Engineering RNA Devices for Gene Regulation, Biosensing, and Higher-Order Cellular Information Processing

Thesis by

Maung Nyan Win

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

California Institute of Technology

Pasadena, California

2008

(Defended May 12, 2008)

Acknowledgements

I would first and foremost like to thank my research advisor, Christina Smolke, for her continuous support and creative and insightful advice provided to me throughout my graduate study time at Caltech. Her research enthusiasm, scholarship, and dedication has inspired and challenged me as a researcher. Her guidance has been invaluable in helping me accomplish my thesis projects. It has been a great honor and a distinct pleasure to work with her at both professional and social levels.

I would also like to thank my thesis committee members, Judith Campbell, Carl Parker, and Anand Asthagiri, for their helpful comments and inputs. I have been privileged and enjoyed working in an environment that is pleasant, wonderful, and full of friendly professionals. Of course, the environment spans from our own laboratory to the Division of Chemistry and Chemical Engineering and the entire Caltech campus. I am grateful to have both professional and social support from my colleagues in the Smolke group. Thanks to Kristy Hawkins and Andrew Babiskin for frequently providing their 'ready-made' plasmid vectors, which saved me from slaving over a few clonings. I have also enjoyed the company of a few 'lunch and coffee' friends, especially, Andrew Babiskin, Joe Liang, and Leo d'Espaux; without their wonderful and fun company, the five-year period would have felt twice as long.

I would like to show my great appreciation to my cousin and her husband, Janice and David Lye. I came to the United States about nine years ago from Burma and started my undergraduate study at Virginia Commonwealth University while staying at my cousin and her husband's house. Without their help and support, I would not have become a graduate student at Caltech. I would also like to attribute this to my undergraduate university and distinct professors, Rachel Chen, Gary Huvard, and Gary Wnek, who have laid a foundation

for my graduate study at Caltech. I would like to acknowledge my partner, Steve Fitzgerald, who has been a wonderful companion for my graduate career at Caltech, and I am very much thankful to have him as my partner.

Finally, none of my achievements would have been remotely possible without the help and support of my parents. I am deeply, deeply grateful for everything that they have done for me.

Engineering RNA Devices for Gene Regulation, Biosensing, and Higher-order Cellular Information Processing

Maung Nyan Win

B.S., Virginia Commonwealth University

M.S., California Institute of Technology

Ph.D., California Institute of Technology

Abstract

The proper regulation of gene expression is critical to many biological processes occurring in the cell. It is becoming increasingly apparent that post-transcriptional processing pathways play significant roles in regulating the expression of various genes in both prokaryotic and eukaryotic organisms, where they direct a variety of complex cellular functions. A striking example of a biological communication and control system directing sophisticated gene expression regulation through precise molecular recognition is the class of RNA regulatory elements, called riboswitches, comprised of distinct sensor (ligand-binding) and actuator (gene-regulatory) functions that control gene expression in response to changing levels of specific target ligand concentrations.

Inspired by these natural examples, numerous synthetic riboswitch systems have been developed and have made profound contribution to the field of riboswitch engineering. However, these early examples of synthetic riboswitches pose one or more challenges, such as portability of the switch design across different cellular systems and modularity and programmability of the components comprising the switch molecule. Therefore, we set out to develop a modular and extensible RNA-based gene-regulatory platform that will provide a

framework for the reliable design and construction of gene regulatory systems that can control the expression of specific target genes in response to effector molecules of interest. The platform is called the "ribozyme switch" and composed of distinct functional components, which are modularly coupled and functionally independent of each other. Through this platform, ribozyme switch devices that enable up- or down-regulation of target gene expression were developed. Design modularity and response programmability of the switch platform were also demonstrated. We also exhibited the versatility of the platform in implementing application-specific control systems for small molecule-mediated regulation of cell growth and non-invasive *in vivo* sensing of metabolite production.

Through the ribozyme switch platform, we further constructed higher-order RNA devices that enable complex cellular information processing operations, including logic control (AND, NOR, and NAND gates), advanced computation (bandpass filter and signal shift in the output swing), and cooperativity (signal gain). Finally, we extended the small ribozyme switch platform responsive to small molecules to a different class of ligand molecules, proteins, by developing protein-responsive gene regulators and cellular biosensors. In addition to engineering RNA devices for programming cellular function, we also developed a high-throughput method for functional characterization of small molecule-binding RNA aptamers, which enables robust, accurate, and rapid characterization of such RNA aptamers. This method can be very useful as we (and others) develop RNA aptamers for small molecules of specific interest, which can be subsequently integrated into the ribozyme switch platform as sensing elements for specific applications. Together, these research developments hold synergistic values for the reliable construction of 'designer' gene-regulatory systems for various biotechnological and medical applications.

Table of Contents

List of Tables

List of Figures and Supplementary Figures

Chapter I: Introduction*

The proper regulation of gene expression is critical to many biological processes occurring in the cell. It is becoming increasingly apparent that post-transcriptional processing pathways play significant roles in regulating gene expression in both prokaryotic and eukaryotic organisms where they direct a variety of complex cellular functions. The directions regulating these processes are encoded within the genome, and many of them act at the level of *cis*- or *trans*-acting RNA elements that bind a wide range of biological molecules such as DNA, proteins, and other RNA molecules. Recent discoveries reveal larger roles of RNA as a sophisticated gene-regulatory molecule, which can implement diverse regulatory forms, such as riboswitches[1-5] and ribozymes[6, 7], in regulating target gene expression. Therefore, RNA represents an attractive platform for the development of novel gene-regulatory tools for a variety of biotechnological and medical applications and present opportunities for the modular design and construction of synthetic RNA switches and sensors to program different cellular functions.

1.1. RNA as a versatile and powerful gene-regulatory element

Ribonucleic acid (RNA) has traditionally been recognized for its role as a passive messenger of genetic information between the genome and the proteome in all living organisms. However, there are an increasing number of discoveries of naturally-occurring RNA molecules that act as regulatory elements, performing various cellular functions including gene expression regulation through sophisticated mechanisms, thereby expanding its traditional role as a genetic messenger and revealing it as a functionally versatile molecule.

*Sections 1.1.-1.4: Reproduced/adapted with permission from: M. N. Win and C. D. Smolke. (2006) "Regulating Gene Expression through Engineered RNA Technologies". Metabolic Engineering Protocols: Synthesis and Design Strategies. In: Walker J, series ed., Methods in Molecular Biology. Totawa: Humana Press. (book chapter in preparation)

Riboswitches[1-5], ribozymes[6, 7], antisense RNAs[8, 9], and small interfering and microRNAs (siRNAs and miRNAs, respectively)[10-13] are examples of RNA elements that exert their regulatory effects at different levels of the gene expression pathway such as transcription, translation, splicing, or decay. Unlike messenger RNAs (mRNAs), these regulatory RNAs are often noncoding RNAs (ncRNAs) or do not encode protein information. In addition, these regulatory elements are implemented through diverse physical compositions that can be grouped generally into *cis*- and *trans*-acting elements. In the former composition, the regulatory element is integrated within the transcript that harbors the target gene, whereas in the latter composition the regulatory element is a separate RNA molecule that acts on the transcript harboring the target gene through RNA-RNA binding interactions between the two individual molecules.

RNA exhibits a wide variety of functional properties, including catalytic, gene-regulatory, and ligand-binding activities. In addition, integrated RNA regulatory molecules have been characterized that achieve more sophisticated control over the expression of target proteins through a combination of these functional properties. These functional properties are encoded within the nucleotide sequence of an RNA molecule, which subsequently dictates its secondary and tertiary structure and ultimately its function. RNA adopts different conformations by folding into secondary and tertiary structures, which interact with various cellular constituents such as DNA, proteins, small molecules, and other RNA molecules[14, 15]. Furthermore, RNA molecules exhibit structural flexibility, which enable them to dynamically adopt different conformations. The binding of cellular and environmental molecules to particular RNA conformations has been demonstrated to regulate the equilibrium distribution between stable conformational states[16-18]. Unlike larger biomolecules such as proteins, the

functional activity of RNA is more directly defined by its secondary structure. This relationship between RNA secondary structure and function, in combination with predictive RNA secondary structure / energetic folding programs and rational and/or combinatorial design strategies, has enabled molecular engineers to construct synthetic 'designer' regulatory RNA elements[17-19]. In addition, technological advances in RNA engineering have demonstrated the programming of regulatory properties through alteration of nucleotide composition and ultimately RNA structure-function relationships[17-19].

Recent research in RNA biology and engineering supports the model of RNA as a versatile and powerful molecule possessing biologically-relevant gene-regulatory properties. Advances in RNA technology and nucleic acid engineering have allowed researchers to apply naturally-occurring RNAs as basic regulatory platforms and to develop more sophisticated regulatory RNA elements that involve integrated designs of multiple platforms. These synthetic riboregulators enable gene expression to be regulated in a more controlled manner and represent powerful tools for fundamental research and exhibit important applications in biotechnology and medical research.

1.2. RNA as sensory elements that exhibit universal sensing/binding properties

RNA molecules are functionally diverse and structurally flexible, and exhibit a wide range of regulatory properties such as catalytic, interactive, and allosteric binding properties. A unique property of RNA molecules is their sensing/binding capability to various types of inputs ranging from temperature to varied molecular ligands. Similar to other regulatory properties, RNA employs this property in exerting its diverse functional roles.

1.2.1. RNA elements that serve as thermosensors

RNA secondary structure is known to be highly dependent on temperature such that RNA assumes different structures in response to changes in temperature. Examples of such temperature-responsive RNA elements have been described[20]. For instance, genes that encode small heat-shock proteins and regulators of heat shock-responsive genes were found to contain sequences in their 5' untranslated regions (UTRs) that are capable of sequestering the Shine-Dalgarno (SD) sequence or prokaryotic ribosomal binding site (RBS) and the start codon AUG. The dependent adoption of RNA secondary structures at different temperatures is used to regulate the accessibility of these sequences by the ribosome, which subsequently modulates target gene expression levels, thereby enabling these RNA elements to serve as thermosensing gene expression regulators.

1.2.2. RNA elements that bind nucleic acids

Instances where RNA molecules exert their regulatory activities through base-pairing interactions with other RNA molecules are widespread in both natural and synthetic biological systems. For instance, a class of RNA molecules, called antisense RNAs, are single-stranded, *trans*-acting non-coding RNA elements, whose sequences are complementary to target transcripts and usually consist of 12-20 complementary nucleotides[8]. These RNA regulatory elements bind to their target transcripts through a sequence-specific hybridization event, resulting in the inhibition of gene expression from the bound mRNA[21]. As a natural example, SgrS is an antisense RNA that down-regulates the expression of the glucose transporter through base-pairing with the *ptsG* transcript when localized to the cell membrane, thereby lowering the accumulation of toxic phosphosugar metabolites[22, 23]. A

good synthetic example of RNA-RNA interactions through which a gene-regulatory function was achieved was described by Isaacs et al.[24]. In their engineered system, an RNA sequence segment was integrated into a location upstream of the RBS of a reporter gene to serve as a nucleic acid-sensing domain. This segment sequesters the RBS in the absence of the target RNA molecule, thereby suppressing the expression of the reporter gene. In the presence of the target RNA molecule, the sensor domain becomes bound to the target through RNA-RNA base-pairing interactions and releases the RBS for efficient translation, thereby serving as a nucleic acid-binding sensor domain.

1.2.3. RNA elements that bind molecular ligands

The ability of RNA structural elements to bind specific molecular ligands has been characterized in several natural systems[1]. However, researchers have also generated many examples of synthetic RNA ligand-binding elements, referred to as aptamers, in the laboratory. Synthetic aptamers can be generated through a standardized *in vitro* selection process known as SELEX (Systematic Evolution of Ligands by EXponential enrichment)[25, 26] (Figure 1.1). SELEX provides a very powerful selection method through which nucleic acid molecules exhibiting rare and specific binding properties to a ligand of interest can be generated *de novo* by selecting for functional binding activities from large randomized nucleic acid pools through iterative *in vitro* selection and amplification cycles. *In vitro* RNA aptamer selection schemes begin with a large pool of single-stranded RNA molecules generated through *in vitro* transcription from a DNA library. Aptamer pools are usually comprised of 30-70 randomized nucleotides, in order to generate an initial sequence diversity between 10^{14} to 10^{15} molecules[27]. The pool is incubated with the target ligand of interest and

subject to a partitioning event to separate bound members from unbound members. The most commonly used partitioning schemes are based on affinity chromatography. Bound (functional) members are recovered and then amplified through reverse transcription and polymerase chain reaction (PCR) to yield a pool enriched for target binding. This enriched pool will become the input pool for the next round of selection.

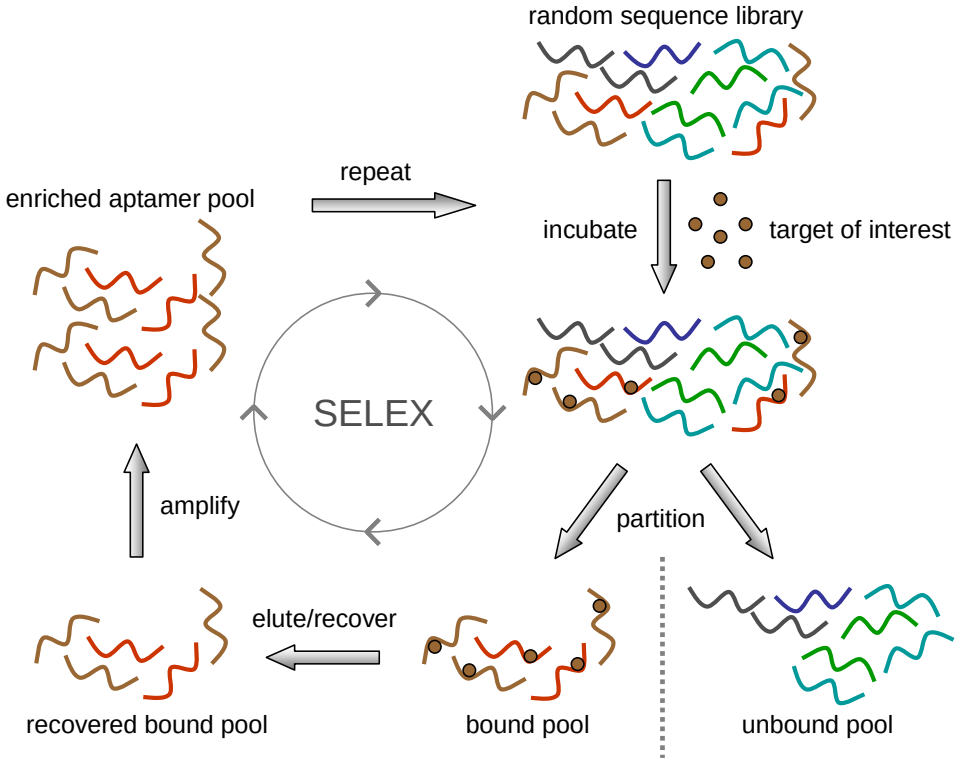


Figure 1.1. A schematic illustration of an *in vitro* selection process known as SELEX. The process starts with a large randomized pool of single-stranded RNA molecules transcribed from their DNA templates. The RNA pool is then incubated with target molecules of interest followed by separation of target-bound pool, which is reverse-transcribed to cDNA and amplified for the next selection cycle. The selection cycle is repeated typically for 8-15 cycles.

SELEX is particularly powerful in that RNA sequences that bind particular ligands can be generated *de novo* and the binding properties of the resulting aptamers can be programmed as desired. Specifically, aptamer binding properties such as affinities and specificities can be programmed by tailoring the stringency and counter-selections during

each selection. Aptamer affinities are tailored through the stringency of each selection cycle, normally by modifying wash volumes and target concentrations, whereas aptamer specificities are tailored through performed counter-selections with molecular analogues to the target. Typically eight to fifteen selection cycles are required to generate aptamers with high binding affinities and specificities. Recent work has demonstrated that protein aptamer selection schemes can be automated using standard robotics[28-30]. In addition, partitioning schemes for protein aptamer selections based on capillary electrophoresis have been recently developed that provide several advantages over conventional affinity-based partitioning schemes[31-33]. In particular, the efficiency of separation between the bound and unbound pools is significantly greater such that aptamers can be generated in fewer selection cycles. Synthetic aptamers have been generated to a wide range of target ligands, including small molecules, antibiotics, carbohydrates, amino acids, peptides, proteins[34], and even organelles such as phospholipid bilayers[35, 36], indicating that synthetic aptamers can be potentially generated to any targets of interest for user-specific applications.

1.3. Riboswitches are ligand-responsive RNA regulators of gene expression

All living organisms must manage the expression of many different genes in response to different signals such as metabolic demands and environmental changes[37]. This type of genetic management requires highly responsive sensors that accurately measure the magnitude of a particular signal and subsequently modulate the amount of the appropriate gene products to be synthesized. Proteins have traditionally been viewed as being the responsible sensor molecules for these signals. However, recent discoveries have

demonstrated that elements within mRNAs, termed riboswitches, are also capable of performing such tasks.

Riboswitches are *cis*-acting RNA elements that modulate the expression of target genes through integrated sensor and regulatory domains. This integration scheme enables riboswitches to sense their target ligands, typically cellular metabolites, through direct binding interactions and thus autonomously mediate their own functional activity in response to changing metabolite levels. While the majority of riboswitches characterized to-date have been discovered in bacteria, it has been shown more recently that these complex RNA regulatory elements are also present in eukaryotes[5, 38]. Riboswitches exhibiting unique mechanistic properties have also been recently identified. For example, the glycine riboswitch exhibits cooperative binding to its metabolite product, in which the metabolite binding turns on the gene expression of the enzyme responsible for the glycine cleavage system[3]. This cooperative-binding feature is present in the glycine riboswitch and is proposed to ensure that the metabolite is indeed in excess after consumption to provide carbon flux through the citric acid cycle while maintaining sufficient amounts of the amino acid available for protein synthesis. In another example, a tandem riboswitch system was characterized that exhibits a Boolean logic ability and functions as a two-input NOR logic gate, in which the two ligands (S-adenosylmethionine and coenzyme B_{12}) can independently suppress the target gene expression by binding their corresponding aptamers located upstream of a structure resembling an intrinsic transcription terminator[4]. Another tandem riboswitch system was very recently discovered that consists of two distinct riboswitches[39]. Unlike the glycine riboswitch and the two input-responsive logic-gate riboswitch systems, this riboswitch system does not exhibit cooperative ligand binding or detect two different metabolites,

respectively. This tandem riboswitch system responds independently to the same metabolite, thiamine pyrophosphate (TPP), and is predicted to function in concert to yield a more 'digital' gene control output response than a single riboswitch system.

1.3.1 General composition and conformational dynamics of riboswitches

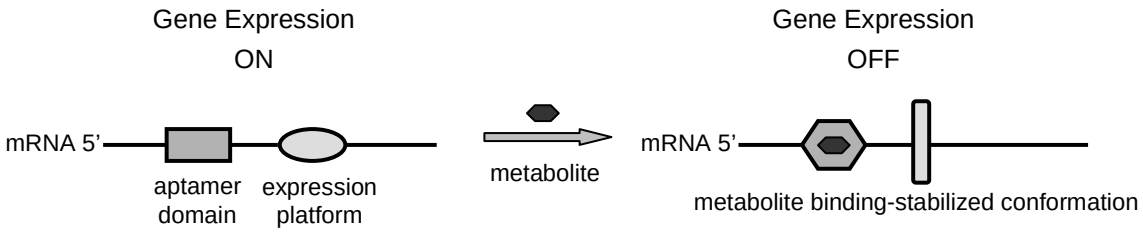


Figure 1.2. A schematic diagram of a typical riboswitch composed of two distinct domains: the ligand-binding domain known as the aptamer domain and the regulatory domain known as the expression platform (adapted from Winkler and Breaker[37]). Metabolite binding to the aptamer domain enables the stabilization of the rearranged conformation of the riboswitch (right), resulting in a shift in the equilibrium distribution between the two regulatory conformations that leads to the metabolite-dependent regulation of target gene expression.

Riboswitches are naturally-occurring, metabolite-responsive gene control elements primarily located within the 5' UTRs of cellular transcripts[37]. A riboswitch is typically comprised of two domains: the ligand-binding or sensor domain known as the aptamer domain, and the gene-regulatory domain known as the expression platform (Figure 1.2)[37]. Both domains are structurally flexible and capable of adopting different conformations. Riboswitches accomplish ligand-controlled regulation of gene expression through targeted dynamic switching between two primary conformations at equilibrium: one in which the regulatory domain is active and the other in which it is inactive. One of these conformational states is associated with the formation of the ligand-binding pocket within the aptamer domain, whereas the other carries an incorrectly formed binding pocket. Therefore, riboswitches can either repress or activate the expression of the target gene by assuming an

appropriate combination of different conformational states adopted by the aptamer and regulatory domains. Ligand binding to the riboswitch shifts the equilibrium distribution between these stable conformations to favor the ligand-bound form, thereby resulting in an allosteric gene regulation event. Most of the riboswitches characterized to date down-regulate the expression of the target gene; however, in a few exceptions such as the glycine[3] and adenine riboswitches[40], target gene expression is activated upon binding of the metabolite-ligand to the riboswitch.

1.3.2. Mechanisms of ligand-controlled gene regulation by riboswitches

Riboswitches regulate target gene expression *in cis* in response to changing metabolite levels through different mechanisms involving transcription termination[40, 41], translation initiation[42], mRNA processing[43], and splicing[5] (Figure 1.3). Transcription termination takes place through the mediated formation of a rho-independent terminator stem, which is usually GC rich, thereby destabilizing the transcription elongation complex[41]. Regulation can also target the disruption of the formation of a terminator stem upon metabolite binding, which allows proper transcription and thus up-regulation of target expression levels[40]. Riboswitches can also mediate translation initiation by undergoing adopting a secondary structure that interferes with ribosomal access to the target gene, such as sequestering the ribosome-binding site (RBS) or Shine-Dalgarno (SD) sequence in prokaryotic cells[42]. Regulation targeting transcript processing or deactivation can be achieved through expression platforms comprised of self-cleaving ribozymes, where the target transcripts undergo a ligand-directed cleavage event[43]. Regulation through splicing has recently been demonstrated in a filamentous fungus, in which metabolite binding to its

riboswitch can either repress or activate the expression of the main protein product by modulating the splice site choice through structural rearrangements[5]. Metabolite-binding domains have also been found within the 3' UTRs of transcripts in certain organisms, suggesting that riboswitch-mediated gene control may also occur through the regulation of mRNA stability[38].

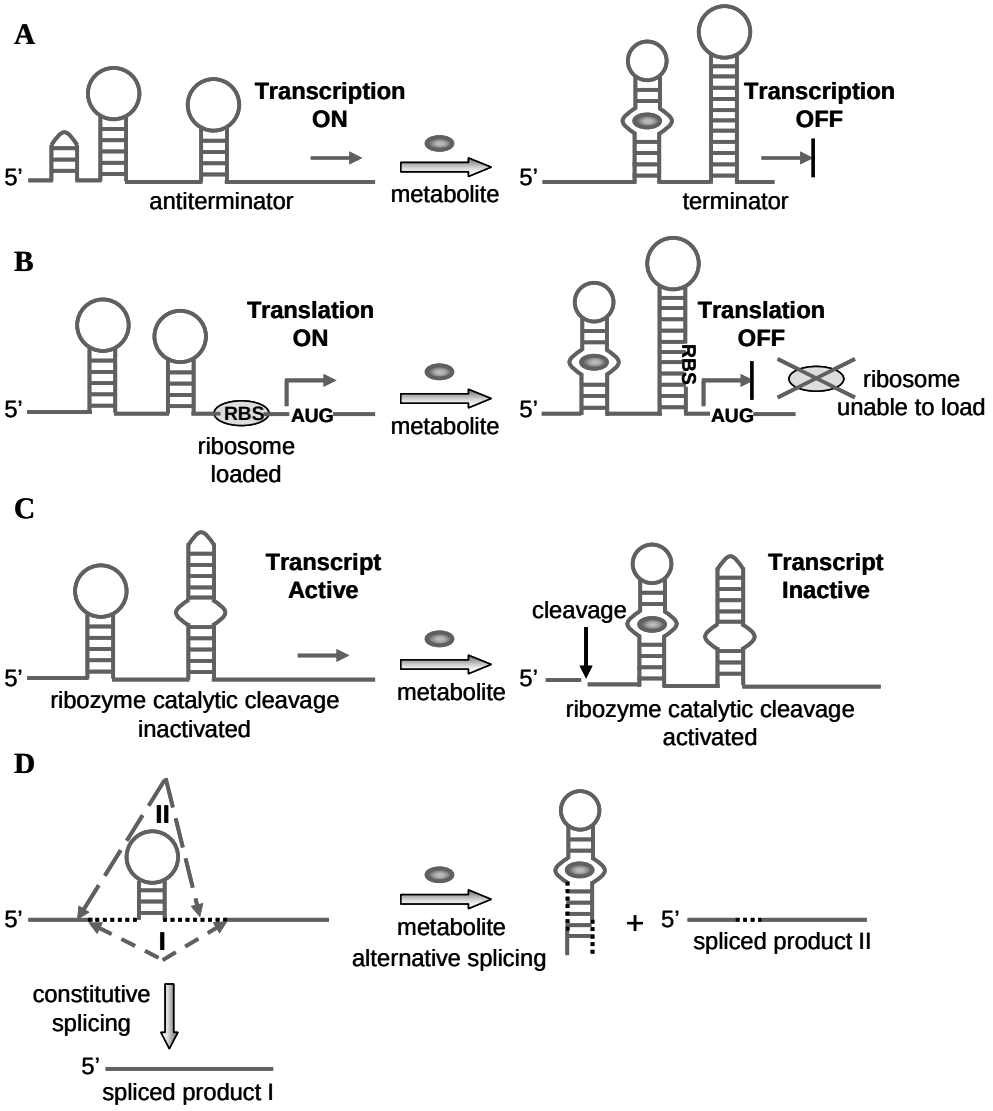


Figure 1.3. A schematic illustration of mechanisms by which riboswitches achieve gene expression regulation in response to their target metabolite binding (A and B adapted from Nudler and Mironov[44]). Ligand-regulated mechanisms involve (A) the formation of a transcription terminator stem, (B) sequestering the RBS and inhibiting translation initiation, (C) mRNA processing through catalytic cleavage of the transcript, and (D) alternative splicing using different sets of splice sites, I and II, respectively.

1.3.3. Riboswitch targets and implementation in metabolic networks

Most of the riboswitches characterized to-date have been discovered in bacteria. Cells employ these elements as genetic regulators in many fundamental metabolic pathways in response to changing metabolite levels. Target metabolites include various classes of small molecules such as amino acids, nucleotide bases, and coenzymes[45]. The presence of the integrated sensor domain enables riboswitches to sense intracellular metabolite concentrations through specific binding interactions and subsequently regulate expression levels of the associated gene product through allosteric conformational changes. Typically, this gene product is an enzyme directly involved in the biosynthesis, biodegradation, and/or transport of the target metabolite[45]. This mode of regulation provides a direct dynamic relationship between the intracellular metabolite concentration and the expression levels of the enzyme responsible for the metabolism, catabolism, and transport of the target metabolite. In addition, riboswitches are capable of binding their target metabolites with high specificities and affinities. Several different classes of natural riboswitches, their corresponding target metabolites, and functional roles in metabolic networks in various organisms are reviewed elsewhere in detail[45].

1.4. Current synthetic riboswitch systems for ligand-mediated regulation of target gene expression

Riboswitches are sophisticated gene control elements that achieve regulation by direct sensing of target metabolite levels and exhibit molecular recognition, high affinities, and precise control. Examples of the level of complexity achieved by these genetic regulatory elements include the self-cleaving ability of the *glmS* riboswitch[43], the alternative splicing

control of the TPP riboswitch[39], the cooperative binding property of the glycine riboswitch[3], and the NOR gate signal processing behavior of the SAM - coenzyme B_{12} riboswitch[4], demonstrating that riboswitches are powerful sensor-actuator control systems for autonomous gene expression control.

Inspired by the natural examples of riboswitches, numerous synthetic riboswitch systems have been engineered for artificial, ligand-mediated control of gene expression of target mRNAs. Many recent engineering efforts have focused on the construction of such synthetic ligand-controlled RNA-based gene regulatory elements through the integration of sensor and regulatory domains[18]. The flexibility in RNA regulatory systems, the programmability inherent in RNA design strategies, and the ability to generate sensor domains to potentially any molecular ligand of interest, enable such synthetic riboswitch systems to hold significant promise in transforming our ability to engineer cellular function. Therefore, engineered riboswitch control elements are attractive molecular tools for applications in synthetic network design where they can serve as powerful regulators of expression levels of genes of interest as well as *in vivo* biosensors of their corresponding target ligands through mechanisms of molecular recognition that go beyond base pairing. An overview of various engineered riboswitch systems that employ diverse gene-expression regulatory platforms is provided below to highlight the current technologies and challenges in the field of constructing synthetic riboswitches.

1.4.1. Riboswitch construction based on aptamer insertion within a transcript

Synthetic riboswitches have been constructed by inserting an aptamer or multiple aptamers directly into the 5’ UTR of a target mRNA in eukaryotes (Figure 1.4). Although the insertion location may be any region in the 5’ UTR, it is often chosen to be in the vicinity of

the cap region or the start codon of the transcript. This insertion strategy is a trail-and-error strategy, as the inserted aptamer(s) often does not result in ligand-mediated regulation of gene expression and may even cause substantial knockdown of the target gene in the absence of ligand. This strategy requires that the insertion of the aptamer itself and its associated secondary structure does not interfere with translation in the absence of the target ligand. The binding of ligand to the aptamer results in structural stabilization due to the molecular binding interaction between the aptamer and its target[34, 46]. Similar to binding of a protein to the 5' UTR[47, 48], this stabilized secondary structure can repress translation[49] presumably by interfering with ribosomal scanning or ribosome-mRNA interactions required for effective translation.

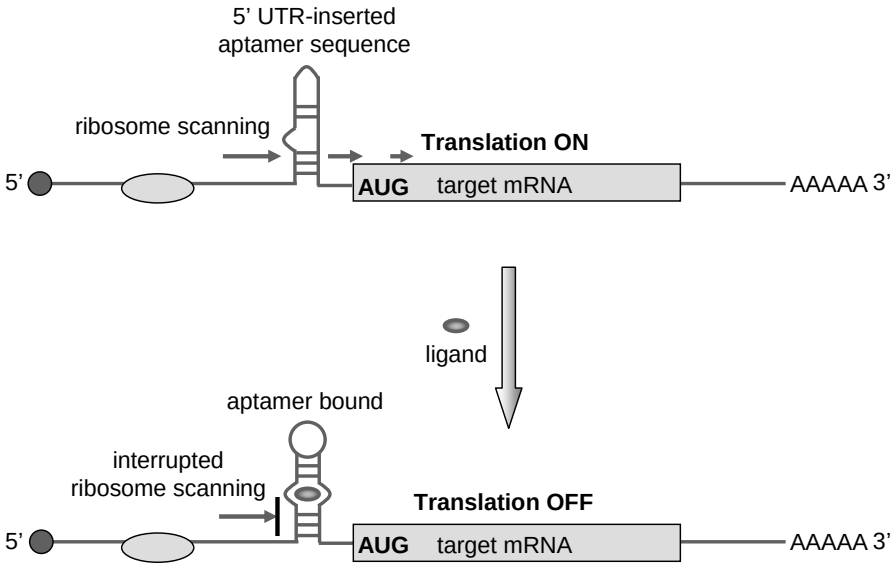


Figure 1.4. A schematic illustration of riboswitches constructed based on aptamer insertion within a target transcript. An aptamer or multiple aptamers can be inserted into the 5' untranslated region of an mRNA transcript near the 5' cap region or the start codon. Such an insertion may allow the aptamer-fused transcript region to adopt two primary conformations, one in which the aptamer binding pocket is disrupted (top) and the other in which the aptamer is correctly formed to reside its ligand (bottom), and the ligand binding shifts the equilibrium towards the latter conformation by stabilizing this conformation. This insertion strategy yet requires that the former conformation not introduce steric hindrance to ribosome for proper translation and that the ligand-bound latter conformation effectively inhibit translation through its stabilized structure.

Werstuck and Green[50] constructed the first examples of such riboswitches by inserting small molecule-binding RNA aptamers into the 5' UTR of transcripts. Translation was demonstrated to be repressed by the addition of the appropriate ligands both *in vitro* and *in vivo* in mammalian cells. Following this initial work, different research groups have constructed synthetic riboswitches through this design strategy using theophylline-[51], biotin-[52], and tetracycline-binding[53] aptamers, and demonstrated similar ligand-controlled gene regulation in different systems, including wheat germ extracts[54], *Xenopus oocyte*[54], and the budding yeast *S. cerevisiae*[55, 56].

Synthetic riboswitches have also been constructed to regulate translation of target genes in prokaryotes through a similar aptamer insertion strategy. Although still located in the 5' UTR of the target transcript, prokaryotes do not exhibit the same type of ribosomal scanning as eukaryotic organisms. Therefore, the physical implementation of these switches requires slightly different design strategies. In bacteria, the sequence distance between the ribosomal binding site (RBS) and the start codon is relatively short and varies between 5 to 13 nucleotides[57]. As a result, targeted insertion of an aptamer in this region to interfere with ribosomal scanning through a ligand-induced secondary structure is generally not applicable. In most bacteria, translation initiation relies on ribosomal accessibility to the RBS and the start codon, and thus mRNA secondary structure in the translational initiation region can dictate the efficiency of translation[58-60]. Desai and Gallivan developed a synthetic riboswitch system in *E. coli* where they inserted the theophylline aptamer[51] to a location five base-pairs upstream of the RBS to modulate ribosomal access to the RBS through ligand binding[61]. The theophylline-dependent up-regulation of gene expression by this synthetic riboswitch was demonstrated through plate-based screening and liquid culture assays. In addition, the extent

of up-regulation from this synthetic riboswitch was observed to be dramatically affected when the aptamer was moved to a slightly different location, two or eight base-pairs upstream of the RBS, indicating the functional sensitivity of this riboswitch system to the aptamer insertion location.

