during the assembly process.

Figure 41 lists the steps that we planned to perform on the library members. After constructing the combinatorial library on a p15a plasmid, the next was to transform this library into cells containing a compatible reporter plasmid (Step 2 of Figure 41). With the cells containing both plasmids, we would sort the library for cells that are dim in GFP fluorescence levels—since we do not actually know what the threshold for "dim" is, these

bin boundaries would be determined empirically. Since the sorter we used provides two-way sorting (i.e., two possible output tubes), we could set two bins that represent possible dim cells. The boundaries would be chosen to be higher in GFP signal than the negative control cells but lower in GFP than positive control cells. The next step would be to transform a pSC101 plasmid that expresses the tRNAse domain into the sorted cells, and sort for cells that are of higher intensity. Finally, we would need to isolate individual library members by mini-prepping the p15a plasmids and re-transforming into cells containing the reporter plasmid. We can then verify in co-culture assays that the selected p15a plasmids represent parameter combinations that give the desired system behavior.

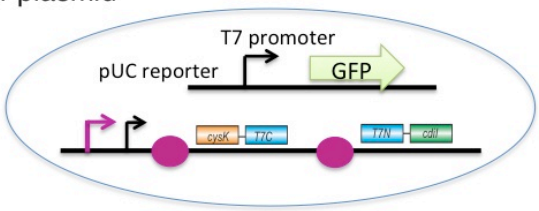
Figure 41. Proposed steps with FACS to find parameters that provide desired system behavior.

1. Make library (32*32*32 members)

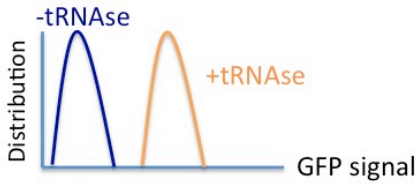


2. Transform into cells containing pUC reporter plasmid

3. Sort for dim cells



4. Transform in pSC101 plasmid expressing tRNAse domain and sort for bright cells



5. Isolate p15a plasmids that passed both screening steps

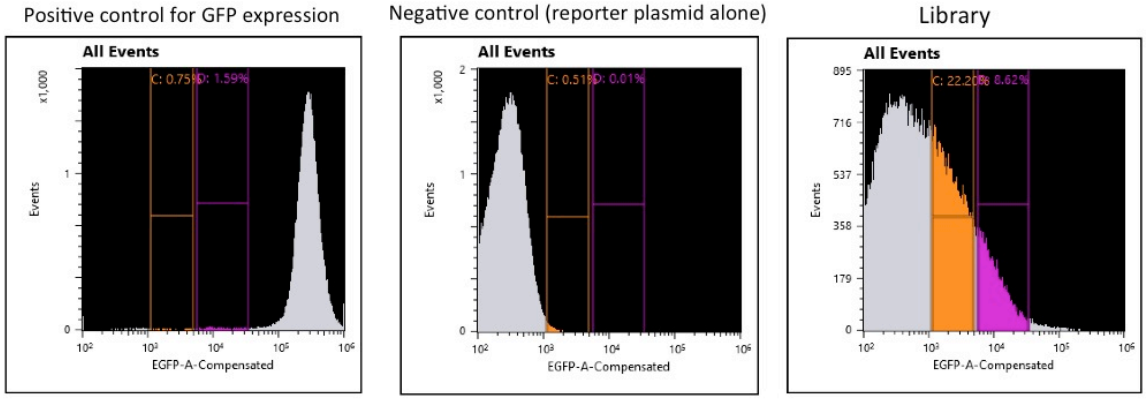
6. Re-transform into cells with pUC reporter plasmid and verify individual clones can turn on in the presence of sender cells

The scheme outlined in Figure 41 represents a two-stage screen: a negative screen followed by a positive screen after introducing a plasmid that encodes for the tRNAse protein that is delivered. The downside with this approach is that we do not guarantee that the positive screen occurs at the tRNAse concentration that is actually delivered during the CDI process. Alternatively, the positive screen could be performed after co-culturing sender cells with the receiver cells containing library members that passed the negative screening stage. The primary downside with this approach is that we actually plan on doing the eventual microscopy experiments on a solid format (i.e., the sender and receiver cells would be growing on an agar pad). We do not know if the concentration of tRNAse that is transferred is comparable between liquid and solid formats, and it

substantially more difficult to experimentally carry out the sorting step on cells that are growing on an agar pad (ie, we would need to develop a protocol to ensure that we are not scraping pieces of agar that could lead to clogging of the cytometer's fluidics). We thus chose a protocol that has much less uncertainty in our ability to carry it out, but our choice does lead to more biological uncertainty.

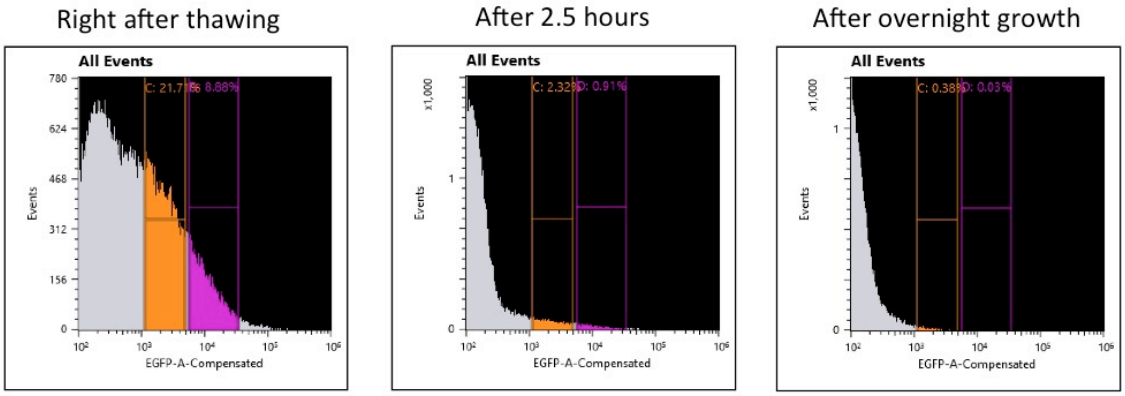
We eventually produced approximately 225,000 transformants using Golden-gate cloning to assemble the library (details in Materials and Methods). Using a Poisson model where $\lambda = \frac{225{,}000}{32^3} = 6.8665$ gives a probability of 0.1% that a member is not present. Assuming every member has an equal probability of being constructed, transformed, and for the host cell to grow, we expect a negligible amount of the possible library members to not be present. Figure 42 shows the distribution of GFP expression for the library after construction. The distribution is consistent with what we expect since a large fraction of the library is essentially off, while some cells are moderately dim, and a small fraction is almost as bright as the positive control cells. The library was then stored as aliquots in a $-80°C$ freezer, so that they could be further investigated.

Figure 42. Initial distribution of fluorescence of the library members. The left panel shows a positive control containing cells that express high levels of full-length T7 RNA polymerase and contain the reporter plasmid (pCDI629). The middle panel shows the fluorescence distribution for cells containing only the reporter plasmid. The right panel shows the distribution of fluorescence for the library after transforming the library into cells containing the reporter plasmid.



Unfortunately, we soon discovered that the cells with lower levels of fluorescence rapidly out-compete the dim and bright cells. Figure 43 shows the GFP distribution if we thaw an aliquot and immediately look at the GFP distribution of the cells, and we find that the distribution is quite similar to the distribution from Figure 42 after we first made the library. If we allow the cells from the aliquot to grow even just 2.5 hours, we find a strong selection for cells that have extremely low levels of GFP. If we allow the cells to grow overnight, the situation is even worse.

