Winter 12-15-2016

Development and Characterization of Genetic Sensors and Regulators for the Construction of Environmentally-Responsive Genetic Circuits

Allison Hoynes-O'Connor
Washington University in St. Louis

WASHINGTON UNIVERSITY IN ST. LOUIS

School of Engineering & Applied Science

Department of Energy, Environmental, and Chemical Engineering

Dissertation Examination Committee:
Tae Seok Moon, Chair
Ursula Goodenough
Himadri Pakrasi
Hani Zaher
Fuzhong Zhang

Development and Characterization of Genetic Sensors and Regulators for the Construction of Environmentally-Responsive Genetic Circuits

by

Allison Hoynes-O'Connor

A dissertation presented to
The Graduate School
of Washington University in
partial fulfillment of the
requirements for the degree
of Doctor of Philosophy

December 2016
St. Louis, Missouri

Table of Contents

List of Figures

List of Tables

Acknowledgments

Over the last four and a half years, I have been overwhelmed by the support, encouragement, and mentorship from my friends, family, and colleagues. First, I would like to thank Dr. Tae Seok Moon, for acting as my advisor and pushing me to think creatively and pursue my scientific interests. I would also like to thank my committee members, Dr. Goodenough, Dr. Pakrasi, Dr. Zaher, and Dr. Zhang, for their mentorship over the years.

I must also acknowledge my funding sources, which include the National Science Foundation [MCB-1331194], and the Mr. and Mrs. Spencer T. Olin Fellowship for Women in Graduate Study. I would also like to thank the former Director of the Olin Fellowship, Nancy Pope, who believed in me from the beginning and supported me through the challenges of graduate school.

Thanks to my friends, who have made the last four and a half years fun, challenging, and exciting, and to my labmates, who have supported me through my failures and celebrated my successes. To my parents, for prioritizing education, encouraging me to pursue a PhD, and providing confidence, strength, and friendship. To my brother and sister; I am inspired by your ambitious pursuit of your own dreams, and humbled by your support of mine. Thanks to Abigail, for reminding me to take lots of walks. And finally, to Matt, for bringing me so much happiness and providing unwavering support.

Allison Hoynes-O'Connor

Washington University in St. Louis

December 2016

Dedicated to my parents.

ABSTRACT OF THE DISSERTATION

Development and Characterization of Genetic Sensors and Regulators for the Construction of Environmentally-Responsive Genetic Circuits

by

Allison Hoynes-O'Connor

Doctor of Philosophy in Energy, Environmental and Chemical Engineering

Washington University in St. Louis, 2016

Professor Tae Seok Moon, Chair

Genetic circuits enable engineers to program complex logical behaviors into living organisms. Organisms can be programmed to optimize the production of fuels and chemicals, diagnose and treat diseases, or remediate environmental pollutants. A well-characterized toolbox of genetic sensors and regulators is needed to construct these circuits. Genetic sensors that respond to environmentally-relevant signals allow circuits to evaluate the cell's conditions, and versatile and designable regulators translate information about the cell's environment into the desired response. In this work, we demonstrate the *de novo* design of RNA thermosensors in *Escherichia coli*, and integrate these sensors into complex genetic circuits. Next, we provide a large-scale analysis of antisense RNA regulators, generate design rules for these regulators, and validate these design rules through the construction of genetic circuits with predictable behaviors. Finally AND and NAND gates are developed that respond to temperature and pH, and utilize protein and RNA regulators. The sensors, regulators, and circuits developed and characterized here represent a substantial contribution to the synthetic biology toolbox. Furthermore, this work constitutes an important step forward in enabling genetic circuits to overcome challenges in chemical synthesis, medicine, and environmental remediation.

Chapter 1: Sensor and regulator development for genetic circuit design

1.1 Accomplishments and challenges in synthetic biology

Synthetic biology is built on the idea that biology can be engineered, and that cells can be redesigned and repurposed to serve our needs in medicine, manufacturing, and a range of other industries. The first description of the lac operon in 1961 introduced the idea of regulatory circuits[1], and advances in PCR in the 1970's and 1980's provided the first methods to design and create novel DNA sequences[2]. However, it was not until the new millennium that the first synthetic regulatory circuits were developed. Genetic circuits are devices that allow logical behavior to be programmed into living organisms, and are comprised of genetic parts such as sensors, regulators, and actuators. There were two notable genetic circuits published in 2000 that laid the groundwork for future circuit development. The first was a toggle switch that used two promoters, each controlling the expression of a repressor for the other promoter[3]. This circuit could switch between two stable states using chemical or temperature induction. The second circuit used a transcriptional repressor to build an oscillating network, and was called a repressilator[4]. These two publications were the first examples of synthetic gene circuits, and were foundational in the field of synthetic biology.

As synthetic biology progressed, genetic circuits became more complex. Circuits were built that demonstrated a wide variety of logical operations, including all sixteen two-input Boolean logic gates[5], an edge-detecting circuit[6], and a genetic counter[7]. Circuits were also developed that integrated novel types of regulators, including RNA regulators[8-11], and CRISPR-based regulators[12-15]. As these increasingly complex circuits were developed, a design paradigm

emerged describing genetic circuits as consisting of parts (e.g. promoters, transcription factors, riboswitches, etc.) that could be combined to generate devices (e.g. toggle switch, repressilator, AND gate, etc.). These devices are genetic modules that perform a specific function.

Many of these genetic devices controlled the expression of a fluorescent protein or some other reporter. Studies of this type are essential in understanding the dynamics of genetic circuits, and enabling the development of complex circuit behaviors. However, as the field of synthetic biology advanced, genetic circuits began to be designed for specific applications in metabolic engineering, medicine, or environmental protection. Synthetic biology has significant overlaps with metabolic engineering. In fact, it has been argued that all applications of synthetic biology that modify metabolic pathways are actually examples of metabolic engineering[16]. Semantics aside, the ability to engineer biology in the context of metabolic engineering has provided important advances. One of the earliest examples used the *glnAp2* promoter, a genetic part that is activated by high glucose flux, to control two genes associated with lycopene production[17]. This simple circuit increased lycopene productivity three-fold. The concept demonstrated in this lycopene production system laid the groundwork for a wide range of dynamic sensor-regulator systems, which are circuits that sense the concentration of important metabolites and regulate the expression of downstream enzymes accordingly. Dynamic sensor-regulator systems were used in to increase the production of biodiesel[18], free fatty acids[19, 20], and amorphadiene[21].

In addition to metabolic engineering applications, synthetic biology has been used in the medical field. In some instances, light has been used induce gene expression in gene therapy applications[22, 23]. Genetic parts have also been integrated into molecular diagnostics, wherein a sensor responds to the presence of a disease biomarker and guides diagnostic decisions[24]. Similar therapeutic circuits have been developed that respond to the presence of a disease biomarker by

expressing the appropriate therapeutic, instead of simply expressing a reporter protein[25]. A significant amount of synthetic biology research has been focused in oncology. Genetic sensors have a unique ability to sense the subtle differences between cancer cells and healthy cells, thus genetic circuits have the potential to be used as effective diagnostic or treatment devices, wherein a genetic circuit is designed to kill cancerous cells, while leaving healthy cells unharmed. Signals such as microRNAs and hypoxia have been used to distinguish the cancerous cells from healthy cells[26, 27]. The integration of multiple genetic sensors within a circuit can improve accuracy and help prevent dangerous side effects[28, 29].

Synthetic biology has also been used to solve problems facing our environment. Genetic circuits have been developed for the detection of environmental pollutants such as arsenic[30], metal ions[29], and dinitrotoluenes[31]. Synthetic biologists have also endeavored to address problems caused by the use of nitrogen fertilizers in agriculture. Though certain organisms have the ability to fix nitrogen, crop plants are generally unable to do so, meaning that farmers often use large quantities of nitrogen fertilizer. Fertilizer runoff can cause a variety of environmental problems. Several engineering efforts have focused on understanding and engineering natural nitrogen fixation pathways. The ultimate goal is to imbue crop plants with the ability to fix their own nitrogen, and alleviate the need for nitrogen fertilizers. In one instance, synthetic biologists rebuilt the nitrogen fixation gene cluster from *Klebsiella oxytoca* using well-characterized synthetic parts, allowing the pathway to be more easily engineered[32]. Another study created a simple genetic circuit in which the nitrogenase genes from *Cyanothece* sp. ATCC 51142 were expressed only in the presence of low oxygen, given that nitrogenase is irreversibly inactivated by oxygen[33]. These fundamental studies may enable important agricultural advances that will positively impact our environment.

Though synthetic biology has advanced quickly, there are several challenges facing the field that are addressed in this work. First, there is an need for a greater diversity of genetic sensors with the ability to sense and respond to a wide range of signals in a reliable and robust manner. Second, there is a need for regulators with simple design rules that are easy to generate, are orthogonal to their target, and can be used to regulate a large number of genes simultaneously. Finally, these genetic parts must be integrated into complex genetic circuits in order to carry out programmed logical behaviors. With a broadened range of genetic sensors, versatile and designable regulators, and devices capable of complex logic, synthetic biology will continue to provide useful solutions to society.

1.1.1 Expanding the toolbox of genetic sensors

Currently, many of the promoters used in complex genetic circuits are inducible promoters that, while they are derived from natural systems, respond to a chemical inducer (e.g. pTet, pBad, pLux)[34]. While these promoters are extremely useful in the development of novel genetic circuits and in understanding circuit behavior, they also have several drawbacks. First, these chemical inducers must be added by the researcher at a set time, requiring a "hands-on" approach[35]. Next, these chemical inducers are generally not relevant to the eventual application of the circuit. So circuits designed to respond to a chemical inducer are responding to a biologically irrelevant signal. Furthermore, some of these chemical inducers cannot be used in medical applications, because they are pharmacologically active[22]. Finally, these chemical inducers can be expensive, and their cost becomes prohibitive in large-scale applications[36]. To overcome the drawbacks of these inducible promoters, genetic sensors should respond to conditions within a cell or in the cellular environment. This would allow a "hands-off" approach from the researcher, it would

enable researchers to choose meaningful signals for a specific application, and by responding to a condition that is already present, there would be no additional inducer cost.

There are many examples of genetic sensors in synthetic biology that respond to either intracellular (intrinsic) conditions or extracellular (extrinsic) conditions[37]. Intrinsic conditions such as metabolite concentration and glucose availability can be extremely important in optimizing carbon flux in metabolic engineering, as has been shown in the dynamic sensor-regulator systems discussed above[17-20]. Extrinsic conditions, such as pH and temperature, can be important to monitor in metabolic engineering because they are often associated with cellular stress, and can affect cellular productivity[38-42]. These extrinsic conditions are also relevant in medical applications as indicators of disease states or cancer microenvironments. However, it can be particularly difficult to integrate genetic sensors that respond to extrinsic conditions into genetic circuits because there are often substantial metabolic changes that occur in the host cell in these disparate environments[43]. This means that any genetic circuit that integrates such signals must be able to function robustly in different environments.

In terms of developing genetic sensors that respond to intrinsic and extrinsic signals, there are two general strategies. First, there is the discovery strategy, in which genetic sensors that exist naturally are mined from genomes, characterized, and implemented in synthetic systems. Most of the existing genetic sensors currently in use have been developed using this method. In one notable study, researchers were seeking a promoter that responded to farnesyl pyrophosphate (FPP), a toxic intermediate of the isoprenoid pathway[21]. Researchers exposed *E. coli* to different levels of FPP by creating a mutant that would accumulate FPP, and used whole-genome transcriptional analysis to identify promoters that were differentially expressed. These promoters

were then implemented in the construction of a dynamic sensor-regulator system which resulted in increased amorphadiene titer.

The second strategy for sensor development is *de novo* design. This strategy involves designing novel sensors that do not exist in nature. In an interesting demonstration of the *de novo* design strategy, a group of researchers devised a model-based approach for the design of aptazyme regulated expression devices (aREDs)[44]. Aptazymes, which are ribozymes that respond to the binding of a ligand to an aptamer, allow these regulators to sense internal metabolite concentrations. This strategy used computational models to predict the kinetics and RNA folding behavior of RNA sequences in response to the presence of a ligand. This strategy was later used to increase the production of p-aminostyrene in *E. coli*[45].

Both the discovery strategy and the *de novo* design strategy are important components of the synthetic biology toolbox. The discovery strategy is effective in identifying sensors that exist in nature, and as sequencing technologies advance, this strategy will become even faster and less expensive. However, these natural sensors can have complex mechanisms that are difficult to understand, confounding downstream engineering efforts. Furthermore, the discovery technique is unlikely to be effective in sensing signals that are not normally present in the environment. On the other hand, the *de novo* design strategy can be used to develop sensors for signals that are not normally encountered in nature. However, the *de novo* design strategy requires a thorough understanding of the regulatory mechanism, which is why *de novo* design has been successful with RNA regulators. RNA has a simple structure which can be relatively accurately predicted with thermodynamic modeling. As computational power increases, the *de novo* design strategy will be even more effective.

In this work, both the discovery strategy and the *de novo* design strategy are utilized for the development and characterization of genetic sensors for extrinsic conditions. First, the *de novo* design strategy is used to develop novel RNA thermosensors, which are RNA devices that regulate gene expression based on temperature. Later, a pH-responsive promoter and a temperature-responsive promoter which have been previously identified are extracted from the genome of *E. coli*, characterized, and integrated into AND and NAND gates. This work expands the availability of extrinsic genetic sensors in the synthetic biology toolbox.

1.1.2 Developing design rules for asRNA regulators

Once genetic sensors have been developed, they must be linked to an output via genetic regulators to produce logical behaviors. Protein regulators such as transcription factors and chaperones are commonly used to construct genetic circuits. They are harvested from natural systems and placed in a new genetic context where they facilitate a programmed behavior. For example, in this work we utilize regulators from the type III secretion systems in *Pseudomonas aeruginosa*[46, 47] and *Salmonella typhimurium*[48, 49]. However, protein regulators are difficult to design, and are not easily repurposed to orthogonally regulate a novel target. This limits the availability of protein regulators, which in turn limits the complexity and diversity of genetic circuits that can be constructed. Protein regulators can also be expensive for the cell to produce, both in terms of energy and resource utilization. While protein regulators serve an important purpose in synthetic biology and are used extensively throughout this work, there remains a need for regulators with simple design rules that can be designed to orthogonally regulate any given target, are inexpensive for the cell to produce, and can be multiplexed in order to regulate several targets at once.

RNA regulators have several advantages over protein regulators in the synthetic biology context[8, 50]. In this work, we focus particularly on antisense RNA (asRNA), which is a type of RNA regulator that works by binding to a target mRNA with a complementary sequence. This binding event disrupts translation and prevents gene expression. The first advantage that asRNA regulators have over protein regulators is the ease with which asRNA can be designed to target a particular gene. Because asRNA binds to its target through complementary base pairing, an asRNA can be designed by simply taking the reverse complement of the target region. Secondly, because of this base pairing, it is relatively simple to develop orthogonal regulators and prevent off-target effects. In addition, thermodynamic modeling software can be used to predict the secondary structure of RNA regulators[51]. Next, because asRNAs are small and are not translated, they are less energy- and resource-expensive for the host cell to produce, and can take effect more quickly than protein regulators. Because of the low host burden and orthogonality, many asRNAs can be expressed simultaneously for multiplexed regulation. Multiplexing is important because for many applications, it is necessary to regulate more than one gene at the same time in order to optimize circuit behavior[35]. Finally, asRNA can be transcribed independently of its target, meaning that it can be used to regulate chromosomal gene expression without modification of the chromosome.

Though asRNA's ability to be reliably designed is a major advantage over traditional protein regulators, there are still no reliable design rules for achieving specific levels of repression. Achieving a particular level of repression allows for optimization of circuit behavior. For example, in metabolic engineering it is necessary to balance the expression levels of multiple enzymes within a metabolic pathway, requiring regulators that could achieve targeted repression levels. Though asRNAs can be designed to target a particular gene by taking the reverse

complement of the target sequence, these regulators are not designed to achieve a particular repression level. Instead, the repression level can be modified by changing the transcription rate of the asRNA through the use of an inducible promoter. This requires the use of two genetic parts instead of one, and encounters barriers when it comes to scale-up due to the use of the inducible promoter, as discussed in section 1.1.1. While a number of studies have attempted to develop design rules to achieve targeted repression levels[52-59], the results offer conflicting design guidelines, and are not particularly informative for metabolic engineers[60]. For example, in some cases researchers have found that longer asRNAs improve repression[54], but in others it is suggested that short asRNAs are preferable to prevent off-target effects[52]. There is also disagreement about the role that thermodynamics plays in repression efficiency[52, 57-59], the necessity of asRNA-ribosome interactions[53, 57, 61], and the presence of the YUNR motif[10, 55, 58], which is a structural motif thought to improve asRNA-target interactions. In this work, the largest-scale study to date is performed to elucidate design rules for asRNA, resolve the conflicts that exist in literature, and enable metabolic engineers and synthetic biologists to simply and easily design asRNAs to achieve a programmed level of repression.

1.1.3 Construction of environmentally-responsive genetic circuits

Genetic sensors and regulators can be linked together in a number of different arrangements to form genetic circuits capable of performing complex logic operations. There have been many demonstrations of complex genetic circuits, including circuits that are able to store information or retain memory[62, 63], layered logic gates[64], and a genetic counter[7]. A key component of complex circuits is their ability to parse information from multiple signals and translate this information into a programmed response. A circuit that relies on multiple inputs will be able to exhibit a certain behavior in a very specific environment.

The ability to respond to multiple signals simultaneously is particularly important in medical applications[28, 29]. For instance, if a genetic circuit is designed to sense a cancer microenvironment and kill the nearby cancer cell, it is vitally important that this genetic circuit is able to accurately sense the proper microenvironment. Failure to do so could result in the destruction of healthy cells, which would be dangerous for the patient. As a method of ensuring this accuracy, multiple signals can be integrated into the circuit. So in order to activate the cell-killing behavior, the circuit would have to experience several different conditions that are characteristic of a cancer cell, instead of just one, reducing the probability of a false positive.

The final chapter of research presented in this work integrates findings from the first two sections, and further enables the use of genetic circuits in real-world applications. In the third research chapter, genetic sensors for extrinsic signals (i.e. temperature and pH), which had been previously identified, are extracted from the *E. coli* genome, capitalizing on previously published work that had used a discovery strategy for genetic sensors. These sensors are characterized, integrated into a simple AND gate, and modified to function in the new genetic context. Finally, an asRNA regulator developed in the second research chapter is integrated into the circuit architecture to invert the logical behavior, forming a set of NAND gates. This is the first demonstration of complex logical behavior based on pH and temperature, two extrinsic signals with relevance in metabolic engineering, medicine, and environmental applications.

The work described in this thesis provides a thorough study into the development of genetic sensors, the characterization of genetic regulators, and the integration of diverse parts into complex genetic circuits. These genetic circuits have the potential to be used in a wide range of real-world applications, and future work will focus on optimizing such circuits for these applications.

Chapter 2: *De novo* design of heat-repressible RNA thermosensors in *E. coli*

Reprinted with permission from Hoynes-O'Connor, A., Hinman, K., Kirchner, L. & Moon, T.S. *De novo* design of heat-repressible RNA thermosensors in *E. coli Nucleic Acids Research* **43**, 6166-6179 (2015).

A key challenge facing synthetic biology is the need for a greater diversity of genetic sensors that can be integrated into complex genetic circuits. This chapter describes the *de novo* design of genetic sensors that respond to temperature. The *de novo* design approach generates sensors with a simple, well-understood mechanism that can be engineered to suit the needs of a genetic circuit. Thus investigations into the mechanism of regulation, as are shown here, provide essential information to future genetic engineering endeavors.

2.1 Abstract

RNA-based temperature sensing is common in bacteria that live in fluctuating environments. Most naturally-occurring RNA thermosensors are heat-inducible, have long sequences, and function by sequestering the ribosome binding site in a hairpin structure at lower temperatures. Here, we demonstrate the *de novo* design of short, heat-repressible RNA thermosensors. These thermosensors contain a cleavage site for RNase E, an enzyme native to *Escherichia coli* and many other organisms, in the 5' untranslated region of the target gene. At low temperatures, the cleavage site is sequestered in a stem-loop, and gene expression is unobstructed. At high temperatures, the stem-loop unfolds, allowing for mRNA degradation and turning off expression. We demonstrated that these thermosensors respond specifically to temperature and provided experimental support for the central role of RNase E in the mechanism. We also demonstrated the modularity of these RNA thermosensors by constructing a three-input composite circuit that

utilizes transcriptional, post-transcriptional, and post-translational regulation. A thorough analysis of the 24 thermosensors allowed for the development of design guidelines for systematic construction of similar thermosensors in future applications. These short, modular RNA thermosensors can be applied to the construction of complex genetic circuits, facilitating rational reprogramming of cellular processes for synthetic biology applications.

2.2 Introduction

The ability to sense and respond to temperature is essential for survival, and accordingly, a variety of mechanisms to achieve this task can be observed in nature. Protein-based regulation systems, such as sigma factors specific to heat-shock proteins[65, 66] and chaperone proteins that aid in a variety of heat- and cold-shock responses[67-71], allow organisms to respond to temperature changes at both transcriptional and post-translational levels. RNA-based, temperature-responsive regulation systems, which function at a translational level, are also common throughout nature[72, 73]. They exploit the natural tendency of single-stranded RNA molecules to change their secondary structure in response to temperature shifts, resulting in altered RNA stability or translation rates.

RNA thermosensors can be described as heat-inducible or heat-repressible, meaning that they turn on or off gene expression at high temperatures, respectively. Most naturally-occurring RNA thermosensors are heat-inducible, and they function by sequestering the ribosome binding site (RBS) in a hairpin structure at low temperatures and exposing the RBS upon hairpin destabilization at high temperatures[73]. One example of such an RNA thermosensor is a regulatory element known as ROSE (Repression Of heat-Shock gene Expression) [74]. The predicted structure of the 5' UTR sequesters not only the Shine-Dalgarno (SD) sequence but also the start codon at low temperatures. Research has shown that the thermosensor hairpin does not

unfold completely at high temperatures, but rather that structural perturbations at high temperatures are sufficient for translation initiation to occur[73, 75]. A computational model has recently been developed to predict temperature-dependent perturbations in RNA secondary structure and might provide insight into the mechanisms of such RNA thermosensors[76].

Heat-repressible RNA thermosensors function by a variety of mechanisms. For example, expression of the RpoS sigma factor in *E. coli* is regulated by a trans-acting asRNA called DsrA[77]. DsrA can take two structural conformations, one of which (the F form) will bind to the target mRNA and expose the RBS. The F form has increased stability at low temperatures, allowing for heat-repressible expression of the target gene. However, the mechanism for the stability difference between the two forms is unknown[77]. Another well-studied heat-repressible RNA thermosensor regulates translation of the *cspA* mRNA in *E. coli*[78]. At low temperatures, the mRNA takes a stable conformation that is more efficiently translated. The entire length of the mRNA (428 nt), not just the 5'UTR, participates in this structural rearrangement, making the coding region an integral part of the mechanism[78, 79].

Naturally-occurring RNA thermosensors, though abundant in nature, can be difficult to implement in engineered systems. For example, the RpoS mechanism is complex and poorly understood, and the *cspA* thermosensor is very large and requires the participation of the *cspA* coding sequence. These characteristics can prevent naturally-occurring thermosensors from being effectively implemented in synthetic biology applications (i.e., limited reusability or modularity). On the other hand, a *de novo* design strategy offers the potential to develop minimal size thermosensors with a simple, well-understood mechanism. RNA regulators lend themselves to *de novo* design because they form predictable secondary structures[51, 80] and have a well-understood structure-function relationship[44, 81], which are characteristics that contribute to their

scalability[82]. Furthermore, synthetic RNA thermosensors can be designed to respond to a pre-determined temperature. However, very few studies have attempted to design synthetic RNA thermosensors thus far. Two studies in 2008 used computational tools and experimental screening to design heat-inducible RNA thermosensors *de novo* that unfolded a stem-loop to expose the SD sequence at high temperatures[83, 84]. To our knowledge, the only example of a designed heat-repressible RNA thermosensor was published more recently, which did not use a *de novo* design strategy, but simply fused naturally-existing RNA sequences[85]. In this work, we demonstrate the first, heat-repressible RNA thermosensors designed *de novo* (Figure 1 and Table 1).

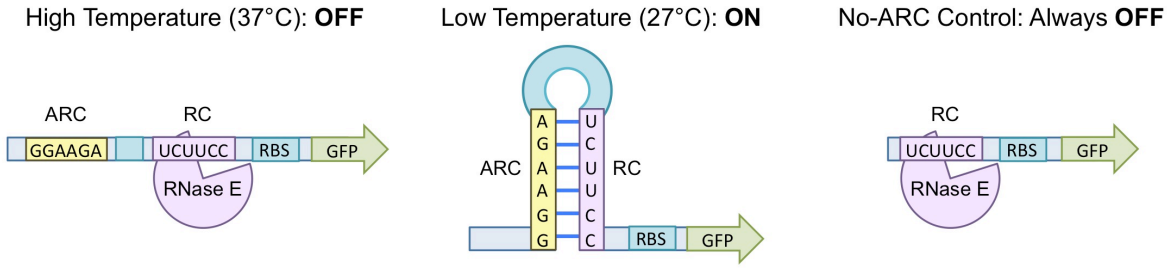

Figure 1: Mechanism of RNA thermosensors. At high temperatures (37°C), the RNase E cleavage site (RC - purple) is exposed, mRNA is cleaved by RNase E, and expression is "off." At low temperatures (27°C), the RC binds to the anti-RNase E cleavage site (ARC - yellow) and forms a hairpin. This structure sequesters the RC, and expression is turned "on." The No-ARC control lacks an ARC, and thus it is unable to form a hairpin structure. This control is expected to be in the "off" state at all temperatures.

The RNA thermosensors developed here have a small size and a simple mechanism, allowing for construction of temperature-responsive, complex genetic circuits and potential implementation in synthetic biology applications. At low temperatures, a hairpin sequesters a cleavage site for the native ribonuclease, RNase E in the 5' untranslated region (UTR) of the target mRNA (Figure 1).

At high temperatures, the hairpin is destabilized, allowing RNase E to degrade the transcript and turn off expression. RNase E was chosen for several reasons among a variety of ribonucleases, including RNase III, PNPase, and RNase P. First, RNase E is an endoribonuclease, with a preference for regions of single-stranded RNA[86]. This allows for targeted degradation of RNA in its unfolded form, which occurs at higher temperatures. Second, RNase E is native to *E. coli*, alleviating the need for the expression of a heterologous protein. Finally, both RNase E and its homologue RNase G are common in β - and γ-proteobacteria[87] as well as cyanobacteria[88], and about half of all sequenced eubacteria outside of these groups have at least one of these enzymes on its chromosome[87]. This provides reason to believe that implementation of these thermosensors in other organisms is possible with host-specific optimization. Our data show that gene expression can be regulated through the use of small, cis-acting, heat-repressible RNA thermosensors designed *de novo*. In addition, by analyzing thermosensor behaviors, we provide insights into their design principles.

2.3 Materials and Methods

2.3.1 *De novo* design and construction of RNA thermosensors

Thermosensors were initially under the control of a strong constitutive Anderson promoter (http://parts.igem.org/Part:BBa_J23104) and were constructed using blunt end ligation. Each thermosensor sequence was inserted downstream of the transcription start site and upstream of the RBS, as shown in Table 1. A template plasmid containing constitutive *gfp* was amplified with primers containing the thermosensor sequence. After digestion with DpnI (New England Biolabs), the amplified fragment was phosphorylated with T4 polynucleotide kinase (New England Biolabs), ligated using T4 DNA ligase (New England Biolabs), and electroporated into *E. coli* DH10B[89]. The subsequent replacement of the constitutive promoter with pTet was

accomplished using Golden Gate assembly[90]. All plasmids, strains, and key DNA sequences (i.e., genes, promoters, and UTRs) used in this work are summarized in Table 2, Table 3, and Table 4 respectively.

Thermosensor structures (Figure 17) and parameters (Table 1) were estimated using the Mfold Web Server[51]. The RNA folding form provided the predicted secondary structures and ΔG values as shown in Table 1. Melting temperatures (T_m = 25.6 to 37.8°C) were estimated using the "Two-state melting (hybridization)" application on the DINAMelt Web Server (http://mfold.rna.albany.edu/?q=DINAMelt/Two-state-melting)[59]. This application predicts the melting temperature of two separate strands of RNA. In this case, the RC and the ARC were considered to be the two separate strands, as shown in Table 1 and Figure 17. The effects of loop size were neglected in this approximation. Table 1 and Figure 17 show that the design parameters are varied within the 24 thermosensors. It is worth noting that the number of RCs and the total number of bulges are linked parameters. Bulges on either side of the stem were introduced in order to tune melting temperatures to the desired range (T_m = 25.6 to 37.8°C). With two RCs, it is necessary to include bulges in the stem because a perfectly complementary stem of that length has an estimated melting temperature of 76°C, which exceeds an appropriate growth temperature for *E. coli*. On the other hand, including bulges in thermosensor stems with only one RC reduces the estimated melting temperature to less than ~15°C, which is below our testing temperatures. For this reason, all of the thermosensors with two RCs contain stem bulges, and all of the thermosensors with one RC contain no stem bulges.

2.3.2 Experimental characterization of RNA thermosensors

Cells were grown overnight in LB media (Miller) and diluted 100X with M9 minimal media with 4 g/L glucose (for cultures of the cells that contain the three-input composite circuit, see below;

given catabolite repression and its effect on the pBAD promoter, glycerol, instead of glucose, was used as a carbon source for characterization of the three-input composite circuit). After 2 hours of growth at 37°C and 250 rpm, cells were centrifuged and resuspended in M9 minimal media with 4 g/L glucose. These cultures were grown with supplementation of anhydrotetracycline (aTc) at the appropriate testing temperature and 250 rpm until stationary phase. Stationary phase occurred 20, 22, and 25 hours after induction for cells growing at 37°C, 32°C, and 27°C, respectively. The same time points were used in all experiments unless otherwise indicated. To determine the optimum level of transcription, cells were induced with a gradient of aTc concentrations, ranging from 3.2 pg/mL to 250 ng/mL (Figure 18). An aTc concentration of 1 ng/mL was used unless otherwise indicated. Kanamycin (20 mg/mL), ampicillin (100 mg/mL), and chloramphenicol (34 mg/mL) were added as appropriate.

Measurements were taken with a Tecan Infinite M200 plate reader. Absorbance (Abs) was measured at 600 nm to monitor cell growth. GFP was measured at excitation = 483 nm and emission = 530 nm. In order to normalize the data, a series of controls were included in each experiment. *E. coli* DH10B was grown to provide a background fluorescence level. First, fluorescence was divided by absorbance (abs) to provide an approximate "per cell" fluorescence measure. Any GFP/Abs value within one standard deviation of the value of DH10B was indicated with an asterisk in all figures. The background GFP/Abs value, determined by measuring fluorescence and absorbance from DH10B, was subtracted from thermosensors' values. To account for differences in promoter activity due to temperature, as well as differences in protein folding and degradation rates at different temperatures, the GFP/Abs value for each culture was divided by that of the positive control. The positive control contained the same promoter as the thermosensor testing construct, but lacked the thermosensor entirely. For the

three-input composite circuit, the positive control was psicA-GFP (lacking a thermosensor), with the same input plasmids (pBAD-sicA* and pTet-invF) (Figure 6D). For all other experiments, the positive control was pTet-GFP with no thermosensor. The final normalized fluorescence was calculated as follows, where TS = thermosensor, + = positive control, and 0 = *E. coli* DH10B or BL21 Star (DE3): Normalized GFP = $[(GFP/Abs)_{TS} - (GFP/Abs)_0] / [(GFP/Abs)_+ - (GFP/Abs)_0]$.

For magnesium and pH testing, cells were grown overnight in LB media and diluted 100X with M9 minimal media with 4 g/L glucose, appropriate antibiotics, and 1 ng/mL aTc. Test conditions for the magnesium experiments were 2 mM Mg^{2+} or 2 μM Mg^{2+}. Magnesium was added in the form of $MgSO_4$, and missing SO_4^{2-} in the 2 μM Mg^{2+} condition was supplemented with Na_2SO_4 to 2 mM. Test conditions for pH experiments were pH=5 or pH=7. Media was acidified with HCl. Cultures were grown at 27°C until stationary phase, and measurements were taken as described above.

To construct the RNase E rescue strain, the coding sequence for RNase E (*rne*) along with its native promoters and 5'UTR was PCR-amplified from the *E. coli* MG1655 genome (2550362-2554197; http://blast.ncbi.nlm.nih.gov) and cloned on a plasmid (Table 2 and Table 4). An alternative version of the plasmid containing no *rne* was used as a control. *E. coli* BL21 Star (DE3) [F^- *ompT* *hsd*S_B ($r_B^- m_B^-$) *gal dcm rne1*31 (DE3)] was co-transformed with one thermosensor plasmid and either the plasmid containing *rne* or the alternative control plasmid with no *rne* (Table 3). Transcriptional scanning was repeated as described in Figure 18, and 2 ng/mL aTc was identified to give the optimum transcription level for thermosensor function in these strains. Fluorescence measurements were taken at stationary phase as described above.

2.3.3 Methods for using RT-qPCR to analyze the role of RNase E

RT-qPCR was performed with four thermosensors (D2, E3, F2, and F3), as well as the No-ARC control and the positive control (pTet-GFP). These four thermosensors were chosen because they had demonstrated significant increases in fold change upon introduction of the RNase E rescue plasmid based on fluorescence data (Figure 22). Each thermosensor and each control was tested in both the BL21 Star (DE3) strain and the RNase E rescue strain, and at 27°C and 37°C, for a total of 24 samples. To prepare samples for RT-qPCR, temperature induction was performed as described above. Samples were treated with rifampicin (300 mg/mL) as described previously [91] at stationary phase. RNA was immediately isolated from two biological replicates of each sample with a total culture volume of 1.5 mL per replicate (48 samples total). RNA isolation was performed using TRIzol Reagent (Life Technologies), followed by DNase treatment using the DNA-free Kit (Life Technologies). Finally, cDNA libraries were generated using the AffinityScript QPCR cDNA synthesis Kit (Agilent Technologies). Details of each of these steps are included in the Supplementary Methods.

Reference genes and their primers for RT-qPCR were chosen based on literature. The *cysG, hcaT*, and *idnT* genes were found to be stably expressed in the BL21 (DE3) strain, specifically at different temperatures, and their primer sequences were taken from literature [92]. Primers for *gfp* were generated using the NCBI Primer-BLAST tool to ensure specificity. All amplicons are 100-150 nt in length. All primers were ordered from Integrated DNA technologies and sequences are shown in Table 5. Details of primer optimization and efficiency can be found in the Supplementary Methods.

RT-qPCR was performed using the Power SYBR Green PCR Master Mix (Life Technologies) according to manufacturer's instructions, using a 50 mL reaction and 50 nM primers. The

CFX96 Touch™ Real-Time PCR Detection System (Bio-Rad Laboratories, Inc.) was used with the following cycling conditions: 95°C for 10 minutes, followed by 40 cycles of 95°C for 15 seconds, 60°C for 1 minutes, and then fluorescent detection. This was immediately followed by a melting curve (65-95°C, incrementing 0.5°C for 5 seconds, plate reading). The melting curve analysis confirmed the absence of non-specific products. For each sample, data are representative of two biological and two technical replicates (qPCR stage).

Quantification cycles (C_q) were determined using The CFX96 Touch™ Real-Time PCR Detection System software (Bio-Rad Laboratories, Inc.). C_q of the target gene (*gfp*) was normalized to the geometric mean of that of the reference genes (*cysG*, *hcaT*, and *idnT*) [92]. The relative expression level of each sample was normalized to that of the positive control (pTet-gfp) in that strain and at that temperature. Corrections were applied (log transformation, mean centering, and autoscaling) to account for variation associated with biological replicates, in accordance with MIQE guidelines [93, 94]. For each sample, biological and technical replicates were averaged and the standard error of the mean was calculated.

2.3.4 Construction and characterization of three-input composite circuits

To construct the three-input composite circuit, seven different thermosensors (B1, C1, D1, E1, E3, F1, F3) along with the common UTR (Table 1) were inserted upstream of *gfp* and downstream of the transcription start site of the psicA promoter (Table 4) using blunt end ligation as described above. The plasmids containing pBAD-sicA* and pTet-invF had been constructed previously [64]. Cells were grown overnight in LB media and diluted 100X with M9 minimal media with 0.4% glycerol, 2 g/L casamino acids, and 0.3 g/L thiamine hydrochloride. After 2 hours of growth at 37°C and 250 rpm, cells were centrifuged and resuspended in M9 minimal media with 0.4% glycerol, 2 g/L casamino acids, and 0.3 g/L thiamine hydrochloride.

These cultures were grown, with supplementation of 2 ng/mL aTc and 0.32 mM arabinose when necessary, at the appropriate testing temperature and 250 rpm until stationary phase. Measurements were taken as described above.

2.4 Results

2.4.1 Design of heat-repressible RNA thermosensors

The RNA thermosensors described here consist of a fluorescence reporter (i.e., GFP), a common RBS, and an RNase E cleavage site (RC) sequestered by an anti-RNase E cleavage site (ARC) in a stem-loop at low temperatures and exposed at high temperatures (Figure 1 and Table 1). At low temperatures, the mRNA will be protected from degradation by the stem-loop formation, and translation will occur unhindered, resulting in an "on" state. As the temperature increases, the stem-loop will unfold, exposing the RNase E cleavage site and allowing the transcript to be degraded. Thus, at high temperatures, expression will be off. The No-ARC control does not contain an ARC and is not expected to form a stem-loop at any temperature. Thus, at all temperatures, this control is expected to be in the "off" state.

RNase E was chosen for several reasons. First, it was necessary to choose an endoribonuclease so that an internal location on the transcript could be cleaved. It was also essential to choose an enzyme that cleaves single-stranded RNA, instead of double stranded RNA, so that the transcript is degraded at high temperatures when the stem-loop is unfolded. Additionally, an enzyme that has some sequence specificity allows for rational design of the thermosensor. Finally, an enzyme that is conserved throughout many bacterial species provides the potential for this thermosensor to be applied in other organisms in the future. RNase E met all these requirements, it is well studied, and an RNase E mutant strain is available [86, 87, 95, 96].

The mechanism shown in Figure 1 is a simplified version of the true behavior of RNA within the cell. While RNA hairpin folding is sometimes treated as a two-state process (folded and unfolded) [97, 98], a wealth of kinetic data indicates that it is actually a multi-state process [99-102]. Furthermore, the structures shown in Figure 17 are one of several potential folded states, and different secondary structures may dynamically coexist within the cell [51]. As a simple shorthand for this complex process, we will consider the folded and unfolded states shown in Figure 1 to be two "model-predicted" states, recognizing that there are various dynamic structures that may occur at any given temperature.

De novo design of RNA thermosensors began with the RNase E cleavage site (RC) (UCUUCC), identified in literature [96]. This sequence does not appear elsewhere within the *gfp* transcript (Table 4). Thermosensor sequences contain either one RC, or two RCs separated by a GC spacer (Table 1). The anti-RC (ARC) was constructed by taking the reverse complement of the RC, and then modifying it to achieve a predicted melting temperature within the 25 - 38°C temperature range. Thermosensors were named such that those that share a letter in their name (e.g., A1, A2, and A3) share a stem structure, and due to our estimation method, also share an estimated melting temperature. The loop region separating the ARC and RC consists of A's and U's and has lengths ranging from 5 to 16 nucleotides. Thermosensors that share a number in their name (e.g., A1, B1, and C1) share an approximate loop size, where X1 = 5-6 nt loop, X2 = 10-11 nt loop, and X3 = 15-16 nt loop. The 24 thermosensors also vary in their estimated ΔG and the number of bulges that they contain in each side of the stem. These parameters are summarized for each thermosensor in Table 1, and all predicted thermosensor structures are shown in Figure 17. To ensure that there was no potential downstream interaction that would prevent sequestration of the

RC, secondary structures were also predicted with the common UTR included (Table 1). It was found that there was no major deviation from the structure shown in Figure 17.

2.4.2 Optimization of thermosensor function

The thermosensors were initially tested with a strong constitutive promoter, Bba_J23104 (http://parts.igem.org/Part:BBa_J23104). Although they demonstrated increased fluorescence at low temperatures, the fold change was very small and in some cases, only observable when the temperature was reduced to 17°C. It was hypothesized that at such a high level of transcription, the large number of thermosensor-containing mRNAs overwhelmed the capacity of the native RNase E. Furthermore, it was reasoned that there would be an optimum transcription level that would allow for the maximum fold-change of the temperature response. In order to find the optimum transcription level, the constitutive promoters were replaced with the inducible pTet promoter. The No-ARC control and the positive control, which contain no thermosensor, were induced at 37°C at a variety of aTc concentrations. Because the No-ARC control does not form a hairpin at any temperature, it was expected to mimic the ideal "off" behavior. Conversely, the positive control does not contain an RC and is expected to mimic the ideal "on" behavior. By scanning expression levels (i.e., measuring reporter fluorescence) for maximum fold change between the two controls, we could identify the optimum transcription level for thermosensor function (Figure 18). There was a strong peak in the fold change (positive control/No-ARC control) between 0.08 and 2 ng/mL aTc. After narrowing down an appropriate range of expression levels, follow-up experiments were performed. Further experimentation identified 1.0 ng/mL as the optimum aTc concentration.

Further optimization was necessary in the measurement and normalization procedures. Because the rates of growth, transcription, translation, and RNA and protein degradation can vary with

temperature, measurements were taken at early stationary phase, when fluorescence and absorbance (at 600 nm) values were stable. Furthermore, all data was normalized to the positive control (pTet-GFP) to account for differences in transcription, translation, and degradation rates (see Materials and Methods).

2.4.3 Thermosensor response to temperatures

Once the transcription level and induction protocols had been optimized, thermosensor function was measured at 27°C and 37°C (Figure 2). Asterisks are shown for expression levels that are "completely off," meaning the fluorescence is within one standard deviation of the background (*E. coli* DH10B). Thermosensors functioned as expected, with a tightly-regulated "off" state at 37°C and a clear "on" state at 27°C. Some thermosensors (D2, E2, E3, F1, F2, H1, I1, and L1) showed leaky expression at 37°C. The No-ARC control confirms the importance of the stem-loop structure for temperature sensing. Because this control is unable to sequester the RC in a stem-loop at low temperatures, expression is off at both temperatures.

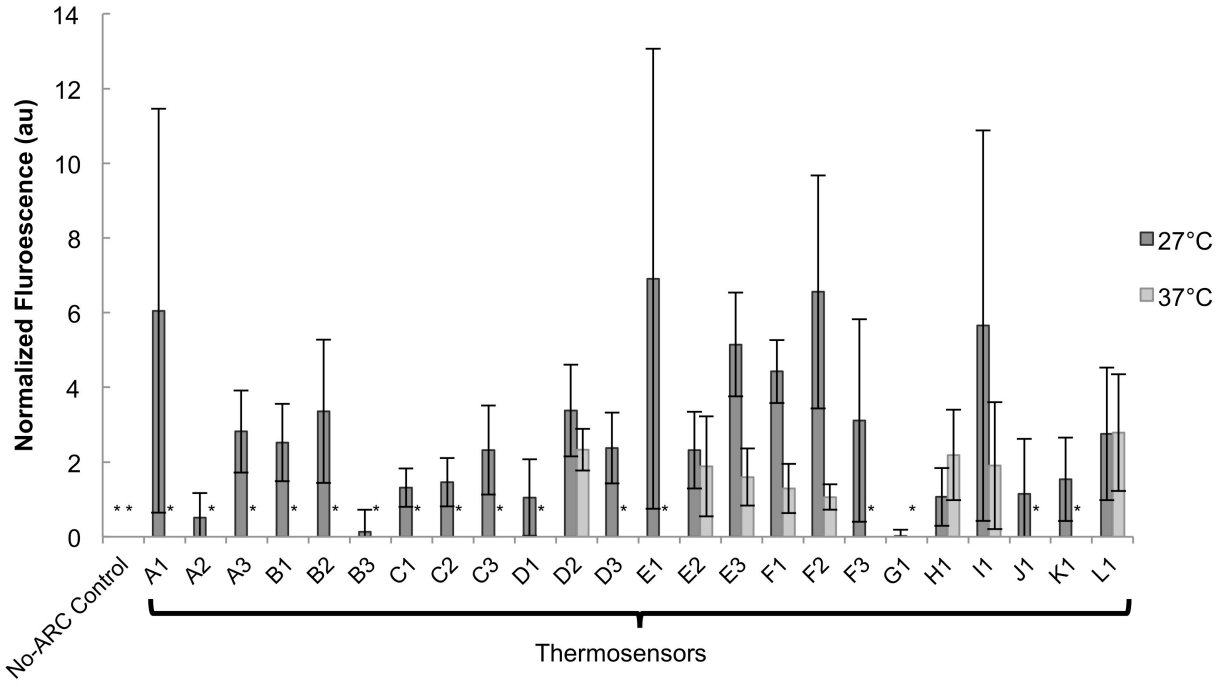

Figure 2: Thermosensor response in *E. coli* DH10B. Normalized fluorescence of thermosensors at 27°C and 37°C is shown (see Figure 19 for raw data, without normalization applied, including data for an intermediate temperature at 32°C). Fluorescence was normalized to pTet-GFP output at each temperature (1 ng/mL aTc). The asterisk (*) indicates that the GFP/Abs value was within one standard deviation of the DH10B GFP/Abs value (Materials and Methods). This means that expression was completely off and that these thermosensors are not leaky at 37°C. As expected, the No-ARC control is completely off at both temperatures. Data is the average of six biological replicates, over two different days. Error bars represent standard error of the mean (s.e.m.). A one-tailed, unpaired, Student's t-test was performed to see if expression was significantly higher at 27°C than at 37°C. The increase in fluorescence was significant for the A3, B1, C1, E3, and F1 thermosensors ($p<0.05$). If the criterion is relaxed, the increase in expression at 27°C from the B2, D3, and F2 thermosensors ($p<0.07$) as well as the K1 thermosensor ($p<0.09$) can also be considered significant. Thermosensors A1, A2, C2, C3, D1, D2, E1, F3, G1, and H1 had p-values less than 0.25, and thermosensors B3, E2, I1, J1, and L1 had p-values greater than 0.25.

The behaviors exhibited in Figure 2 were analyzed with respect to the design characteristics of each thermosensor, shown in Table 1. Statistical analysis showed that certain design parameters could be correlated with thermosensor behavior *in vivo* (Figure 3). A total of 24 thermosensors

were constructed and tested at 37°C and 27°C, and all 24 thermosensors were included in the statistical analysis. Most importantly, a tradeoff was observed between reduced leakiness and a high maximum expression level in the "on" state. In other words, a thermosensor that had a very tight "off" state would have a lower expression level in the "on" state, which suggests the importance of a delicate balance when selecting design parameters. To provide guidelines for thermosensor design, detailed statistical analysis was performed as discussed below.

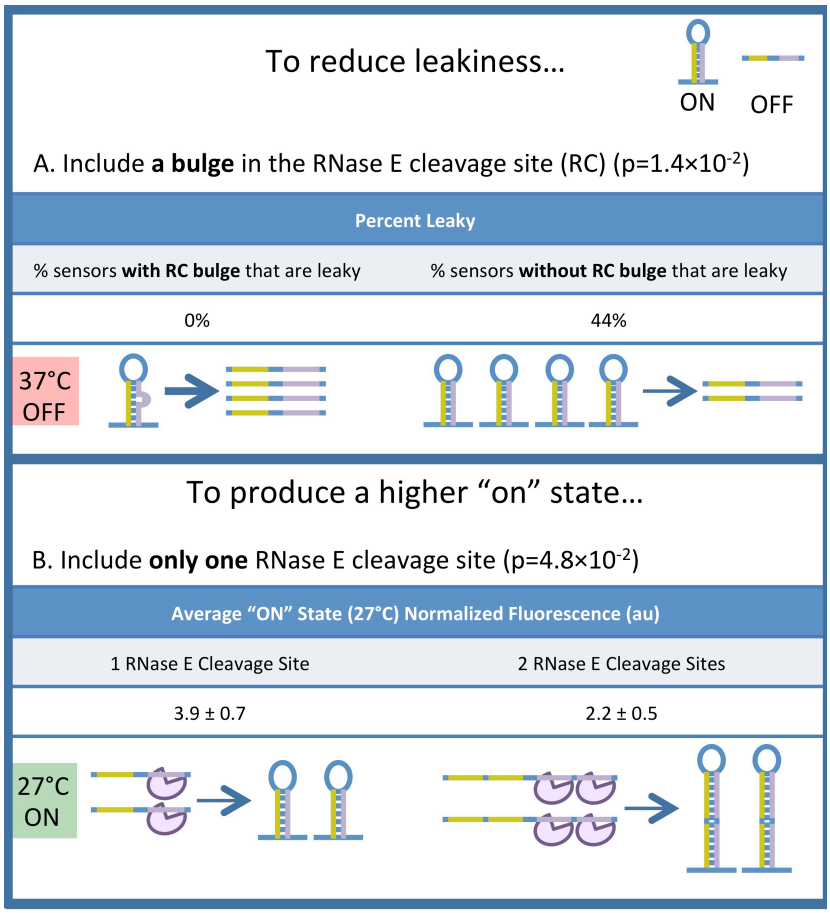

Figure 3: Analysis of thermosensor parameters. Analysis of thermosensor parameters. The parameters describing all 24 thermosensors were analyzed with a two-tailed, unpaired Student's t-test to determine which parameters had a significant impact ($p<0.05$) on thermosensor function. The parameters described here are correlated with reduced leakiness (A), or a higher "on" state expression level (B). (A) Reduced leakiness is correlated with the presence of a bulge in the RC. Stem bulges contribute to instability of the secondary structure, which will cause equilibrium to shift more transcripts to the model-predicted unfolded "off" state at high temperatures, reducing

leakiness. The weights of the arrows indicate the relative abundance of the two structures at equilibrium. The p-value ($1.4x10^{-2}$) indicates that there is a significant difference in the average number of RC bulges between "leaky" and "not leaky" thermosensors. (B) A higher "on" state is correlated with a single RC. The purple object represents RNase E. Increasing the number of RCs will increase the probability that an unfolded transcript will be degraded and turned off. The p-value ($4.8x10^{-2}$) indicates that there is a significant difference in average "on" state normalized fluorescence between thermosensors with 1 RC and thermosensors with 2 RCs.