1.4.2. Riboswitch construction based on direct attachment of the aptamer to a regulatory element

Synthetic riboswitches have been constructed by directly attaching an aptamer to the regulatory domain, such that ligand binding to the aptamer inhibits the activity of the regulatory domain through some mechanism (Figure 1.5). This construction strategy is also a trial-and-error strategy, since the desired ligand-responsive regulatory activity may be highly specific to the location of attachment, the mechanism of action, and the specific aptamer-regulator pair.

A riboswitch system based on this strategy was developed by An et al. to modulate Dicer processing of an shRNA molecule through a small molecule ligand-aptamer interaction within this RNAi substrate[62]. In this example, the theophylline aptamer[51] was directly fused to an shRNA molecule, in place of the loop sequence. This shRNA construct silenced reporter gene expression in mammalian cells in a theophylline dose-dependent manner. Dicer cleavage of the aptamer-fused shRNA for subsequent generation of siRNAs was demonstrated to be modulated *in vitro* and *in vivo* by theophylline. This ligand-mediated regulation of Dicer processing was likely achieved due to locating the ligand-binding site of the aptamer sufficiently close to the Dicer processing site, such that theophylline binding to its aptamer blocks Dicer cleavage of the shRNA molecule, resulting in regulatable siRNA-based gene silencing. This proposed mechanism is supported by the observation that the

ligand-mediated regulatory effect was abolished when the shRNA stem was extended by one or two base-pairs, resulting in a small shift in the aptamer fusion point compared to the initial fusion design.

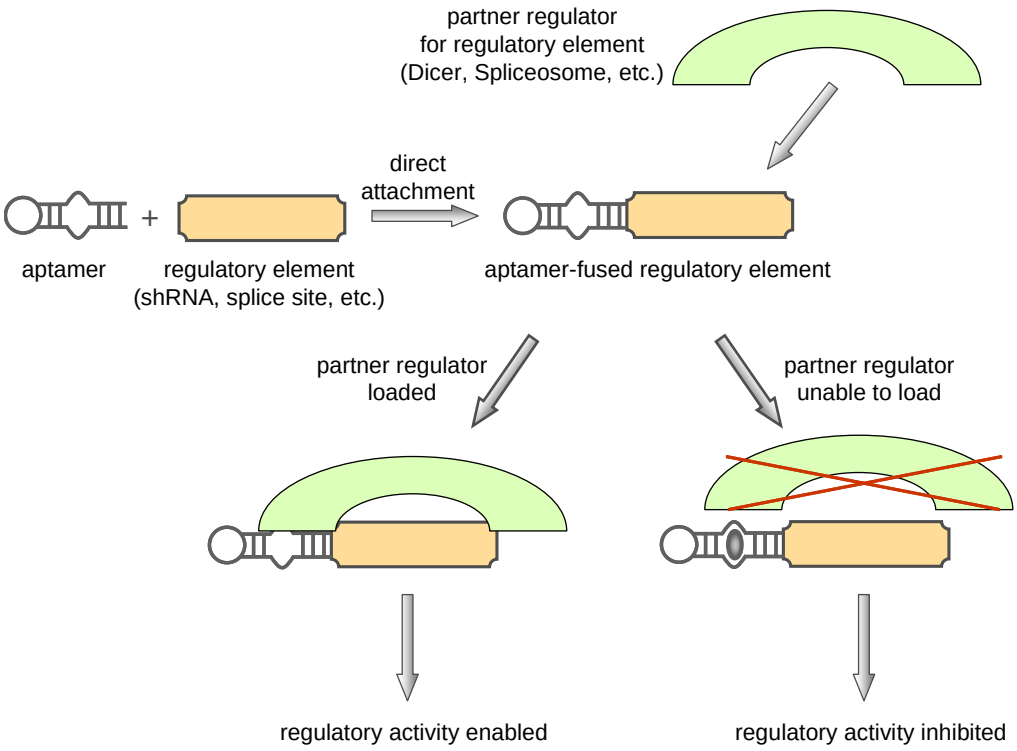


Figure 1.5. A schematic illustration of riboswitches constructed based on direct attachment between the aptamer and regulatory domains. An aptamer is directly attached to the regulatory platform in a way that the ligand-binding pocket within the aptamer is sufficiently close to the regulatory platform. In the absence of ligand, the partner regulatory element is capable of loading onto its platform and enabling the corresponding regulatory event to occur. In the presence of ligand, ligand binding to the aptamer and residing within the binding pocket create steric hindrance for the partner regulatory element loading to its platform, resulting in the inhibition of the normal regulatory event.

This direct attachment strategy has also been employed in constructing a synthetic riboswitch that regulates gene expression at the level of splicing. This riboswitch was designed by insertion of the theophylline aptamer[51] near the 3' consensus splice site region of a model pre-mRNA to modulate the splicing of the pre-mRNA through ligand-aptamer complex interactions[63]. The addition of theophylline was shown to repress the *in vitro*

splicing of the pre-mRNA harboring the aptamer and have no regulatory effect on the pre-mRNA without the aptamer. In addition, the aptamer's effect on splicing was demonstrated to be location-dependent and explained by modulating the spliceosome's accessibility to the splice site. The pre-mRNA harboring an aptamer with a stable base stem, inserted to encompass the 3' splice site AG within the ligand-binding sequence exhibited the most efficient splicing inhibition, as the splice site becomes less accessible when theophylline resides in the aptamer binding pocket.

Synthetic riboswitches constructed through the direct attachment strategy between the aptamer and regulatory domains exhibit functional dependence on the attachment location of the aptamer to the regulatory domain. This is because the ligand-mediated regulatory mechanism relies solely on how effective the ligand-aptamer complex interaction is in affecting the functional activity of the regulatory domain. This mechanism is not standardized and is highly specific to the particular aptamer-regulator pair. Therefore, such engineered riboswitches lack a reliable composition framework for integrating sensor and regulatory domains that results in allosteric binding properties.

1.4.3. Riboswitch construction based on an evolved linker between the aptamer and regulatory domains

Synthetic riboswitches have also been constructed by using a linker region that couples the aptamer and regulatory domains and serves as an element that translates the ligand-binding event in the aptamer domain to the adjacent regulatory domain. Early examples of evolved linker regions implemented a mechanism of information transmission known as 'helix slipping', in which a nucleotide shift event within the element is translated to a small-scale change in the conformation of the regulatory domain in a ligand-dependent

manner[64]. Such functional elements are often evolved through *in vitro* selection procedures and referred to as 'communication modules'[64]. These dynamic elements are typically three to five base-pairs long and often contain non-Watson-Crick base pairing.

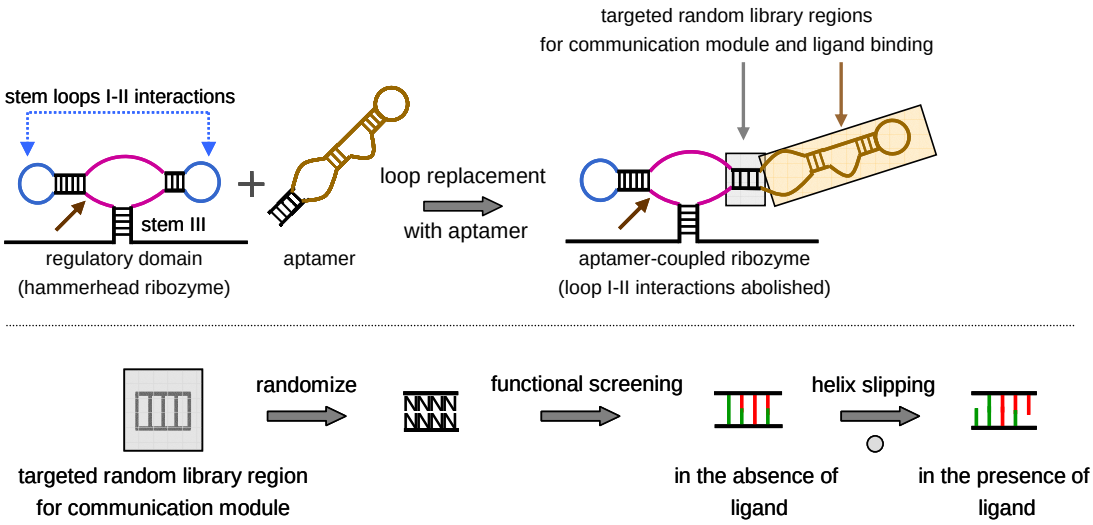


Figure 1.6. A schematic illustration of riboswitches constructed based on an evolved linker between the aptamer and regulatory domains. An *in vitro* functional communication module-based allosteric hammerhead ribozyme system is shown as an example. In general, an aptamer is attached to the regulatory domain through a linker (top), whose function is evolved to be ligand-dependent through selection from a random sequence library (bottom). The functional linkers are called 'communication modules', which employ the helix slipping mechanism in mediating the activity of the regulatory domain. An existing functional linker can also be used to mediate the activities of other regulatory platforms such as the RBS. In addition, an existing functional linker can also be used to couple a regulatory domain to an aptamer domain comprised of a random sequence library to evolve the latter to bind a new ligand of interest and function in this ligand-dependent manner (top). In this particular example of the *in vitro* allosteric ribozyme system, part of the aptamer domain replaces one loop of the ribozyme domain, thereby abolishing loop I-II interactions required for *in vivo* functionality.

Significant effort has been directed towards the construction of communication module-based riboswitches that use a hammerhead ribozyme as the regulatory domain (Figure 1.6), as ribozymes have proven to be a powerful platform for controlling gene expression. Several research groups have engineered a class of *in vitro* riboswitches called allosteric ribozymes[64-70]. Allosteric ribozymes resemble allosteric enzymes in that binding of

specific effectors, typically small molecule ligands, modulate the functional activities of the molecule[71]. An allosteric ribozyme contains two separate domains, a catalytic, or regulatory, domain and a ligand-binding, or aptamer, domain, which interact in a ligand-dependent manner to control the catalytic activity of the molecule[71]. Thus, the allosteric property of these ribozymes enables their catalytic activity to be regulated through specific ligands, and therefore may represent a modular design platform that can directly make use of different ligand-aptamer pairs.

Different strategies including rational design, library screening, and combinatorial approaches have been employed to generate allosteric hammerhead ribozymes[71, 72]. Rational design strategies involve integration of an existing aptamer domain directly to the catalytic domain of the ribozyme through different linkers, followed by examination of the activities of the resulting integrated constructs[65-67]. Library screening strategies involve screening randomized sequence libraries for novel aptamer domains (sometimes including communication modules) that function allosterically with the attached catalytic domain[73, 74]. Finally, the combined approach involves integration of an existing aptamer domain to the ribozyme's catalytic domain through a randomized linker region, and screening for functional linker sequences from this library that result in allosteric binding[64, 68, 69]. The majority of synthetic allosteric hammerhead ribozymes constructed to-date are responsive to small molecule ligands such as theophylline[51, 64], adenosine triphosphate (ATP)[65, 75], and flavin mononucleotide (FMN)[66, 76].

Researchers have also developed communication module-based riboswitches using a different regulatory platform such as the RBS. Suess et al. engineered a synthetic riboswitch comprised of a theophylline aptamer[51] and a previously developed communication module[71]

placed at a position proximal to the RBS[77]. This linker element had been proposed to perform helix slipping by one nucleotide between the ligand-bound and unbound states[71]. In their design, the communication module served as a helix bridge between the aptamer and the RBS such that binding of theophylline to its aptamer causes a single-nucleotide shift in the communication module, thereby enabling ribosome binding to the RBS without steric interference, and thus efficient translation in the presence of theophylline. This design scheme is similar to a direct coupling design between a theophylline aptamer and RBS described above, except that a distinct communication module was incorporated between the aptamer and regulatory domains.

Linker regions have also been evolved, which implemented a different mechanism of information transmission known as 'strand displacement', a functionally similar mechanism to 'helix slipping'. Gallivan and colleagues developed a second riboswitch system as an extension of their initial direct attachment riboswitch design, using a combined rational and library screening design strategy[78]. A linker region adjoining the theophylline aptamer and the RBS was randomized, and sequences that translated a ligand-binding event in the aptamer domain to a structural change in the RBS, thereby regulating ribosomal access to the RBS, were screened through plate-based assays. These sequences are functionally similar to communication modules in that they translate ligand-binding events at the aptamer domain to the regulatory domain, but they are compositionally and mechanistically distinct. Gene expression regulation through this second class of linker regions takes place through the strand-displacement mechanism instead of a helix slipping mechanism. The functional sequences are complementary to regions of the theophylline aptamer such that base-pairing with a region of the aptamer sequesters the RBS and thus inhibits ribosomal access to the

RBS, whereas binding of theophylline to its aptamer disrupts this conformation and releases the RBS, resulting in up-regulation of target gene expression. As such, this regulatory mechanism is specific to the theophylline aptamer employed in this system and is not functionally independent. Consequently, this riboswitch system is not readily amenable to the insertion of different aptamers and thus lacks modularity, such that new linker regions would need to be generated by screening for specific aptamer-regulator pairs.

1.5. Further advancing the current field of engineering synthetic riboswitch systems

Numerous synthetic riboswitch systems have been developed for ligand-mediated regulation of functional activities both *in vitro* and *in vivo* as described above. These systems have made remarkable contributions in advancing the fielding of engineering synthetic RNA switches and sensors. During the past decade, the field has rapidly emerged and gained tremendous interest, as engineered riboswitches are effective regulatory elements that hold significant promise in transforming our ability to engineer cellular functions for various biotechnological applications through the functionally versatile biological substrate, RNA. Nevertheless, previous examples of synthetic riboswitches face one or more limitations such as failure to function in the cellular environment (*in vivo* functionality), requirement of specific cellular machinery (portability across different cellular systems), functional dependence among components within the switch molecule (modularity), programmability of the components (response tunability), and limited availability of molecular ligands or inputs that can be employed (scalability).

In order to further advance the promising field of riboswitch engineering, we set out to develop a riboswitch platform that exhibits the above-described functional properties.

Allosteric hammerhead ribozymes with *in vivo* functionality are highly attractive substrates for the development of a functionally versatile riboswitch platform for the following reasons: first, their regulatory mechanism employs self-cleavage and thus does not require cell-specific machinery (portability); and second, since RNA aptamers can be *de novo* generated for potentially any target ligands of interest using a standardized procedure known as SELEX[25, 26] and the sensor domain of allosteric hammerhead ribozymes is comprised of an RNA aptamer, it is amenable to many user-specified molecular inputs that can be employed (scalability). Therefore, we employed allosteric hammerhead ribozymes as platform substrates, incorporated engineering design principles and strategies into our design, and developed an *in vivo* functional riboswitch platform, called the ribozyme switch[79], that is modular in design, tunable in regulation, scalable in molecular input, and portable in regulatory mechanism across diverse cellular systems, thereby further advancing the current field of riboswitch engineering.

1.6. Interrelationship among the thesis projects

Chapter I provides an overview of RNA as a functionally versatile molecule, and current technologies and challenges in the field of riboswitch engineering, from which my thesis projects were developed. Chapter II describes work on one of my thesis projects, which provides a high-throughput method for functional characterization of small molecule-binding RNA aptamers. This method will enable robust, accurate, and rapid characterization of such RNA aptamers and can be very useful as we (and others) develop RNA aptamers for small molecules of specific interest that are to be integrated into the ribozyme switch platform as sensing elements for specific applications. Chapter III describes the detailed

work in the development of the above-mentioned extensible ribozyme switch platform for the reliable design and construction of ligand-controlled gene-regulatory systems applicable across different cellular systems. Chapter IV describes a sophisticated application aspect of our ribozyme switch platform through which higher-order RNA devices were built to achieve complex cellular information processing operations, including logic control, advanced computation, and cooperativity. Chapter V describes the functional extension of the small molecule-responsive ribozyme switch platform to respond to a different class of ligand molecules, proteins, in developing protein-responsive gene regulators and cellular biosensors. This extension broadens the platform utility to a wider range of biotechnological applications. These research projects synergistically support each other such that 'designer' gene-regulatory systems for various biotechnological and medical applications can be reliably and effectively constructed.

References

1. Mandal, M. & Breaker, R. R. Gene regulation by riboswitches. Nat Rev Mol Cell Biol 5, 451-63 (2004).
2. Breaker, R. R. Complex riboswitches. Science 319, 1795-7 (2008).
3. Mandal, M. et al. A glycine-dependent riboswitch that uses cooperative binding to control gene expression. Science 306, 275-9 (2004).
4. Sudarsan, N. et al. Tandem riboswitch architectures exhibit complex gene control functions. Science 314, 300-4 (2006).

5. Cheah, M. T., Wachter, A., Sudarsan, N. & Breaker, R. R. Control of alternative RNA splicing and gene expression by eukaryotic riboswitches. Nature 447, 497-500 (2007).

6. Doudna, J. A. & Cech, T. R. The chemical repertoire of natural ribozymes. Nature 418, 222-8 (2002).

7. Fedor, M. J. & Williamson, J. R. The catalytic diversity of RNAs. Nat Rev Mol Cell Biol 6, 399-412 (2005).

8. Simons, R. W. Naturally occurring antisense RNA control--a brief review. Gene 72, 35-44 (1988).

9. Inouye, M. Antisense RNA: its functions and applications in gene regulation--a review. Gene 72, 25-34 (1988).

10. Meister, G. & Tuschl, T. Mechanisms of gene silencing by double-stranded RNA. Nature 431, 343-9 (2004).

11. Dykxhoorn, D. M., Novina, C. D. & Sharp, P. A. Killing the messenger: short RNAs that silence gene expression. Nat Rev Mol Cell Biol 4, 457-67 (2003).

12. Novina, C. D. & Sharp, P. A. The RNAi revolution. Nature 430, 161-4 (2004).

13. Tang, G. siRNA and miRNA: an insight into RISCs. Trends Biochem Sci 30, 106-14 (2005).

14. Breaker, R. R. Natural and engineered nucleic acids as tools to explore biology. Nature 432, 838-45 (2004).

15. Schroeder, R., Barta, A. & Semrad, K. Strategies for RNA folding and assembly. Nat Rev Mol Cell Biol 5, 908-19 (2004).

16. Soukup, J. K. & Soukup, G. A. Riboswitches exert genetic control through metabolite-induced conformational change. Curr Opin Struct Biol 14, 344-9 (2004).

17. Bauer, G. & Suess, B. Engineered riboswitches as novel tools in molecular biology. J Biotechnol 124, 4-11 (2006).

18. Isaacs, F. J., Dwyer, D. J. & Collins, J. J. RNA synthetic biology. Nat Biotechnol 24, 545-54 (2006).

19. Davidson, E. A. & Ellington, A. D. Engineering regulatory RNAs. Trends Biotechnol 23, 109-12 (2005).

20. Lai, E. C. RNA sensors and riboswitches: self-regulating messages. Curr Biol 13, R285-91 (2003).

21. Kurreck, J. Antisense technologies. Improvement through novel chemical modifications. Eur J Biochem 270, 1628-44 (2003).

22. Kawamoto, H., Morita, T., Shimizu, A., Inada, T. & Aiba, H. Implication of membrane localization of target mRNA in the action of a small RNA: mechanism of post-transcriptional regulation of glucose transporter in Escherichia coli. Genes Dev 19, 328-38 (2005).

23. Vanderpool, C. K. & Gottesman, S. Noncoding RNAs at the membrane. Nat Struct Mol Biol 12, 285-6 (2005).

24. Isaacs, F. J. et al. Engineered riboregulators enable post-transcriptional control of gene expression. Nat Biotechnol 22, 841-7 (2004).

25. Tuerk, C. & Gold, L. Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. Science 249, 505-10 (1990).

26. Ellington, A. D. & Szostak, J. W. In vitro selection of RNA molecules that bind specific ligands. Nature 346, 818-22 (1990).

27. Rimmele, M. Nucleic acid aptamers as tools and drugs: recent developments. Chembiochem 4, 963-71 (2003).

28. Cox, J. C. & Ellington, A. D. Automated selection of anti-protein aptamers. Bioorg Med Chem 9, 2525-31 (2001).

29. Cox, J. C. et al. Automated acquisition of aptamer sequences. Comb Chem High Throughput Screen 5, 289-99 (2002).

30. Cox, J. C. et al. Automated selection of aptamers against protein targets translated in vitro: from gene to aptamer. Nucleic Acids Res 30, e108 (2002).

31. Berezovski, M. et al. Nonequilibrium capillary electrophoresis of equilibrium mixtures: a universal tool for development of aptamers. J Am Chem Soc 127, 3165-71 (2005).

32. Drabovich, A., Berezovski, M. & Krylov, S. N. Selection of smart aptamers by equilibrium capillary electrophoresis of equilibrium mixtures (ECEEM). J Am Chem Soc 127, 11224-5 (2005).

33. Krylov, S. N. Nonequilibrium capillary electrophoresis of equilibrium mixtures (NECEEM): A novel method for biomolecular screening. J Biomol Screen 11, 115-22 (2006).

34. Hermann, T. & Patel, D. J. Adaptive recognition by nucleic acid aptamers. Science 287, 820-5 (2000).

35. Khvorova, A., Kwak, Y. G., Tamkun, M., Majerfeld, I. & Yarus, M. RNAs that bind and change the permeability of phospholipid membranes. Proc Natl Acad Sci U S A 96, 10649-54 (1999).

36. Vlassov, A., Khvorova, A. & Yarus, M. Binding and disruption of phospholipid bilayers by supramolecular RNA complexes. Proc Natl Acad Sci U S A 98, 7706-11 (2001).

37. Winkler, W. C. & Breaker, R. R. Genetic control by metabolite-binding riboswitches. Chembiochem 4, 1024-32 (2003).

38. Sudarsan, N., Barrick, J. E. & Breaker, R. R. Metabolite-binding RNA domains are present in the genes of eukaryotes. Rna 9, 644-7 (2003).

39. Welz, R. & Breaker, R. R. Ligand binding and gene control characteristics of tandem riboswitches in Bacillus anthracis. Rna 13, 573-82 (2007).

40. Mandal, M. & Breaker, R. R. Adenine riboswitches and gene activation by disruption of a transcription terminator. Nat Struct Mol Biol 11, 29-35 (2004).

41. Winkler, W. C., Nahvi, A., Sudarsan, N., Barrick, J. E. & Breaker, R. R. An mRNA structure that controls gene expression by binding S-adenosylmethionine. Nat Struct Biol 10, 701-7 (2003).

42. Winkler, W., Nahvi, A. & Breaker, R. R. Thiamine derivatives bind messenger RNAs directly to regulate bacterial gene expression. Nature 419, 952-6 (2002).

43. Winkler, W. C., Nahvi, A., Roth, A., Collins, J. A. & Breaker, R. R. Control of gene expression by a natural metabolite-responsive ribozyme. Nature 428, 281-6 (2004).

44. Nudler, E. & Mironov, A. S. The riboswitch control of bacterial metabolism. Trends Biochem Sci 29, 11-7 (2004).

45. Winkler, W. C. & Breaker, R. R. Regulation of bacterial gene expression by riboswitches. Annu Rev Microbiol 59, 487-517 (2005).

46. Patel, D. J. et al. Structure, recognition and adaptive binding in RNA aptamer complexes. J Mol Biol 272, 645-64 (1997).

47. Stripecke, R., Oliveira, C. C., McCarthy, J. E. & Hentze, M. W. Proteins binding to 5' untranslated region sites: a general mechanism for translational regulation of mRNAs in human and yeast cells. Mol Cell Biol 14, 5898-909 (1994).

48. Paraskeva, E., Atzberger, A. & Hentze, M. W. A translational repression assay procedure (TRAP) for RNA-protein interactions in vivo. Proc Natl Acad Sci U S A 95, 951-6 (1998).

49. Pelletier, J. & Sonenberg, N. Insertion mutagenesis to increase secondary structure within the 5' noncoding region of a eukaryotic mRNA reduces translational efficiency. Cell 40, 515-26 (1985).

50. Werstuck, G. & Green, M. R. Controlling gene expression in living cells through small molecule-RNA interactions. Science 282, 296-8 (1998).

51. Jenison, R. D., Gill, S. C., Pardi, A. & Polisky, B. High-resolution molecular discrimination by RNA. Science 263, 1425-9 (1994).

52. Wilson, C., Nix, J. & Szostak, J. Functional requirements for specific ligand recognition by a biotin-binding RNA pseudoknot. Biochemistry 37, 14410-9 (1998).

53. Berens, C., Thain, A. & Schroeder, R. A tetracycline-binding RNA aptamer. Bioorg Med Chem 9, 2549-56 (2001).

54. Harvey, I., Garneau, P. & Pelletier, J. Inhibition of translation by RNA-small molecule interactions. Rna 8, 452-63 (2002).

55. Suess, B. et al. Conditional gene expression by controlling translation with tetracycline-binding aptamers. Nucleic Acids Res 31, 1853-8 (2003).

56. Hanson, S., Berthelot, K., Fink, B., McCarthy, J. E. & Suess, B. Tetracycline-aptamer-mediated translational regulation in yeast. Mol Microbiol 49, 1627-37 (2003).

57. Kozak, M. Initiation of translation in prokaryotes and eukaryotes. Gene 234, 187-208 (1999).

58. de Smit, M. H. & van Duin, J. Control of prokaryotic translational initiation by mRNA secondary structure. Prog Nucleic Acid Res Mol Biol 38, 1-35 (1990).

59. de Smit, M. H. & van Duin, J. Secondary structure of the ribosome binding site determines translational efficiency: a quantitative analysis. Proc Natl Acad Sci U S A 87, 7668-72 (1990).

60. de Smit, M. H. & van Duin, J. Control of translation by mRNA secondary structure in Escherichia coli. A quantitative analysis of literature data. J Mol Biol 244, 144-50 (1994).

61. Desai, S. K. & Gallivan, J. P. Genetic screens and selections for small molecules based on a synthetic riboswitch that activates protein translation. J Am Chem Soc 126, 13247-54 (2004).

62. An, C. I., Trinh, V. B. & Yokobayashi, Y. Artificial control of gene expression in mammalian cells by modulating RNA interference through aptamer-small molecule interaction. Rna 12, 710-6 (2006).

63. Kim, D. S., Gusti, V., Pillai, S. G. & Gaur, R. K. An artificial riboswitch for controlling pre-mRNA splicing. Rna 11, 1667-77 (2005).

64. Soukup, G. A. & Breaker, R. R. Engineering precision RNA molecular switches. Proc Natl Acad Sci U S A 96, 3584-9 (1999).

65. Tang, J. & Breaker, R. R. Rational design of allosteric ribozymes. Chem Biol 4, 453-9 (1997).

66. Araki, M., Okuno, Y., Hara, Y. & Sugiura, Y. Allosteric regulation of a ribozyme activity through ligand-induced conformational change. Nucleic Acids Res 26, 3379-84 (1998).

67. Soukup, G. A. & Breaker, R. R. Design of allosteric hammerhead ribozymes activated by ligand-induced structure stabilization. Structure 7, 783-91 (1999).

68. Soukup, G. A., Emilsson, G. A. & Breaker, R. R. Altering molecular recognition of RNA aptamers by allosteric selection. J Mol Biol 298, 623-32 (2000).

69. Kertsburg, A. & Soukup, G. A. A versatile communication module for controlling RNA folding and catalysis. Nucleic Acids Res 30, 4599-606 (2002).

70. Link, K. H. et al. Engineering high-speed allosteric hammerhead ribozymes. Biol Chem 388, 779-86 (2007).

71. Soukup, G. A. & Breaker, R. R. Nucleic acid molecular switches. Trends Biotechnol 17, 469-76 (1999).

72. Wilson, D. S. & Szostak, J. W. In vitro selection of functional nucleic acids. Annu Rev Biochem 68, 611-47 (1999).

73. Koizumi, M., Soukup, G. A., Kerr, J. N. & Breaker, R. R. Allosteric selection of ribozymes that respond to the second messengers cGMP and cAMP. Nat Struct Biol 6, 1062-71 (1999).

74. Piganeau, N., Thuillier, V. & Famulok, M. In vitro selection of allosteric ribozymes: theory and experimental validation. J Mol Biol 312, 1177-90 (2001).

75. Sassanfar, M. & Szostak, J. W. An RNA motif that binds ATP. Nature 364, 550-3 (1993).

76. Burgstaller, P. & Famulok, M. Isolation of RNA apatmers for biological cofactors by in vitro selection. Angew. Chem. Int. Ed. Engl. 33, 1084-1087 (1994).

77. Suess, B., Fink, B., Berens, C., Stentz, R. & Hillen, W. A theophylline responsive riboswitch based on helix slipping controls gene expression in vivo. Nucleic Acids Res 32, 1610-4 (2004).

78. Lynch, S. A., Desai, S. K., Sajja, H. K. & Gallivan, J. P. A high-throughput screen for synthetic riboswitches reveals mechanistic insights into their function. Chem Biol 14, 173-84 (2007).

79. Win, M. N. & Smolke, C. D. From the Cover: A modular and extensible RNA-based gene-regulatory platform for engineering cellular function. Proc Natl Acad Sci U S A 104, 14283-8 (2007).

Chapter II: Codeine-binding RNA aptamers and rapid determination of their binding constants using a direct coupling surface plasmon resonance assay*

Abstract

RNA aptamers that bind the opium alkaloid codeine were generated using an iterative *in vitro* selection process. The binding properties of these aptamers, including equilibrium and kinetic rate constants, were determined through a rapid, high-throughput approach using surface plasmon resonance analysis to measure real-time binding. The approach involves direct coupling of the target small molecule onto a sensor chip without utilization of a carrier protein. Two highest binding aptamer sequences, FC5 and FC45 with K_d values of 2.50 μM and 4.00 μM, respectively, were extensively studied. Corresponding mini-aptamers for FC5 and FC45 were subsequently identified through the described direct coupling Biacore assays. These assays were also employed to confirm the proposed secondary structures of the mini-aptamers. Both aptamers exhibit high specificity to codeine over morphine, which differs from codeine by a methyl group. Finally, the direct coupling method was demonstrated to eliminate potential non-specific interactions that may be associated with indirect coupling methods in which protein linkers are commonly employed. Therefore, in addition to presenting the first RNA aptamers to a subclass of benzylisoquinoline alkaloid molecules, this work highlights a method for characterizing small-molecule aptamers that is more robust, precise, rapid, and high-throughput than other commonly employed techniques.

*Reproduced with permission from: M. N. Win, J. S. Klein, and C. D. Smolke. (2006) *Nucleic Acids Res.*, 34, 5670-5682.

2.1. Introduction

Codeine is a naturally-occurring opium alkaloid, part of the larger class of benzylisoquinoline alkaloids (BIAs), found in the opium poppy, *Papaver somniferum*, and constitutes approximately 0.5% of opium[1]. It is one of the most widely used narcotic drugs for the treatment of mild to moderate pain, diarrhea, and cough with relatively low side effects[2]. Despite its extensive medical applications, codeine is often abused for its euphoric and depressant effects as well as to prevent opiate withdrawal[3]. Due to increasing misuse, codeine has been incorporated into workplace and military drug testing programs, and a screening and confirmation cutoff concentration of 40 μg/L has been suggested for federally-mandated testing in oral fluid by the Substance Abuse and Mental Health Services Administration[3]. Therefore, a sensor system that can precisely measure the concentration of codeine and effectively discriminate against its structural analogues is highly desired.

Aptamers are nucleic acid molecules that bind ligands with high specificity and affinity[4]. There is increasing interest in utilizing aptamers as the target recognition elements in various sensing applications[5-8]. In addition to the drug detection applications of a codeine-binding aptamer, there are other potential biotechnology applications for this aptamer. Codeine is a member of the BIA family and is a key product metabolite in the opium alkaloid biosynthesis pathway[9]. The BIAs comprise a structurally diverse group of pharmacologically important compounds[10] and efforts are ongoing to engineering microbial and plant hosts for the production of some of the important BIA intermediates in the codeine synthesis pathway such as (S)-reticuline and thebaine[9-11].

Aptamers to BIA molecules may prove to be useful tools for such engineering efforts. Recent research has highlighted the application of aptamers as components of synthetic and

naturally-occurring cellular sensors and switches[12-17], which can regulate enzyme levels in response to small-molecule ligand concentrations. Therefore, aptamer-based cellular sensors may be generated to act as 'intelligent' regulatory tools for metabolic engineering efforts to provide dynamic regulation of gene expression at specific enzymatic steps so that pathway fluxes are rewired to enable the accumulation of desired intermediate metabolites, which has proven to be difficult to achieve in natural plant hosts[9]. A codeine-binding aptamer may be used to construct tools such as synthetic riboswitches that can be employed to redirect flux through an engineered BIA metabolic pathway or in setting up rapid functional screens of pathway variants. In addition, while aptamers have been developed to several of the far upstream metabolites in this pathway such as dopamine[18] and tyrosine[19], they have not yet been developed against any BIA compounds, which harbor bulky, nitrogen-containing ring structures. Prior work has demonstrated that aptamers to specific molecules within a family of compounds may be used to design doped libraries for the selection of aptamers to similar compounds within that family from smaller library sequence spaces[19], and thus codeine aptamers would be potentially useful for selecting aptamers to diverse BIA molecules.