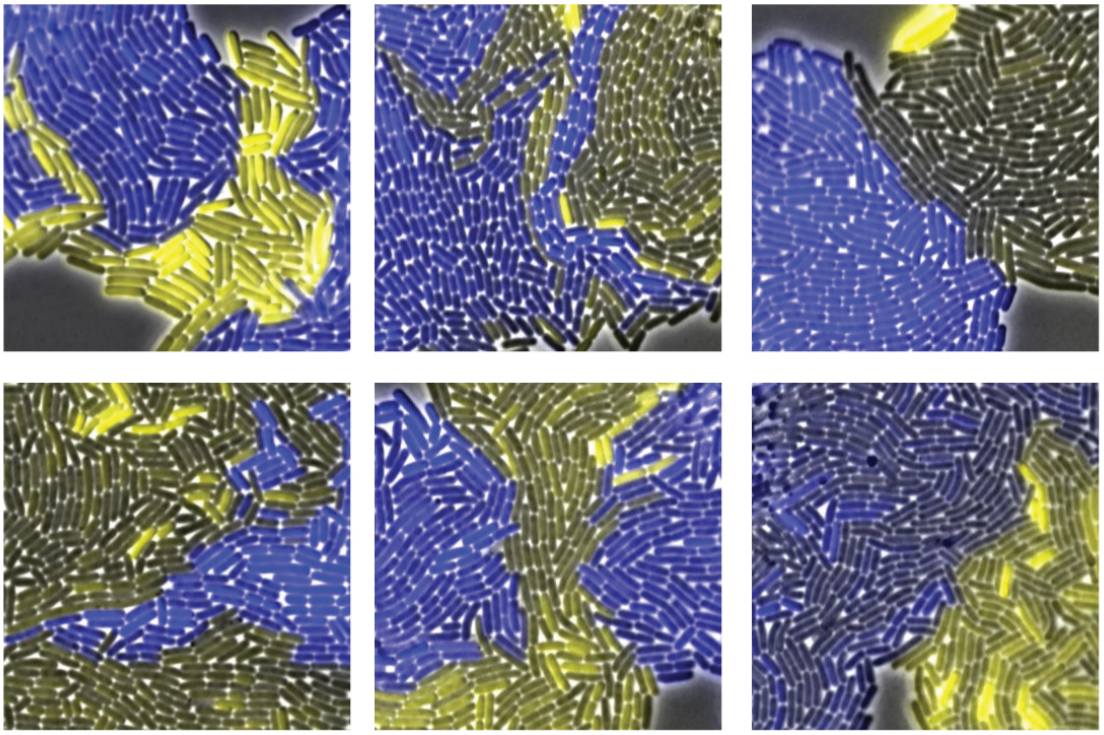
Figure 43. Strong selection for members of the library that don't express GFP. The left panel shows the fluorescence distribution of an aliquot immediately after thawing from the -80°C freezer. The middle panel shows the fluorescence distribution after growing for 2.5 hours after recovering from the freezer, and the right panel shows the distribution after allowing an aliquot to grow overnight.



The strong growth advantage for cells that have extremely low levels of GFP makes our proposed plan non-feasible because cells that never express GFP quickly take over the population at any stage of the screening process. For example, when we want to screen for dim cells before the tRNAse is expressed, we hope to identify cells that are capable of turning bright once tRNAse is later introduced, but the population will be dominated by cells that are OFF. Even if we manage to successfully pull out the desired sub-population of cells, once we introduce the tRNAse, we want to select out the cells that strongly express GFP. Again, the population will be overwhelmingly dominated by cells that do not express GFP, so it becomes extremely difficult to find cells with high GFP levels.

Although the library approach did not work, we thought that a randomly selected subset of the library members could result in a combination of parameters with the desired behavior. We thus isolated six colonies from the library, and allowed these candidates to grow on solid format near sender cells. Figure 44 shows the results of this experiment by examining the micro-colony development at a single time-point after approximately six hours of growth. The results do not qualitatively match our desired behavior of receiver cells at the boundary expressing much higher levels of GFP. Instead, consistent with the stochastic simulations, we find that noise can be problematic since it allows receiver cells that are not near a sender cell to switch to the ON state. We find many receiver cells that are not near a sender cell and express high levels of GFP while many receiver cells that are touching sender cells remain OFF.

Figure 44. Checking six library members on solid format co-culture with sender cells. The receiver cells (candidate library plasmids with pCDI629) are colored in yellow and the sender cells (pCDI596 and pCDI616) are colored in blue. Ideally, the receiver cells touching the sender cells would be much brighter.



3.3. Discussion

We attempted to utilize the recently discovered contact-dependent inhibition (CDI) system as a platform for inter-cellular communication. The challenge with the CDI system, however, is that much of the biology is still unknown. We still do not know what are the rules that govern the transfer process. For example, if we were to simply insert a few amino acids into the C-terminus (C-T) domain at the end of the toxin domain, can this modified C-T domain still make it into receiver cells? What controls the transfer rate of the C-T domains and how can we improve it? The most challenging aspect with answering these questions is that we can not simply fuse a fluorescent reporter tag to the C-T domain. For example, we attempted to fuse a small GFP fragment (< 20 amino acids) from the split GFP as a reporter for the cdiA-CT [96], but we failed to observe a GFP signal (unpublished results). We could not determine if the GFP signal was below our detection limit or if the fusion itself had disrupted the transfer process.

Development of a fluorescent reporter that measures the transfer rate of C-T molecules would allow one to quickly answer many of these types of biological questions. For example, one could generate a library of mutant C-T's that are tagged with the fluorescent reporter and then ask which of these can still be successfully delivered into receiver cells by sorting for fluorescent receiver cells with fluorescence-activated cell-

sorting (FACS) technology. Developing the fluorescent reporter system is a challenging step, however, since the only way to debug the process is to resort to immunofluorescence staining, such as used by Webb *et al* [79] to determine if the C-T domain made it into the cytoplasm of the receiver cells.

Once a fluorescent reporter system is developed, however, one could rapidly use this to build a system that controls gene expression. For example, we could use the reporter system to optimize the transfer rate to deliver as many molecules as possible. After that is accomplished, we could fuse on a small peptide that controls gene expression (such as the small split T7 RNA polymerase fragment we used) onto the fluorescent reporter. Assuming this fusion does not disrupt the fluorescent reporter, we could then assay for changes in gene expression caused by delivery of the small peptide and correlate to the output of the fluorescent reporter system. Thus, one could quickly determine if the fusion of the small peptide had significantly disrupted the efficiency of transfer. Since so much technology has been built around fluorescent reporters [97], we think that the development of a fluorescent reporter system to measure the CDI transfer process is a likely development to study biological questions, and this will dramatically enable efforts to engineer contact-based communication by methods such as our first approach of directly fusing a peptide to the C-T domain.

We also designed an alternative strategy for using the CDI system that does not rely on fusing any peptide sequences to the C-T domain. We exploit the known property that the delivered tRNAse co-localizes two proteins (CysK and CdiI) that can be expressed in the receiver cells [78]. If we fuse the two halves of a split protein to the CysK and CdiI proteins, co-localization of the CysK and CdiI proteins increases the probability that the split molecules can interact and leads to an increase in activity. In order to amplify the response we obtain from co-localization, we designed a circuit with positive feedback to provide an ultrasensitive response upon delivery of the tRNAse molecules. The key challenge with this scheme is that the two split proteins can bind and interact in the absence of the tRNAse scaffold, so we can not simply crank up the expression levels of the split proteins. We show with analytical models and numerical simulations that the desired response is achievable with biologically feasible parameters.

We attempted to probe a large region of the parameter space for our proposed circuit by searching three different parameters at once ($K_a, K_b,$ and $basal_{mRNA}$ from Equations 13). To determine which parameter sets provide the desired functionality, we designed a procedure using fluorescence-activated cell-sorting (FACS) to identify the individual cells that contained plasmids with the desired parameter values. Unfortunately, our scheme has a fatal flaw that cells without any fluorescence have a substantial growth advantage that allows them to quickly take over the population. There are two possible routes to combat this problem. We suspect that the growth disadvantage comes from the T7 RNA polymerase stealing away too many of the host *E. coli* cell's resources, so one approach is to re-engineer the host cell or the T7 RNAp to mitigate this. For example, if we are depleting the cell of a metabolite because of the T7 RNAp, we could over-express the enzymes used for production of the limited metabolite. In addition, we could express lower levels of the split T7 RNAp molecules to minimize the burden on the host cell,

although issues with stochasticity could become a bigger problem. An alternative approach is to avoid pooling any of the different parameter combinations by building and monitoring individual clones separately. One could choose promoters and 5' UTR regions of known strengths from the BioFAB parts to probe the parameter space in a guided manner [86]. The primary downside with this approach is that we would need to order the DNA for each parameter set whereas our library approach allowed us to order one set of degenerate oligos that represented all the possible library members. With current costs of DNA synthesis, the cost of building a reasonable number of parameter combinations is prohibitive today, although we think such a methodology could become routine in the next decade.

The key parameter with the co-localization scheme is the binding affinity between the two split proteins without the tRNAse scaffold compared to the binding affinities of CysK and CdiI to the tRNAse. Future improvements in protein engineering may make our problem dramatically easier. For example, mutations to the tRNAse that improve the binding affinities to CysK and CdiI would directly improve the system's performance. We might also be able to rationally design domains that reduce the binding affinity in a specified manner. For example, Shekhawat *et al* designed coiled-coil domains to fuse on to split firefly luciferase proteins so that the reporters are only active once a TEV protease is present in the cell to cleave off one of the autoinhibitory coiled-coil domains [98]. While this strategy has not been shown with the split T7 RNA polymerase yet, we may be able to understand the design rules for the autoinhibitory domains so that we can predictably apply them to new proteins.