Reduced leakiness was analyzed as a potentially desirable thermosensor characteristic. A thermosensor was considered "not leaky" if the fluorescence of the "off" state at 37°C was within one standard deviation of the white cells (DH10B). A thermosensor was considered "leaky" if the fluorescence of the "off" state exceeded one standard deviation of the white cells. Thermosensors that were not leaky were likely to contain a bulge in their RC ($p=1.4\times10^{-2}$; unpaired, 2-tailed Student's t-test). In fact, none of the thermosensors that contained a bulge in their RC were leaky, while 44% of the thermosensors lacking a bulge in their RC were leaky (Figure 3A and Figure 20). A possible explanation for this trend relates to stability. Thermosensors containing a bulge are less stable, meaning that the equilibrium between the two model-predicted structures would be driven towards the model-predicted unfolded "off" state, especially at high temperatures. It is interesting to note that bulges in the ARC had no effect on leakiness; only mismatches that would cause bulges on the RC side of the stem were correlated with reduced leakiness. This suggests that the bulge may improve RNase E access to the RC, providing an alternative explanation for this correlation.

Depending on the ultimate application, thermosensors with high expression in the "on" state may be more useful than thermosensors with reduced leakiness. The inclusion of a single RC instead of two was correlated with a higher "on" state ($p=4.8\times10^{-2}$) (Figure 3B and Figure 21). Increasing the number of RCs will increase the probability of RNase E-mediated cleavage,

regardless of the proportion of transcripts in the model-predicted unfolded state. Thus, as the number of RCs is reduced, the chance of cleavage decreases, more transcripts are left intact, and expression increases. Because all of the thermosensors with one RC also have perfectly complementary stems, it is difficult to say whether the higher "on" state is due to the lower number of RCs or the lack of bulges in the stems. Thermosensor variants with these parameters decoupled (e.g., with one RC and stem bulges, or two RCs and no bulges) were not tested because the estimated melting temperatures of such designs were outside the range of temperatures that can be tested *in vivo*.

Some parallels can be drawn between the synthetic heat-repressible RNA thermosensors developed here and the naturally-occurring heat-inducible ROSE thermosensors. It has been observed that ROSE thermosensors contain a conserved G-bulge across from the SD sequence when it is sequestered in a stem-loop at low temperatures. When this bulge was eliminated, two changes in thermosensor function were observed [103]. First, the thermosensor is not de-repressed at high temperatures. Second, there is increased repression at low temperatures. Both of these behaviors can be explained by an increased proportion of transcripts in the model-predicted folded state at each temperature. Similarly, we observed that thermosensors containing no RC bulge had leaky expression and a higher "on" state. Again, this behavior would suggest that an increased proportion of transcripts are in the model-predicted folded state, consistent with observations of the ROSE thermosensors. Thus, conserved features observed in nature can provide insights into the function of synthetic RNA devices.

2.4.4 Specificity of temperature response

While temperature is known to cause changes in RNA secondary structure, there are other environmental conditions that can also affect RNA stability. In order to confirm that the RNA

thermosensors respond specifically to temperature, and are not inadvertently activated by other variations in RNA stability, the effects of pH and magnesium starvation were assessed. It has been observed that some RNA thermosensors, such as the *Salmonella* fourU-type RNA thermometer, behave differently as magnesium concentration is altered [104]. However, other RNA thermosensors, such as ROSE thermosensors, are unaffected by magnesium concentration [105]. RNA is a negatively charged molecule, and intramolecular repulsive forces can prevent correct RNA folding. Divalent cations, especially Mg^{2+}, play an important role in alleviating these repulsive forces and allowing RNA structures to form [106, 107]. Thus, a magnesium-limited environment may prevent the hairpin structure from forming at 27°C, causing transcripts to remain in the model-predicted unfolded "off" state, leading to a lower "on" state. In order to test this hypothesis, responses of thermosensors were measured at 27°C at 2 mM Mg^{2+} or 2 mM Mg^{2+}.

The second stability variable tested was pH. RNA shows increased stability in slightly acidic environments [108]. Furthermore, although cells are known to maintain homeostasis, research has shown that *E. coli* can reduce its intracellular pH in an acidic environment [109-111]. Thus, it is plausible to suggest that these thermosensors, though they function intracellularly, might be influenced by media pH in addition to temperature. In order to test this hypothesis, cells were grown in neutral and acidic media (pH=7 and pH=5, respectively). Based on a range of $[H+]_{internal}/[H+]_{external}$ ratios (0.025 – 6.3) reported by literature and supported by thermodynamic modelling [112], the intracellular pH is expected to be between 6.2 and 8.6 for an external pH of 7 and between 4.2 and 6.6 for an external pH of 5. If the thermosensors were to respond to low pH, it would be expected that a higher proportion of thermosensors would be in the stable, model-predicted folded state at 27°C, resulting in a higher "on" state at pH=5.

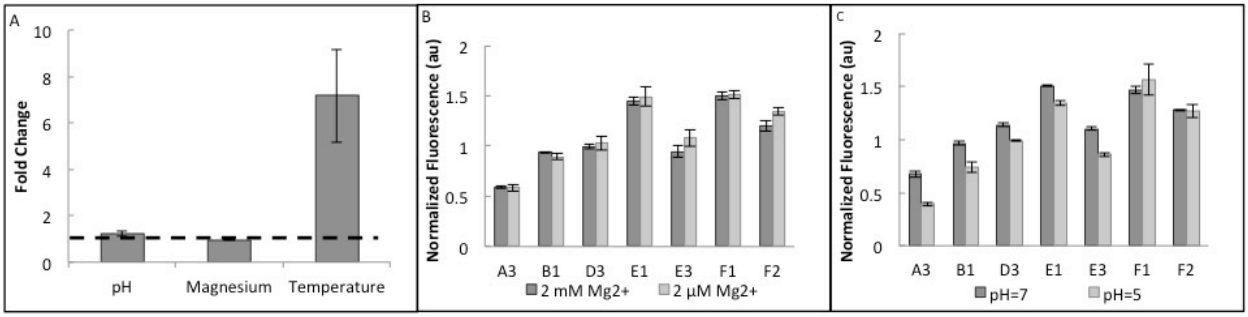

Figure 4: Specificity of RNA thermosensors. (A) Average fold changes are shown for the A3, B1, D3, E1, E3, F1, and F2 thermosensors. Fold changes are the ratios of normalized fluorescence at the two conditions (pH fold change = pH7/pH5; magnesium fold change = 2 mM Mg^{2+}/2 mM Mg^{2+}; and temperature fold change = 27°C/37°C). A fold change of one, shown by the dashed line, is the expected fold change if there is no response to the stimulus. A one-mean, 2-tailed Student's t-test was used to determine if the average fold change was significantly different from one. The average fold change was not significantly different from one for low pH and magnesium starvation ($p>0.05$), but it was significantly different from one for the temperature change ($p<0.05$). (B) Thermosensor response to magnesium starvation at 27°C for individual thermosensors. There is no apparent response to magnesium starvation. (C) Thermosensor response to pH change at 27°C for individual thermosensors. There is no apparent response to pH changes. Data for pH and magnesium experiments is the average of three biological replicates; temperature data is the average of six biological replicates, over two separate days. Error bars represent standard error of the mean (s.e.m.).

In Figure 4A, the average fold changes for the A3, B1, D3, E1, E3, F1, and F2 thermosensors are compared when exposed to three different stimuli. A fold change of one (shown by the dashed line) indicates that there is no response to the stimulus in question. The fold changes for both the pH experiment (pH=7/pH=5, 2 mM Mg^{2+}, 27°C) and the magnesium experiment (2 mM Mg^{2+}/2 mM Mg^{2+}, pH=7, 27°C) were not significantly different from one, ($p>0.05$; one-mean, 2-tailed Student's t-test). On the other hand, the fold change for the temperature response (27°C/37°C, 2 mM Mg^{2+}, pH7) was significantly different from one ($p<0.05$; one-mean, 2-tailed Student's t-test). This means that the thermosensors respond specifically to temperature and are unlikely to be activated by RNA stability variations brought on by magnesium starvation or low pH (Figure 4). However, because intracellular pH was not directly measured or manipulated in this

experiment, the possibility that the intracellular pH remained constant between the two cultures cannot be disregarded.

2.4.5 Confirmation of RNase E participation in the temperature-sensing mechanism

In order to confirm that RNase E does in fact play a central role in the temperature-sensing mechanism (Figure 1), the thermosensors were tested in the BL21 Star (DE3) strain of *E. coli*. This strain contains a truncated RNase E that is unable to cut RNA [113]. According to the hypothesized mechanism, the thermosensors would not be able to turn off without a fully-functional RNase E. Thus, in BL21 Star (DE3), the RNA thermosensors would be expected to lose their ability to respond to temperature and instead remain in the "on" state at all temperatures. Given that extensive optimization was required to see a clear temperature response in DH10B, it was important to verify that any failure to sense temperature in the BL21 Star (DE3) strain was due to the absence of a functional RNase E, not a lack of optimization. To ensure that this was the case, an RNase E rescue strain was constructed by expressing the wild-type RNase E gene (*rne*) on a plasmid in BL21 Star (DE3). The RNase E rescue strain would be expected to show a temperature response due to the presence of a functional version of RNase E, providing a control for thermosensor behavior in BL21 Star (DE3).

Thermosensor behavior was initially assessed with fluorescence data, as had been done for the DH10B strain. However, the results from the fluorescence data were unclear. A complete loss of temperature sensing in BL21 Star (DE3) would result in a fold change of one, which was not observed (Figure 5A). Though 27°C/37°C fold changes were consistently higher in the RNase E rescue strain than in the BL21 Star (DE3) strain, as would be expected, fold changes in BL21 Star (DE3) were greater than one for all tested thermosensors (Figure 5A). Furthermore, small

changes in fluorescence were observed in the No-ARC control, which was expected to display a constant fold change close to one in both strains.

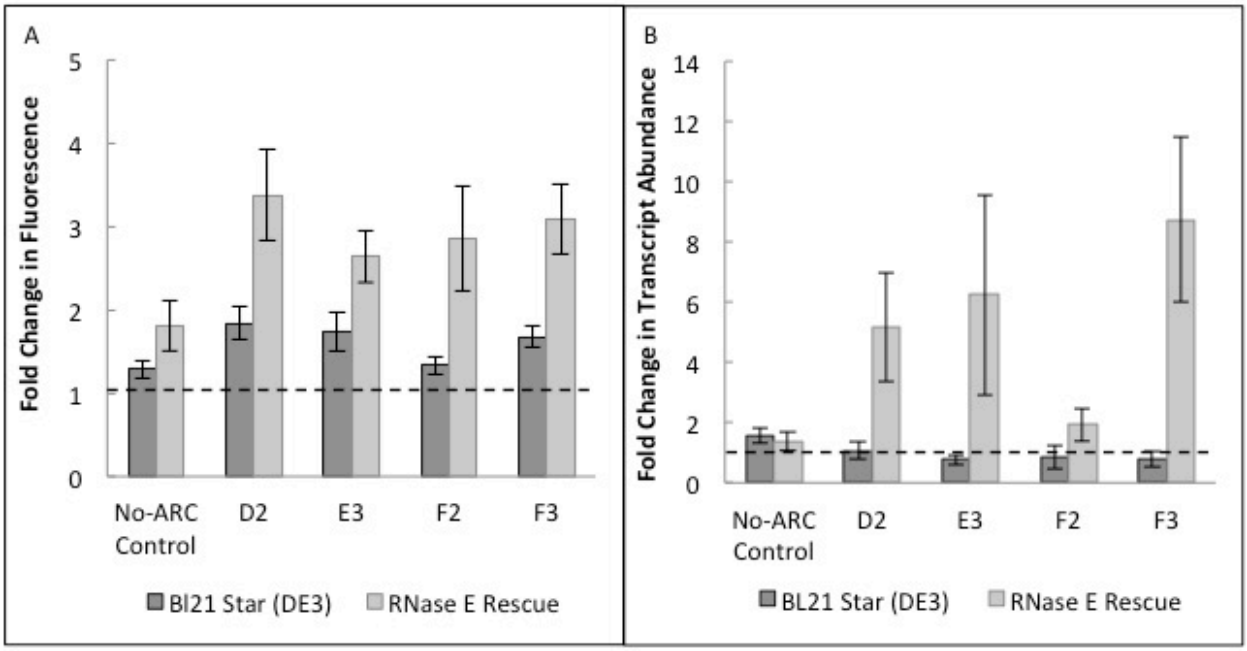

Figure 5: Thermosensor response in BL21 Star (DE3) and RNase E rescue strains. BL21 Star (DE3) contains a mutated version of RNase E that is unable to cut RNA. The RNase E rescue strain was constructed by expressing RNase E on a plasmid in BL21 Star (DE3). Fold changes are expected to be equal to one in the BL21 Star (DE3) strain (shown by dashed lines), indicating a loss of the ability to sense temperature, and greater than one in the RNase E rescue strain, demonstrating the recovery of the ability to sense temperature. (A) Fluorescence fold change (27°C fluorescence / 37°C fluorescence) of the D2, E3, F2, and F3 thermosensors, all of which demonstrate a significant increase in the 27°C/37°C fluorescence ratio from the BL21 Star (DE3) strain to the RNase E rescue strain ($P<0.05$; two-tailed, unpaired, Student's t-test). P-values are as follows: No-ARC control = 0.13, D2 = 0.02, E3 = 0.03, F2 = 0.04, and F3 = 0.01. Data is the average of 14 biological replicates, over a total of three different days. Data for the fluorescence in each strain at each temperature is provided in Figure 22. (B) Fold change in transcript abundance (27°C transcript abundance / 37°C transcript abundance) in the D2, E3, F2 and F3 thermosensors as well as the No-ARC control, based on RT-qPCR data. The thermosensors demonstrate a low 27°C/37°C fold change in BL21 Star (DE3), indicating that the temperature response has been removed. The response is recovered by introducing a functional version of RNase E, as evidenced by the increased fold change in the RNase E rescue strain. As expected, the No-ARC control has a low fold change in both strains. Data was normalized to the positive control (pTet-gfp) in that strain and at that temperature, and corrections were applied

(log transformation, mean centering, and autoscaling) in accordance with MIQE guidelines [93, 94]. The data shown is from two biological and two technical replicates. Error bars represent standard error of the mean (s.e.m.). P-values were calculated with a one-tailed, unpaired students t-test, and are as follows: No-ARC = 0.33, D2 = 0.05, E3 = 0.10, F2 = 0.08, and F3 = 0.03. Data for the relative transcript levels in each strain at each temperature is provided in Figure 23.

To gain a clearer understanding of the effect of RNase E on the stability of thermosensor-containing transcripts, RT-qPCR was performed with the two strains. Because the hypothesized mechanism functions on the transcript stability level, directly measuring transcript levels provides better experimental support for the mechanism than does fluorescence data, which measures protein levels. Furthermore, it has been reported that there is no clear linear relationship between mRNA levels and protein levels [114, 115], so by examining transcript abundance directly with RT-qPCR, we can neglect differences in translation rate or protein stability that may contribute to unexpected differences between strains in the fluorescence data. This will allow us to directly observe the impacts of RNase E on thermosensor-containing transcript abundance.

RT-qPCR analysis was performed with the D2, E3, F2, and F3 thermosensors because they each showed a significant increase in fold change between the BL21 Star (DE3) strain and the RNase E rescue strain based on fluorescence data (Figure 5A). Four other thermosensors that were tested did not show a significant change in fold change, possibly due to a lack of optimization in the new strain (Figure 22). Transcript abundance levels based on RT-qPCR analysis are shown in Figure 5B. The 27°C/37°C fold change is expected to be one in the BL21 Star (DE3) strain, because this strain is lacking a functional RNase E and is not expected to respond to temperature. This level is shown by the dotted line. As expected, in each of the four thermosensors as well as the No-ARC control, the reported fold change is very close to one in BL21 Star (DE3). Upon introduction of a functional RNase E in the RNase E Rescue strain, all four thermosensors

showed an increase in fold change, though the increase was small in the F2 thermosensor (fold changes in RNase E rescue strain: D2 = 5.2, E3 = 6.2, F2 = 1.9, and F3 = 8.7). This indicates that RNase E does in fact play an important role in the temperature sensing mechanism. Additionally, the No-ARC control showed a low fold change in the RNase E rescue strain. Although this fold change was not exactly equal to one (No-ARC fold change = 1.6 in BL21 Star (DE3) and 1.4 in RNase E rescue strain), the difference in fold change between BL21 Star DE3 and the RNase E rescue strain was not significant ($p=0.33$, Figure 5B), indicating the RNase E does not have a major impact on the 27°C/37°C fold change of the No-ARC transcript abundance.

While RT-qPCR provides the most relevant data for elucidating the mechanism of these thermosensors on the level of transcript stability, fluorescence data is more relevant for assessing whether or not these thermosensors have the potential for implementation in a real system. For most synthetic biology applications, protein expression level is the central outcome, whether that protein is an enzyme in a metabolic pathway, a pathogen-killing toxin, or a transcription factor in a complex genetic circuit. This means that while transcript abundance provides important insights into the mechanism, fluorescence data would be more important for determining the potential of these thermosensors to be applied in engineered systems. Because thermosensors had been optimized to function in DH10B, behavioral differences between thermosensors in DH10B and BL21 Star (DE3) strains are not unexpected. Further optimization would be required to show that these thermosensors can function as well on the protein (fluorescence) level in BL21 Star (DE3) (the RNase E rescue strain) as they do in DH10B.

2.4.6 Construction and characterization of multi-input composite circuits

A benefit of regulators that act on the transcript stability level is that they can be combined with regulation at multiple levels to build complex logic operations. In order to demonstrate that the

RNA thermosensors developed in this study can be used in complex circuits, the existing construct was first demonstrated to function as a two-input composite circuit. We define a composite circuit as a circuit that utilizes regulation mechanisms at more than one level (e.g., transcription and transcript stability). Because the thermosensors are under the control of pTet (Figure 6A), we can consider this construct a two-input composite circuit, with temperature and aTc as the two inputs. A response would only be expected when the two inputs are present (aTc = 1 ng/mL and temperature = 27°C) (Figure 6B). Using the B1 thermosensor, which was the best performing two-input composite circuit that was also tested in a three-input composite circuit, we can see that the two-input composite circuit functions as expected (Figure 6C). Results for two-input composite circuits using all 24 thermosensors are shown in Figure 24.

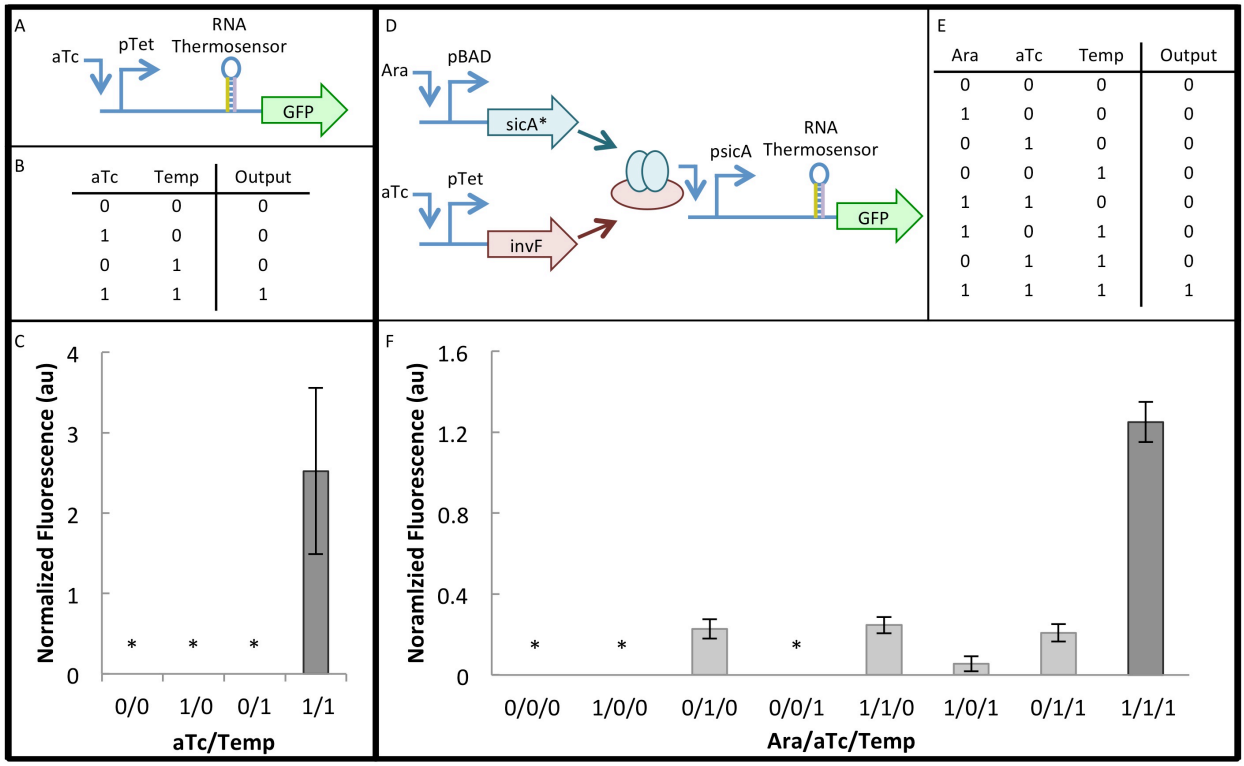

Figure 6: Implementation of RNA thermosensors in genetic circuits. (A) Diagram for two-input composite circuit shows that *gfp* is under the control of the pTet promoter and an RNA

thermosensor. (B) Truth table for two-input composite circuit. For temperature, "0" = 37°C and "1" = 27°C. aTc was used at a concentration of 1 ng/mL. (C) Results of two-input composite circuit with the B1 thermosensor. Data is the average of six biological replicates, over two separate days. The asterisk (*) indicates that the GFP/Abs value was within one standard deviation of the background DH10B GFP/Abs value (Materials and Methods). Error bars represent standard error of the mean (s.e.m.). (D) Circuit diagram for three-input composite circuit. pBAD and pTet control the expression of SicA* (chaperone) and InvF (transcription factor), respectively. These two proteins form a complex that is necessary to activate the psicA promoter, which controls the transcription of *gfp* with an RNA thermosensor. (E) Truth table for three-input composite circuit. For temperature, "0" = 37°C and "1" = 27°C. aTc was used at a concentration of 2 ng/mL and arabinose (Ara) was used at a concentration of 0.32 mM. (F) Results of three-input composite circuit with the E3 thermosensor. Data is the average of three biological replicates, over two separate days. The asterisk (*) indicates that the GFP/Abs value was within one standard deviation of the background DH10B GFP/Abs value (Materials and Methods). Error bars represent standard error of the mean (s.e.m.).

To build on this concept and further demonstrate the modularity and composability of the RNA thermosensors, a three-input composite circuit was constructed (Figure 6D). This circuit used components from the type III secretion system in *Salmonella typhimurium*[48]. In this system, the chaperone (SicA) and transcription factor (InvF) form a complex which is required to activate transcription from the sicA promoter. This system has been previously optimized to function in a two-input AND gate[64]. The previously published AND gate was modified to form a three-input composite circuit by adding a thermosensor downstream of the psicA promoter to control the expression of the reporter, GFP (Figure 6D). Seven variants of this circuit were generated by inserting seven different thermosensors. These thermosensors were selected so as to represent a wide variety of behaviors, specifically with respect to three criteria (see Figure 2): (i) a high “on” state fluorescence level (>3 normalized fluorescence units), (ii) a low p-value (<0.05), and (iii) leakiness. Two thermosensors (E3 and F1) met all three criteria, two thermosensors (E1 and F3)

met criterion (i) only, two thermosensors (B1 and C1) met criterion (ii) only, and one thermosensor (D1) met none of the three criteria.

The three-input composite circuit operates on three regulatory levels: the transcriptional (promoter-mediated), the post-translational (protein-protein interaction-mediated), and the transcript stability (RNA-mediated) level. Results for the E3 thermosensor circuit, which was the best performing three-input composite circuit, are shown in Figure 6F. This circuit performed as expected, with the highest level of expression occurring when all three inputs are present (Ara = 0.32 mM; aTc = 2 ng/mL; temperature = 27°C) (Figure 6E and Figure 6F). A 5.1-fold change was achieved between the leakiest "off" state [110] and the "on" state [111]. It is notable that this circuit demonstrates leakiness under certain conditions. This behavior is consistent with previous data, since the E3 thermosensor was slightly leaky in the two-input composite circuit (Figure 24). However, thermosensors that did not exhibit leakiness in the two-input composite circuit exhibited levels of leakiness in the three-input composite circuit similar to that of the E3 thermosensor circuit (Figure 24 for two-input and Figure 25 for three-input). Still, because the thermosensors are functioning in a different genetic context (i.e., under the control of the different promoter psicA, instead of pTet), slight differences in behavior would not be unexpected. The behavior of each thermosensor was compared in the three-input and two-input circuits. A slight correlation (r^2=0.72) was found between the fold change of the 3-input circuit and the "on" state of the two-input circuit (Figure 26).

2.5 Discussion

RNA thermosensors have a wide variety of potential applications in synthetic biology and metabolic engineering. As synthetic biology transitions from lab-scale genetic circuit demonstrations to industrial, medical, and environmental applications, chemical inducers such as

arabinose and aTc will have to be replaced. Not only are these inducers irrelevant in situ (e.g., inside the human body or at a potential bioremediation site), but their cost is also inhibitory when it comes to scale-up. On the other hand, temperature can be a meaningful signal in the environment, and temperature-responsive systems will not require expensive chemical inducers. Furthermore, metabolic engineers often need to consider temperature variations within large-scale bioreactors and adjust host cells' metabolism accordingly. For example, hotspots are a common problem in solid-state fermentation[116], and photobioreactors can overheat in the afternoon[38]. Heat-repressible RNA thermosensors with customized melting temperatures could be implemented to down-regulate product synthesis pathways that divert the cells' resources away from survival during such undesirable periods.

In this study, heat-repressible RNA thermosensors were designed *de novo* and demonstrated to function in *E. coli* (Figure 2). They have a stem-loop in the 5' UTR upstream of the RBS that unfolds at high temperatures to expose an RNase E cleavage site. The exposed RNase E cleavage site allows for degradation of the transcript, turning off expression (Figure 1). Several experiments were conducted to confirm the hypothesized mechanism outlined in Figure 1. The No-ARC control is not expected to form a stem-loop at any temperature, leaving it consistently prone to RNase E degradation. This control is off at both 27°C and 37°C, demonstrating the importance of the stem-loop structure for a functional thermosensor (Figure 2). The magnesium and pH experiments summarized in Figure 4 ensure that the thermosensor responds specifically to temperature and does not respond to conditions that affect RNA stability in general. Finally, the RT-qPCR experiments with BL21 Star (DE3) support the proposed mechanism outlined in Figure 1. These experiments show that thermosensors show no temperature response in a strain with a non-functional RNase E, but regain function upon introduction of a fully functional

RNase E (Figure 5). This result supports the hypothesis that RNase E plays an important role in the temperature-sensing mechanism. A thorough knowledge of the mechanism will allow for streamlined implementation of these thermosensors in other genetic systems and potentially in other organisms.

A potential complication to the mechanism presented in Figure 1 is the participation of the ribosome. Each of the 24 thermosensors was designed to share an SD sequence (Table 1) in order to maintain relatively consistent translation rates. However, because context effects are well known to affect translation rates[117], it is likely that the actual RBS strengths varied among thermosensors. One factor that may cause differences in translation rates among thermosensors is the secondary structure of the thermosensors themselves. A recent paper investigates the effect of secondary structure of long 5' UTRs on ribosome binding[118]. The authors found that the ribosome can bind to standby sites, which are structurally similar to the model-predicted folded state of the thermosensor. However, because a structured 5' UTR introduces a binding free energy penalty, it is energetically more difficult for a ribosome to bind to a structured 5' UTR than an unstructured 5' UTR. Thus, higher translation rates would be expected in the unfolded "off" state, which counters the prediction made by our proposed mechanism (Figure 1), and is not supported by experimental data (Figure 2). While the structure of the 5' UTR may impact translation rates, it is likely that the structural effect is overshadowed by the effects of RNase E-mediated transcript degradation.

However, differences in translation rates due to secondary structure can provide an alternative explanation for the higher observed "on" state in thermosensors containing only one RC site, instead of two. While this trend could be explained by considering that a transcript with two RCs will be less stable than a transcript with only one RC (Figure 3B), it is also possible that the

longer hairpins in two-RC thermosensors reduces translation rates[118]. Still, RT-qPCR data indicates that transcript stability plays a major role in the observed changes in gene expression. If differential translation rates were solely responsible for temperature-induced changes in expression, we would not expect to see the changes in mRNA abundance that are shown in Figure 5. Furthermore, the correlation between RC bulges and a reduction in leakiness indicates that RNase E access may play a more important role than secondary structure itself (Figure 3A), since no correlation was observed between leakiness and bulges in the ARC.

Another factor to consider is the impact of ribosome binding on the structure of the transcript. The ribosome footprint is approximately 30 nt[119, 120]. Though a distance of about 20 nt was left between the SD sequence and the thermosensor hairpin to account for ribosome binding, the possibility exists that the binding of the ribosome would cause partial unfolding of the thermosensor hairpin[118]. Furthermore, in the model-predicted unfolded state at high temperatures, it is possible that either the ribosome or RNase E could bind to the transcript, though simultaneous binding is unlikely due to steric hindrance. Thus, translation can occur at high temperatures (in the "off" state) until RNase E binds to the transcript. Once RNase E binds and cleaves the transcript at the RC site, the remaining transcript fragments will be quickly degraded to nucleotides by exoribonucleases and accessory factors including RNase II, RNase R, PNPase, and RhlB, among others[121]. Because RNase E cleavage is the limiting factor in the RNA degradation process, it is unlikely that the mRNA fragments would be stable enough for significant translation to occur after RNase E cleavage. Although these factors should be considered to gain a more complex understanding of the mechanism, the simple mechanism presented in Figure 1 can still be used to describe the general behavior of these thermosensors.

An understanding of this mechanism, in addition to the results shown in Figure 2, can be used to evaluate assumptions made during the design phase. For instance, the assumption was made that loop size could be neglected when estimating melting temperature (Methods). Current RNA secondary structure prediction programs calculate stem stability using nearest-neighbor approximations, and account for hairpin loops with an additional free energy term that reduces the stability of the structure[122, 123]. In general, the stability of the hairpin decreases with increasing loop size[124], which is reflected in the ΔG estimations obtained from Mfold that considered loop size (Table 1). Had loop size been considered in melting temperature approximations, it is expected that melting temperature estimates would have decreased with increasing loop size. However, the analysis of experimental data showed no relationship between thermosensor behavior and loop size.

Besides the "on" and "off" states shown in Figure 2, another potentially important characteristic of these thermosensors is their response time. To determine how quickly a thermosensor would respond to a change in temperature, the response time of the F1 thermosensor was measured. The F1 thermosensor was chosen due to its high "on" state (normalized fluorescence = 4.4 au) and low p-value ($p = 1.5 \times 10^{-2}$; Figure 2). In order to characterize the response time, Welch's t-test was used to find the time at which the fluorescence of the thermosensor at 27°C was significantly higher than that of the No-ARC control at 27°C (Figure 27 and Supplementary Methods) [125]. Based on this method, the response time for the F1 thermosensor was 1.2 hours (72 minutes) after the temperature shift from 37°C to 27°C. This is in line with a recent study on an *in vivo* RNA regulator controlling RFP expression, which reported response times of 41.7 to 72.7 minutes, depending on the speed of the response element [125]. The lower growth temperature in our experiment may have contributed to a slow response time.

After characterizing "on" and "off" states (Figure 2), response specificity (Figure 4), thermosensor mechanism (Figure 5), and the response time (Figure 27), the ability of thermosensors to participate in more complex systems can be evaluated. Construction of a three-input composite circuit demonstrates the modularity and composability of the RNA thermosensors (Figure 6). They remain functional in a different genetic context (i.e., with two different promoters pTet and psicA) and can be combined with other genetic devices to form complex logic operations. Modularity and composability are fundamental to the scalability of any genetic device[126]. Considering these characteristics alongside the designability of the RNA thermosensors and their proposed generality to diverse hosts, it is reasonable to suggest that there is wide potential applicability for these RNA thermosensors in synthetic biology.

There are several recent examples of RNA-based devices being implemented in genetic circuits, including the construction of an RNA-mediated transcriptional cascade[11], cotranscriptional *in vitro* RNA circuits[127], logic gates consisting of RNA toehold switches[128], and a pathway diverter utilizing an RNA transducer together with a promoter to determine cell fate[129]. In addition, several recent studies have used the CRISPR-Cas system, which utilizes small guide RNAs (sgRNAs) and associated proteins, in the design of genetic circuits[13, 14, 130]. The composite circuit in this study uses a unique combination of genetic controls on a variety of regulatory levels. Namely, this circuit implements inducible promoters at the transcriptional level, interacting proteins at the post-translational level, and an RNA thermosensor at the transcript stability level. By diversifying the levels of circuit regulation, the potential for different circuit architectures is expanded while the metabolic burden and overall circuit size for a given logical operation are reduced. For example, an analogous three-input AND gate demonstrated previously requires 3 layers, 5 transcription units, and 7 protein regulators (chaperones and transcription factors)[64].

The three-input composite circuit demonstrated here requires only 2 layers, 3 transcription units, and 4 protein regulators. While the circuit complexity was reduced, the same logical operation was maintained. By extension, a much larger genetic circuit could be similarly simplified by using a combination of transcriptional and post-transcriptional controls, providing opportunities for future circuit development.

In this work, we have demonstrated that a simple stem-loop structure could be designed to act as a heat-repressible RNA thermosensor in *E. coli*. These thermosensors are small, have a simple mechanism, and can be designed to have a very tightly regulated "off" state. Because of these characteristics, they can be more easily implemented into complex genetic circuits than can natural RNA thermosensors. To our knowledge, this is the first demonstration of heat-repressible RNA thermosensors designed *de novo*. Although the structure of these thermosensors is simple, design and optimization were not trivial. Insights gained from this study regarding design choices and optimization protocol will be invaluable in the implementation of these or similar thermosensors in future work.

Chapter 3: Development of design rules for reliable antisense RNA behavior in *E. coli*

While the development of genetic sensors is an invaluable component of genetic circuit design, the development of simple and programmable regulators is an equally important concern. These regulators are necessary to link inputs (e.g. extrinsic signals) with outputs (e.g. reporters, or ultimately application-specific outputs) . This work outlines the development of design rules for a versatile class of regulators known as antisense RNA (asRNA).

3.1 Abstract

A key driver of synthetic biology is the development of designable genetic parts with predictable behaviors that can be quickly implemented in complex genetic systems. However, the intrinsic complexity of gene regulation can make the rational design of genetic parts challenging. This challenge is apparent in the design of antisense RNA (asRNA) regulators. Though asRNAs are well-known regulators, the literature governing their design is conflicting and leaves the synthetic biology community without clear asRNA design rules. The goal of this study is to perform a comprehensive experimental characterization and statistical analysis of 121 unique asRNA regulators in order to resolve the conflicts that currently exist in the literature. asRNAs usually consist of two regions, the Hfq binding site and the target binding region (TBR). First, the behaviors of several high-performing Hfq binding sites were compared, in terms of their ability to improve repression efficiencies and their orthogonality. Next, a large-scale analysis of TBR design parameters identified asRNA length, thermodynamics of asRNA-mRNA complex

formation, and the percent of target mismatch as key parameters for TBR design. These parameters were used to develop simple asRNA design rules. Finally, these design rules were applied to construct both a simple and a complex genetic circuit containing different asRNAs, and predictable behavior was observed in both circuits. The results presented in this study will drive synthetic biology forward by providing useful design guidelines for the construction of asRNA regulators with predictable behaviors.

3.2 Introduction

In the recent past, engineered genetic systems have been developed to perform increasingly complex tasks, such as sense and kill pathogens[131], genomically record exposure to certain chemical or light environments[132], detect edges[6], and perform a variety of other functions[28]. As these engineered genetic systems become more and more complex, the need for rationally designed regulators with predictable behavior becomes increasingly important[126]. Antisense RNA (asRNA) is a well-studied category of RNA regulators that has been used extensively in engineered systems[133]. Synthetic asRNA regulation has been used in many metabolic engineering studies to optimize expression levels of genes within a target pathway or to down-regulate competing pathways[52, 60, 134, 135]. asRNA has also been used as an antagonistic regulator that sequesters a small guide RNA (sgRNA)[12], as a component of a counter-selection method[136], and as a tool to study essential gene knockdown[137]. Furthermore, work is underway to use asRNA in various pharmaceutical applications[138, 139]. Despite its wide range of applications, there remains no consensus regarding the design rules governing asRNA behavior. The goal of this study is to develop rules for designing asRNA regulators with predictable behaviors and minimal off target effects, which can be implemented in complex genetic circuits.

Though a large number of studies have been done to develop asRNA design rules[52-59], their conclusions are often in conflict, which leaves researchers without a clear set of guidelines when designing these regulators for metabolic engineering applications[60]. For example, some studies suggest designing long asRNAs to improve repression[54], but others suggest designing short asRNAs to prevent off-target effects.[52] In addition, while many studies recommend using thermodynamics to guide asRNA design[52, 58, 59], others have found no correlation between thermodynamic parameters and asRNA repression efficiency[57]. Additional conflicts exist regarding the necessity of asRNA-ribosome interactions and the impact of including a YUNR motif. These conflicts are discussed more thoroughly in the Results and Discussion section.

A challenge in developing asRNA design rules is the range of parameters that can be considered, and the interdependence of each of these parameters. Several small-scale studies have looked at a relatively limited number of asRNAs and sought to glean design rules based on correlations within these data sets. One drawback to this approach is that small-scale studies cannot analyze variations in a large number of parameters and are usually limited to studying the effect of just one parameter, such as binding location, asRNA length, or Hfq binding site. When studying one parameter in isolation, it can be difficult to keep all other parameters constant. For instance, when varying asRNA length, varying the thermodynamics of the asRNA-mRNA interaction as well is generally unavoidable. Disentangling these contributing factors is not possible with a small data set. Thus, varying one parameter alone may lead to results that are difficult to analyze, and may fail to capture the complexity of asRNA design.

Furthermore, many studies look at the effects of asRNA design on only one target gene. This strategy may lead to the identification of design rules that are not generalizable, but only happen to be applicable to one particular target sequence. Comparing asRNAs that were designed to

target multiple different genes will eliminate trends that are unique to one particular gene target. Finally, most asRNA design studies consider how well each regulator represses its target gene. An equally important consideration is the effect that asRNA has on non-target genes. Not only can off-target effects result in unintended consequences, but they can also reduce repression of the target gene by providing alternative targets for the asRNA[140]. Challenges such as these have led to a number of inconsistencies among studies intended to provide consistent asRNA design rules[52, 54, 57-59].

A smaller number of studies have endeavored to provide asRNA design rules using a large-scale approach and have laid the groundwork for the research presented here[52, 58]. Our work builds on these studies by performing a large-scale study of rationally designed, rather than randomly generated, asRNA regulators, providing a more comprehensive look into asRNA-target orthogonality and generating design rules that can be implemented without modifying the target gene sequence.

In this work, we seek to resolve inconsistencies in the literature by performing a large-scale analysis of 121 unique asRNA regulators. asRNAs can be thought to consist of two regions: a target binding region (TBR) containing a sequence that is complementary to the target gene, and an Hfq binding site which allows for binding of the Hfq protein. Hfq is a native chaperone protein that mediates RNA-RNA interactions by binding to a particular RNA binding site on the asRNA molecule[91, 141]. In this study, five high-performing Hfq binding sites were collected from the literature and compared in terms of their ability to improve gene repression while maintaining target specificity. Though several of these binding sites performed well, the MicF M7.4 Hfq binding site was selected as the best performer, based on its high gene repression and its limited off-target effects. Using the MicF M7.4 Hfq binding site, a comprehensive analysis of

96 unique TBRs was completed by methodically varying TBR parameters. A statistical analysis of the resulting data identified three rules for designing effective asRNAs. These design rules were implemented and verified by analyzing their effects on the expression of regulatory proteins in a simple and a complex genetic circuit. This work provides simple rules for asRNA design that will result in predictable asRNA behavior in complex genetic circuits and other applications.

3.3 Results and Discussion

3.3.1 Hfq Sites

The first aim of this study was to test several Hfq binding sites and select the best performing site for use in downstream investigations. A few studies have looked at naturally occurring Hfq binding sites in an attempt to identify the high-performers[52, 142, 143]. Here, the high-performing binding sites from several studies were analyzed in terms of both their gene regulation capabilities and their ability to avoid off-target effects. First, Sakai et al. sought to improve the function of a previously designed taRNA-crRNA system[10] by fusing Hfq sites to the taRNA[142]. They found that the MicF binding site was the highest performer. Additionally, they demonstrated that by mutating nucleotides in the hairpin structures of the MicF binding site, they could further increase the fold change of the taRNA-crRNA system. Both the MicF binding site and the mutated MicF binding site (MicF M7.4) were included in this study. Another recent study also identified the MicF binding site as a top candidate, in addition to MicC and SgrS[52]. Though among these three choices, MicC was found to be the highest performer, all three were included in this study. Finally, the Spot42 binding site has been previously identified through a library screening process as the binding site that is potentially most tolerant to variation in the fused TBR[143]. This modularity allows for design flexibility, which is a desirable characteristic in

engineered biological systems. Thus, Spot42 was included as the final Hfq binding site to be tested in this study.

Once these five Hfq binding sites had been identified, they were each fused to five distinct TBRs as shown in Figure 7a. Each of these five TBRs (TBR10, TBR15, TBR21, TBR26, and TBR30; see Table 11 for their sequences) was designed to be complementary to *rfp*, which was constitutively expressed. Additionally, each of the five TBRs was constructed without an Hfq binding site (No-Hfq) as a control. Each of the 30 asRNAs was put under the control of the pTet promoter and induced with aTc (anhydrotetracycline). By comparing induced and un-induced fluorescence levels, the percent repression could be calculated using the formula $100\%\times(F_- - F_+)/F_-$, where F_- is the un-induced normalized fluorescence (on state) and F_+ is the induced normalized fluorescence (off state) (see Materials and Methods for detail).

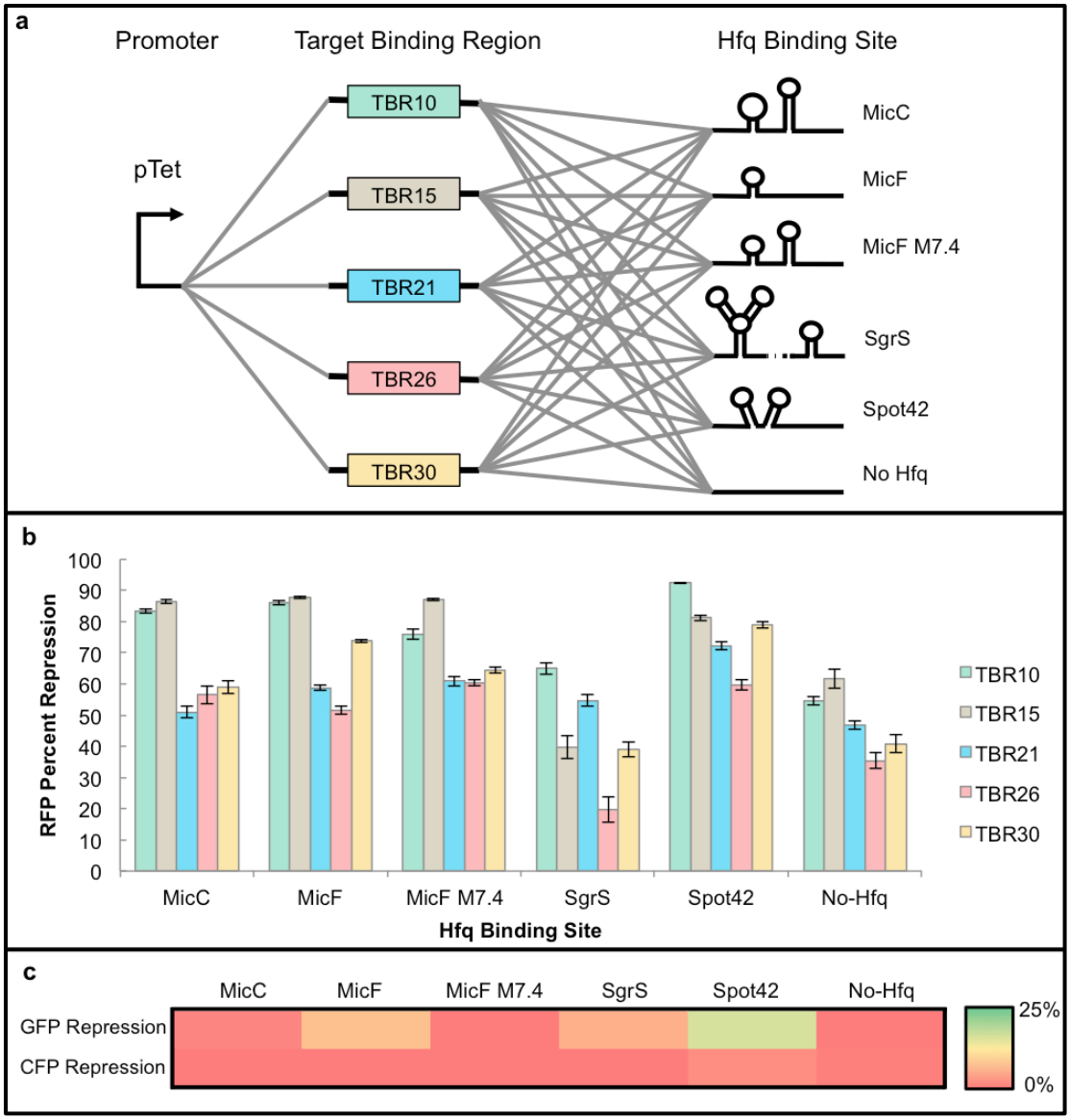

Figure 7: Hfq binding site selection. (a) Five separate TBRs (TBR10, TBR15, TBR21, TBR26, and TBR30; for their sequences, see Table 11), each targeting *rfp*, were put under the control of pTet. Each of these TBRs was fused to one of five Hfq binding sites that had been identified as high performing sites in literature. In addition, each TBR was tested without an Hfq binding site as a control, resulting in a total of 30 constructs. (b) The percent repression of each of the 30 constructs described in (a) is shown. Cells were grown in the presence of aTc (250 ng/mL) to induce expression of the asRNA. The percent repression was calculated by comparing the fluorescence of induced and uninduced cells, using the following equation: Percent Repression = $(F_{aTc-} - F_{aTc+})/F_{aTc-} \times 100$. F refers to normalized fluorescence as described in the Methods. SgrS performed poorly, and in fact had lower repression than the No-Hfq control. Spot42 had the highest average repression, and the other three Hfq binding sites (MicC, MicF, and MicF M7.4) all performed at an approximately equivalent level. Data is the average of six replicates, over two different days. Error bars represent standard error of the mean (SEM). (c) To evaluate the

orthogonality of each of the Hfq binding sites, the repression of *gfp* and *cfp* was measured. Because each of the TBRs shown in (a) was designed to target *rfp*, no repression would be expected for either *gfp* or *cfp*. As expected, repression levels were generally low for these two genes. However, Spot42 had high off-target effects for *gfp*, ruling out the use of this Hfq binding site for the remainder of the study. Data is the average of six replicates, over two different days. Detailed data is shown in Figure 28.

Spot42 had the highest overall repression with an average of 76.9% repression for all five TBRs (Figure 7b). The MicC, MicF, and MicF M7.4 binding sites also performed well, with average repressions of 67.2%, 71.6%, and 69.8%, respectively. The SgrS binding site performed poorly, only achieving 43.7% average repression, which was lower than that of the No-Hfq control (47.9% repression). Though this data alone indicates that Spot42 is the highest performing Hfq binding site, a closer look into the orthogonality of each of the binding sites provides further insight into their behavior. In order to test the orthogonality of these 30 constructs, the fluorescence of constitutively expressed *gfp* and *cfp* was also measured in the presence and absence of each asRNA. These asRNAs are not expected to repress *gfp* or *cfp*, since their TBRs were complementary to *rfp*, and not either of the other two genes. The average off-target repression for each Hfq site for *gfp* and *cfp* is shown in Figure 7c. Though most of the constructs showed low off-target effects, Spot42 showed substantial repression of *gfp*, including two TBRs that showed almost 30% *gfp* repression (Figure 28). Spot42 was the only Hfq site for which the average off-target repression (all TBRs, both GFP and CFP) was significantly higher than zero ($p=0.015$, one-tailed z-test). Because orthogonality is critical to the function of complex gene circuits, Spot42 was discarded, in anticipation of these asRNAs ultimately being implemented in more complex systems. The next three Hfq sites, MicC, MicF, and MicF M7.4, had almost equivalent *rfp* repression levels, but because MicF M7.4 had the lowest off-target effects (Figure 28), this Hfq binding site was used in the remainder of the study.

To verify that Hfq was truly involved in this interaction, two of these Hfq binding sites and the control without any Hfq binding site (TBR10-MicF M7.4, TBR10-Spot42, and TBR10-No Hfq) were tested in an Hfq deficient strain (JW4130-1) and the background strain containing the intact Hfq gene (BW25113).[144] The target plasmid was modified to contain the ampicillin resistance gene, for compatibility with JW4130-1, which contains the kanamycin resistance gene in its genome. The asRNA with no Hfq site (TBR10-No Hfq) showed no change in repression between the two strains. On the other hand, both strains containing an Hfq site experienced reduced repression efficiency in JW4130-1 (Figure 29; MicF M7.4 with 16% efficiency decrease; Spot42 with 21% efficiency decrease). This finding confirms that Hfq interacts with these sites to improve repression efficiency, which is in agreement with previous studies[145-147].

3.3.2 Target Binding Region Design

The asRNAs in this study are composed of two regions: the TBR and the Hfq binding site. Once the highest performing Hfq binding site had been identified (MicF M7.4), the focus of this study shifted to TBR design. Several studies have investigated the effects of TBR design on asRNA activity, yet no consensus exists on the design parameters that most strongly influence asRNA function. Here, we identified six categories of design parameters that are thought to influence asRNA activity, methodically varied these factors, and analyzed the effects on both target gene repression and off-target effects. The goal is to develop a list of design rules that will allow for the *de novo* design of asRNAs that can achieve high levels of target repression, high orthogonality, and generally predictable behavior. The following categories of design parameters were investigated: (1) target location, (2) mismatch, (3) length, (4) thermodynamics, (5) ribosome interactions, and (6) YUNR motif.

A total of 96 asRNAs were designed. Each asRNA contained the MicF M7.4 Hfq binding site and was under the control of the pTet promoter (Figure 8a). Each TBR targeted either *rfp*, *gfp*, or *cfp*, with 32 asRNAs targeting each gene. These three fluorescent protein genes were constitutively expressed on the target plasmid, which was co-transformed with one of the 96 asRNA plasmids. Each set of 32 asRNAs (each set targeting a different gene) contained equivalent variations in the six categories of design parameters listed above.