This work describes the generation of novel RNA aptamers to the small molecule codeine and highlights a robust, high-throughput assay method for measuring small molecule-aptamer binding properties. RNA aptamers that bind codeine with high affinities were selected from a combinatorial library containing a 30 nucleotide randomized region using an iterative *in vitro* selection procedure or SELEX (Systematic Evolution of Ligands by EXponential enrichment)[20, 21]. The binding properties of the generated codeine aptamers were measured by surface plasmon resonance (SPR) through a real-time binding assay (Biacore), similar to previously reported methods[22, 23] where the small-molecule ligand is

directly coupled to a sensor chip through chemical modification of the ligand, eliminating the need to use protein linkers between the target small molecule and the sensor surface as described in other methods[24-26]. This direct coupling method limits potential non-specific interactions or binding artifacts arising from the presence of the linker protein observed in previous studies[24, 25], which may alter the determined binding affinities. Therefore, this method may provide a more accurate assessment of small molecule-aptamer binding affinities since the measured interaction more closely mimics the binding environment of the *in vitro* selection process.

2.2. Results

2.2.1. Selection of codeine-binding RNA aptamers

A slightly modified *in vitro* selection procedure was used to isolate codeine-binding RNA aptamers from a library of RNA molecules containing a 30-nucleotide random region flanked by constant primer-binding sequences (Figure 2.1A). Aptamers were selected on a codeine affinity column, which was made by immobilizing codeine to the epoxy-activated agarose through its hydroxyl group (Figure 2.2A). To enhance the stringency of the selection process, the wash volume was increased incrementally from cycles 6 to 15. To increase the specificity of the selected pool, a counter-selection with a 5 mM morphine solution was performed at cycle 10 prior to elution with codeine. In addition, a total of three error-prone PCR steps were carried out for the DNA template pools of cycles 11, 12, and 13, respectively, to potentially introduce sequences that are of slightly diverse nucleotide composition and search a larger sequence space for higher affinity binders. After cycle 15, the enriched pool was cloned and approximately 60 colonies were sequenced (Figure 2.1B).

A

TTC*TAATACGACTCACTATA* (GGGACAGGGCTAGC) $(N)_{30}$ (GAGGCAAAGCTTCCG)

T7 promoter 5’ constant region 3’ constant region

B

Clone	Sequence	
FC21(2)	GGGACAGGGCTAGC AAAAGGGTGGTTGAAGGGACAGCTGGTGTG	GAGGCAAAGCTTCCG *
A25 (3)	GGGACAGGGCTAGC ACAAGAATTAGGGTCGGGAAATGGTGTGTG	GAGGCAAAGCTTCCG *
C4 (2)	GGGACAGGGCTAGC CACAAGTGTGAAGGGATGGGAGTAGTGGTG	GAGGCAAAGCTTCCG *
C9 (2)	GGGACAGGGCTAGC AAGAATAGGATGTGGGTAAAGGTGCTGGTG	GAGGCAAAGCTTCCG *
C12 (3)	GGGACAGGGCTAGC ACATGGAGGCTTATAGGGATTCGTGCTGGG	GAGGCAAAGCTTCCG *
FC5 (1)	GGGACAGGGCTAGC AGTAGGATTGGGTGAGGGGATGTGCTGTG	GAGGCAAAGCTTCCG *
B10 (1)	GGGACAGGGCTAGC AGTAGGATTAGGGTGAGGGGATGTGCTGTG	GAGGCAAAGCTTCCG
A28 (1)	GGGACAGGGCTAGC ACATTGTGGGAAAGGGAATTGAGTGTGGTG	GAGGCAAAGCTTCCG
B11 (1)	GGGACAGGGCTAGC ACATTGAGGGAAAGGGAATTGAGTGTGGTG	GAGGCAAAGCTTCCG *
FC1	GGGACAGGGCTAGC CACGAAATGGGTGAAGGGAAACGTGGTGGG	GAGGCAAAGCTTCCG
FC3	GGGACAGGGCTAGC ACCAAAAATAGGGGTAAGGGCATGGGGGTG	GAGGCAAAGCTTCCG *
FC13	GGGACAGGGCTAGC AGGGTAAGGGGATTGGAGTAGTGCCGTGTG	GAGGCAAAGCTTCCG *
FC17	GGGACAGGGCTAGC GGACAAGAAGTGGGTAAGGGAATCCGTGGG	GAGGCAAAGCTTCCG *
FC23	GGGACAGGGCTAGC CAATAAATAAGGCGAAGTAAGGGATGGGGTG	GAGGCAAAGCTTCCG *
FC27	GGGACAGGGCTAGC TACTAATGTACGCACTAAGGGATTGGGGTG	GAGGCAAAGCTTCCG *
FC33	GGGACAGGGCTAGC GAAAGCGGTTTGGGAAAGTAAAGGGTGGTG	GAGGCAAAGCTTCCG
FC34	GGGACAGGGCTAGC TACAGAATAAGCGAATTAAGGGTTGGGGTG	GAGGCAAAGCTTCCG *
FC36	GGGACAGGGCTAGC AAAGTGAGGGTTATGGGGATACGTGGCGTG	GAGGCAAAGCTTCCG
FC41	GGGACAGGGCTAGC ATTAGGGTAATCGATCAAGAGGGAGTGGTG	GAGGCAAAGCTTCCG
FC45	GGGACAGGGCTAGC TTAGTGCTATGTGAGAAAAGGGTGTGGGGG	GAGGCAAAGCTTCCG *
A2	GGGACAGGGCTAGC ACGTTAGGATGAGGGTAATGGCGTTGTAGAAGA	GAGGCAAAGCTTCCG
A3	GGGACAGGGCTAGC GTAATAAGTAGGGAAAGGGTTCCCGCTGGG	GAGGCAAAGCTTCCG *
A5	GGGACAGGGCTAGC TTTAAAGTGAGGGGTTATGGGCAGTGTGGT	GAGGCAAAGCTTCCG
A7	GGGACAGGGCTAGC TTTTAAGCACAATAACAGGGTGGGGATGGT	GAGGCAAAGCTTCCG
A15	GGGACAGGGCTAGC ACCATTAGGGATTATCCAACGGGGGGTGTG	GAGGCAAAGCTTCCG
A20	GGGACAGGGCTAGC CTATAGTGAGGCTATTAAGGGTTGTGGGGG	GAGGCAAAGCTTCCG *
A22	GGGACAGGGCTAGC AGTTGAATAGGGTTGGAGAAAGGACGTGGT	GAGGCAAAGCTTCCG
A23	GGGACAGGGCTAGC TTATTTAGGGTTGGAGGGTAGTTAGCGGTG	GAGGCAAAGCTTCCG
A30	GGGACAGGGCTAGC TAATGAAGGGCAAGGGAATAGTGGCTAGGG	GAGGCAAAGCTTCCG
B1	GGGACAGGGCTAGC GAGTAAAAAGGGTTGGGAAAATCGCATGGT	GAGGCAAAGCTTCCG
B2	GGGACAGGGCTAGC GCAGAACAGAGGGTAGGGAATTTGCGTGTG	GAGGCAAAGCTTCCG *
B4	GGGACAGGGCTAGC TCAGAACGCTAGATTAGGATGTGGGTGGTG	GAGGCAAAGCTTCCG
B6	GGGACAGGGCTAGC AAAAGGGTGGTTGAAGGGACAGCTGGTGTG	GAGGCAAAGCTTCCG
B8	GGGACAGGGCTAGC TACAATAGGGCAATTAATGGGGAGTGTGTG	GAGGCAAAGCTTCCG
B9	GGGACAGGGCTAGC ATCGGTGTAGGGAAGGGATATGATGTGGTG	GAGGCAAAGCTTCCG *
B10	GGGACAGGGCTAGC AGTAGGATTAGGGTGAGGGGATGTGCTGTG	GAGGCAAAGCTTCCG
B12	GGGACAGGGCTAGC AGCGGTAAGGGTGGGGAGAATGGTGCTGTG	GAGGCAAAGCTTCCG *
C1	GGGACAGGGCTAGC ATAGCATGGAGCGACTATGCGTTGATGGGT	GAGGCAAAGCTTCCG
C3	GGGACAGGGCTAGC CGTTGTAACGGTGAATTTAGGGTAAGGGGG	GAGGCAAAGCTTCCG
C7	GGGACAGGGCTAGC CCGTCCCTATAGTGAGTCGTATTAGAACGG	AAGCTTCCG
C10	GGGACAGGGCTAGC TTTACAGTGAAAAATTAAGGGAAGGGGGTG	GAGGCAAAGCTTCCG

Figure 2.1. Codeine-binding RNA aptamer clone sequences. (A) DNA template from which the initial RNA pool was generated. (B) Sequences of clones from the final aptamer pool. The codeine-binding properties of the sequences marked with an asterisk were characterized by the described direct coupling SPR assay. The number in parenthesis represents the frequency of a particular clone in the sequenced pool.

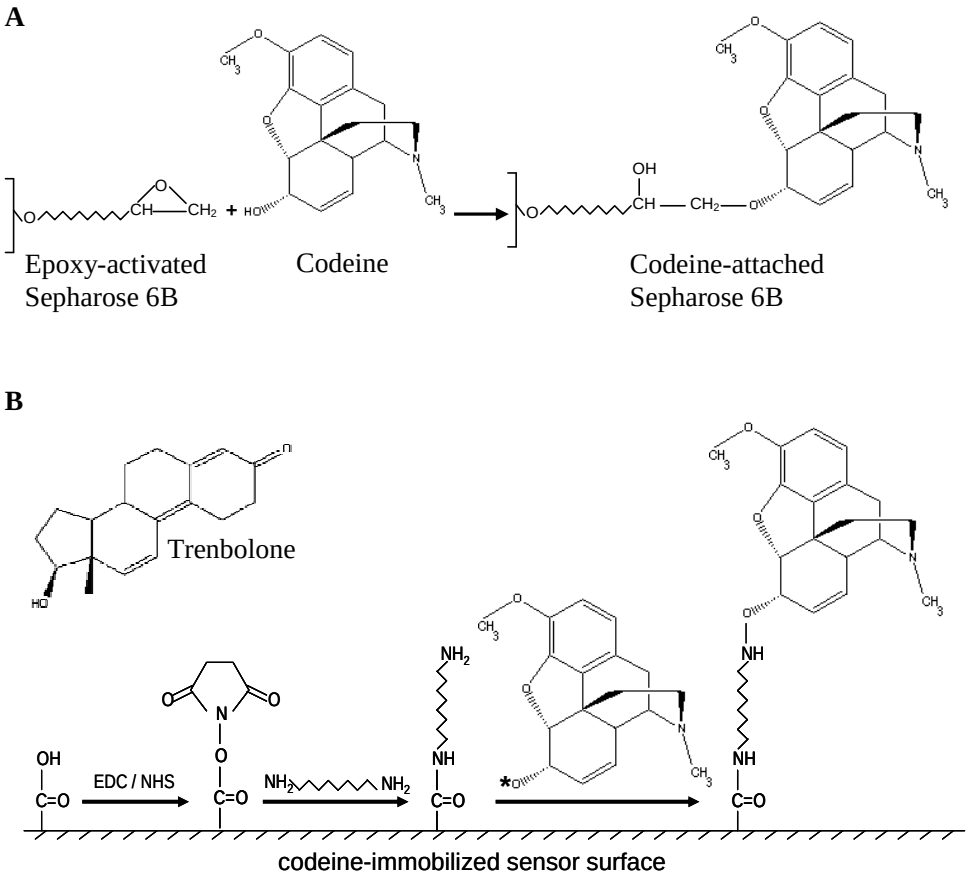


Figure 2.2. Schematics of the codeine-immobilized surfaces used in the *in vitro* selection process and SPR binding property assay. Illustration of the chemistries used for codeine coupling to the (A) Sepharose matrix and (B) Biacore CM5 sensor chip surface. Note that the codeine-immobilized sensor surface more closely mimics that of the affinity matrix used during the aptamer selection process versus coupling methods that employ a protein linker. The asterisk next to the oxygen group of codeine in (B) represents a succinimidyl group (the same group that is covalently attached to the carboxyl group of the sensor surface after EDC/NHS activation), which reacts with the amine group of the 1,8-diaminooctane linker. Codeine is thereby immobilized onto the chip surface through the same functional group used to attach it to the affinity matrix during the selection process. Trenbolone, the negative control molecule, is immobilized to the chip surface through the same chemistry and its structure is shown in (B).

2.2.2. Qualitative assessment of codeine-binding affinity of the enriched final pool

The codeine-binding affinity of the final pool was qualitatively assessed by monitoring eluted levels of the radiolabeled aptamer pool using codeine affinity chromatography. Radiolabeled RNA from the enriched pool was incubated with codeine-

modified and unmodified columns. The eluted RNA from each column was run on a polyacrylamide gel and visualized with a phosphorimager (Supplementary Figure 2.1). Significantly stronger radioactive signals were detected in the sample eluted from the codeine affinity column than that eluted from the unmodified column, indicating that the RNA aptamers in the final pool are highly enriched in codeine-binding affinity.

2.2.3. Determination of small molecule-aptamer binding constants using a direct coupling surface plasmon resonance assay

Quantitative assessment of the codeine-binding properties of the final pool, the initial pool, and several aptamers from the final pool was performed using a modified SPR assay developed on a Biacore 2000. Previous studies where SPR was used to determine binding affinities between aptamers and non-protein targets involved the use of BSA and biotin/streptavidin as intermediate linkers between the sensor surface and the target molecules[24, 26]. Here we employ a direct coupling approach, similar to a previously described method[22, 23], in which the small-molecule target is directly coupled to the sensor surface without a supporting intermediate such as BSA or biotin/strepavidin. Previous direct coupling strategies have used target molecules that contain an amine group[22, 23], which is a commonly used functional group in Biacore sensor chip immobilization strategies. However, since codeine does not contain an amine group, a chemical modification strategy was developed to directly couple codeine to the chip surface through its hydroxyl group. In this coupling strategy codeine is first modified at its hydroxyl group with an amine-reactive succinimidyl group. This chemical modification enables codeine molecules to readily react with the amine groups attached to the activated chip surface (Figure 2.2B). Trenbolone was

also immobilized onto the sensor surface in the same manner and used as a negative control molecule. Following the immobilization of codeine and trenbolone in their respective flow cells of the sensor chip, serial dilutions of RNA samples were injected into these flow cells. The response detected from the trenbolone-immobilized flow cell was used as the background subtraction in evaluating the binding constants. An equilibrium binding curve was generated from concentration-dependent binding response data for each sample to determine the corresponding K_d value.

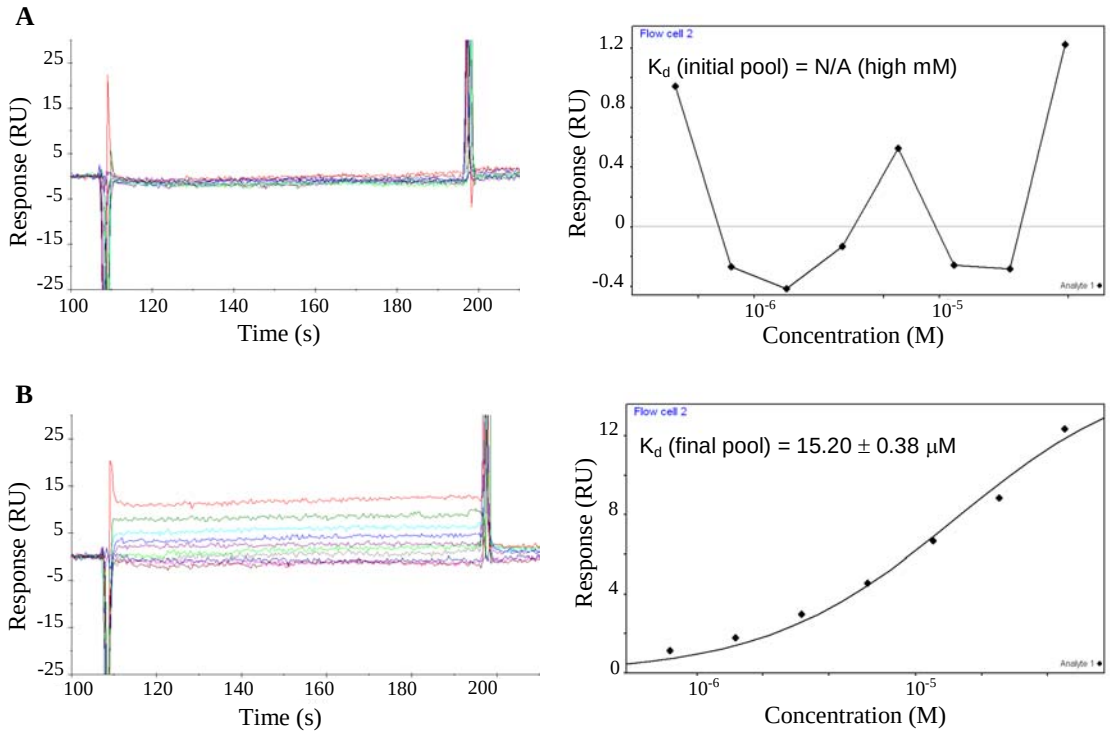


Figure 2.3. Concentration-dependent codeine-binding responses (left) and the corresponding equilibrium binding curve (right) of (A) the initial pool and (B) the enriched final pool. Codeine was coupled to the sensor chip as described. Serial dilutions of the appropriate RNA sample were injected across the sensor surface and binding responses were recorded over time. Kinetic rate constants were determined by examining the rate of change of binding response when the RNA samples were initially injected over the surface until equilibrium responses were reached (k_{on}) and when a solution lacking the RNA sample was injected over the surface once equilibrium levels were bound to the chip surface (k_{off}). Equilibrium binding constants (K_d) were determined by plotting the equilibrium binding response versus the RNA sample concentration and calculating the corresponding RNA concentration at which half of the maximal response was achieved. Binding responses were adjusted for background

binding by subtracting responses of the corresponding RNA samples determined from a trenbolone-coupled sensor surface.

The binding data from the SPR assay supports the qualitative binding data obtained from the chromatography-based assay. The data indicate that there was little to no detectable binding (Figure 2.3A) between the initial pool and codeine, whereas the final pool bound codeine with significant binding responses (Figure 2.3B). The overall K_d value of the final pool was evaluated to be approximately 15 μM, whereas that of the initial pool was estimated to be in the high millimolar range. This latter value is only an estimate, as no binding curve could be established for the initial pool due to its insufficient binding response. Therefore, codeine-binding affinity of the final pool was enhanced over 1000-fold from that of the initial pool.

Table 2.1. Codeine-binding affinities of the full-length aptamer sequences as determined from the direct coupling SPR assay.

RNA sample	K_d	RNA sample	K_d	RNA sample	K_d
final pool	15.20 ± 0.38 μM	FC27	10.90 ± 0.95 μM	B11	5.80 ± 0.29 μM
initial pool	N/A (high mM)	FC34	28.00 ± 1.42 μM	B12	8.80 ± 0.44 μM
FC3	28.60 ± 0.96 μM	FC45	2.50 ± 0.06 μM	C4	78.00 ± 4.15 μM
FC5	4.00 ± 0.13 μM	A3	11.50 ± 0.27 μM	C9	9.17 ± 0.33 μM
FC13	14.50 ± 0.58 μM	A20	13.00 ± 0.65 μM	C12	8.88 ± 0.39 μM
FC17	43.70 ± 1.23 μM	A25	7.23 ± 0.34 μM	C15	7.67 ± 0.26 μM
FC21	23.60 ± 1.22 μM	B2	4.75 ± 0.32 μM	C23	8.18 ± 0.23 μM
FC23	19.10 ± 1.49 μM	B9	5.77 ± 0.36 μM		

The K_d values of the analyzed aptamer clones are listed in Table 2.1. Several of the aptamer sequences have K_d values that are much lower than that of the enriched final pool. Two of the highest binding aptamers FC45 and FC5, with K_d values of 2.50 ± 0.06 μM and 4.00 ± 0.13 μM, respectively, were subject to further characterization studies (Figure 2.4).

Despite their similar affinities for codeine, FC5 and FC45 may form different binding pockets since their corresponding mini-aptamers adopt different predicted secondary structures supported by structural studies described in a later section. In addition, FC5 and FC45 exhibit fairly different binding kinetics (Table 2.2), where the latter has faster kinetics (both binding and dissociation) than that of the former. Some clones such as FC3, FC13, FC34, and C9 have observed dissociation constants on the same order as that of FC5, while other clones such as FC23, A3, A20, B11, C15, and C23 exhibit similar dissociation kinetics to FC45 (data not shown). The kinetic data of the modified FC5 and FC45 sequences discussed in later sections are also reported in Table 2.2.

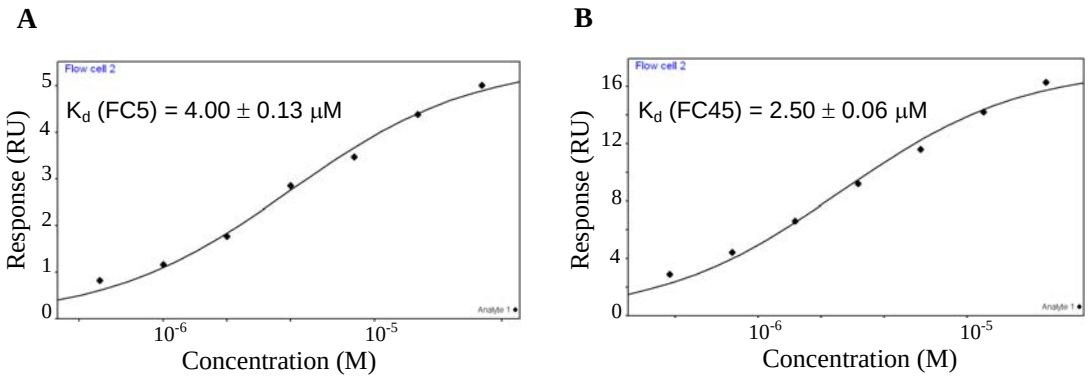


Figure 2.4. Equilibrium codeine-binding response curves of (A) FC5 and (B) FC45.

Table 2.2. Dissociation rate constants (k_{off}) for codeine binding of the final pool, FC5, FC45, and their corresponding truncated sequences. The corresponding association rate constant (k_{on}) is equivalent to k_{off}/K_d.

RNA sample	k_{off} (1/s)	RNA sample	k_{off} (1/s)
final pool	7.62e-3±5.81%	initial pool	N/A
FC5	6.50e-3±3.78%	FC45	1.14e-2±3.35%
FC5L	6.70e-3±2.80%	FC45L	1.03e-2±3.61%
FC5L-S1	6.65e-3±2.79%	FC45L-S1	6.84e-3±3.98%
FC5L-S2	6.54e-3±2.44%	FC45L-S2	2.43e-3±3.29%
FC5L-S3	4.79e-3±5.00%	FC45L-S3	2.69e-3±2.23%

The affinities of the two highest binding aptamers, FC5 and FC45, to codeine in solution were also determined through a standard isocratic affinity elution method[27, 28]. This control enables the comparison of the surface-based binding affinities determined with the described SPR assays to the solution-based affinities. The determined solution-binding affinity of FC45 (K_d = 4.5 μM) was very similar to its surface-binding affinity (K_d = 2.5 μM), whereas FC5 was determined to bind free codeine with an approximately 10-fold lower affinity (K_d = 47 μM) than that to surface-immobilized codeine (K_d = 4.0 μM). For a given aptamer-ligand pair, the binding affinities for free target in solution and a target immobilized onto a solid support may differ, as has been observed in previous studies[19, 29, 30]. For the aptamers studied here, FC5 shows differing affinities for free and immobilized codeine, whereas FC45 exhibits similar binding affinities.

2.2.4. Assays reveal distinct specificities of the codeine-binding aptamers to other benzylisoquinoline alkaloid targets

The ability of FC5 and FC45 to distinguish between three similar BIA molecules, codeine, thebaine, and morphine, was determined using a chromatography-based assay. Radiolabeled RNA aptamers were eluted with codeine, morphine, and thebaine, which are all closely related structural analogues (Figure 2.5A). Eluted FC5 and FC45 demonstrated approximately 4-fold and 6-fold increases in radioactivity counts, respectively (Figure 2.5B) when eluted with codeine versus morphine. The semi-quantitative molecular specificities of these aptamers were supported by isocratic affinity elution experiments in which the solution affinities of these aptamers were determined and observed to differ by similar magnitudes. The solution affinity for FC45 was determined to be approximately 4.5 μM to codeine and 25

μM to morphine, whereas the solution affinity for FC5 was determined to be approximately 47 μM to codeine and 212 μM to morphine. These results demonstrate that the single morphine counter-selection performed during the *in vitro* selection process was effective at enhancing the specificity of the aptamers in the final pool to codeine over morphine. While aptamers that discriminate between molecules that differ by a single methyl group have been described previously for purine alkaloid targets[31, 32], these results indicate that aptamers can exhibit this level of molecular discrimination in spite of the presence of the bulky 4 six-membered rings in the BIA targets examined here.

These assays also demonstrate that these two aptamers exhibit differing specificities to thebaine. The eluted FC5 exhibited nearly a two-fold increase in radioactivity counts when eluted with thebaine versus codeine, whereas FC45 exhibited an approximately 30% decrease in signal. These results indicate that FC5 exhibits higher specificity for thebaine over codeine, whereas FC45 exhibits higher specificity for codeine over thebaine. It should be noted that during the selection process codeine was coupled to the Sepharose column in such a way that there was no differentiable functional group between codeine and thebaine. With the attachment chemistry used in these studies through the functional group at C5, these two molecules exhibit conformational differences in that the former has one double bond in the C5-six-membered ring, whereas the latter contains two (Figure 2.5A). These results suggest that aptamers can potentially perform molecular discrimination at the level of conformation, as the difference between these two targets is at the level of torsional structure of the ring backbone.

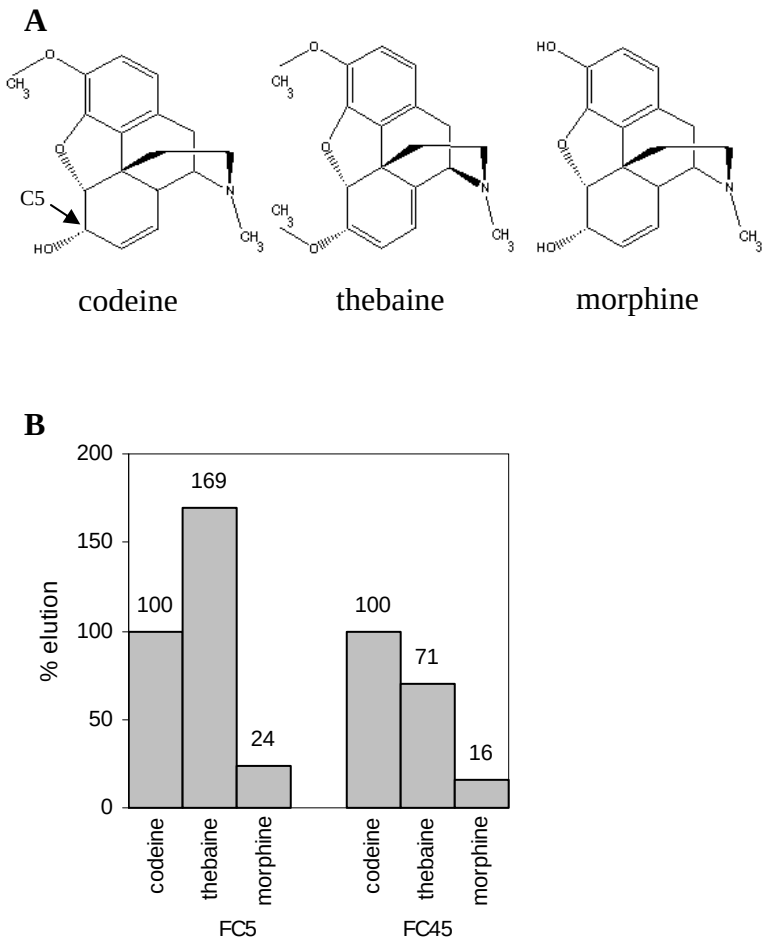


Figure 2.5. The FC5 and FC45 aptamers exhibit differing specificities to BIA structural analogues. (A) Structures of the three BIA molecules, codeine, thebaine, and morphine, used in examining aptamer specificity. (B) Specificity elution profiles of the FC5 and FC45 aptamers. Radiolabeled aptamers were incubated with a codeine-modified Sepharose matrix. The bound aptamers were subsequently eluted with the different BIA targets and radioactivity levels in the eluted fractions were measured. Radioactivity levels were normalized with respect to values obtained from the codeine elutions for each aptamer.

2.2.5. Characterization of mini-aptamers that demonstrate binding affinities similar to the full-length aptamers

Truncation experiments were systematically performed on the full-length FC5 and FC45 aptamers to identify minimal aptamer domains, or mini-aptamers. Various truncated aptamer sequences were characterized for their codeine-binding properties. Truncated sequences that form well-defined secondary structures as predicted by mfold or

RNAstructure were selected for further analysis. The described SPR small molecule-aptamer binding assays were employed to determine the codeine-binding affinities of these truncated sequences.

An FC5 mini-aptamer was identified by characterizing three truncated sequences of the FC5 full-length aptamer. The codeine-binding properties of the random region (FC5Ran), which is the N30 region of the aptamer library; the cloning region (FC5Cln), which includes the random region, most of the 3' constant terminus, and part of the 5' constant terminus; and FC5L, which includes the random region and the 5' constant terminus, were analyzed using the described SPR binding assay. No binding was observed between FC5Ran and the codeine-immobilized sensor surface, indicating that the FC5 random region is not sufficient for the codeine-binding properties of this aptamer. FC5Cln demonstrated a significantly reduced affinity to codeine (K_d = 39.50 ± 2.27 μM), suggesting that the remainder of the 5' constant terminus of FC5 may play an important role in the formation of the correct binding pocket for codeine. FC5L binds codeine with an affinity similar to that of its full-length (59 nucleotides) parent sequence (K_d = 4.55 ± 0.14 μM) despite its significantly reduced length (41 nucleotides). These results indicate that FC5L, referred to as FC5 mini-aptamer, contains the necessary and sufficient sequence within FC5 for binding codeine (Figure 2.6, A and C).

An FC45 mini-aptamer was identified by characterizing two truncated sequences of the FC45 full-length aptamer. The codeine-binding properties of the cloning region (FC45Cln), which includes the random region, most of the 3' constant terminus, and part of the 5' constant terminus; and FC45L, which includes the random region and the 5' constant terminus, were analyzed using the described SPR binding assay. FC45Ran, harboring the N30 region of the library, was not analyzed in this set of truncation experiments, as there was

no well-defined secondary structure predicted for this sequence by mfold or RNAstructure. FC45Cln did not exhibit binding to codeine, suggesting that the codeine binding pocket was not correctly formed within the secondary structure adopted by this sequence. However, FC45L (44-nt) binds codeine with an affinity (2.59 ± 0.09 μM) that is almost identical to that of the full-length FC45 sequence (Figure 2.6, B and D). Therefore, this FC45 mini-aptamer includes the sequence within FC45 required to form the correct binding pocket for codeine in contrast to that of FC45Cln.

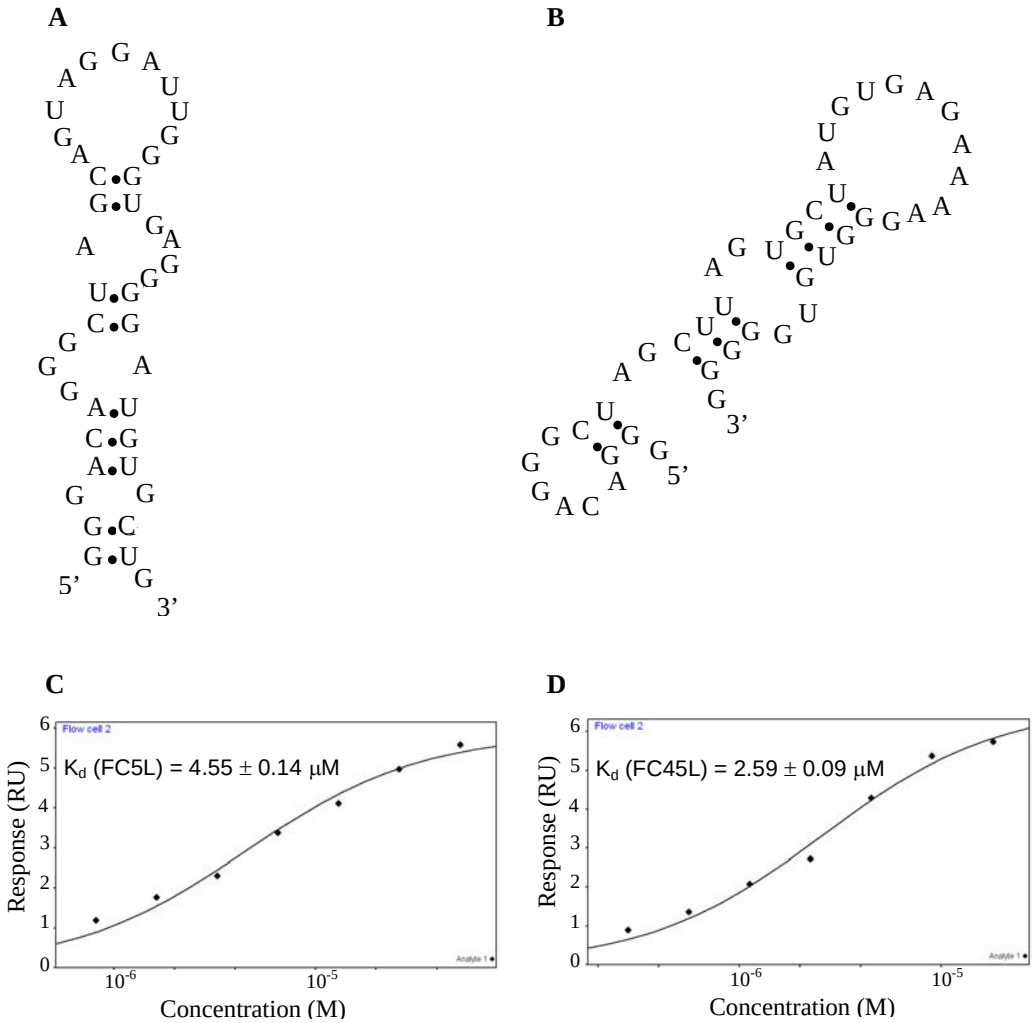


Figure 2.6. Codeine-binding mini-aptamer characterization. Proposed secondary structures from mfold of (A) the FC5 mini-aptamer (FC5L) and (B) the FC45 mini-aptamer (FC45L), and the corresponding equilibrium codeine-binding curves of (C) FC5L and (D) FC45L.