Despite the challenges we faced with engineering contact-based communication in *E. coli*, this type of technology would provide a spatial resolution almost three orders of magnitude better than what can be achieved with the quorum sensing molecules that are commonly used currently. For example, Basu *et al* used acyl-homoserine lactone (AHL) molecules with a band-pas filter to demonstrate bulls-eye patterns, and the authors could tune the radius of their patterns from two to eight millimeters [68]. Ortiz and Endy recently demonstrated a novel communication channel based on delivering phage particles from one *E. coli* cell to another [99]. Ortiz and Endy chose to use Bacteriophage M13 since the sender cells can package M13-based viral particles that contain a desired DNA payload, and the sender cells can then secrete the viral particles without destroying themselves. Ortiz and Endy calculate a diffusion constant for their M13 system that is two orders of magnitude slower compared to the AHL quorum-sensing molecule. In the future, we envision that a contact-based communication channel can complement these other methods so that one could mix and match the appropriate communication methods to match the spatial requirements.

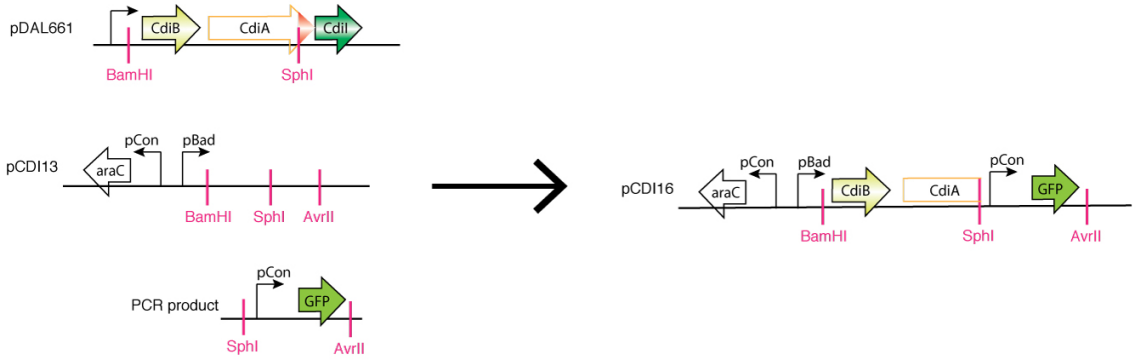
3.4. Materials and Methods

3.4.1. Molecular cloning and construction of strains/plasmids

We constructed pCDI16 to serve as a base into which we can insert the desired domain downstream of the VENN domain. Because the *cdiA* gene is so large, we avoid having to

PCR a 9-kb gene and instead resort to traditional restriction enzyme-based cloning. We exploit the presence of an SphI site that is located 89 bp upstream of the VENN motif (there is no other SphI site present in the *cdiA* or *cdiB*). Figure 45 provides an overview of the steps to build pCDI16. We obtained pDAL661, which contains the EC93 *cdiBAI* operon, from David Low's laboratory and used BamHI/SphI to move everything from *cdiB* up to the SphI site of *cdiA* into our own vector. Starting from the pBad vector on pUC that was previously described (pDC298), we modified the restriction sites so that a unique BamHI site is located after the pBad promoter to make pCDI13. Finally, we insert a GFP cassette between the SphI site and the unique AvrII site on the vector to make pCDI16.

Figure 45. Overview of cloning to build pCDI16, which allows us to easily insert in-frame translational fusions for the tRNAse domain or other desired payload.



The key feature of pCDI16 is that we can easily replace the GFP cassette with whatever payload we wish to fuse in-frame to the EC93 CdiA main-body. In general, we build these constructs by using Splicing by Overlap Extension (SOE) PCR to fuse a PCR product that contains the 89 bp between the SphI site and the VENN motif to another PCR product that starts from the VENN motif to ends at the translational stop that marks the end of the modified *cdiA*. For example, to build pCDI233 that encodes for the natural tRNAse domain, we use o25/o379 to PCR off a plasmid that contains the original EC93 *cdiA* (e.g. pDAL661) and use o378/o387 to PCR the tRNAse domain off a plasmid that contains the UPEC536 *cdiA* (e.g. pDAL858). The two PCR products contain 20 bp that overlap, so we can then fuse the PCR products together via SOE-PCR where we use both of the previous PCR products as the template. All PCR reactions used Phusion Hot-start II High-Fidelity Polymerase (Thermo Scientific F-549L). The key for SOE PCR to work properly is that the PCR products from the first PCR reactions must be purified with gel extraction to minimize carryover of the oligos from the first reaction to the SOE PCR reaction. Finally, we cut pCDI16 and the final fused PCR product with SphI/AvrII and ligate the PCR product into pCDI16. Note that we do not need to purify pCDI16 vector from the GFP cassette by gel extraction since we can simply identify colonies the next day that do not glow (i.e., express GFP). We verify that the sequence between SphI and AvrII is correct with commercial Sanger sequencing and verify that the remaining vector is the correct length by restriction digest mapping.

We also built pCDI51, which contains a strong consensus RBS between the end of *cdiB* and the start of *cdiA* by using a BamHI/HindIII replacement and o90-o93. We also made a p15a equivalent plasmid for pCDI16 by using a BamHI/AvrII cut to move the *cdiB*, *cdiA*, and the GFP cassette that is contained between the SphI and AvrII sites. This results in pCDI554, which contains pBad driving both *cdiB* and *cdiA* (with the GFP cassette that can be replaced with a SphI/AvrII cut). We then built pBad-cdiBA(H178A), where the conserved histidine residue has been mutated to alanine to destroy the tRNAse catalytic activity, with o378/o401 off pCDI484 and inserting into pCDI554. Finally, we introduce the H178A mutation into a high-copy plasmid that expresses cdiBA by moving the SphI/AvrII region of pCDI554 into pCDI51 to result in pCDI616.

To make pCDI657, which contains the small T7 RNAp fragment fused to 103aa of the tRNAse domain and is shown in Figure 27, we use SOE PCR with o1278/o1319 off pCDI328 (which contains the entire T7 RNA polymerase) and o25/o1277 off pCDI233. To make pCDI658, which contains the small T7 RNAp fragment fused to the end of the tRNAse domain, we use SOE PCR with o1294/o1319 off pCDI328 and o362/o25 off pCDI233.

In order to use pCDI657 or pCDI658, the receiver cells must also constitutively express the large T7 RNAp fragment via pCDI679 or pCDI680. We express the large T7 RNAp fragment with a strong RBS as designed using the Salis RBS calculator [95] and also use the strong J23100 constitutive promoter [94]. For pCDI680, we fuse the large T7 RNAp fragment to the entire CdiI with a (GS)x5 linker sequence. We note that we also incorporate the R632S mutation as previously described by the Voigt Lab [100]. Both pCDI679 and pCDI680 were built using oligos designed using J5 and constructed with Golden-gate cloning as per the protocol suggested by J5 [101].

To make the reporter plasmid for the receiver cells of the T7 experiments, we originally constructed pCDI624. pCDI624 contains an operon with the T7 promoter driving sfGFP and a separate operon with a strong constitutive promoter driving BamA, AcrB, and OmpF. BamA and AcrB were previously shown to be required for delivery of the EC93 toxin into target cells [102], and we thought that their over-expression could help improve transfer rates. Over-expression of OmpF was suspected to improve transfer rates of the tRNAse toxin (email from Christina Beck, a graduate student in Christopher Hayes' laboratory). However, a later email from Christina Beck indicated that with our construct with the EC93 CdiA main body and the UPEC536 tRNAse domain, over-expression of OmpF will not help. In addition, they found that over-expression of AcrB did not improve delivery, but that over-expression of BamA did improve binding and delivery. Thus, we built pCDI629 that contains only the T7 promoter driving sfGFP and a separate operon with a constitutive promoter driving only BamA (ie, pCDI629 was derived from pCDI624). Both pCDI624 and pCDI629 were built using oligos designed using J5 and constructed with Golden-gate cloning as per the protocol suggested by J5 [101].

In order to test the scaffolding behavior in a single cell, we constructed pCDI484 in order to induce low amounts of the tRNAse domain. We started with pSB4K5, which contains

the BioBrick restriction sites on a pSC101 plasmid with a kanamycin resistance marker. We then used an EcoRI/PstI restriction digest to move the araC-pBad cassette into the pSB4K5 vector. This cassette includes a BamHI site after the pBad promoter. This allowed us to construct pCDI434, which has pBad controlling expression of the tRNAse domain, by using o401/o682 off pCDI233. In this construct, we do not explicitly incorporate a sequence that resembles a ribosome binding site since we want low expression levels of the tRNAse. Finally, we incorporate the H178A mutation to remove the tRNAse catalytic activity with o874/o401 and o875/o682 off pCDI434 and use SOE PCR to fuse the two PCR products together.