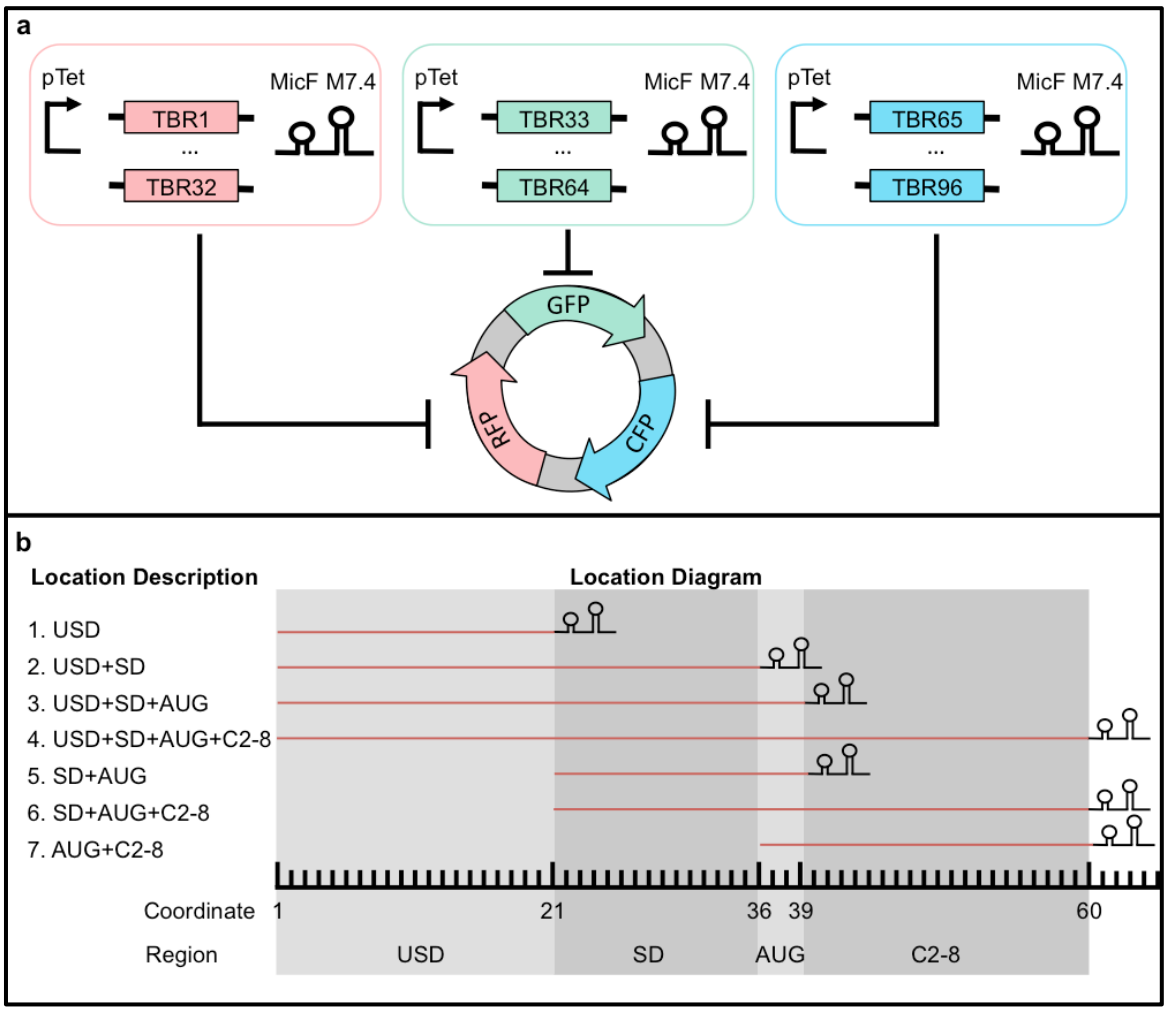

Figure 8: Experimental design for asRNA testing. (a) 96 different TBRs were designed to target one of three fluorescent proteins. Each TBR was fused with the MicF M7.4 Hfq binding site, and each asRNA was placed under the control of pTet. A third of the asRNAs targeted the *rfp* mRNA (TBR1-32), a third targeted the *gfp* mRNA (TBR33-64) and the final third targeted the *cfp* mRNA (TBR65-96). Each of the 96 plasmids was co-expressed with the target plasmid, which contained all three target genes expressed constitutively (Bba_J23105-RFP, Bba_J23116-

CFP, and Bba_J23110-GFP; http://parts.igem.org/Promoters/Catalog/Anderson). The plasmid schematic is not an exact representation of the actual target plasmid. To avoid polycistronic expression, these three genes were either oriented in opposite directions, or separated by another plasmid component (Figure 32). (b) Target binding regions (TBRs) were designed to target one of seven locations. These locations are described in the figure above. The coordinates begin at the transcription start site (TSS = 1). The coordinates vary slightly between the three target genes due to differences in the UTRs, and the coordinates shown above are only applicable to *rfp*. USD (Upstream of the Shine-Dalgarno) refers to the region of the mRNA on the 5' end of the SD sequence (coordinates =1-20 (*rfp*), 1-25 (*gfp*), and 1-21 (*cfp*)). The SD region contains the SD sequence, in addition to the nucleotides between the SD sequence and the start codon (coordinates =21-35 (*rfp*), 26-35 (*gfp*), and 22-34 (*cfp*)). The start codon is simply referred to as AUG (coordinates = 36-38 (*rfp* and *gfp*) and 35-37 (*cfp*)). The subsequent seven codons are referred to as C2-8 (Codons 2-8) (coordinates = 39-59 (*rfp* and *gfp*) and 38-58 (*cfp*)). See Table 11 for sequence information for each TBR.

3.3.3 Target location

The first design characteristic that was varied was the binding location of the asRNA. Several previous studies have shown that asRNAs are most effective when they target the translation initiation region (TIR)[52, 59, 148, 149]. However, it is not clear whether there is a particular portion of the TIR that is most important to target for effective repression. In this study, the TIR was first divided into four regions (USD, SD, AUG, and C2-8), and then each TBR was designed to target one of seven different combinations of these four regions, as shown in Figure 8b. The coordinates used to describe these locations begin with the transcription start site (TSS = +1). The first region is the portion of the 5' untranslated region (5' UTR) that is upstream of the Shine Dalgarno sequence, and is hereafter referred to as the Upstream Shine Dalgarno (USD) (Coordinates +1 to +20 in Figure 8b). The second region contains the Shine Dalgarno (SD) sequence in addition to the nucleotides between the SD Sequence and the start codon. This region is simply referred to as the Shine Dalgarno region (SD), to indicate that it includes the SD sequence (Coordinates +21 to +35 in Figure 8b). The third region is the three-nucleotide start codon (AUG), and the fourth region is codons 2-8 in the coding region (C2-8) (Coordinates +36 to +38, and +39 to +59 in Figure 8b, respectively). These regions were targeted in seven different

combinations, as described in Figure 8b. Four asRNAs targeted the same region for each of the three genes, for a total of twelve asRNAs targeting each location (12x7 = 84 total). An additional four asRNAs per gene (12 total) were designed to contain the YUNR motif as described below, resulting in a total of 96 asRNAs with the MicF M7.4 Hfq binding site.

3.3.4 Mismatch

Though most engineered TBRs are perfectly complementary to their target gene, asRNAs in natural systems bind imperfectly to their target, resulting in several shorter (8-9 nt) regions of dsRNA[150]. To determine what effect imperfect binding may have on asRNA function, the TBR sequences were designed such that the four TBRs targeting the same location would either have 0% mismatch, 1-5% mismatch, 6-15% mismatch, or 16-25% mismatch. Mismatch was introduced by deleting nucleotides from the TBR sequence in order to introduce mismatches into the asRNA-mRNA complex. Mismatch percentage was calculated by counting the number of mismatched nucleotides and dividing it by the total length of the TBR. Increasing the mismatch percentage created a range of ΔG values among otherwise similar TBRs. Mismatch location for each TBR is shown in Table 12, Table 13, and Table 14.

3.3.5 Length

The length of the TBR has long been a parameter of interest. While some studies indicate that a longer TBR is more effective in gene repression[54], other studies suggest that long asRNAs (30 nt) may lead to off-target effects[52]. In this study, the effect of length on both the target repression and off-target repression was investigated. The target location (Figure 8b) was the primary determinant of the length of the asRNA, which was altered slightly by introducing mismatch into the sequence. The lengths of the TBRs ranged from 11-59 nt. This does not include the length of the Hfq binding site or of the YUNR hairpin. Because many constructs had bulges in the asRNA-

mRNA complex, the region of continuous dsRNA varied in length. Thus, not only was the total length of the TBR considered, but the maximum length of continuous dsRNA present in the asRNA-mRNA complex was also considered as a parameter in the "length" category.

3.3.6 Thermodynamics

Perhaps one of the most commonly considered design parameters is the thermodynamics of the asRNA-mRNA interaction. This is often quantified as the ΔG of the asRNA-mRNA complex. Many studies have identified this as an important parameter[52, 58, 59], while other studies have found that there is no correlation between asRNA repression and the free energy of the asRNA-mRNA complex[57]. In this study, two thermodynamic parameters were considered. First, the ΔG of the asRNA-mRNA complex was taken into consideration (ΔG Complex). This accounted for intermolecular forces between the asRNA and mRNA, and because this structure was designed to be stable, this value was very negative. This parameter was estimated using Mfold, as described previously[51, 59]. Second, the difference between the ΔG of the TBR (ΔG TBR), which was unstructured and generally very close to zero, and the ΔG of the asRNA-mRNA complex was calculated and called ΔG Complex Formation (ΔG CF). This represents the change in free energy that results from the formation of the asRNA-mRNA complex.

3.3.7 Ribosome Interactions

Though an asRNA molecule does not contain an open reading frame, the potential still exists for the ribosome to interact with the asRNA. In certain cases, it has been shown that the inclusion of a Shine-Dalgarno sequence in the asRNA is essential for the asRNA to function[53, 61]. It has been hypothesized that interactions between the asRNA and the ribosome increase the stability of the asRNA, thus increasing the intracellular concentration of the asRNA, which contributes to its effectiveness. However, other studies have suggested that ribosome interactions may actually

prevent asRNAs from functioning properly. One possible explanation is that asRNAs that interact with the ribosome would be unable to repress their target genes due to steric hindrance[57]. In this study, asRNAs were not designed to contain a Shine Dalgarno sequence, but the potential for ribosome interaction was predicted in two ways. First, the number of start codons in each asRNA molecule was counted, including AUG, GUG, and UUG. Next, the ΔG of the complex formed by the asRNA and the anti-Shine-Dalgarno sequence in the ribosome was calculated using Mfold. These parameters were both analyzed in terms of their effect on repression efficiency.

3.3.8 YUNR Motif

A final factor that may influence the functionality of an asRNA regulator is the kinetics of the interaction between the asRNA and mRNA. Many natural asRNA regulators contain a YUNR motif. This allows the asRNA to form a secondary structure consisting of a stem-loop, wherein the nucleotides in the loop region are exposed and can easily interact with their complementary sequence[55]. It has been shown that mutating the YUNR motif in naturally occurring RNA regulators reduces the interaction kinetics[55]. This motif was used in the development of the taRNA-crRNA regulation system[10], but was shown to be unnecessary for the function of orthogonal asRNA-target pairs[58]. In this study, there were four asRNAs that contained a YUNR motif for each of the three target genes, for a total of 12 asRNAs containing YUNR motifs. Each of the asRNAs that contained a YUNR motif targeted location 3 (USD+SD+AUG; Figure 8b) so that an easy comparison could be made between asRNAs with and without the YUNR motif. The results of this comparison are shown in Figure 30. There was no clear trend in terms of the effect of the YUNR motif. For asRNAs targeting *rfp*, there was a marked decrease in repression upon the addition of the YUNR motif, while the opposite was true for asRNAs targeting *gfp*. For

asRNAs targeting *cfp*, there was a smaller magnitude change, though asRNAs without a YUNR motif had slightly higher repression levels. One possible explanation for this pattern is that the secondary structure may have been unable to properly form in some of the asRNAs. The presence of a stem-loop structure with the exposed YUNR motif in the loop region was predicted using Mfold for all of the asRNAs that contained the YUNR motif [51]. However, no experiments were performed to confirm that these structures were indeed forming *in vivo*. Future studies on the effects of the YUNR motif should look closely at the *in vivo* structure of the asRNA, and may take advantage of emerging techniques, such as in-cell SHAPE-Seq[151, 152] or other fluorescence-based techniques to characterize RNA secondary structure *in vivo*[153]. Because the addition of the YUNR structure did not appear to have a consistent effect on the repression efficiency, these twelve asRNAs were removed from the overall statistical analysis and were considered to be an independent dataset.

3.3.9 SRCC and Normalization

Once all 96 asRNAs had been designed and constructed, they were co-transformed with the target plasmid containing the three constitutively expressed target genes, resulting in 96 testing strains (Table 9, Table 10). Each of the asRNAs was expressed upon pTet induction, and the fluorescence levels of induced and un-induced cells were compared in order to calculate percent repression. On average, *cfp* repression was not as high as repression for the other two genes (Figure 31). This could be due to differences in transcript abundance and the ratio of asRNA to the target mRNA. Though *gfp*, *cfp*, and *rfp* were all located on the same plasmid (pAH197; Table 9 and Figure 32), they all had different promoters and different UTRs (Table 9, Table 10, and Table 11), which may have led to varying ratios of asRNA to mRNA within the cell. In addition, secondary structures formed by these different UTRs may have provided different levels of

access to asRNAs targeting different mRNAs. To account for these differences, percent repression was normalized to the maximum repression for that particular gene. This was done for the overall analysis that included data from all three genes, but this correction was not applied when data sets for each target gene were analyzed independently. The data was analyzed using the Spearman Rank Correlation Coefficient (SRCC). The value of the SRCC ranges between -1 and 1. An SRCC of negative one indicates a strong negative correlation, an SRCC of positive one indicates a strong positive correlation, and an SRCC of zero indicates that there is no correlation. The Student's t-test was performed to determine significance, as even correlations that are small in magnitude may be statistically significant. In addition, structures with YUNR motifs were disregarded, as mentioned previously, resulting in a final sample size of 28 asRNAs per gene, and 84 asRNAs overall.

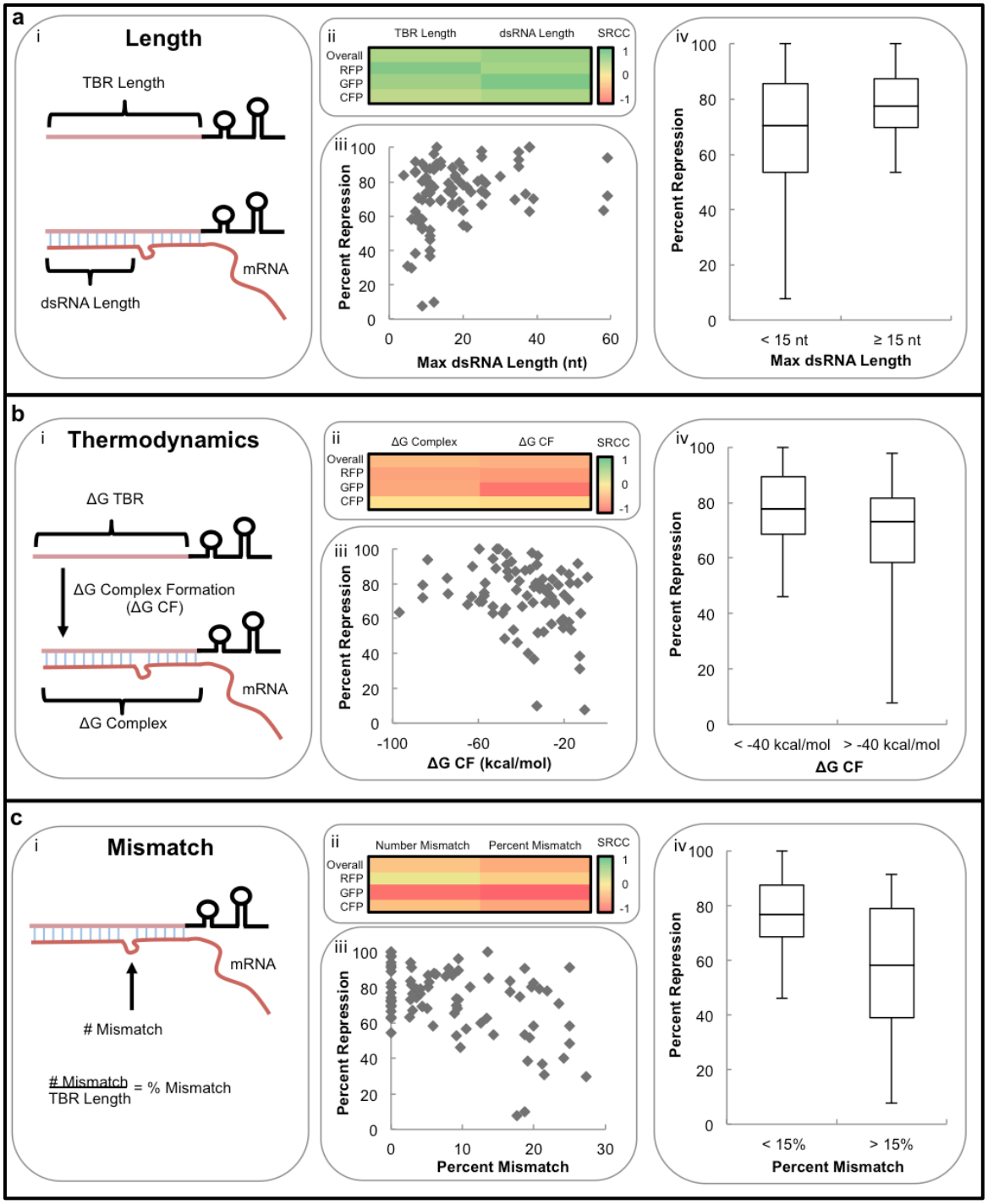

Figure 9: asRNA Parameter analysis. Several design parameters were found to have a significant impact on repression level. The Spearman Rank Correlation Coefficient (SRCC) was used to assess the degree to which each parameter is correlated with repression efficiency (a(ii), b(ii), c(ii)). A SRCC of 1 indicates a strong positive correlation, a SRCC of -1 indicates a strong negative correlation, and a SRCC of zero indicates no correlation. SRCCs were determined separately for each gene target, and repression data for all the three target genes was pooled for

the overall data. Overall data was normalized to the maximum repression for the given target gene to account for inherent differences in transcript abundance among the three target genes (Figure 31). SRCC values and p-values are shown in Table 6. For each parameter, box-and-whisker plots are used to display the distribution of the data points (a(iv), b(iv), c(iv)). The ends of the "whiskers" show the maximum and minimum values, the midpoint shows the median, and the "boxes" extend from the end of the first quartile to the end of the third quartile. (a) Two separate length parameters were considered. The results show that higher repression is correlated with an increase in TBR length, and an increase in the maximum length of dsRNA in the asRNA-mRNA complex. Data is the average of six replicates, over two different days.(i) The TBR is the region of the asRNA expected to bind directly with the mRNA target. Because not all TBRs were directly complementary to the mRNA, some asRNA-mRNA complexes had bulges that interrupted the regions of double stranded RNA. The longest uninterrupted strand of dsRNA in the asRNA-mRNA complex was identified as the maximum dsRNA length. (ii) Both length parameters, TBR length and dsRNA length, had a significant positive correlation with target repression, based on their SRCC ($p<0.05$, overall data). (iii) Scatterplot showing a positive correlation between the maximum length of a dsRNA region and repression. Longer dsRNA regions are correlated with higher target gene repression. Once the dsRNA length reaches 15 nt, percent repression appears to level off. Gene-normalized percent repression data for all target genes is shown. (iv) Box-and-whisker plot showing repression distribution of asRNAs with greater than or less than 15 nt of dsRNA. asRNA-mRNA complexes that have at least 15 nt of dsRNA in a row have a higher percent repression as compared to complexes that have less than 15 nt of dsRNA. ($p = 4.3\times10^{-3}$, two-tailed, unpaired Student's t-test). Gene-normalized percent repression data for all target genes is shown. (b) Two thermodynamic parameters were found to have negative correlation with repression. The results show that higher repression is correlated with a lower (more negative) ΔG Complex, and ΔG Complex Formation (ΔG CF; the difference between ΔG Complex and ΔG TBR). Data is the average of six replicates, over two different days.(i) The ΔG of the asRNA-mRNA complex, called ΔG Complex, was generally very negative, a characteristic of a thermodynamically favorable complex. Because TBRs were generally unstructured, the ΔG of the TBR was very close to zero in most cases. Thus, the difference (ΔG CF) between ΔG Complex and ΔG TBR, which represents the free energy change due to complex formation, was also generally negative and was strongly correlated to ΔG Complex (ii) Both ΔG Complex and ΔG CF had a negative correlation with target repression, based on their SRCC ($p<0.05$, overall data). Because these two parameters are closely correlated with one another, it is not possible to determine which particular relationship was more impactful. (iii) Scatterplot showing a negative correlation between ΔG CF and repression. A lower ΔG CF, which indicates that complex formation is thermodynamically favorable, is correlated with higher gene repression. There does not appear to be a gain in gene repression after the ΔG CF falls below -40 kcal/mol. Gene-normalized percent repression data for all target genes is shown. (iv) Box-and-whisker plot showing repression distribution of asRNAs with a ΔG CF greater than or less than -40 kcal/mol. A ΔG CF that is less than -40 kcal/mol indicates that the complex formation is strongly favored. Thus, there is a higher percent repression for TBRs with ΔG CF values less than -40 kcal/mol than for TBRs with ΔG CF values that are greater than -40 kcal/mol. ($p = 1.5\times10^{-2}$, two-tailed, unpaired Student's t-test). Gene-normalized percent repression data for all target genes is shown. (c) Two parameters associated with mismatch were considered. Both the number of mismatched nucleotides and the percent of the TBR length that

was mismatched were found to negatively correlate with percent repression. Data is the average of six replicates, over two different days. (i) Many TBRs were not perfectly complementary to their target mRNA, which resulted in bulges forming in the asRNA-mRNA complex. Both the number of bulges present in a complex (measured in nucleotides) and the percent of the length of the TBR that those bulges comprised were analyzed. (ii) Both number mismatch and percent mismatch were found to negatively correlate with percent repression, indicating that TBR sequences with a greater degree of complementarity to their target mRNA more effectively repressed their target gene ($p<0.05$, overall data). (iii) Scatterplot showing a negative correlation between the percent mismatch and repression. Higher percentages of mismatch are correlated with reduced gene repression. A substantial reduction in gene repression is not observed until approximately 15% mismatch. Gene-normalized percent repression data for all target genes is shown. (iv) Box-and-whisker plot showing repression distribution of asRNAs with TBRs containing greater than or less than 15% mismatch with their mRNA target. TBRs with less than 15% mismatch have a higher percent repression than TBRs with greater than 15% mismatch ($p = 1.9\times10^{-3}$, two-tailed, unpaired Student's t-test). Gene-normalized percent repression data for all target genes is shown.

3.3.10 Development of asRNA Design Rules

Of the six categories of design parameters that were investigated, three categories had significant effects on target repression (Figure 9). First, both the total length of the TBR and the maximum length of the dsRNA region in the asRNA-mRNA complex were positively correlated with repression (Figure 9a(ii)). That is, a longer TBR, or a longer stretch of dsRNA, was associated with an increase in target gene repression. A scatterplot showing the correlation between percent repression and maximum dsRNA length in the asRNA-mRNA complex shows a sharp increase in repression, followed by a plateau in repression after approximately 15 nt of dsRNA (Figure 9a(iii)). The asRNAs can be divided into two groups based on dsRNA length. The box-and-whisker plot in Figure 9a(iv) shows the distribution of the percent repression in both of these groups. asRNAs that had at least 15 nt of dsRNA had significantly higher target gene repression than asRNAs with less than 15 nt of dsRNA ($p=4.3\times10^{-3}$, two-tailed, unpaired, Student's t-test).

Thermodynamics were also found to have a significant impact on asRNA function. This result is expected, since several studies have already identified the importance of thermodynamics to the function of RNA regulators.[52, 58, 59] Both of the thermodynamic parameters that were evaluated,

ΔG Complex and ΔG CF, had a significant effect on target repression when looking at the overall data (Figure 9b(ii), Table 6). The stronger correlation was between percent repression and ΔG CF (Figure 9b(iii)). Lower ΔG values, which indicate more thermodynamically favorable structures, were correlated with higher repression, as expected. asRNAs with ΔG CF values that were less than -40 kcal/mol had significantly higher repression than asRNAs whose ΔG CF values were higher than -40 kcal/mol (Figure 9b(iv); $p=1.5\times10^{-2}$, two-tailed, unpaired, Student's t-test).

In addition to length and thermodynamics, the number of mismatched nucleotides also had an impact on target gene repression (Figure 9c(ii)). There was a negative correlation between the percent mismatch and target gene repression, where an increase in mismatch was associated with a decrease in target gene repression (Figure 9c(iii)). Though this trend may have been expected, it was surprising that mismatch did not result in a significant drop in gene repression unless the percent mismatch exceeded 15%. asRNAs with less than 15% mismatch had significantly higher target gene repression than asRNAs with more than 15% mismatch (Figure 9c(iv); $p=1.9\times10^{-3}$, two-tailed, unpaired, Student's t-test). Though there was no observable benefit to designing a TBR with greater than 0% mismatch, this finding is interesting. As mentioned previously, many naturally occurring asRNAs have some degree of mismatch[150], which raises the question as to what the maximum mismatch tolerance is for effective repression. The knowledge that TBRs with up to 15% mismatch can still effectively repress gene expression will be a helpful design guideline in preventing off-target effects. Ensuring that newly designed TBRs have greater than 15% mismatch with non-target genes will help reduce unwanted off-target effects.

As discussed previously, one challenge to developing rules for asRNA design is the interdependence of design parameters. To ensure that the trends observed in Figure 9 were due to each of the parameters mentioned on their own, and not the residual effects of another independent variable, the variance inflation factors were calculated for each pair of independent variables. The variance inflation factor is a measure of multicollinearity, or the extent to which multiple independent variables are correlated. Generally, a variance inflation factor greater than 10 indicates that two independent variables cannot be simultaneously used in the same analysis, because their effects cannot be separated. Variance inflation factors for all variables included in the mismatch, length, and thermodynamics categories are shown in Table 7. There is a strong correlation between ΔG Complex and ΔG CF, which makes sense because these parameters measure very similar values. This finding indicates that it is only meaningful to consider one of these two parameters when designing asRNAs. Other than ΔG Complex and ΔG CF, each of the other variables can be considered independently from one another. Thus, an asRNA with strong target repression will have a dsRNA length of at least 15 nt, a ΔG CF value of less than -40 kcal/mol, and less than 15% mismatch (though a perfectly complementary TBR is not necessary).

Two of the six categories of design parameters that were investigated did not have a significant impact on asRNA repression. First, ribosome interactions, as measured by both the number of start codons in each asRNA molecule and the ΔG of the complex formed by the asRNA and the anti-Shine-Dalgarno sequence, did not appear to have a notable impact on the repression level (Figure 33a, Figure 33b). There was no significant relationship between the ΔG of the asRNA-rRNA complex and percent repression (Figure 33a, SRCC = -0.07, $p = 0.485$). A significant relationship between the number of start codons and percent repression was detected using a one-

way analysis of variance (ANOVA), indicating that asRNAs containing two start codons had the highest average repression level (Figure 33b, $p=2.67\times10^{-4}$). However, this did not appear to be part of a larger trend, as repression decreased as the number of start codons was increased to three or more. The binding location also appeared to have no significant impact on percent repression. The average percent repression was calculated for each of the seven binding locations, and each location resulted in an approximately equal percent repression, indicating that any of the seven locations tested are effective asRNA targets (Figure 33c, ANOVA, p=0.51).

Finally, transcript abundance, which may be influenced by the stability of the asRNA, was considered as another potential influencer of repression efficiency. Six asRNAs were selected with varying repression levels, and RT-qPCR was performed to determine the abundance of these transcripts in the cell. No significant correlation was found between target gene repression and transcript abundance (Figure 34). Given the experimental design, this is unsurprising. asRNAs were expressed from the fully induced pTet promoter on a high copy plasmid (ColE1 origin, approximately 50-70 copies per cell), and target genes were expressed from weak constitutive promoters on a medium copy plasmid (p15a origin, approximately 20-30 copies per cell)[154]. Because the experiment was designed such that the number of asRNA transcripts would be in stoichiometric excess of the target transcripts, abundance was not expected to be a determining factor. However, it is important to note that in cases where there is not an excess of asRNA, abundance may begin to play an important role.

3.3.11 Off-Target Effects

In order to determine which design parameters were important to consider in avoiding off-target effects, the SRCC analysis was repeated with the maximum off-target repression as the dependent variable. The results are provided in Table 8. Though the binding location did not

have an impact on repression efficiency, binding location was the only parameter that appeared to have an impact on off-target repression. Though all of the asRNAs had a high degree of orthogonality, with repression generally limited to the target gene (Figure 10a), higher off-target repression was observed for asRNAs that targeted regions further downstream in the mRNA (Figure 10b). It was hypothesized that this trend was due to the sequence similarity between the three fluorescent protein genes. For each of the seven locations, the sequence identities were determined using Vector NTI for *gfp* and *rfp*, *rfp* and *cfp*, and *gfp* and *cfp*. The average of these three sequence identity values was plotted against the average off-target repression for each location. A positive correlation was observed ($R^2 = 0.496$, $p = 0.08$; Figure 10c), providing some explanation for the trend observed in Figure 10b. Because these three fluorescent proteins shared similar coding DNA sequences, off-target binding was more likely in the coding region than in the 5' UTR of the mRNAs. Thus, for future design of orthogonal asRNAs, it will be important to select asRNA target regions whose sequences are dissimilar to key non-target genes. Based on the results of this study, the sequence similarity between the target and non-target genes is an important design consideration. This conclusion is supported by a recent computational study demonstrating a relationship between sequence identity and off-target effects of asRNA[140]. Additionally, it is worthwhile to note that there was no significant difference in off-target effects between small (shorter than 30 nt) and large (30 nt or longer) TBRs, as had been suggested previously[52] ($p=0.73$, two-tailed, unpaired, Student's t-test).

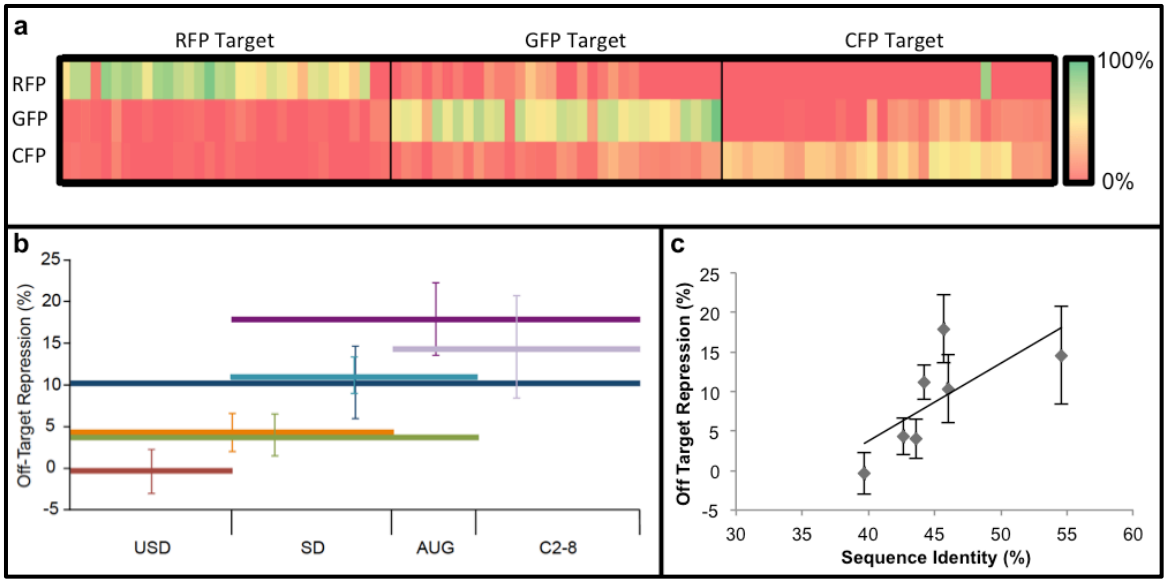

Figure 10: Orthogonality analysis. Off-target repression was measured for each of the 96 asRNAs. Because orthogonality is a desirable characteristic, repression results were analyzed to determine which design parameters are associated with lower levels of off-target repression. (a) Heat maps showing the percent repression for each of the 96 asRNAs, sorted by their gene target. The three groupings along the top show the genes that were being targeted by the asRNA. The rows represent the fluorescence that was measured. The asRNAs were generally orthogonal, with high target gene repression and low repression for the non-target genes. Experiments were conducted in triplicate, on two separate days, for a total of six replicates. (b) An SRCC analysis was completed to determine which design parameters were most associated with orthogonality. The start and end coordinates of the TBR were the only characteristics that were significantly correlated with off-target repression (Table 8, $p<0.05$, overall data). asRNAs were designed to target seven separate locations, as outlined in Figure 8b. The average off-target repression for each of these seven locations is shown, with the horizontal length and placement of the bar indicating the target region, and the vertical location of the bar indicating the maximum off-target repression. asRNAs targeting regions closer to the 5' end of the mRNA had lower off-target effects, and asRNAs targeting regions closer to the coding region of the mRNA had higher off-target effects. Data for asRNAs containing a YUNR motif is excluded, as discussed in the main text. Experiments were conducted in triplicate, on two separate days, for a total of six replicates. Error bars represent standard error of the mean (SEM). (c) In order to explain the trend observed in Figure 10b, the sequence identity for each of the seven locations was calculated, and plotted against the average off-target repression for each location. Because the three genes were all fluorescent proteins, it is expected that there will be higher sequence identity within the coding regions than within the 5' UTRs. A positive correlation is shown between sequence identity and off-target repression ($R^2 = 0.496$, $p = 0.08$).

3.3.12 Validation of Design Rules

In order to validate the design rules presented in Figure 9, a simple genetic circuit was

constructed that utilizes an asRNA as a key regulator (Figure 11b). This circuit also utilizes

regulatory proteins from the type III secretion system of *Pseudomonas aeruginosa*[46, 63, 64]. The transcription factor ExsA, which is under the control of the pBad promoter, activates transcription from the pExsD promoter, resulting in expression of *gfp*. An asRNA was designed to target the *exsA* mRNA transcript, and placed under the control of pTet. Six versions of this asRNA were generated. Two separate locations were targeted (locations 2 and 6; Figure 8b). These regions were chosen because they were similar in length, and had minimal overlap with one another. For both of these locations, three types of asRNAs were designed (Figure 11a). Type I asRNAs are perfectly complementary to the target region with 0% mismatch, and have the lowest ΔG CF of the three types (<<-40 kcal/mol). Type II asRNAs follow the design rules, but have higher levels of mismatch and higher ΔG CF values than Type I asRNAs (<-40 kcal/mol). Both Type I and Type II asRNAs followed the three design rules outlined in Figure 9, and were both expected to function well. On the other hand, Type III asRNAs were designed to be as effective as possible, with the constraint that they must break all three design rules. Type III asRNAs were expected to have lower repression efficiency than Type I and Type II asRNAs. An overview of Type I, II, and III asRNAs is found in Figure 11a, and detailed descriptions for these six asRNAs are provided in Table 16.

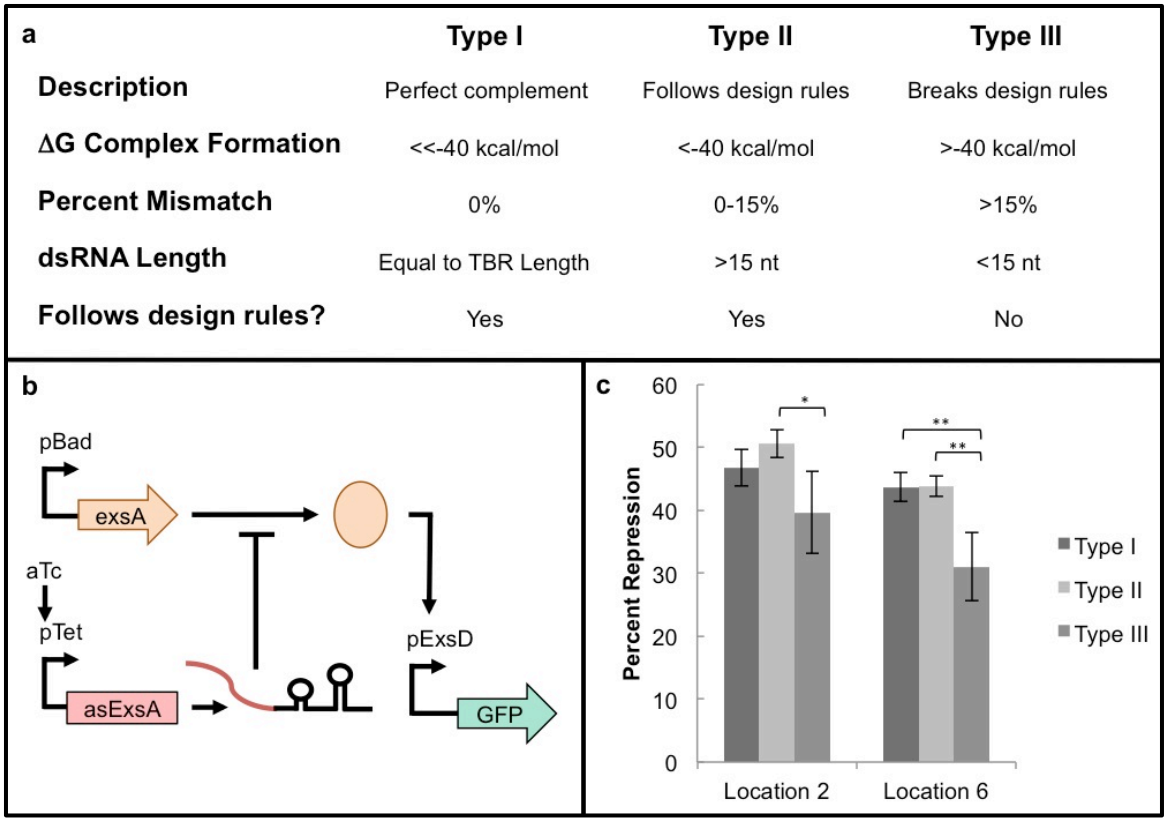

Figure 11: Validation of asRNA design rules in a simple genetic circuit. (a) In order to validate the asRNA design rules presented in Figure 9, three types of asRNA regulators were designed according to the guidelines presented here. One set of three asRNAs targets location 2, and a second set targets location 6 (Figure 8b), for a total of six asRNAs that were tested. It was expected that Type I and Type II asRNAs would perform equally well, but Type III would have reduced repression efficiency. (b) A simple genetic circuit was constructed that uses each of these six asRNAs. ExsA is a regulatory protein that activates transcription of *gfp* from the pExsD promoter. ExsA was expressed constitutively by maintaining a constant arabinose concentration (10 μM). All six asRNAs were designed to target *exsA*, and placed under the control of pTet. Thus, addition of aTc is expected to repress *gfp* expression. (c) Percent repression is shown for each of the six asRNAs tested. Significance was determined for each pair of asRNAs, and is indicated on the graph, (*, $p < 0.1$; **, $p < 0.05$; one-tailed, unpaired Student's t-test). As expected, there was no significant difference between Type I and Type II repression for either of the two target locations, suggesting that asRNAs do not exhibit diminished repression until the design rules are broken. While no significant difference was detected between Type I and Type III for location 2, significant differences in repression were found for all other pairs, as indicated on the graph, providing experimental support for the design rules presented in Figure 9. Experiments were conducted in triplicate, on two separate days, for a total of six replicates. Error bars represent standard error of the mean (SEM).

If these design rules are valid, it would be expected that both Type I and Type II asRNAs would perform equally well, since both types follow all three design rules. Type III asRNAs are expected to achieve lower levels of repression than Type I and Type II asRNAs, while still achieving an observable level of repression. As expected, Type I and Type II asRNAs both perform well, and the Type III asRNA shows diminished repression (Figure 11c). For both target locations, there was no significant difference between the Type I and Type II asRNAs ($p > 0.1$, one-tailed, unpaired Student's t-test). This data supports the hypothesis that Type I and Type II asRNAs perform equally well. Though Type I asRNAs have a lower ΔG CF, lower mismatch percent, and a longer dsRNA region, they do not perform significantly better than Type II asRNAs, which follow the same three design rules. In addition, both locations also showed the lowest level of repression for the Type III asRNA. The Type III asRNA showed reduced repression efficiency, with varying levels of significance (Figure 11c). Though these differences were small, this data demonstrates that breaking the asRNA design rules will result in a less effective asRNA, while following the design rules will result in a more effective asRNA, regardless of whether the asRNA is perfectly complementary to its target.

3.3.13 Complex Genetic Circuit Construction

To test the validity of these design rules in an even more complex circuit, asRNAs were designed as components of an AND Gate. In this AND Gate, as in Figure 11, ExsA is under the control of the pBad promoter, and *gfp* is transcribed from the pExsD promoter. However, to construct this AND gate, an additional regulatory protein was included[63, 64]. The anti-activator ExsD, which is expressed constitutively in this system by maintaining a constant 3OC6 (N-(β-ketocaproyl)-L-homoserine lactone) concentration, sequesters ExsA and prevents it from activating *gfp* transcription. An asRNA, under the control of pTet, was designed to post-transcriptionally

repress *exsD*, freeing ExsA to activate transcription of *gfp*. Thus, expression of *gfp* is only expected in the presence of both ExsA and the asRNA, which occurs when both arabinose and aTc are present (Figure 12a).

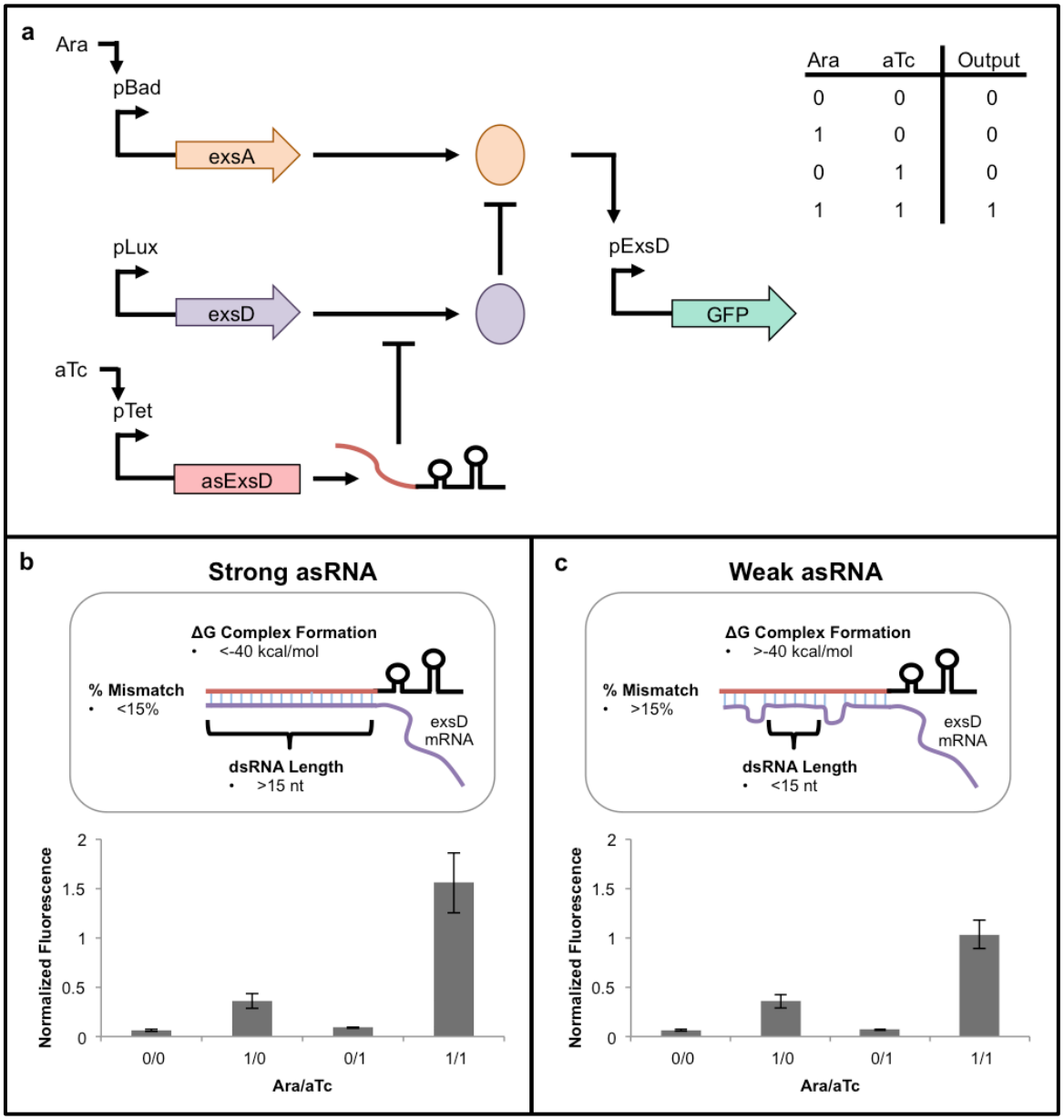

Figure 12: Application of asRNA design rules in constructing complex genetic circuits. (a) An AND Gate was constructed that utilizes asRNA to repress translation of a protein involved in gene regulation, rather than directly repress the reporter gene. ExsA activates transcription of *gfp* from the pExsD promoter, but ExsA is sequestered by ExsD. Translation into ExsD is repressed

in the presences of the asRNA (asExsD), preventing the sequestration of ExsA and allowing the expression of *gfp*. ExsD was expressed constitutively by maintaining a constant concentration of 3OC6 (5 nM). The truth table shows the expected circuit output. *gfp* expression is only expected in the presence of both arabinose and aTc. (b) Two versions of asExsD were constructed to demonstrate the effectiveness of the design rules presented in Figure 9, both targeting location 4 (Figure 8b). The "Strong" asRNA (asExsD-S) had a maximum dsRNA length of 45 bp, a ΔG CF value of -81.9 kcal/mol, and 0% mismatch. The asExsD-S circuit was tested with varying concentrations of aTc (0 ng/mL or 250 ng/mL) and arabinose (0 mM or 1 mM). The circuit behaved as expected, with a 4.3-fold change between the "on" state and the leakiest "off" state. Experiments were conducted in triplicate, on two separate days, for a total of six replicates. Error bars represent standard error of the mean (SEM). (c) A second version of asExsD was designed to demonstrate that failing to follow the design rules presented in Figure 9 would result in a less effective asRNA, and thus a lower expression level in the "on" state. The "Weak" asRNA (asExsD-W) had a maximum dsRNA length of 10 bp, a ΔG CF value of -27.7 kcal/mol, and 19% mismatch. The asExsD-W circuit was tested with varying concentrations of aTc (0 ng/mL or 250 ng/mL) and arabinose (0 mM or 1 mM). The circuit behaved as expected, with only a 2.9-fold change between the "on" state and the leakiest "off" state, which was lower than that of the asExsD-S circuit. The "on" states of the asExsD-W circuit and asExsD-S circuit were significantly different ($p<0.05$, two-tailed, paired Student's t-test). Experiments were conducted in triplicate, on two separate days, for a total of six replicates. Error bars represent standard error of the mean (SEM).

Two versions of the *exsD*-targeting asRNA (asExsD, targeting location 4) were designed to demonstrate the effects of the design rules outlined in Figure 9. The strong asExsD (asExsD-S) follows each of the three design rules presented in Figure 9 and was predicted to strongly repress *exsD* expression. The weak asExsD (asExsD-W) does not follow the design rules, but still targets the *exsD* transcript, and was expected to repress *exsD* expression to a lesser degree than does asExsD-S. In accordance with the rules presented in Figure 9, asExsD-S has a maximum dsRNA length of greater than 15 nt (i.e., 45 nt), a ΔG CF of less than -40 kcal/mol (i.e., -81.9 kcal/mol), and less than 15% mismatch (i.e., 0%) (Figure 12b). asRNAs targeting *gfp*, *cfp,* or *rfp* that met these same criteria (n=28) achieved an average of 71% repression of their target gene. On the other hand, asExsD-W has a maximum dsRNA length of less than 15 nt (i.e., 10 nt), a ΔG CF greater than -40 kcal/mol (i.e., -27.7 kcal/mol), and more than 15% mismatch (i.e., 19%) (Figure 12c). asRNAs targeting *gfp*, *cfp,* or *rfp* that met these same criteria (n=20) achieved an average

of 50% repression of their target gene. If these trends hold for both versions of asExsD, one would expect asExsD-S to perform approximately 1.4 times better than asExsD-W.

When asExsD-S is implemented into the genetic circuit, the system behaves as expected, showing a 4.3-fold increase in the "on" state [1 1] over the leakiest "off" state [1 0] (Figure 12b). When asExsD-W is implemented into the genetic circuit instead, AND gate behavior is still observed, but the "on" state [1 1] is significantly lower than it is in the asRNA-S circuit, and the asExsD-W circuit only achieves a 2.9-fold increase over the leakiest "off" state [1 0] (Figure 12c). Overall, the fold change for the asExsD-S circuit is 1.5 times higher than the fold change for the asExsD-W circuit. Though the complexity of the circuit makes it difficult to compare the difference in fold change to the difference in repression efficiency, the improvement that results from applying the design rules is in the expected range. These results show that the design rules developed in this study can be used to guide the design of effective asRNAs that target regulatory genes in a complex genetic circuit.

3.4 Conclusions

This study represents a step forward in the design of reliable asRNA regulators. Though two recent studies have utilized similar strategies to develop asRNA design rules[52, 58], our work provides a more comprehensive design-based look at both repression and off-target effects, and our system does not require target sequence modification. In one of these two studies, researchers designed 13 unique asRNA target binding regions, and then generated a library of variants (~30) to examine the relationship between ΔG and repression efficiency[52]. Another large-scale study analyzed repression and orthogonality using 23 RNA-IN mutants and 23 antisense RNA-OUT mutants[58]. A key difference between the RNA-IN/OUT system and the asRNAs studied here is that the RNA-IN/OUT system relies on modification of the target gene

sequence (RNA-IN) to match alterations in the asRNA sequence (RNA-OUT), resulting in target-regulator pairs. Although this study examined a slightly different regulatory system, it also identified thermodynamic parameters as important design considerations. Our study builds on both reports by providing the comprehensive set of rationally designed asRNA regulators (total 121). Our target genes do not require sequence modification, enabling easy implementation in the regulation of chromosomal gene targets, or any gene target that may be difficult to modify. Furthermore, we provide a comprehensive look at orthogonality by analyzing the effects of both the Hfq binding site and the TBR on two non-target genes, resulting in a large data set (i.e., repression abilities of 121 asRNAs were assessed for one target gene and two non-target genes; 363 total combinations).