These truncation experiments support the importance of the formation of the correct binding pocket for aptamer molecular recognition capabilities. In addition, both FC5 and FC45 mini-aptamers lack the 3' constant terminal sequence, indicating that the 3' terminus is not involved in binding codeine. Secondary structure predictions from mfold and RNAstructure indicate that the 3' terminus forms a small hairpin (Supplementary Figure 2.2C), isolating itself from the remaining sequences of FC5 and FC45. The proposed secondary structures of the FC5 and FC45 mini-aptamers (Figure 2.6) are supported by the structural modification and structural probing experiments described in the next section.

2.2.6. Characterization of modified mini-aptamer sequences supports the proposed secondary structures

The proposed secondary structures of the FC5 and FC45 mini-aptamers do not possess a strong base stem (Figure 2.6, A and B) in comparison to other reported aptamer structures. For instance, the tetracycline minimer[33] has a base stem that is comprised of five base-pairs, which contribute to the stability of the overall secondary structure of the minimer. Sequences lacking strong or stabilized base stems may adopt a number of possible secondary structures, whereas a stabilized base stem can significantly reduce presumed structural variability and therefore restrict a given aptamer sequence to adopt a very few, and in some cases just one, distinct structures. Therefore, the proposed secondary structures of the FC5 and FC45 mini-aptamers may be evaluated by examining the binding properties of these aptamers modified with stabilized base stems.

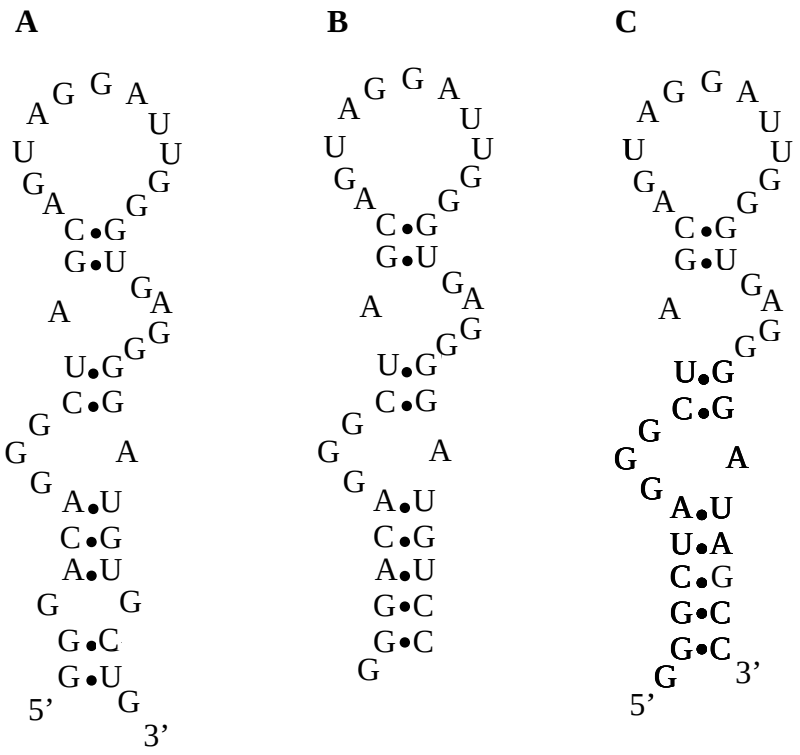


Figure 2.7. Structural stabilization and sequence requirements of the FC5 mini-aptamer stems. Proposed secondary structures from mfold of (A) the original FC5 mini-aptamer (FC5L), (B) the FC5 mini-aptamer with a stabilized base stem (FC5L-S1), (C) the FC5 mini-aptamer with a stabilized base stem composed of randomly-selected nucleotides (FC5L-S2).

The base stems of the mini-aptamers were modified with an extension of GC base-pairs to stabilize the proposed structures of these mini-aptamers. The FC5 mini-aptamer (FC5L) was stabilized by extending the existing three base-pair stem with two GC base-pairs (Figure 2.7B), based on the assumption that a few nucleotides present on each end of the original mini-aptamer are unessential for codeine binding. Similarly, the FC45 mini-aptamer (FC45L) was stabilized by extending the base stem formed by the 5’-CUU and 3’-GGG pairing with two GC base-pairs (Figure 2.8B), excluding several nucleotides from the 5’ end.

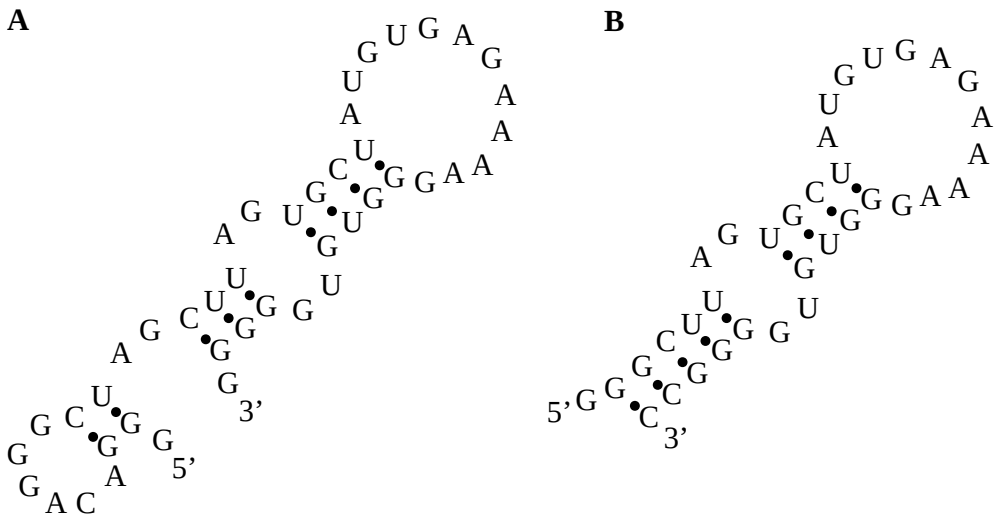


Figure 2.8. Structural stabilization of the FC45 mini-aptamer. Proposed secondary structures from mfold of (A) the original FC45 mini-aptamer (FC45L) and (B) the FC45 mini-aptamer with a stabilized base stem (FC45L-S1) in which several nucleotides at the termini of the original mini-aptamer are truncated.

Following the modification, the structures of these stabilized mini-aptamers were further analyzed in mfold using the DotPlot Partition Function, which confirms these structures to be the most favorable ones to adopt among others. The codeine-binding properties of the resulting mini-aptamers, referred to as FC5L-S1 and FC45L-S1, respectively, were determined using the described SPR assay. FC5L-S1 and FC45L-S1 were determined to bind codeine with K_d values of 5.51 ± 0.23 μM and 4.18 ± 0.48 μM, respectively (Supplementary Figure 2.3, A and B). These results indicate that the modified mini-aptamers bind the target molecule codeine with affinities similar to the corresponding unmodified mini-aptamers. Therefore, these results support the proposed secondary structures of the FC5 and FC45 mini-aptamers (Figure 2.6) and that their codeine-binding affinities were minimally affected by extending the original base stems. Structural probing studies were performed on FC5 and FC45 full-length aptamers using a standard lead-based cleavage assay to confirm the structures predicted through the SPR analysis. Lead-induced

and RNase T1 cleavage patterns were observed to be in agreement with the corresponding proposed structures (Supplementary Figure 2.4).

The sequence requirements and flexibility of the mini-aptamer base stems were examined with directed mutational analysis coupled with characterization of the effects of these sequence changes on the codeine-binding properties of these aptamers by the described SPR assays. Two of the three original base-pairs in the base stem of the FC5 mini-aptamer were replaced with randomly selected base-pairs (Figure 2.7C). This new sequence (FC5L-S2) was determined to bind codeine with an affinity (K_d = 5.39 ± 0.28 μM) (Supplementary Figure 2.3C) comparable to that of the original aptamer sequence, indicating that while the presence of the base stem is essential for codeine-binding, its sequence is not. The sequence space flexibility demonstrated for the aptamer base stem of FC5L has been reported in other aptamers such as the theophylline aptamer[31].

Studies were also conducted to demonstrate that the formation of the correct binding pocket within a given aptamer sequence is highly dictated by the formation of the correct base stem. The base stems of two alternative secondary structures for the FC45 mini-aptamer were extended with two GC-pairs to stabilize these proposed secondary structures (Supplementary Figure 2.2, A and B), in the same way as previously described for FC45L-S1. Binding assays revealed that these structures did not bind codeine with as high affinity as the initially proposed structure. The K_d values of these alternative FC45 mini-aptamer structures were increased approximately 10-fold (~25 μM), indicating that the codeine-binding pocket may be somewhat disrupted in these structures. These results indicate that the formation of the correct base stem can have significant influence on the formation of the correct binding pocket for aptamer recognition events.

2.2.7. Validation of the direct coupling SPR assay for characterization of small molecule-aptamer binding properties

Biacore assays are widely used to study a variety of molecular interactions such as RNA-protein and protein-protein interactions. While these assays are applicable to a broad range of target molecules, proteins have most often been used as the primary targets. Although Biacore assays have been used to measure the interaction between aptamers and non-protein targets, these assays often include a carrier or linker protein between the target and the sensor surface[24, 26]. However, significant discrepancies have been observed in K_d values determined from these assays and other commonly used methods potentially due to the use of a linker protein between the dextran surface and the small molecule. While SPR assays in which the small-molecule target is directly coupled to the sensor surface without inclusion of a linker protein have been previously reported[22, 23], the observed binding properties have not been validated or proven to potentially eliminate non-specific interactions or artifacts that may arise from the presence of a linker protein used in the assay. Therefore, experiments were conducted to examine the reproducibility, accuracy, and versatility of these direct coupling assays.

Table 2.3. Codeine-binding affinities of several full-length aptamers determined from replicate SPR binding assays for method reproducibility assessment.

RNA sample	K_d (first trial)	K_d (second trial)
initial pool	no binding response	no binding response
final pool	15.20 ± 0.38 μM	15.00 ± 0.30 μM
FC5	4.00 ± 0.13 μM	4.15 ± 0.10 μM
FC34	28.00 ± 1.42 μM	27.80 ± 1.37 μM
FC45	2.50 ± 0.06 μM	2.45 ± 0.10 μM
A25	7.23 ± 0.34 μM	7.17 ± 1.12 μM

The reproducibility of the assay method was confirmed through several means. Binding assays were repeated for several samples: two randomly-selected sequences (FC34 and A25), the initial pool, FC5, and FC45. The codeine-binding affinities determined from these replicate experiments were nearly identical, thereby confirming the reproducibility of the assay method (Table 2.3). In addition, the assay was repeated for the initial pool, FC5, and FC45 such that the concentration series sets of these samples were injected into the flow cells in a random order. Consistent K_d values (data not shown) were obtained from the random-injection experiments for all three of the tested RNA samples when compared to the values obtained from injecting them sequentially from lowest to highest concentrations. These results demonstrate the reproducibility and the robustness of this direct coupling assay method. FC45 was used as a positive control when performing the described assays on the remainder of the RNA aptamer sequences.

The potential elimination of non-specific interactions between an aptamer and the linker protein by the direct coupling small molecule-aptamer binding assay was demonstrated on a previously characterized RNA aptamer to a different small-molecule target. The described SPR binding assay was performed on a previously characterized dopamine aptamer (dopa2)[18], whose reported K_d value was determined through commonly used solution-based affinity methods. Dopamine was immobilized onto the sensor chip through the same coupling chemistry that was used in the original selection of this dopamine-binding aptamer. The binding affinity determined through the direct coupling SPR assay of the dopa2 RNA aptamer to dopamine ($K_d = 2.71 \pm 0.06$ μM) was nearly identical to the reported value of 2.8 μM[18] (Figure 2.9A).

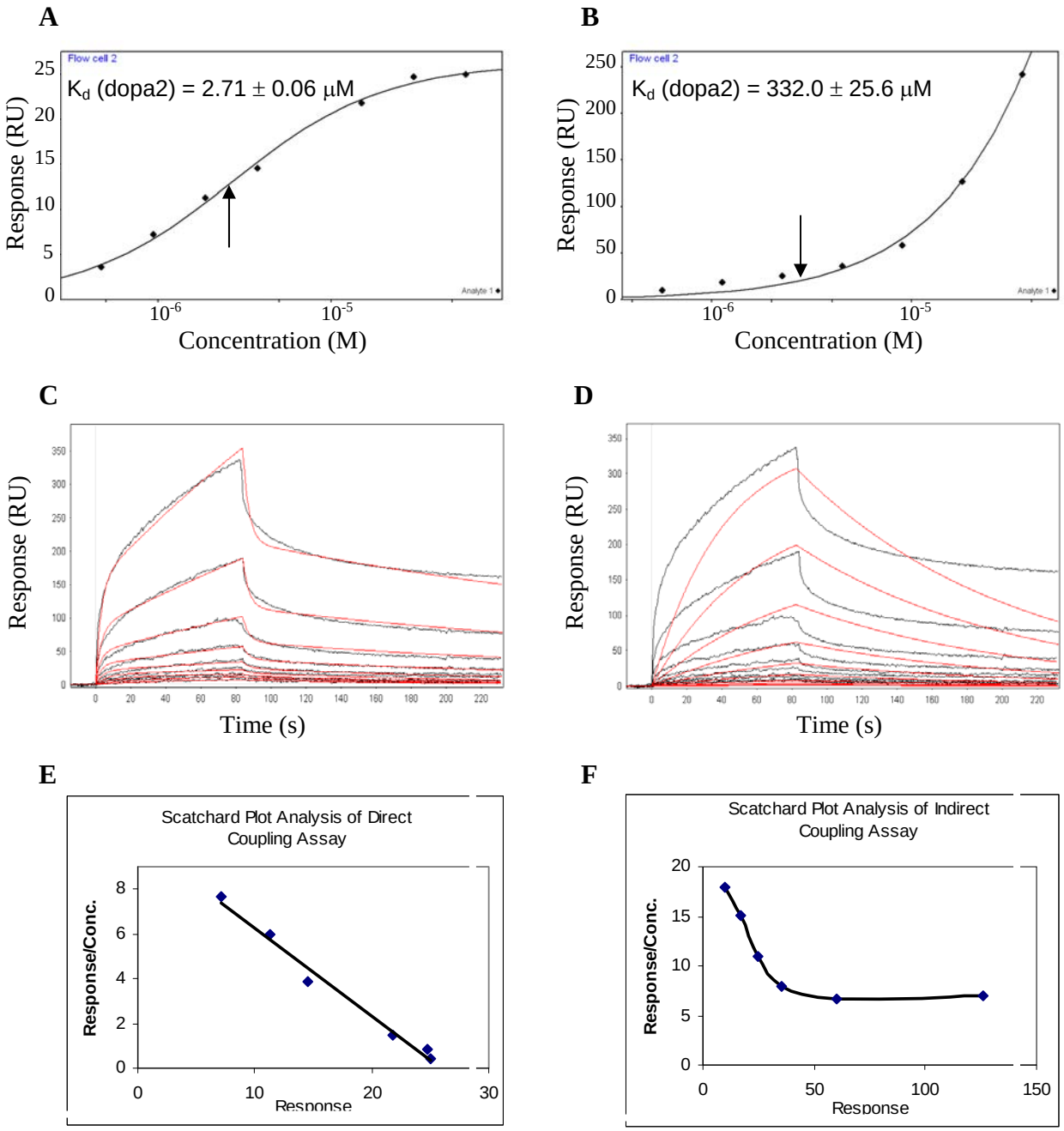


Figure 2.9. Validation of the direct coupling SPR assay. Equilibrium binding curve of the dopamine-dopa2 RNA aptamer interaction obtained from (A) the direct coupling binding assay and (B) the indirect coupling binding assay using BSA as a protein linker for dopamine. The arrow in (A) indicates the inflection point of the equilibrium binding curve obtained from the direct coupling method and the determined K_d value is nearly identical to the reported value obtained from a commonly used characterization assay. The arrow in (B) indicates where the inflection point should have been if there were no non-specific interactions or artifacts arising from the presence of the protein linker. Kinetic data analysis shows that the kinetics of the binding responses obtained from the indirect coupling method satisfy (C) a multiple binding site model rather than (D) a one-to-one binding model.

Scatchard plot analysis suggests (E) the direct coupling system as a single binding site system and (F) the indirect coupling system as a multiple binding site system.

An indirect coupling assay was performed on the dopamine aptamer using BSA as a protein linker to demonstrate that the presence of a linker protein in a SPR small molecule-aptamer binding assay may generate non-specific interactions or artifacts. It was observed that aptamer samples at the same concentrations take considerably longer to reach an equilibrium binding response in the BSA linker assay versus the direct coupling assay, which is indicative of non-specific interactions. The two highest concentration samples reached a near equilibrium response after 50 min of injection, approximately 33 times longer than that employed in the direct coupling assay. Data analysis revealed that the aptamer binding affinity was significantly affected and resulted in a false assessment as the observed K_d value was substantially higher than the reported value of 2.8 μM (Figure 2.9B).

To better analyze the data, Clamp[33] was used to fit the kinetic binding responses, as the two highest concentration samples did not completely reach equilibrium. Kinetic data analysis suggested that multiple binding events are present in the BSA linker assay since the data were well-fit with a multiple binding site model and did not satisfy a one-to-one binding model (Figure 2.9, C and D). This finding was further supported by Scatchard plot analysis, which also suggested this indirect coupling assay as a multiple binding site system represented by a curvature in this plot, a hallmark of a multiple binding site model (Figure 2.9F). In contrast, the direct coupling data fit a single binding site system represented by a linear fit to this data (Figure 2.9E). These results indicate that the presence of a protein linker can cause an aptamer to bind to its surface-immobilized target molecule in a non-specific

manner, leading to an inaccurate assessment of the binding affinity of the aptamer to its small-molecule target.

2.3. Discussion

In this study, we employed *in vitro* selection strategies to isolate RNA aptamers with high affinity and specificity to a subclass of BIA molecules, including codeine, within 15 selection cycles. A counter-selection with morphine and three error-prone PCR steps were incorporated into the selection process to enhance the specificity and affinity of the selected aptamers for their target molecule. The qualitative binding assays revealed that the final aptamer pool was highly enriched with codeine-binding affinity. The binding affinity of the enriched aptamer pool was determined to be 15 μM when characterized through the described Biacore assay; however, several of its member sequences, including FC5 and FC45, were determined to have higher affinities to codeine. In addition both of these aptamers were shown to be highly specific to codeine over morphine, indicating that the morphine counter-selection performed during the selection process was effective at enhancing the desired target specificity of the aptamers. Interestingly, while FC45 maintains codeine-binding specificity over another structural analogue, thebaine, FC5 demonstrates higher specificity to the latter. Therefore, these aptamers exhibit differing specificities to BIA alkaloid molecules, displaying molecular discrimination between targets differing by a single methyl group or structural conformation.

This work also highlights a direct coupling SPR binding assay for accurately and robustly determining the binding properties of aptamers to small-molecule ligands. The described method is based on the direct immobilization of the target small molecule onto the

sensor chip surface without inclusion of a linker protein as is commonly used. This direct coupling may provide a more accurate assessment of the binding affinity between the small-molecule target and the aptamer by eliminating potential non-specific binding between the nucleic acid aptamer and the protein linker and more accurately reproducing conditions used in the selection process. Significant discrepancies have been observed between reported K_d values obtained from Biacore assays that employ a protein linker connecting the target molecule to the sensor surface and other methods that involve direct target coupling. In one example, BSA was used as a linker between a target carbohydrate and the sensor surface[24]. The binding affinity of a selected aptamer was reported as 85 pM using this assay method. However, when the aptamer was immobilized onto the sensor surface and target molecules were injected over the surface, the binding affinity to the BSA-linked target was similar to that observed with the earlier experimental setup (K_d = 57 pM), whereas the binding affinity to the target molecule alone was determined to be approximately 60-fold lower (K_d = 3.3 nM). In another example, an existing tobramycin aptamer, characterized with a Biacore binding assay using a streptavidin linker, showed a lower degree of selectivity and significantly reduced affinity[25] from previously reported binding properties for this aptamer determined using a number of different assay methods[34-37]. These results indicate that the presence of a protein linker may introduce artifacts or non-specificity in the small molecule-aptamer interaction, preventing an accurate assessment of the intact affinity of the aptamer to its target molecule. Direct coupling of the small-molecule target onto the sensor surface may provide a more accurate assessment of small molecule-aptamer binding properties by eliminating potential non-specific interactions or artifacts introduced when using a linker protein.

The direct coupling SPR small molecule-aptamer binding assay has the additional benefit of providing a rapid characterization assay. In comparison to other commonly used binding assays, such as isocratic elution or equilibrium filtration, Biacore assays offer a rapid, high-throughput platform, which provides information about both equilibrium and kinetic binding properties. Using the Biacore 2000 and the serial dilution method described in this work, the binding properties of as many as eight aptamer sequences may be accurately and precisely determined in one day on a single chip. It should be noted, that the binding properties determined through this assay correspond to ligand-immobilized binding properties, which may differ from free ligand binding properties depending on the particular aptamer-ligand pair as demonstrated in this and previous work. However, this high-throughput assay strategy may be particularly useful when applied to the screening of libraries for aptamers that exhibit particular binding properties. From this initial screen, those aptamers exhibiting desired binding affinities for surface-immobilized targets may be further analyzed with standard solution affinity assays to determine and verify the corresponding binding affinities of those selected aptamers to free target in solution. Furthermore, the high-throughput nature of this platform may be used to rapidly determine the mini-aptamers for selected aptamers through truncation experiments, eliminating the need to perform time-consuming and labor-intensive chemical probing experiments[18, 38]. In addition, while traditional binding assays involve the use of radiolabeled aptamers or often rare and expensive radiolabeled target molecules, Biacore assays eliminate this requirement. The same assay methodology may be employed to perform structural stabilization studies, which were used to develop mini-aptamers with stabilized base stems. Aptamers with stabilized, modifiable, and extendable base stems are more functionally attractive for applications in

downstream molecular design strategies that involve exploiting structural rearrangements associated with the base stem formation[17]. Therefore, the FC5 and FC45 mini-aptamers may be readily employed in molecular engineering applications as their stems are extendable and modifiable. Finally, the versatility of this direct coupling SPR assay to the study of small molecule-aptamer interactions was demonstrated through several means and validated on a previously characterized dopamine RNA aptamer. Elimination of non-specific interactions was demonstrated in the direct coupling assay compared to the indirect coupling assay for the same aptamer, where non-specific interactions or binding artifacts arose in the presence of the linker protein. Therefore, the SPR assay discussed here is proven to be a rapid, versatile, accurate, and robust method for quantitative measurement of small molecule-RNA interactions.

2.4. Materials and Methods

2.4.1. DNA template library preparation

A random DNA library was generated through PCR using the following oligonucleotide sequences: a 59-nt DNA template 5'-GGGACAGGGCTAGC(N_{30})GAGGCAAAGCTT CCG-3', primer1 5'-TTC*TAATACGACTCACTATA*GGGACAGGGCTAGC-3', and primer2 5'-CGGAAGCTTTGCCTC-3'. All DNA synthesis was performed by Integrated DNA Technologies, Inc. The template contains a 30-nt randomized region flanked by two fixed primer-binding regions (Figure 2.1A). Primer1 contains a 17-nt T7 promoter sequence (italic). NheI and HindIII restriction endonuclease sites (underlined) were included in primer1 and primer2, respectively, for cloning of aptamer sequences.

2.4.2. Codeine coupling and affinity chromatography matrix preparation

Approximately 300 mg of epoxy-activated Sepharose 6B (GE Healthcare) was hydrated and incubated with 2.5 mM codeine in coupling buffer (0.05 M Na_2PO_4, pH 13) overnight at 37°C according to the manufacturer's instructions. The coupled medium was washed three times with 2 ml of coupling buffer to remove uncoupled codeine. The medium was then incubated overnight with 1 M Tris-HCl, pH 8 at 40°C to block any remaining active groups. Finally, the medium was washed with a solution containing 0.1 M NaOAc, pH 4 and 0.5 M NaCl followed by a second solution containing 0.1 M Tris-HCl, pH 8 and 0.5 M NaCl. The wash was repeated twice and the matrix was resuspended in 10 mM Tris-HCl, pH 8 and stored at 4°C. The codeine affinity chromatography matrix was prepared by packing the coupled medium (500 μl) into a column following the manufacturer's instructions (Pierce). The packed column was washed with 10 column volumes of binding buffer (250 mM NaCl, 20 mM Tris-HCl, pH 7.4, 5 mM $MgCl_2$) and equilibrated prior to the selection process.

2.4.3. Initial RNA library pool preparation

The initial DNA library pool was generated by PCR conducted for 12 cycles on a mixture (100 μl) containing 20 pmol DNA template, 300 pmol each primer1 and primer2, 200 μM each dNTPs, 1.6 mM $MgCl_2$, and 10 U Taq DNA polymerase (Roche). This DNA library pool (~$1.2x10^{14}$ molecules) was transcribed into an initial RNA library pool by incubating overnight at 37°C in the presence of 40 mM Tris-HCl, pH 7.9, 16 mM $MgCl_2$, 10 mM DTT, 2 mM spermidine, 3 mM each rNTPs, 50 μCi α-[^{32}P] UTP (GE Healthcare), 500 U RNase inhibitor, and 50 U T7 RNA polymerase (New England Biolabs). The DNA template was subsequently degraded by incubating the reaction mixture with 10 U of DNase

I (Invitrogen) at 37°C for 15 min. The unincorporated nucleotides were removed with a NucAway spin column (Ambion) following the manufacturer's instructions and binding buffer was added to the flow-through RNA to bring the total volume up to 500 μl.

2.4.4. In vitro selection of codeine-binding aptamers

Prior to incubation with the codeine-modified affinity column, the RNA pool was denatured at 70°C for 3 min and allowed to renature at room temperature for 30 min. To eliminate RNA molecules that non-specifically bind to the column matrix, the initial pool was first incubated with an unmodified column. The flow-through fraction from this incubation was subsequently transferred to a codeine-modified affinity column and incubated for 45 min. Following the incubation period, the affinity column was washed with 10 column volumes of binding buffer for cycles 1 to 5 to remove unbound RNAs. This wash volume was increased 10 column volumes for each of the subsequent cycles. Bound RNA was eluted with 7 column volumes of 5 mM codeine in binding buffer. The eluted RNA was recovered by ethanol precipitation in the presence of 20 μg/ml glycogen. Reverse transcription and cDNA amplification (15 PCR cycles) were performed in a single step using 200 U of SuperScript III reverse transcriptase (Invitrogen) and 5 U of Taq DNA polymerase in a 50 μl reaction volume. One-fifth of this DNA library was transcribed into an RNA library pool for the subsequent selection cycle. A total of 15 selection cycles were carried out during the *in vitro* selection process.

At the tenth cycle, a counter-selection against morphine was performed by eluting the bound RNA with 3 column volumes of 5 mM morphine in binding buffer prior to elution with codeine. Only RNA eluted with codeine was used to make the input DNA library pool

for the subsequent selection cycle. Following the reverse transcription step of cycles 11, 12, and 13, an error-prone PCR was performed in a mutagenic buffer containing 40 pmol each primer1 and primer2, 7 mM $MgCl_2$, 50 mM KCl, 10 mM Tris-HCl, pH 8.3, 0.2 mM dGTP, 0.2 mM dATP, 1 mM dCTP, 1 mM dTTP, and 0.5 mM $MnCl_2$. One fifth of the error-prone PCR product from each of these cycles was used as the input DNA library pool for the subsequent selection cycle.

2.4.5. Aptamer library sequence analysis

The DNA pool from cycle 15 was amplified by PCR and cloned into a plasmid using the NheI and HindIII restriction sites present in the fixed regions of the aptamer sequence and the plasmid construct. This plasmid library was transformed into an electrocompetent *Escherichia coli* strain, DH10B (Invitrogen; F- *mcr*A Δ(*mrr-hsd*RMS-*mcr*BC) ϕ80d*lac*ZΔM15 Δ*lac*X74 *deo*R *rec*A1 *end*A1 *ara*D139 Δ(*ara, leu*)7697 *gal*U *gal*K λ- *rps*L *nup*G). Subcloning was confirmed by colony PCR, and a total of 58 positive colonies were sequenced by Laragen, Inc. The resulting sequences were aligned using the ClustalX sequence alignment program.

2.4.6. Qualitative binding affinity assay

Radiolabeled RNA was prepared from approximately 1 μg of the final DNA pool (cycle 15) in the presence of 40 mM Tris-HCl, pH 7.9, 14 mM $MgCl_2$, 10 mM DTT, 2 mM spermidine, 3 mM each rAGC mix, 150 μM rUTP, 50 μCi α-[^{32}P] UTP, 40 U RNase inhibitor, and 50 U T7 RNA polymerase. After allowing the transcription reaction to proceed for 3 h at 37°C, 5 U of DNase I were added to the mixture and the reaction was incubated for

15 min. The unincorporated nucleotides were removed with a NucAway spin column and the flow-through RNA was divided equally into two volumes. One of the radiolabeled RNA pools was incubated with a codeine-modified column, whereas the other pool was incubated with an unmodified column. After a 15 min incubation, each column was washed with 3 column volumes of binding buffer followed by elution with 7 column volumes of 5 mM codeine in binding buffer. The eluted RNA from each column was separated by electrophoresis on an 8% polyacryamide/7 M urea gel in 1X Tris-borate buffer. The gel was dried and the recovered radiolabeled RNA was imaged on a FX phosphorimager (BioRAD).

2.4.7. Quantitative direct coupling small molecule-aptamer binding assay

A CM5 sensor chip was primed with RNase-free water followed by preconditioning with a 50 mM sodium hydroxide, 0.1% hydrochloric acid, 0.1% (w/v) sodium dodecyl sulfate, 0.085% phosphoric acid solution prior to immobilization of codeine onto the chip surface. The chip was subsequently activated with a 0.2 M N-ethyl-N'-(dimethylaminopropyl)carbodiimide (EDC), 0.05 M N-hydroxysuccinimide (NHS) solution. An amine surface was created by injecting a solution of 0.1 M 1,8-diaminooctane dissolved in 50 mM sodium borate, pH 8.5 over the activated sensor chip at 5 μl/min for 10 min. In order to couple codeine to the amine surface, codeine was modified at its hydroxyl group with a succinimidyl group by placing 10 mM codeine in a pyridine solution containing 40 mM disuccinimidyl carbonate and 40 mM 4-dimethylamino pyridine. This modification reaction was allowed to take place for 30 min and the reaction mixture was subsequently diluted with 100 mM sodium borate, pH 7.0 in a 1:1 v/v ratio. Trenbolone (Figure 2.2B), a small molecule structurally distinct from codeine, was modified in the same manner for use

as a background response. The modified trenbolone and codeine molecules were separately coupled onto flow cells 1 and 2 of the sensor chip, respectively, by alternating injections for 7 min at 5 μl/min for a total of 28 min for each molecule. After ligand coupling, the chip was deactivated with 1 M ethanolamine, pH 8.5 and primed twice with binding buffer.

RNA samples (initial pool, final pool, and randomly-selected individual sequences from the final pool) were prepared for Biacore analysis using the Ampliscribe T7 High Yield Transcription Kit (Epicentre) following the manufacturer's instructions. Samples were sequentially injected over the sensor surface for 1.5 min at 5 μl/min with a 2 min dissociation time. For each sample, various RNA concentrations were injected by serially diluting samples from 48 μM to 0.375 μM along with two blank samples containing just binding buffer for use as double referencing. After each run, the surface was regenerated with 10 mM EGTA for 2 min at 5 μl/min. The raw data were processed and analyzed to determine the binding constant for each aptamer using Scrubber (Biologic Software, Pty, Australia, http://www.cores.utah.edu/interaction/).

2.4.8. Isocratic affinity elution and specificity assays

Radiolabeled FC5 and FC45 RNA were prepared using the Ampliscribe T7 High Yield Transcription Kit with minor modifications to the manufacturer's instructions (3 mM each rATP, rCTP, rUTP, 150 μM rGTP, and 50 μCi α-[^{32}P] GTP). After 3 h of incubation, DNase I was added to the transcription mixture and the reaction was incubated at 37°C for 15 min. Unincorporated nucleotides were removed with a NucAway spin column.