We built an intermediate plasmid, pCDI590, which constitutively expresses tetR repressor and contains the Ptet promoter controlling expression of both split proteins: CysK fused via GSGSDGSGSG to T7C and T7N fused via (GS)x5 to CdiI. pCDI590 was built using oligos designed in J5 and constructed with golden-gate cloning.

We also built pCDI604 (the p15a plasmid shown schematically in Figure 29) to test the spatial co-localization idea. pCDI604 was built using oligos designed using J5 with pCDI590 serving as a template for the CysK-T7C and T7N-CdiI fusion proteins. The mRFP cassette was obtained from pSB4K5-BBa_J04450 that is available in the iGEM parts registry. The drop-in pieces were constructed using o1313, o1315, o1316, and o1318 for apFAB311, apFAB347, apFAB117, and apFAB115, respectively. For example, to build pCDI651, which contains apFAB311 and apFAB117, we used o1313 and o1316 to PCR a fragment (off a template containing the double terminator), and then digested this fragment and pCDI604 with EcoRI/XhoI. We then selected for colonies that were not red the next day.

In order to build the combinatorial library for an ultrasensitive response to spatial co-localization (pCDI685), we PCR'd three separate parts: o1367/o1368 with pCDI629 contains the promoter library and the T7 promoter, o1369/o1370 off pCDI590 contains CysK-T7C with an RBS library, and o1371/o1372 off pCDI590 contains T7N-CdiI with its own RBS library. The three parts were combined via Golden-gate cloning and transformed into CDI299 cells containing pCDI629 that were prepared to be electrochemically competent. After recovery in SOC for 45 minutes in a shaking, 37°C incubator, the cells were spun down and re-suspended in the growth medium used for experiments (M9 minimal medium) and then allowed to grow overnight in the shaking, 37°C incubator. To make a positive control for the library (pCDI685), we ordered oligos forcing a strong constitutive promoter and a strong RBS for both split proteins. Thus, pCDI686 was made using o1373/o1374 off pCDI629, o1375/o1376 off pCDI590, and o1377/o1378 off pCDI590, and then the PCR reactions were combined via Golden-gate cloning as was done for the library. pCDI596 was built using oligos designed using J5 and contains a strong promoter driving both mRFP and the CdiI protein that provides immunity against tRNAse; this plasmid was co-transformed into sender cells that contain pBad-cdiBA on a high-copy plasmid.

Since CysK is normally expressed in *E. coli* cells, experiments with co-localization require a strain that does not contain a chromosomal copy. We obtained strain JW2407,

which contains a Kanamycin resistance cassette integrated inside the *cysK* gene, from the Keio collection [103]. We used P1 phage transduction to transfer the *cysK::kanR* knockout into BW27783 and then cured the kanamycin resistance cassette by transiently expressing FLP recombinase using pKD46 [104]. This resulted in strain CDI299, which we transformed our plasmids into to conduct our experiments. Propagation of the plasmids for cloning was done in Tg1 except constructs that contained the non-mutant tRNAse domain were propagated in CDI299 to prevent toxicity.

Table 5 lists the plasmids and strains that were previously described in cloning for the circuits that were used to demonstrate communication with the CDI system. Table 6 lists the oligos that were described above in cloning of the plasmids. Note that the numbering for the oligos starts separately from the oligos shown for the sequestration-based bistability project and that we do not list oligos that were designed using the J5 program. We used the following parameters for J5: golden_gate_cut_site=GGTCTC, Terminus_extra_seq=CACAACGGTCTCA, Max_identities_gg_overhangs_compatible=2, Min_pcr_product_bps=70, Primer_gc_clamp=2, Primer_min=2, Primer_max=36, Primer_min_tm=60, Primer_max_tm=70, and Primer_max_diff_tm=5. The oligos for the combinatorial library (pCDI685) were designed by hand since J5 does not currently support degenerate oligos.

Table 5. Plasmids and strains used for the cell communication circuits

Plasmids

Name	Resistance	Origin	Description
pJT026	ampR	pMB1	Construct that was made previously in the Arkin Lab for strongly expressing mRFP in *B. subtilis* cells via the hyperspank promoter. Also expresses strongly in *E. coli*
pCDI16	ampR	pMB1	Contains pBad driving both *cdiB* and *cdiA*. Only contains *cdiA* up to the SphI site and then contains a GFP cassette
pCDI51	ampR	pMB1	Same as pCDI16 except the RBS for *cdiA* has been changed to a consensus AAGGAGGAA followed by a six bp spacer before the ATG start for *cdiA*
pCDI233	ampR	pMB1	pBad driving *cdiB* and *cdiA* with the wild-type tRNAse downstream of the VENN site
pCDI328	cmR	p15a	Constitutive promoter driving full-length T7 RNA polymerase with a C-terminal TEV cleavage site and a degradation tag.
pCDI434	kanR	pSC101	Contains pBad driving the wild-type tRNAse domain (from VENN to end)
pCDI484	kanR	pSC101	Contains pBad driving the tRNAse domain with H178A mutation (same as pCDI434 except for one changed residue)
pCDI558	cmR	p15a	Contains pBad driving both *cdiB* and *cdiA*. Only contains cdiA up to the SphI site and then contains a GFP cassette (same as pCDI16 but on a p15a origin)
pCDI590	cmR	p15a	Ptet driving both CysK-T7C and T7N-CdiI
pCDI596	cmR	p15a	Strong constitutive promoter driving *cdiI* and *mRFP*

pCDI604	cmR	p15a	Shown schematically in Figure 29. Contains an mRFP cassette between the EcoRI and XhoI sites to allow easy drop-in replacement of new promoters.
pCDI616	ampR	pMB1	pBad driving both *cdiB* and *cdiA* (H178A)
pCDI624	ampR	pMB1	Contains one operon with the T7 promoter driving *sfGFP* and a separate operon with a strong constitutive promoter driving *bamA*, *acrB*, and *ompF*
pCDI629	ampR	pMB1	Contains one operon with the T7 promoter driving *sfGFP* and a separate operon with a strong constitutive promoter driving *bamA*
pCDI651	cmR	p15a	pCDI604 with apFAB311+apFAB117 promoters dropped in
pCDI652	cmR	p15a	pCDI604 with apFAB311+apFAB115 promoters dropped in
pCDI653	cmR	p15a	pCDI604 with apFAB347+apFAB117 promoters dropped in
pCDI654	cmR	p15a	pCDI604 with apFAB347+apFAB115 promoters dropped in
pCDI657	ampR	pMB1	Contains pBad driving both *cdiB* and modified *cdiA*; CdiA contains only first 103aa of the tRNAse domain, (GS)x5 linker, and then a fusion to the first 67 amino acids of T7 RNAp
pCDI658	ampR	pMB1	Contains pBad driving both *cdiB* and modified *cdiA*; cdiA contains the entire wild-type tRNAse domain, (GS)x5 linker, and then a fusion to the first 67 amino acids of T7 RNAp
pCDI679	cmR	p15a	Strong constitutive promoter driving large T7 RNAp fragment (amino acid 68 until the end)
pCDI680	cmR	p15a	Strong constitutive promoter driving CdiI fused via a (GS)x5 linker to the large T7 RNAp fragment (amino acid 68 until the end)
pCDI685	cmR	p15a	Contains two promoters that drive a two-gene operon. The first promoter is built using an oligo that contains 32 possible members and the second promoter is the T7 promoter. The first gene encodes for CysK-T7C and allows for 32 possible RBS. The second gene encodes for T7N-CdiI and also allows for 32 possible RBS.
pCDI686	cmR	p15a	Similar to pCDI685: instead of degenerate oligos for the first promoter and the two RBS, we choose a strong promoter and strong RBS for both genes.

Strains

Name	Genotype/Plasmids	Description
BW25113		Starting strain used in Keio collection
BW27783	BW25113 $DE(araFGH)\ \Delta araEp$ $P_{CP8} - araE$	Deletion of araFGH and araE is constitutively expressed

CDI299	BW27783 with $\Delta CysK$	Strain used for co-localization experiments

Table 6. Oligos used in cloning for the cell communication circuits. For denotes forward primer and rev denotes reverse primer.