3.5 Materials and Methods

3.5.1 Sequence Design

Designing each TBR sequence began with analysis of the target mRNA sequence. The portion of the mRNA that was targeted by the TBR started at the predicted transcription start site (TSS) and extended 24 nucleotides into the coding region. This sequence was divided into the USD, SD, AUG, and C2-8, as described in the main text. To quantify the target location for downstream analysis, gene coordinates were used. The predicted TSS was set as "1", and coordinates were applied to each nucleotide by counting in the 5' to 3' direction. Coordinates differ slightly for each gene due to differences within the UTR sequences, but for each gene, a coordinate of "1" refers to the TSS. USD coordinates were 1-20 (*rfp*), 1-25 (*gfp*), and 1-21 (*cfp*). SD coordinates were 21-35 (*rfp*), 26-35 (*gfp*), and 22-34 (*cfp*). AUG coordinates were 36-38 (*rfp* and *gfp*) and 35-37 (*cfp*). C2-8 coordinates were 39-59 (*rfp* and *gfp*) and 38-58 (*cfp*). The TBR sequence was determined by taking the reverse complement of the target location (locations 1-7, Figure 8b).

For each target gene, there were four TBRs per location, for a total of 28 TBRs per gene, and 84 TBRs total. An additional four TBRs were designed for each target gene by adding a YUNR structure to the TBRs targeting location 3 (Figure 8b). The YUNR structure was introduced by identifying a YUNR motif in the 3' end of the TBR, and then adding complementary nucleotides that would force the YUNR motif into the loop section of a hairpin loop. The addition of the YUNR-bearing TBRs brought the total number of TBRs to 32 per gene and 96 total. The four TBRs that targeted the same location differed in their percent mismatch, which was altered by deleting nucleotides within the TBR to cause bulges to form in the asRNA-mRNA complex. Each of the four TBRs had either 0%, 1-5%, 6-15%, or 16-25% mismatch with the mRNA target.

Once each of the 96 sequences had been designed, several characteristics of each TBR were determined. Mfold was used to estimate the free energy of the secondary structure of the TBR (ΔG TBR)[51]. The free energy of the asRNA-mRNA complex (ΔG Complex) was estimated using the DINAMelt application available online, as described previously[59]. The TBR sequence and the target mRNA sequence (including only the region that was specifically targeted) were entered as the two sequences, energy rules were set to RNA, and the application predicted the structure and ΔG of the complex. This ΔG value was recorded as ΔG Complex. The predicted structure provided by Mfold was used to predict the total number of mismatched nucleotides and the maximum uninterrupted dsRNA length. The total number of mismatched nucleotides was estimated by counting the total number of unpaired nucleotides, including bulges on either side of the predicted structure. The maximum uninterrupted dsRNA length was identified by counting the number of nucleotides in the longest region of dsRNA in the predicted structure. Once these values had been recoded, several other values could be determined through simple calculations. The length of the TBR was recorded by counting the number of nucleotides in the TBR. The

actual percent mismatch was calculated by dividing the total number of mismatched nucleotides by the length of the TBR. The difference between the unbound and bound TBR ΔG values (ΔG Complex Formation, or ΔG CF) was determined by subtracting ΔG TBR from ΔG Complex. Mfold was also used to predict the secondary structure of the entire asRNA (TBR and Hfq binding site) to ensure that the TBR was expected to appear in a region of single stranded RNA.

The DINAMelt application was also used to predict the binding energy between the TBR and the anti-Shine Dalgarno site in the ribosome. Here, the TBR and the anti-SD sequence for DH10B (ACCTCCTTA) were entered as the two sequences and the ΔG was recorded. Finally, the number of start codons in each asRNA molecule was counted, including AUG, GUG, and UUG. These parameters were used as a proxy for the likelihood that the TBR would interact with the ribosome.

To construct the simple circuit shown in Figure 11, six asRNAs were designed to target the *exsA* transcript according to the guidelines presented in Figure 11a. The parameters associated with each of these asRNAs can be found in Table 16. To construct the AND gate, two asRNAs were designed to target the *exsD* transcript, and are referred to as "asExsD-S" and "asExsD-W". Following the design rules laid out in Figure 9, "asExsD-S" had a maximum dsRNA length of 45 bp, a ΔG CF of -81.9 kcal/mol, and 0% mismatch. "asExsD-W" had a maximum dsRNA length of 10 bp, a ΔG CF of -27.7 kcal/mol, and 19% mismatch.

3.5.2 Plasmid Construction

Five TBRs targeting *rfp* were chosen to take part in the Hfq binding site analysis. These five TBRs were placed under the control of pTet via blunt end ligation as described previously[49]. A terminator was included downstream of the Hfq binding site (T0 Terminator, Table 11). Hfq

binding sites that had been identified from literature were inserted adjacent to the TBR sequence using either blunt end ligation or Golden Gate assembly, depending on the size of the binding site.[90] Potential off-target binding between each Hfq binding site sequence and each target gene's translation initiation region was assessed using Mfold, and no significant interactions were identified. Once MicF M7.4 had been identified as the highest performing Hfq site, the remaining 91 TBRs were inserted upstream of the Hfq binding site via blunt end ligation. For experiments in BW25113 and JW4130-1, the kanamycin resistance gene in the target plasmid (pAH197) was replaced with an ampicillin resistance gene using the Golden Gate assembly method (generating pAH332, Table 9). For the six asRNAs for the simple circuit (Figure 11) and the two asRNAs for the AND gate (Figure 12), each asRNA TBR was inserted upstream of the MicF M7.4 Hfq binding site via blunt end ligation. pAH148 (pLux-exsD+pBad-exsA) was constructed using the Golden Gate assembly method, and pTS118 (pExsD-GFP) and pTS001 (pBad-exsA) were taken from previously published work[63, 90]. All plasmids, strains, and key DNA sequences (i.e. genes, promoters, UTRs, Hfq binding sites, terminators, and TBRs) are summarized in Table 9, Table 10, and Table 11, respectively.

3.5.3 Reporter Fluorescence Measurements to Determine Percent Repression

Testing strains were generated by co-transforming the asRNA plasmid with the target plasmid (pAH197; Table 9), which contained three fluorescent protein genes constitutively expressed (Bba_J23105-RFP, Bba_J23116-CFP, and Bba_J23110-GFP; http://parts.igem.org/Promoters/Catalog/Anderson). To avoid polycistronic mRNA, *gfp* and *cfp* had opposite orientations, *rfp* and *cfp* were separated by the origin of replication, and *rfp* and *gfp* had opposite orientations and were separated by the kanamycin resistance gene (Figure 32). *E. coli* DH10B was used throughout the study unless otherwise indicated. For testing in the Hfq

deficient strain, asRNA plasmids were co-transformed with the alternative target plasmid (pAH332; Table 9) in both BW25113 and JW4130-1[144]. For the simple circuit shown in Figure 11, cells were co-transformed with three plasmids: pTS001, pTS118, and one of the six plasmids containing an asRNA targeting *exsA* (pAH333-pAH335, pAH339-pAH341) (Table 9). For the AND gate, cells were co-transformed with either the asExsD-S (pAH146) or the asExsD-W (pAH290) plasmid, pLux-exsD+pBad-exsA (pAH148), and pExsD-GFP (pTS118) (Table 10). Cells were grown overnight in LB media (Miller) with the appropriate antibiotics and diluted to OD_{600}=0.2. After 2 h of growth at 37°C and 250 rpm, cells were diluted 600X into 600 μL of media with anhydrotetracycline (aTc) supplementation (0 ng/mL or 250 ng/mL), and grown for 8 hours at 37°C. For the ExsA circuit (Figure 11), cells were diluted 600X into 600 μL of media with 10 μM arabinose for constitutive *exsA* expression, in addition to aTc (0 ng/mL or 250 ng/mL). For the AND gate induction, cells were diluted 600X into 600 μL of media with 3OC6 (N-(β-ketocaproyl)-L-homoserine lactone, 5 nM) for constitutive *exsD* expression, in addition to aTc (0 ng/mL or 250 ng/mL) and arabinose (0 mM or 1 mM). After 8 hours of growth, cells were centrifuged and resuspended in phosphate-buffered saline for fluorescence measurements. Measurements were taken with a Tecan Infinite M200 plate reader. Absorbance (Abs) was measured at 600 nm to monitor cell growth. GFP was measured at excitation = 483 nm and emission = 530 nm, RFP was measured at excitation = 535 nm and emission = 620 nm, and CFP was measured at excitation = 435 nm and emission = 483 nm. *E. coli* DH10B was grown to provide a background fluorescence level, and a strain of *E. coli* containing the target reporter plasmid alone was grown to provide a positive control for the fluorescence values. To normalize the data, fluorescence was calculated according to the following equation, where F = normalized fluorescence, f = raw fluorescence, and A = absorbance: $F = ((f_{asRNA\ cell}/A_{asRNA\ cell}) -$

$(f_{DH10B}/A_{DH10B}))$ / $((f_{+\ control}/A_{+\ control}) - (f_{DH10B}/A_{DH10B}))$. Repression was calculated by comparing the fluorescence of cells grown with and without aTc supplementation, using the following formula: Percent Repression = $(F_{aTc-} - F_{aTc+})/F_{aTc-} \times 100$.

Chapter 4: Enabling Complex Genetic Circuits to Respond to Extrinsic Environmental Signals

By expanding the range of genetic sensors available to synthetic biologists, we hope to broaden the possibilities for genetic circuit construction. In contrast to Chapter 2, which utilized a *de novo* design approach, this Chapter utilizes genetic sensors that have been identified through the discovery approach. In other words, in this work we harvest sensors that exist in nature, and characterize and modify these sensors to suit the needs of the circuit. These pH and temperature sensors are integrated with the protein regulators used in Chapter 2 to generate a set of AND gates. Finally, an asRNA from Chapter 3 is added to the circuit architecture to generate a set of NAND gates. Thus by integrating concepts and parts developed in the previous two chapters, this final chapter successfully demonstrates the construction of environmentally-responsive genetic circuits.

4.1 Abstract

Synthetic biology has the potential to improve a broad range of metabolic engineering processes and address a variety of medical and environmental challenges. However, in order to engineer genetic circuits that can meet the needs of these real-world applications, further characterization of genetic sensors that respond to relevant extrinsic and intrinsic signals is required, and these sensors must be implemented in complex genetic circuits. In this work, we construct the first AND and NAND gates that can respond to temperature and pH, two signals that have relevance in a variety of real-world applications. A pH-responsive promoter and a temperature-responsive promoter were extracted from the *E. coli* genome, characterized, and modified to suit the needs of the genetic circuits. These promoters were combined with components of the type III secretion

system in *Salmonella typhimurium* and used to construct a set of AND gates with up to 23-fold change. Next, an antisense RNA was integrated into the circuit architecture to invert the logic of the AND gate and generate a set of NAND gates with up to 1,168-fold change. These circuits provide the first demonstration of complex pH- and temperature-responsive genetic circuits, and lay the groundwork for the use of similar circuits in real-world applications.

4.2 Introduction

Since its inception, synthetic biology has produced an increasingly wide variety of complex circuits that carry out programmed behaviors in a predictable manner. Notable accomplishments such as the genetic toggle switch[3], repressilator[4], layered logic gates[64], circuits capable of storing information or retaining robust genetic memory[62, 63] and a diversity of CRISPR-based circuits[12, 14, 15, 130] have laid the groundwork for transformational real-world applications. While the overarching goal of synthetic biology remains focused on transitioning genetic circuits to these applications[155], synthetic biologists are faced with several formidable challenges. One key challenge is the discovery and characterization of genetic sensors capable of responding to both extrinsic (extracellular) and intrinsic (intracellular) signals, and a second challenge is the integration of these sensors into robust genetic circuits capable of sensing complex environments and responding in a predictable and reliable manner[35, 156-160].

Many of the existing sensors used in genetic circuits have been adapted from natural systems, and respond to chemical inducers (e.g., aTc, IPTG, or arabinose)[34]. Though these sensors have been essential tools in the development of synthetic biology, the need for chemical inducers adds cost and requires active monitoring of the system[35, 36]. Furthermore, these inducers can have pharmacological side effects, which precludes their use in medical applications[22]. To replace these more traditional inducible promoters, there is a need for genetic sensors that can sense and

respond to the natural environment of the cell. This environment can be defined by both the extrinsic conditions (e.g., temperature, pH, oxygen levels, and light) and the intrinsic conditions (e.g., metabolite concentrations and redox state)[37]. Genetic sensors that are able to sense and respond to extrinsic and intrinsic conditions will facilitate the development of genetic circuits that are sensitive to their environment and able to perform useful functions in real-world applications.

Several of these sensors have been developed, and a few have been integrated into genetic circuits. Metabolite sensors have been used as a directed evolution tool in several ways. For example, metabolite sensors have been used to monitor product formation in real time with fluorescent proteins, enabling efficient screening of enzyme variants[161, 162]. In a similar strategy, metabolite sensors were used to imbue a fitness advantage to productive cells, allowing researchers to easily select for the highest producers[163]. Metabolite sensors have also been used in dynamic regulation systems that modify gene expression levels to optimize process efficiency, overcome burden limitations, or avoid buildup of toxic intermediates[18, 21, 36, 159, 164]. In addition, genetic sensors that respond to both intrinsic and extrinsic signals have been used to monitor cellular stress in bioprocess development[37], and as components of a fatty acid- and copper presence/phosphate starvation-responsive AND gate[165]. Because such strategies are limited by the breadth of available sensors, platform techniques for biosensor development are valuable. One such technique uses G-protein coupled receptors to sense chemicals secreted from the cell, and was implemented in the production of medium-chain fatty acids[166].

Extrinsic sensors are playing an important role in the development of synthetic biology tools for medical applications. For example, light-responsive genetic circuits have vast potential in gene therapy applications[22, 23]. In addition, circuits have been constructed that can sense disease-

specific biomarkers and respond by releasing an appropriate therapeutic[25]. Perhaps, the most widely-researched medical application for synthetic biology is the treatment of cancer. Many studies have endeavored to design genetic circuits that can distinguish cancerous cells from healthy cells by sensing signals such as microRNAs and hypoxia[26, 27]. It is essential for such circuits to accurately identify the cancer microenvironment in order to avoid false positives and potentially dangerous side effects. Including multiple sensors in these circuits can improve their accuracy and specificity, but requires complex and robust logic[28, 29]. Additional circuits have been developed that incorporate both extrinsic and intrinsic signals for a variety of environmental applications. These include a complex circuit for the detection of metal ions[29], a circuit that expresses nitrogenase genes in response to low oxygen[33], and several complex light-sensitive circuits[6, 167].

An ongoing challenge in synthetic biology is to expand the range of well characterized sensors that can be integrated into complex genetic circuits[35, 156-160]. These efforts support the overarching goal of transitioning genetic circuits to real-world applications. In this work, we focus on the characterization of genetic sensors for temperature and pH, and the construction of complex genetic circuits that respond to these two extrinsic conditions. These signals can be difficult to integrate into genetic circuits because large variations in temperature and pH can engender broad changes to the host cell's metabolism and physiology. However, these extrinsic signals are relevant in both metabolic engineering and medical applications,[38-43] and thus the sensors and circuits developed here are necessary components of the synthetic biology toolbox.

In this study, we construct the first AND and NAND gates that respond to both temperature and pH in *E. coli*. After characterizing previously-identified pH- and temperature-sensitive promoters from the *E. coli* genome, libraries of mutants were generated to provide a broad toolbox of

sensors for downstream genetic circuit construction. These promoters were placed to control the transcription of a chaperone (SicA) and transcription factor (InvF) from the type III secretion system in *Salmonella typhimurium*, generating a functioning AND gate[48]. Next, an antisense RNA was integrated into the circuit in order to invert the behavior and create a functioning NAND gate. This work demonstrates the characterization of genetic sensors for pH and temperature, and the first integration of these sensors into complex genetic circuits.

4.3 Materials and Methods

4.3.1 Plasmid Construction

The pCspA promoter and pAsr promoter were both amplified from the *E. coli* MG1655 genome and placed upstream of *gfp* on a plasmid using Golden Gate assembly[90], resulting in pTS048 and pAH005, respectively (Table 17, Table 18, and Table 19). The AND and NAND logic gates were constructed using genetic components of the type III secretion system from *Salmonella typhimurium*, which had been previously optimized in genetic circuits[64]. Using Golden Gate assembly, the pCspA promoter was inserted upstream of *sicA*, forming pAH305. Using the same method, pAsr was inserted upstream of *invF*, resulting in pAH143 (Table 17). To prepare pAH305 for saturation mutagenesis, an additional ATG was inserted directly upstream of the *sicA* coding region, generating pAH293.

Saturation mutagenesis was performed for both pAH005 and pAH293 by amplifying the template plasmid with primers containing random bases at the target location, digesting with DpnI (New England Biolabs), phosphorylating with T4 polynucleotide kinase (New England Biolabs), and ligating using T4 DNA ligase (New England Biolabs). The ligated plasmid library was then electroporated into *E. coli* DH10B[89], the resulting colonies were pooled and the plasmid library was extracted (Zymo Research Zyppy™ Plasmid Miniprep Kit), and the sequence was

verified. This sequence-verified plasmid library was electroporated into *E. coli* DH10B. The pAH293 plasmid library was co-transformed with pAH143 and pSicA-*gfp*. Individual colonies were then screened for activity, as described below. pAsr variants that produced a range of "on" states were sequenced, and identified as pAH353-pAH356. pAH293 variants that resulted in working AND gates were sequenced, and identified as pAH312, 314-317, and 320-322 (Table 17, Table 18).

To construct the NAND gate, a constitutively expressed *gfp* was inserted onto pAH143 using Golden Gate assembly to generate pAH325. The JC1H1 variant of pAH005 (Table 20) was inserted into pAH325 via unimolecular ligation as described above, generating pAH346. To construct pAH326, the *gfp* on the pSicA-*gfp* plasmid was replaced with a previously published asRNA targeting *gfp*[47].

4.3.2 Growth Conditions and Fluorescence Measurements

Cultures were grown overnight in LB at pH=7 and 37°C. They were then diluted 10X into M9 minimal media with appropriate antibiotics in a deep well plate with a total culture volume of 300 μL. The minimal media was supplemented with 0.4% glycerol (Fisher Scientific), 0.105 g/L leucine (Sigma), 2 g/L casamino acids (BD Bacto™ Casamino Acids), and 0.3 g/L thiamine hydrochloride (Sigma), and was acidified with HCl to either pH=7 or pH=5. Cultures were grown at the appropriate temperature until early stationary phase, which occurred at 20 hours at 37°C and 25 hours at 27°C. The pH of the culture was measured at the end of the growth phase to ensure that it had not changed from the initial pH.

For the temperature pulse experiments, plates were initially grown in pH=7 media at 27°C, but were transferred to a 37°C incubator after 2, 4, 10, 15, 18, or 20 hours, and then measurements

were done after 25 hours of total growth. For the pH pulse experiments, cultures were initially grown in pH=5 media, but were centrifuged and resuspended in pH=7 media after 2, 4, 10, 15, or 18 hours and returned to the 37°C incubator, and then measurements were done after 20 hours of total growth. All growth for the pH pulse experiments was completed at 37°C.

Once cultures had reached early stationary phase (which occurred at 20 hours at 37°C and 25 hours at 27°C), cultures were centrifuged, resuspended in phosphate buffered saline, and absorbance and fluorescence were measured. A Tecan Infinite M200 plate reader was used to take these measurements. Absorbance was measured at 600 nm, and GFP fluorescence was measured at excitation = 483 nm and emission = 530 nm. Fluorescence was divided by absorbance, and the fluorescence levels of *E. coli* DH10B were subtracted to account for cellular autofluorescence. GFP fluorescence was reported as follows: $GFP/Abs - GFP_{DH10B}/Abs_{DH10B}$.

Culture conditions were changed slightly to allow for growth improvements in the NAND gates. Improved growth and fold change were observed when seed cultures were grown in the "off" state (pH=5 and 27°C). Using a seed culture in the "off" state is preferable, because if "on" state cells (already expressing *gfp*) are inoculated into fresh media which is subsequently grown in the "off" condition (pH=5, 27°C), residual GFP left in the system from the seed culture would limit the observable fold change between the "on" and "off" states. This problem is exacerbated at lower dilutions, thus modifying the culture conditions to allow for growth at a higher dilution from an "off" state seed culture improved the behavior of the NAND gates. NAND cultures were grown overnight in LB at pH=5 and 27°C (the "off" state). They were then diluted 100X into M9 minimal media with appropriate antibiotics in 14 mL culture tubes with a total culture volume of 1 mL. Fluorescence and absorbance were measured after 48 hours at each temperature. At the

end of the growth period, 300 μL of culture was transferred to a deep well plate and measurements were completed as described above.

4.4 Results

4.4.1 pH-Responsive Promoter Characterization

In order to develop a pH-responsive circuit, a pH-responsive promoter was first obtained from the *E. coli* genome. The promoter of *asr* (acid shock RNA) has been previously shown to be induced in response to low pH[168]. The *asr* promoter (pAsr) is thought to be under the control of the RstBA two-component system, and pAsr sequence contains a binding site for the response regulator RstA. To characterize this promoter, pAsr was extracted from the genome of *E. coli* MG1655 and placed upstream of *gfp*. The strain of *E. coli* containing this plasmid was grown at pH=7 and pH=5, and fluorescence was measured at stationary phase. As reported previously, this promoter showed a strong pH response. We observed 278-fold higher expression in cells exposed to pH=5 as compared to cells exposed to pH=7 (Figure 13A). To further characterize this promoter, the effects of pH pulses of varying durations were investigated. Cells were grown in pH=5 media for varying lengths of time. With increasing duration of pH=5 exposure, *gfp* expression at stationary phase increased, but leveled off after approximately 15 hours of exposure (Figure 35).

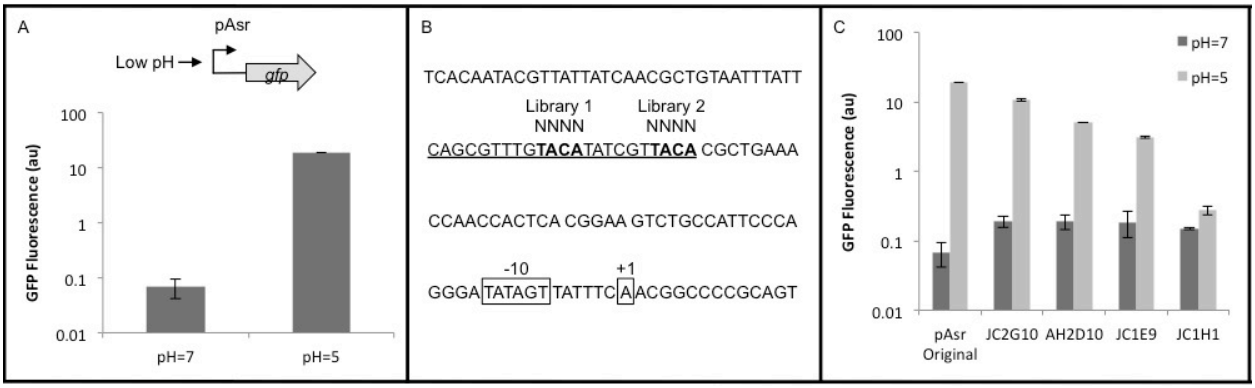

Figure 13: pH-Responsive Promoters. (a) *gfp* is under the control of the pAsr promoter, which turns on in response to low pH (pH = 5), and turns off in response to neutral pH (pH=7). This promoter showed 278-fold change (expression at pH=5/expression at pH=7). Experiments were conducted in triplicate and error bars represent standard error of the mean (SEM). (b) Saturation mutagenesis was used to generate a library of pAsr variants. The sequence of the pAsr promoter is shown. Because the pH response is thought to result from the binding of RstA, one of the two conserved repeats of the RstA binding sites were mutated while the other repeat remained intact. The RstA binding site is underlined, and the conserved repeats are in bold. The -10 site and transcription start site are indicated. (c) Several pAsr variants were generated, with varying levels of *gfp* expression in the "on" state. Experiments were conducted in triplicate and error bars represent standard error of the mean (SEM).

The "on" state of this promoter was very high, and though this provided a robust response, it was hypothesized that this level of expression may be excessive for downstream circuit design. Considering that our ultimate goal was to integrate this promoter into a genetic circuit, a library of promoter variants was generated. This library of pAsr variants would provide design flexibility downstream with respect to the "on" expression level of the promoter. Because this pH-responsive behavior is thought to result from the binding of RstA, saturation mutagenesis was performed at the two conserved direct repeats of the RstA box (Figure 13B)[168]. The two libraries were screened, and several variants of the pAsr promoters were identified (Figure 13C). The most apparent variation was in the expression level of the "on" state of these variants. One variant with a low "on" state and only 2-fold change was identified (Figure 13C, JC1H1). The sequences of each variant are shown in Table 20.

4.4.2 Temperature-Responsive Promoter Characterization

A temperature-responsive promoter was identified in the literature. The major cold shock protein in *E. coli*, CspA, can comprise up to 2% of total soluble proteins in the cell during cold shock[169]. For this reason, the genetic parts that regulate *cspA* expression were utilized in this study. The mechanism for the temperature-responsive expression of *cspA* is thought to have both transcriptional and post-transcriptional components, though post-transcriptional regulation is believed to play a larger role[170]. The *cspA* mRNA can form two alternative secondary structures depending on the temperature[78]. At high temperatures, the Shine-Dalgarno (SD) sequence and start codon are occluded in a region of dsRNA, preventing translation. However, at low temperatures, the SD sequence and start codon are located in an unstable helix, and are more accessible to ribosomes. To account for this mechanism, not only was the *cspA* promoter region used, but the 5' UTR and the first 39 nucleotides (nt) of the *cspA* coding region were also included (Figure 14)[171]. The coding region was included because it is believed to play a role in the two alternative secondary structures of the mRNA, which is integral to the temperature response[78]. We will refer to this entire region as pCspA for the rest of the paper.

When pCspA was placed upstream of *gfp*, a 16-fold increase in expression at 27°C was observed when compared to expression at 37°C (Figure 14). The pCspA promoter was also characterized in terms of its pulse response. Cells were exposed to low temperatures for varying lengths of time, and the resulting fluorescence at stationary phase was measured. Fluorescence increased with increasing time at 27°C, but leveled off after approximately 10 hours (Figure 36).

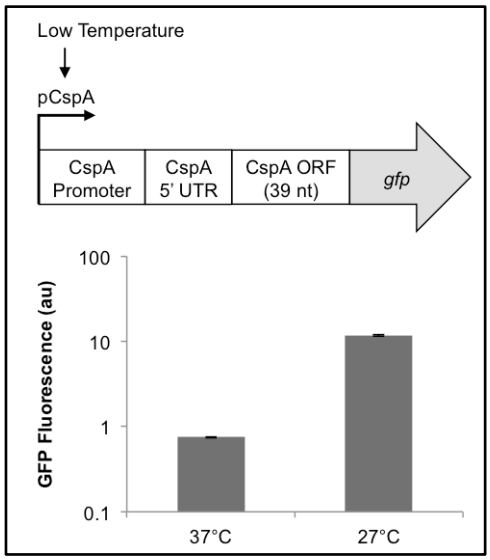

Figure 14: Temperature-responsive promoter. *gfp* is under the control of the pCspA promoter, which turns on in response to low temperature (27°C), and turns off in response to high temperature (37°C). Because the 5' UTR of the *cspA* transcript and the coding region are thought to play an important role in the temperature response[78], the native 5' UTR from *cspA*, as well as the first 39 nt of the *cspA* ORF were fused upstream of the *gfp* ORF. This promoter showed 16-fold change (expression at 27°C /expression at 37°C). Experiments were conducted in triplicate and error bars represent standard error of the mean (SEM).

4.4.3 AND Gate Construction

Once the pH- and temperature-responsive promoters had been characterized, they were first tested for orthogonality. In order to construct a working AND gate, the pAsr promoter would have to respond to pH at both high and low temperatures without showing a temperature response, and the pCspA promoter would have to respond to temperature at both high and low pH, without showing a pH response. Both promoters performed well, showing a response only to their corresponding signal (Figure 37).

Next, these two promoters were incorporated into an AND gate. This AND gate consists of a transcription factor (InvF) and chaperone (SicA) from the type III secretion system in *Salmonella typhimurium*[48]. These proteins have been previously integrated in a variety of genetic circuits[33,

[49, 64]. In this system, the presence of both SicA and InvF is required for transcription from the pSicA promoter to occur. A previously published circuit was modified to include the pAsr promoter and the pCspA promoter (Figure 15A). Unfortunately, this initial version of the AND gate did not work as expected (Figure 38). At both temperatures, expression was higher at pH=5 than at pH=7, implying that the pAsr-*invF* plasmid was working as expected, with higher levels of *invF* at pH=5 than at pH=7. However, even at pH=7, *gfp* expression was high, indicating that there was still some InvF in the system in the predicted "off" state. This suggests that the pAsr promoter is somewhat leaky in this genetic context. Though leakiness is not ideal, it was thought to be the minor contributor to the unexpected behavior. Rather, the larger problem with this circuit was hypothesized to be that the pCspA promoter was not turning off at high temperatures, as shown by the unexpected result that *gfp* expression at 37°C and pH=7 was even higher than *gfp* expression at 27°C and pH=7.

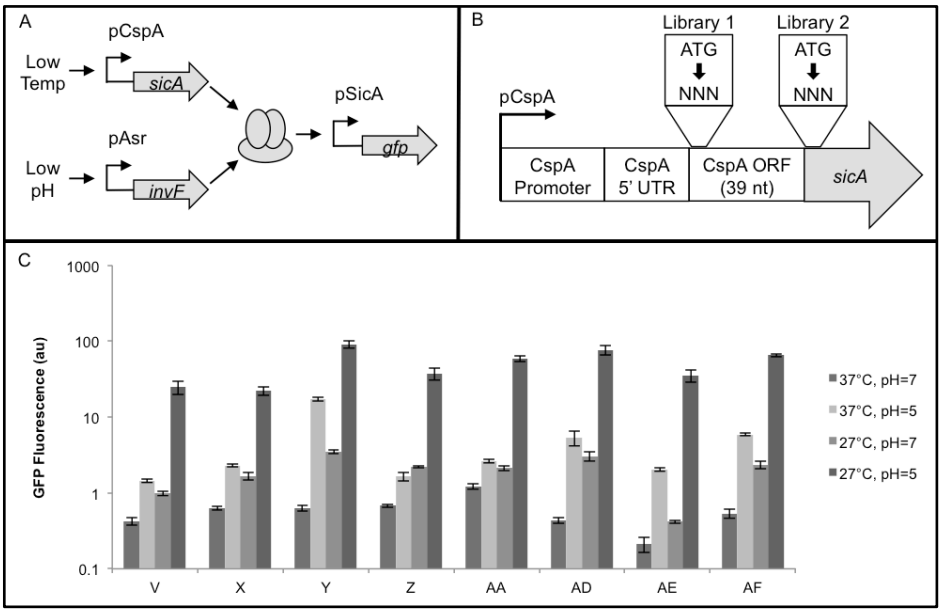

Figure 15: pH- and temperature-responsive AND gate. (a) An AND gate was constructed using components of the type III secretion system in *Salmonella typhimurium*. In this version of the circuit, pCspA controls the expression of *sicA*, and pAsr controls the expression of *invF*.

Only when both of these proteins are present in the cell, *gfp* will be expressed from pSicA[48]. (b) To obtain a working version of the pCspA-*sicA* plasmid, two start codons were mutated via saturation mutagenesis. Only one start codon was mutated at a time so that translation from either start codon would still be able to occur. Two libraries were generated via saturation mutagenesis of either the *cspA* ATG or the *sicA* ATG. (c) Several working AND gates were generated from both of the libraries. The AA AND gate was the best functioning AND gate, with 23-fold change. Each of these AND gates behaved as expected, with expression only in pH=5 media at 27°C. Sequences of each variant are shown in Table 21. Experiments were conducted in triplicate and error bars represent standard error of the mean (SEM).

To restore the temperature-sensitive behavior of this promoter, we considered the proposed mechanism[78]. In the native CspA system, the accessibility of the SD sequence and start codon in the secondary structure of the mRNA impacts the expression level. At high temperatures, the SD sequence and start codon are trapped in a stable dsRNA structure, resulting in low expression levels. In contrast, at low temperatures, the SD sequence and start codon are in an unstable helix which can be relatively easily accessed by the translation machinery, leading to higher expression at low temperatures. Because the pCspA-*gfp* and pCspA-*sicA* mRNAs were only different in their coding regions, it was hypothesized that the secondary structure of the mRNA near the translation initiation region was not forming properly at each temperature due to the change in the coding region. In order to introduce variation into this structure with the hopes of restoring the proper secondary structures, saturation mutagenesis was used to mutate the start codon. The start codon was targeted instead of the SD sequence because it is a short sequence, which would generate a small, manageable library, and we would avoid significantly changing the translation rate by mutating the SD sequence. However, the start codon is still necessary for translation to occur, so a second start codon was introduced downstream, and each of these two start codons were mutated independently of one another. The first start codon was located at the 5' end of the first 39 nt of the *cspA* coding region (start codon for *cspA* ORF; Figure 15B), while the second start codon was introduced between the first 39 nt of the *cspA* coding region and the

sicA coding region (start codon for *sicA* ORF; Figure 15B). Library 1 was generated by mutating the *cspA* start codon, and Library 2 was generated by mutating the *sicA* start codon. These two libraries were screened, and several working AND gates were identified (Figure 15C; Table 21). The best functioning AND gate (Circuit AA) demonstrated a 23-fold change from the leakiest "off" state to the "on" state. Sequences for the functioning pCpsA-*sicA* variants are shown in Table 21.

4.4.4 NAND Gate Construction

Once several working AND gates had been constructed, the same regulatory components were used to construct a NAND gate, which required an additional level of circuit complexity. The NAND gate architecture was similar to that of the AND gate. In both cases, pAsr controlled the expression of *invF* and pCspA controlled the expression of *sicA*. However, while the pSicA promoter directly controlled *gfp* expression in the AND gate, in the NAND gate, the pSicA promoter controlled the expression of an antisense RNA (asRNA) targeting *gfp,* which had been previously shown to effectively repress *gfp*[47]. Meanwhile, *gfp* was expressed constitutively from a low-copy number plasmid (Figure 16A). Because the asRNA is only expressed at 27°C and pH=5, *gfp* would be repressed in this condition and expressed in all other conditions, resulting in the characteristic NAND gate behavior.

Several functioning variants of the pCspA-*sicA* plasmid had been identified from the AND gate screening, and four of these plasmids were used in the NAND gate construction (pAH312, pAH314, pAH317, and pAH320 from AND gates V, X, AA, and AD, respectively; Figure 15C, Table 21). These plasmids were each co-transformed with a version of the pAsr-*invF* plasmid containing the JC1H1 mutation in the pAsr promoter and a constitutively expressed *gfp,* in addition to the pSicA-asRNA plasmid (pAH346 and pAH326; Table 18).

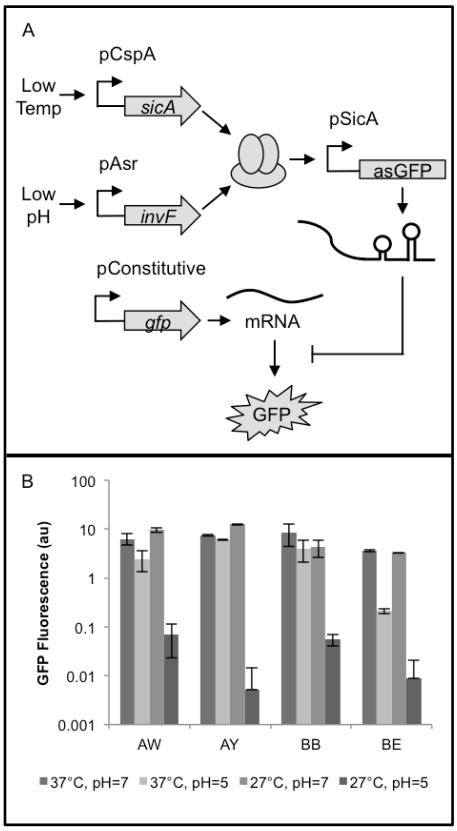

Figure 16: pH- and temperature-responsive NAND gate. (a) A NAND gate was constructed by placing an asRNA complementary to *gfp* under the control of the pSicA promoter. *gfp* was expressed constitutively on a low-copy plasmid, and induction of the asRNA in pH=5 media at 27°C would prevent translation of *gfp*. (b) Four functioning NAND gates were constructed by combining the lowest-expression pAsr variant (JC1H1; Figure 13C) and four functioning pCspA-*sicA* variants. The AY NAND gate was the best functioning NAND gate, with 1,168-fold change. It behaved as expected, with expression in every condition, except pH=5 media at 27°C. Experiments were conducted in triplicate and error bars represent standard error of the mean (SEM).

To improve the growth in the NAND gates, culture conditions were modified slightly as described in the Materials and Methods. All four NAND gates (AW, AY, BB, and BE) behaved as expected, with high expression in the three "on" conditions, and low expression in the "off" condition at 27°C and pH=5 (Figure 16B). The highest fold change was observed in the AY circuit, with 1,168-fold change between the minimum "on" state and the "off" state. The other three circuits showed high fold changes as well (Circuits AW, BB, and BE showed 36-fold, 73-

fold, and 24-fold change, respectively). These NAND gates demonstrate that extrinsic signals (i.e. temperature and pH) can be sensed by *E. coli* and converted into a programmed response via a complex genetic circuit.

4.5 Discussion

As synthetic biology moves towards its ultimate goal of building genetic circuits with real-world applications, there is a need to characterize genetic sensors that respond to cellular conditions, and to integrate these sensors into complex circuits[35, 156-160]. There are potential applications for such circuits in metabolic engineering[18, 21, 36, 161-164], medicine[22, 23, 25-27], and environmental engineering[29, 33]. In this work, we construct the first AND and NAND gates that respond to pH and temperature, two extrinsic signals with relevance in a wide variety of applications.

First, two genetic sensors were extracted from the genome of *E. coli* MG1655. The pAsr promoter, which has been shown to respond to pH[168], showed a 278-fold increase in *gfp* expression when exposed to pH=5 as opposed to pH=7. The pCspA promoter, which includes the 5'UTR and first 13 codons of the cspA coding region, demonstrated 16-fold higher *gfp* expression at 27°C than at 37°C.

These two genetic sensors were then integrated into a two-input AND gate that utilized regulatory proteins from the Type III secretion system in *Salmonella typhimurium*[48]. In order to generate functional versions of the AND gate, saturation mutagenesis was used to mutate one of the two start codons in the *sicA* transcript. Previous work has shown that the *cspA* 5' UTR, which was fused upstream of the *sicA* coding region, will form alternative secondary structures wherein the SD sequence and start codon are more accessible to the ribosome at low temperatures[78]. It was expected that saturation mutagenesis would introduce variation into the secondary structure

of the *sicA* transcript, enabling these alternative secondary structures to form correctly in the new genetic context (i.e., with the *sicA* coding region replacing the *gfp* or *cspA* coding region). Though mutating these start codons did result in temperature-responsive AND gates, it is not clear whether this effect can be attributed to significant changes in the secondary structure of the *sicA* transcript. Once each of the functional pCspA-*sicA* mutants had been sequenced, Mfold was used to predict the expected secondary structure of each transcript at 27°C and 37°C[51]. There were no major structural differences predicted between each sequence at the two temperatures. Furthermore, there were no major structural differences between the original *sicA* transcript and the mutated *sicA* transcripts. It is possible that subtle differences were not observable via Mfold, especially in light of the fact that these variant sequences only differed by 2-3 nucleotides. Future work may delve more deeply into the *in vivo* secondary structures of these transcript variants, or potential alternative mechanisms for temperature-sensitivity in the pCspA promoter.

Components of the AND gates were then repurposed to form a NAND gate. pSicA was placed upstream of an asRNA targeting constitutively expressed *gfp*, and was only activated at pH=5 and 27°C, resulting in an "off" state only in that condition. A maximum of 1,168-fold change was achieved. Notably, the genetic parts that comprised each circuit led to different circuit behaviors when utilized in either the AND gate or NAND gate. For example, the pCspA-*sicA* variant that resulted in the highest fold change circuit was not the same between the AND and NAND gates, and different pAsr variants resulted in functional versions of the two circuit types. There are several potential reasons why different genetic parts might lead to improved behavior in each circuit type. These explanations are centered around the mechanistic difference between the AND and NAND gates. First, in the AND gate, the pSicA promoter expresses *gfp* directly, but in the NAND gate, the pSicA promoter expresses an asRNA targeting *gfp*. Due to context

effects, equivalent levels of SicA and InvF may result in different transcription levels of *gfp* or asRNA from the pSicA promoter[172-174]. Secondly, it has been shown that there is not a linear relationship between RNA and protein levels[114, 115]. Therefore, even if transcription levels from the pSicA promoter were equivalent in the AND and NAND gates, because asRNA is not translated, there may not be a correlation between the amount of asRNA in the system in the NAND gate, and the amount of GFP in the system in the AND gate. In addition, impedance matching plays an important role in the mechanism of the NAND gate[156, 173]. Because the asRNA prevents translation of the *gfp* transcript, the appropriate level of asRNA expression is dependent on the *gfp* transcript level within the cell. The optimum asRNA expression level in the NAND gate may differ from the *gfp* expression level that leads to the highest fold change AND gate. Finally, because the NAND gate has an additional level of regulation, heightened metabolic burden or increasing demands on limited cellular resources (e.g., RNA polymerase, ribosomes, and energy) may contribute to performance differences between circuits or even affect overall growth rates[175-177]. Ultimately, behavioral differences between the AND and NAND gates may have a variety of mechanistic explanations, and a broad range of well-characterized genetic parts can be helpful in constructing high fold change circuits.

This work addresses two major challenges for engineering complex genetic circuits to solve real-world problems. Namely, this work characterizes genetic sensors capable of responding to extrinsic signals, and integrates these sensors into complex genetic circuits. The pH- and temperature-responsive AND and NAND gates developed here respond entirely to extrinsic conditions, and are capable of performing complex logical operations in a programmed manner. This study marks a step forward in transitioning synthetic biology away from the bench and towards real-world applications.

Chapter 5: Future Directions

This work is motivated by the potential applications of genetic circuits in metabolic engineering, medicine, and environmental protection. In this work, we have demonstrated the *de novo* development of RNA thermosensors, the characterization and modification of naturally occurring genetic sensors, and the development of design rules for versatile RNA regulators. While these parts were combined to form genetic circuits capable of performing complex logic operations, the output in each case was a reporter gene. As these circuits move into real-world applications, those reporter genes will be replaced with genes that serve any number of purposes. In order to enable these circuits to behave properly in these real-world contexts, we must develop techniques for ensuring circuit robustness, continue to characterize versatile regulators, and advance techniques for automated, high-throughput design, assembly, and screening.

5.1 Ensuring circuit robustness and diversifying host organisms

Robustness is the ability of a genetic circuit to maintain its functionality even in the face of external and internal perturbations[178]. In a highly controlled lab setting, robustness is not a major concern because perturbations can be relatively easily avoided. However, as genetic circuits are designed to solve real-world problems, they will encounter settings where perturbations are more likely to be encountered, such as in a large-scale bioreactor or inside the human gastrointestinal tract. Thus research into the mechanisms underlying biological robustness, and the development of techniques for constructing robust genetic circuits are important future research directions.

There are several strategies for maintaining biological robustness in genetic circuits[179]. The use of a selective marker, or chromosomal integration of the circuit, prevents the loss of the circuit as the population grows. Network motifs such as feedback can also be used to improve the

robustness of a genetic circuit[179-181]. Parts can also be designed to be orthogonal to the host machinery, so that the circuit does not interact with the native host genes. In addition, building circuits that do not compete for resources needed by the host is a way to reduce burden and increase robustness[179]. For example, the T7 phage polymerase can be expressed to transcribe circuit genes and avoid competition for host polymerase enzymes[182]. There are a number of other strategies for constructing robust circuits, and continued research in this field will allow genetic circuits to reliably function in a wide range of settings.

Not only is it important to ensure that circuits behave robustly in the face of external and internal perturbations, it is also essential to develop genetic parts that can function predictably in a wide range of organisms. The primary host organism used throughout this work is *E. coli* DH10B. While this strain is useful in a lab setting, other species have important capabilities, such as the ability to photosynthesize, the ability to tolerate toxic compounds, and the ability to synthesize value-added products, that provide a motivation for the development of genetic circuits in these hosts as well.

For instance, *Synechocystis* sp. PCC 6803 is a cyanobacteria that can perform both heterotrophic growth and photoautotrophic growth, it is genetically tractable, and has been engineered to produce a range of fuels and chemicals[183]. Furthermore, there is evidence for native genome-wide antisense regulation in *Synechocystis*[184], indicating that design rules similar to the ones developed here could be developed in additional hosts. Another cyanobacteria, *Cyanothece* sp. ATCC 51142, is able to perform nitrogen fixation and perform photosynthesis by temporally separating the processes[185]. The process of nitrogen fixation, and the complex regulatory mechanism that allows photosynthesis and nitrogen fixation to occur in the same cell, have important applications in agriculture. Developing circuits that utilize regulatory principles from

Cyanothece may be a step forward in engineering heterologous nitrogen fixation pathways, with the goal of ultimately replacing harmful synthetic nitrogen fertilizers.

There are a number of other organisms that may ultimately be used for synthetic biology applications. *Pseudomonas putida* is genetically tractable, is able to withstand stressful environmental conditions, and is used in many studies on bioremediation[186]. *Bacillus subtilis* is a gram positive bacterium with secretory pathways that allow products to be transported outside of the cell, making *B. subtilis* an attractive host for the production of enzymes and antibiotics[186]. *Rhodococcus opacus* can tolerate and metabolize phenolic compounds that result from lignin hydrolysis, making it a promising host for the production of value-added chemicals from biomass[187]. Further development of the parts and devices constructed here, in addition to the development of other genetic tools, will enable synthetic biology to provide solutions to real-world problems.

5.2 Characterizing versatile regulators

While it is necessary to develop techniques to ensure circuit robustness and characterize parts in new hosts, the synthetic biology toolbox will be further strengthened by the characterization of new and distinct types of regulators. In Chapter 3 of this work we have focused on the development of design rules for asRNA. However, asRNA has some inherent weaknesses. For instance, asRNA only acts as a repressor, and cannot be used to activate gene expression without adding another layer of regulation. In addition, most of the asRNAs shown here did not reach 100% repression, and may not repress expression as efficiently as would other types of regulators. In addition, the use of the Hfq binding site presents a challenge when using diverse host organisms. Hfq has not been found in all organisms, and native Hfq proteins are likely to behave differently across unrelated bacterial groups. While asRNA is still a powerful tool despite

these weaknesses, the development of similar design rules for alternative types of regulators would strengthen the synthetic biology toolbox.

The CRISPR-Cas system is a versatile tool that has provided the synthetic biology community with a powerful method to regulate gene expression. This system has been adapted to both repress (via CRISPR interference, or CRISPRi) and activate (via CRISPR activation, or CRISPRa) gene expression[188]. In CRISPRi, a nuclease deficient version of the Cas9 protein (dCas9) binds to a particular DNA sequence as directed by a single guide RNA (sgRNA) and interferes with transcription by blocking RNA polymerase, or by causing premature transcription termination[189]. In CRISPRa, dCas9 is fused to the omega subunit of the RNA polymerase and binds to the DNA to activate transcription[189]. CRISPRi has been used extensively in bacteria, whereas there are fewer reports of the use of CRISPRa in bacteria[188]. CRISPRi can achieve extremely high repression levels, and it is both orthogonal and capable of being multiplexed because of the specificity of the sgRNA targeting. However, the dCas9 protein is burdensome to the cell, and high levels can be toxic. There have been several research efforts utilizing CRISPR to construct genetic circuits[12-15]. Looking towards the future, continued development of this tool will allow for increasingly complex and diverse circuits to be constructed.

Another missing piece in the synthetic biology toolbox is a small, low-burden, easily designable regulator that can orthogonally activate gene expression. While CRISPRa has been shown to activate gene expression, there are limited bacterial examples, and the mechanism requires fusion of the omega subunit to dCas9, which may make it difficult to translate this mechanism to different organisms. A novel gene regulation system has been recently developed that may be able to fill this gap. Small transcription activating RNAs (STARs) work by binding to an anti-anti-terminator structure in the 5' end of a nascent transcript, preventing transcription

termination and effectively activating gene expression at the transcriptional level[190]. These regulators do require a particular sequence in the target gene, meaning that regulation of native genes would require chromosomal modification. However, STARs are orthogonal and can be multiplexed, and were demonstrated to function in a complex RNA-only genetic circuit. In future work, the continued development and characterization of asRNA, CRISPRi and CRISPRa, and novel regulators such as STARs, will result in a well-developed synthetic biology toolbox that will enable diverse applications of genetic circuits.

5.3 Advancing automation in synthetic biology

A final challenge facing synthetic biologists is developing the ability to design, build, and test genetic circuits in a high-throughput and automated manner. This will allow for rapid characterization of genetic sensors and regulators in a variety of contexts, further broadening the potential of synthetic biology. High-throughput techniques are required at each step of the engineering process to prevent a bottleneck from forming at the design, build, or test step. Automation will be an important component of these high throughput processes in order to reduce costs, increase speeds, and reduce errors.

There have been several efforts to provide platforms for automating genetic circuit design. In addition to requiring a library of well-characterized parts with robust functions, these platforms require a mechanism for insulating these parts to buffer against context effects. Several techniques for insulating genetic parts have been developed[191]. For example, one study showed that placing self-cleaving ribozymes in the 5' UTR of an mRNA could effectively reduce context effects[192]. A programming language called Eugene was developed to describe genetic parts and enable the development of synthetic biology design software[193]. One example of such design software is Cello, which was designed to accelerate and simplify circuit design[194]. Cello was

recently used to design 60 genetic circuits in *E. coli*, 45 of which performed correctly. As this software continues to improve, the current method of building small numbers of genetic circuits by hand is likely to be replaced by automated design methods such as this one.