Isocratic affinity elution assays were performed on radiolabeled FC5 and FC45 as previously described[27, 28]. The binding affinities to codeine and morphine in solution were

determined using the following equation: $K_d = [L_{el}] \times (V_{el}–V_n) / (V_e–V_{el})$, where L_{el} is the free ligand concentration used to elute bound RNA, V_{el} and V_e are the elution volumes for RNA in the presence and absence of free ligand in binding buffer, respectively, and V_n is the column void volume.

Specificity assays were performed by equally dividing the flow-through radiolabeled FC5 and FC45 into three Sepharose columns (300 μl) modified with codeine. After a 30 min incubation, each column was washed with 7 column volumes of binding buffer. Columns were then eluted with a 5 mM solution of the different targets (codeine, morphine, or thebaine) in binding buffer, and 5 column volumes of the elution were collected. Collected samples were added to 10 ml of Safety-Solve scintillation liquid (Reseach Products International Corp.) and radioactivity levels were measured on a liquid scintillation counter (Beckman Coulter).

2.4.9. Truncation experiments

Two full-length aptamers with the lowest determined K_d values were truncated primarily into four different sequences containing distinct regions of their parent sequences: (1) the random region (Ran), (2) the cloning region (Cln), (3) the random region and the 5' constant terminus (L), and (4) the random region and the 3' constant terminus (R). Predicted secondary structures formed by these truncated sequences were examined using mfold[39] and RNAstructure (http://rna.chem.rochester.edu/RNAstructure.html). Sequences that adopt well-defined secondary structures were selected for subsequent K_d determination through the described small molecule-aptamer binding affinity SPR assay.

2.4.10. Structural probing assay

Structural probing of the FC5 and FC45 full-length aptamers was performed using a lead ion cleavage assay as described by Berens et al.[38] with the following slight modifications. 5'-end labeled RNA was incubated in binding buffer containing 0-250 μM codeine and 0.5 mM lead (II) acetate. After a 15 min incubation, the cleavage reactions were stopped by adding 0.5 mM EDTA and 1 μg/μl glycogen and the cleaved RNA was recovered by ethanol precipitation. Radiolabeled RNA was also subject to RNase T1 cleavage (Ambion) and alkaline hydrolysis (Ambion) following the manufacturer's instructions to be used as ladders. The recovered RNA samples were separated by electrophoresis on a 10% polyacryamide/8 M urea gel in 1x Tris-borate buffer. The gel was dried and the RNA cleavage patterns were imaged on a FX phosphorimager (BioRAD).

2.4.11. Dopamine aptamer binding assay

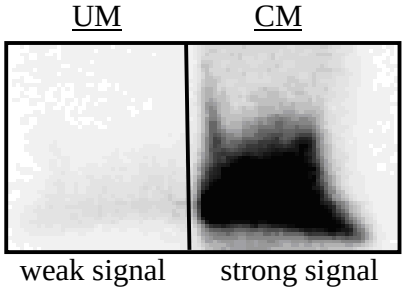
For direct coupling of dopamine to the sensor surface, a CM5 sensor chip was activated with EDC/NHS as described above. Following the EDC/NHS activation step, a 10 mM dopamine, 50 mM sodium borate, pH 8.5 solution was injected over the activated sensor surface for 30 min at 5 μl/min to couple dopamine to the surface through its amino group. This is the same chemistry used in the selection of dopamine-binding aptamers described by Mannironi et al.[18]. After dopamine immobilization, the sensor surface was deactivated with 1 M ethanolamine for 10 min at 5 μl/min to block the remaining unreacted succinimidyl groups. The previously selected dopamine-binding dopa2 RNA aptamer[18] was synthesized using a similar transcription procedure as described above. Various concentrations of this RNA sample were injected over the dopamine-coupled sensor surface and concentration-

dependent binding responses were recorded and subsequently analyzed for binding properties as described previously.

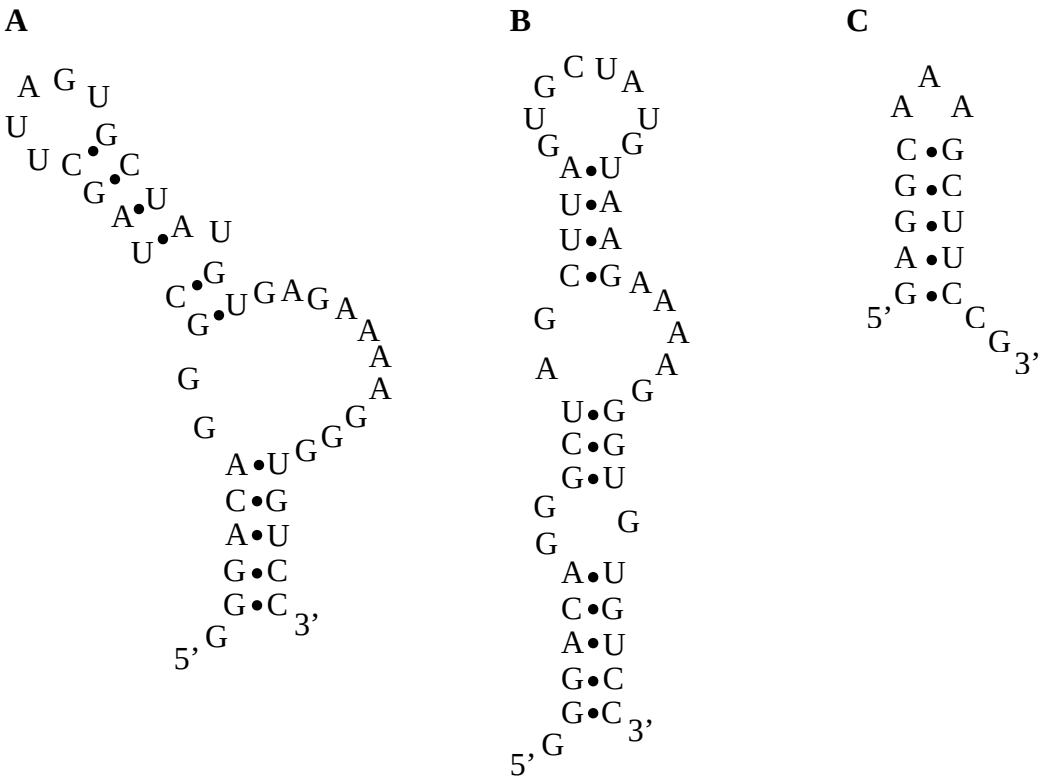
For indirect coupling of dopamine to the sensor surface through a BSA protein linker, a CM5 sensor chip was activated with EDC/NHS as described above. Following the EDC/NHS activation step, BSA was injected over the activated surface at 5 μl/min until a signal of 12,500 response units (RU) was reached. A 0.2 M EDC and 0.1 M dopamine solution was injected over the BSA-immobilized surface for 30 min at 5 μl/min to couple dopamine to BSA. This chemistry couples dopamine to BSA through the same functional group as in the direct coupling chemistry. The remaining steps in the indirect coupling method are identical to those used in the direct coupling method.

2.5. Supplementary Information

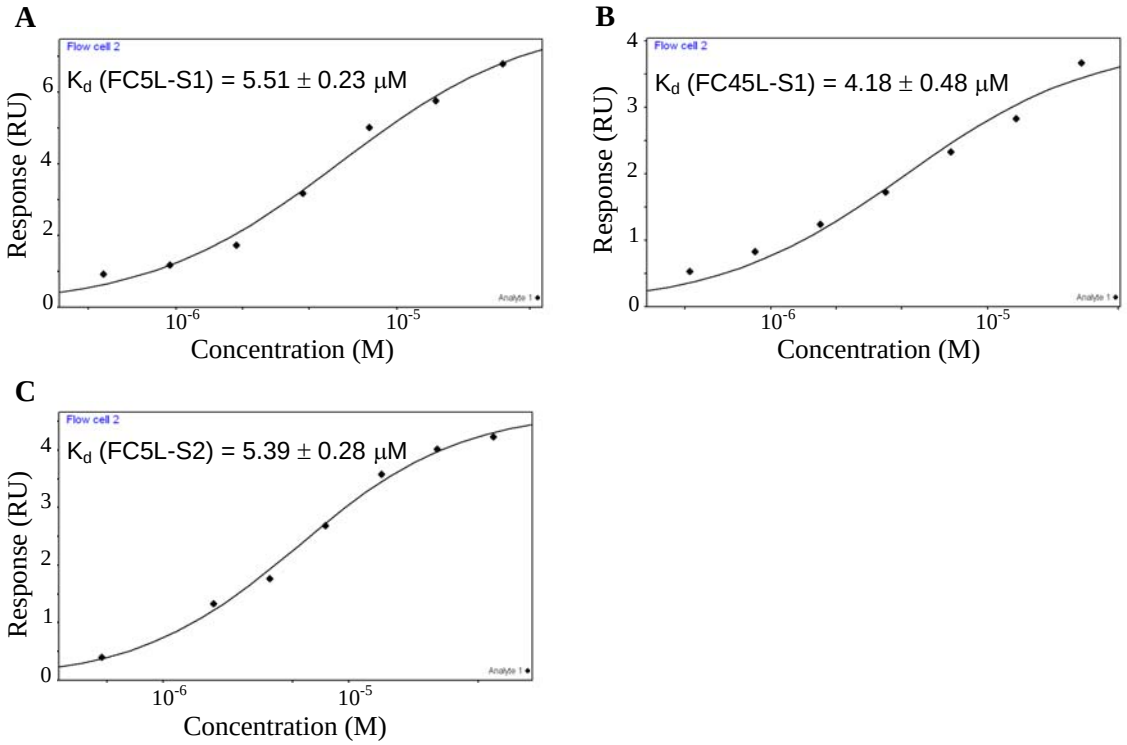
Supplementary Figures



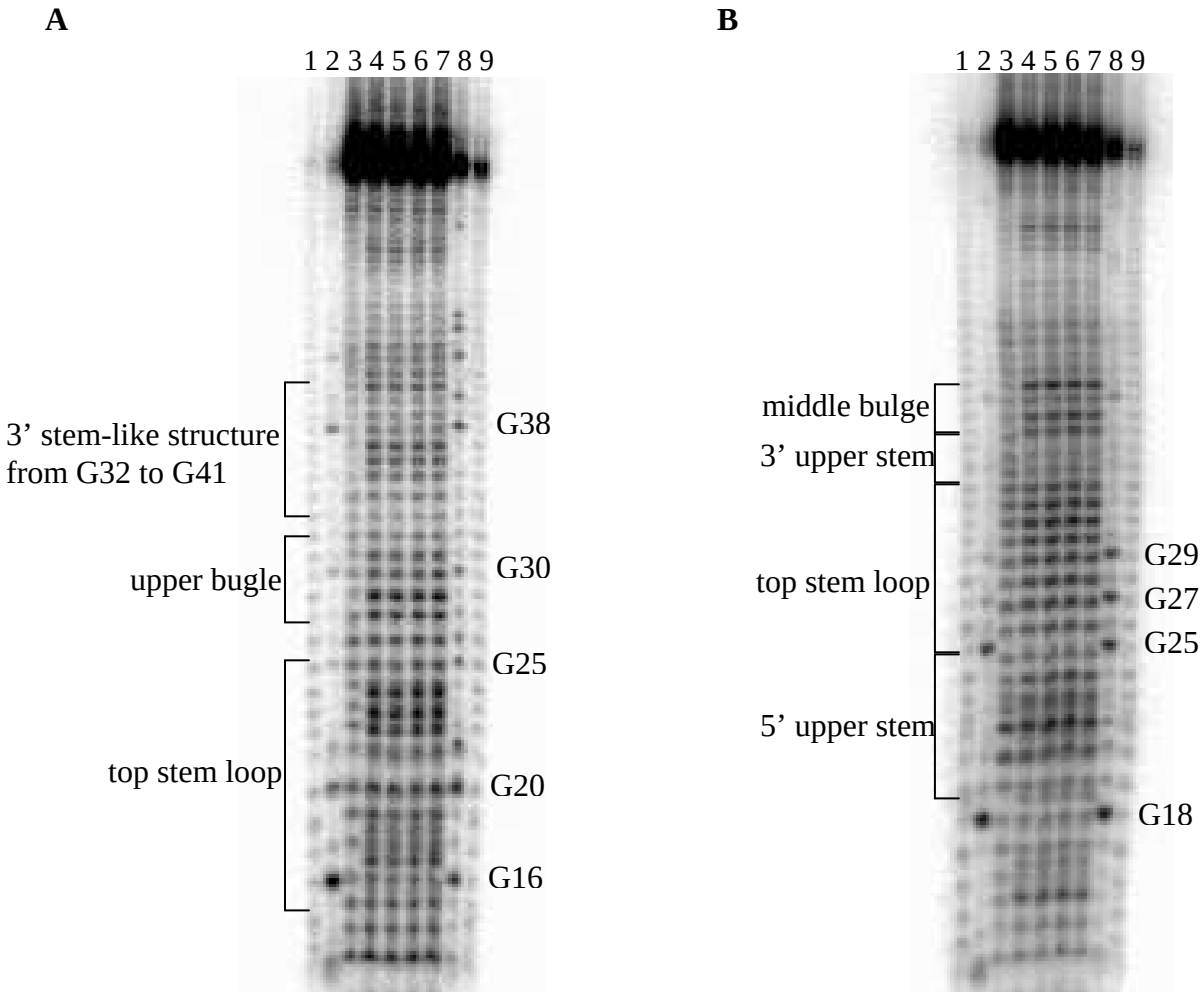
Supplementary Figure 2.1. Qualitative assessment of the enrichment in codeine-binding affinity of the final aptamer pool. The left column (weak) and the right column (strong) represent the signals obtained from the eluted pool incubated with an unmodified (UM) and codeine-modified column (CM), respectively.



Supplementary Figure 2.2. Alternative FC45 mini-aptamer structures support the proposed structure of the FC45 mini-aptamer. Alternative secondary structures proposed by mfold for the FC45 mini-aptamer sequence were examined by stabilizing the base stems of these structures with an extension of two GC base-pairs as illustrated in (A) and (B). SPR assays demonstrated that these stabilized alternative structures did not exhibit significant codeine-binding affinities and therefore do not allow for the formation of the correct codeine-binding pocket. (C) Formation of a small hairpin by the 3' constant terminus proposed by mfold.



Supplementary Figure 2.3. Structural stabilization studies of FC5 and FC45 mini-aptamers. Equilibrium codeine-binding curves of (A) FC5L-S1, (B) FC45L-S1, and (C) FC5L-S2 from their corresponding Biacore binding assays.



Supplementary Figure 2.4. Structural probing results of **(A)** FC5 and **(B)** FC45 through lead-induced and RNase T1 cleavage patterns. Lanes 1, 9: alkaline hydrolysis with 5 min and 15 min reaction times, respectively; lanes 2, 8: RNase T1 cleavage patterns using 1 μl and 0.1 μl of the enzyme, respectively; lane 3: intact RNA; lanes 4-7: lead-induced cleavage patterns in the presence of increasing codeine concentrations from 0-250 μM.

Acknowledgements

The authors thank Professor Pamela J. Bjorkman for generously allowing us the use of the Biacore 2000 instrument, which was used to develop the direct small molecule-aptamer characterization assays. The authors also gratefully acknowledge Dr. Laure Jason-Moller (Biacore Life Sciences) and Katie Saliba (Caltech Chemistry Department) for their

technical support and advice regarding codeine coupling chemistry to the sensor chip. The authors also thank Dr. Kevin Hoff (Caltech Chemical Engineering Department) for his critical reading of the manuscript. This work was funded by Caltech startup funds, the Center for Biological Circuit Design at Caltech (fellowship support for M.N.W.), and the National Institutes of Health (training grant for J.S.K.).

References

1. William, D. G., Hatch, D. J. & Howard, R. F. Codeine phosphate in paediatric medicine. Br J Anaesth 86, 413-21 (2001).
2. Williams, D. G., Patel, A. & Howard, R. F. Pharmacogenetics of codeine metabolism in an urban population of children and its implications for analgesic reliability. Br J Anaesth 89, 839-45 (2002).
3. Kim, I. et al. Plasma and oral fluid pharmacokinetics and pharmacodynamics after oral codeine administration. Clin Chem 48, 1486-96 (2002).
4. Jayasena, S. D. Aptamers: an emerging class of molecules that rival antibodies in diagnostics. Clin Chem 45, 1628-50 (1999).
5. Silverman, S. K. Rube Goldberg goes (ribo)nuclear? Molecular switches and sensors made from RNA. Rna 9, 377-83 (2003).
6. O'Sullivan, C. K. Aptasensors--the future of biosensing? Anal Bioanal Chem 372, 44-8 (2002).
7. Hesselberth, J. R., Robertson, M. P., Knudsen, S. M. & Ellington, A. D. Simultaneous detection of diverse analytes with an aptazyme ligase array. Anal Biochem 312, 106-12 (2003).

8. Seetharaman, S., Zivarts, M., Sudarsan, N. & Breaker, R. R. Immobilized RNA switches for the analysis of complex chemical and biological mixtures. Nat Biotechnol 19, 336-41 (2001).

9. Facchini, P. J. ALKALOID BIOSYNTHESIS IN PLANTS: Biochemistry, Cell Biology, Molecular Regulation, and Metabolic Engineering Applications. Annu Rev Plant Physiol Plant Mol Biol 52, 29-66 (2001).

10. Allen, R. S. et al. RNAi-mediated replacement of morphine with the nonnarcotic alkaloid reticuline in opium poppy. Nat Biotechnol 22, 1559-66 (2004).

11. Rathbone, D. A. & Bruce, N. C. Microbial transformation of alkaloids. Curr Opin Microbiol 5, 274-81 (2002).

12. Suess, B. et al. Conditional gene expression by controlling translation with tetracycline-binding aptamers. Nucleic Acids Res 31, 1853-8 (2003).

13. Suess, B., Fink, B., Berens, C., Stentz, R. & Hillen, W. A theophylline responsive riboswitch based on helix slipping controls gene expression in vivo. Nucleic Acids Res 32, 1610-4 (2004).

14. Werstuck, G. & Green, M. R. Controlling gene expression in living cells through small molecule-RNA interactions. Science 282, 296-8 (1998).

15. Desai, S. K. & Gallivan, J. P. Genetic screens and selections for small molecules based on a synthetic riboswitch that activates protein translation. J Am Chem Soc 126, 13247-54 (2004).

16. Winkler, W. C. & Breaker, R. R. Regulation of bacterial gene expression by riboswitches. Annu Rev Microbiol 59, 487-517 (2005).

17. Bayer, T. S. & Smolke, C. D. Programmable ligand-controlled riboregulators of eukaryotic gene expression. Nat Biotechnol 23, 337-43 (2005).

18. Mannironi, C., Di Nardo, A., Fruscoloni, P. & Tocchini-Valentini, G. P. In vitro selection of dopamine RNA ligands. Biochemistry 36, 9726-34 (1997).

19. Mannironi, C., Scerch, C., Fruscoloni, P. & Tocchini-Valentini, G. P. Molecular recognition of amino acids by RNA aptamers: the evolution into an L-tyrosine binder of a dopamine-binding RNA motif. Rna 6, 520-7 (2000).

20. Tuerk, C. & Gold, L. Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. Science 249, 505-10 (1990).

21. Ellington, A. D. & Szostak, J. W. In vitro selection of RNA molecules that bind specific ligands. Nature 346, 818-22 (1990).

22. Kwon, M., Chun, S. M., Jeong, S. & Yu, J. In vitro selection of RNA against kanamycin B. Mol Cells 11, 303-11 (2001).

23. Gebhardt, K., Shokraei, A., Babaie, E. & Lindqvist, B. H. RNA aptamers to S-adenosylhomocysteine: kinetic properties, divalent cation dependency, and comparison with anti-S-adenosylhomocysteine antibody. Biochemistry 39, 7255-65 (2000).

24. Jeong, S., Eom, T., Kim, S., Lee, S. & Yu, J. In vitro selection of the RNA aptamer against the Sialyl Lewis X and its inhibition of the cell adhesion. Biochem Biophys Res Commun 281, 237-43 (2001).

25. Verhelst, S. H., Michiels, P. J., van der Marel, G. A., van Boeckel, C. A. & van Boom, J. H. Surface plasmon resonance evaluation of various aminoglycoside-RNA

hairpin interactions reveals low degree of selectivity. Chembiochem 5, 937-42 (2004).

26. Davis, J. H. & Szostak, J. W. Isolation of high-affinity GTP aptamers from partially structured RNA libraries. Proc Natl Acad Sci U S A 99, 11616-21 (2002).

27. Arnold, F. H. & Blanch, H. W. Analytical affinity chromatography. II. Rate theory and the measurement of biological binding kinetics. J Chromatogr 355, 13-27 (1986).

28. Brockstedt, U., Uzarowska, A., Montpetit, A., Pfau, W. & Labuda, D. In vitro evolution of RNA aptamers recognizing carcinogenic aromatic amines. Biochem Biophys Res Commun 313, 1004-8 (2004).

29. Lozupone, C., Changayil, S., Majerfeld, I. & Yarus, M. Selection of the simplest RNA that binds isoleucine. Rna 9, 1315-22 (2003).

30. Majerfeld, I. & Yarus, M. Isoleucine:RNA sites with associated coding sequences. Rna 4, 471-8 (1998).

31. Jenison, R. D., Gill, S. C., Pardi, A. & Polisky, B. High-resolution molecular discrimination by RNA. Science 263, 1425-9 (1994).

32. Haller, A. A. & Sarnow, P. In vitro selection of a 7-methyl-guanosine binding RNA that inhibits translation of capped mRNA molecules. Proc Natl Acad Sci U S A 94, 8521-6 (1997).

33. Myszka, D. G. & Morton, T. A. CLAMP: a biosensor kinetic data analysis program. Trends Biochem Sci 23, 149-50 (1998).

34. Wang, Y. & Rando, R. R. Specific binding of aminoglycoside antibiotics to RNA. Chem Biol 2, 281-90 (1995).

35. Wang, Y., Killian, J., Hamasaki, K. & Rando, R. R. RNA molecules that specifically and stoichiometrically bind aminoglycoside antibiotics with high affinities. Biochemistry 35, 12338-46 (1996).

36. Hamasaki, K., Killian, J., Cho, J. & Rando, R. R. Minimal RNA constructs that specifically bind aminoglycoside antibiotics with high affinities. Biochemistry 37, 656-63 (1998).

37. Cho, J., Hamasaki, K. & Rando, R. R. The binding site of a specific aminoglycoside binding RNA molecule. Biochemistry 37, 4985-92 (1998).

38. Berens, C., Thain, A. & Schroeder, R. A tetracycline-binding RNA aptamer. Bioorg Med Chem 9, 2549-56 (2001).

39. Zuker, M. Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res 31, 3406-15 (2003).

Chapter III: A modular and extensible RNA-based gene-regulatory platform for engineering cellular function*

Abstract

Engineered biological systems hold promise in addressing pressing human needs in chemical processing, energy production, materials construction, and maintenance and enhancement of human health and the environment. However, significant advancements in our ability to engineer biological systems have been limited by the foundational tools available for reporting on, responding to, and controlling intracellular components in living systems. Portable and scalable platforms are needed for the reliable construction of such communication and control systems across diverse organisms. We report an extensible RNA-based framework for engineering ligand-controlled gene regulatory systems, called ribozyme switches, that exhibit tunable regulation, design modularity, and target specificity. These switch platforms contain a sensor domain, comprised of an aptamer sequence, and an actuator domain, comprised of a hammerhead ribozyme sequence. We examined two modes of standardized information transmission between these domains and demonstrate a mechanism that allows for the reliable and modular assembly of functioning synthetic RNA switches and regulation of ribozyme activity in response to various effectors. In addition to demonstrating the first examples of small molecule-responsive, *in vivo* functional allosteric hammerhead ribozymes, this work describes a general approach for the construction of portable and scalable gene-regulatory systems. We demonstrate the versatility of the platform in implementing application-specific control systems for small molecule-mediated regulation of cell growth and non-invasive *in vivo* sensing of metabolite production.

*Reproduced with permission from: M. N. Win and C. D. Smolke. (2007) *Proc. Natl. Acad. Sci. U. S. A.*, 104, 14283-14288.

3.1. Introduction

Basic and applied biological research and biotechnology are limited by our ability to get information into and out from living systems, and to act on information inside living systems[1, 2]. For example, there are only a small number of inducible promoter systems available to provide control over gene expression in response to exogenous molecules[3, 4]. Many of the molecular inputs to these systems are not ideal for broad implementation, as they can be expensive and introduce undesired pleiotropic effects. In addition, broadly-applicable methods for getting information out of cells non-invasively has been limited to strategies that rely on protein and promoter fusions to fluorescent proteins, which enable researchers to monitor protein levels and localization and transcriptional outputs of networks, leaving a significant amount of the cellular information content currently inaccessible.

To address these challenges scalable platforms are needed for reporting on, responding to, and controlling any intracellular component in a living system. A striking example of a biological communication and control system is the class of RNA regulatory elements called riboswitches, comprised of distinct sensor and actuation (gene regulatory) functions, that control gene expression in response to specific ligand concentrations[5]. Building on these natural examples, engineered riboswitch elements have been developed for use as synthetic ligand-controlled gene regulatory systems[6-9]. However, these early examples of riboswitch engineering do not address the challenges posed above because they lack portability across organisms and systems, and their designs and construction do not support modularity and component reuse.

We set out to develop a universal and extensible RNA-based platform that will provide a framework for the reliable design and construction of gene regulatory systems that

can control the expression of specific target genes in response to various effector molecules. We implemented five engineering design principles (DPs) in addressing this challenge: *scalability* (DP1: a sensing platform enabling *de novo* generation of ligand-binding elements for implementation within the sensor domain); *portability* (DP2: a regulatory element that is independent of cell-specific machinery or regulatory mechanisms for implementation within the actuator domain); *utility* (DP3: a mechanism through which to modularly couple the control system to functional level components); *composability* (DP4: a mechanism by which to modularly couple the actuator and sensor domains without disrupting the activities of these individual elements); and *reliability* (DP5: a mechanism through which to standardize the transmission of information from the sensor domain to the actuator domain). A glossary of terms is available in Supplementary Text 3.1.

3.2. Results

3.2.1. Component specification for a scalable and portable gene-regulatory system

To satisfy the engineering design principle of scalability (DP1) we chose RNA aptamers[10], nucleic acid ligand-binding molecules, as the sensing platform for the universal control system. Our choice of sensing platform was driven by the proven versatility of RNA aptamers. Standard *in vitro* selection strategies or SELEX[11, 12] have been used to generate RNA aptamers *de novo* to a wide variety of ligands, including small molecules, peptides, and proteins[13]. In addition, the specificity and affinity of an aptamer can be tuned through the selection process to meet the specific performance requirements of a given application. The continued selection of new aptamers to appropriate cellular molecules that function under *in*

vivo conditions will enable these elements to be implemented as sensors in RNA-based control systems.

To satisfy the engineering design principle of portability (DP2) we chose the hammerhead ribozyme, a catalytic RNA, as the regulatory element in the universal control system. Our choice of regulatory element was driven by the ability of the hammerhead ribozyme to exhibit self-cleavage activity across various organisms and its demonstrated potential in biomedical and biotechnological applications owing to its small size, relative ease of design, and rapid kinetics[14]. The utility of hammerhead ribozymes as gene regulatory elements has been demonstrated in various systems[15-17]. In addition, several research groups have engineered a special class of synthetic hammerhead ribozymes referred to as allosteric hammerhead ribozymes that contain separate catalytic and ligand-binding domains, which interact in a ligand-dependent manner to control the activity of the ribozyme[18-21]. While this class of ribozymes enables a better control system due to the presence of the integrated ligand-binding domain, there has been no success in translating them to *in vivo* environments.

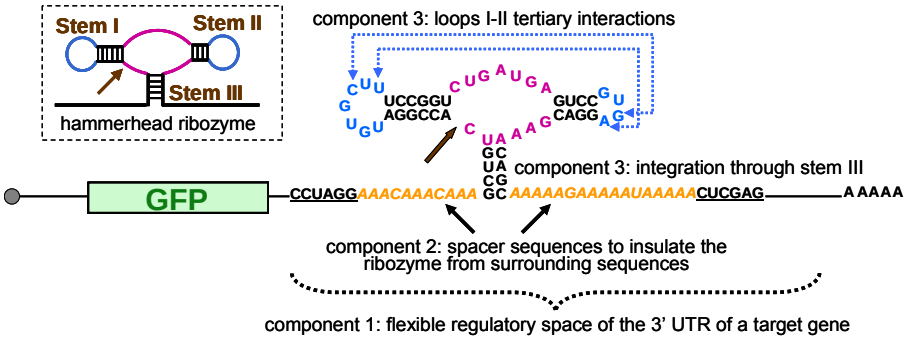
3.2.2. Design strategies for engineering portability, utility, and composability into a biological control system

To support a framework for engineering ligand-controlled gene regulatory systems, we specified a design strategy that is in accordance with our engineering principles stated above (Figure 3.1, A and B). This strategy is comprised of three components that address mechanisms for the portability (DP2), utility (DP3), and composability (DP4) of the control system and are critical to the development of a general ribozyme switch platform. First, the

cis-acting hammerhead ribozyme constructs are integrated into the flexible and portable regulatory space of the 3' UTR (Figure 3.1A). We chose to locate the synthetic ribozymes within the 3' UTR of their target gene as opposed to the 5' UTR in order to isolate their specific cleavage effects on transcript levels from their non-specific structural effects on translation initiation, as secondary structures have been demonstrated to repress efficient translation when placed in the 5' UTR[22]; K. Hawkins and C.D.S., unpublished observations). In addition, cleavage within the 3' UTR is a universal mechanism for transcript destabilization in eukaryotic and prokaryotic organisms. Second, each ribozyme construct is insulated from surrounding sequences, which may disrupt its structure and therefore its activity, by incorporating spacer sequences immediately 5' and 3' of stem III (Figure 3.1A). By implementing these two components, we ensure that these control systems will be portable across organisms and modular to coupling with different coding regions (Y. Chen and C.D.S., manuscript in preparation). The third component was necessitated by the fact that previous engineered *in vitro* allosteric ribozyme systems, which replace stem loops I or II with part of the aptamer domain (Figure 3.1B, lower right), do not function *in vivo*. From previous studies on the satellite RNA of tobacco ringspot virus (sTRSV) hammerhead ribozyme[16], we suspect that this lack of *in vivo* functionality in earlier designs results from removal of stem loop sequences that may play a critical role in tertiary interactions that stabilize the catalytically active conformation under physiological Mg^{2+} concentrations. To develop ribozyme switches that function *in vivo*, we chose to integrate the hammerhead ribozyme into the target transcript through stem III and couple the sensor domain directly to the ribozyme through stem loops I or II to maintain these potentially essential sequence elements (Figure 3.1B, upper right). We constructed a series of ribozyme controls

(Supplementary Text 3.2 and Supplementary Figure 3.1), which consist of loop coupling and stem integration controls. Implementation and characterization of the gene regulatory activity of these ribozyme constructs within a modular plasmid system in the eukaryotic model organism *Saccharomyces cerevisiae* (Figure 3.1A) indicate that maintenance of loop I and II sequences and thus integration through stem III are essential for *in vivo* functionality of such gene regulatory elements (Supplementary Figure 3.1D).

A



B

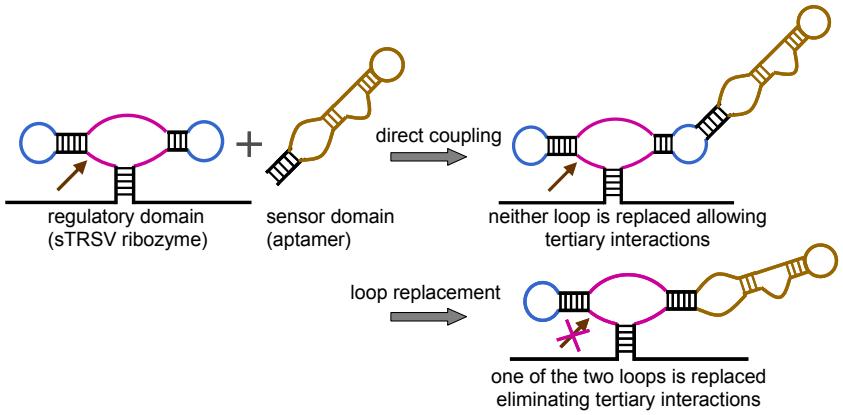


Figure 3.1. General design strategy for engineering ribozyme switches. Color schemes: catalytic core, purple; aptamer sequences, brown; loop sequences, blue; spacer sequences, yellow; brown arrow, cleavage site. (A) General compositional framework and design strategy for engineering universal cis-acting hammerhead ribozyme-based regulatory systems; restriction enzyme sites are underlined. (B) Modular coupling strategies of the sensor and regulatory domains to maintain *in vivo* activity of the individual domains.

3.2.3. Engineering mechanisms for information transmission between the modular switch domains

The final design challenge in building a universal switch platform is to develop a standardized means of transmitting information (encoded within an information transmission domain) from the sensor (aptamer) domain to the regulatory (ribozyme) domain (DP5). There are two different strategies for transmitting information between the aptamer and ribozyme domains: strand displacement and helix slipping. We constructed and characterized ribozyme switch platforms based on both mechanisms.