Name	Description	Sequence
o25	Upstream of SphI site on CdiA main-body (for)	tgagtgaacaggaacgtcagca
o90	BamHI site (for)	ATTTCTGggatccCACCCCA
o91	End of CdiB (rev)	tgcatagatccTTCCTCCTTttaaaacgcgacggcaacg
o92	Puts consensus RBS in front of CdiA (for)	AAGGAGGAAggatctatgcatcagcctcccgtt
o93	Downstream of HindIII site (rev)	agctgcgtcggattaagctta
o362	end of tRNAse (rev, includes (GS)x5 linker)	gctgccgctaccgctaccgctgccgctaccTATTCCATATCCTTTCAAGGCTG
o378	tRNAse (for, includes overhang to match VENN motif from EC93 CdiA)	actcggttgagaataatgcaCTGAGTCTGGTTGCCAGAGG
o379	CdiA main body up to VENN motif (rev)	CCTCTGGCAACCAGACTCAGtgcattattctcaaccgagt
o387	tRNAse (rev, includes AvrII site)	tatgtCCTAGGttatattccatatcctttcaaggctg
o401	tRNAse (rev, includes XhoI site)	cattactcgagttatattccatatcctttcaaggctg
o682	tRNAse domain (for, starting at VENN sequence)	acataGGATCCatgGTTGAGAATAATgcaCTGAG
o874	tRNAse H178A mutation (for)	GAGAATGGAGGATATTGGGATGCTATGCAGGAAATGCAAAATAC
o875	tRNAse H178A mutation (rev)	GTATTTTGCATTTCCTGCATAGCATCCCAATATCCTCCATTCTC
o1277	Reverse of tRNAse at 103aa	ttagcgatgttaatcgtgtttcctgagccggaaccg
o1278	Small T7 RNAp (for)	aacacgattaacatcgctaagaacg
o1294	Small T7 RNAp fragment (for, includes (GS)x3 linker)	ggtagcggtagcggcagcAACACGATTAACATCGCTAAG
o1313	Promoter apFAB311	acatagaattcTCCACACAACCTACGAGCCGGATGATTAATTGTCAA
o1315	Promoter apFAB347	acatagaattcTCCACACATATTAAGAGCCGGATGATTAATTGTCAA
o1316	Promoter apFAB117	tacagCTCGAGATGAAATAATTATGCAGAAAAATTTTCCTGATGTCGA

o1318	Promoter apFAB115	tacagCTCGAGATCCACACATTATACAGAAAAATTTTCCTGATGTCGA
o1319	Small T7 RNAp fragment (rev)	cattaCTCGAGcctaggttagttatccgcaacctcaccag
o1367	Pcon.lib+ pT7 (for)	GCATGGTCTCAgtccTAGGGAYTATGCTAGCCGCAACGCAATTAATGTAAG
o1368	pT7 (rev), includes part of RBS lib for CysK-T7C	ATGCGGTCTCActttCSKCCTTTRATATGSGATTTTCTCCCTATAGTGAGTCGT
o1369	CysK-T7C (for, with RBS lib)	GCATGGTCTCAaaagGYAATAATATGAGTAAGATTTTTGAAGATAACTCGCTGAC
o1370	CysK-T7C (rev, includes pat of RBS lib for T7N-CdiI)	ATGCGGTCTCAAATCCSGTGGGCTGGAGTTACGCGAACGCGAAGTC
o1371	T7N-CdiI (for, with RBS lib)	GCATGGTCTCAGATTAATTAARCAGSTAGSCSATGAACACGATTAACATCGCTAAG
o1372	T7N-CdiI (rev), includes part of Pcon.lib	ATGCGGTCTCAggacTGAGCTAGCYRTMARagccggaagcataaagtg
o1373	Pcon (strong) with pT7 (for)	GCATGGTCTCACCTAGGTACAGTGCTAGCCGCAACGCAATTAATGTAAG
o1374	pT7 (rev, includes strong RBS for CysK-T7C)	ATGCGGTCTCACTCATttttatcctccttctccctatagtgagtcg
o1375	CysK-T7C (for, with strong RBS)	GCATGGTCTCATGAGTAAGATTTTTGAAGATAACTCGCTGAC
o1376	CysK-T7C (rev, includes strong RBS for T7N-CdiI)	ATGCGGTCTCAgttcatattattacctcctTTACGCGAACGCGAAGTC
o1377	T7N-CdiI	GCATGGTCTCAgaacacgattaacatcgctaagaacg
o1378	T7N-CdiI (rev, includes part of strong Pcon)	ATGCGGTCTCATAGGactgagctagccgtcaaagccggaagcataaagtg

3.4.2. Growth conditions

Experiments in liquid media were conducted in M9 minimal medium, which were prepared as described for the sequestration-based bistability experiments. Cells were grown in a shaking incubator at 37°C.

Experiments in solid format were also done in M9 minimal medium. Before addition of any inducers or antibiotics, 5 mL of the M9 minimal medium and 76 mg of agarose (Fisher BP164) were microwaved for a few seconds at a time. After every few seconds, we checked that the agarose was fully dissolved. Once fully dissolved, the tube containing the M9 minimal medium with 1.5% agarose was placed in a warm water bath at 60°C. At this time, appropriate antibiotics and inducers were added. For time-lapse experiments, the medium was placed in a silicon insulator (Invitrogen P18174, 20x0.5 mm) and sealed with coverslips on both sides of the silicon insulator. After allowing half an hour for the agar pad to dry, a small drop (~0.7 uL of cells) was placed on the pad and allowed to dry, and then the pad was transferred to a Wilco dish (Wilco Wells D3522P, 35/22 mm, #1.5). For experiments where we did not allow the cells to grow overnight and

wanted higher throughput, the M9 minimal medium with 1.5% agarose was transferred to a silicone isolator with 8 wells (Invitrogen P24744, 9x1.0mm) and sealed with 60x40mm coverslips.

3.4.3. Fluorescence microscopy

Time-lapse experiments were imaged on a Deltavision system with an Olympus IX70 and UPlanApo 100x/1.35 oil objective. Illumination was with a Xenon lamp module (API part # 34-100390-002-9) with a Chroma 89006 dichroic and filter set. Cells were imaged every 15 minutes, typically with the following settings: ND filter=50% and .05 sec exposure for the bright-field image, ND filter=32% and .25 sec exposure for the GFP image, and ND=10% and .08 sec exposure for the RFP channel (RFP was expressed at a high concentration to keep track of sender cells). The cells were kept at 30°C so that we could image the cells twice per doubling (we previously had issues with photo-toxicity if imaging was done with a shorter period than 15 minutes).

Experiments that did not require time-lapse were done on a Zeiss Axio Observer D1 using a 63x plan-apochromat Ph3 oil-immersion objective. Cells were imaged with phase contrast and with the appropriate fluorescence filter set (38 for GFP/YFP and 45 for RFP) with the following typical exposure times (40 ms for RFP and 145 ms for GFP). Fluorescence excitation was provided with an X-CITE 120XL. Images were captured with an AxioCam MRm using the provided Zeiss Axiovision software. Images were processed using the same Matlab software that was described for analyzing the sequestration-based bistability experiments.

3.4.4. Flow sorting

Flow sorting was performed on a Sony SH800 with the 100 micron sorting chip. Cells were analyzed with the following parameters: FSC 9, BSC 30.0%, Threshold with FSC and value of 1.50%, and FL1 for GFP at 40.4% gain and acquiring area information. Cells were sorted into 15 mL conical tubes until either output tube was full (typically at 3 million cells using Purity sorting mode). The cells were sorted based on GFP intensity with bin boundaries set by comparing to positive and negative cells' GFP intensities.

3.4.5. Other instruments

Bulk fluorescence measurements were performed on a Tecan Safire 2 (OD measured at 600nm; GFP measured with 488/510 excitation and emissions wavelengths and 5nm bandwidths, 60 gain, 10 reads, high sensitivity flash mode, and 100 μs integration).

3.4.6. Computational

Analytical models were investigated in Mathematica. Nullcline plots were made using StreamPlot and manually setting coordinates to draw the StreamPoints. Numerical simulations were performed in Matlab. The bifurcation curves were computed by finding the steady-state for the ODE system. We probed the system by varying $tRNAse_{tot}$ from 10^{-3} up to $10^{2.5}$ in 100 logarithmic increments and computing up to two steady states for each value of $tRNAse_{tot}$. We probed the system starting at $tRNAse_{tot} = 10^{-3}$ by trying to identify a low steady state with all seven species starting at a concentration of .01 and

trying to identify a high steady state with all seven species starting at a concentration of 100. At each increment of $tRNAse_{tot}$, we attempt to find the next low steady-state by setting all the initial concentrations at 0.9 the values of the previous low steady-state. Similarly, we attempt to find the next high steady-state by setting all initial concentrations at 1.1 times the value of the previous high steady-state. To compute the steady-state, we iterate between two methods. In the first approach, we use Matlab's ode15s to simulate the system forward and check if the system is no longer moving. In the second approach, we use Matlab's fsolve with the default trust-region-dogleg solver with a Jacobian computed via finite-differences. If we successfully identify a steady-state where all concentrations are not negative and all real eigenvalues are not positive, we return the steady-state solution. Otherwise, we use the final concentrations of the previous ode15s run as the input to the next ode15s simulation to run for a longer period of time and attempt the two methods again.