As circuit design software improves, DNA assembly must become easier, faster, and less expensive to keep pace with design. Though advances in molecular cloning techniques have enabled faster construction, the process is still often done by hand. As DNA synthesis becomes cheap enough to displace molecular cloning methods, genetic circuit construction will be able to keep pace with the large numbers of circuits being rapidly generated. Over the past 20 years, the number of bases that can be chemically synthesized in one day has increased by five orders of magnitude, and the price of synthesizing this DNA has dropped dramatically[195]. It is expected that the cost of DNA synthesis will continue to drop, following the trend of DNA sequencing. It has been suggested that there is less of an economic motivation to develop inexpensive techniques for DNA synthesis than there was for DNA sequencing, because synthetized DNA is only a research tool, while DNA sequencing can have a variety of commercial applications[196]. However, the commercial potential of DNA synthesis is yet to be seen, and improvements in the cost and reliability of DNA are likely to aid the synthetic biology community.

The final challenge in automating synthetic biology is developing high-throughput methods for screening large numbers of circuit variants. The strategies to address this particular challenge depend on the nature of the output. When screening circuit behaviors, fluorescent proteins are often used as reporters. In this case it is relatively easy to assess circuit behaviors in a high-throughput manner. Similarly, in some metabolic engineering applications, the product generates a visible signal. For instance, multiplex automated genome engineering (MAGE) is a high-throughput technique to introduce genome mutations and perform directed evolution[197]. This

technique was used to optimize the production of lycopene, which is red in color and amenable to visual screening. However in cases where the product cannot be screened visually, innovative techniques have been developed. For instance, metabolite sensors have been used to monitor product formation in real time and translate product formation into a visual signal[161, 162]. Metabolite sensors have also been used to endow the most productive variants with a fitness advantage, allowing for these variants to be selected for in the population[163]. These strategies require the development and characterization of a wide range of genetic sensors.

5.4 Conclusions

As synthetic biology continues to move towards application-based design, it is essential to develop a number of capabilities that will allow genetic circuits to reliably perform the necessary tasks. First, circuits will need to function robustly in a wide range of different host organisms, genetic contexts, and environmental conditions. Next, synthetic biologists must continue to develop diverse genetic sensors and regulators that are well characterized and predictable. Finally, automation techniques are being developed that use these well-characterized, robust genetic parts to design, build, and test genetic circuits in a high throughput manner. As advances are made in each of these areas, synthetic biology will fulfill its potential and provide a meaningful contribution to society.

Appendix A: Supplementary Data for RNA Thermosensors

Table 1: Parameters of thermosensors designed in this work."# RC" is the number of RNase E cleavage sites. "# Bulges" refers to the number of groups of unpaired nucleotides bulging out from the stem on either the side of the RNase E cleavage site (RC) or the anti-RNase E cleavage site (ARC). T_m and ΔG are estimated from Mfold [51] as described in the Materials and Methods.

Thermosensor	# RC	T_m (°C)	ΔG (kcal/mol)	Loop Size (nt)	# Bulges	
					RC	ARC
A1	2	27.5	-4.5	6	1	2
A2	2	27.5	-3.3	11	1	2
A3	2	27.5	-2.9	16	1	2
B1	2	29.5	-6.6	5	0	3
B2	2	29.5	-5.7	10	0	3
B3	2	29.5	-5.3	15	0	3
C1	2	29.2	-4.4	5	2	1
C2	2	29.2	-3.5	10	2	1
C3	2	29.2	-3.1	15	2	1
D1	1	37.8	-7.8	5	0	0
D2	1	37.8	-6.9	10	0	0
D3	1	37.8	-6.5	15	0	0
E1	1	32.4	-6.9	5	0	0
E2	1	32.4	-6.0	10	0	0
E3	1	32.4	-5.6	15	0	0
F1	1	37.8	-7.8	5	0	0
F2	1	37.8	-6.9	10	0	0
F3	1	37.8	-6.5	15	0	0
G1	2	34.0	-5.9	5	0	4
H1	2	36.5	-8.4	5	0	3
I1	2	35.9	-8.6	5	0	4
J1	2	25.6	-5.4	5	0	4
K1	2	25.6	-5.4	5	0	4
L1	2	25.6	-5.4	5	0	4

Table 2: Plasmids used in this study.

Name	Parts
Controls	
pAH016	ColE1 ori; cm-R; Bba_J23104 – *gfp*
pAH021	ColE1 ori; cm-R; Bba_J23104 - No-ARC Control - *gfp*
pAH034	p15A ori; kan-R; Bba_J23105 – *rfp*
pAH048 (pTet-GFP)	ColE1 ori; cm-R; pTet – *gfp*
pAH049	ColE1 ori; cm-R; pTet - No-ARC Control – *gfp*
Thermosensors	
pAH050	ColE1 ori; cm-R; pTet - A1 Thermosensor - *gfp*
pAH051	ColE1 ori; cm-R; pTet - A2 Thermosensor - *gfp*
pAH052	ColE1 ori; cm-R; pTet - A3 Thermosensor - *gfp*
pAH053	ColE1 ori; cm-R; pTet - B1 Thermosensor - *gfp*
pAH054	ColE1 ori; cm-R; pTet - B2 Thermosensor - *gfp*
pAH055	ColE1 ori; cm-R; pTet - B3 Thermosensor - *gfp*
pAH056	ColE1 ori; cm-R; pTet - C1 Thermosensor - *gfp*
pAH057	ColE1 ori; cm-R; pTet - C2 Thermosensor - *gfp*
pAH058	ColE1 ori; cm-R; pTet - C3 Thermosensor - *gfp*
pAH059	ColE1 ori; cm-R; pTet - D1 Thermosensor - *gfp*
pAH060	ColE1 ori; cm-R; pTet - D2 Thermosensor - *gfp*
pAH061	ColE1 ori; cm-R; pTet - D3 Thermosensor - *gfp*
pAH062	ColE1 ori; cm-R; pTet - E1 Thermosensor - *gfp*
pAH063	ColE1 ori; cm-R; pTet - E2 Thermosensor - *gfp*
pAH064	ColE1 ori; cm-R; pTet - E3 Thermosensor - *gfp*
pAH065	ColE1 ori; cm-R; pTet - F1 Thermosensor - *gfp*
pAH066	ColE1 ori; cm-R; pTet - F2 Thermosensor - *gfp*
pAH067	ColE1 ori; cm-R; pTet - F3 Thermosensor - *gfp*
pAH068	ColE1 ori; cm-R; pTet - G1 Thermosensor - *gfp*
pAH069	ColE1 ori; cm-R; pTet - H1 Thermosensor - *gfp*
pAH070	ColE1 ori; cm-R; pTet - I1 Thermosensor - *gfp*
pAH071	ColE1 ori; cm-R; pTet - J1 Thermosensor - *gfp*
pAH072	ColE1 ori; cm-R; pTet - K1 Thermosensor - *gfp*
pAH073	ColE1 ori; cm-R; pTet - L1 Thermosensor - *gfp*
RNase E Rescue	
pAH045	p15A ori; kan-R; Bba_J23105 – *rfp;* Native promoters - *rne*
Genetic Circuits	
psicA-gfp [198]	ColE1 ori; cm-R; psicA - *gfp*

pTet-invF [64]	pSC101* ori; kan-R; pTet - *invF*
pBAD-sicA*[64]	p15A ori; amp-R; pBAD - *sicA**
pAH134	ColE1 ori; cm-R; psicA - D1 Thermosensor - *gfp*
pAH135	ColE1 ori; cm-R; psicA - E1 Thermosensor - *gfp*
pAH136	ColE1 ori; cm-R; psicA - E3 Thermosensor - *gfp*
pAH137	ColE1 ori; cm-R; psicA - F1 Thermosensor - *gfp*
pAH138	ColE1 ori; cm-R; psicA - F3 Thermosensor - *gfp*
pAH152	ColE1 ori; cm-R; psicA - B1 Thermosensor - *gfp*
pAH153	ColE1 ori; cm-R; psicA - C1 Thermosensor - *gfp*

Table 3. *E. coli* strains used in this study.

Name	Host Strain	Plasmids
Controls		
DH10B	DH10B	None
BL21 Star (DE3)	BL21 Star (DE3)	None
Constitutive Positive Control	DH10B	pAH016
pTet Positive Control	DH10B	pAH048, pAH034
pTet No-ARC Control	DH10B	pAH049, pAH034
BL21 Star (DE3) pTet Positive Control	BL21 Star (DE3)	pAH048, pAH034
BL21 Star (DE3) No-ARC Control	BL21 Star (DE3)	pAH049, pAH034
RNase E Rescue pTet Positive Control	BL21 Star (DE3)	pAH048, pAH045
RNase E Rescue No-ARC Control	BL21 Star (DE3)	pAH049, pAH045
No Thermosensor Circuit Control	DH10B	pBAD-sicA*, pTet-invF, psicA-gfp
Thermosensors		
A1 Inducible	DH10B	pAH050, pAH034
A2 Inducible	DH10B	pAH051, pAH034
A3 Inducible	DH10B	pAH052, pAH034
B1 Inducible	DH10B	pAH053, pAH034
B2 Inducible	DH10B	pAH054, pAH034
B3 Inducible	DH10B	pAH055, pAH034
C1 Inducible	DH10B	pAH056, pAH034
C2 Inducible	DH10B	pAH057, pAH034
C3 Inducible	DH10B	pAH058, pAH034
D1 Inducible	DH10B	pAH059, pAH034
D2 Inducible	DH10B	pAH060, pAH034
D3 Inducible	DH10B	pAH061, pAH034
E1 Inducible	DH10B	pAH062, pAH034
E2 Inducible	DH10B	pAH063, pAH034
E3 Inducible	DH10B	pAH064, pAH034
F1 Inducible	DH10B	pAH065, pAH034
F2 Inducible	DH10B	pAH066, pAH034
F3 Inducible	DH10B	pAH067, pAH034
G1 Inducible	DH10B	pAH068, pAH034
H1 Inducible	DH10B	pAH069, pAH034
I1 Inducible	DH10B	pAH070, pAH034
J1 Inducible	DH10B	pAH071, pAH034
K1 Inducible	DH10B	pAH072, pAH034
L1 Inducible	DH10B	pAH073, pAH034
RNase E Rescue		
BL21 Star (DE3) D1	BL21 Star (DE3)	pAH059, pAH034

Thermosensor		
BL21 Star (DE3) D2 Thermosensor	BL21 Star (DE3)	pAH060, pAH034
BL21 Star (DE3) D3 Thermosensor	BL21 Star (DE3)	pAH061, pAH034
BL21 Star (DE3) E2 Thermosensor	BL21 Star (DE3)	pAH063, pAH034
BL21 Star (DE3) E3 Thermosensor	BL21 Star (DE3)	pAH064, pAH034
BL21 Star (DE3) F1 Thermosensor	BL21 Star (DE3)	pAH065, pAH034
BL21 Star (DE3) F2 Thermosensor	BL21 Star (DE3)	pAH066, pAH034
BL21 Star (DE3) F3 Thermosensor	BL21 Star (DE3)	pAH067, pAH034
RNase E Rescue D1 Thermosensor	BL21 Star (DE3)	pAH059, pAH045
RNase E Rescue D2 Thermosensor	BL21 Star (DE3)	pAH060, pAH045
RNase E Rescue D3 Thermosensor	BL21 Star (DE3)	pAH061, pAH045
RNase E Rescue E2 Thermosensor	BL21 Star (DE3)	pAH063, pAH045
RNase E Rescue E3 Thermosensor	BL21 Star (DE3)	pAH064, pAH045
RNase E Rescue F1 Thermosensor	BL21 Star (DE3)	pAH065, pAH045
RNase E Rescue F2 Thermosensor	BL21 Star (DE3)	pAH066, pAH045
RNase E Rescue F3 Thermosensor	BL21 Star (DE3)	pAH067, pAH045
Genetic Circuits		
D1 Thermosensor Circuit	DH10B	pBAD-sicA*, pTet-invF, pAH134
E1 Thermosensor Circuit	DH10B	pBAD-sicA*, pTet-invF, pAH135
E3 Thermosensor Circuit	DH10B	pBAD-sicA*, pTet-invF, pAH136
F1 Thermosensor Circuit	DH10B	pBAD-sicA*, pTet-invF, pAH137
F3 Thermosensor Circuit	DH10B	pBAD-sicA*, pTet-invF, pAH138
B1 Thermosensor Circuit	DH10B	pBAD-sicA*, pTet-invF, pAH152
C1 Thermosensor Circuit	DH10B	pBAD-sicA*, pTet-invF, pAH153

Table 4: Genetic parts used in this study.

Part name	Type and source	DNA sequence
gfp	Gene [198]	atgagtaaaggagaagaacttttcactggagttgtcccaattcttgttgaattagatggtgatgttaatgggcac aaattttctgtcagtggagagggtgaaggtgatgcaacatacggaaaacttacccttaaatttatttgcactact ggaaaactacctgttccatggccaacacttgtcactactttgacttatggtgttcaatgcttttcaagatacccag atcatatgaaacggcatgactttttcaagagtgccatgcccgaaggttatgtacaggaaagaactatatttttca aagatgacgggaactataagacacgtgctgaagtcaagtttgaaggtgatacacttgttaatagaatcgagtta aaaggtattgattttaaagaagatggaaacattcttggacacaagttggaatacaactataactcacacaatgta tacatcatggcagacaaacaaaagaatggaatcaaagttaacttcaaaattagacacaacattgaagatggaa gcgttcaactagcagaccattatcaacaaaatactccaattggcgatggccctgtccttttaccagacaaccatt acctgtccacacaatctgccctttcgaaagatcccaacgaaaagagagaccacatggtccttcttgagtttgta acagctgctgggattacacatggcatggatgaactatacaaaaggcctgcagcaaacgacgaaaactacgc ttaagtagcttaa
rfp	Gene [199]	atggcgagtagcgaagacgttatcaaagagttcatgcgtttcaaagttcgtatggaaggttccgttaacggtca cgagttcgaaatcgaaggtgaaggtgaaggtcgtccgtacgaaggtacccagaccgctaaactgaaagtta ccaaaggtggtccgctgccgttcgcttgggacatcctgtccccgcagttccagtacggttccaaagcttacgtt aaacacccggctgacatcccggactacctgaaactgtccttcccggaaggtttcaaatgggaacgtgttatga acttcgaagacggtggtgttgttaccgttacccaggactcctccctgcaagacggtgagttcatctacaaagtt aaactgcgtggtaccaacttcccgtccgacggtccggttatgcagaaaaaaaccatgggttgggaagcttcc accgaacgtatgtacccggaagacggtgctctgaaaggtgaaatcaaaatgcgtctgaaactgaaagacgg tggtcactacgacgctgaagttaaaaccacctacatggctaaaaaaccggttcagctgccgggtgcttacaa aaccgacatcaaactggacatcacctcccacaacgaagactacaccatcgttgaacagtacgaacgtgctg aaggtcgtcactccaccggtgcttaa
rne	Gene *E. coli* MG1655 [200]	atgaaaagaatgttaatcaacgcaactcagcaggaagagttgcgcgttgcccttgtagatgggcagcgtctgt atgacctggatatcgaaagtccagggcacgagcagaaaaaggcaaacatctacaaaggtaaaatcacccg cattgaaccgagtctggaagctgcttttgttgattacggcgctgaacgtcacggtttcctcccactaaaagaaat tgcccgcgaatatttccctgctaactacagtgctcatggtcgtcccaacattaaagatgtgttgcgtgaaggtca ggaagtcattgttcagatcgataaagaagagcgcggcaacaaaggcgcggcattaaccacctttatcagtct ggcgggtagctatctggttctgatgccgaacaacccgcgcgcgggtggcatttctcgccgtatcgaaggcg acgaccgtaccgaattaaaagaagcactggcaagccttgaactgccggaaggcatggggcttatcgtgcgc accgctggcgtcggcaaatctgctgaggcgctgcaatgggatttaagcttccgtctgaaacactgggaagcc atcaaaaaagccgctgaaagccgcccggccccgttcctgattcatcaggagagcaacgtaatcgttcgcgc attccgcgattacttacgtcaggacatcggcgaaatccttatcgataacccgaaagtgctcgaactggcacgt cagcatatcgctgcattaggtcgcccggatttcagcagcaaaatcaaactgtacaccggcgagatcccgctg ttcagccactaccagatcgagtcacagatcgagtccgccttccagcgtgaagttcgtctgccgtctggtggttc cattgttatcgacagcaccgaagcgttaacggccatcgacatcaactccgcacgcgcgacccgcggcggc gatatcgaagaaaccgcgtttaacactaacctcgaagctgccgatgagattgctcgtcagctgcgcctgcgt gacctcggcggcctgattgttatcgacttcatcgacatgacgccagtacgccaccagcgtgcggtagaaaac cgtctgcgtgaagcggtgcgtcaggaccgtgcgcgtattcaaatcagccatatttctcgctttggcctgctgga aatgtcccgtcagcgcctgagcccatcactgggtgaatccagtcatcacgtttgtccgcgttgttctggtactg gcaccgtgcgtgacaacgaatcgctgtcgctctctattctgcgtctgatcgaagaagaagcgctgaaagaga acacccaggaagttcacgccattgttcctgtgccaatcgcttcttacctgctgaatgaaaaacgttctgcggta aatgccattgaaactcgtcaggacggtgtgcgctgtgtaattgtgccaaacgatcagatggaaaccccgcac taccacgtgctgcgcgtgcgtaaaggggaagaaaccccaaccttaagctacatgctgccgaagctgcatga agaagcgatggcgctgccgtctgaagaagagttcgctgaacgtaagcgtccggaacaacctgcgctggca acctttgccatgccggatgtgccgcctgcgccaacgccagctgaacctgccgcgcctgttgtagctccagca ccgaaagctgcaccggcaacaccagcagctcctgcacaacctgggctgttgagccgcttcttcggcgcact gaaagcgctgttcagcggtggtgaagaaaccaaaccgaccgagcaaccagcaccgaaagcagaagcga aaccggaacgtcaacaggatcgtcgcaagcctcgtcagaacaaccgccgtgaccgtaatgagcgccgcg acacccgtagtgaacgtactgaaggcagcgataatcgcgaagaaaaccgtcgtaatcgtcgccaggcaca gcagcagactgccgagacgcgtgagagccgtcagcaggctgaggtaacggaaaaagcgcgtaccgccg acgagcagcaagcgccgcgtcgtgaacgtagccgccgccgtaatgatgataaacgtcaggcgcaacaag aagcgaaggcgctgaatgttgaagagcaatctgttcaggaaaccgaacaggaagaacgtgtacgtccggtt cagccgcgtcgtaaacagcgtcagctcaatcagaaagtgcgttacgagcaaagcgtagccgaagaagcgg tagtcgcaccggtggttgaagaaactgtcgctgccgaaccaattgttcaggaagcgccagctccacgcaca gaactggtgaaagtcccgctgccagtcgtagcgcaaactgcaccagaacagcaagaagagaacaatgctg ataaccgtgacaacggtggcatgccgcgtcgttctcgccgctcgcctcgtcacctgcgcgtaagtggtcagc gtcgtcgtcgctatcgtgacgagcgttatccaacccagtcgccaatgccgttgaccgtagcgtgcgcgtctcc

		ggaactggcctctggcaaagtctggatccgctatccaattgtacgtccgcaagatgtacaggttgaagagca gcgcgaacaggaagaagtacatgtgcagccgatggtgactgaggtccctgtcgccgccgctatcgaaccg gttgttagcgcgccagttgttgaagaagtggccggtgtcgtagaagcccccgttcaggttgccgaaccgcaa ccggaagtggttgaaacgacgcatcctgaagtgatcgctgccgcggtaactgaacagccgcaggtgattac cgagtctgatgttgccgtagcccaggaagttgcagaacaagcagaaccggtggttgaaccgcaggaagag acggcagatattgaagaagttgtcgaaactgctgaggttgtagttgctgaacctgaagttgttgctcaacctgc cgcgccagtagtcgctgaagtcgcagcagaagttgaaacggtagctgcggtcgaacctgaggtcaccgttg agcataaccacgctaccgcgccaatgacgcgcgctccagcaccggaatatgttccggaggcaccgcgtca cagtgactggcagcgccctacttttgccttcgaaggtaaaggtgccgcaggtggtcatacggcaacacatca tgcctctgccgctcctgcgcgtccgcaacctgttgagtaa
rne upstream region	Promoters and 5' UTR *E. coli* MG1655 [200]	acaggattcgcgccactcatttttctatgcttatatttacttttgcaccttattacttcactgcgtgatcactttattgat ggttattaaaccaatcaccagcaagaagtgaaaaaactgtgagtaagcgggtgataaatggtaaaagtcatct tgctataacaaggcttgcagtggaataatgaggccgtttccgtgtccatccttgttaaaacaagaaatttacgg aataacccattttgcccgaccgatcatccacgcagcaatggcgtaagacgtattgatctttcaggcagttagcg ggctgcgggttgcagtccttaccggtagatggaaatatttctggagagtaatacccagtctgtttctttgataatt gcgctgtttttccgcatgaaaaacgggcaaccgacactctgcgcctctttgagctgacgataaccgtgaggtt ggcgacgcgactagacacgaggccatcggttcacacccggaaaggcgttactttgcccgcagcttagtcgt caatgtaagaataatgagtaagttacg
invF	Gene [64]	atgtcattttctgaaagccgacacaatgaaaattgcctgattcaggaaggcgcgctgctttttgcgagcaggc cgttgtcgcaccagtatcaggagacctggtttttcgaccgttaaaaattgaagtactcagcaaattactggcatt tatcgatggcgcaggattagtggacacgacatatgctgaatccgataaatgggttttgctgagtcctgagtttc gcgctatttggcaagatcgtaaacgctgcgagtactggtttttgcagcaaattattacgccttctccggccttca ataaggtactggcgctgttacgaaaaagcgagagttactggttggttggctatttactcgctcagtcaaccagc ggcaacacgatgagaatgctgggagaagactatggcgtttcttataccccattttcgtcgtttgtgcagcagagc gttgggcggaaaagcgaagagtgaattacgaaactggcgtatggcgcaatcgctgctgaatagtgtagaag gccacgagaacatcacccaattagccgttaatcatggttactcatcgccttcacatttttctagtgagatcaaag agctgatcggcgtttcgccgcggaaattatcaaatattattcaattggcagacaaatga
*sicA**	Gene [64]	atggattatcaaaataatgtcagcgaagaacgtgttgcggaaatgatttgggatgccgttagtgaaggcgcca cgctaaaagacgttcatgggatccctcaagatatgatggacggtttatatgctcatgcttatgagttttataacca gggacgactggatgaagctgagacgttctttcgttacttatgcatttatgatttttacaatcccgattacaccatg ggactggcggcagtatgccaactgaaaaaacaatttcagaaagcatgtgacctttatgcagtagcgtttacgtt acttaaaaatgattatcgccccgttttttttaccgggcagtgtcaattattaatgcgtaaggcagcaaaagccag acagtgttttgaacttgtcaatgaacgtactgaagatgagtctctgcgggcaaaagcgttggtctatctggagg cgctaaaaacggcggagacagagcagcacagtgaacaagaaaaggaataa
psicA	Promoter [198]	ccacaagaaacgaggtacggcattgagccgcgtaaggcagtagcgatgtattcattgggcgtttttgaatgtt cactaaccaccgtcggggtttaataactgca
pTet	Promoter [64]	ttttcagcaggacgcactgacctccctatcagtgatagagattgacatccctatcagtgatagagatactgagc acatct
pBAD	Promoter [64]	agaaaccaattgtccatattgcatcagacattgccgtcactgcgtcttttactggctcttctcgctaaccaaaccg gtaaccccgcttattaaaagcattctgtaacaaagcgggaccaaagccatgacaaaaacgcgtaacaaaagt gtctataatcacggcagaaaagtccacattgattatttgcacggcgtcacactttgctatgccatagcatttttatc cataagattagcggatcctacctg
Bba_J23104	Promoter (http://parts.igem.org/Part:BBa_J23104)	ttgacagctagctcagtcctaggtattgtgctagc
Bba_J23105	Promoter (http://parts.igem.org/Part:BBa_J23105)	tttacggctagctcagtcctaggtactatgctagc

Table 5: Primers for RT-qPCR

Primer	Sequence	Source
cysGF	ttgtcggcggtggtgatgtc	[92]
cysGR	atgcggtgaactgtggaataaacg	[92]
hcaTF	gctgctcggctttctcatcc	[92]
hcaTR	ccaaccacgctgaccaacc	[92]
idnTF	ctgtttagcgaagaggagatgc	[92]
idnTF	acaaacggcggcgatagc	[92]
gfpF	ctgtccacacaatctgccct	This Study
gfpR	gtttgctgcaggccttttgt	This Study

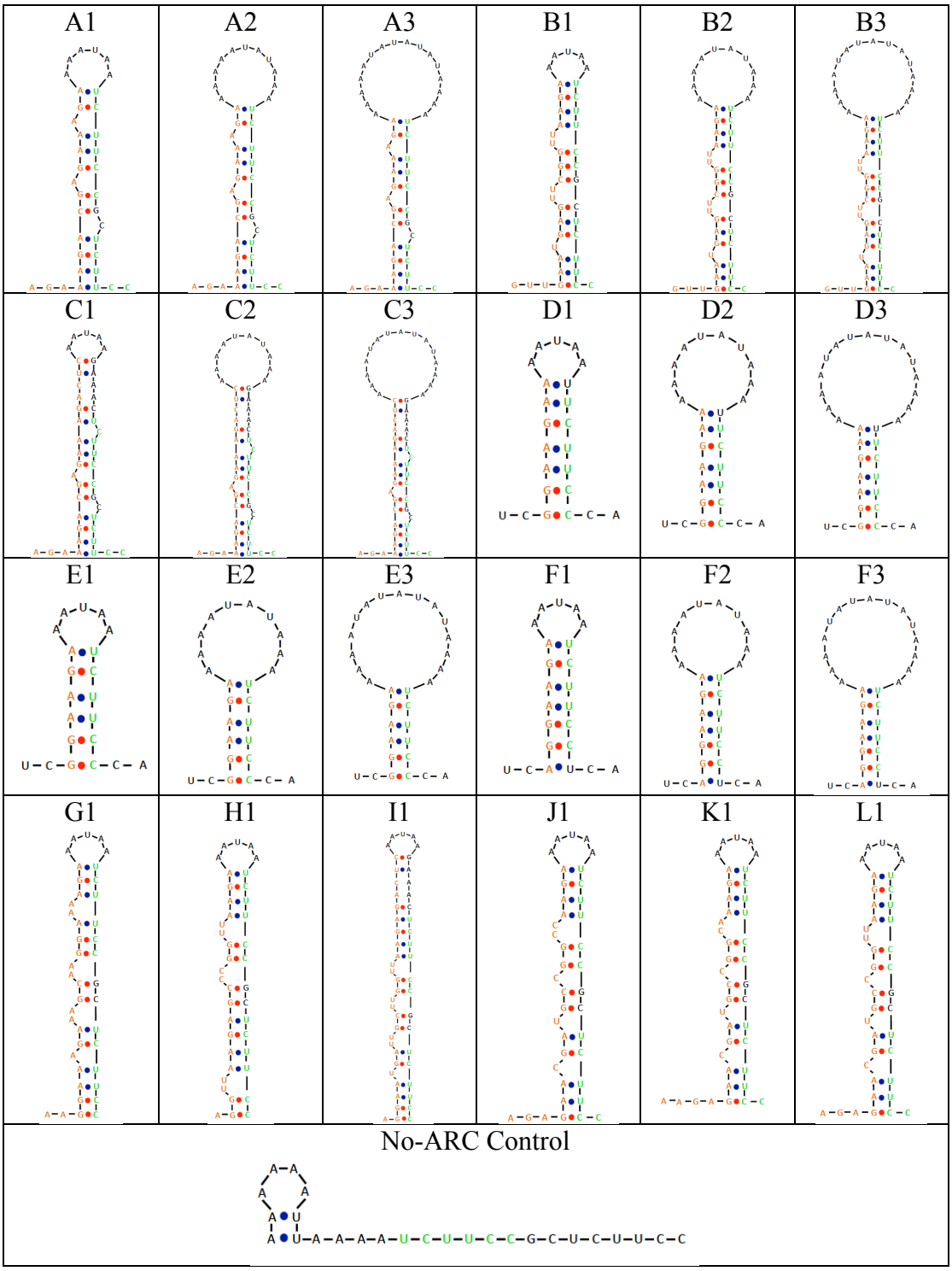

Figure 17: Thermosensor structures predicted by Mfold. The structure includes the sequence ranging from the ARC to RC, but does not extend into the SD and coding sequence. The RC is in green and the ARC is in orange. When two RNase E cleavage sites are included in the structure, a GC spacer was placed between the two sequences, as has been observed in nature [96].

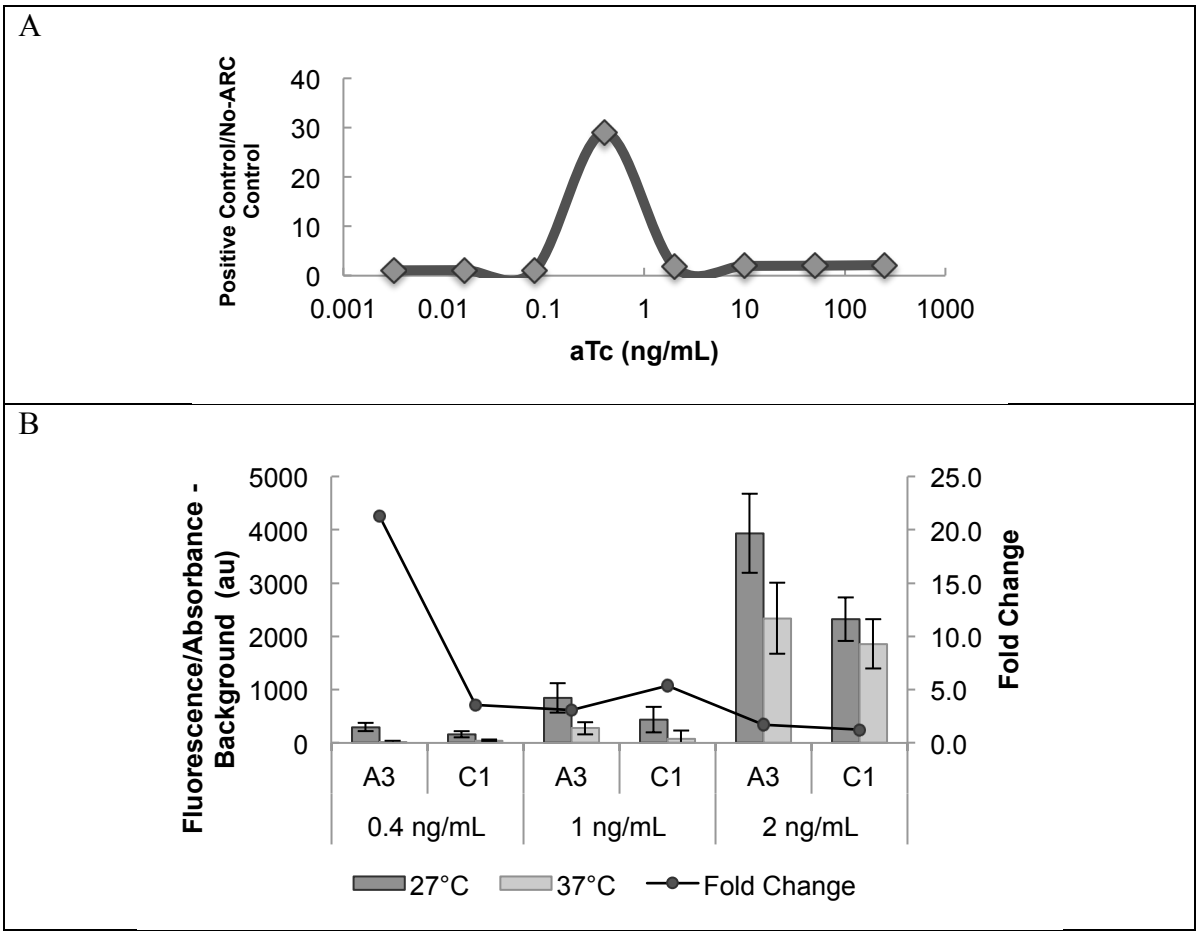

Figure 18: Transcription level optimization. Optimization of transcript level is necessary to ensure that the thermosensors will function correctly within the cell. An aTc level of 1 ng/mL was used in all experiments unless otherwise indicated. (A) Ratio of the fluorescence output of positive control (pTet-GFP) to that of No-ARC control at 37°C at 0.0032, 0.016, 0.08, 0.4, 2, 10, 50, and 250 ng/mL aTc. The line is added as a guide to the eye. Thermosensors were under the control of pTet, which allowed for control of the transcript level within the cell. A transcript level that is too low will be undetectable, and a transcript level that is too high will overwhelm the capacity of RNase E. (B) A3 and C1 thermosensors were tested at 0.4 ng/mL, 1 ng/mL, and 2 ng/mL aTc to identify the optimum aTc level for thermosensor response. Per cell fluorescence is shown by dividing fluorescence by absorbance, and subtracting the autofluorescence levels, divided by absorbance, of background cells (DH10B). GFP expression is compared at 0.4 ng/mL, 1 ng/mL, and 2 ng/mL, at both 27°C and 37°C. Fold change between 27°C and 37°C is shown by the circles, and the line is added as a guide to the eye. At 0.4 ng/mL, high fold changes are likely due to extremely low fluorescence levels. These fluorescence levels are too low to be clearly distinguished from the background. At 2 ng/mL, fluorescence levels were high, but fold changes decreased. At 1 ng/mL, fluorescence levels were high enough that they could be clearly distinguished from the background, and fold changes were sufficiently high. Data is the average of three biological replicates. Error bars represent standard error of the mean (s.e.m.).

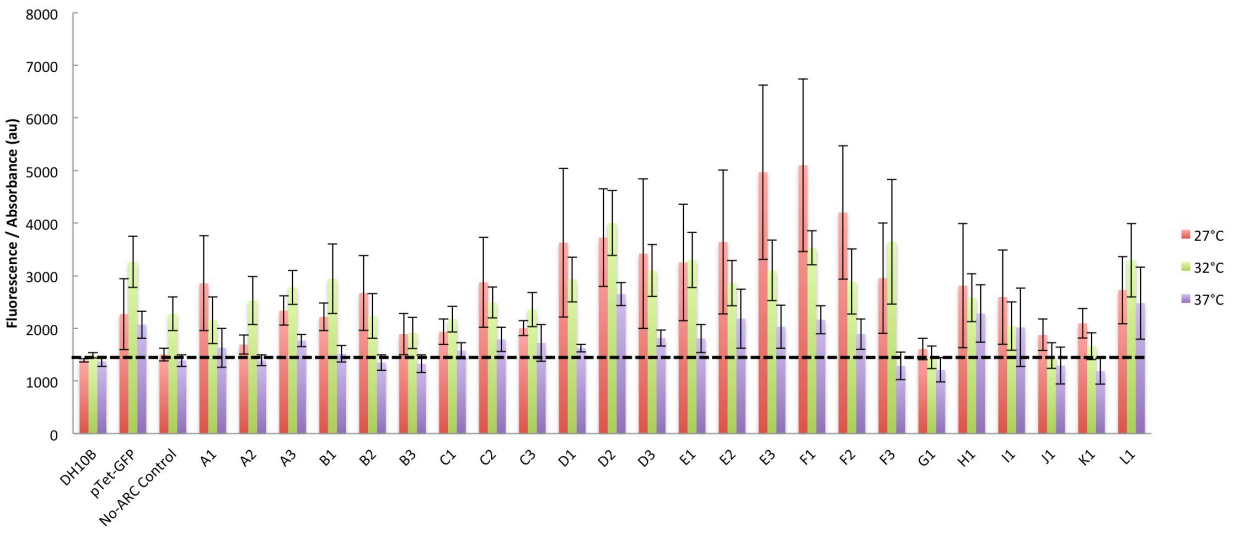

Figure 19: Thermosensor response in *E. coli* DH10B. Fluorescence is shown for all 24 thermosensors, as well as a variety of controls, at 27°C, 32°C, and 37°C. Fluorescence is divided by absorbance to give an approximate "per cell" expression level. The dashed line represents the autofluorescence level (divided by absorbance) measured in DH10B. Though in some cases expression appears to be higher at 32°C than at 27°C, slightly higher expression in the positive control (pTet-GFP) at 32°C can account for this trend. Normalized values show that for the majority of thermosensors, expression is highest at 27°C, and the expression level at 32°C is between expression levels at 27°C and 37°C. Data is the average of six biological replicates, over two different days. Error bars represent standard error of the mean (s.e.m.).

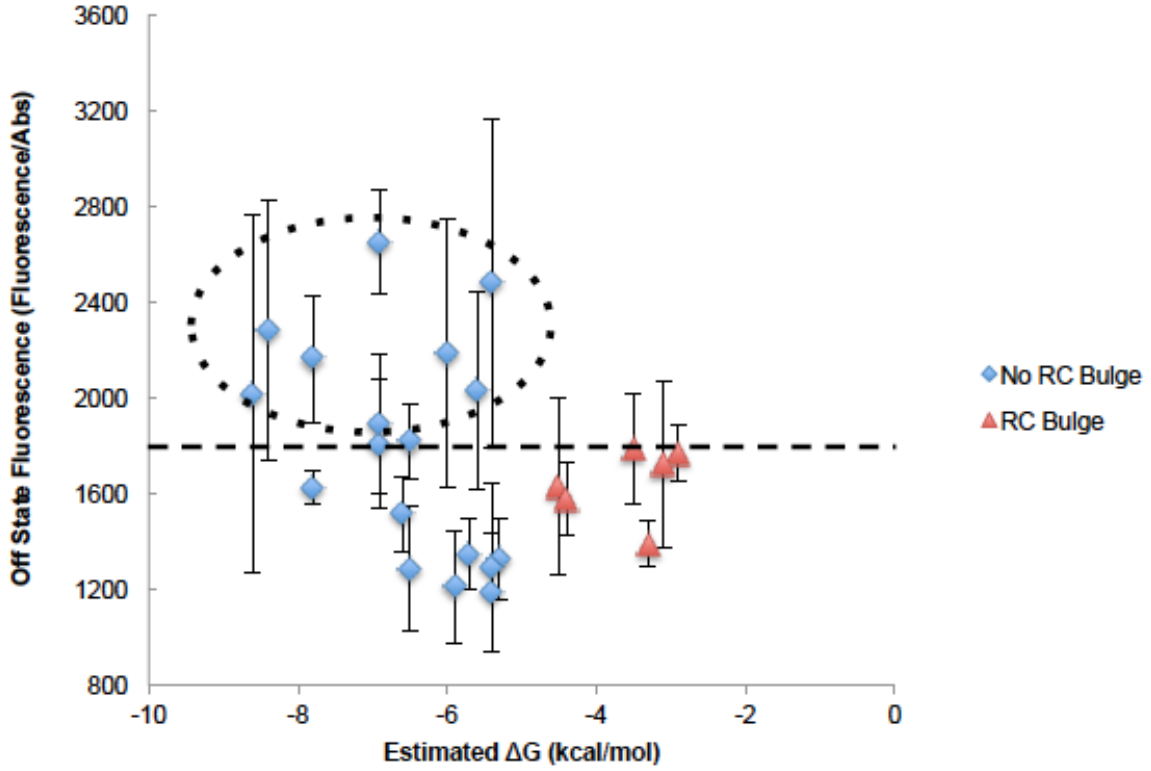

Figure 20: Correlation between "off" state fluorescence and estimated ΔG. "Off" state fluorescence is a measure of leakiness. The horizontal dashed line shows the threshold fluorescence, determined by DH10B autofluorescene divided by absorbance. Thermosensors with an "off" state fluorescence that exceeds the threshold are considered to be leaky. The threshold fluorescence is equal to one standard deviation above the average autofluorescence (devided by absorbance), as measured in the white cells (DH10B). The dotted circle indicates points representing the leaky thermosensors, which are all above the dashed line. Data is the average of six biological replicates, over two different days. Error bars represent standard error of the mean (s.e.m.).

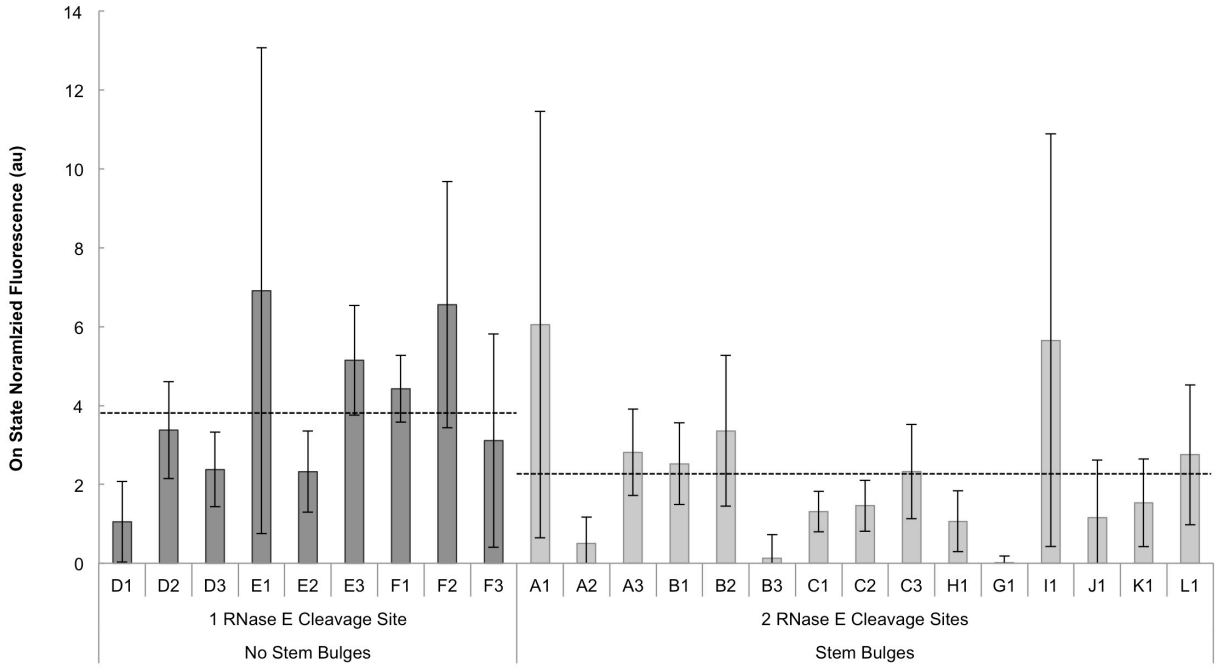

Figure 21: "On" state of thermosensors with 1 RC and no stem bulges, or 2 RCs and stem bulges. The dashed lines show the average "on" state for each group. Data is the average of six biological replicates, over two different days. Error bars represent standard error of the mean (s.e.m.).

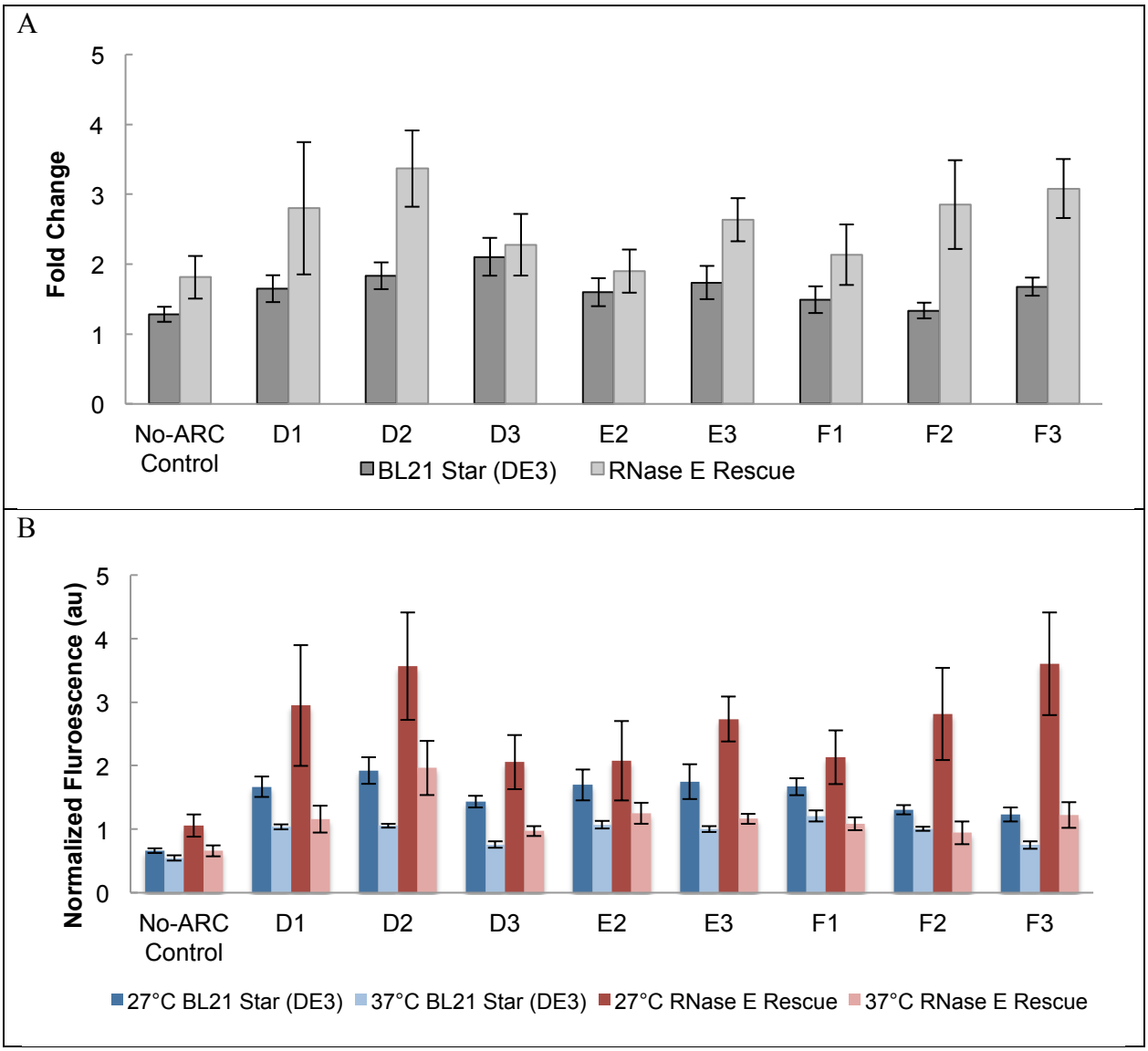

Figure 22: Fluorescence results for thermosensors in the BL21 Star (DE3) and RNase E Rescue Strains. (A) Fold change (27°C fluorescence / 37°C fluorescence) of selected thermosensors in BL21 Star (DE3) stain and RNase E Rescue strain. The D2, E3, F2, and F3 thermosensors demonstrate a significant increase in the 27°C/37°C expression ratio from the BL21 Star (DE3) strain to the RNase E rescue strain ($P<0.05$; two-tailed, unpaired, Student's t-test). P-values are as follows: No-ARC control = 0.13, D1 = 0.25, D2 = 0.02, D3 = 0.74, E2 = 0.42, E3 = 0.03, F1 = 0.19, F2 = 0.04, and F3 = 0.01. (B) Normalized fluorescence of selected thermosensors at 27°C and 37°C in BL21 Star (DE3) stain and RNase E Rescue strain. Fluorescence was normalized to pTet-GFP output under each condition in each strain. Raw fluorescence values were generally higher in the BL21 Star (DE3) strain than in the RNase E Rescue strain, although normalized values do not represent this trend due to differences in pTet-GFP output. Data is the average of 14 biological replicates, over a total of three different days. Error bars represent standard error of the mean (s.e.m.).

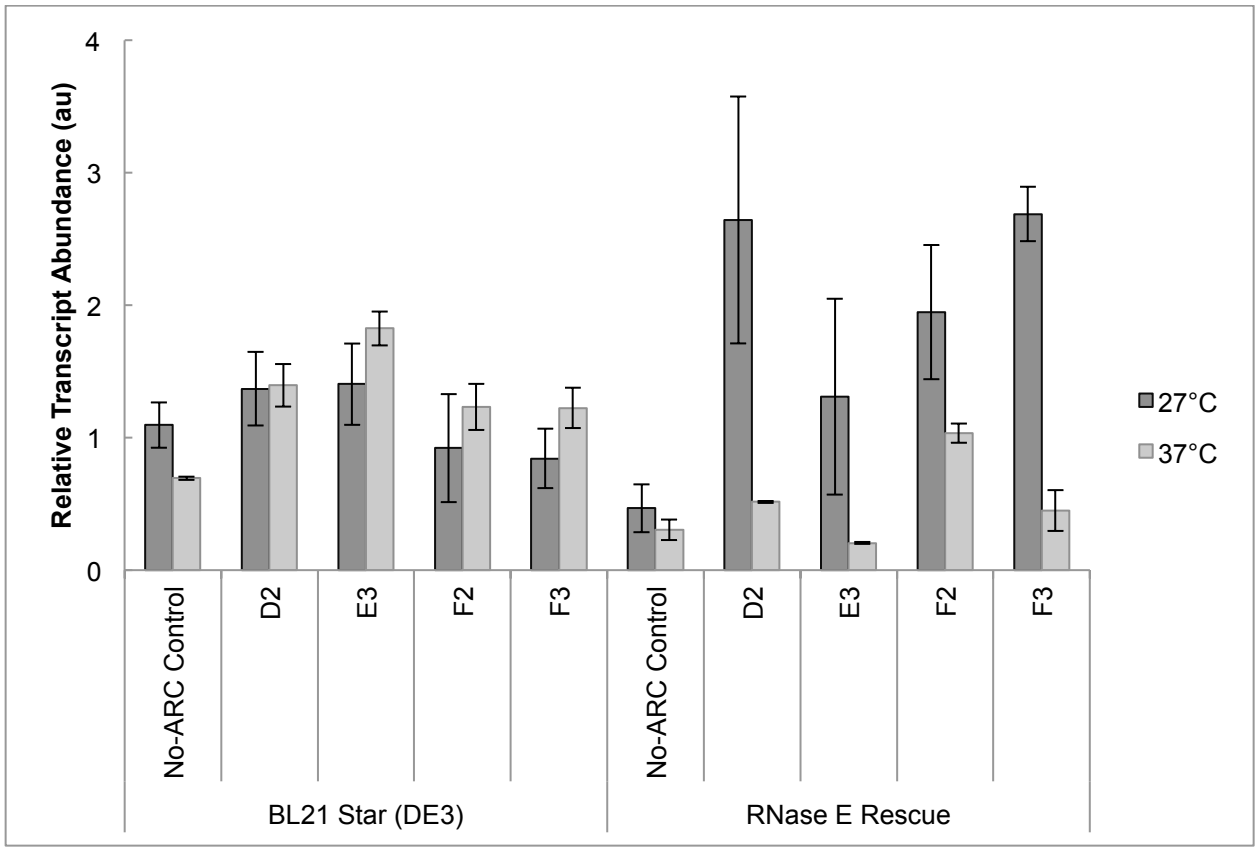

Figure 23: Relative transcript abundance of selected thermosensors at 27°C and 37°C in BL21 Star (DE3) stain and RNase E Rescue strain based on RT-qPCR data. Transcript abundances do not display major changes in response to temperature in the BL21 Star (DE3) strain, but show higher transcript abundance at 27°C than at 37°C in the RNase E Rescue strain. Data was normalized to the positive control (pTet-gfp) in that strain and at that temperature, and corrections were applied (log transformation, mean centering, and autoscaling) in accordance with MIQE guidelines [93, 94]. The data shown is from two biological and two technical replicates. Error bars represent standard error of the mean (s.e.m.).