The first information transmission domain that we developed is based on a strand displacement mechanism, which involves the rational design of two sequences that compete for binding to a general transmission region (the base stem of the aptamer) (Figure 3.2, A and B). We employed this mechanism in engineering a ribozyme switch platform that enables both up- and down-regulation of gene expression in response to increasing effector concentrations ('ON' and 'OFF' switches, respectively). An initial ribozyme switch, L2bulge1, was constructed to up-regulate gene expression through the corresponding base platform (L2Theo, Supplementary Figure 3.1C) by incorporating a competing strand following the 3' end of the theophylline aptamer[23] (Figure 3.2A). This competing strand is perfectly complementary to the base stem of the aptamer at the 5' end. Using the same design principles, we engineered another ribozyme switch, L2bulgeOff1 (Figure 3.2B), for down-regulating gene expression. Our strand displacement strategy is based on the conformational dynamics characteristic of RNA molecules that enables them to distribute between at least two different conformations at equilibrium: one conformation in which the competing strand is not base-paired or base-paired such that the ligand-binding pocket is not formed, and the

other conformation in which the competing strand is base-paired with the aptamer base stem, displacing the switching strand and thus allowing the formation of the ligand-binding pocket. Strand displacement results in the disruption (L2bulge1) or restoration (L2bulgeOff1) of the ribozyme's catalytic core. Binding of theophylline to the latter conformation shifts the equilibrium distribution between these two conformations to favor the aptamer-bound form as a function of increasing theophylline concentration.

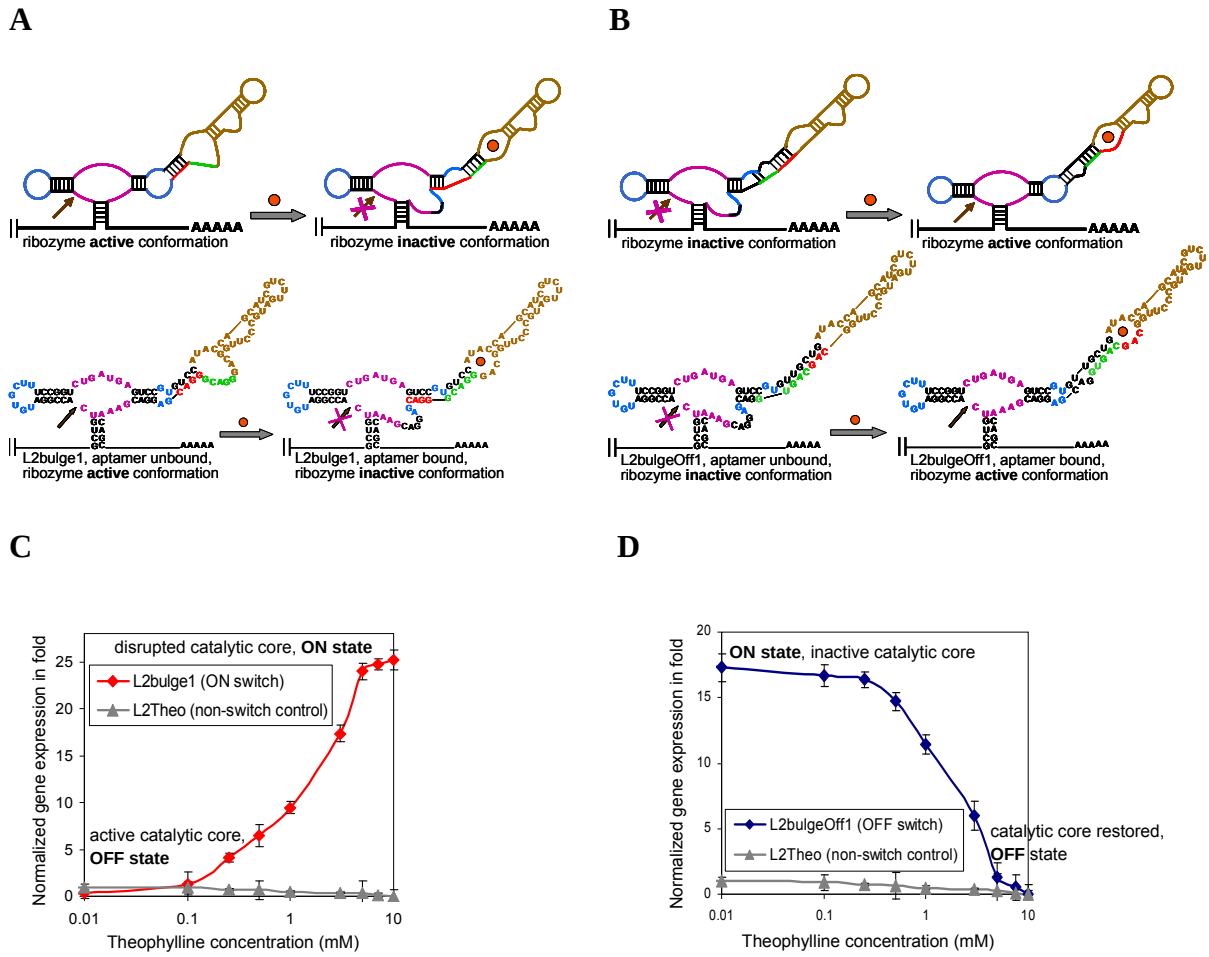


Figure 3.2. Regulatory properties of the strand displacement information transmission mechanism. Color schemes: switching strand, red; competing strand, green; all other schemes correspond to those used in Figure 3.1. (A) Gene expression 'ON' ribozyme switch platform, L2bulge1. (B) Gene expression 'OFF' ribozyme switch platform, L2bulgeOff1. The theophylline-dependent gene regulatory behavior of (C) L2bulge1 ('ON' switch), (D) L2bulgeOff1 ('OFF' switch), and L2Theo (non-switch control). Gene expression levels are

reported in fold as defined in Materials and Methods and normalized to the expression levels in the absence of effector.

Increased binding of theophylline to L2bulge1 resulted in an approximate 25 folds increase in target expression levels at 5 mM theophylline relative to those in the absence of effector (Figure. 3.2C and Supplementary Figure 3.2). In contrast, increased binding of theophylline to L2bulgeOff1 resulted in an approximate 18 folds reduction in expression levels at 5 mM theophylline relative to those in the absence of effector (Figure. 3.2D and Supplementary Figure 3.2). Through our strand displacement mechanism, we have engineered ribozyme switches *de novo*, L2bulge1 and L2bulgeOff1, that provide allosteric regulation of gene expression and function as 'ON' and 'OFF' switches, respectively.

We engineered a second class of ribozyme switch platforms to examine an alternative information transmission domain based on a helix slipping mechanism, which does not allow for rational design (Figure. 3.3A). This mechanism involves the functional screening of 'communication modules'[19-21] within the base stem of the aptamer. Communication modules are dynamic elements capable of transmitting the binding state of an aptamer domain to an adjacent regulatory domain through a 'slip-structure' mechanism[19], in which a nucleotide shift event within the element is translated to a small-scale change in the conformation of the regulatory domain in a ligand-dependent manner. These elements have been developed through *in vitro* screening processes, and their dynamic and communicative properties have been demonstrated *in vitro* in engineered allosteric ribozymes[18-21]. We screened the *in vivo* functionality of previously *in vitro* selected communication modules[19-21] by assaying the activity of these sequences within L1Theo and L2Theo. A critical difference between the design of the previously developed *in vitro* allosteric ribozymes, from which these

communication modules were generated, and that of our engineered ribozyme switches is the coupling strategies between the aptamer and ribozyme domains and their effects on the *in vivo* activity of the ribozyme domain as described previously (Figure 3.1B).

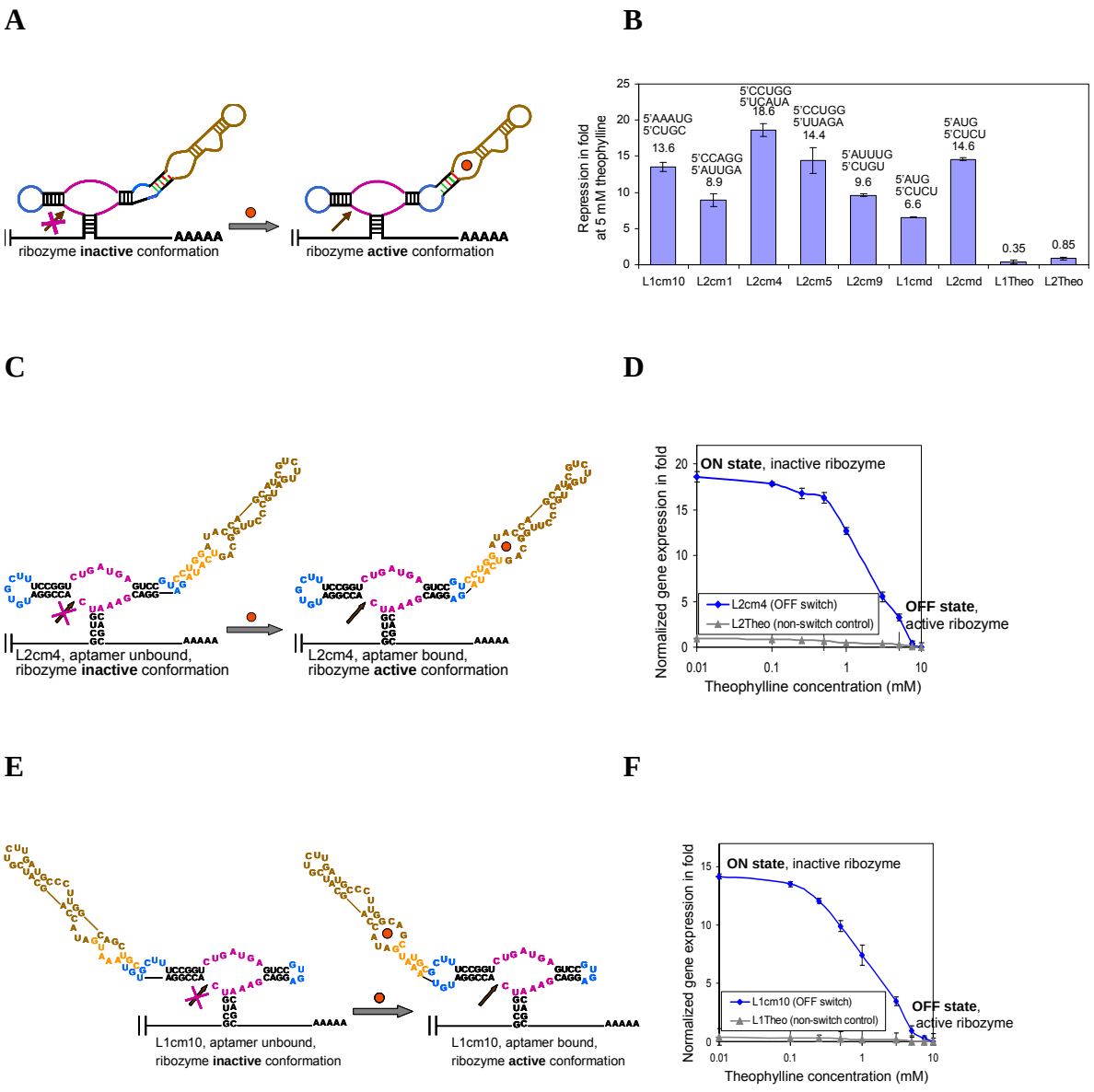


Figure 3.3. Regulatory properties of the helix slipping information transmission mechanism. Color schemes: communication module schematic, red and green; communication module sequence, orange; all other schemes correspond to those used in Figure 3.1. (A) Gene expression 'OFF' ribozyme switch platform based on helix slipping. The base stem of the aptamer is replaced with a communication module. (B) Regulatory activities of helix slipping-based ribozyme switches. Gene regulatory effects of the 'OFF' switches at 5 mM

theophylline are reported in fold repression relative to expression levels in the absence of effector. The corresponding communication module sequences are indicated. Sequence and structure of representative helix slipping ribozyme switches, (C) L2cm4 and (E) L1cm10. The theophylline-dependent gene regulatory behavior of (D) L2cm4 and (F) L1cm10. Gene expression levels are reported as described in Figure 3.2, except that in (F) L1Theo is used as a non-switch control.

Among the thirteen communication modules[19-21] screened for *in vivo* activity, five (cm1, cm4, cm5, cm9, and cmd) exhibit down-regulation of expression levels through loop II, whereas only two (cm10 and cmd) exhibit such regulation through loop I (Figure 3.3B). The regulatory activities of two helix slipping-based ribozyme switches, L2cm4 (Figure 3.3, C and D, and Supplementary Figure 3.3) and L1cm10 (Figure 3.3, E and F, and Supplementary Figure 3.3), were characterized across a range of theophylline concentrations and exhibit substantial regulatory effects. Although the helix slipping constructs are comprised of identical aptamer and catalytic core sequences, they exhibit different extents of regulation. This variability suggests that each construct contains a different equilibrium distribution between the adoptable conformations and that the energy required for structural switching between the conformations is also different.

We validated the regulatory mechanisms of representative strand displacement- and helix slipping-based switches. Relative steady-state transcript levels in the absence and presence of effector are consistent with corresponding fluorescent protein levels (Supplementary Table 3.1), indicating that cleavage in the 3' UTR results in rapid decay and inactivation of the target transcript. In addition, we demonstrated that changes in expression levels are induced shortly after effector addition (Supplementary Figure 3.4), indicating that the response of the regulatory elements to changes in effector levels is relatively rapid.

3.2.4. Rational tuning strategies enable programming of switch regulatory response

The ability to program the regulatory response of a universal switch platform is an important property in tuning the platform performance to comply with the design specifications for a particular application. We demonstrate that our strand displacement-based ribozyme switch platform incorporates an information transmission mechanism that is amenable to rational tuning strategies for programming regulatory response properties. Programming of new regulatory information is achieved by sequence alteration resulting in a change in the molecule's structural stability, which may affect its conformational switching dynamics if the molecule can adopt multiple conformations. These rational sequence modification tuning strategies are not applicable to communication module-based switches due to an inability to predict their activities.

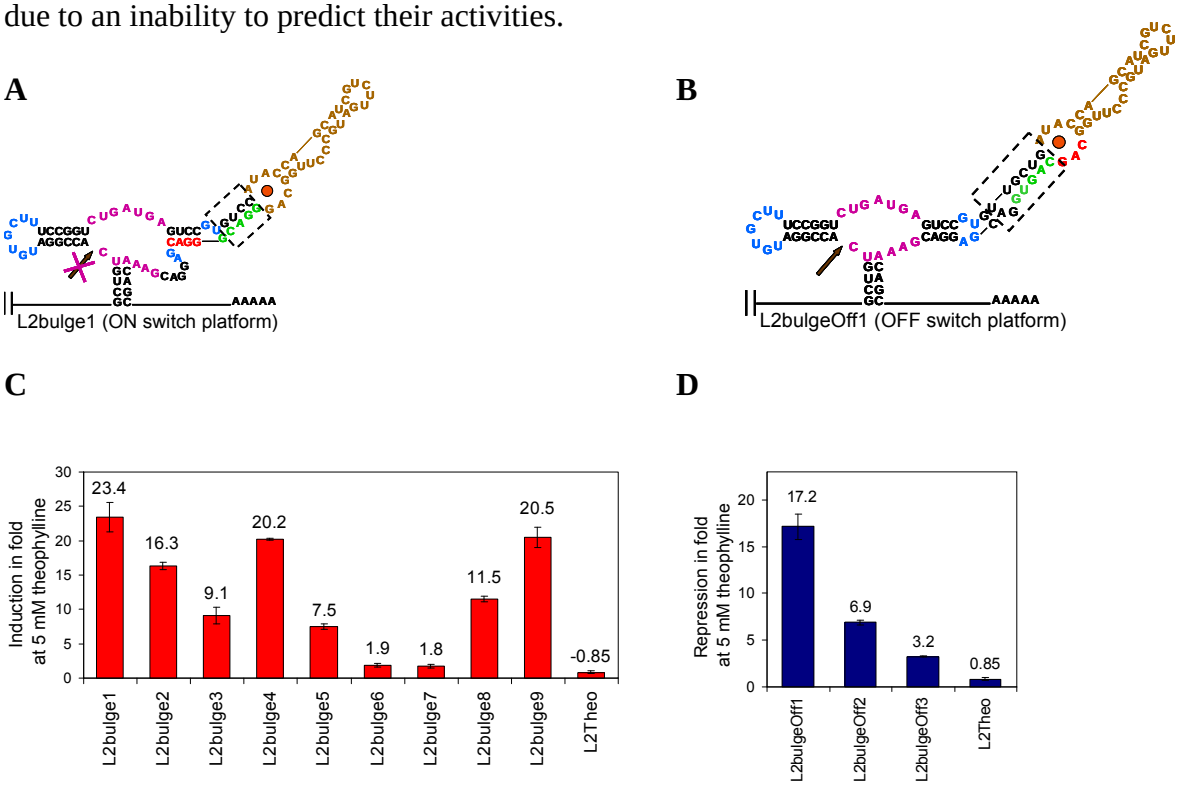


Figure 3.4. Tunability of the strand displacement-based ribozyme switches. Sequences targeted by the rational tuning strategies are indicated in the dashed boxes on the effector-bound conformations of (A) L2bulge1 (ribozyme inactive) and (B) L2bulgeOff1 (ribozyme

active). Regulatory activities of tuned strand displacement-based (C) 'ON' and (D) 'OFF' ribozyme switches. Gene regulatory effects of these switches at 5 mM theophylline are reported in fold induction for 'ON' switches and fold repression for 'OFF' switches relative to the expression levels in the absence of theophylline as described in Figure 3.2.

A more complete description of our tuning strategies is provided in Supplementary Text 3.3, Supplementary Figure 3.5, and Supplementary Table 3.2. Briefly, our rational tuning strategies target alteration of the nucleotide composition of the base stem of the aptamer domain to affect the stabilities of individual switch constructs and the energies required for the construct to switch between two adoptable conformations. Using these strategies, we rationally engineered a series of tuned 'ON' and 'OFF' switches from L2bulge1 and L2bulgeOff1, respectively (Figure 3.4, A and B). These tuned switches exhibit different regulatory ranges in accordance with our rational energetic tuning strategies (Figure 3.4, C and D, and Supplementary Figure 3.6).

3.2.5. The ribozyme switch platform exhibits component modularity and specificity

In implementing a standardized mechanism through which to transmit information between the domains of a switch platform (DP5), we needed to confirm that the modular coupling between the aptamer and ribozyme domains is maintained (DP4). We performed modularity studies on our strand displacement-based ribozyme switch platform, in which aptamers possessing sequence flexibility in their base stems can be swapped into the sensor domain. To begin to demonstrate that ribozyme switch activity may be controlled by different effector molecules we replaced the theophylline aptamer of L2bulge1 with a tetracycline mini-aptamer[24] to construct a tetracycline-repsonsive ON switch (L2bulge1tc) (Figure 3.5A). Despite similar aptamer ligand affinities[23, 24], the extent of up-regulation with

L2bulge1tc was greater than that with L2bulge1 at the same extracellular concentration of their respective ligands (Figure 3.5B). This is likely due to the high cell permeability of tetracycline[25] compared to theophylline[26]. These results demonstrate that our strand displacement-based ribozyme switch platform maintains modularity between the aptamer and ribozyme domains. We also performed similar modularity studies on the helix slipping-based switch platform by replacing the theophylline aptamer of L1cm10, L2cm4 and L2cm5 with the tetracycline mini-aptamer (L1cm10, L2cm4tc, and L2cm5tc, respectively). These constructs do not exhibit effector-mediated gene-regulatory effects (data not shown).

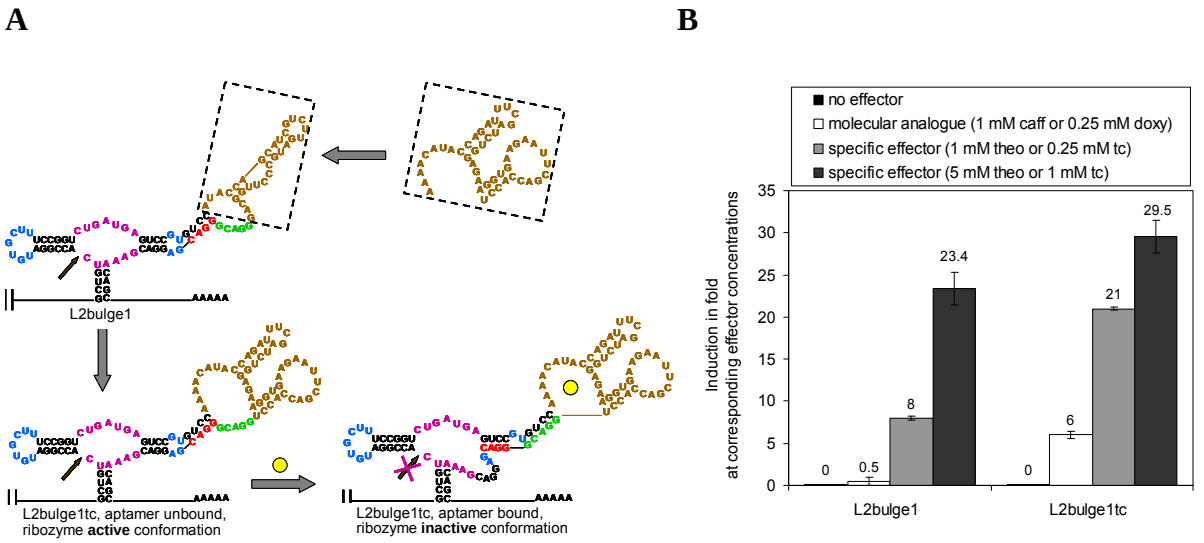


Figure 3.5. Modularity and specificity of the strand displacement-based ribozyme switches. (A) Modular design strategies for the construction of new ribozyme switches. The theophylline (left dashed box) and tetracycline (right dashed box) aptamers are shown. (B) Regulatory activities of the modular ribozyme switch pair, L2bulge1 and L2bulge1tc, in response to their respective ligands, theophylline (theo) and tetracycline (tc), and closely-related analogues, caffeine (caff) and doxycycline (doxy). Regulatory effects are reported in fold induction relative to the expression levels in the absence of effector as described in Figure 3.2.

We also demonstrated that the aptamer sequences (theophylline and tetracycline) incorporated into our ribozyme switch platforms maintain highly specific target recognition capabilities *in vivo* similar to their *in vitro* specificities generated during the selection process

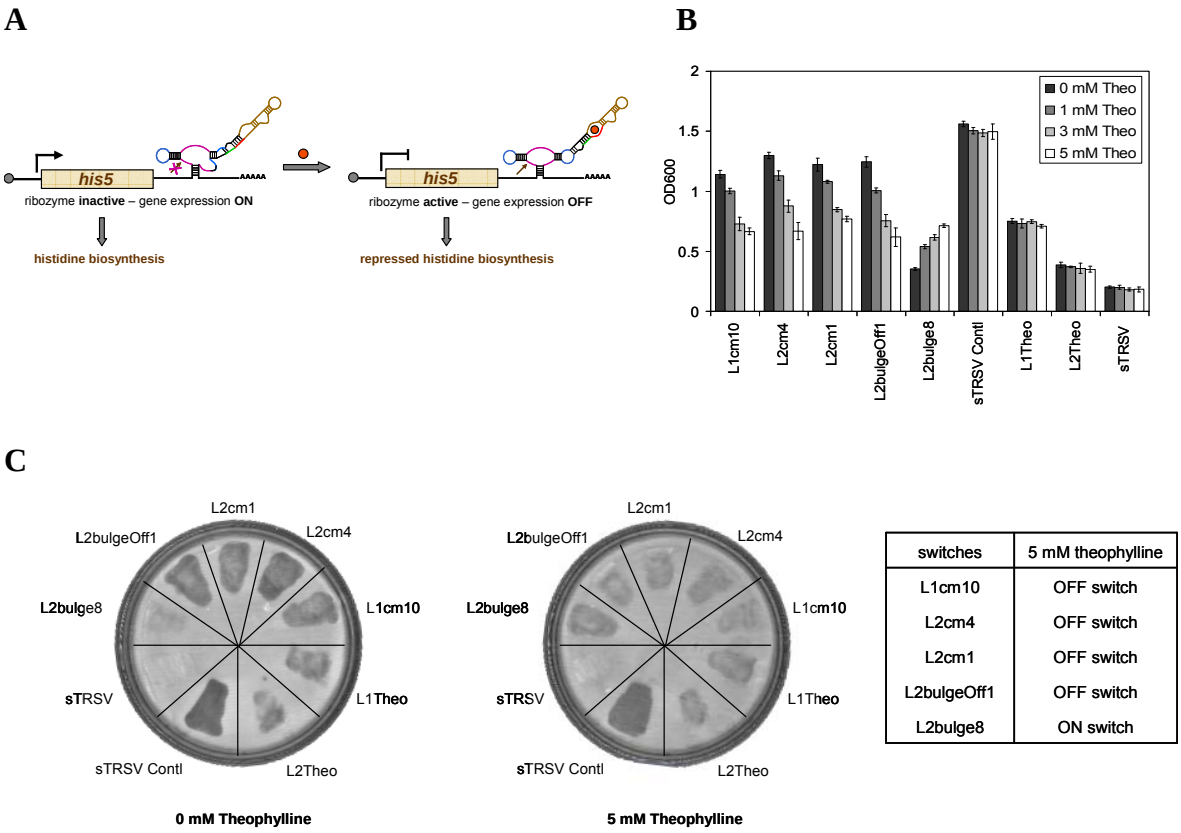
against corresponding molecular analogues (caffeine and doxycycline, respectively)[23, 24] (Figure 3.5B). This is an important property in implementing these platforms in cellular engineering applications that involve complex environments where molecular species similar to the target ligand may be present.

3.2.6. Component modularity enables implementation of ribozyme switches as regulatory systems in diverse applications

To demonstrate the scalability and utility of these switch platforms as application-specific control systems, we demonstrate the implementation of ribozyme switches in two distinct cellular engineering application areas. First, utility (DP3) and the ability to respond to and control cellular information is demonstrated by the application of ribozyme switches to small molecule-mediated regulation of cell growth. Second, scalability (DP1) and the ability to respond to and report on cellular information is demonstrated by the implementation of ribozyme switches as non-invasive *in vivo* sensors of metabolite production.

The first system explores the application of our ribozyme switches to the regulation of a survival gene, where modification of expression levels is expected to produce an observable and titratable phenotypic effect on cell growth. The reporter gene within the original constructs was replaced with a growth-associated gene (*his5*) responsible for the biosynthesis of histidine in yeast[27] (Figure 3.6A). We performed growth regulation assays across various effector concentrations using representative switch constructs and demonstrated that these switches mediate cell growth in a highly effector-dependent manner (Figure 3.6B). Plate-based assays confirm the theophylline-dependent ribozyme switch-based

regulation of cell growth (Figure 3.6C). This application demonstrates the utility (DP3) of our switch platform, in which the control system exhibits modularity to the functional level components in the regulatory system.



switches	5 mM theophylline
L1cm10	OFF switch
L2cm4	OFF switch
L2cm1	OFF switch
L2bulgeOff1	OFF switch
L2bulge8	ON switch

Figure 3.6. System modularity of ribozyme switches enables implementation in programmed cell growth. (A) System design for ribozyme switch-based regulation of cell growth. Small molecule-mediated regulation of a gene required for cell growth is illustrated for a strand displacement-based 'OFF' switch. (B) Theophylline-mediated ribozyme switch-based regulation of cell growth. Changes in growth are reported as OD_{600} values for cells grown in 5 mM 3AT in media lacking histidine. (C) Demonstration of theophylline-regulated cell growth by ribozyme switches through plate-based assays. Cells harboring ribozyme switches and control constructs were streaked on two plates containing the same medium except different effector concentrations (0 mM versus 5 mM theophylline). OFF switches (L1cm10, L2cm4, L2cm1, L2bulgeOff1) exhibit suppressed cell growth on the plate containing 5 mM theophylline while an ON switch (L2bulge8) exhibits a higher growth level on the plate containing 5 mM theophylline. The control constructs (L1Theo, L2Theo, sTRSV Contl, and sTRSV) exhibit similar growth levels on both plates. sTRSV exhibits no cell growth due to its efficient cleavage activity and sTRSV Contl exhibits the highest levels of growth due to its lack of cleavage activity.

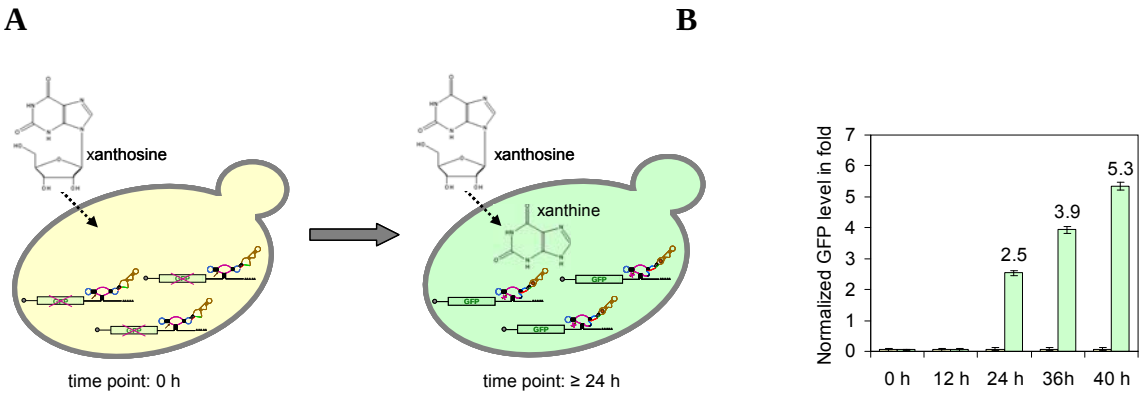


Figure 3.7. System modularity of ribozyme switches enables implementation in non-invasive detection of metabolite biosynthesis. (A) System design for ribozyme switch-based *in vivo* sensing of metabolite production. Xanthine is converted from fed xanthosine through an activity endogenous to yeast and product accumulation over time is detected through a strand displacement-based xanthine-responsive 'ON' switch coupled to the regulation of a reporter protein. (B) Ribozyme switch-based xanthine synthesis detection through L2bulge9. Metabolite sensing through L2bulge9 is reported in fold induction of GFP levels relative to the expression levels in the absence of xanthosine feeding as described in Figure 3.2. Expression data for experiments performed with L2bulge1 exhibit similar induction profiles and levels (data not shown).

The second system explores the application of these ribozyme switches to the *in vivo* sensing of metabolite production to demonstrate that these switches provide a non-invasive mechanism through which to transmit molecular information from cells. Nucleoside phosphorylase activities resulting in *N*-riboside cleavage of purine nucleosides have been identified in various organisms[28]. We observe that feeding xanthosine to our yeast cultures results in the production of xanthine, a product synthesized through riboside cleavage of xanthosine. Relatively high xanthine accumulation was detected in cell extracts between 24-48 h after substrate feeding by HPLC analysis (Supplementary Figure 3.7). Xanthosine accumulation was detected in cell extracts at earlier times, indicating that specific levels of intracellular xanthosine accumulation may be required for efficient conversion to xanthine, possibly due to a high K_m value for this enzyme. The theophylline aptamer employed in our switch platforms possesses a reduced binding affinity for xanthine (27-fold lower than

theophylline)[23]. We employed two 'ON' switch constructs (L2bulge1 and L2bulge9) for the *in vivo* detection of xanthine production in cultures fed the precursor xanthosine (Figure 3.7A). GFP levels in cells fed xanthosine rose steadily between 24-40 h post-feeding in correlation with HPLC data (Figure 3.7B), illustrating the non-invasive metabolite-sensing capabilities of these switches through transmitting changes in metabolite accumulation to changes in reporter expression levels. This application demonstrates the scalability (DP1) of our switch platform, in which the unique properties of the sensing platform employed in this control system enable broad implementation in diverse applications not generally accessible by other regulatory systems.

3.3. Discussion

A key component in the development of an RNA-based framework for engineering ligand-controlled gene regulatory systems is captured within DP5: a mechanism through which to reliably transmit information between distinct domains of the molecule. The strand-displacement and helix-slipping mechanisms demonstrate different strengths and weaknesses as standardized means of transmitting information from the aptamer domain to the ribozyme domain. Only 7 out of the 26 tested communication modules exhibited regulatory activity in our system. In addition, all of the functional communication module sequences demonstrate 'OFF' activity in our *in vivo* system, whereas one of these sequences (cmd) exhibited 'ON' activity in an *in vitro* system[19]. These results indicate that *in vitro* functionality of these elements is selectively translated to *in vivo* activity due to their sensitivity to surrounding sequences. Furthermore, modularity studies performed on this platform indicate that the helix slipping mechanism is not amenable to modular domain swapping strategies. In contrast, we

have demonstrated that strand displacement exhibits greater reliability as an information transmission mechanism in our platform and is characterized by engineering properties such as modular assembly, rational *de novo* design, flexible induction and repression profiles, and response programmability. Although not preferred for the rational design strategies presented here, our helix slipping platform can be employed for the effective generation of new ribozyme switches by *in vivo* screening for helix slipping elements that function with new aptamer sequences, different regulatory ranges, and flexible regulatory profiles. In addition, screening strategies may represent a powerful alternative when rational design strategies fail. For example, we applied our rational design strategies to the construction of strand displacement-based ribozyme switches that modulate cleavage through stem I (L1bulge1-6 in Supplementary Table 3.3). Although these design strategies were successfully applied to stem II, they did not result in functional switches when applied to stem I. These results indicate that screening strategies may be more effective in generating ribozyme switches that modulate activity through stem I.