Stochastic simulations were performed in Matlab using the exact Gillespie algorithm [92]. The ODE system shown in Equations 13 can be converted to a stoichiometric matrix containing 21 reactions and eight species; we add unbound tRNAse as another species and convert each binding reaction into two reactions for each direction. At the start of the simulation, we assume $tRNAse_{tot}$ is the concentration of unbound tRNAse. We ran each simulation until 100,000 reactions have fired, storing the concentration of all species after every 10 reactions.

4. Conclusion

The goal of engineering is to build useful solutions to real-world problems. While the previous sentence sounds simple enough, this process is composed of many steps with possible subtleties. For most problems, we must begin by formalizing the problem and abstracting it; this involves writing down the assumptions and constraints. With this understanding of the problem, we can express the problem mathematically and test possible designs on a computer. For most problems, we do not fully trust the *in silico* results, so we build a scale model that we use to thoroughly look for issues that have not yet been considered. After we are confident that our solution will work at scale, we build the actual solution to the problem that we originally desired to solve.

While we would like to apply the previously described sequence of steps to the engineering of genetic circuits, biology is fraught with challenges that are unique from other engineering fields [8,105]. Genetic engineers are faced with issues such as parasitic interactions of the circuit with the host, stochastic fluctuations of concentrations, and evolutionary stability of the circuit. The real crux of the challenges in biology is uncertainty about what is happening [12]; we are working in extremely complex systems with many moving parts, but we are extremely limited in our knowledge of the underlying mechanisms. Furthermore, debugging biological circuits is a non-trivial endeavor; recent developments allow us to monitor expression levels across the genome, but these experimental procedures are extremely labor-intensive today [106–108].

Despite these challenges, our desire was to bring genetic engineering closer to the workflow as practiced in other engineering disciplines. Towards this goal, we designed genetic circuits for memory and communications. We were more successful in our demonstration of the memory elements with sequestration-based bistability, and we attribute some of this to luck that the possible unknown biological interactions did not over-whelm the circuit's ability to function as designed. For example, sigma factors from *B. subtilis* can kill *E. coli* host cells [109], so we were fortunate to have chosen a pair of sigma factor with a cognate anti-sigma factor that does not exhibit unknown toxic interactions. We were also fortunate that the sigma factor and anti-sigma factor bind tightly inside *E. coli* cells and can strongly activate the cognate promoter.

We were less successful in our demonstration of circuits for contact-based communication. In the first strategy of directly fusing a small protein fragment to the end of the tRNAse that is delivered, we failed to see any indications that the small fragment was being delivered to the receiver cells. We suspect that our modifications to the tRNAse prevent its delivery via the CDI system. In the second strategy of using the tRNAse as a scaffold for co-localizing two split proteins, we were hamstrung by the parasitic growth interactions of our circuit with the host and the binding of the split proteins in the absence of the tRNAse.

For a field to be engineerable in a rational manner, there must be modules with specified interfaces that allow the designer to abstract away everything on the other side of the module boundary. The incredibly complex software systems that are built today are enabled by the widespread use of modules that other engineers can re-use and can be sure will function as desired [110]. Inspired by these successes in other engineering fields, synthetic biology seeks to bring the reusable plug-and-play workflow to genetic engineering. The current trend is to define genetic modules as parts or devices, where devices are composed of parts.

Our work provides candidate modules at both the device and parts level. For our sequestration-based bistable circuit, we can consider the memory device as a modular unit, where the inputs are expression of the sigma factor and anti-sigma factor and the output is transcriptional activity from the promoter recognized by the sigma factor. Alternatively, the pair of sigma factor and anti-sigma factor proteins may serve as parts to implement molecular sequestration for other uses besides bistability. For example, if we have a system with constitutive expression of the anti-sigma factor and then begin the system at time 0 with sigma factor expression, there is a delay before the sigma factor concentration surpasses the threshold set by the anti-sigma factor concentration. Ongoing research in the Arkin Lab will show if we can use this delay to program a gene to begin expression a set amount of time after another gene.

The contact-based communication system also provides possible modules that can serve as reusable building blocks. We can define all the proteins used in the CDI process, including any proteins necessary in the sender and receiver cells, as a communication module. The input would be expression of the proteins necessary for the CDI transfer process, and the output would be expression of a gene that the CDI transfer system ultimately can control. By using memory-based switches to control the contact-based communication device, we could program bacteria to form spatial communities based on complex decisions that depend on past events.

We also think that the internal components could be re-used in other projects. In particular, the design we presented for an ultrasensitive response to spatial co-localization may find uses in other areas. For example, two-hybrid technologies are currently in development for use in drug discovery by detecting binding of a ligand, but these technologies are hampered by limited sensitivity [111]. Our design with positive feedback, while originally designed for use in contact-based communication, could also find use in these other applications since the common goal is enhancing a response to protein fragments that are spatially co-localized.

Ultimately, we hope that the circuits we have designed in this dissertation will provide modules for other engineers to build upon. While our knowledge of the underlying pieces is still incomplete, we have provided some experimental evidence for their functionality. In addition, we provide detailed investigations with analytical models and numerical simulations that can guide future researchers as they design more complex genetic circuits.

References

1. Berg, P. & Mertz, J. E. Personal Reflections on the Origins and Emergence of Recombinant DNA Technology. *Genetics* **184,** 9–17 (2010).
2. Arkin, A. P. Synthetic cell biology. *Curr. Opin. Biotechnol.* **12,** 638–44 (2001).
3. McAdams, H. H. & Arkin, A. Gene regulation: Towards a circuit engineering discipline. *Curr. Biol.* **10,** R318–R320 (2000).
4. Haseltine, E. L. & Arnold, F. H. Synthetic gene circuits: design with directed evolution. *Annu. Rev. Biophys. Biomol. Struct.* **36,** 1–19 (2007).
5. Gardner, T. S. & Hawkins, K. Synthetic Biology: evolution or revolution? A co-founder's perspective. *Curr. Opin. Chem. Biol.* **17,** 871–7 (2013).
6. Hartwell, L. H., Hopfield, J. J., Leibler, S. & Murray, A. W. From molecular to modular cell biology. *Nature*
7. Alon, U. Biological Networks: The Tinkerer as an Engineer. *Science* **301,** 1866–7 (2003).
8. Brophy, J. A. N. & Voigt, C. A. Principles of genetic circuit design. *Nat. Methods* **11,** 508–20 (2014).
9. Gardner, T. S., Cantor, C. R. & Collins, J. J. Construction of a genetic toggle switch in Escherichia coli. *Nature* **403,** 339–42 (2000).
10. Elowitz, M. B. & Leibler, S. A synthetic oscillatory network of transcriptional regulators. *Nature* **403,** 335–8 (2000).
11. Lucks, J. B. & Arkin, A. P. The hunt for the biological transistor. *IEEE Spectr.* **48,** 38–43 (2011).
12. Lucks, J. B., Qi, L., Whitaker, W. R. & Arkin, A. P. Toward scalable parts families for predictable design of biological circuits. *Curr. Opin. Microbiol.* **11,** 567–73 (2008).
13. Purnick, P. E. M. & Weiss, R. The second wave of synthetic biology: from modules to systems. *Nat. Rev. Mol. Cell Biol.* **10,** 410–22 (2009).
14. Burrill, D. R. & Silver, P. A. Making Cellular Memories. *Cell* **140,** 13–8 (2010).
15. You, L., Cox, R. S., Weiss, R. & Arnold, F. H. Programmed population control by cell–cell communication and regulated killing. *Nature* **428,** 868–71 (2004).
16. Saeidi, N. *et al.* Engineering microbes to sense and eradicate Pseudomonas aeruginosa, a human pathogen. *Mol. Syst. Biol.* **7,** (2011).
17. Strogatz, S. H. *Nonlinear Dynamics And Chaos: With Applications To Physics, Biology, Chemistry, And Engineering*. (Westview Press, 2001).
18. Ferrell, J. E. Self-perpetuating states in signal transduction: positive feedback, double-negative feedback and bistability. *Curr. Opin. Cell Biol.* **14,** 140–8 (2002).
19. C.D., T. A model for a bistable biochemical trigger of mitosis. *Biophys. Chem.* **57,** 239–51 (1996).
20. Pomerening, J. R., Sontag, E. D. & Ferrell, J. E. Building a cell cycle oscillator: hysteresis and bistability in the activation of Cdc2. *Nat Cell Biol* **5,** 346–51 (2003).
21. Xiong, W. & Ferrell, J. E. A positive-feedback-based bistable `memory module' that governs a cell fate decision. *Nature* **426,** 460–5 (2003).