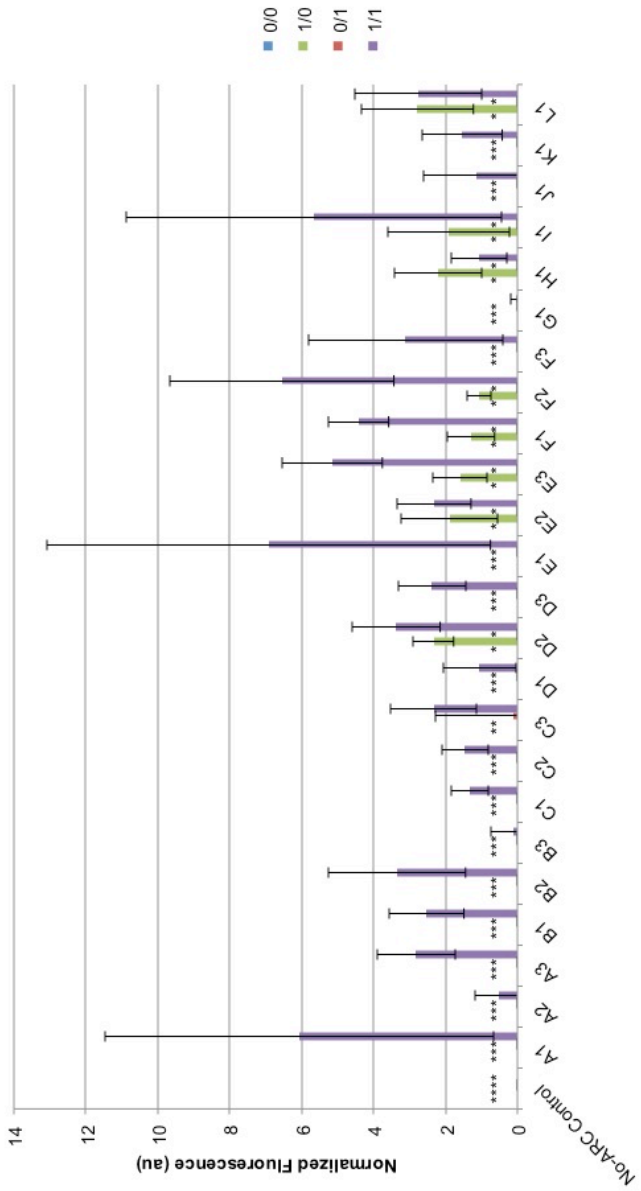

Figure 24: Two-input composite circuits with all 24 thermosensors. Conditions are reported as "aTc/Temperature". For temperature, "0" = 37°C and "1" = 27°C. aTc was used at a concentration of 1 ng/mL. Data is the average of six biological replicates, over two separate days. The asterisk (*) indicates that the GFP/Abs value was within one standard deviation of the background DH10B GFP/Abs value (Materials and Methods). Error bars represent standard error of the mean (s.e.m.).

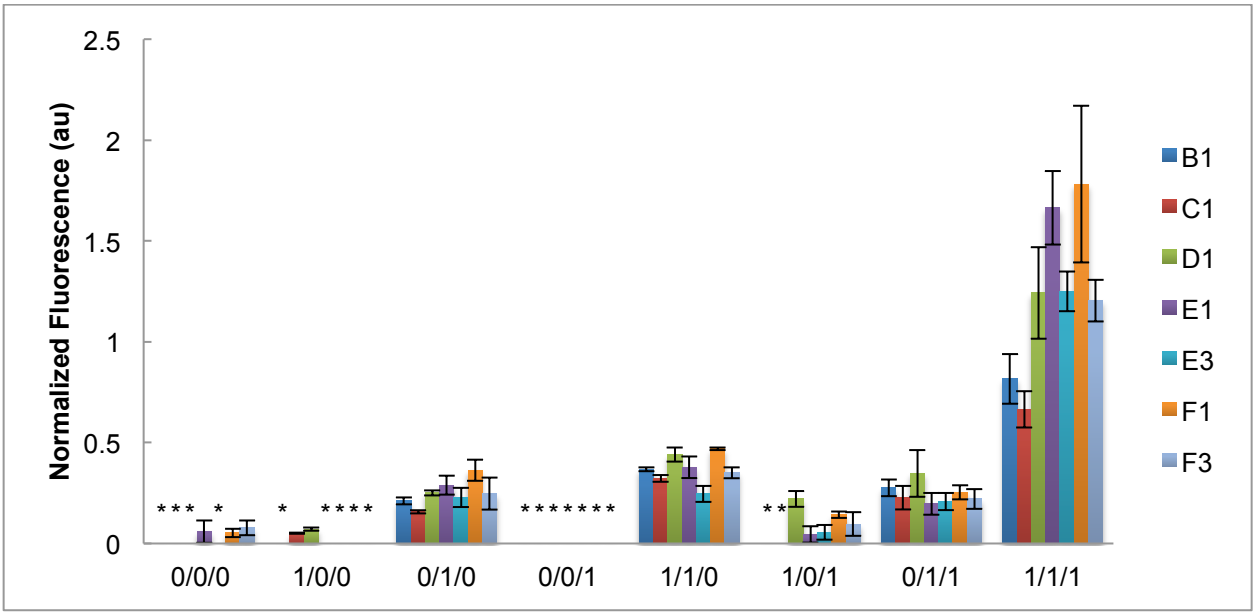

Figure 25: Three-input composite circuits with the B1, C1, D1, E1, E3, F1 and F3 thermosensors. Conditions are reported as "Ara/aTc/Temp". For temperature, "0" = 37°C and "1" = 27°C. aTc was used at a concentration of 2 ng/mL, and arabinose (Ara) was used at a concentration of 0.32 μM. Fold changes were calculated by dividing normalized expression in the [111] condition by that of the [110] condition, which was the leakiest condition in all cases. Fold changes are as follows: B1 = 2.2; C1 = 2.1; D1 = 2.8; E1 = 4.4; E3 = 5.1; F1 = 3.8; F3 = 3.4. P-values were found with a one-tailed, unpaired Student's t-test, which compared expression in the [111] condition to expression in the [110] condition. P-values are as follows: B1 = 0.03; C1 = 0.03; D1 = 0.03; E1 = 0.01; E3 = 0.002; F1 = 0.04; F3 = 0.005. The change in expression was found to be significant ($p<0.05$) for all seven circuits. Data is the average of three biological replicates, over two separate days. The asterisk (*) indicates that the GFP/Abs value was within one standard deviation of the background DH10B GFP/Abs value (Materials and Methods). Error bars represent standard error of the mean (s.e.m.).

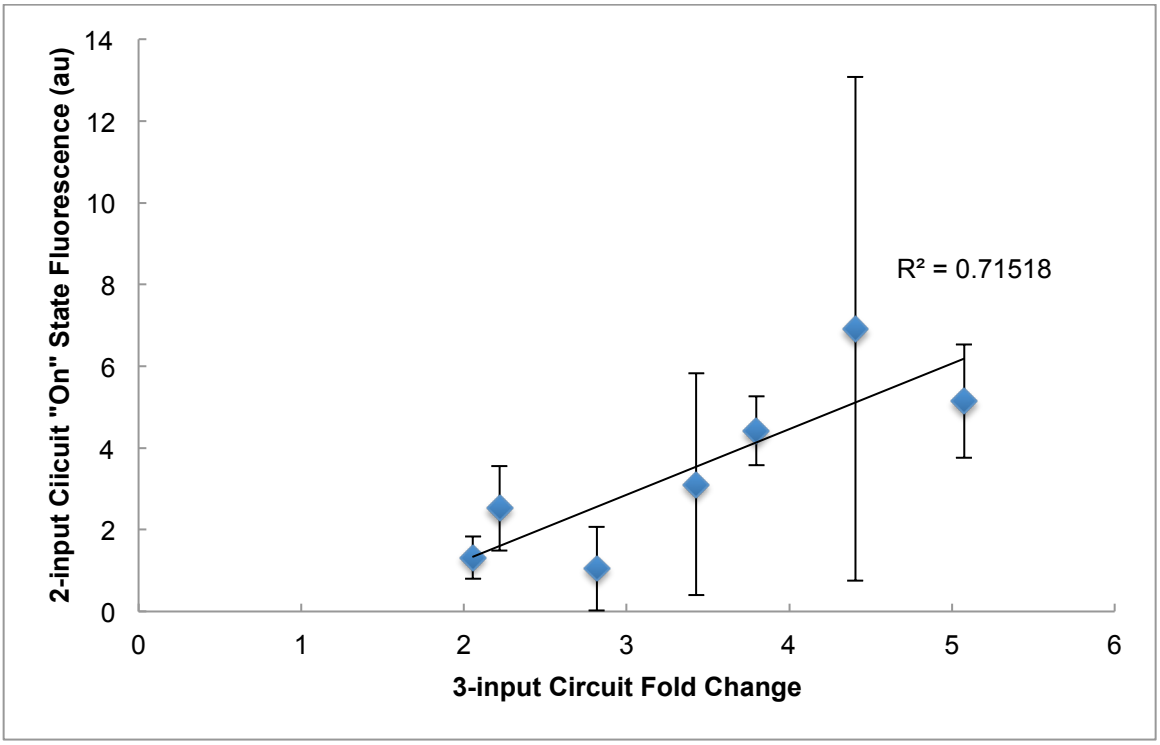

Figure 26: Correlation of 2-input circuit "on" state with 3-input circuit fold change. Results are shown for the B1, C1, D1, E1, E3, F1, and F3 thermosensors. Fold changes were calculated by dividing normalized expression in the [111] condition by expression in the [110] condition, which was the leakiest condition in all cases (see Figure 25). Error bars represent standard error of the mean (s.e.m.).

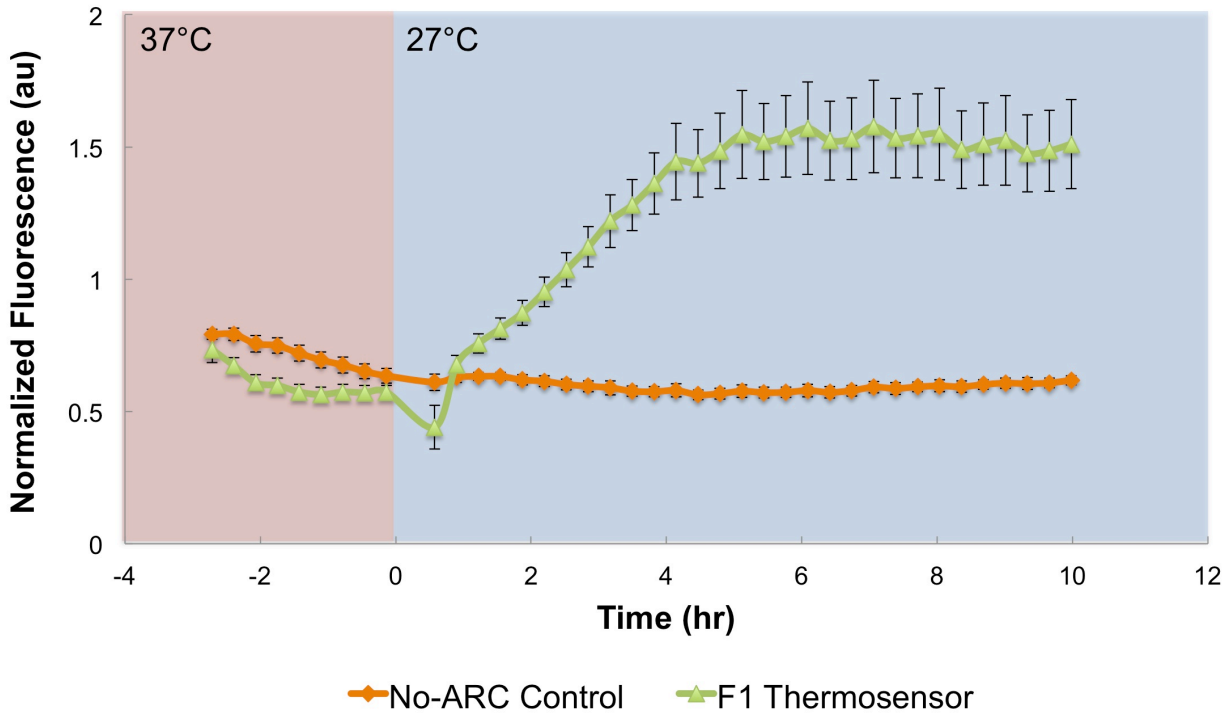

Figure 27: Time response of the RNA thermosensor F1 is shown. Cells were grown at 37°C for four hours, then the temperature was reduced to 27°C. The time at which the temperature change occurs is set as t=0. Upon exposure to a colder temperature, fluorescence for the F1 thermosensor increases relative to the fluorescence of the No-ARC control. Fluorescence is normalized to the positive control (pTet-GFP). These fluorescence values cannot be directly compared to the data shown in Figure 2 because the experimental conditions were different (e.g. temporal separation of chemical and temperature induction; see the Methods below). Data shown is the average of eight biological replicates, and error bars represent standard error of the mean (s.e.m.).

Supplementary Methods

RNA Extraction

RNA isolation was performed using TRIzol Reagent (Life Technologies), according to manufacturer's instructions with the following modifications. After resuspension with 1 mL of TRIzol, cells were incubated at 95°C for 5 minutes, then incubated on ice for 5 minutes before phase separation. After the addition of chloroform, tubes were vigorously shaken and incubated at room temperature for 10-15 minutes prior to centrifugation. After the addition of isopropanol to the aqueous phase following chloroform extraction, samples were incubated at room temperature for 10 minutes and centrifuged at 12,000g for 30 minutes at 4°C. The RNA wash was performed with ice cold ethanol, and after air-drying the RNA pellet, the RNA was resuspended in DPEC-treated water. Concentrations and purities (A260/A280) were measured with a Nanodrop 2000 UV-Vis Spectrophotometer. Purities (A_{260}/A_{280} values) ranged from 1.80 to 2.05, with an average value of 1.94.

DNase Treatment

DNA was removed using the DNA-free Kit (Life Technologies) according to manufacturer's instructions. To check for DNA contamination, PCR was completed with RT-qPCR primers (50 nM), and DNase-treated RNA samples as the template. GoTaq DNA Polymerase (Promega Corporation) was used according to manufacturer's instructions. PCR reaction volumes were 50 mL, and were held at 95°C for 2 minutes, then underwent 40 cycles of 45 seconds at 95°C, 45 seconds at 60°C, and 20 seconds at 72°C, followed by 5 minutes at 72°C. No bands were detected when visualized in a gel, indicating that DNA had been completely removed from the

sample. Concentrations and purities (A_{260}/A_{280}) were measured. Purities (A_{260}/A_{280} values) ranged from 1.67 to 2.02, with an average value of 1.91.

Reverse Transcription

The DNase-treated RNA samples were converted to cDNA libraries using the AffinityScript QPCR cDNA synthesis Kit (Agilent Technologies). To generate the cDNA library, 10 mL of 2X cDNA synthesis master mix, 3 mL random primers, 1 mL AffinityScript RT/RNase block enzyme mixture, and 6 mL of RNA (0.8 – 2.5 mg of RNA, depending on concentration) were combined for a total reaction volume of 20 mL. Samples were cycled at 25°C for 5 minutes, 42°C for 15 minutes, and 95°C for 5 minutes. The concentrations of the cDNA libraries ranged from 1400 – 3200 ng/mL. Samples were stored at -20°C for up to 2 days.

RT-qPCR Primer Optimization and Efficiency

Primer concentration was optimized by performing PCR with a gradient of primer concentrations and identifying the concentration at which no primer dimers or non-specific binding occurred. PCR for this optimization step was performed using 0.5 mL of GoTaq polymerase and either plasmid DNA (containing GFP), gDNA (containing the reference genes), or water (no-template control) as the template. PCR reaction volumes were 50 mL, and were held at 95°C for 2 minutes, then underwent 40 cycles of 45 seconds at 95°C, 45 seconds at 60°C, and 20 seconds at 72°C, followed by 5 minutes at 72°C. Optimal primer concentration was found to be 50 nM. No bands were observed in the no-template control.

Calibration curves were generated for each set of primers to determine primer efficiency in RT-qPCR (>90% for each primer set) and ensure that cDNA concentration was within the linear range for each gene target. RT-qPCR conditions are described in the Methods. 6 mg of cDNA

was used for both *idnT* and *hcaT*, 12 mg of cDNA was used for *cysG*, and 400 ng of cDNA was used for *gfp*. The concentrations of cDNA were different because reference genes expressed from the genome have a much lower copy number than *gfp*, which is expressed from a high-copy plasmid.

Characterization of Response Time

To determine the speed with which the thermosensor responds to a change in temperature, cells were prepared as described in the Methods. After resuspension in 1 ng/mL aTc-containing M9 minimal media with 4 g/L glucose, cells were cultured at 37°C and 250 rpm in the plate reader, with absorbance and fluorescence measurements taken every 15 minutes. After 4 hours, the temperature was decreased to 27°C, and measurements were continued until stationary phase was reached.

Appendix B: Supplementary Data for asRNA Design Rules

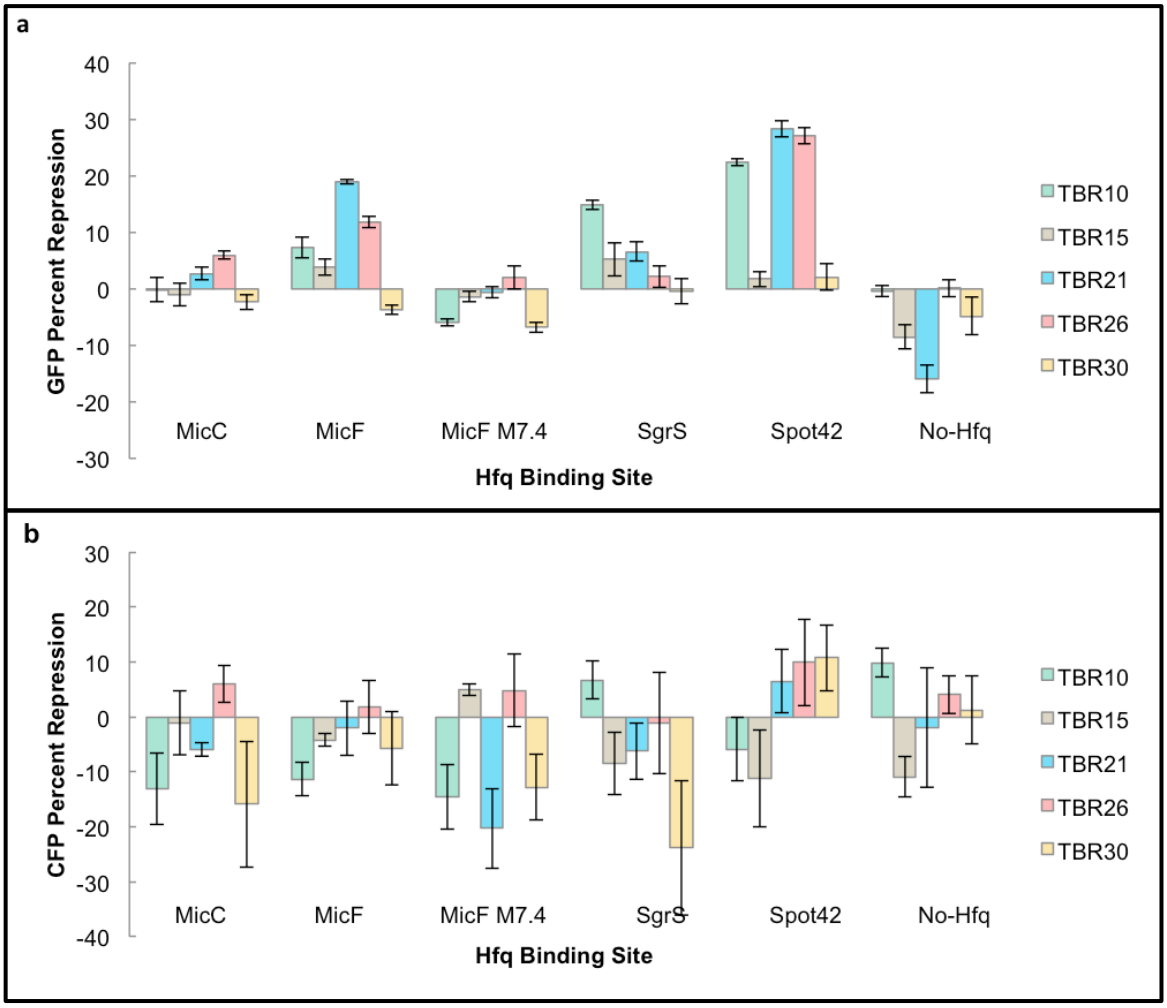

Figure 28: Off-target effects for each Hfq site. Five unique TBRs designed to target *rfp* were paired with five different Hfq binding sites, as well as a No-Hfq control, as shown in Figure 1. Off-target effects on both *gfp* expression (a) and *cfp* expression (b) are shown. Spot42 had high off-target repression for *gfp*, ruling out the use of this Hfq binding site for the remainder of the study. Experiments were conducted in triplicate, on two separate days, for a total of six replicates. Error bars represent standard error of the mean (SEM).

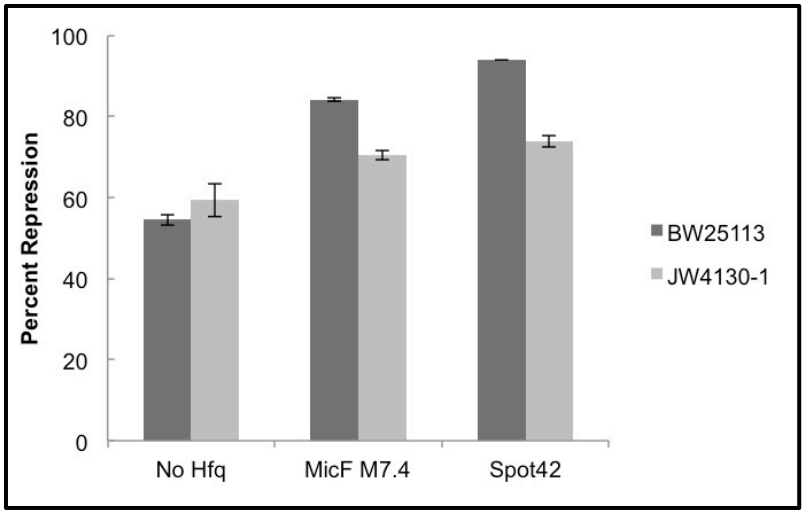

Figure 29: Repression efficiency in an Hfq deficient strain. To demonstrate that the Hfq protein plays a role in target gene repression, three sets of constructs were tested in an Hfq deficient strain (JW4130-1) and the corresponding strain containing the intact Hfq gene (BW25113). TBR10 was tested with no Hfq site, with MicF M7.4, and with Spot42. As expected, there was no change in repression efficiency in the "No Hfq" construct ($p=0.36$, two-tailed, unpaired Student's t-test). However, both TBR10-MicF M7.4 and TBR10-Spot42 showed a marked decrease in repression efficiency when expressed in JW4130-1 as compared to BW25113 (MicF M7.4: $p=0.004$, 16% reduction; Spot42: $p=0.0042$, 21% reduction; two-tailed, unpaired Student's t-test). Experiments were conducted in triplicate. Error bars represent standard error of the mean (SEM).

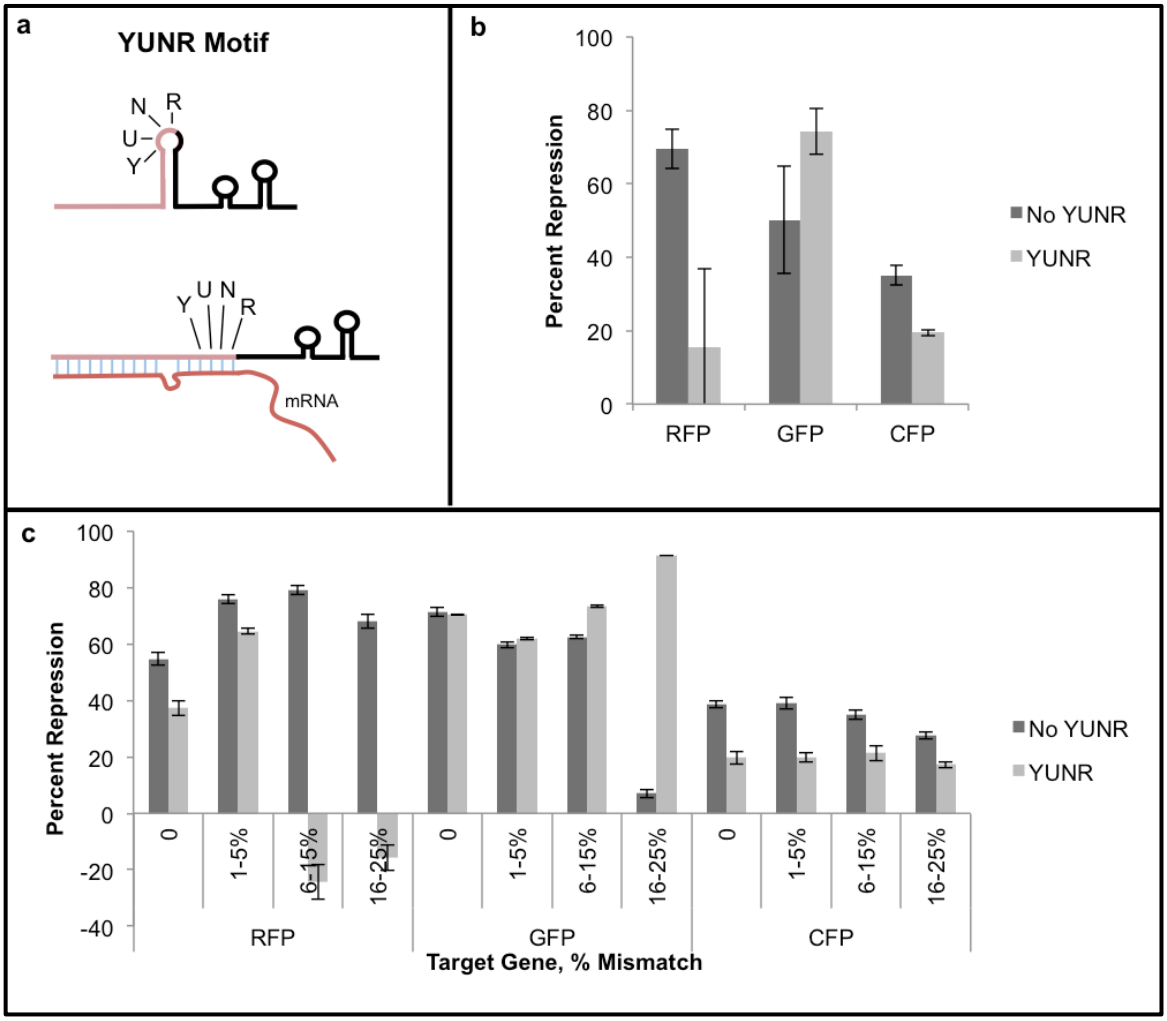

Figure 30: Effect of YUNR motif on repression efficiency. (a) A YUNR motif was included in a total of twelve constructs. The YUNR motif is thought to improve the kinetics of the interaction between the asRNA and the mRNA. To construct these TBRs, a YUNR motif was identified in the 3' end of the TBR, and a complementary sequence was introduced to cause the YUNR motif to fall in the loop of a hairpin. (b) For each target gene, four TBRs were designed with the YUNR motif, and four equivalent TBRs were designed that lacked a YUNR motif. The average repression that each of these four TBRs was able to achieve is shown. For *rfp* and *cfp*, the addition of a YUNR motif reduced the repression efficiency of the TBR. However, the addition of a YUNR motif in *gfp*-targeting TBRs improved repression efficiency. No clear trend is observable from this data. Experiments were conducted in triplicate, on two separate days, for a total of six replicates. Error bars represent standard error of the mean (SEM). (c) A comparison between otherwise equivalent TBRs with and without a YUNR motif. The TBR of each of these twelve constructs targeted location 3 (USD+SD+AUG, Figure 2b) with varying levels of

mismatch. Each of the twelve pairs of constructs targeted the same location, with the same levels of mismatch, and differed only regarding the presence (or absence) of the YUNR motif. For two *rfp*-targeting TBRs and all *cfp*-targeting TBRs, the addition of a YUNR motif led to a modest decrease in repression efficiency. For the remaining two *rfp*-targeting TBRs, there was a complete loss of repression efficiency upon addition of the YUNR motif. For three of the four *gfp*-targeting TBRs, there was no change or a modest increase in repression upon addition of the YUNR motif. For the final *gfp*-targeting TBR, the TBR was unable to function without the addition of a YUNR motif. This study did not identify the *in vivo* secondary structure of these constructs, and the secondary structure was simply predicted using Mfold. Thus, it is conceivable that in some cases the YUNR structure did not form as expected. This data could be clarified with further research into the *in vivo* secondary structures of these TBRs. Experiments were conducted in triplicate, on two separate days, for a total of six replicates. Error bars represent standard error of the mean (SEM).

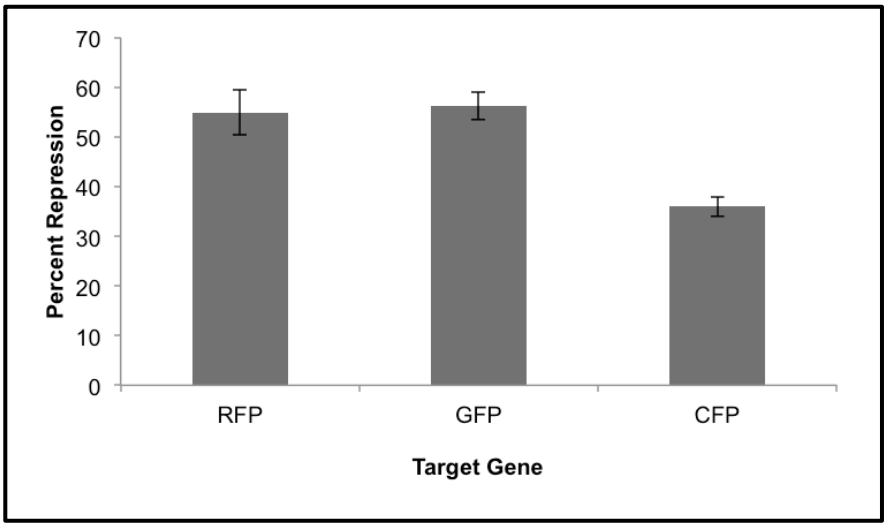

Figure 31: Average repression for each target gene. The average percent repression for each of the three genes is shown. This data includes all low-performing TBRs, as well as the TBRs containing a YUNR motif, for a total of 32 TBRs per gene. Though the average repression for *gfp* and *rfp* was relatively close, *cfp* had a lower average repression value. Because *cfp* repression was low, and not comparable to the repression of *gfp* and *rfp*, gene-normalized repression values were used when analyzing the overall data in Figure 3 and Figure 4. Experiments were conducted in triplicate, on two separate days, for a total of six replicates. Error bars represent standard error of the mean (SEM).

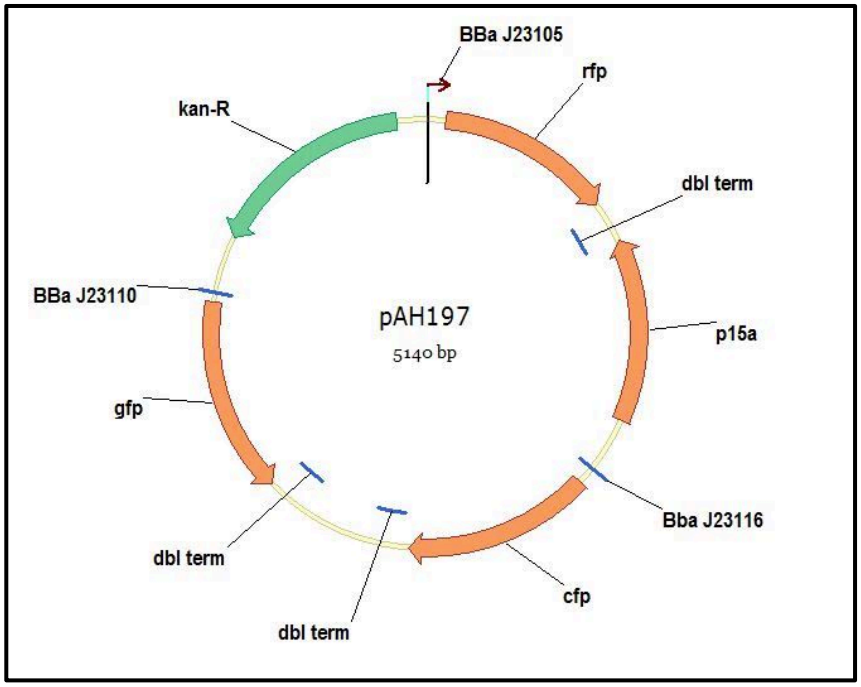

Figure 32: Plasmid map for pAH197, the target plasmid. Each of the 96 plasmids was co-expressed with the target plasmid, which contained all three target genes expressed constitutively (Bba_J23105-rfp, Bba_J23116-cfp, and Bba_J23110-gfp; http://parts.igem.org/Promoters/Catalog/Anderson).

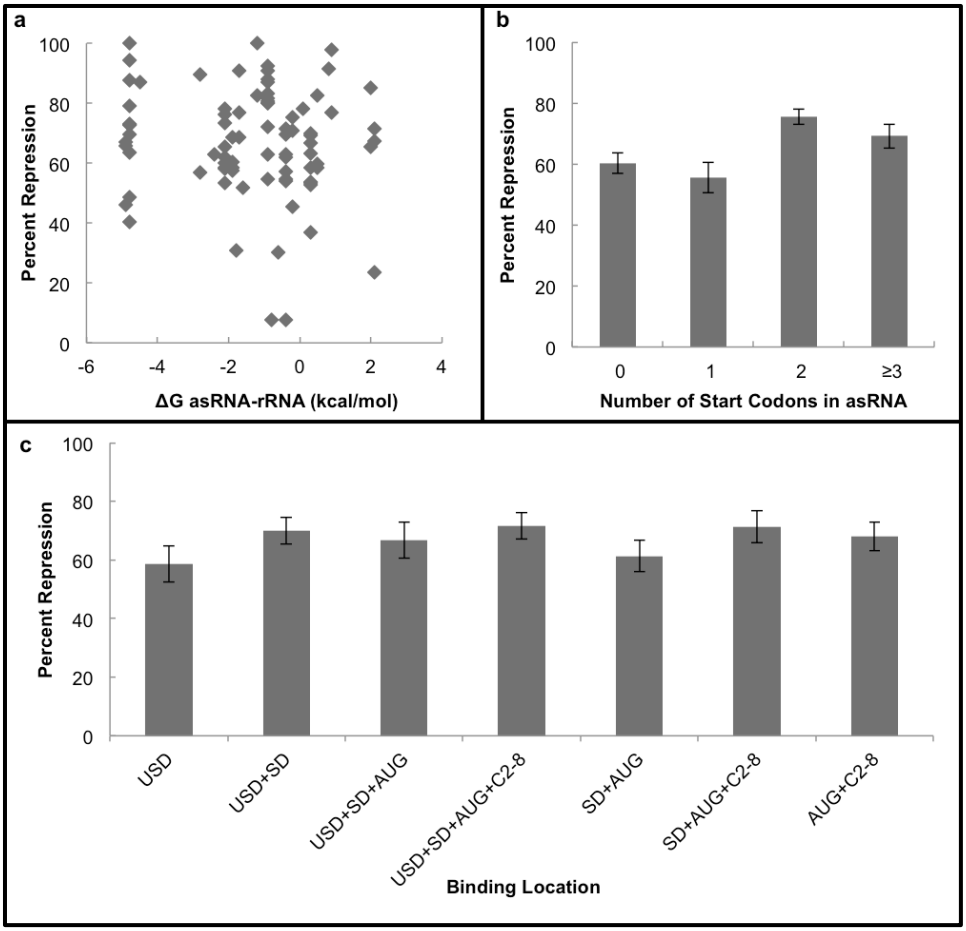

Figure 33: Effects of ribosome interactions and target location on gene repression. Two of the six design categories (ribosome interactions and target location) were found to have no impact on the repression efficiency of asRNA. Ribosome interactions were measured by the ΔG of the asRNA-rRNA complex and the number of start codons in the TBR. (a) Scatterplot showing the relationship between the ΔG of the complex formed by the asRNA and the anti-Shine-Dalgarno sequence and repression efficiency. No relationship is evident. Experiments were conducted in triplicate, on two separate days, for a total of six replicates. (b) Average percent repression of TBRs containing a varying number of start codons. Though a one-way analysis of variance (ANOVA) detected a significant increase in repression for TBRs containing two start codons ($p=2.67\times10^{-4}$), this does not appear to be part of a larger trend. Experiments were conducted in triplicate, on two separate days, for a total of six replicates. Error bars represent standard error of the mean (SEM). (c) TBRs were designed to target one of seven mRNA locations (Figure 2b). The binding location has no impact on the repression efficiency of the asRNA, as each location showed approximately equal percent repression (ANOVA, p=0.51). Experiments were conducted in triplicate, on two separate days, for a total of six replicates. Error bars represent standard error of the mean (SEM).

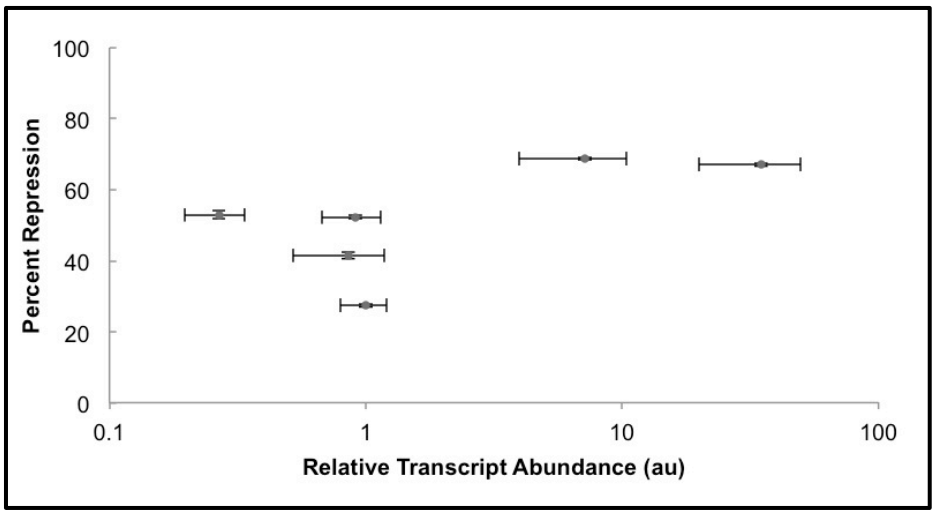

Figure 34: Relationship between asRNA transcript abundance and repression efficiency.
The cellular abundance of six asRNAs with varying repression levels (asRNA36, asRNA39, asRNA45, asRNA46, asRNA55, and asRNA60) was measured using RT-qPCR. There is no significant correlation using linear regression analysis ($p=0.23$), or using SRCC analysis ($p=0.4$). The data shown is from two biological and two technical replicates (total four replicates). Error bars represent standard error of the mean (SEM). au, arbitrary unit.

Table 6. SRCC and p-values for asRNA design parameters. Spearman Rank Correlation Coefficients (SRCC) and p-values for the asRNA design parameters included in Figure 3 are shown in the table below. Cells containing p-values that are less than 0.05 are highlighted in green.

		Overall	RFP Only	GFP Only	CFP Only
Length	TBR Length	SRCC =0.284 p-value =0.009	SRCC =0.468 p-value =0.012	SRCC =0.329 p-value =0.087	SRCC =0.190 p-value =0.332
	dsRNA Length	SRCC =0.360 p-value =0.001	SRCC =0.320 p-value =0.097	SRCC =0.486 p-value =0.009	SRCC =0.273 p-value =0.160
Thermodynamic Parameters	ΔG Complex	SRCC =-0.284 p-value =0.009	SRCC =-0.383 p-value =0.044	SRCC =-0.360 p-value =0.060	SRCC =-0.124 p-value =0.529
	ΔG CF	SRCC =-0.322 p-value =0.003	SRCC =-0.406 p-value =0.032	SRCC =-0.559 p-value =0.004	SRCC =-0.139 p-value =0.480
Mismatch	Number Mismatch	SRCC =-0.236 p-value =0.031	SRCC =-0.007 p-value =0.971	SRCC =-0.573 p-value =0.001	SRCC =-0.259 p-value =0.183
	Percent Mismatch	SRCC =-0.345 p-value =0.001	SRCC =-0.217 p-value =0.268	SRCC =-0.628 p-value =0.000	SRCC =-0.355 p-value =0.063

Table 7. Variance inflation factors for all asRNA design parameters found to have significant effects on asRNA repression capabilities.

		Length		Thermodynamic Parameters		Mismatch	
		TBR Length	dsRNA Length	ΔG Complex	ΔG CF	Number Mismatch	Percent Mismatch
Length	TBR Length		1.65	7.86	7.34	1.04	1.05
	dsRNA Length			2.34	2.53	1.25	1.68
Thermodynamic Parameters	ΔG Complex				38.48	1.00	1.21
	ΔG CF					1.00	1.24
Mismatch	Number Mismatch						3.25
	Percent Mismatch						

Table 8. Spearman Rank Correlation Coefficients for each parameter with respect to off-target repression. Values are listed as "SRCC, p-value". Location parameters (start coordinate and end coordinate) had the highest SRCCs, and they were the most consistently significant. This indicates that the target location is the most important parameter in determining orthogonality. Cells containing p-values that are less than 0.05 are highlighted in green.

	Data Set	Overall, No YUNR	RFP Only	GFP Only	CFP Only
Length	TBR Length	0.015, 0.892	-0.229, 0.242	0.484, 0.009	0.013, 0.947
	dsRNA Length	-0.025, 0.82	-0.009, 0.963	0.364, 0.057	-0.134, 0.497
Thermodynamic Parameters	ΔG Complex	0.012, 0.914	0.188, 0.338	-0.566, 0.002	0.107, 0.589
	ΔG CF	-0.076, 0.501	0.168, 0.393	-0.873, 0	0.083, 0.676
Mismatch	Number Mismatch	-0.044, 0.689	-0.326, 0.091	-0.053, 0.788	0.175, 0.372
	Percent Mismatch	-0.086, 0.435	-0.32, 0.097	-0.184, 0.348	0.092, 0.64
Ribosome Interactions	Number of start codons in TBR	-0.155, 0.158	0.067, 0.734	-0.031, 0.876	-0.169, 0.389
	ΔG asRNA-rRNA	0.056, 0.614	0.067, 0.736	0.063, 0.752	0.2, 0.307
Location	Start Coordinate	0.435, 0	-0.016, 0.937	0.277, 0.154	0.766, 0
	End Coordinate	0.328, 0.002	-0.237, 0.225	0.729, 0	0.693, 0

Table 9. Plasmids used in this study.

Name	Parts
	Hfq Binding Site Study
pAH197	p15a ori; kan-R; Bba_J23105-RFP, Bba_J23116-CFP, Bba_J23110-GFP
pAH332	p15a ori; amp-R; Bba_J23105-RFP, Bba_J23116-CFP, Bba_J23110-GFP
pAH158	ColE1 ori; cm-R; pTet-TBR10-MicC
pAH179	ColE1 ori; cm-R; pTet-TBR15-MicC
pAH180	ColE1 ori; cm-R; pTet-TBR21-MicC
pAH181	ColE1 ori; cm-R; pTet-TBR26-MicC
pAH182	ColE1 ori; cm-R; pTet-TBR30-MicC
pAH133	ColE1 ori; cm-R; pTet-TBR10-MicF
pAH168	ColE1 ori; cm-R; pTet-TBR15-MicF
pAH132	ColE1 ori; cm-R; pTet-TBR21-MicF
pAH169	ColE1 ori; cm-R; pTet-TBR26-MicF
pAH170	ColE1 ori; cm-R; pTet-TBR30-MicF
pAH155	ColE1 ori; cm-R; pTet-TBR10-MicF M7.4
pAH171	ColE1 ori; cm-R; pTet-TBR15-MicF M7.4
pAH172	ColE1 ori; cm-R; pTet-TBR21-MicF M7.4
pAH173	ColE1 ori; cm-R; pTet-TBR26-MicF M7.4
pAH174	ColE1 ori; cm-R; pTet-TBR30-MicF M7.4
pAH160	ColE1 ori; cm-R; pTet-TBR10-SgrS
pAH183	ColE1 ori; cm-R; pTet-TBR15-SgrS
pAH184	ColE1 ori; cm-R; pTet-TBR21-SgrS
pAH185	ColE1 ori; cm-R; pTet-TBR26-SgrS
pAH186	ColE1 ori; cm-R; pTet-TBR30-SgrS
pAH156	ColE1 ori; cm-R; pTet-TBR10-Spot42
pAH175	ColE1 ori; cm-R; pTet-TBR15-Spot42

pAH176	ColE1 ori; cm-R; pTet-TBR21-Spot42
pAH177	ColE1 ori; cm-R; pTet-TBR26-Spot42
pAH178	ColE1 ori; cm-R; pTet-TBR30-Spot42
pAH130	ColE1 ori; cm-R; pTet-TBR10
pAH164	ColE1 ori; cm-R; pTet-TBR15
pAH165	ColE1 ori; cm-R; pTet-TBR21
pAH166	ColE1 ori; cm-R; pTet-TBR26
pAH167	ColE1 ori; cm-R; pTet-TBR30
	TBR Study
pAH198	ColE1 ori; cm-R; pTet-TBR1-MicF M7.4
pAH199	ColE1 ori; cm-R; pTet-TBR2-MicF M7.4
pAH200	ColE1 ori; cm-R; pTet-TBR3-MicF M7.4
pAH201	ColE1 ori; cm-R; pTet-TBR4-MicF M7.4
pAH202	ColE1 ori; cm-R; pTet-TBR5-MicF M7.4
pAH203	ColE1 ori; cm-R; pTet-TBR6-MicF M7.4
pAH204	ColE1 ori; cm-R; pTet-TBR7-MicF M7.4
pAH205	ColE1 ori; cm-R; pTet-TBR8-MicF M7.4
pAH206	ColE1 ori; cm-R; pTet-TBR9-MicF M7.4
pAH207	ColE1 ori; cm-R; pTet-TBR11-MicF M7.4
pAH208	ColE1 ori; cm-R; pTet-TBR12-MicF M7.4
pAH209	ColE1 ori; cm-R; pTet-TBR13-MicF M7.4
pAH210	ColE1 ori; cm-R; pTet-TBR14-MicF M7.4
pAH211	ColE1 ori; cm-R; pTet-TBR16-MicF M7.4
pAH212	ColE1 ori; cm-R; pTet-TBR17-MicF M7.4
pAH213	ColE1 ori; cm-R; pTet-TBR18-MicF M7.4
pAH214	ColE1 ori; cm-R; pTet-TBR19-MicF M7.4
pAH215	ColE1 ori; cm-R; pTet-TBR20-MicF M7.4

pAH216	ColE1 ori; cm-R; pTet-TBR22-MicF M7.4
pAH217	ColE1 ori; cm-R; pTet-TBR23-MicF M7.4
pAH218	ColE1 ori; cm-R; pTet-TBR24-MicF M7.4
pAH219	ColE1 ori; cm-R; pTet-TBR25-MicF M7.4
pAH220	ColE1 ori; cm-R; pTet-TBR27-MicF M7.4
pAH221	ColE1 ori; cm-R; pTet-TBR28-MicF M7.4
pAH222	ColE1 ori; cm-R; pTet-TBR29-MicF M7.4
pAH223	ColE1 ori; cm-R; pTet-TBR31-MicF M7.4
pAH224	ColE1 ori; cm-R; pTet-TBR32-MicF M7.4
pAH225	ColE1 ori; cm-R; pTet-TBR33-MicF M7.4
pAH226	ColE1 ori; cm-R; pTet-TBR34-MicF M7.4
pAH227	ColE1 ori; cm-R; pTet-TBR35-MicF M7.4
pAH228	ColE1 ori; cm-R; pTet-TBR36-MicF M7.4
pAH229	ColE1 ori; cm-R; pTet-TBR37-MicF M7.4
pAH230	ColE1 ori; cm-R; pTet-TBR38-MicF M7.4
pAH231	ColE1 ori; cm-R; pTet-TBR39-MicF M7.4
pAH232	ColE1 ori; cm-R; pTet-TBR40-MicF M7.4
pAH233	ColE1 ori; cm-R; pTet-TBR41-MicF M7.4
pAH234	ColE1 ori; cm-R; pTet-TBR42-MicF M7.4
pAH235	ColE1 ori; cm-R; pTet-TBR43-MicF M7.4
pAH236	ColE1 ori; cm-R; pTet-TBR44-MicF M7.4
pAH237	ColE1 ori; cm-R; pTet-TBR45-MicF M7.4
pAH238	ColE1 ori; cm-R; pTet-TBR46-MicF M7.4
pAH239	ColE1 ori; cm-R; pTet-TBR47-MicF M7.4
pAH240	ColE1 ori; cm-R; pTet-TBR48-MicF M7.4
pAH241	ColE1 ori; cm-R; pTet-TBR49-MicF M7.4
pAH242	ColE1 ori; cm-R; pTet-TBR50-MicF M7.4

pAH243	ColE1 ori; cm-R; pTet-TBR51-MicF M7.4
pAH244	ColE1 ori; cm-R; pTet-TBR52-MicF M7.4
pAH245	ColE1 ori; cm-R; pTet-TBR53-MicF M7.4
pAH246	ColE1 ori; cm-R; pTet-TBR54-MicF M7.4
pAH247	ColE1 ori; cm-R; pTet-TBR55-MicF M7.4
pAH248	ColE1 ori; cm-R; pTet-TBR56-MicF M7.4
pAH249	ColE1 ori; cm-R; pTet-TBR57-MicF M7.4
pAH250	ColE1 ori; cm-R; pTet-TBR58-MicF M7.4
pAH251	ColE1 ori; cm-R; pTet-TBR59-MicF M7.4
pAH252	ColE1 ori; cm-R; pTet-TBR60-MicF M7.4
pAH253	ColE1 ori; cm-R; pTet-TBR61-MicF M7.4
pAH254	ColE1 ori; cm-R; pTet-TBR62-MicF M7.4
pAH255	ColE1 ori; cm-R; pTet-TBR63-MicF M7.4
pAH256	ColE1 ori; cm-R; pTet-TBR64-MicF M7.4
pAH257	ColE1 ori; cm-R; pTet-TBR65-MicF M7.4
pAH258	ColE1 ori; cm-R; pTet-TBR66-MicF M7.4
pAH259	ColE1 ori; cm-R; pTet-TBR67-MicF M7.4
pAH260	ColE1 ori; cm-R; pTet-TBR68-MicF M7.4
pAH261	ColE1 ori; cm-R; pTet-TBR69-MicF M7.4
pAH262	ColE1 ori; cm-R; pTet-TBR70-MicF M7.4
pAH263	ColE1 ori; cm-R; pTet-TBR71-MicF M7.4
pAH264	ColE1 ori; cm-R; pTet-TBR72-MicF M7.4
pAH265	ColE1 ori; cm-R; pTet-TBR73-MicF M7.4
pAH266	ColE1 ori; cm-R; pTet-TBR74-MicF M7.4
pAH267	ColE1 ori; cm-R; pTet-TBR75-MicF M7.4
pAH268	ColE1 ori; cm-R; pTet-TBR76-MicF M7.4
pAH269	ColE1 ori; cm-R; pTet-TBR77-MicF M7.4

pAH270	ColE1 ori; cm-R; pTet-TBR78-MicF M7.4
pAH271	ColE1 ori; cm-R; pTet-TBR79-MicF M7.4
pAH272	ColE1 ori; cm-R; pTet-TBR80-MicF M7.4
pAH273	ColE1 ori; cm-R; pTet-TBR81-MicF M7.4
pAH274	ColE1 ori; cm-R; pTet-TBR82-MicF M7.4
pAH275	ColE1 ori; cm-R; pTet-TBR83-MicF M7.4
pAH276	ColE1 ori; cm-R; pTet-TBR84-MicF M7.4
pAH277	ColE1 ori; cm-R; pTet-TBR85-MicF M7.4
pAH278	ColE1 ori; cm-R; pTet-TBR86-MicF M7.4
pAH279	ColE1 ori; cm-R; pTet-TBR87-MicF M7.4
pAH280	ColE1 ori; cm-R; pTet-TBR88-MicF M7.4
pAH281	ColE1 ori; cm-R; pTet-TBR89-MicF M7.4
pAH282	ColE1 ori; cm-R; pTet-TBR90-MicF M7.4
pAH283	ColE1 ori; cm-R; pTet-TBR91-MicF M7.4
pAH284	ColE1 ori; cm-R; pTet-TBR92-MicF M7.4
pAH285	ColE1 ori; cm-R; pTet-TBR93-MicF M7.4
pAH286	ColE1 ori; cm-R; pTet-TBR94-MicF M7.4
pAH287	ColE1 ori; cm-R; pTet-TBR95-MicF M7.4
pAH288	ColE1 ori; cm-R; pTet-TBR96-MicF M7.4
	Genetic Circuit Plasmids
pAH333	ColE1 ori; cm-R; pTet- L2-TI-MicF M7.4
pAH334	ColE1 ori; cm-R; pTet- L2-TII-MicF M7.4
pAH335	ColE1 ori; cm-R; pTet- L2-TIII-MicF M7.4
pAH339	ColE1 ori; cm-R; pTet- L6-TI-MicF M7.4
pAH340	ColE1 ori; cm-R; pTet- L6-TII-MicF M7.4
pAH341	ColE1 ori; cm-R; pTet- L6-TIII-MicF M7.4
pTS001	p15a ori; amp-R; pBad-exsA

pAH148	p15a ori; amp-R; pLux-exsD, pBad-exsA
pAH290	ColE1 ori; cm-R; pTet-asExsD-W
pAH146	ColE1 ori; cm-R; pTet-asExsD-S
pTS118	psc101* ori; kan-R; pExsD-GFP

Table 10: *E. coli* strains used in this study.