We have developed and demonstrated universal RNA-based regulatory platforms called ribozyme switches using engineering design principles. This work describes a framework for the reliable *de novo* construction of modular, portable, and scalable control systems that can be used to achieve flexible regulatory properties, such as up- and down-regulation of target expression levels and tuning of regulatory response to fit application-specific performance requirements, thereby expanding the utility of our platforms to a broader range of applications. For example, these switch platforms may be applied to the construction of transgenic regulatory control systems that are responsive to cell-permeable, exogenous molecules of interest for a given cellular network. In regulating sets of functional

proteins, these switches can act to rewire information flow through cellular networks and reprogram cellular behavior in response to changes in the cellular environment. In regulating reporter proteins, ribozyme switches can serve as synthetic cellular sensors for diverse input molecules to monitor temporal and spatial fluctuations in the levels of their target molecules. The switch platforms described here represent powerful tools for constructing ligand-controlled gene regulatory systems tailored to respond to specific effector molecules and enable regulation of target genes in various living systems, and due to their general applicability our platforms offer broad utility for applications in synthetic biology, biotechnology, and health and medicine.

3.4. Materials and Methods

3.4.1. Plasmid, switch construction, and cell strains

Using standard molecular biology techniques[29], a modular characterization plasmid, pRzS, harboring the yeast-enhanced green fluorescence protein (yEGFP)[30] under control of a GAL1-10 promoter, was constructed and employed as a universal vector for the characterization of all ribozyme switches. For the ribozyme switch-mediated growth studies, the *yegfp* gene was replaced with the *his5* gene[27]. The engineered ribozyme constructs were generated by PCR amplification using the appropriate oligonucleotide templates and primers. All oligonucleotides were synthesized by Integrated DNA Technologies. All engineered ribozyme constructs were cloned into two unique restriction sites, *Avr*II and *Xho*I, 3 nucleotides downstream of the yEGFP stop codon and upstream of an ADH1 terminator.

Cloned plasmids were transformed into an electrocompetent *Escherichia coli* strain, DH10B (Invitrogen) and all ribozyme constructs were confirmed by subsequent sequencing

(Laragen, Inc). Confirmed plasmid constructs were transformed into a *Saccharomyces cerevisiae* strain (W303 *MATα his3-11,15 trp1-1 leu2-3 ura3-1 ade2-1*) using standard lithium acetate procedures[31].

3.4.2. RNA secondary structure prediction and free energy calculation

RNAstructure 4.2 (http://rna.urmc.rochester.edu/rnastructure.html) was used to predict the secondary structures of all switch constructs and their thermodynamic properties. RNA sequences that are predicted to adopt at least two stable equilibrium conformations (ribozyme inactive and active) were constructed and examined for functional activity.

3.4.3. Ribozyme characterization assays

S. cerevisiae cells harboring the appropriate plasmids were grown in synthetic complete medium supplemented with an appropriate dropout solution and sugar (2% raffinose, 1% sucrose) overnight at 30°C. Overnight cultures were back diluted into fresh medium to an optical density at 600 nm (OD_{600}) of approximately 0.1 and grown at 30°C. An appropriate volume of concentrated effector stock (to the appropriate final concentration of theophylline or tetracycline) dissolved in medium or an equivalent volume of the medium (no effector control) was added to the cultures at the time of back dilution. In addition, at this time an appropriate volume of galactose (2% final concentration) or an equivalent volume of water were added to the cultures for the induced and non-induced controls, respectively. For specificity assays, an appropriate volume of a concentrated caffeine or doxycycline stock (final concentrations of 1 mM and 250 μM, respectively) was added to a separate culture. Cells were grown to an OD_{600} of 0.8-1.0 or for a period of approximately 6 h before

measuring GFP levels on a Safire fluorescent plate reader (Tecan) and/or on a Cell Lab Quanta SC flow cytometer (Beckman Coulter). For temporal response assays, cultures were grown as described above in the absence of the appropriate effector and fluorescence data were taken every 30 min. After 4 h growth, appropriate volumes of the concentrated effector stock or plain medium were added to the cultures and fluorescence was monitored for several hours thereafter.

3.4.4. Cell growth regulation assays

For liquid culture assays, *S. cerevisiae* cells carrying appropriate plasmids were back diluted and grown according to procedures described above with minor modifications. A competitive inhibitor of the *his5* gene product, 3-amino-triazole (3AT), was added to a final concentration of 5 mM to increase the threshold level of histidine required for cell growth. Cultures were grown in various theophylline concentrations and the growth of each sample was monitored over a 24 h period. The theophylline-regulated growth at 24 h is reported in terms of OD_{600} readings measured on the Tecan. For plate-based assays, 10 μL of the back diluted culture samples was streaked on plates containing 0 and 5 mM theophylline. A higher concentration of 3AT (25 mM) was used in the plate-based assays to optimize visual assessment of theophylline-regulated cell growth.

3.4.5. Metabolite sensing assays

S. cerevisiae cells carrying appropriate plasmids were back diluted and grown according to procedures described above with minor modifications. Cultures were grown in the absence and presence of xanthosine (250 μM final concentration). To account for inducer

depletion, galactose was added to the cultures at 8 h time intervals to a 2% final concentration. Fluorescence levels of the samples were monitored over a 48 h period according to procedures described above. For HPLC analysis, cell extracts were prepared after appropriate growth periods following xanthosine feeding by rapid freezing of cell cultures in liquid nitrogen in the form of beads. Frozen cell beads were subsequently lysated by grinding using a mortar and pestle followed by extraction with methanol. Intracellular metabolite levels were analyzed using an HPLC system integrated with a mass spectrometer (HPLC-MS) (Agilent Technologies), which enables confirmation of metabolite peaks based on their corresponding molecular weights.

3.4.6. Fluorescence quantification

The population-averaged fluorescence of each sample was measured on a Safire fluorescence plate reader with the following settings: excitation wavelength of 485 nm, an emission wavelength of 515 nm, and a gain of 100. Fluorescence readings were normalized to cell number by dividing fluorescence units by the OD_{600} of the cell sample and subtracting the background fluorescence level to eliminate autofluorescence.

Fluorescence distributions within the cell populations were measured on a Quanta flow cytometer with the following settings: 488 nm laser line, 525 nm bandpass filter, and PMT setting of 5.83. Fluorescence data was collected under low flow rates for approximately 30,000 cells. Viable cells were selected and fluorescence levels were determined from 10,000 counts in this selected population. A non-induced cell population was used to set a ‘negative GFP’ gate. Cells exhibiting fluorescence above this negative gate are defined as the ‘positive GFP’ cell population.

Similar to previous reports[17, 32], we report gene expression levels as 'fold', where 1 fold is defined as the reporter gene expression level of sTRSV relative to the background fluoresence level. Ligand-directed regulatory effects are reported as fold gene expression levels normalized to the levels in the absence of effector. All fluorescence data and mean ±s.d. are reported from at least three independent experiments.

3.4.7. Quantification of cellular transcript levels

Briefly, total RNA was extracted employing standard acid phenol extraction methods[33] followed by cDNA synthesis and PCR amplification. cDNA was synthesized using gene-specific primers (Supplementary Table 3.3) and Superscript III Reverse Transcriptase (Invitrogen) following manufacturer's instructions. Relative transcript levels were quantified from the cDNA samples by employing an appropriate primer set and the iQ SYBR Green Supermix (BioRAD) according to manufacturer's instructions on an iCycler iQ qRT-PCR machine (BioRAD). The resulting data were analyzed with the iCycler iQ software according to manufacturer's instructions. Transcript levels of switch constructs were normalized to that of the endogenous *actI* gene[34] using *actI*-specific primers.

3.5. Supplementary Information

Supplementary Text 3.1: *Glossary of terms*

actuator domain	A switch domain that encodes the system control function. As used here, the actuator domain encodes the gene regulatory function and is comprised of a hammerhead ribozyme sequence.

communication module	A sequence element that typically forms an imperfectly paired double-stranded stem that can adopt different base pairs between nucleotides through a 'slip-structure' mechanism. As used here, a communication module is a type of information transmission domain that transmits the binding state of the aptamer domain to the adjacent actuator domain through a helix slipping mechanism. As demonstrated in this work, a communication module does not act in a modular fashion with other switch domains. The term is retained here from earlier work in the field of nucleic acid engineering.
competing strand	The nucleic acid sequence within a strand displacement domain that is bound to the general transmission region of the switch when the sensor domain is in the restored conformation (i.e., in the presence of ligand). The competing strand competes for binding with the switching strand, which is initially bound to this transmission region in the absence of ligand.
component	A part of a system that encodes a distinct activity or function.
composability	A property of a system that indicates its ability to be comprised of components that can be selected and

	assembled in a modular fashion to achieve a desired system performance. As used here, composability refers to the ability of the individual domains of the control system to be modularly linked without disrupting their activities.
engineering design principle	A required property of a constructed system that enables use by others.
framework	A basic conceptual structure that is used to solve a complex product design issue. As used here, the framework is used to reliably design and construct specific instances of RNA switches. The conceptual structure of our framework is comprised of specified engineering design principles and design strategies that enable extensible and reusable system design.
helix slipping domain	A subset of information transmission domains that act through a helix slipping mechanism. The helix slipping domain is also referred to as the communication module.
helix slipping mechanism	An information transmission mechanism that is based on an information transmission domain that functions through a helix slipping event and does not allow for rational design. Such a helix slipping event utilizes a communication module (or helix slipping domain) within

	the general transmission region of the switch (the base stem of the aptamer) to result in disruption or restoration of the actuator domain in response to restoration of the sensor domain.
information transmission domain	A switch domain that encodes the function of transmitting information between the sensor domain and the actuator domain.
information transmission mechanism	A general mechanism for transmitting information between the sensor domain and the actuator domain of a switch. As used here, this mechanism regulates the activity of the actuator domain in response to the binding state of the sensor domain.
modular	A property of a system comprised of modules, which indicates that the modules can be interchanged as parts without changing the interface between modules or the modules themselves.
module	A self-contained system component that has a well-defined interface with other system components.
platform	A general framework on which specific applications can be implemented. As used here, the platform enables specific instances of switches to be built in a standardized

	manner.
portability	A property of a system that indicates its ability to be implemented in environments different from that which it was originally designed. As used here, portability refers to the ability of the control system to be implemented in different organisms.
reliability	A property of a system that indicates its ability to perform and maintain its functions under a set of specified conditions. As used here, reliability refers to the ability of the information transmission domain to standardize the transmission of information between the sensor and actuator domains.
scalability	A property of a system that indicates its ability to handle increasing work. As used here, scalability refers to the ability of the control system to be implemented across broad application space by being able to forward design its response to different molecular information.
switch	A molecule that can adopt at least two different conformational states, where each state is associated with a different activity of the molecule. Often a ligand can bind to one or more conformations of the switch, such that

	the presence of the ligand shifts the equilibrium distribution across the adoptable conformations and therefore regulates the activity of the switch molecule. As used here, switch refers to an RNA molecule that can adopt different structures that correspond to different gene-regulatory activities. An RNA switch is then a ligand-controlled gene regulatory system.
switch domain	A component of a switch that encodes a distinct activity or function.
switching strand	The nucleic acid sequence within a strand displacement domain that is bound to the general transmission region of the switch when the sensor domain is in the disrupted conformation (i.e., in the absence of ligand). The switching strand is displaced by the competing strand in the presence of ligand.
sensor domain	A switch domain that encodes a ligand-binding function. As used here, the sensor domain is comprised of an RNA aptamer sequence.
strand displacement domain	A subset of information transmission domains that act through a strand displacement mechanism.

strand displacement mechanism	An information transmission mechanism that is based on the rational design of an information transmission domain that functions through a strand displacement event. Such a strand displacement event utilizes competitive binding of two nucleic acid sequences (the competing strand and the switching strand) to a general transmission region of the switch (the base stem of the aptamer) to result in disruption or restoration of the actuator domain in response to restoration of the sensor domain.
universal	A system property that indicates its ability to maintain function across different applications, environments, and component interfaces. As used here, a universal system is composed of the five engineering design principles (scalability, portability, utility, composability, and reliability) and results in the specified extensible platform for RNA switch construction.
utility	A property of a system that indicates its ability to be of practical use. As used here, utility refers to the ability of the control system to interface with different functional level components to enable forward design of the function that is being controlled by the system.

Supplementary Text 3.2: *Ribozyme control constructs for loop sequence coupling and stem integration controls*

To establish and make useful our design strategy we constructed and characterized a series of ribozyme controls. We characterized the regulatory activity of our ribozyme constructs within a modular ribozyme characterization system in the eukaryotic model organism *Saccharomyces cerevisiae* (Figure 3.1A). First, an inactive ribozyme control (sTRSV Contl, Supplementary Figure 3.1A) was constructed to adopt the same structural motif as sTRSV (Figure 3.1A), while carrying a scrambled catalytic core sequence. Second, a synthetic sTRSV ribozyme (hhRz I) that contains closed loops in stems II and III and is embedded through stem I was constructed as a stem integration control (Supplementary Figure 3.1A). Finally, we constructed four loop sequence controls. In one set, stem loops I and II (L1R and L2R, respectively) were replaced by the theophylline aptamer TCT8-4[23] (Supplementary Figure 3.1B), and in another set, the theophylline aptamer was coupled directly to sequences in loops I and II (L1Theo and L2Theo, respectively) (Supplementary Figure 3.1C). sTRSV exhibits a 50-fold reduction in target expression levels relative to sTRSV Contl (Supplementary Figure 3.1D). HhRz I, L1R, and L2R exhibit similar target expression levels to that of sTRSV Contl, suggesting that ribozyme activity was abolished in these constructs. In contrast, L1Theo and L2Theo exhibit significantly lower target expression levels relative to sTRSV Contl. L1Theo and L2Theo were employed as the primary base constructs in engineering our synthetic ribozyme switch platforms. In addition, scrambled core versions of L1Theo and L2Theo exhibit no theophylline-dependent shifts in gene expression (data not shown), indicating that theophylline binding in that region of the transcript alone is not responsible for the observed regulatory effects. Taken together, we

find that our design strategy enables the construction of a universal ribozyme switch platform that satisfies the design principles of portability, utility, and composability.

Supplementary Text 3.3: *Rational tuning strategies for strand displacement-based switches*

A series of nine tuned 'ON' switches were constructed from L2bulge1 as a base structure by employing rational energetic tuning strategies developed in this work. This strategy is based on the effects of altering the predicted free energies of a particular conformation ($-\Delta G$) and the predicted difference between the free energies of two conformations ($\Delta\Delta G$) on RNA conformational dynamics, or the ability of the RNA molecule to distribute between these two conformational states. Supplementary Table 3.2 lists free energies ($-\Delta G$) of ribozyme active and inactive conformations and the energy difference ($\Delta\Delta G$) between the free energies of these two conformations. Specifically, lowering values for either of these energetic measurements ($-\Delta G$ or $\Delta\Delta G$) is expected to make it easier for a particular RNA molecule to switch between the conformational states in question. Therefore, there is an anticipated optimum conformational energy and energetic difference between conformations to achieve the desired range of switching in response to effector concentration (i.e., energy measurements too high will result in stable non-switch designs, and energy measurements or energy difference measurements too low will result in fairly equal distributions between the two conformational states and lower switching capabilities). It is also expected, then, that one can "push" switches into a non-switch state by moving away from this energetic optimum. This strategy was examined in a series of tuning experiments described below.

L2bulge2 and L2bulge3 (Supplementary Figure 3.5) replace canonical base pairs in the aptamer base stem of the ribozyme inactive conformation of L2bulge1 with U-G wobble pairs. As a result of these destabilizing alterations, both equilibrium conformations (ribozyme active and ribozyme inactive) become less thermodynamically stable than those of L2bulge1, as estimated from their predicted free energies (-ΔG). In addition, the energy required to switch between the two equilibrium conformations was maintained similar to that of L2bulge1, as estimated by the difference between the free energies of the two conformations (ΔΔG). Ribozyme assays indicate that both L2bulge2 and L2bulge3 exhibit smaller dynamic ranges than that of L2bulge1 (Figure 3.4C and Supplementary Figure 3.8). It is proposed that the lower stabilities of the conformational states enable more frequent dynamic switching between the two equilibrium conformations and therefore lower the difference in distribution favoring one state over the other.

L2bulge4 (Supplementary Figure 3.5) incorporates an additional G-U wobble pair within the aptamer base stem of the ribozyme inactive conformation of L2bulge1. However, this aptamer stem extension does not result in an appreciable predicted change in the thermodynamic stabilities of the equilibrium conformations or the energy required to switch between the two equilibrium conformations when compared to L2bulge1. Ribozyme assays indicate that L2bulge4 exhibits a dynamic range in response to theophylline levels similar to that of L2bulge1 (Figure 3.4C and Supplementary Figure 3.8).

L2bulge5 (Supplementary Figure 3.5) incorporates an additional canonical base pair (A-U) within the aptamer base stem of L2bulge1. As a result of this stabilizing alteration, the conformation of the ribozyme switch, in which the aptamer structure is formed and the catalytic core is disrupted (ribozyme inactive), is increased in stability and as stable as the

conformation in which the catalytic core is not disrupted (ribozyme active). The increased stability of the ribozyme inactive conformation in L2bulge5 in comparison to L2bulge1 and L2bulge4 indicates that the equilibrium distribution between these two conformations will shift to favor the ribozyme inactive conformation. Ribozyme assays indicate that L2bulge5 exhibits significantly higher GFP expression levels in the absence and presence of theophylline compared to those of L2bulge1 and L2bulge4, such that the theophylline-regulated increase in gene expression is similar to that of L2bulge3 but different in regulatory dynamic ranges (Figure 3.4C and Supplementary Figure 3.8).

Two switches in this series, L2bulge6 and L2bulge7, were constructed to demonstrate the ability of this tuning strategy to "push" the ribozyme switch constructs out of a switchable energetic range and approach non-switching extremes. L2bulge6 (Supplementary Figure 3.5) was designed to energetically favor the conformation, in which the aptamer structure is formed and the catalytic core is disrupted, (ribozyme inactive) in the absence of theophylline by introducing a stabilizing G-C base pair into the aptamer stem of this conformation. Since the aptamer conformation is expected to be favored in L2bulge6, the presence of theophylline is expected to have little or no effect on the conformational dynamics of this switch. L2bulge7 (Supplementary Figure 3.5) was designed to energetically favor the conformation, in which the aptamer structure is not formed and the catalytic core is undisrupted (ribozyme active), by introducing a stabilizing U-A base pair into the stem extending from loop II in this conformation. As the stability of the ribozyme active conformation is significantly higher than that of the ribozyme inactive conformation for L2bulge7, the presence of theophylline is expected to have little effect on the conformational dynamics of this ribozyme switch. Ribozyme assays indicate that L2bulge7 exhibits very low

GFP expression levels and L2bulge6 exhibits very high GFP expression levels in the presence and absence of theophylline (Supplementary Figure 3.8). As rationally designed, both constructs exhibit little increase in target expression levels in response to theophylline by energetically favoring one of the two conformational states (Figure 3.4C).

L2bulge 8 (Supplementary Figure 3.5) was modified from L2bulge7 by replacing the canonical base pair (U-A) with a wobble base pair (U-G), thereby reducing the stability of the ribozyme active conformation of L2bulge7 and allowing it to adopt the ribozyme inactive conformation. Similarly, L2bulge 9 (Supplementary Figure 3.5) was modified in such a way to reduce the energy difference between the two conformations of L2bulge7. Ribozyme assays indicate that L2bulge8 and L2bulge9 exhibit theophylline-dependent up-regulation of target gene expression in accordance with the reduced stabilities of the ribozyme active conformations and energy differences between the two adoptable conformations for each of these switch constructs (Figure 3.4C and Supplementary Figure 3.8).

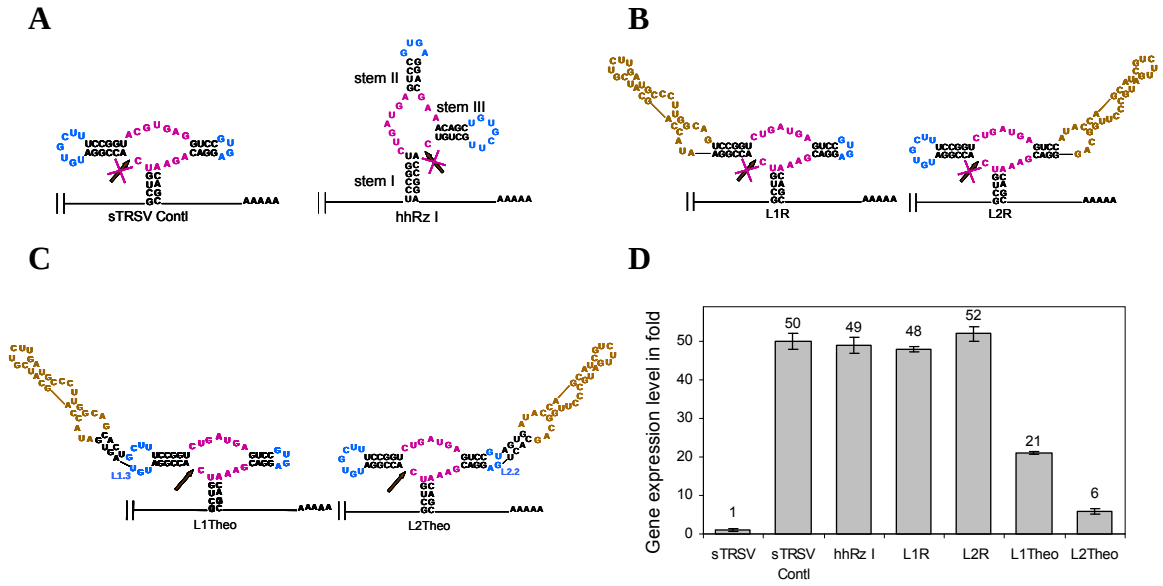
In addition, a series of three tuned ‘OFF’ switches were constructed by using rational energetic tuning strategies from L2bulgeOff1 as a base structure. L2bulgeOff2 and L2bulgeOff3 were constructed to demonstrate tunability of the ‘OFF’ switch platform using similar energetic design strategies (Supplementary Figure 3.5). These switch variants exhibit different theophylline-responsive dynamic ranges from that of L2bulgeOff1 (Figure 3.4D and Supplementary Figure 3.8).

Flow cytometry analysis of the tuned ribozyme switch series demonstrate that the tuned switches exhibit corresponding shifts in the mean fluorescence of the cell populations in the presence and absence of theophylline (Supplementary Figure 3.6). The relative dynamic ranges of the switches across the full regulatory range bracketed by the ribozyme

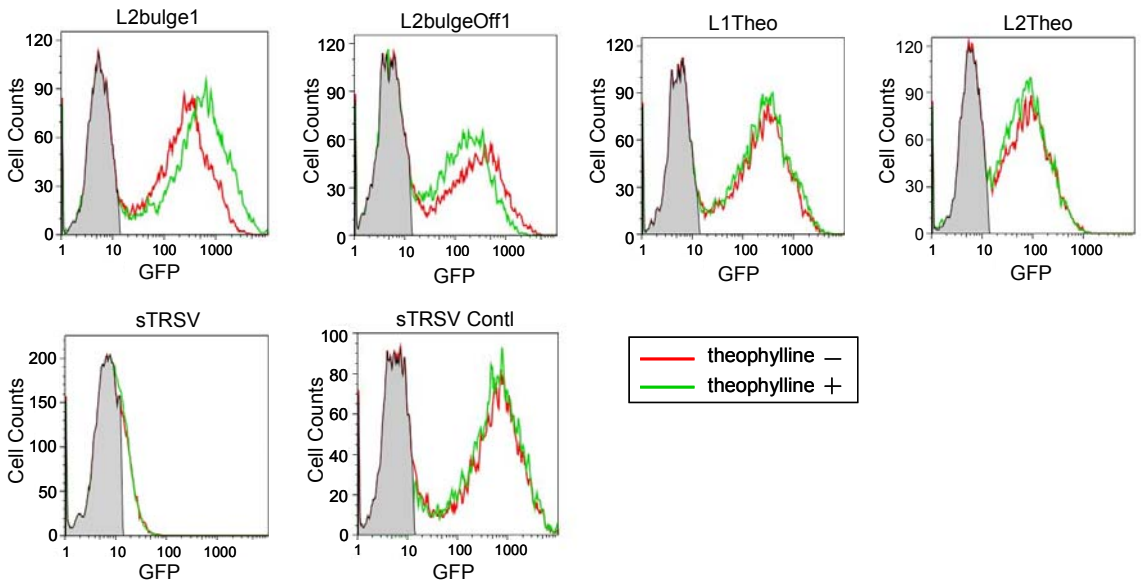
active and inactive controls, sTRSV and sTRSV Contl respectively, are presented in Supplementary Figure 3.8.

Among the twelve tuned switches (both 'ON' and 'OFF'), the dynamic regulatory ranges of most of these switches are in agreement with our rational tuning strategies based on the -ΔG and ΔΔG values predicted by RNAstructure 4.2. Two exceptions are noted: L2bulge9 and L2bulgeOff3. L2bulge9 exhibits a larger dynamic regulatory effect despite its higher ΔΔG than L2bulge8. L2bulgeOff3 exhibits a smaller dynamic regulatory effect despite its smaller ΔΔG than L2bulgeOff2. However, it is more difficult to make a direct comparison between L2bulgeOff2 and L2bulgeOff3, as both conformations of L2bulgeOff3 are significantly more stable than those of L2bulgeOff2, likely resulting in L2BulgeOff3 less frequently switching between its two conformations and thus enabling this molecule to get 'trapped' in its lower free energy states. In addition, outliers may also arise because the RNAstructure program predicts these energy values based on the secondary structure of a particular conformation and does not take into consideration energy contributions from tertiary interactions (that have been observed in prior work[16]) in estimating these energies. Nevertheless, we demonstrate that energetic predictions based solely on secondary structure are useful for our rational tuning design strategies. The different dynamic regulatory ranges exhibited by our tuned switches in response to their specific effector (Supplementary Figure 3.8) validate that such response programming can be achieved by altering the nucleotide composition of the information transmission region within a switch, thereby demonstrating the interdependence between RNA sequence, structure, and function.

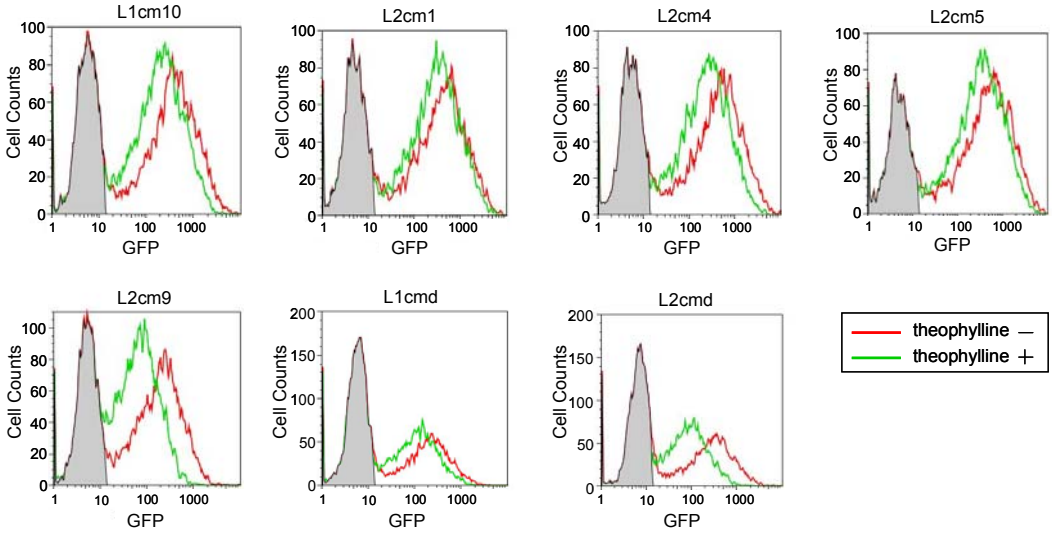
Supplementary Figures and Tables



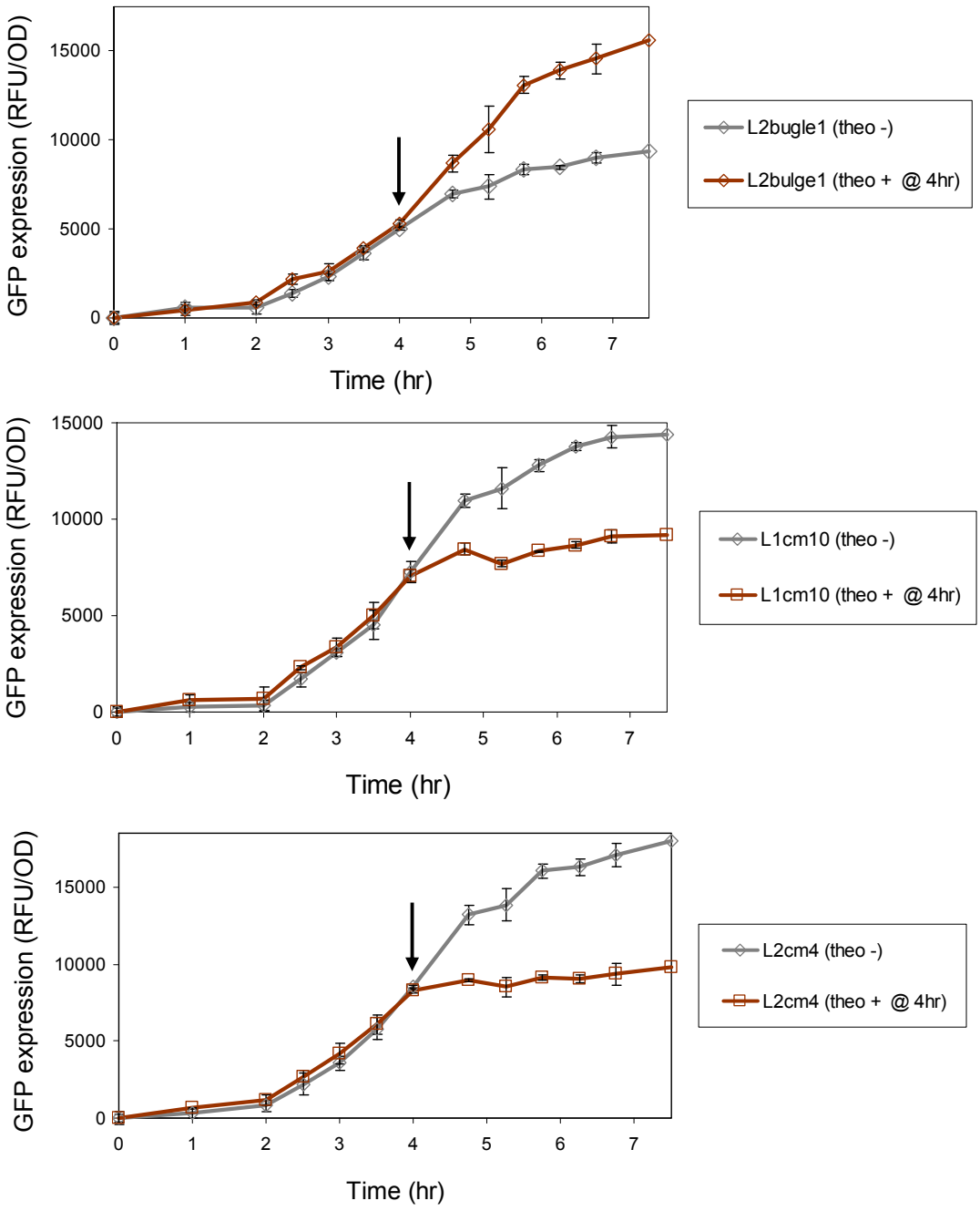
Supplementary Figure 3.1. Control constructs supporting the design strategy for engineering ligand-regulated ribozyme switches. Color schemes are as follows: catalytic core, purple; aptamer sequences, brown; loop sequences, blue; brown arrow, cleavage site. (A) Sequences of the ribozyme (sTRSV Contl) and stem integration (hhRz I) controls. (B) Sequences of the loop sequence controls in which the loop I and II sequences are replaced by the theophylline aptamer (L1R and L2R, respectively). (C) Sequences of the loop sequence controls in which the theophylline aptamer is connected directly to the loop I nucleotides through L1.3 and L1.4 (L1Theo) and the loop II nucleotides through L2.2 and L2.3 (L2Theo). (D) Gene expression levels (in fold) of the control constructs. 1-fold is defined as the reporter gene expression level of sTRSV relative to that of the background fluorescence level. The mean ±s.d. from at least three independent experiments is shown.