22. Yao, G., Lee, T. J., Mori, S., Nevins, J. R. & You, L. A bistable Rb-E2F switch underlies the restriction point. *Nat Cell Biol* **10,** 476–82 (2008).
23. Anderson, J. C., Clarke, E. J., Arkin, A. P. & Voigt, C. A. Environmentally Controlled Invasion of Cancer Cells by Engineered Bacteria. *J. Mol. Biol.* **355,** 619–27 (2006).
24. Anderson, J. C., Voigt, C. A. & Arkin, A. P. Environmental signal integration by a modular AND gate. *Mol. Syst. Biol.* **3,** (2007).
25. Soule, C. Graphic Requirements for Multistationarity. *Complexus* **1,** 123–33 (2003).
26. Alon, U. *An Introduction to Systems Biology: Design Principles of Biological Circuits*. (Chapman & Hall/CRC, 2006).
27. Ferrell, J. E. Tripping the switch fantastic: how a protein kinase cascade can convert graded inputs into switch-like outputs. *Trends Biochem. Sci.* **21,** 460–6 (1996).
28. Angeli, D., Ferrell, J. E. & Sontag, E. D. Detection of multistability, bifurcations, and hysteresis in a large class of biological positive-feedback systems. *Proc. Natl. Acad. Sci. U. S. A.* **101,** 1822 –1827 (2004).
29. Ray, J. C. J., Tabor, J. J. & Igoshin, O. A. Non-transcriptional regulatory processes shape transcriptional network dynamics. *Nat. Rev. Microbiol.* **9,** 817–28 (2011).
30. Becskei, A., Seraphin, B. & Serrano, L. Positive feedback in eukaryotic gene networks: cell differentiation by graded to binary response conversion. *EMBO J* **20,** 2528–35 (2001).
31. Atkinson, M. R., Savageau, M. A., Myers, J. T. & Ninfa, A. J. Development of Genetic Circuitry Exhibiting Toggle Switch or Oscillatory Behavior in Escherichia coli. *Cell* **113,** 597–607 (2003).
32. Kramer, B. P. & Fussenegger, M. Hysteresis in a synthetic mammalian gene network. *Proc. Natl. Acad. Sci. U. S. A.* **102,** 9517 –9522 (2005).
33. Ajo-Franklin, C. M. *et al.* Rational design of memory in eukaryotic cells. *Genes Dev.* **21,** (2007).
34. Kim, S. Y. & Ferrell Jr., J. E. Substrate Competition as a Source of Ultrasensitivity in the Inactivation of Wee1. *Cell* **128,** 1133–45 (2007).
35. Buchler, N. E. & Louis, M. Molecular Titration and Ultrasensitivity in Regulatory Networks. *J. Mol. Biol.* **384,** 1106–19 (2008).
36. Bashor, C. J., Helman, N. C., Yan, S. & Lim, W. A. Using Engineered Scaffold Interactions to Reshape MAP Kinase Pathway Signaling Dynamics. *Science* **319,** 1539 –1543 (2008).
37. Buchler, N. E. & Cross, F. R. Protein sequestration generates a flexible ultrasensitive response in a genetic network. *Mol Syst Biol* **5,** 272 (2009).
38. Lee, T.-H. & Maheshri, N. A regulatory role for repeated decoy transcription factor binding sites in target gene expression. *Mol. Syst. Biol.* **8,** (2012).
39. Tan, C., Marguet, P. & You, L. Emergent bistability by a growth-modulating positive feedback circuit. *Nat. Chem. Biol.* **5,** 842–8 (2009).
40. Nevozhay, D., Adams, R. M., Van Itallie, E., Bennett, M. R. & Balázsi, G. Mapping the Environmental Fitness Landscape of a Synthetic Gene Circuit. *PLoS Comput Biol* **8,** e1002480 (2012).
41. Rhodius, V. A. *et al.* Design of orthogonal genetic switches based on a crosstalk map of σs, anti-σs, and promoters. *Mol. Syst. Biol.* **9,** (2013).

42. Kim, J., White, K. S. & Winfree, E. Construction of an in vitro bistable circuit from synthetic transcriptional switches. *Mol. Syst. Biol.* **2,** 68 (2006).
43. Francois, P. & Hakim, V. Design of genetic networks with specified functions by evolution in silico. *Proc. Natl. Acad. Sci.* **101,** 580–5 (2004).
44. Tiwari, A., Balázsi, G., Gennaro, M. L. & Igoshin, O. A. The interplay of multiple feedback loops with post-translational kinetics results in bistability of mycobacterial stress response. *Phys. Biol.* **7,** 036005 (2010).
45. Palani, S. & Sarkar, C. A. Synthetic conversion of a graded receptor signal into a tunable, reversible switch. *Mol Syst Biol* **7,** 480 (2011).
46. Lane, W. J. & Darst, S. A. The Structural Basis for Promoter −35 Element Recognition by the Group IV σ Factors. *PLoS Biol* **4,** e269 (2006).
47. Staroń, A. *et al.* The third pillar of bacterial signal transduction: classification of the extracytoplasmic function (ECF) σ factor protein family. *Mol. Microbiol.* **74,** 557–81 (2009).
48. Klumpp, S., Zhang, Z. & Hwa, T. Growth Rate-Dependent Global Effects on Gene Expression in Bacteria. *Cell* **139,** 1366–75 (2009).
49. Lou, C. *et al.* Synthesizing a novel genetic sequential logic circuit: a push-on push-off switch. *Mol. Syst. Biol.* **6,** (2010).
50. Ham, T. S., Lee, S. K., Keasling, J. D. & Arkin, A. P. Design and Construction of a Double Inversion Recombination Switch for Heritable Sequential Genetic Memory. *PLoS ONE* **3,** e2815 (2008).
51. Friedland, A. E. *et al.* Synthetic Gene Networks That Count. *Science* **324,** 1199 – 1202 (2009).
52. Siuti, P., Yazbek, J. & Lu, T. K. Synthetic circuits integrating logic and memory in living cells. *Nat. Biotechnol.* (2013). doi:10.1038/nbt.2510
53. Bonnet, J., Subsoontorn, P. & Endy, D. Rewritable digital data storage in live cells via engineered control of recombination directionality. *Proc. Natl. Acad. Sci.* (2012). doi:10.1073/pnas.1202344109
54. Schöbel, S., Zellmeier, S., Schumann, W. & Wiegert, T. The Bacillus subtilis σW anti-sigma factor RsiW is degraded by intramembrane proteolysis through YluC. *Mol. Microbiol.* **52,** 1091–105 (2004).
55. Zellmeier, S., Schumann, W. & Wiegert, T. Involvement of Clp protease activity in modulating the Bacillus subtilisσW stress response. *Mol. Microbiol.* **61,** 1569–82 (2006).
56. Qiu, J. & Helmann, J. D. The -10 Region Is a Key Promoter Specificity Determinant for the Bacillus subtilis Extracytoplasmic-Function σ Factors σX and σW. *J Bacteriol* **183,** 1921–7 (2001).
57. Li, M. Z. & Elledge, S. J. Harnessing homologous recombination in vitro to generate recombinant DNA via SLIC. *Nat Meth* **4,** 251–6 (2007).
58. Cox, R. S., Dunlop, M. J. & Elowitz, M. B. A synthetic three-color scaffold for monitoring genetic regulation and noise. *J. Biol. Eng.* **4,** 10 (2010).
59. Schilke, B. A. & Donohue, T. J. ChrR positively regulates transcription of the Rhodobacter sphaeroides cytochrome c2 gene. *J. Bacteriol.* **177,** 1929–37 (1995).