Name	Host Strain	Plasmids
	Hfq Binding Site Study	
TBR10-MicC	DH10B	pAH197 + pAH158
TBR15-MicC	DH10B	pAH197 + pAH179
TBR21-MicC	DH10B	pAH197 + pAH180
TBR26-MicC	DH10B	pAH197 + pAH181
TBR30-MicC	DH10B	pAH197 + pAH182
TBR10-MicF	DH10B	pAH197 + pAH133
TBR15-MicF	DH10B	pAH197 + pAH168
TBR21-MicF	DH10B	pAH197 + pAH132
TBR26-MicF	DH10B	pAH197 + pAH169
TBR30-MicF	DH10B	pAH197 + pAH170
TBR10-MicF M7.4	DH10B	pAH197 + pAH155
TBR15-MicF M7.4	DH10B	pAH197 + pAH171
TBR21-MicF M7.4	DH10B	pAH197 + pAH172
TBR26-MicF M7.4	DH10B	pAH197 + pAH173
TBR30-MicF M7.4	DH10B	pAH197 + pAH174
TBR10-SgrS	DH10B	pAH197 + pAH160
TBR15-SgrS	DH10B	pAH197 + pAH183
TBR21-SgrS	DH10B	pAH197 + pAH184

TBR26-SgrS	DH10B	pAH197 + pAH185
TBR30-SgrS	DH10B	pAH197 + pAH186
TBR10-Spot42	DH10B	pAH197 + pAH156
TBR15-Spot42	DH10B	pAH197 + pAH175
TBR21-Spot42	DH10B	pAH197 + pAH176
TBR26-Spot42	DH10B	pAH197 + pAH177
TBR30-Spot42	DH10B	pAH197 + pAH178
TBR10-No-Hfq	DH10B	pAH197 + pAH130
TBR15-No-Hfq	DH10B	pAH197 + pAH164
TBR21-No-Hfq	DH10B	pAH197 + pAH165
TBR26-No-Hfq	DH10B	pAH197 + pAH166
TBR30-No-Hfq	DH10B	pAH197 + pAH167
TBR10-No-Hfq	BW25113	pAH332 + pAH130
TBR10-MicF M7.4	BW25113	pAH332 + pAH155
TBR10-Spot42	BW25113	pAH332 + pAH156
TBR10-No-Hfq	JW4130-1	pAH332 + pAH130
TBR10-MicF M7.4	JW4130-1	pAH332 + pAH155
TBR10-Spot42	JW4130-1	pAH332 + pAH156
TBR Study		
TBR1-MicF M7.4	DH10B	pAH197 + pAH198
TBR2-MicF M7.4	DH10B	pAH197 + pAH199
TBR3-MicF M7.4	DH10B	pAH197 + pAH200
TBR4-MicF M7.4	DH10B	pAH197 + pAH201
TBR5-MicF M7.4	DH10B	pAH197 + pAH202
TBR6-MicF M7.4	DH10B	pAH197 + pAH203
TBR7-MicF M7.4	DH10B	pAH197 + pAH204
TBR8-MicF M7.4	DH10B	pAH197 + pAH205

TBR9-MicF M7.4	DH10B	pAH197 + pAH206
TBR11-MicF M7.4	DH10B	pAH197 + pAH207
TBR12-MicF M7.4	DH10B	pAH197 + pAH208
TBR13-MicF M7.4	DH10B	pAH197 + pAH209
TBR14-MicF M7.4	DH10B	pAH197 + pAH210
TBR16-MicF M7.4	DH10B	pAH197 + pAH211
TBR17-MicF M7.4	DH10B	pAH197 + pAH212
TBR18-MicF M7.4	DH10B	pAH197 + pAH213
TBR19-MicF M7.4	DH10B	pAH197 + pAH214
TBR20-MicF M7.4	DH10B	pAH197 + pAH215
TBR22-MicF M7.4	DH10B	pAH197 + pAH216
TBR23-MicF M7.4	DH10B	pAH197 + pAH217
TBR24-MicF M7.4	DH10B	pAH197 + pAH218
TBR25-MicF M7.4	DH10B	pAH197 + pAH219
TBR27-MicF M7.4	DH10B	pAH197 + pAH220
TBR28-MicF M7.4	DH10B	pAH197 + pAH221
TBR29-MicF M7.4	DH10B	pAH197 + pAH222
TBR31-MicF M7.4	DH10B	pAH197 + pAH223
TBR32-MicF M7.4	DH10B	pAH197 + pAH224
TBR33-MicF M7.4	DH10B	pAH197 + pAH225
TBR34-MicF M7.4	DH10B	pAH197 + pAH226
TBR35-MicF M7.4	DH10B	pAH197 + pAH227
TBR36-MicF M7.4	DH10B	pAH197 + pAH228
TBR37-MicF M7.4	DH10B	pAH197 + pAH229
TBR38-MicF M7.4	DH10B	pAH197 + pAH230
TBR39-MicF M7.4	DH10B	pAH197 + pAH231
TBR40-MicF M7.4	DH10B	pAH197 + pAH232

TBR41-MicF M7.4	DH10B	pAH197 + pAH233
TBR42-MicF M7.4	DH10B	pAH197 + pAH234
TBR43-MicF M7.4	DH10B	pAH197 + pAH235
TBR44-MicF M7.4	DH10B	pAH197 + pAH236
TBR45-MicF M7.4	DH10B	pAH197 + pAH237
TBR46-MicF M7.4	DH10B	pAH197 + pAH238
TBR47-MicF M7.4	DH10B	pAH197 + pAH239
TBR48-MicF M7.4	DH10B	pAH197 + pAH240
TBR49-MicF M7.4	DH10B	pAH197 + pAH241
TBR50-MicF M7.4	DH10B	pAH197 + pAH242
TBR51-MicF M7.4	DH10B	pAH197 + pAH243
TBR52-MicF M7.4	DH10B	pAH197 + pAH244
TBR53-MicF M7.4	DH10B	pAH197 + pAH245
TBR54-MicF M7.4	DH10B	pAH197 + pAH246
TBR55-MicF M7.4	DH10B	pAH197 + pAH247
TBR56-MicF M7.4	DH10B	pAH197 + pAH248
TBR57-MicF M7.4	DH10B	pAH197 + pAH249
TBR58-MicF M7.4	DH10B	pAH197 + pAH250
TBR59-MicF M7.4	DH10B	pAH197 + pAH251
TBR60-MicF M7.4	DH10B	pAH197 + pAH252
TBR61-MicF M7.4	DH10B	pAH197 + pAH253
TBR62-MicF M7.4	DH10B	pAH197 + pAH254
TBR63-MicF M7.4	DH10B	pAH197 + pAH255
TBR64-MicF M7.4	DH10B	pAH197 + pAH256
TBR65-MicF M7.4	DH10B	pAH197 + pAH257
TBR66-MicF M7.4	DH10B	pAH197 + pAH258
TBR67-MicF M7.4	DH10B	pAH197 + pAH259

TBR68-MicF M7.4	DH10B	pAH197 + pAH260
TBR69-MicF M7.4	DH10B	pAH197 + pAH261
TBR70-MicF M7.4	DH10B	pAH197 + pAH262
TBR71-MicF M7.4	DH10B	pAH197 + pAH263
TBR72-MicF M7.4	DH10B	pAH197 + pAH264
TBR73-MicF M7.4	DH10B	pAH197 + pAH265
TBR74-MicF M7.4	DH10B	pAH197 + pAH266
TBR75-MicF M7.4	DH10B	pAH197 + pAH267
TBR76-MicF M7.4	DH10B	pAH197 + pAH268
TBR77-MicF M7.4	DH10B	pAH197 + pAH269
TBR78-MicF M7.4	DH10B	pAH197 + pAH270
TBR79-MicF M7.4	DH10B	pAH197 + pAH271
TBR80-MicF M7.4	DH10B	pAH197 + pAH272
TBR81-MicF M7.4	DH10B	pAH197 + pAH273
TBR82-MicF M7.4	DH10B	pAH197 + pAH274
TBR83-MicF M7.4	DH10B	pAH197 + pAH275
TBR84-MicF M7.4	DH10B	pAH197 + pAH276
TBR85-MicF M7.4	DH10B	pAH197 + pAH277
TBR86-MicF M7.4	DH10B	pAH197 + pAH278
TBR87-MicF M7.4	DH10B	pAH197 + pAH279
TBR88-MicF M7.4	DH10B	pAH197 + pAH280
TBR89-MicF M7.4	DH10B	pAH197 + pAH281
TBR90-MicF M7.4	DH10B	pAH197 + pAH282
TBR91-MicF M7.4	DH10B	pAH197 + pAH283
TBR92-MicF M7.4	DH10B	pAH197 + pAH284
TBR93-MicF M7.4	DH10B	pAH197 + pAH285
TBR94-MicF M7.4	DH10B	pAH197 + pAH286

TBR95-MicF M7.4	DH10B	pAH197 + pAH287
TBR96-MicF M7.4	DH10B	pAH197 + pAH288
Genetic Circuits		
Location 2 Type I Circuit	DH10B	pAH333 + pTS001 + pTS118
Location 2 Type II Circuit	DH10B	pAH334 + pTS001 + pTS118
Location 2 Type III Circuit	DH10B	pAH335 + pTS001 + pTS118
Location 6 Type I Circuit	DH10B	pAH339 + pTS001 + pTS118
Location 6 Type II Circuit	DH10B	pAH340 + pTS001 + pTS118
Location 6 Type III Circuit	DH10B	pAH341 + pTS001 + pTS118
asExsD-S Circuit	DH10B	pAH146 + pAH148 + pTS118
asExsD-W Circuit	DH10B	pAH290 + pAH148 + pTS118
RT-qPCR		
asRNA36	DH10B	pAH228
asRNA39	DH10B	pAH231
asRNA45	DH10B	pAH237
asRNA46	DH10B	pAH238
asRNA55	DH10B	pAH247
asRNA60	DH10B	pAH252

Table 11: Genetic parts used in this study. For asExsA variants, L=Location and T=Type; for example, L2-TI = Location 2 and Type I.

Part name	Type and source	DNA sequence
gfp	Gene[198]	atgagtaaaggagaagaacttttcactggagttgtcccaattcttgttgaattagatggtgatgttaatgggcac aaattttctgtcagtggagagggtgaaggtgatgcaacatacggaaaacttacccttaaatttatttgcactact ggaaaactacctgttccatggccaacacttgtcactactttgacttatggtgttcaatgcttttcaagatacccag atcatatgaaacggcatgactttttcaagagtgccatgcccgaaggttatgtacaggaaagaactatatttttca aagatgacgggaactataagacacgtgctgaagtcaagtttgaaggtgatacacttgttaatagaatcgagtta aaaggtattgattttaaagaagatggaaacattcttggacacaagttggaatacaactataactcacacaatgta tacatcatggcagacaaacaaaagaatggaatcaaagttaacttcaaaattagacacaacattgaagatggaa gcgttcaactagcagaccattatcaacaaaatactccaattggcgatggccctgtccttttaccagacaaccatt acctgtccacacaatctgccctttcgaaagatcccaacgaaaagagagaccacatggtccttcttgagtttgta acagctgctgggattacacatggcatggatgaactatacaaaaggcctgcagcaaacgacgaaaactacgc ttaagtagcttaa
rfp	Gene[199]	atggcgagtagcgaagacgttatcaaagagttcatgcgtttcaaagttcgtatggaaggttccgttaacggtca cgagttcgaaatcgaaggtgaaggtgaaggtcgtccgtacgaaggtacccagaccgctaaactgaaagtta ccaaaggtggtccgctgccgttcgcttgggacatcctgtccccgcagttccagtacggttccaaagcttacgtt aaacacccggctgacatcccggactacctgaaactgtccttcccggaaggtttcaaatgggaacgtgttatga acttcgaagacggtggtgttgttaccgttacccaggactcctccctgcaagacggtgagttcatctacaaagtt aaactgcgtggtaccaacttcccgtccgacggtccggttatgcagaaaaaaaccatgggttgggaagcttcc accgaacgtatgtacccggaagacggtgctctgaaaggtgaaatcaaaatgcgtctgaaactgaaagacgg tggtcactacgacgctgaagttaaaaccacctacatggctaaaaaaccggttcagctgccgggtgcttacaa aaccgacatcaaactggacatcacctcccacaacgaagactacaccatcgttgaacagtacgaacgtgctg aaggtcgtcactccaccggtgcttaa
cfp	Gene[201]	atgactagcaaaagaagcaaaggtgaagaactgttcactggtgttgttccaattctggttgaactggatggtga tgttaatggtcacaaattttctgtctctggtgagggtgaaggtgatgcaacctacggtaaactgaccctgaaatt tatttgcactactggtaaactgcctgttccgtggccaaccctggtcactactctgacttggggtgttcaatgcttt gctcgttacccagatcacatgaaacagcatgacttttttcaagtctgccatgccggaaggttatgttcaggaacg tactatctttttcaaagatgacggtaactacaagacccgtgctgaagtcaagtttgaaggtgataccctggttaa tcgtatcgagctgaaaggtattgattttaaagaagatggtaacattctgggtcacaaactggaatacaacgctat ttctgataatgtatacatcactgctgacaaacaaaagaatggtatcaaagctaatttcaaaattcgtcacaacatt gaagatggtagcgttcaactggcagaccattatcaacaaaatactccaattggcgatggccctgtcctgctgc cagacaaccattacctgtccacccaatctcgtctgtctaaagatccgaacgaaaagcgcgatcacatggtcct gctggagtttgtaaccgctgctggtattaccctgggcatggatgaactgtataaataatag
exsA	Gene[202]	atgcaaggagccaaatctcttggccgaaagcagataacgtcttgtcattggaacattccaactttcgaatacag ggtaaacaaggaagagggcgtatatgttctgctcgagggcgaactgaccgtccaggacatcgattccacttttt tgcctggcgcctggcgagttgcttttcgtccgccgcggaagctatgtcgtaagtaccaagggaaaggacag ccgaatactctggattccattatctgcccagtttctacaaggcttcgtccagcgcttcggcgcgctgttgagtga agtcgagcgttgcgacgagcccgtgccgggcatcatcgcgttcgctgccacgcctctgctggccggttgcg tcaaggggttgaaggaattgcttgtgcatgagcatccgccgatgctcgcctgcctgaagatcgaggagttgct gatgctcttcgcgttcagtccgcaggggccgctgctgatgtcggtcctgcggcaactgagcaaccggcatgt cgagcgtctgcagctattcatggagaagcactacctcaacgagtggaagctgtccgacttctcccgcgagtt cggcatggggctgaccaccttcaaggagctgttcggcagtgtctatggggtttcgccgcgcgcctggatcag cgagcggagaatcctctatgcccatcagttgctgctcaacagcgacatgagcatcgtcgacatcgccatgga ggcgggcttttccagtcagtcctatttcacccagagctatcgccgccgtttcggctgcacgccgagccgctcg cggcaggggaaggacgaatgccgggctaaaaataactga

exsD	Gene[202]	atggagcaggaagacgataagcagtactcccgagaagcggtgttcgctggcaggcgggtatccgtggtgg gctcggacgcccgctcgcggggtcgggtgccgggttacgcatcgagcagtttgtatcgtgagtccggaatc atcagtgcgcggcaactggcgttgctgcagcggatgctgccgcgcctgcggctggagcaactgttccgctg cgagtggttgcagcagcgcctggcgcgcggcctggcgctggggcgcgaagaggtgcggcagattctcct ctgcgcggcgcaggacgacgacggctggtgctccgaactgggcgaccgggtcaacctcgccgtgccgca gtcgatgatcgactgggtcctgctgccggtctatggctggtgggaaagcctgctcgaccaggcgatccccg gctggcgcctgtcgctggtggagctggagacccagtcccggcaactgcgagtcaagtccgaattctggtcc cgcgtggccgagctggagccggagcaggcccgcgaggaactggccagggtcgccaagtgccaggcgc gcacccaggaacaggtggccgaactggccggcaagctggagacggcttcggcactggcgaagagcgcc tggccgaactggcagcggggcatggcgacgctgctcgccagcggcgggctggccggcttcgagccgatc cccgaggtcctcgaatgcctctggcaacctctctgccggctggacgacgacgtcggcgcggcggacgccg tccaggcctggctgcacgaacgcaacctgtgccaggcacaggatcacttctactggcagagctga
pTet	Promoter[64]	ttttcagcaggacgcactgacctccctatcagtgatagagattgacatccctatcagtgatagagatactgagc acatct
pBad	Promoter[198]	agaaaccaattgtccatattgcatcagacattgccgtcactgcgtcttttactggctcttctcgctaaccaaaccg gtaaccccgcttattaaaagcattctgtaacaaagcgggaccaaagccatgacaaaaacgcgtaacaaaagt gtctataatcacggcagaaaagtccacattgattatttgcacggcgtcacactttgctatgccatagcatttttatc cataagattagcggatcctacctg
pLux	Promoter[64]	acctgtaggatcgtacaggtttacgcaagaaaatggtttgttactttcgaataaa
pExsD	Promoter[202]	gaaggacgaatgccgggctaaaaataactgacgtttttttgaaagcccggtagcggctgcatgagtagaatcg gcccaaat
Bba_J23105	Promoter (http://parts.igem.org/Part:BBa_J23105)	tttacggctagctcagtcctaggtactatgctagc
Bba_J23116	Promoter (http://parts.igem.org/wiki/index.php/Part:BBa_J23116)	ttgacagctagctcagtcctagggactatgctagc
Bba_J23110	Promoter (http://parts.igem.org/wiki/index.php/Part:BBa_J23110)	tttacggctagctcagtcctaggtacaatgctagc
T0 Terminator	Terminator[64, 203]	ttgttcagaacgctcggttgccgccgggcgtttttttattggtgagaatcca
dbl term	Terminator	ccaggcatcaaataaaacgaaaggctcagtcgaaagactgggcctttcgttttatctgttgtttgtcggtgaac gctctctactagagtcacactggctcaccttcgggtgggcctttctgcgtttata
rfp UTR	UTR	gaattcaaaagatcttttaagaaggagatatacat
gfp UTR	UTR	tagcgaattcacttattaaagaacaggagtaagta

cfp UTR	UTR	attcgagctcggtactcacacaggaaaggcctcg
exsD UTR	UTR	aagcgcgcaatgtagtgaggaagccaaggcagaga
MicC	Hfq Binding Site[52]	tttctgttgggccattgcattgccactgattttccaacatataaaaagacaagcccgaacagtcgtccgggctttttttctcgag
MicF	Hfq Binding Site[52]	tcatttctgaatgtctgtttacccctatttcaaccggatgcctcgcattcggtttttttt
MicF M7.4	Hfq Binding Site[142]	cgtcccgcaaggatgcgggtctgtttacccctatttcaaccggccgcctcgcggccggtttttttttt
SgrS	Hfq Binding Site[52]	gatgaagcaaggggggtgccccatgcgtcagttttatcagcactattttaccgcgacagcgaagttgtgctggttgcgttggttaagcgtcccacaacgattaaccatgcttgaaggactgatgcagtgggatgaccgcaattctgaaagttga...atcacccgccagcagattatacctgctggttttttttt
Spot42	Hfq Binding Site[143]	atttggctgaatattttagccgccccagtcagtaatgactggggcgttttttta
TBR1	TBR	ttaaaagatcttttgaattc
TBR2	TBR	ttaaaagatctttgaattc
TBR3	TBR	ttaaagatctttgaattc
TBR4	TBR	ttaaagtctttgaattc
TBR5	TBR	atgtatatctccttcttaaaagatcttttgaattc
TBR6	TBR	atgtatatctccttcttaaaagatctttgaattc
TBR7	TBR	atgtatatctccttcttaaagatctttgaattc
TBR8	TBR	agtatatctcctcttaagatctttgaattc
TBR9	TBR	catatgtatatctccttcttaaaagatcttttgaattc
TBR10	TBR	catatgtatatctccttcttaaagatcttttgaattcc
TBR11	TBR	catatgtatatctcttcttaaagatctttgaattc
TBR12	TBR	catatgtatatctccttcttagatcttgaatc
TBR13	TBR	gataacgtcttcgctactcgccatatgtatatctccttcttaaaagatcttttgaattc
TBR14	TBR	gataacgtcttcgctactcgccatatgtattctccttcttaaagatctttgaattc
TBR15	TBR	gataacgtccgctactcgcatatgttatctccttcttaagatctttgaattc
TBR16	TBR	gatacgtcttcgctactcgcattgtattccttcttaaaagatctttgattc

TBR17	TBR	catatgtatatctccttc
TBR18	TBR	catatgtattctccttc
TBR19	TBR	catatgtatctccttc
TBR20	TBR	catatgatatctcctc
TBR21	TBR	gataacgtcttcgctactcgccatatgtatatctccttc
TBR22	TBR	gataacgtcttcgctactcgcatatgtatatctccttc
TBR23	TBR	gatacgtcttcctctcgccatatgtatatctcctc
TBR24	TBR	gatacgtcttcgctctcgccaatgttatccctc
TBR25	TBR	gataacgtcttcgctactcgccat
TBR26	TBR	gataacgtctcgctactcgccat
TBR27	TBR	gataacgcttcgctaccgccat
TBR28	TBR	gatacgtctcgctctgccat
TBR29	TBR	catatgtatatctccttcttaaaagatcttttgaaccaagatcttaagagagatacatg
TBR30	TBR	catatgtatatctcctcttaaaagatcttttgaaccaagtcttaagaggatatcattg
TBR31	TBR	catatgttatctccttcttaaagatcttttgaaccaagatcttaagagagatacattg
TBR32	TBR	catatgttatctctcttaagatcttttgaaccaagatctaaagagatacaatg
TBR33	TBR	gttctttaataagtgaattcgcta
TBR34	TBR	gttctttataagtgaattcgcta
TBR35	TBR	gttcttaataagtgattcgcta
TBR36	TBR	gtcttaataatgaatcgcta
TBR37	TBR	tacttactcctgttctttaataagtgaattcgcta
TBR38	TBR	tacttactcctgttcttaataagtgaattcgcta
TBR39	TBR	tacttactctgttctttaatagtgaatcgcta
TBR40	TBR	tactaccctgttcttaataagtgattgcta
TBR41	TBR	cattacttactcctgttctttaataagtgaattcgcta
TBR42	TBR	cattacttactcctgttcttaataagtgaattcgcta
TBR43	TBR	cattactactcctgttcttaataagtgattcgcta

TBR44	TBR	cattacactcctgttcttataagtattcgcta
TBR45	TBR	gaaaagttcttctcctttactcattacttactcctgttctttaataagtgaattcgcta
TBR46	TBR	gaaaagttcttctccttactcattacttactcctgttcttaataagtgaattcgcta
TBR47	TBR	gaaagttctctcctttacattacttactcctgttcttaataagtgaattcgcta
TBR48	TBR	gaagttcttcctttactctacttactcctgttctaataagtgattcgcta
TBR49	TBR	cattacttactcct
TBR50	TBR	cattactactcct
TBR51	TBR	cattcttatcct
TBR52	TBR	cattacaccct
TBR53	TBR	gaaaagttcttctcctttactcattacttactcct
TBR54	TBR	gaaaagttcttctccttactcattacttactcct
TBR55	TBR	gaaagttctttcctttactcatacttactcct
TBR56	TBR	gaaagttctccttactcattactactcct
TBR57	TBR	gaaaagttcttctcctttactcat
TBR58	TBR	gaaaagttctctcctttactcat
TBR59	TBR	gaaaagttctctccttactcat
TBR60	TBR	gaaaagcttctcttactcat
TBR61	TBR	cattacttactcctgttctttaataagtgaattcgctcttcacttataagaacaggtaaaatg
TBR62	TBR	cattacttactcctgttcttaataagtgaattcgctcttcacttataagaacaggtaaaatg
TBR63	TBR	cattactactcctgttctttatagtgaattcgctcttcacttataagaacaggtaaaatg
TBR64	TBR	cattatactctgttctaatatgaattcgctcttcacttataagaacaggtaaaatg
TBR65	TBR	gtgtgagtaccgagctcgaat
TBR66	TBR	gtgtgagtacgagctcgaat
TBR67	TBR	gtgtggtaccggctcgaat
TBR68	TBR	gtgtgaaccgctcgaat
TBR69	TBR	cgaggcctttcctgtgtgagtaccgagctcgaat
TBR70	TBR	cgaggcctttcctgtggagtaccgagctcgaat

TBR71	TBR	cgaggccttcctgtggagtacgagctcgaat
TBR72	TBR	cgaggccttcctgtgagacgagctcgaat
TBR73	TBR	catcgaggcctttcctgtgtgagtaccgagctcgaat
TBR74	TBR	catcgaggccttcctgtgtgagtaccgagctcgaat
TBR75	TBR	catcgaggccttcctgtggagtaccgagctgaat
TBR76	TBR	catcggccttctgtgtgagtccgactcgaat
TBR77	TBR	acctttgcttcttttgctagtcatcgaggcctttcctgtgtgagtaccgagctcgaat
TBR78	TBR	acctttgcttctttgctagtcatcgaggccttcctgtgtgagtaccgagctcgaat
TBR79	TBR	acctttgctcttttgctgtcatcgagcctttcctgtgtggtaccggctcgaat
TBR80	TBR	acctttgtctttgctagtcacgaggcttcctgtgagtacgactcgaat
TBR81	TBR	catcgaggcctttcct
TBR82	TBR	catcgaggccttcct
TBR83	TBR	catcgggctttcct
TBR84	TBR	catcaggcttcct
TBR85	TBR	acctttgcttcttttgctagtcatcgaggcctttcct
TBR86	TBR	acctttgcttctttgctagtcatcgaggcctttcct
TBR87	TBR	acctttgcttctttgctagtcatgaggccttcct
TBR88	TBR	accttgctttttgctagtctcgagccttcct
TBR89	TBR	acctttgcttcttttgctagtcat
TBR90	TBR	acctttgcttctttgctagtcat
TBR91	TBR	accttgcttctttgctatcat
TBR92	TBR	accttgctttttgcgtcat
TBR93	TBR	catcgaggcctttcctgtgtgagtaccgagctcgaatctcgtatcacacagaagcctcgtg
TBR94	TBR	catcgaggccttcctgtgtgagtaccgagctcgaatctcgtatcacacagaagcctcgtg
TBR95	TBR	catcaggccttcctgtgtagtaccgagctcgaatctcgtatcacacagaagcctcgtg
TBR96	TBR	catcggccttctgtgtgagacgagctcgaatctcgtatcacacagaagcctcgtg
asExsA, L2-TI	TBR	attataagaaccccaAcactgtaccgagctcgaattc

asExsA, L2-TII	TBR	attataagacccaAcactgtaccgagccgaattc
asExsA, L2-TIII	TBR	attatagaacccacactgtacgagtcgaattc
asExsA, L6-TI	TBR	gccaagagatttggctccttgcatattataagaacccca
asExsA, L6-TII	TBR	gccaagaatttggctccttgcatatataagacccca
asExsA, L6-TIII	TBR	gccaaggatttctccttcatattataaacccca
asExsD-S	TBR	ctgcttatcgtcttcctgctccattctctgccttggcttcctcac
asExsD-W	TBR	tcttcctgctattctcccttg

Table 12: Alignment of asRNAs targeting RFP. The top, bold sequence in each alignment is the mRNA location that is targeted by the TBR. The four aligned sequences are the four TBRs targeting the location. TBRs are aligned as reverse complements, as indicated by the arrows, which indicate the 5' to 3' directionality of each sequence. TBRs containing a YUNR motif include only the region of the TBR expected to directly bind the mRNA. Additional nucleotides that participate in the YUNR structure, but do not bind to the mRNA, are excluded.

Location, TBR	Alignment of mRNA Target Location and Four TBRs
Location 1	**>gaattcaaaagatcttttaa>**
TBR1	<gaattcaaaagatcttttaa<
TBR2	<gaattcaaa-gatcttttaa<
TBR3	<gaattcaaa-gatc-tttaa<
TBR4	<gaattcaaa-ga-c-tttaa<
Location 2	**>gaattcaaaagatcttttaagaaggagatatacat>**
TBR5	<gaattcaaaagatcttttaagaaggagatatacat<
TBR6	<gaattcaaa-gatcttttaagaaggagatatacat<
TBR7	<gaattcaaa-gatc-tttaagaaggagatatacat<
TBR8	<gaattcaaa-gatc--ttaag-aggagatatac-t<
Location 3	**>gaattcaaaagatcttttaagaaggagatatacatatg>**
TBR9	<gaattcaaaagatcttttaagaaggagatatacatatg<
TBR10	<gaattcaaaagatcttt-aagaaggagatatacatatg<
TBR11	<gaattcaaa-gatc-tttaagaa-gagatatacatatg<
TBR12	<ga-ttc--aagatc---taagaaggagatatacatatg<
Location 4	**>gaattcaaaagatcttttaagaaggagatatacatatggcgagtagcgaagacgttatc>**
TBR13	<gaattcaaaagatcttttaagaaggagatatacatatggcgagtagcgaagacgttatc<
TBR14	<gaattcaaa-gatc-tttaagaaggaga-atacatatggcgagtagcgaagacgttatc<
TBR15	<gaattcaaa-gatc--ttaagaaggagata-acatat-gcgagtagcg--gacgttatc<
TBR16	<gaat-c-aaagatcttttaagaagga---ataca-at-gcgagtagcgaagacgt-atc<
Location 5	**>gaaggagatatacatatg>**
TBR17	<gaaggagatatacatatg<
TBR18	<gaaggaga-atacatatg<
TBR19	<gaaggagata--catatg<
TBR20	<ga-ggagatat-catatg<
Location 6	**>gaaggagatatacatatggcgagtagcgaagacgttatc>**
TBR21	<gaaggagatatacatatggcgagtagcgaagacgttatc<
TBR22	<gaaggagatatacatatg-cgagtagcgaagacgttatc<
TBR23	<ga-ggagatatacatatggcgag-ag-gaagacg-tatc<
TBR24	<ga-gg-gata-acat-tggcgag-agcgaagacg-tatc<
Location 7	**>atggcgagtagcgaagacgttatc>**
TBR25	<atggcgagtagcgaagacgttatc<
TBR26	<atggcgagtagcga-gacgttatc<
TBR27	<atggcg-gtagcgaag-cgttatc<
TBR28	<atggc-ag-agcg-agacg-tatc<
Location 3 + YUNR	**>gaattcaaaagatcttttaagaaggagatatacatatg>**
TBR29	<~~~ttcaaaagatcttttaagaaggagatatacatatg<
TBR30	<~~~ttcaaaagatcttttaaga-ggagatatacatatg<
TBR31	<~~~ttcaaaagatc-tttaagaaggagata-acatatg<
TBR32	<~~~ttcaaaagatc--ttaaga--gagata-acatatg<

Table 13: Alignment of asRNAs targeting GFP. The top, bold sequence in each alignment is the mRNA location that is targeted by the TBR. The four aligned sequences are the four TBRs targeting the location. TBRs are aligned as reverse complements, as indicated by the arrows, which indicate the 5' to 3' directionality of each sequence. TBRs containing a YUNR motif include only the region of the TBR expected to directly bind the mRNA. Additional nucleotides that participate in the YUNR structure, but do not bind to the mRNA, are excluded.

Location, TBR	Alignment of mRNA Target Location and Four TBRs
Location 1	**>tagcgaattcacttattaaagaac>**
TBR33	<tagcgaattcacttattaaagaac<
TBR34	<tagcgaattcacttat-aaagaac<
TBR35	<tagcgaat-cacttatt-aagaac<
TBR36	<tagcga-ttca-ttatt-aag-ac<
Location 2	**>tagcgaattcacttattaaagaacaggagtaagta>**
TBR37	<tagcgaattcacttattaaagaacaggagtaagta<
TBR38	<tagcgaattcacttattaa-gaacaggagtaagta<
TBR39	<tagcga-ttcac-tattaaagaaca-gagtaagta<
TBR40	<tagc-aa-tcacttatt-aagaacagg-gt-agta<
Location 3	**>tagcgaattcacttattaaagaacaggagtaagtaatg>**
TBR41	<tagcgaattcacttattaaagaacaggagtaagtaatg<
TBR42	<tagcgaattcacttattaa-gaacaggagtaagtaatg<
TBR43	<tagcgaat-cacttatt-aagaacaggagta-gtaatg<
TBR44	<tagcgaat--acttat--aagaacaggagt--gtaatg<
Location 4	**>tagcgaattcacttattaaagaacaggagtaagtaatgagtaaaggagaagaactttc>**
TBR45	<tagcgaattcacttattaaagaacaggagtaagtaatgagtaaaggagaagaactttc<
TBR46	<tagcgaattcacttattaa-gaacaggagtaagtaatgagtaa-ggagaagaactttc<
TBR47	<tagcgaattcacttattaa-gaacaggagtaagtaatg--taaaggag-agaac-tttc<
TBR48	<tagcgaat-cacttatt--agaacaggagtaagta--gagtaaag--gaagaac--ttc<
Location 5	**>aggagtaagtaatg>**
TBR49	<aggagtaagtaatg<
TBR50	<aggagta-gtaatg<
TBR51	<agga-taag-aatg<
TBR52	<agg-gt--gtaatg<
Location 6	**>aggagtaagtaatgagtaaaggagaagaactttc>**
TBR53	<aggagtaagtaatgagtaaaggagaagaactttc<
TBR54	<aggagtaagtaatgagtaa-ggagaagaactttc<
TBR55	<aggagtaagta-tgagtaaagga-aagaac-tttc<
TBR56	<aggagta-gtaatgagt-aagg---agaac-tttc<
Location 7	**>atgagtaaaggagaagaactttc>**
TBR57	<atgagtaaaggagaagaactttc<
TBR58	<atgagtaaaggaga-gaactttc<
TBR59	<atgagtaa-ggag-agaactttc<
TBR60	<atgagtaa--gagaag--ctttc<
Location 3 + YUNR	**>tagcgaattcacttattaaagaacaggagtaagtaatg>**
TBR61	<~agcgaattcacttattaaagaacaggagtaagtaatg<
TBR62	<~agcgaattcacttatt-aagaacaggagtaagtaatg<
TBR63	<~agcgaattcac-ta-taaagaacaggagta-gtaatg<
TBR64	<~agcgaattca--tatt--agaaca-gagta--taatg<

Table 14: Alignment of asRNAs targeting CFP. The top, bold sequence in each alignment is the mRNA location that is targeted by the TBR. The four aligned sequences are the four TBRs targeting the location. TBRs are aligned as reverse complements, as indicated by the arrows, which indicate the 5' to 3' directionality of each sequence. TBRs containing a YUNR motif include only the region of the TBR expected to directly bind the mRNA. Additional nucleotides that participate in the YUNR structure, but do not bind to the mRNA, are excluded.

Location, TBR	Alignment of mRNA Target Location and Four TBRs
Location 1	**>attcgagctcggtactcacac>**
TBR65	<attcgagctcggtactcacac<
TBR66	<attcgagctcg-tactcacac<
TBR67	<attcgagc-cggtac-cacac<
TBR68	<attcgagc--ggt--tcacac<
Location 2	**>attcgagctcggtactcacacaggaaaggcctcg>**
TBR69	<attcgagctcggtactcacacaggaaaggcctcg<
TBR70	<attcgagctcggtactc-cacaggaaaggcctcg<
TBR71	<attcgagctcg-tactc-cacagg-aaggcctcg<
TBR72	<attcgagctcg-t-ct--cacagg-aaggcctcg<
Location 3	**>attcgagctcggtactcacacaggaaaggcctcgatg>**
TBR73	<attcgagctcggtactcacacaggaaaggcctcgatg<
TBR74	<attcgagctcggtactcacacaggaa-ggcctcgatg<
TBR75	<attc-agctcggtactc-cacagg-aaggcctcgatg<
TBR76	<attcgag-tcgg-actcacacag--aaggc--cgatg<
Location 4	**>attcgagctcggtactcacacaggaaaggcctcgatgactagcaaaagaagcaaaggt>**
TBR77	<attcgagctcggtactcacacaggaaaggcctcgatgactagcaaaagaagcaaaggt<
TBR78	<attcgagctcggtactcacacaggaa-ggcctcgatgactagcaaa-gaagcaaaggt<
TBR79	<attcgagc-cggtac-cacacaggaaaggc-tcgatgac-agcaaaag-agcaaaggt<
TBR80	<attcgag-tc-gtact--cacaggaa--gcctcg-tgactagc-aaaga--caaaggt<
Location 5	**>aggaaaggcctcgatg>**
TBR81	<aggaaaggcctcgatg<
TBR82	<aggaa-ggcctcgatg<
TBR83	<aggaaag-cc-cgatg<
TBR84	<aggaa--gcct-gatg<
Location 6	**>aggaaaggcctcgatgactagcaaaagaagcaaaggt>**
TBR85	<aggaaaggcctcgatgactagcaaaagaagcaaaggt<
TBR86	<aggaaaggcctcgatgactagcaaa-gaagcaaaggt<
TBR87	<aggaa-ggcctc-atgactagc-aaagaagcaaaggt<
TBR88	<aggaa-gg-ctcga-gactagcaaa--aagc-aaggt<
Location 7	**>atgactagcaaaagaagcaaaggt>**
TBR89	<atgactagcaaaagaagcaaaggt<
TBR90	<atgactagcaaa-gaagcaaaggt<
TBR91	<atga-tagc-aaagaagcaa-ggt<
TBR92	<atgac--gcaaa--aagc-aaggt<
Location 3 + YUNR	**>attcgagctcggtactcacacaggaaaggcctcgatg>**
TBR93	<attcgagctcggtactcacacaggaaaggcctcgatg<
TBR94	<attcgagctcggtactcacacaggaa-ggcctcgatg<
TBR95	<attcgagctcggtact-acacagg-aaggcct-gatg<
TBR96	<attcgagctcg-t-ctcacacag--aaggc--cgatg<

Table 15: Primers for RT-qPCR

Primer	Sequence	Source
cysGF	ttgtcggcggtggtgatgtc	[92]
cysGR	atgcggtgaactgtggaataaacg	[92]
hcaTF	gctgctcggctttctcatcc	[92]
hcaTR	ccaaccacgctgaccaacc	[92]
idnTF	ctgtttagcgaagaggagatgc	[92]
idnTR	acaaacggcggcgatagc	[92]
MicFR2	cggccggttgaaataggggtaaac	This Study
asRNA36F	gtcttaataatgaatcgctacgtcccgcaag	This Study
asRNA39F	cttactctgttctttaatagtgaatcgctacgtccc	This Study
asRNA45F	gttcttctcctttactcattacttactcctgttctttaataagtg	This Study
asRNA46F	gaaaagttcttctccttactcattacttactcctgttcttaataagtg	This Study
asRNA55F	gttctttcctttactcatacttactcctcgtccc	This Study
asRNA60F	gaaaagcttctcttactcatcgtcccgcaag	This Study

Table 16: Design parameters for asRNAs targeting mRNA of *exsA* or *exsD*. Details for each of the asRNAs used in both the simple and the complex genetic circuits are shown below. Type I, II, and III asRNAs are defined in Figure 5a.

Target Location	Type	Figure #	Maximum dsRNA Length (nt)	ΔG CF (kcal/mol)	Percent Mismatch (%)	Follows Design Rules?
2	I	5	37	-60.2	0.0	Yes
2	II	5	18	-48.2	8.8	Yes
2	III	5	8	-31.8	15.6	No
6	I	5	39	-60.2	0.0	Yes
6	II	5	17	-43.7	8.3	Yes
6	III	5	9	-31.2	18.2	No
4	I (asExsD-S)	6	45	-81.9	0.0	Yes
4	III (asExsD-W)	6	10	-27.7	19.0	No

Supplementary Methods: RT-qPCR for asRNA

RT-qPCR was performed to determine whether transcript abundance was correlated with repression efficiency. Six strains containing six different asRNAs (Supplementary Table 5; asRNA36, asRNA39, asRNA45, asRNA46, asRNA55, and asRNA60) were included in this experiment. These six asRNAs were included because they had achieved varying levels of repression efficiency, and all targeted *gfp*. Induction was performed as described in the main text. Samples were treated with 300 ug/mL rifampicin as described previously.[91] RNA was then isolated from two biological replicates for each strain, with a total culture volume of 1.5 mL per culture. RNA was extracted using TRIzol Reagent (Life Technologies), as described previously.[49] A Nanodrop 2000 UV-Vis Spectrophotometer was used to measure concentrations and purities, (A260/A280 values) which ranged from 1.96 to 2.05. RNA samples were then treated with DNase (DNA-free Kit, Life Technologies). PCR with the DNase treated samples confirmed the absence of DNA contamination. DNase-treated RNA samples were then used to generate cDNA libraries according to manufacturer's instructions (AffinityScript QPCR cDNA synthesis kit, Agilent Technologies).

Three reference genes, *cysG*, *hcaT,* and *idnT*, were chosen based on literature.[92] One reverse primer was used for all the asRNAs (MicFR2), since each asRNA contained the same MicF M7.4 sequence. Each asRNA had a unique forward primer. All primer sequences can be found in Supplementary Table 8. All amplicons were 65-150 nt in length. Though some amplicons were unusually small (e.g. 65 nt), amplicon size was limited by the length of the asRNA. Primer concentrations were optimized, and calibration curves were generated for each set of primers as described previously.[49] Efficiencies were found to be >90% for each set of primers. Varying

amounts of cDNA were required for the different target genes due to different expression levels (6 μg for *idnT* and *hcaT*, 12 μg for *cysG*, and 400 ng for each asRNA).

RT-qPCR was performed with 50 nM primers and Power SYBR Green PCR Master Mix (Life Technologies). Cycling conditions were as follows: 95°C for 10 min, then 40 cycles of 95°C for 15s and 60°C for 1 min, then fluorescent detection. A melting curve analysis (65-95°C, 0.5°C increments for 5s) was used to confirm the absence of nonspecific products. This was performed with a Real-Time PCR Detection System (Bio-Rad Laboratories, Inc.). The CFX96 TouchTM Real-Time PCR Detection System software (Bio-Rad Laboratories, Inc.) was used to determine quantification cycles (Cq). Cq for each asRNA was normalized to the geometric mean of the three reference genes. Corrections (log transformation, mean centering, and autoscaling) were applied to the data as described previously,[94] in order to account for expected variation between biological replicates. This process, including data analysis, was performed in accordance with MIQE guidelines.[93] Two biological and two technical replicates (total four replicates) were averaged and the standard error of the mean was calculated.

Appendix C: Supplementary Data for pH- and Temperature-Responsive Circuits

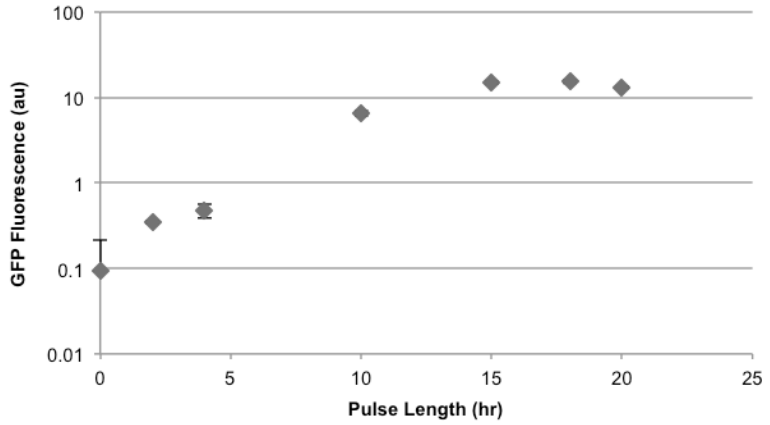

Figure 35: Response of pAsr to pulses of pH=5 media. Cells were grown in pH=5 media for varying lengths of time (pulse length), and then centrifuged and resuspended in pH=7 media. g*fp* expression increases with increasing pulse length, but levels off after approximately 15 hours. Experiments were conducted in triplicate and error bars represent standard error of the mean (SEM)

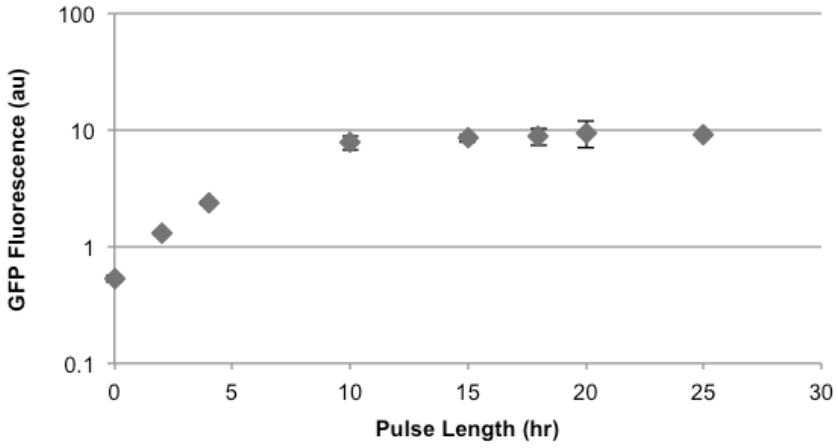

Figure 36: Response of pCspA promoter to pulses of 27°C. Cells were grown at 27°C for varying lengths of time (pulse length), and then moved to 37°C. *gfp* expression increases with increasing pulse length, but levels off after approximately 10 hours. Experiments were conducted in triplicate and error bars represent standard error of the mean (SEM).

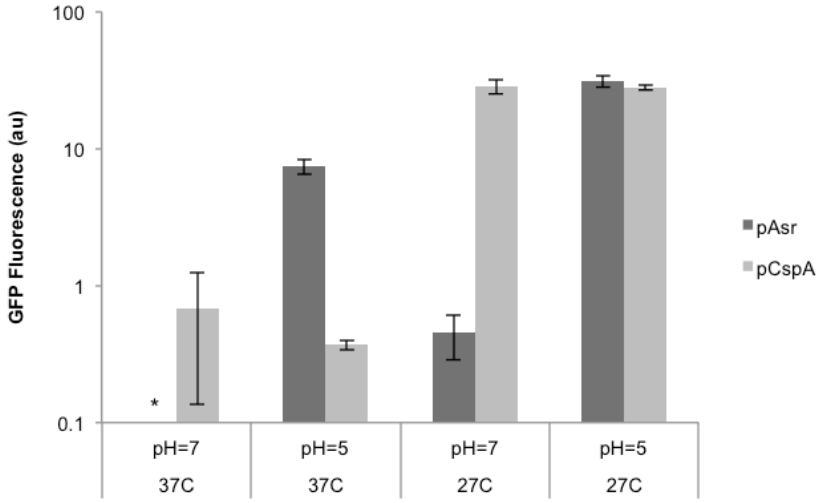

Figure 37: Orthogonality of pAsr and pCspA promoters. The pAsr promoter turns on at pH=5 and off at pH = 7. The pAsr promoter does not respond to temperature, and can function at either 27°C or 37°C. The pCspA promoter turns on at 27°C and off at 37°C. The pCspA promoter does not respond to pH, and can function at either pH=5 or pH=7. The "*" indicates that the *gfp* expression level for this culture was within one standard deviation of the DH10B autofluorescence level, and was considered completely "off". Experiments were conducted in triplicate and error bars represent standard error of the mean (SEM)

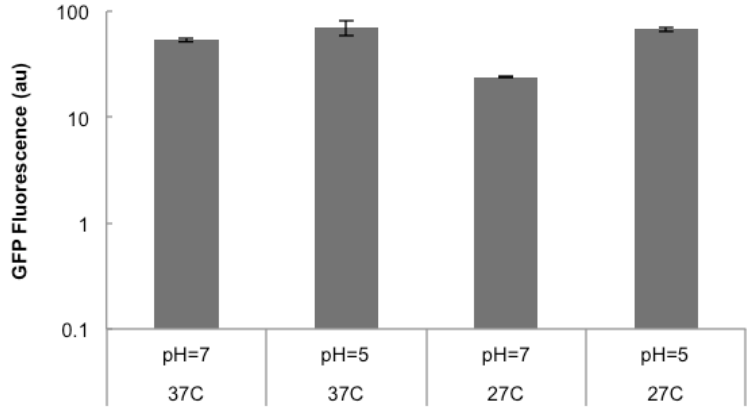

Figure 38: Results for initial temperature/pH AND Gate (Circuit N). This circuit did not function as expected. Experiments were conducted in triplicate and error bars represent standard error of the mean (SEM).

Table 17: Plasmids used in this study

Name	Parts
pAH005	p15a ori; Amp-R; pAsr-*gfp*
pAH353	p15a ori; Amp-R; pAsr_JC2G10-*gfp*
pAH354	p15a ori; Amp-R; pAsr_AH2D10-*gfp*
pAH355	p15a ori; Amp-R; pAsr_JC1E9-*gfp*
pAH356	p15a ori; Amp-R; pAsr_JC1H1-*gfp*
pTS048	ColE1 ori; Cm-R; pCspA-*gfp*
pAH305	p15a ori; Amp-R; pCspA_ATG$_{cspA}$-*sicA*
pAH293	p15a ori; Amp-R; pCspA_ATG$_{cspA}$_ ATG$_{sicA}$-*sicA*
pAH312	p15a ori; Amp-R; pCspA_ATG→GGC$_{cspA}$_ ATG$_{sicA}$-*sicA*
pAH314	p15a ori; Amp-R; pCspA_ATG→GGG$_{cspA}$_ ATG$_{sicA}$-*sicA*
pAH315	p15a ori; Amp-R; pCspA_ATG$_{cspA}$_ ATG→TAG$_{sicA}$-*sicA*
pAH316	p15a ori; Amp-R; pCspA_ATG$_{cspA}$_ ATG→GGT$_{sicA}$-*sicA*
pAH317	p15a ori; Amp-R; pCspA_ATG$_{cspA}$_ ATG→TTA$_{sicA}$-*sicA*
pAH320	p15a ori; Amp-R; pCspA_ATG$_{cspA}$_ ATG→CAG$_{sicA}$-*sicA*
pAH321	p15a ori; Amp-R; pCspA_ATG$_{cspA}$_ ATG→CCC$_{sicA}$-*sicA*
pAH322	p15a ori; Amp-R; pCspA_ATG$_{cspA}$_ ATG→GCG$_{sicA}$-*sicA*
pAH143	pSC101* ori; Kan-R; pAsr-*invF*
pSicA-*gfp*	ColE1 ori; Cm-R; pSicA-*gfp*
pAH325	pSC101* ori; Kan-R; pAsr-*invF,* BBa_J23110-*gfp*
pAH346	pSC101* ori; Kan-R; pAsr_JC1H1-*invF,* BBa_J23110-*gfp*
pAH326	ColE1 ori; Cm-R; pSicA-asRNA64-MicF M7.4

Table 18: *E. coli* strain used in this study.

Name	Host Strain	Plasmids
pH- and Temperature-Responsive Strains		
pAsr-*gfp*	DH10B	pAH005
pAsr_JC2G10-*gfp*	DH10B	pAH353
pAsr_AH2D10-*gfp*	DH10B	pAH354
pAsr_JC1E9-*gfp*	DH10B	pAH355
pAsr_JC1H1-*gfp*	DH10B	pAH356
pCspA-*gfp*	DH10B	pTS048
AND Gates		
Circuit N	DH10B	pAH305 + pAH143 + pSicA-*gfp*
Circuit V	DH10B	pAH312 + pAH143 + pSicA-*gfp*
Circuit X	DH10B	pAH314 + pAH143 + pSicA-*gfp*
Circuit Y	DH10B	pAH315 + pAH143 + pSicA-*gfp*
Circuit Z	DH10B	pAH316 + pAH143 + pSicA-*gfp*
Circuit AA	DH10B	pAH317 + pAH143 + pSicA-*gfp*
Circuit AD	DH10B	pAH320 + pAH143 + pSicA-*gfp*
Circuit AE	DH10B	pAH321 + pAH143 + pSicA-*gfp*
Circuit AF	DH10B	pAH322 + pAH143 + pSicA-*gfp*
NAND Gates		
Circuit AW	DH10B	pAH312 + pAH346 + pAH326
Circuit AY	DH10B	pAH314 + pAH346 + pAH326
Circuit BB	DH10B	pAH317 + pAH346 + pAH326
Circuit BE	DH10B	pAH320 + pAH346 + pAH326

Table 19: Genetic parts used in this study.

Part Name	Type and Source	DNA Sequence
gfp	Gene [198]	atgagtaaaggagaagaacttttcactggagttgtcccaattcttgttgaattagatggtgatgttaatgggcacaaatttt ctgtcagtggagagggtgaaggtgatgcaacatacggaaaacttacccttaaatttatttgcactactggaaaactacct gttccatggccaacacttgtcactactttgacttatggtgttcaatgcttttcaagatacccagatcatatgaaacggcatg actttttcaagagtgccatgcccgaaggttatgtacaggaaagaactatatttttcaaagatgacgggaactataagaca cgtgctgaagtcaagtttgaaggtgatacacttgttaatagaatcgagttaaaaggtattgattttaaagaagatggaaac attcttggacacaagttggaatacaactataactcacacaatgtatacatcatggcagacaaacaaaagaatggaatca aagttaacttcaaaattagacacaacattgaagatggaagcgttcaactagcagaccattatcaacaaaatactccaatt ggcgatggccctgtccttttaccagacaaccattacctgtccacacaatctgccctttcgaaagatcccaacgaaaaga gagaccacatggtccttcttgagtttgtaacagctgctgggattacacatggcatggatgaactatacaaaaggcctgc agcaaacgacgaaaactacgcttaagtagcttaa
*sicA**	Gene [64]	atggattatcaaaataatgtcagcgaagaacgtgttgcggaaatgatttgggatgccgttagtgaaggcgccacgctaa aagacgttcatgggatccctcaagatatgatggacggtttatatgctcatgcttatgagttttataaccagggacgactgg atgaagctgagacgttctttcgttacttatgcatttatgatttttacaatcccgattacaccatgggactggcggcagtatgc caactgaaaaaacaatttcagaaagcatgtgaccttatgcagtagcgtttacgttacttaaaaatgattatcgccccgtttt ttttaccgggcagtgtcaattattaatgcgtaaggcagcaaaagccagacagtgttttgaacttgtcaatgaacgtactga agatgagtctctgcgggcaaaagcgttggtctatctggaggcgctaaaaacggcggagacagagcagcacagtgaa caagaaaaggaataa
invF	Gene [64]	atgtcattttctgaaagccgacacaatgaaaattgcctgattcaggaaggcgcgctgctttttgcgagcaggccgttgt cgcaccagtatcaggagacctggtttttcgaccgttaaaaattgaagtactcagcaaattactggcatttatcgatggcg caggattagtggacacgacatatgctgaatccgataaatgggttttgctgagtcctgagtttcgcgctatttggcaagatc gtaaacgctgcgagtactggtttttgcagcaaattattacgccttctccggccttcaataaggtactggcgctgttacgaa aaagcgagagttactggttggttggctatttactcgctcagtcaaccagcggcaacacgatgagaatgctgggagaag actatggcgtttcttatacccattttcgtcgtttgtgcagcagagcgttgggcggaaaagcgaagagtgaattacgaaac tggcgtatggcgcaatcgctgctgaatagtgtagaaggccacgagaacatcacccaattagccgttaatcatggttact catcgccttcacatttttctagtgagatcaaagagctgatcggcgtttcgccgcggaaattatcaaatattattcaattggc agacaaatga
pAsr	Promoter [168]	gatcaagactactattattggtagctaaatttcccttaagtcacaatacgttattatcaacgctgtaatttattcagcgtttgta catatcgttacacgctgaaaccaaccactcacggaagtctgccattcccagggatatagttatttcaacggccccgcag tggggttaaatgaaaaaacaaattgagggtatgaca
pCspA	Promoter [171]	ctgatgacaggaccgttttccaaccgattaatcataaatatgaaaaataattgttgcatcacccgccaatgcgtggcttaa tgcacatcaacggtttgacgtacagaccattaaagcagtgtagtaaggcaagtcccttcaagagttatcgttgataccc tcgtagtgcacattccttaacgcttcaaaatctgtaaagcacgccatatcgccgaaaggcacacttaattattaaaggta atacactatgtccggtaaaatgactggtatcgtaaaatggttcaac
pSicA	Promoter [198]	ccacaagaaacgaggtacggcattgagccgcgtaaggcagtagcgatgtattcattgggcgtttttgaatgttcactaa ccaccgtcggggtttaataactgca
BBa_J23110	Promoter (http://parts.igem.org/wiki/index.php/Part:BBa_J23110)	tttacggctagctcagtcctaggtacaatgctagc
asRNA64	asRNA [47]	cattatactctgttctaatatgaattcgctcttcacttataagaacaggtaaaatgcgtcccgcaaggatgcgggtctgttta cccctatttcaaccggccgcctcgcggccggttttttttt
MicF M7.4	Hfq binding site [142]	cgtcccgcaaggatgcgggtctgtttacccctatttcaaccggccgcctcgcggccggttttttttt

Table 20: Sequences of pAsr variants.

Variant	Mutated Library Site	Library 1 Site Sequence	Library 2 Site Sequence
pAsr Original	None	TACA	TACA
JC2G10	Library 2 Site	TACA	TAAA
AH2D10	Library 2 Site	TACA	CGGT
JC1E9	Library 1 Site	TAAG	TACA
JC1H1	Library 1 Site	AGGG	TACA

Table 21: Sequences of pCspA-sicA variants.

Plasmid	AND Gate	NAND Gate	Mutated Start Codon	CspA ATG Sequence	SicA ATG Sequence	AND Gate Fold Change	NAND Gate Fold Change
pAH312	V	AW	CspA	GGC	ATG	17	36
pAH314	X	AY	CspA	GGG	ATG	10	1168
pAH315	Y		SicA	ATG	TAG	5	
pAH316	Z		SicA	ATG	GGT	17	
pAH317	AA	BB	SicA	ATG	TTA	23	73
pAH320	AD	BE	SicA	ATG	CAG	14	24
pAH321	AE		SicA	ATG	CCC	18	
pAH322	AF		SicA	ATG	GCG	11	

References

1. Monod, J. & Jacob, F. Teleonomic mechanisms in cellular metabolism, growth, and differentiation. *Cold Spring Harbor symposia on quantitative biology* **26**, 389-401 (1961).

2. Cameron, D.E., Bashor, C.J. & Collins, J.J. A brief history of synthetic biology. *Nature reviews. Microbiology* **12**, 381-390 (2014).

3. Gardner, T.S., Cantor, C.R. & Collins, J.J. Construction of a genetic toggle switch in Escherichia coli. *Nature* **403**, 339-342 (2000).

4. Elowitz, M.B. & Leibler, S. A synthetic oscillatory network of transcriptional regulators. *Nature* **403**, 335-338 (2000).

5. Tamsir, A., Tabor, J.J. & Voigt, C.A. Robust multicellular computing using genetically encoded NOR gates and chemical 'wires'. *Nature* **469**, 212-215 (2011).

6. Tabor, J.J. et al. A Synthetic Genetic Edge Detection Program. *Cell* **137**, 1272-1281 (2009).

7. Friedland, A.E. et al. Synthetic gene networks that count. *Science* **324**, 1199-1202 (2009).

8. Isaacs, F.J., Dwyer, D.J. & Collins, J.J. RNA synthetic biology. *Nature Biotechnology* **24**, 545-554 (2006).

9. Bayer, T.S. & Smolke, C.D. Programmable ligand-controlled riboregulators of eukaryotic gene expression. *Nature Biotechnology* **23**, 337-343 (2005).

10. Isaacs, F.J. et al. Engineered riboregulators enable post-transcriptional control of gene expression. *Nature Biotechnology* **22**, 841-847 (2004).

11. Lucks, J.B., Qi, L., Mutalik, V.K., Wang, D. & Arkin, A.P. Versatile RNA-sensing transcriptional regulators for engineering genetic networks. *Proceedings of the National Academy of Sciences of the United States of America* **108**, 8617-8622 (2011).

12. Lee, Y.J., Hoynes-O'Connor, A., Leong, M. & Moon, T.S. Programmable control of bacterial gene expression with the combined CRISPR and antisense RNA system. *Nucleic Acids Research* **44**, 2462-2473 (2016).

13. Liu, Y. et al. Synthesizing AND gate genetic circuits based on CRISPR-Cas9 for identification of bladder cancer cells. *Nature Communications* **5**, 10.1038/ncomms6393 (2014).

14. Nielsen, A.A.K. & Voigt, C.A. Multi-input CRISPR/Cas genetic circuits that interface host regulatory networks. *Molecular systems biology* **10**, 1-11 (2014).

15. Zalatan, J.G. et al. Engineering complex synthetic transcriptional programs with CRISPR RNA scaffolds. *Cell* **160**, 339-350 (2015).

16. Stephanopoulos, G. Synthetic Biology and Metabolic Engineering. *ACS Synth. Biol.* **1**, 514-525 (2012).

17. Farmer, W.R. & Liao, J.C. Improving lycopene production in Escherichia coli by engineering metabolic control. *Nature Biotechnology* **18**, 533-537 (2000).

18. Zhang, F., Carothers, J.M. & Keasling, J.D. Design of a dynamic sensor-regulator system for production of chemicals and fuels derived from fatty acids. *Nature Biotechnology* **30**, 354-359 (2012).

19. Liu, D., Xiao, Y., Evans, B.S. & Zhang, F. Negative Feedback Regulation of Fatty Acid Production Based on a Malonyl-CoA Sensor-Actuator. *ACS Synth Biol* **4**, 132-140 (2015).

20. Xu, P., Li, L., Zhang, F., Stephanopoulos, G. & Koffas, M. Improving fatty acids production by engineering dynamic pathway regulation and metabolic control. *Proc Natl Acad Sci U S A* **111**, 11299-11304 (2014).

21. Dahl, R.H. et al. Engineering dynamic pathway regulation using stress-response promoters. *Nature Biotechnology* **31**, 1039-1046 (2013).

22. Bacchus, W., Aubel, D. & Fussenegger, M. Biomedically relevant circuit-design strategies in mammalian synthetic biology. *Mol Syst Biol* **9**, 1-12 (2013).

23. Chen, X. et al. Synthetic dual-input mammalian genetic circuits enable tunable and stringent transcription control by chemical and light. *Nucleic Acids Res* **44**, 2677-2690 (2016).

24. Slomovic, S., Pardee, K. & Collins, J.J. Synthetic biology devices for in vitro and in vivo diagnostics. *Proc Natl Acad Sci U S A* **112**, 14429-14435 (2015).

25. Schukur, L. & Fussenegger, M. Engineering of synthetic gene circuits for (re-)balancing physiological processes in chronic diseases. *Wiley interdisciplinary reviews. Systems biology and medicine* **8**, 402-422 (2016).

26. Xie, Z., Wroblewska, L., Prochazka, L., Weiss, R. & Benenson, Y. Multi-Input RNAi-Based Logic Circuit for Identification of Specific Cancer Cells. *Science* **333**, 1307-1311 (2011).

27. Ding, M. et al. p53 activated by AND gate genetic circuit under radiation and hypoxia for targeted cancer gene therapy. *Cancer science* **106**, 1163-1173 (2015).

28. Khalil, A.S. & Collins, J.J. Synthetic biology: applications come of age. *Nature Reviews Genetics* **11**, 367-379 (2010).

29. Wang, B., Barahona, M. & Buck, M. A modular cell-based biosensor using engineered genetic logic circuits to detect and integrate multiple environmental signals. *Biosens Bioelectron* **40**, 368-376 (2013).

30. Stocker, J. et al. Development of a set of simple bacterial biosensors for quantitative and rapid measurements of arsenite and arsenate in potable water. *Environ Sci Technol* **37**, 4743-4750 (2003).

31. Garmendia, J., de las Heras, A., Galvao, T.C. & de Lorenzo, V. Tracing explosives in soil with transcriptional regulators of Pseudomonas putida evolved for responding to nitrotoluenes. *Microb Biotechnol* **1**, 236-246 (2008).

32. Temme, K., Zhao, D. & Voigt, C.A. Refactoring the nitrogen fixation gene cluster from *Klebsiella oxytoca*. *Proceedings of the National Academy of Sciences of the United States of America* **109**, 7085-7090 (2012).

33. Immethun, C.M. et al. Oxygen-responsive genetic circuits constructed in Synechocystis sp. PCC 6803. *Biotechnol Bioeng* **113**, 433-442 (2016).

34. Holtz, W.J. & Keasling, J.D. Engineering static and dynamic control of synthetic pathways. *Cell* **140**, 19-23 (2010).

35. Hoynes-O'Connor, A. & Moon, T.S. Programmable genetic circuits for pathway engineering. *Curr Opin Biotechnol* **36**, 115-121 (2015).

36. Keasling, J.D. Manufacturing molecules through metabolic engineering. *Science* **330**, 1355-1358 (2010).

37. Polizzi, K.M. & Kontoravdi, C. Genetically-encoded biosensors for monitoring cellular stress in bioprocessing. *Curr Opin Biotechnol* **31**, 50-56 (2015).

38. Bechet, Q., Shilton, A., Fringer, O.B., Munoz, R. & Guieysse, B. Mechanistic Modeling of Broth Temperature in Outdoor Photobioreactors. *Environmental Science & Technology* **44**, 2197-2203 (2010).

39. Akkermans, S., Noriega Fernandez, E., Logist, F. & Van Impe, J.F. Introducing a novel interaction model structure for the combined effect of temperature and pH on the microbial growth rate. *International journal of food microbiology* **10.1016/j.ijfoodmicro.2016.06.011.** (2016).

40. Zhang, J. et al. Metabolic engineering of Escherichia coli for the biosynthesis of 2-pyrrolidone. *Metabolic Engineering Communications* **3**, 1-7 (2016).

41. Demuth, C., Varonier, J., Jossen, V., Eibl, R. & Eibl, D. Novel probes for pH and dissolved oxygen measurements in cultivations from millilitre to benchtop scale. *Appl Microbiol Biotechnol* **100**, 3853-3863 (2016).

42. Guedes, A.C., Amaro, H.M., Pereira, R.D. & Malcata, F.X. Effects of temperature and pH on growth and antioxidant content of the microalga Scenedesmus obliquus. *Biotechnol Prog* **27**, 1218-1224 (2011).

43. Kalyoncu, E., Olmez, T.T., Ozkan, A.D. & Sarioglu, O.F. Biosystems Engineering of Prokaryotes with Tumor-Killing Capacities. *Current pharmaceutical design* **22**, 1521-1528 (2016).

44. Carothers, J.M., Goler, J.A., Juminaga, D. & Keasling, J.D. Model-driven engineering of RNA devices to quantitatively program gene expression. *Science* **334**, 1716-1719 (2011).

45. Stevens, J.T. & Carothers, J.M. Designing RNA-Based Genetic Control Systems for Efficient Production from Engineered Metabolic Pathways. *ACS Synth Biol* **4**, 107-115 (2015).

46. Thibault, J., Faudry, E., Ebel, C., Attree, I. & Elsen, S. Anti-activator ExsD forms a 1:1 complex with ExsA to inhibit transcription of type III secretion operons. *The Journal of biological chemistry* **284**, 15762-15770 (2009).

47. Hoynes-O'Connor, A. & Moon, T.S. Development of Design Rules for Reliable Antisense RNA Behavior in E. coli. *ACS Synth Biol*, 10.1021/acssynbio.1026b00036 (2016).

48. Darwin, K.H. & Miller, V.L. Type III secretion chaperone-dependent regulation: activation of virulence genes by SicA and InvF in *Salmonella typhimurium*. *Embo Journal* **20**, 1850-1862 (2001).

49. Hoynes-O'Connor, A., Hinman, K., Kirchner, L. & Moon, T.S. *De novo* design of heat-repressible RNA thermosensors in *E. coli Nucleic Acids Research* **43**, 6166-6179 (2015).

50. Liang, J.C., Bloom, R.J. & Smolke, C.D. Engineering Biological Systems with Synthetic RNA Molecules. *Molecular Cell* **43**, 915-926 (2011).

51. Zuker, M. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Research* **31**, 3406-3415 (2003).

52. Na, D. et al. Metabolic engineering of *Escherichia coli* using synthetic small regulatory RNAs. *Nature Biotechnology* **31**, 170-174 (2013).

53. Alessandra, S., Alessandro, T., Flavio, S. & Alejandro, H. Artificial antisense RNAs silence lacZ in *E-coli* by decreasing target mRNA concentration. *Bmb Reports* **41**, 568-574 (2008).

54. Ellison, M.J., Kelleher, R.J. & Rich, A. Thermal regulation of beta-galactosidase synthesis using anti-sense RNA directed against the coding portion of the messenger-RNA. *Journal of Biological Chemistry* **260**, 9085-9087 (1985).

55. Franch, T., Petersen, M., Wagner, E.G.H., Jacobsen, J.P. & Gerdes, K. Antisense RNA regulation in prokaryotes: Rapid RNA/RNA interaction facilitated by a general U-turn loop structure. *Journal of Molecular Biology* **294**, 1115-1125 (1999).

56. Hjalt, T.A.H. & Wagner, E.G.H. Bulged-out nucleotides protect an antisense RNA from RNase-III cleavage. *Nucleic Acids Research* **23**, 571-579 (1995).

57. Man, S. et al. Artificial trans-encoded small non-coding RNAs specifically silence the selected gene expression in bacteria. *Nucleic Acids Research* **39**, e50 (2011).

58. Mutalik, V.K., Qi, L., Guimaraes, J.C., Lucks, J.B. & Arkin, A.P. Rationally designed families of orthogonal RNA regulators of translation. *Nat Chem Biol* **8**, 447-454 (2012).

59. Yoo, S.M., Na, D. & Lee, S.Y. Design and use of synthetic regulatory small RNAs to control gene expression in *Escherichia coli*. *Nature Protocols* **8**, 1694-1707 (2013).

60. Solomon, K.V., Sanders, T.M. & Prather, K.L. A dynamic metabolite valve for the control of central carbon metabolism. *Metab Eng* **14**, 661-671 (2012).

61. Stefan, A. et al. Silencing of the gene coding for the epsilon subunit of DNA polymerase III slows down the growth rate of *Escherichia coli* populations. *FEBS Lett* **546**, 295-299 (2003).

62. Bonnet, J., Subsoontorn, P. & Endy, D. Rewritable digital data storage in live cells via engineered control of recombination directionality. *Proc Natl Acad Sci U S A* **109**, 8884-8889 (2012).

63. Shopera, T. et al. Robust, tunable genetic memory from protein sequestration combined with positive feedback. *Nucleic Acids Res* **43**, 9086-9094 (2015).

64. Moon, T.S., Lou, C., Tamsir, A., Stanton, B.C. & Voigt, C.A. Genetic programs constructed from layered logic gates in single cells. *Nature* **491**, 249-253 (2012).

65. Grossman, A.D., Erickson, J.W. & Gross, C.A. The *htpR* gene product of *E. coli* is a sigma factor for heat-shock proteins *Cell* **38**, 383-390 (1984).

66. Yura, T., Nagai, H. & Mori, H. Regulation of the heat-shock response in bacteria *Annual Review of Microbiology* **47**, 321-350 (1993).

67. Craig, E.A., Weissman, J.S. & Horwich, A.L. Heat shock proteins and molecular chaperones: mediators of protein conformation and turnover in the cell. *Cell* **78**, 365-372 (1994).

68. Muga, A. & Moro, F. Thermal Adaptation of Heat Shock Proteins. *Current Protein & Peptide Science* **9**, 552-566 (2008).

69. Guisbert, E., Yura, T., Rhodius, V.A. & Gross, C.A. Convergence of molecular, modeling, and systems approaches for an understanding of the *Escherichia coli* heat shock response. *Microbiol. Mol. Biol. Rev.* **72**, 545-554 (2008).

70. Jones, P.G. & Inouye, M. The cold-shock response - a hot topic. *Molecular Microbiology* **11**, 811-818 (1994).

71. Horn, G., Hofweber, R., Kremer, W. & Kalbitzer, H.R. Structure and function of bacterial cold shock proteins. *Cellular and Molecular Life Sciences* **64**, 1457-1470 (2007).

72. Kortmann, J. & Narberhaus, F. Bacterial RNA thermometers: molecular zippers and switches. *Nature Reviews Microbiology* **10**, 255-265 (2012).

73. Narberhaus, F., Waldminghaus, T. & Chowdhury, S. RNA thermometers. *Fems Microbiology Reviews* **30**, 3-16 (2006).

74. Narberhaus, F., Kaser, R., Nocker, A. & Hennecke, H. A novel DNA element that controls bacterial heat shock gene expression. *Mol Microbiol* **28**, 315-323 (1998).

75. Chowdhury, S., Ragaz, C., Kreuger, E. & Narberhaus, F. Temperature-controlled structural alterations of an RNA thermometer. *Journal of Biological Chemistry* **278**, 47915-47921 (2003).

76. Chursov, A., Kopetzky, S.J., Bocharov, G., Frishman, D. & Shneider, A. RNAtips: analysis of temperature-induced changes of RNA secondary structure. *Nucleic Acids Research* **41**, W486-W491 (2013).

77. Repoila, F. & Gottesman, S. Signal transduction cascade for regulation of RpoS: Temperature regulation of DsrA. *Journal of Bacteriology* **183**, 4012-4023 (2001).

78. Giuliodori, A.M. et al. The cspA mRNA is a thermosensor that modulates translation of the cold-shock protein CspA. *Mol Cell* **37**, 21-33 (2010).

79. Freischmidt, A., Hiltl, J., Kalbitzer, H.R. & Horn-Katting, G. Enhanced in vitro translation at reduced temperatures using a cold-shock RNA motif. *Biotechnology Letters* **35**, 389-395 (2013).

80. Liu, C.C. & Arkin, A.P. Cell biology. The case for RNA. *Science* **330**, 1185-1186 (2010).

81. Chappell, J. et al. The centrality of RNA for engineering gene expression. *Biotechnology Journal* **8**, 1379-1395 (2013).

82. 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**, 567-573 (2008).

83. Waldminghaus, T., Kortmann, J., Gesing, S. & Narberhaus, F. Generation of synthetic RNA-based thermosensors. *Biological Chemistry* **389**, 1319-1326 (2008).

84. Neupert, J., Karcher, D. & Bock, R. Design of simple synthetic RNA thermometers for temperature-controlled gene expression in Escherichia coli. *Nucleic Acids Research* **36**, e124 (2008).

85. Saragliadis, A., Krajewski, S.S., Rehm, C., Narberhaus, F. & Hartig, J.S. Thermozymes Synthetic RNA thermometers based on ribozyme activity. *Rna Biology* **10**, 1009-1016 (2013).

86. Bouvier, M. & Carpousis, A.J. A tale of two mRNA degradation pathways mediated by RNase E. *Molecular Microbiology* **82**, 1305-1310 (2011).

87. Condon, C. & Putzer, H. The phylogenetic distribution of bacterial ribonucleases. *Nucleic Acids Res* **30**, 5339-5346 (2002).

88. Cameron, J.C., Gordon, G.C. & Pfleger, B.F. Genetic and genomic analysis of RNases in model cyanobacteria. *Photosynthesis research*, 10.1007/s11120-11015-10076-11122 (2015).

89. Durfee, T. et al. The complete genome sequence of Escherichia coli DH10B: insights into the biology of a laboratory workhorse. *J Bacteriol* **190**, 2597-2606 (2008).

90. Engler, C., Kandzia, R. & Marillonnet, S. A One Pot, One Step, Precision Cloning Method with High Throughput Capability. *Plos One* **3**, e3647 (2008).

91. Moller, T. et al. Hfq: a bacterial Sm-like protein that mediates RNA-RNA interaction. *Molecular Cell* **9**, 23-30 (2002).

92. Zhou, K. et al. Novel reference genes for quantifying transcriptional responses of Escherichia coli to protein overexpression by quantitative PCR. *BMC molecular biology* **12**, 1-9 (2011).

93. Bustin, S.A. et al. The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. *Clinical chemistry* **55**, 611-622 (2009).

94. Willems, E., Leyns, L. & Vandesompele, J. Standardization of real-time PCR gene expression data from independent biological replicates. *Analytical biochemistry* **379**, 127-129 (2008).

95. Carpousis, A.J., Luisi, B.F. & McDowall, K.J. Endonucleolytic initiation of mRNA decay in Escherichia coli. *Progress in molecular biology and translational science* **85**, 91-135 (2009).

96. Caron, M.-P. et al. Dual-acting riboswitch control of translation initiation and mRNA decay. *Proceedings of the National Academy of Sciences of the United States of America* **109**, E3444-E3453 (2012).

97. Kuznetsov, S.V., Ren, C.C., Woodson, S.A. & Ansari, A. Loop dependence of the stability and dynamics of nucleic acid hairpins. *Nucleic Acids Res* **36**, 1098-1112 (2008).

98. Ansari, A., Kuznetsov, S.V. & Shen, Y. Configurational diffusion down a folding funnel describes the dynamics of DNA hairpins. *Proc Natl Acad Sci U S A* **98**, 7771-7776 (2001).

99. Jung, J. & Van Orden, A. A three-state mechanism for DNA hairpin folding characterized by multiparameter fluorescence fluctuation spectroscopy. *J Am Chem Soc* **128**, 1240-1249 (2006).

100. Bowman, G.R. et al. Structural insight into RNA hairpin folding intermediates. *Journal of the American Chemical Society* **130**, 9676-9678 (2008).

101. Zhang, W.B. & Chen, S.J. Exploring the complex folding kinetics of RNA hairpins: I. General folding kinetics analysis. *Biophysical Journal* **90**, 765-777 (2006).

102. Melnykov, A.V., Nayak, R.K., Hall, K.B. & Van Orden, A. Effect of loop composition on the stability and folding kinetics of RNA hairpins with large loops. *Biochemistry* **54**, 1886-1896 (2015).

103. Nocker, A. et al. A mRNA-based thermosensor controls expression of rhizobial heat shock genes. *Nucleic Acids Res* **29**, 4800-4807 (2001).

104. Rinnenthal, J., Klinkert, B., Narberhaus, F. & Schwalbe, H. Modulation of the stability of the Salmonella fourU-type RNA thermometer. *Nucleic Acids Res* **39**, 8258-8270 (2011).

105. Chowdhury, S., Maris, C., Allain, F.H. & Narberhaus, F. Molecular basis for temperature sensing by an RNA thermometer. *The EMBO journal* **25**, 2487-2497 (2006).

106. Draper, D.E. A guide to ions and RNA structure. *RNA (New York, N.Y.)* **10**, 335-343 (2004).

107. Cole, P.E., Yang, S.K. & Crothers, D.M. Conformational changes of transfer ribonucleic acid. Equilibrium phase diagrams. *Biochemistry* **11**, 4358-4368 (1972).

108. Oivanen, M., Kuusela, S. & Lonnberg, H. Kinetics and Mechanisms for the Cleavage and Isomerization of the Phosphodiester Bonds of RNA by Bronsted Acids and Bases. *Chemical reviews* **98**, 961-990 (1998).

109. Foster, J.W. *Escherichia coli* acid resistance: tales of an amateur acidophile. *Nature reviews. Microbiology* **2**, 898-907 (2004).

110. Richard, H. & Foster, J.W. *Escherichia coli* glutamate- and arginine-dependent acid resistance systems increase internal pH and reverse transmembrane potential. *J Bacteriol* **186**, 6032-6041 (2004).

111. Diez-Gonzalez, F. & Russell, J.B. The ability of *Escherichia coli* O157:H7 to decrease its intracellular pH and resist the toxicity of acetic acid. *Microbiology (Reading, England)* **143**, 1175-1180 (1997).

112. Henry, C.S., Broadbelt, L.J. & Hatzimanikatis, V. Thermodynamics-based metabolic flux analysis. *Biophys J* **92**, 1792-1805 (2007).

113. Lopez, P.J., Marchand, I., Joyce, S.A. & Dreyfus, M. The C-terminal half of RNase E, which organizes the *Escherichia coli* degradosome, participates in mRNA degradation but not rRNA processing *in vivo*. *Mol Microbiol* **33**, 188-199 (1999).

114. Lu, P., Vogel, C., Wang, R., Yao, X. & Marcotte, E.M. Absolute protein expression profiling estimates the relative contributions of transcriptional and translational regulation. *Nature Biotechnology* **25**, 117-124 (2007).

115. Gygi, S.P., Rochon, Y., Franza, B.R. & Aebersold, R. Correlation between protein and mRNA abundance in yeast. *Molecular and Cellular Biology* **19**, 1720-1730 (1999).

116. Wang, E., Han, B. & Li, S. Numerical simulation of transient radial temperature distribution in rotating drum bioreactor for solid state fermentation. *2013 International Conference on Materials for Renewable Energy and Environment*, 291-294 (2013).

117. Salis, H.M. The ribosome binding site calculator. *Methods Enzymol* **498**, 19-42 (2011).

118. Espah Borujeni, A., Channarasappa, A.S. & Salis, H.M. Translation rate is controlled by coupled trade-offs between site accessibility, selective RNA unfolding and sliding at upstream standby sites. *Nucleic Acids Res* **42**, 2646-2659 (2014).

119. Ingolia, N.T., Ghaemmaghami, S., Newman, J.R. & Weissman, J.S. Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science* **324**, 218-223 (2009).

120. Li, G.-W., Oh, E. & Weissman, J.S. The anti-Shine-Dalgarno sequence drives translational pausing and codon choice in bacteria. *Nature* **484**, 538-U172 (2012).

121. Carpousis, A.J., Luisi, B.F. & McDowall, K.J. Endonucleolytic Initiation of mRNA Decay in Escherichia coli. *Molecular Biology of Rna Processing and Decay in Prokaryotes* **85**, 91-135 (2009).

122. Mathews, D.H., Sabina, J., Zuker, M. & Turner, D.H. Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. *Journal of Molecular Biology* **288**, 911-940 (1999).

123. Giese, M.R. et al. Stability of RNA hairpins closed by wobble base pairs. *Biochemistry* **37**, 1094-1100 (1998).

124. Groebe, D.R. & Uhlenbeck, O.C. Characterization of RNA hairpin loop stability *Nucleic Acids Research* **16**, 11725-11735 (1988).

125. Takahashi, M.K. et al. Rapidly Characterizing the Fast Dynamics of RNA Genetic Circuitry with Cell-Free Transcription-Translation (TX-TL) Systems. *ACS Synth Biol*, 10.1021/sb400206c (2014).

126. Qi, L.S. & Arkin, A.P. A versatile framework for microbial engineering using synthetic non-coding RNAs. *Nature Reviews Microbiology* **12**, 341-354 (2014).

127. Bhadra, S. & Ellington, A.D. Design and application of cotranscriptional non-enzymatic RNA circuits and signal transducers. *Nucleic Acids Research* **42**, e58 (2014).

128. Green, A.A., Silver, P.A., Collins, J.J. & Yin, P. Toehold Switches: De-Novo-Designed Regulators of Gene Expression. *Cell* **159**, 925-939 (2014).

129. Galloway, K.E., Franco, E. & Smolke, C.D. Dynamically reshaping signaling networks to program cell fate via genetic controllers. *Science* **341**, 1235005 (2013).

130. Nissim, L., Perli, S.D., Fridkin, A., Perez-Pinera, P. & Lu, T.K. Multiplexed and Programmable Regulation of Gene Networks with an Integrated RNA and CRISPR/Cas Toolkit in Human Cells. *Molecular Cell* **54**, 698-710 (2014).

131. Saeidi, N. et al. Engineering microbes to sense and eradicate *Pseudomonas aeruginosa*, a human pathogen. *Mol Syst Biol* **7**, 521 (2011).

132. Farzadfard, F. & Lu, T.K. Synthetic biology. Genomically encoded analog memory with precise in vivo DNA writing in living cell populations. *Science* **346**, 1256272 (2014).

133. Copeland, M.F., Politz, M.C. & Pfleger, B.F. Application of TALEs, CRISPR/Cas and sRNAs as trans-acting regulators in prokaryotes. *Curr Opin Biotechnol* **29**, 46-54 (2014).

134. Yang, Y., Lin, Y., Li, L., Linhardt, R.J. & Yan, Y. Regulating malonyl-CoA metabolism via synthetic antisense RNAs for enhanced biosynthesis of natural products. *Metab Eng* **29**, 217-226 (2015).

135. Tummala, S.B., Welker, N.E. & Papoutsakis, E.T. Design of antisense RNA constructs for downregulation of the acetone formation pathway of *Clostridium acetobutylicum*. *J Bacteriol* **185**, 1923-1934 (2003).

136. Tsukuda, M., Nakashima, N. & Miyazaki, K. Counterselection method based on conditional silencing of antitoxin genes in *Escherichia coli*. *J Biosci Bioeng* **120**, 591-595 (2015).

137. Goh, S. et al. Silencing of Essential Genes within a Highly Coordinated Operon in *Escherichia coli*. *Appl Environ Microbiol* **81**, 5650-5659 (2015).

138. Phillips, M.I., Costales, J., Lee, R.J., Oliveira, E. & Burns, A.B. Antisense Therapy for Cardiovascular Diseases. *Current pharmaceutical design* **21**, 4417-4426 (2015).

139. Kole, R., Krainer, A.R. & Altman, S. RNA therapeutics: beyond RNA interference and antisense oligonucleotides. *Nature reviews. Drug discovery* **11**, 125-140 (2012).

140. Worley-Morse, T.O. & Gunsch, C.K. A computational analysis of antisense off-targets in prokaryotic organisms. *Genomics* **105**, 123-130 (2015).

141. Pulvermacher, S.C., Stauffer, L.T. & Stauffer, G.V. Role of the *Escherichia coli* Hfq protein in GcvB regulation of oppA and dppA mRNAs. *Microbiology (Reading, England)* **155**, 115-123 (2009).

142. Sakai, Y. et al. Improving the Gene-Regulation Ability of Small RNAs by Scaffold Engineering in *Escherichia coli*. *ACS Synth. Biol.* **3**, 152-162 (2014).

143. Sharma, V., Yamamura, A. & Yokobayashi, Y. Engineering Artificial Small RNAs for Conditional Gene Silencing in *Escherichia coli*. *ACS Synth. Biol.* **1**, 6-13 (2012).

144. Baba, T. et al. Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol Syst Biol* **2**, 1-11 (2006).

145. Urban, J.H. & Vogel, J. Translational control and target recognition by Escherichia coli small RNAs in vivo. *Nucleic Acids Res* **35**, 1018-1037 (2007).

146. Kim, T., Bak, G., Lee, J. & Kim, K.S. Systematic analysis of the role of bacterial Hfq-interacting sRNAs in the response to antibiotics. *The Journal of antimicrobial chemotherapy* **70**, 1659-1668 (2015).

147. Vogel, J. & Luisi, B.F. Hfq and its constellation of RNA. *Nature reviews. Microbiology* **9**, 578-589 (2011).

148. Coleman, J., Green, P.J. & Inouye, M. The use of RNAs complementary to specific messenger RNAs to regulats the expression of individual bacterial genes. *Cell* **37**, 429-436 (1984).

149. Park, H., Yoon, Y., Suk, S., Lee, J.Y. & Lee, Y. Effects of different target sites on antisense RNA-mediated regulation of gene expression. *BMB Rep* **47**, 619-624 (2014).

150. Gottesman, S. Micros for microbes: non-coding regulatory RNAs in bacteria. *Trends in genetics : TIG* **21**, 399-404 (2005).

151. Watters, K.E., Abbott, T.R. & Lucks, J.B. Simultaneous characterization of cellular RNA structure and function with in-cell SHAPE-Seq. *Nucleic Acids Res* **44**, e12 (2015).

152. Loughrey, D., Watters, K.E., Settle, A.H. & Lucks, J.B. SHAPE-Seq 2.0: systematic optimization and extension of high-throughput chemical probing of RNA secondary structure with next generation sequencing. *Nucleic Acids Res* **42**, e165 (2014).

153. Sowa, S.W. et al. Exploiting post-transcriptional regulation to probe RNA structures in vivo via fluorescence. *Nucleic Acids Res* **43**, e13 (2015).

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

155. Ma, K.C., Perli, S.D. & Lu, T.K. Foundations and Emerging Paradigms for Computing in Living Cells. *J Mol Biol* **428**, 893-915 (2016).

156. Bradley, R.W., Buck, M. & Wang, B. Tools and Principles for Microbial Gene Circuit Engineering. *J Mol Biol* **428**, 862-888 (2016).

157. Bradley, R.W. & Wang, B. Designer cell signal processing circuits for biotechnology. *New biotechnology* **32**, 635-643 (2015).

158. Rogers, J.K., Taylor, N.D. & Church, G.M. Biosensor-based engineering of biosynthetic pathways. *Curr Opin Biotechnol* **42**, 84-91 (2016).

159. Cress, B.F., Trantas, E.A., Ververidis, F., Linhardt, R.J. & Koffas, M.A. Sensitive cells: enabling tools for static and dynamic control of microbial metabolic pathways. *Curr Opin Biotechnol* **36**, 205-214 (2015).

160. Voigt, C.A. Genetic parts to program bacteria. *Curr Opin Biotechnol* **17**, 548-557 (2006).

161. Rogers, J.K. et al. Synthetic biosensors for precise gene control and real-time monitoring of metabolites. *Nucleic Acids Res* **43**, 7648-7660 (2015).

162. Rogers, J.K. & Church, G.M. Genetically encoded sensors enable real-time observation of metabolite production. *Proc Natl Acad Sci U S A* **113**, 2388-2393 (2016).

163. Raman, S., Rogers, J.K., Taylor, N.D. & Church, G.M. Evolution-guided optimization of biosynthetic pathways. *Proc Natl Acad Sci U S A* **111**, 17803-17808 (2014).

164. Wu, G. et al. Metabolic Burden: Cornerstones in Synthetic Biology and Metabolic Engineering Applications. *Trends Biotechnol* **34**, 652-664 (2016).

165. Teo, W.S. & Chang, M.W. Development and characterization of AND-gate dynamic controllers with a modular synthetic GAL1 core promoter in Saccharomyces cerevisiae. *Biotechnol Bioeng* **111**, 144-151 (2014).

166. Mukherjee, K., Bhattacharyya, S. & Peralta-Yahya, P. GPCR-Based Chemical Biosensors for Medium-Chain Fatty Acids. *ACS Synth Biol* **4**, 1261-1269 (2015).

167. Tabor, J.J., Levskaya, A. & Voigt, C.A. Multichromatic Control of Gene Expression in Escherichia coli. *Journal of Molecular Biology* **405**, 315-324 (2011).

168. Ogasawara, H. et al. Genomic SELEX search for target promoters under the control of the PhoQP-RstBA signal relay cascade. *J Bacteriol* **189**, 4791-4799 (2007).

169. Brandi, A., Spurio, R., Gualerzi, C.O. & Pon, C.L. Massive presence of the Escherichia coli 'major cold-shock protein' CspA under non-stress conditions. *The EMBO journal* **18**, 1653-1659 (1999).

170. Gualerzi, C.O., Giuliodori, A.M. & Pon, C.L. Transcriptional and post-transcriptional control of cold-shock genes. *J Mol Biol* **331**, 527-539 (2003).

171. Yamanaka, K., Mitta, M. & Inouye, M. Mutation analysis of the 5' untranslated region of the cold shock cspA mRNA of Escherichia coli. *J Bacteriol* **181**, 6284-6291 (1999).

172. Naryshkin, N., Revyakin, A., Kim, Y., Mekler, V. & Ebright, R.H. Structural organization of the RNA polymerase-promoter open complex. *Cell* **101**, 601-611 (2000).

173. Brophy, J.A.N. & Voigt, C.A. Principles of genetic circuit design. *Nature Methods* **11**, 508-520 (2014).

174. Davis, J.H., Rubin, A.J. & Sauer, R.T. Design, construction and characterization of a set of insulated bacterial promoters. *Nucleic Acids Res* **39**, 1131-1141 (2011).

175. Dong, H., Nilsson, L. & Kurland, C.G. Gratuitous overexpression of genes in Escherichia coli leads to growth inhibition and ribosome destruction. *J Bacteriol* **177**, 1497-1504 (1995).

176. Glick, B.R. Metabolic load and heterologous gene expression. *Biotechnol Adv* **13**, 247-261 (1995).

177. Grigorova, I.L., Phleger, N.J., Mutalik, V.K. & Gross, C.A. Insights into transcriptional regulation and sigma competition from an equilibrium model of RNA polymerase binding to DNA. *Proc Natl Acad Sci U S A* **103**, 5332-5337 (2006).

178. Kitano, H. Biological robustness. *Nature reviews. Genetics* **5**, 826-837 (2004).

179. Randall, A., Guye, P., Gupta, S., Duportet, X. & Weiss, R. Design and connection of robust genetic circuits. *Methods Enzymol* **497**, 159-186 (2011).

180. Alon, U. Network motifs: theory and experimental approaches. *Nature reviews. Genetics* **8**, 450-461 (2007).

181. Zhang, J., Yuan, Z., Li, H.X. & Zhou, T. Architecture-dependent robustness and bistability in a class of genetic circuits. *Biophys J* **99**, 1034-1042 (2010).

182. Meyer, A.J., Ellefson, J.W. & Ellington, A.D. Directed Evolution of a Panel of Orthogonal T7 RNA Polymerase Variants for in Vivo or in Vitro Synthetic Circuitry. *ACS Synth Biol* **4**, 1070-1076 (2015).

183. Yu, Y. et al. Development of Synechocystis sp. PCC 6803 as a phototrophic cell factory. *Marine drugs* **11**, 2894-2916 (2013).

184. Mitschke, J. et al. An experimentally anchored map of transcriptional start sites in the model cyanobacterium Synechocystis sp PCC6803. *Proceedings of the National Academy of Sciences of the United States of America* **108**, 2124-2129 (2011).

185. Stoeckel, J. et al. Global transcriptomic analysis of Cyanothece 51142 reveals robust diurnal oscillation of central metabolic processes. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 6156-6161 (2008).

186. Adams, B.L. The Next Generation of Synthetic Biology Chassis: Moving Synthetic Biology from the Laboratory to the Field. *ACS Synth Biol* (2016).

187. Yoneda, A. et al. Comparative transcriptomics elucidates adaptive phenol tolerance and utilization in lipid-accumulating Rhodococcus opacus PD630. *Nucleic Acids Res* **44**, 2240-2254 (2016).

188. Dominguez, A.A., Lim, W.A. & Qi, L.S. Beyond editing: repurposing CRISPR-Cas9 for precision genome regulation and interrogation. *Nature reviews. Molecular cell biology* **17**, 5-15 (2016).

189. Bikard, D. et al. Programmable repression and activation of bacterial gene expression using an engineered CRISPR-Cas system. *Nucleic Acids Res* **41**, 7429-7437 (2013).

190. Chappell, J., Takahashi, M.K. & Lucks, J.B. Creating small transcription activating RNAs. *Nat Chem Biol* **11**, 214-220 (2015).

191. Nielsen, A.A.K., Segall-Shapiro, T.H. & Voigt, C.A. Advances in genetic circuit design: novel biochemistries, deep part mining, and precision gene expression. *Curr. Opin. Chem. Biol.* **17**, 878-892 (2013).

192. Lou, C., Stanton, B., Chen, Y.-J., Munsky, B. & Voigt, C.A. Ribozyme-based insulator parts buffer synthetic circuits from genetic context. *Nature Biotechnology* **30**, 1137-+ (2012).

193. Bilitchenko, L. et al. Eugene--a domain specific language for specifying and constraining synthetic biological parts, devices, and systems. *PLoS One* **6**, e18882 (2011).

194. Nielsen, A.A. et al. Genetic circuit design automation. *Science* **352**, aac7341 (2016).

195. Carlson, R. The changing economics of DNA synthesis. *Nat Biotechnol* **27**, 1091-1094 (2009).

196. Kosuri, S. & Church, G.M. Large-scale de novo DNA synthesis: technologies and applications. *Nat Methods* **11**, 499-507 (2014).

197. Wang, H.H. et al. Programming cells by multiplex genome engineering and accelerated evolution. *Nature* **460**, 894-899 (2009).

198. Temme, K. et al. Induction and relaxation dynamics of the regulatory network controlling the type III secretion system encoded within Salmonella pathogenicity island 1. *J Mol Biol* **377**, 47-61 (2008).

199. Salis, H.M., Mirsky, E.A. & Voigt, C.A. Automated design of synthetic ribosome binding sites to control protein expression. *Nat Biotechnol* **27**, 946-950 (2009).

200. Hayashi, K. et al. Highly accurate genome sequences of Escherichia coli K-12 strains MG1655 and W3110. *Molecular Systems Biology* **2**, 10.1038/msb4100049 (2006).

201. Cox, R.S., 3rd, Dunlop, M.J. & Elowitz, M.B. A synthetic three-color scaffold for monitoring genetic regulation and noise. *Journal of biological engineering* **4**, 10 (2010).

202. Stover, C.K. et al. Complete genome sequence of Pseudomonas aeruginosa PAO1, an opportunistic pathogen. *Nature* **406**, 959-964 (2000).

203. Anderson, J.C., Voigt, C.A. & Arkin, A.P. Environmental signal integration by a modular AND gate. *Mol Syst Biol* **3**, 133 (2007).