60. Newman, J. D., Anthony, J. R. & Donohue, T. J. The importance of zinc-binding to the function of Rhodobacter sphaeroides ChrR as an anti-sigma factor. *J. Mol. Biol.* **313,** 485–99 (2001).
61. Khlebnikov, A., Datsenko, K. A., Skaug, T., Wanner, B. L. & Keasling, J. D. Homogeneous expression of the PBAD promoter in Escherichia coli by constitutive expression of the low-affinity high-capacity AraE transporter. *Microbiology* **147,** 3241 –3247 (2001).
62. Wang, Q., Niemi, J., Tan, C., You, L. & West, M. Image segmentation and dynamic lineage analysis in single-cell fluorescence microscopy. *Cytometry A* **77A,** 101–10 (2010).
63. Guzman, L. M., Belin, D., Carson, M. J. & Beckwith, J. Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. *J. Bacteriol.* **177,** 4121 –4130 (1995).
64. Brenner, K., You, L. & Arnold, F. H. Engineering microbial consortia: a new frontier in synthetic biology. *Trends Biotechnol.* **26,** 483–9 (2008).
65. Minty, J. J. *et al.* Design and characterization of synthetic fungal-bacterial consortia for direct production of isobutanol from cellulosic biomass. *Proc. Natl. Acad. Sci.* 201218447 (2013). doi:10.1073/pnas.1218447110
66. Regot, S. *et al.* Distributed biological computation with multicellular engineered networks. *Nature* **469,** 207–11 (2011).
67. Tamsir, A., Tabor, J. J. & Voigt, C. A. Robust multicellular computing using genetically encoded NOR gates and chemical `wires`. *Nature* **469,** 212–5 (2011).
68. Basu, S., Gerchman, Y., Collins, C. H., Arnold, F. H. & Weiss, R. A synthetic multicellular system for programmed pattern formation. *Nature* **434,** 1130–4 (2005).
69. Fagotto, F. & Gumbiner, B. M. Cell Contact-Dependent Signaling. *Dev. Biol.* **180,** 445–54 (1996).
70. Bassler, B. L. & Losick, R. Bacterially Speaking. *Cell* **125,** 237–46 (2006).
71. Rufino Ferreira, A. & Arcak, M. A Graph Partitioning Approach to Predicting Patterns in Lateral Inhibition Systems. *SIAM J. Appl. Dyn. Syst.* **12,** 2012–31 (2013).
72. Turing, A. M. The Chemical Basis of Morphogenesis. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **237,** 37–72 (1952).
73. Kondo, S. & Miura, T. Reaction-Diffusion Model as a Framework for Understanding Biological Pattern Formation. *Science* **329,** 1616–20 (2010).
74. Aoki, S. K. *et al.* Contact-Dependent Inhibition of Growth in Escherichia coli. *Science* **309,** 1245 –1248 (2005).
75. Aoki, S. K., Webb, J. S., Braaten, B. A. & Low, D. A. Contact-Dependent Growth Inhibition Causes Reversible Metabolic Downregulation in Escherichia coli. *J. Bacteriol.* **191,** 1777–86 (2009).
76. Aoki, S. K. *et al.* A widespread family of polymorphic contact-dependent toxin delivery systems in bacteria. *Nature* **468,** 439–42 (2010).
77. Ruhe, Z. C., Wallace, A. B., Low, D. A. & Hayes, C. S. Receptor Polymorphism Restricts Contact-Dependent Growth Inhibition to Members of the Same Species. *mBio* **4,** (2013).

78. Diner, E. J., Beck, C. M., Webb, J. S., Low, D. A. & Hayes, C. S. Identification of a target cell permissive factor required for contact-dependent growth inhibition (CDI). *Genes Dev.* **26,** 515–25 (2012).
79. Webb, J. S. *et al.* Delivery of CdiA Nuclease Toxins into Target Cells during Contact-Dependent Growth Inhibition. *PLoS ONE* **8,** e57609 (2013).
80. Karimova, G., Pidoux, J., Ullmann, A. & Ladant, D. A bacterial two-hybrid system based on a reconstituted signal transduction pathway. *Proc. Natl. Acad. Sci.* **95,** 5752–6 (1998).
81. Joung, J. K., Ramm, E. I. & Pabo, C. O. A bacterial two-hybrid selection system for studying protein–DNA and protein–protein interactions. *Proc. Natl. Acad. Sci.* **97,** 7382–7 (2000).
82. Hu, J. C., Kornacker, M. G. & Hochschild, A. Escherichia coli One- and Two-Hybrid Systems for the Analysis and Identification of Protein–Protein Interactions. *Methods* **20,** 80–94 (2000).
83. Studier, F. W. & Moffatt, B. A. Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *J. Mol. Biol.* **189,** 113–30 (1986).
84. Kirkeby, S. & Thomsen, C. E. Quantitative immunohistochemistry of fluorescence labelled probes using low-cost software. *J. Immunol. Methods* **301,** 102–13 (2005).
85. Shis, D. L. & Bennett, M. R. Library of synthetic transcriptional AND gates built with split T7 RNA polymerase mutants. *Proc. Natl. Acad. Sci.* (2013). doi:10.1073/pnas.1220157110
86. Mutalik, V. K. *et al.* Precise and reliable gene expression via standard transcription and translation initiation elements. *Nat. Methods* **10,** 354–60 (2013).
87. Martin, C. T. & Coleman, J. E. Kinetic analysis of T7 RNA polymerase-promoter interactions with small synthetic promoters. *Biochemistry (Mosc.)* **26,** 2690–6 (1987).
88. García, L. R. & Molineux, I. J. Rate of translocation of bacteriophage T7 DNA across the membranes of Escherichia coli. *J. Bacteriol.* **177,** 4066–76 (1995).
89. McAdams, H. H. & Arkin, A. Stochastic mechanisms in gene expression. *Proc. Natl. Acad. Sci.* **94,** 814–9 (1997).
90. Arkin, A., Ross, J. & McAdams, H. H. Stochastic Kinetic Analysis of Developmental Pathway Bifurcation in Phage λ-Infected Escherichia coli Cells. *Genetics* **149,** 1633–48 (1998).
91. Rao, C. V., Wolf, D. M. & Arkin, A. P. Control, exploitation and tolerance of intracellular noise. *Nature* **420,** 231–7 (2002).
92. Gillespie, D. T. Exact stochastic simulation of coupled chemical reactions. *J. Phys. Chem.* **81,** 2340–61 (1977).
93. Dalbow, D. G. & Young, R. Synthesis time of beta-galactosidase in Escherichia coli B/r as a function of growth rate. *Biochem. J.* **150,** 13–20 (1975).
94. Anderson, J. C. *et al.* BglBricks: A flexible standard for biological part assembly. *J. Biol. Eng.* **4,** 1 (2010).
95. Salis, H. M., Mirsky, E. A. & Voigt, C. A. Automated design of synthetic ribosome binding sites to control protein expression. *Nat. Biotechnol.* **27,** 946–50 (2009).
96. Cabantous, S., Terwilliger, T. C. & Waldo, G. S. Protein tagging and detection with engineered self-assembling fragments of green fluorescent protein. *Nat. Biotechnol.* **23,** 102–7 (2005).

97. Naylor, L. H. Reporter gene technology: the future looks bright. *Biochem. Pharmacol.* **58,** 749–57 (1999).
98. Shekhawat, S. S., Porter, J. R., Sriprasad, A. & Ghosh, I. An Autoinhibited Coiled-Coil Design Strategy for Split-Protein Protease Sensors. *J. Am. Chem. Soc.* **131,** 15284–90 (2009).
99. Ortiz, M. E. & Endy, D. Engineered cell-cell communication via DNA messaging. *J. Biol. Eng.* **6,** 16 (2012).
100. Temme, K., Hill, R., Segall-Shapiro, T. H., Moser, F. & Voigt, C. A. Modular control of multiple pathways using engineered orthogonal T7 polymerases. *Nucleic Acids Res.* gks597 (2012). doi:10.1093/nar/gks597
101. Hillson, N. J., Rosengarten, R. D. & Keasling, J. D. j5 DNA Assembly Design Automation Software. *ACS Synth. Biol.* **1,** 14–21 (2012).
102. Aoki, S. K. *et al.* Contact-dependent growth inhibition requires the essential outer membrane protein BamA (YaeT) as the receptor and the inner membrane transport protein AcrB. *Mol. Microbiol.* **70,** 323–40 (2008).
103. Baba, T. *et al.* Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol. Syst. Biol.* **2,** (2006).
104. Datsenko, K. A. & Wanner, B. L. One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. *Proc. Natl. Acad. Sci.* **97,** 6640–5 (2000).
105. Arkin, A. P. & Fletcher, D. A. Fast, cheap and somewhat in control. *Genome Biol.* **7,** 114 (2006).
106. Taniguchi, Y. *et al.* Quantifying E. coli Proteome and Transcriptome with Single-Molecule Sensitivity in Single Cells. *Science* **329,** 533–8 (2010).
107. Wang, Z., Gerstein, M. & Snyder, M. RNA-Seq: a revolutionary tool for transcriptomics. *Nat. Rev. Genet.* **10,** 57–63 (2009).
108. Zaslaver, A. *et al.* A comprehensive library of fluorescent transcriptional reporters for Escherichia coli. *Nat. Methods* **3,** 623–8 (2006).
109. Yudkin, M. D. The sigma-like product of sporulation gene spoIIAC of Bacillus subtilis is toxic to Escherichia coli. *Mol. Gen. Genet. MGG* **202,** 55–7 (1986).
110. Parnas, D. L. On the criteria to be used in decomposing systems into modules. *Commun ACM* **15,** 1053–8 (1972).
111. Suter, B., Kittanakom, S. & Stagljar, I. Two-hybrid technologies in proteomics research. *Curr. Opin. Biotechnol.* **19,** 316–23 (2008).