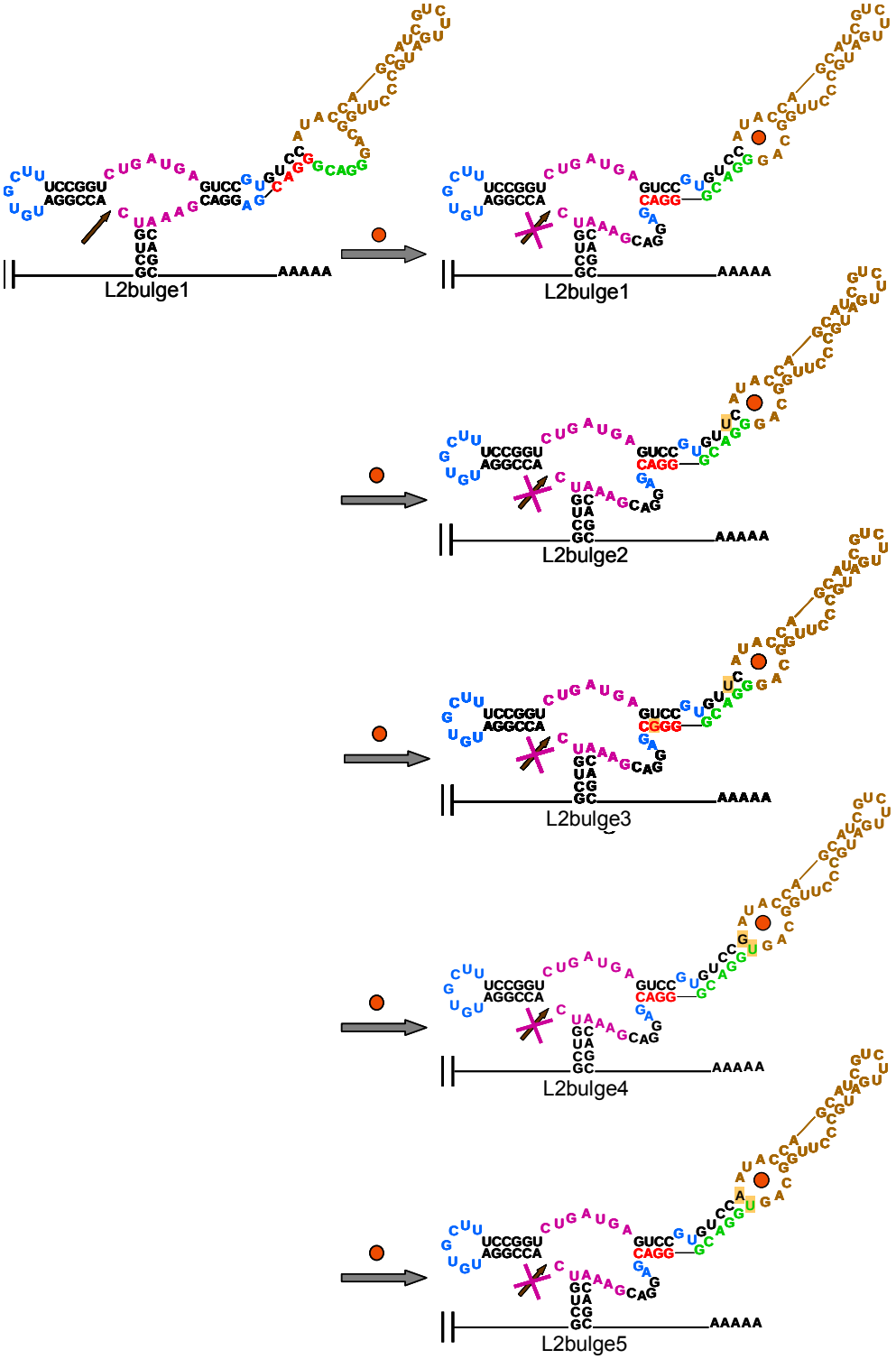
Supplementary Figure 3.2. Flow cytometry histograms of L2bulge1, L2bulgeOff1, and the ribozyme control cell populations grown in the presence (+) and absence (-) of 5 mM theophylline. Red line: cell populations grown in the absence of theophylline; green line: cell populations grown in 5 mM theophylline; shaded population: cell populations indicative of the non-induced cell population, shaded here to indicate the portion of cells in the population that have lost the plasmid and exhibit non-induced, or background, levels of autofluorescence. Histograms are representative of three independent experiments.

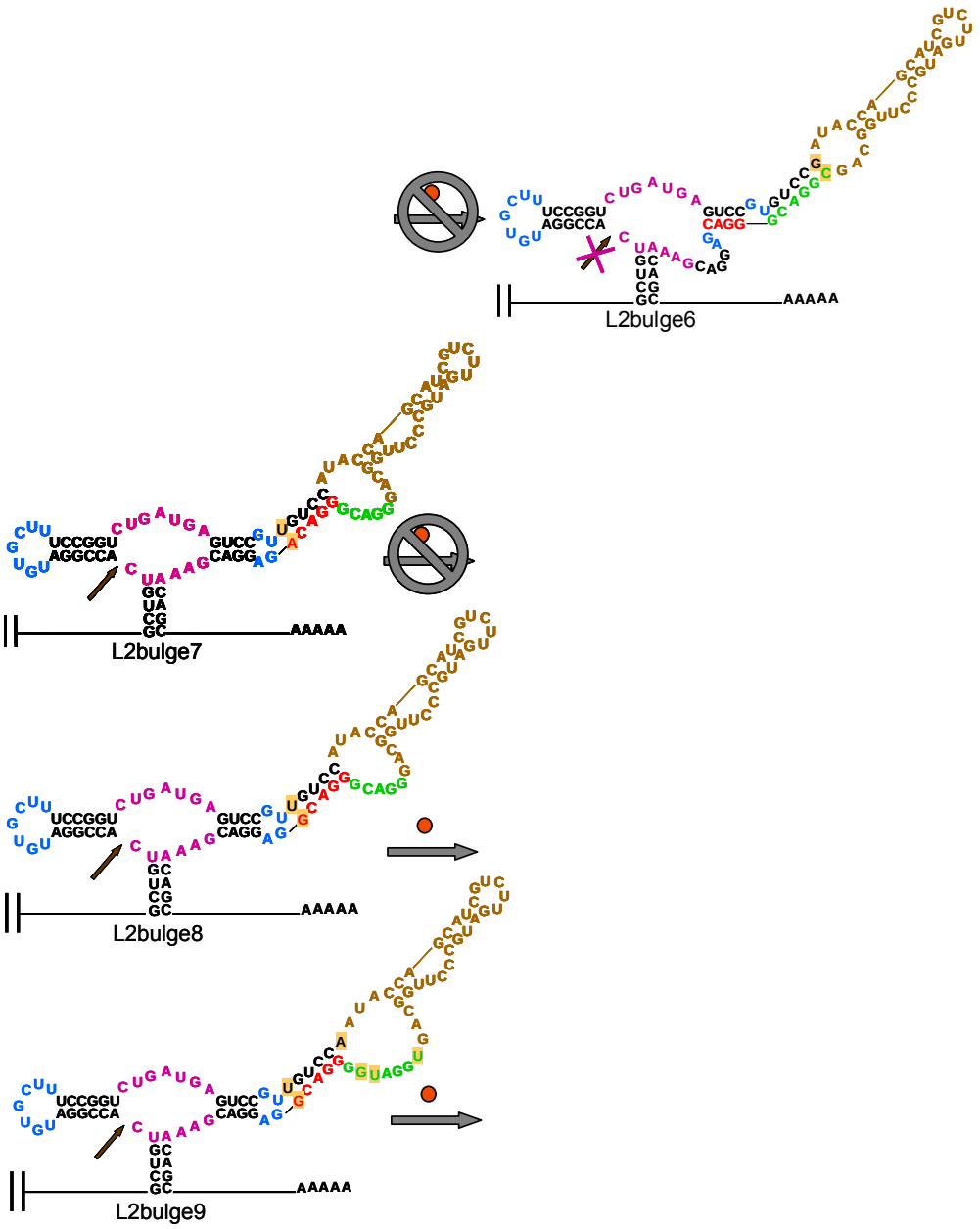


Supplementary Figure 3.3. Flow cytometry histograms of the helix slipping-based ribozyme switch cell populations grown in the presence (+) and absence (-) of 5 mM theophylline. Population data is measured and reported as described in Supplementary Figure 3.2. Histograms are representative of three independent experiments.

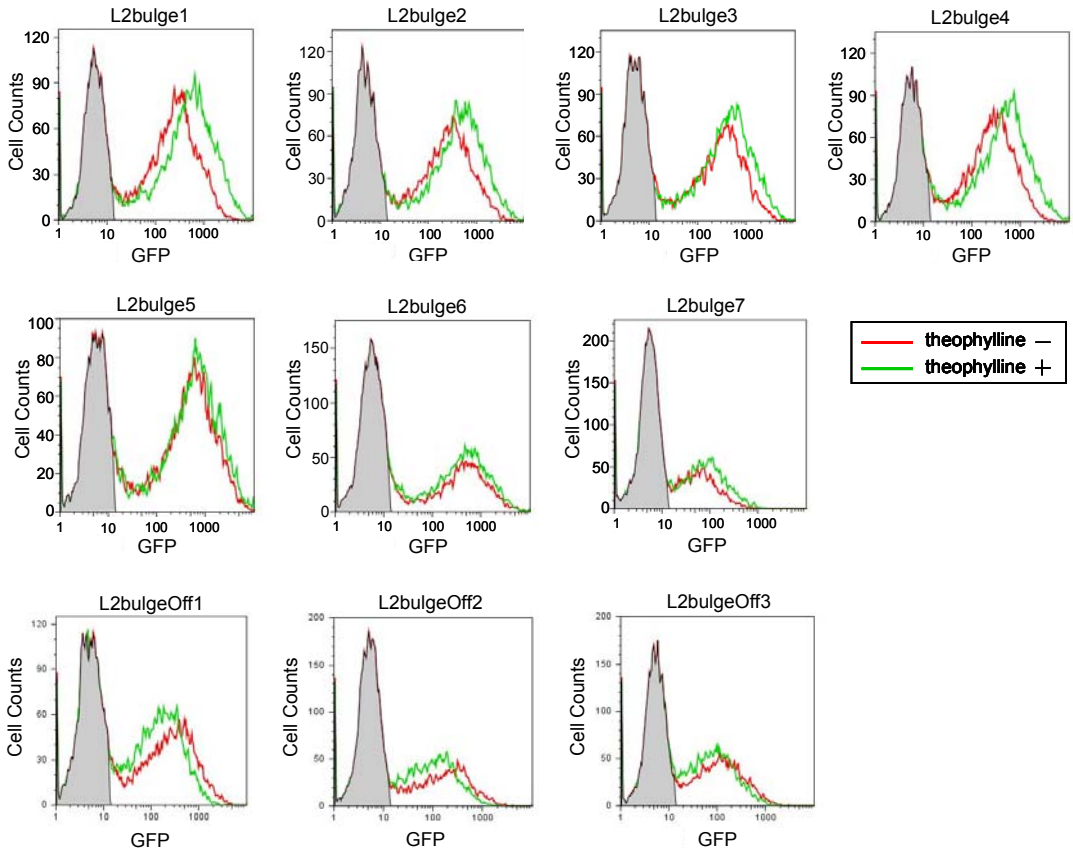


Supplementary Figure 3.4. Temporal responses of L2bulge1, L1cm10, and L2cm4 in response to the addition of 5 mM theophylline (final concentration). The time point at which theophylline was added to the cultures is indicated by an arrow. Brown: 5 mM theophylline added to growing cultures; gray: no theophylline added to growing cultures. Gene expression levels are reported as RFU/OD by dividing fluorescence units by the OD600 of the cell sample and subtracting the background fluorescence level. L2bulge1 exhibits up-regulation of GFP levels in response to the addition of theophylline; L1cm10 and L2cm4 exhibit down-regulation of GFP levels in response to theophylline addition. The mean ± s.d. from at least three independent experiments is shown for all graphs.

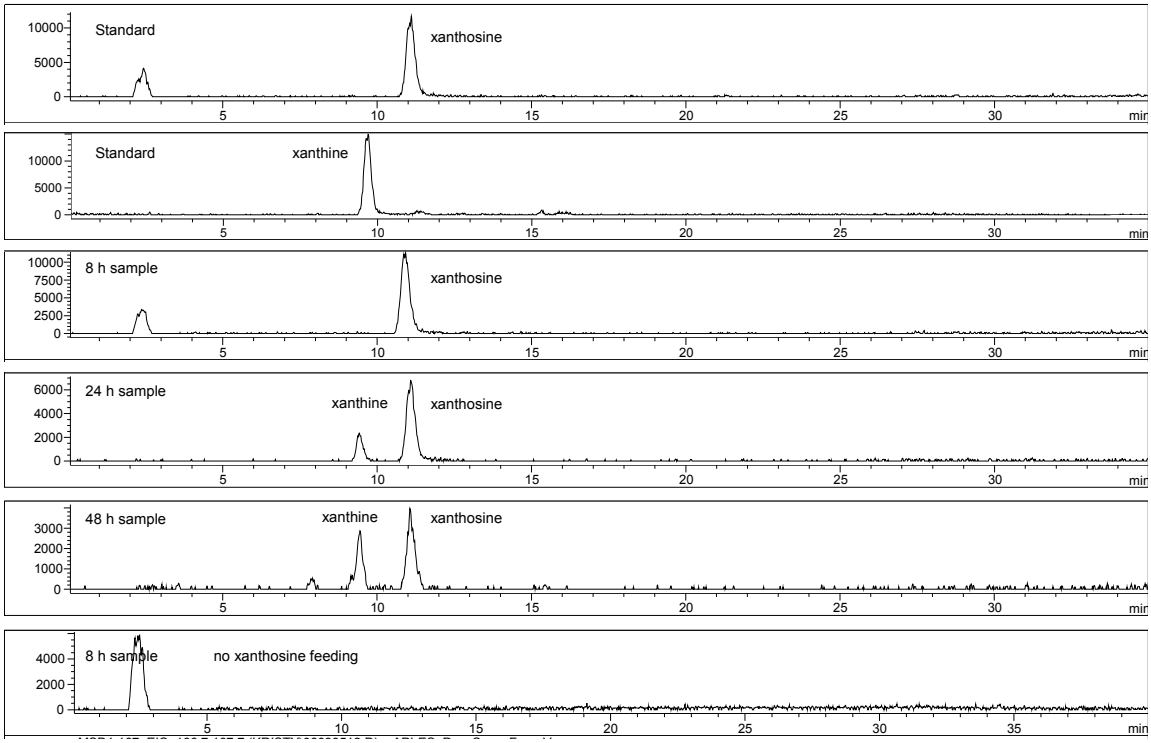

L2bulge1
L2bulge1
L2bulge2
L2bulge3
L2bulge4
L2bulge5



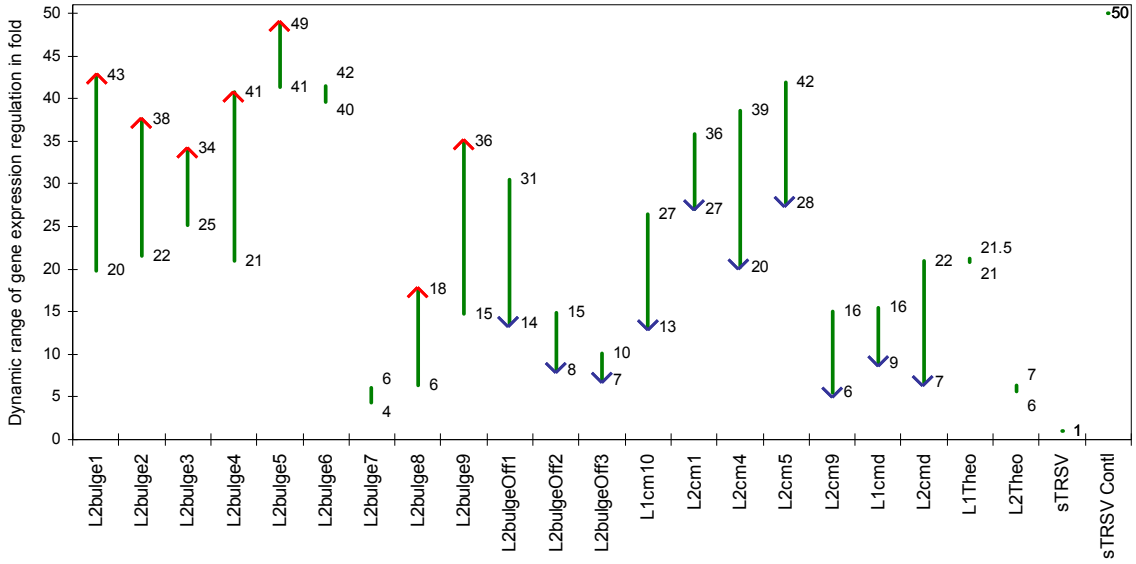
Supplementary Figure 3.5. Sequences and structures of tuned ribozyme switches in the L2bulge series. The nucleotides altered from the parent construct, L2bulge1, are highlighted. The two stable equilibrium conformations, ribozyme active and inactive conformations, are indicated for the parent ribozyme switch. The ribozyme active conformations of L2bulge2-5 are not shown as they are similar to L2bulge1. L2bulge6 and L2bulge7 assume a single predominant conformation, ribozyme inactive and ribozyme active, respectively, and do not undergo theophylline-induced conformational switching. L2bulge8 and L2bulge9, modified from L2bulge7 by reducing the stability of the ribozyme active conformation and the energy difference between the two conformations of L2bulge7, now become capable of switching. For these two modified switch constructs, only the ribozyme active conformations are shown, as their ribozyme inactive conformations are similar to those of the other switches illustrated.



Supplementary Figure 3.6. Flow cytometry histograms of the tuned ribozyme switch series cell populations grown in the presence (+) and absence (-) of 5 mM theophylline. Population data is measured and reported as described in Supplementary Figure 3.2. Histograms are representative of three independent experiments.



Supplementary Figure 3.7. Detection of intracellular accumulation of the substrate xanthosine and the product xanthine over three different time points. Accumulation of xanthosine is observed at earlier time points. Conversion of xanthosine to xanthine was detected at 24 h after substrate feeding and a higher accumulation of xanthine was detected at 48 h after substrate feeding.



Supplementary Figure 3.8. Dynamic ranges of regulation of the ribozyme switches and controls engineered in this work. The regulatory effects at 5 mM theophylline are reported on a full transcriptional range spectrum scale without normalization to the corresponding base expression level of each switch in the absence of effector (0 mM). Little or no effector-mediated gene regulatory effect is observed in the non-switch control constructs. Gene expression fold is defined as previously where 1 fold is equivalent to the reporter gene expression level of sTRSV relative to the background fluorescence level. sTRSV is the most active ribozyme construct exhibiting the lowest gene expression level and sTRSV Contl is the most inactive ribozyme construct exhibiting the highest gene expression level, providing a 50-fold range as the full spectrum equivalent to a total of 50 folds. Arrows indicate the direction of regulation as an increasing concentration of theophylline. These switches offer diverse dynamic ranges of regulation and thus provide a broader utility to fit specific applications of interest. Data are reported from three independent experiments and the ± s.d. is the same as that reported in the manuscript figures.

Supplementary Table 3.1. Relative steady-state ribozyme switch and ribozyme control transcript levels in the presence and absence of theophylline.

constructs	0 mM theopylline	5 mM theophylline	regulatory effect
sTRSV	0.08±0.01	0.11±0.01	little
sTRSV Contl	1.00±0.06	1.10±0.04	little
L2bulge1	0.49±0.04	0.77±0.10	up-regulation
L1cm10	0.66±0.05	0.43±0.06	down-regulation
L2cm4	0.67±0.05	0.38±0.06	down-regulation

Quantitative RT-PCR (qRT-PCR) analysis was performed on L2bulge1, L1cm10, L2cm4, satellite RNA of tobacco ringspot virus (sTRSV), and sTRSV control (sTRSV Contl). Transcript levels in the presence or absence of theophylline are reported as fractions relative to those of sTRSV Contl. L2bulge1 exhibits a higher steady-state level of target transcript, while L1cm10 and L2cm4 exhibited lower steady-state target transcript levels in the presence of 5 mM theophylline than in the absence of theophylline. The ribozyme controls, sTRSV and sTRSV Contl, exhibited little effect on steady-state transcript levels due to the presence of theophylline. In addition, relative steady-state levels of these switches corresponded to the relative GFP expression levels as determined through the functional ribozyme switch assays. All data are reported from three independent experiments.

Supplementary Table 3.2. Free energies (-ΔG, kcal/mol) of individual conformations (ribozyme-active and -inactive) and the energy difference (ΔΔG, kcal/mol) between the free energies of these two conformations predicted by RNAstructure 4.2.

switch constructs	free energy of aptamer-unbound conformation (-ΔG)	free energy of aptamer-bound conformation (-ΔG)	free energy difference (ΔΔG)
ON switches	ribozyme active	ribozyme inactive	
L2bulge1	38.9	38.1	0.8
L2bulge2	36.0	35.2	0.8
L2bulge3	35.5	34.6	0.9
L2bulge4	39.5	38.8	0.7
L2bulge5	39.5	39.5	0.0
L2bulge6	39.2	40.5	-1.3
L2bulge7	40.2	36.5	3.7
L2bulge8	39.4	38.0	1.4
L2bulge9	39.3	37.7	1.6
OFF switches	ribozyme inactive	ribozyme active	
L2bulgeOff1	39.3	38.6	0.7
L2bulgeOff2	39.3	37.2	2.1
L2bulgeOff3	39.9	38.2	1.7

Supplementary Table 3.3. Primer and additional ribozyme construct sequences.

name	oligonucleotide sequences	comments
5' spacer	5' AAACAAACAAA	spacer sequence preceding each construct
3' spacer	5' AAAAAGAAAAATAAAAA	spacer sequence following each construct
L1-2 fwd	5' GACCTAGGAAACAAACAAAGCTGTCACC	forward primer for all constructs
L1-2 rev	5' GGCTCGAGTTTTTATTTTTCTTTTTGCTGTTTCG	reverse primer for all constructs
yEGFP fwd	5' CGGTGAAGGTGAAGGTGATGCTACT	forward primer specific for qRT-PCR of *yegfp*
yEGFP rev	5' GCTCTGGTCTTGTAGTTACCGTCATCTTTG	reverse primer specific for cDNA synthesis and qRT-PCR of *yegfp*
ActI fwd	5' CGGTGAAGGTGAAGGTGATGCTACT	forward primer specific for qRT-PCR of *actI*
ActI rev	5' GCTCTGGTCTTGTAGTTACCGTCATCTTTG	reverse primer specific for cDNA synthesis and qRT-PCR of *actI*
L1bulge1	5' GCTGTCACCGGATGTACCGGAATACCAGCATCGTCTTGAT GCCCTTGGCAGTCTGGTCCGGTGCTTTCCGGTCTGATGAGT CCGTGAGGACGAAACAGC	no functional activity observed
L1bulge2	5' GCTGTCACCGGATGTACCGGAATACCAGCATCGTCTTGAT GCCCTTGGCAGTCCGGTCCGGTGCTTTCCGGTCTGATGAGT CCGTGAGGACGAAACAGC	no functional activity observed
L1bulge3	5' GCTGTCACCGGATGTACCGGAATACCAGCATCGTCTTGAT GCCCTTGGCAGTCTGGGTCCGGTGCTTTCCGGTCTGATGAGT CCGTGAGGACGAAACAGC	no functional activity observed
L1bulge4	5' GCTGTCACCGGATGTACCGGAATACCAGCATCGTCTTGAT GCCCTTGGCAGTCCGGGTCCGGTGCTTTCCGGTCTGATGAGT CCGTGAGGACGAAACAGC	no functional activity observed
L1bulge5	5' GCTGTCACCGGATGTACCGGAATACCAGCATCGTCTTGAT GCCCTTGGCAGTCTGGTTCCGGTGCTTTCCGGTCTGATGAGT CCGTGAGGACGAAACAGC	no functional activity observed
L1bulge6	5' GCTGTCACCGGATGTACCGGAATACCAGCATCGTCTTGAT GCCCTTGGCAGTCTGGATCCGGTGCTTTCCGGTCTGATGAGT CCGTGAGGACGAAACAGC	no functional activity observed
L1cm10	5'CCTAGGAAACAAACAAAGCTGTCACCGGATGTGCTTTCCGG TCTGATGAGTCCGTAAATGATACCAGCATCGTCTTGATGCCCTT GGCAGCTGCGAGGACGAAACAGCAAAAAGAAAAATAAAAA CTCGAG	This sequence represents a template for other communication module constructs through L1 by replacing the colored sequences with corresponding modules
L2cm4	5'CCTAGGAAACAAACAAAGCTGTCACCGGATGTCCTGGATACC AGCATCGTCTTGATGCCCTTGGCAGTCATAGCTTTCCGGTCTG ATGAGTCCGTGAGGACGAAACAGCAAAAAGAAAAATAAAAA CTCGAG	This sequence represents a template for other communication module constructs through L2 by replacing the colored sequences with corresponding modules
cma	5'CCTT 5'ACGT	ref. 19, "class I induction element"
cm2	5'CCAGG 5'TTTGA	ref. 20
cm3	5'TCTGG 5'TCTTA	ref. 20
cm6	5'GGATG 5'CAAT	ref. 21
cm7	5'GGAGG 5'CCTT	ref. 21
cm8	5'ATACG 5'CGGT	ref. 21
cm11	5'TCGAG 5'CTCTA	ref. 21
cm12	5'AGGG 5'CTCTA	ref. 21

Acknowledgements

We thank K. Hawkins for assistance in controls and HPLC experiments and data analysis; A. Babiskin for supply of pRzS and assistance in qRT-PCR assays; K. Dusinberre, J. Michener, J. Liang for assistance in controls; E. Kelsic for assistance in image presentation; Y. Chen, K. Hoff for critical reading of the manuscript. This work was

supported by the Arnold and Mabel Beckman Foundation, the National Institutes of Health, and the Center for Biological Circuit Design at Caltech (fellowship to M.N.W.).

References

1. Endy, D. Foundations for engineering biology. Nature 438, 449-53 (2005).
2. Voigt, C. A. Genetic parts to program bacteria. Curr Opin Biotechnol 17, 548-57 (2006).
3. Gossen, M. & Bujard, H. Tight control of gene expression in mammalian cells by tetracycline-responsive promoters. Proc Natl Acad Sci U S A 89, 5547-51 (1992).
4. 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, 1203-10 (1997).
5. Mandal, M. & Breaker, R. R. Gene regulation by riboswitches. Nat Rev Mol Cell Biol 5, 451-63 (2004).
6. Kim, D. S., Gusti, V., Pillai, S. G. & Gaur, R. K. An artificial riboswitch for controlling pre-mRNA splicing. Rna 11, 1667-77 (2005).
7. An, C. I., Trinh, V. B. & Yokobayashi, Y. Artificial control of gene expression in mammalian cells by modulating RNA interference through aptamer-small molecule interaction. Rna 12, 710-6 (2006).
8. Bayer, T. S. & Smolke, C. D. Programmable ligand-controlled riboregulators of eukaryotic gene expression. Nat Biotechnol 23, 337-43 (2005).
9. Isaacs, F. J., Dwyer, D. J. & Collins, J. J. RNA synthetic biology. Nat Biotechnol 24, 545-54 (2006).

10. Bunka, D. H. & Stockley, P. G. Aptamers come of age - at last. Nat Rev Microbiol 4, 588-96 (2006).

11. Tuerk, C. & Gold, L. Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. Science 249, 505-10 (1990).

12. Ellington, A. D. & Szostak, J. W. In vitro selection of RNA molecules that bind specific ligands. Nature 346, 818-22 (1990).

13. Hermann, T. & Patel, D. J. Adaptive recognition by nucleic acid aptamers. Science 287, 820-5 (2000).

14. Birikh, K. R., Heaton, P. A. & Eckstein, F. The structure, function and application of the hammerhead ribozyme. Eur J Biochem 245, 1-16 (1997).

15. Marschall, P., Thomson, J. B. & Eckstein, F. Inhibition of gene expression with ribozymes. Cell Mol Neurobiol 14, 523-38 (1994).

16. Khvorova, A., Lescoute, A., Westhof, E. & Jayasena, S. D. Sequence elements outside the hammerhead ribozyme catalytic core enable intracellular activity. Nat Struct Biol 10, 708-12 (2003).

17. Yen, L. et al. Exogenous control of mammalian gene expression through modulation of RNA self-cleavage. Nature 431, 471-6 (2004).

18. Koizumi, M., Soukup, G. A., Kerr, J. N. & Breaker, R. R. Allosteric selection of ribozymes that respond to the second messengers cGMP and cAMP. Nat Struct Biol 6, 1062-71 (1999).

19. Soukup, G. A. & Breaker, R. R. Engineering precision RNA molecular switches. Proc Natl Acad Sci U S A 96, 3584-9 (1999).

20. Soukup, G. A., Emilsson, G. A. & Breaker, R. R. Altering molecular recognition of RNA aptamers by allosteric selection. J Mol Biol 298, 623-32 (2000).

21. Kertsburg, A. & Soukup, G. A. A versatile communication module for controlling RNA folding and catalysis. Nucleic Acids Res 30, 4599-606 (2002).

22. Pelletier, J. & Sonenberg, N. Insertion mutagenesis to increase secondary structure within the 5' noncoding region of a eukaryotic mRNA reduces translational efficiency. Cell 40, 515-26 (1985).

23. Jenison, R. D., Gill, S. C., Pardi, A. & Polisky, B. High-resolution molecular discrimination by RNA. Science 263, 1425-9 (1994).

24. Berens, C., Thain, A. & Schroeder, R. A tetracycline-binding RNA aptamer. Bioorg Med Chem 9, 2549-56 (2001).

25. Hanson, S., Berthelot, K., Fink, B., McCarthy, J. E. & Suess, B. Tetracycline-aptamer-mediated translational regulation in yeast. Mol Microbiol 49, 1627-37 (2003).

26. Koch, A. L. The metabolism of methylpurines by Escherichia coli. I. Tracer studies. J Biol Chem 219, 181-8 (1956).

27. Nishiwaki, K. et al. Structure of the yeast HIS5 gene responsive to general control of amino acid biosynthesis. Mol Gen Genet 208, 159-67 (1987).

28. Ogawa, J. et al. Purification, characterization, and gene cloning of purine nucleosidase from Ochrobactrum anthropi. Appl Environ Microbiol 67, 1783-7 (2001).

29. Sambrook, J. & Russell, D. W. Molecular cloning: a laboratory manual (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 2001).

30. Mateus, C. & Avery, S. V. Destabilized green fluorescent protein for monitoring dynamic changes in yeast gene expression with flow cytometry. Yeast 16, 1313-23 (2000).

31. Gietz, R. & Woods, R. in Guide to Yeast Genetics and Molecular and Cell Biology, Part B. (eds. Guthrie, C. & Fink, G.) 87-96 (Academic Press, San Diego, 2002).

32. Isaacs, F. J. et al. Engineered riboregulators enable post-transcriptional control of gene expression. Nat Biotechnol 22, 841-7 (2004).

33. Caponigro, G., Muhlrad, D. & Parker, R. A small segment of the MAT alpha 1 transcript promotes mRNA decay in Saccharomyces cerevisiae: a stimulatory role for rare codons. Mol Cell Biol 13, 5141-8 (1993).

34. Ng, R. & Abelson, J. Isolation and sequence of the gene for actin in Saccharomyces cerevisiae. Proc Natl Acad Sci U S A 77, 3912-6 (1980).

Chapter IV: Higher-order cellular information processing using synthetic RNA devices*

Abstract

The engineering of biological systems is critical to developing effective solutions to many societal challenges including energy and food production, environmental quality, and health and medicine. Programmed cellular information processing and control devices are needed to engineer biological systems[1]. Here, we demonstrate synthetic RNA devices that perform a variety of higher-order cellular information processing operations, including logic (AND, NOR, NAND, OR gates), signal filters, and signal gain (cooperativity). RNA devices process and transmit molecular input signals to targeted protein level outputs, linking computation and logic to gene expression and thus cellular function. The devices are assembled from modular RNA components through a first-generation composition framework, highlighting the potential of such synthetic biology strategies to support the rapid engineering of cellular behavior.

*Reproduced with permission from: M. N. Win and C. D. Smolke. (2008) Manuscript submitted.

4.1. Introduction

Our ability to transmit information to and from living systems and to act on information inside living systems is critical to advancing the scale and complexity at which we can engineer, manipulate, and probe biological systems. There is a need for higher-order cellular information processing and control devices that produce new cellular functions from the diverse molecular information present within biological systems, such as small molecules, proteins, and RNA. For example, logic operations that process and translate multiple molecular inputs into prescribed levels of new molecular outputs are critical to a cell's ability to integrate diverse environmental and intracellular signals to a smaller number of phenotypic responses. As another example, basic computation operations such as signal gain, amplification, restoration, and filtering enable useful manipulation of molecular information through cellular networks.

Researchers have demonstrated many examples of molecular information processing systems that perform computation and logic with biological substrates. For example, protein-based systems that perform logic operations to convert molecular inputs to the regulation of transcriptional events have been demonstrated[2-5]. However, systems based on protein components have faced limitations in the molecular inputs that can be processed, programmability of the components themselves, functional dependence on cell-specific machinery, and the variety of information processing operations that can be readily obtained. As a second example, inspired by the diverse functions exhibited by nucleic acids[6] and the predictability of Watson-Crick base pairing interactions, researchers have built many *in vitro* information processing systems comprised of nucleic acid components, including DNA computing machines that implement logic operations and signaling cascades based solely on

nucleic acid hybridization events[7, 8], molecular automata comprised of deoxyribozymes regulated by nucleic acid inputs that perform various computation and logic functions[9-11], and molecular computers that utilize protein enzymes to regulate sequence-specific cleavage and joining of nucleic acids[12, 13]. Allosteric ribozymes that implement logic functions in response to small molecule[14, 15] and nucleic acid[16, 17] inputs have also been demonstrated *in vitro*.

Significantly, researchers have constructed a variety of single-input RNA switches that process nucleic acid and small molecule inputs to regulate gene expression events *in vivo*[18, 19]. RNA-based systems that integrate multiple RNAi substrates for combinatorial regulation of gene expression i*n vivo* have also been demonstrated [20, 21]. However, an important next challenge is to combine the inherent richness that nucleic acid substrates possess for performing information processing and control operations with the design advantages expected from the relative ease by which RNA structures can be modeled and thus designed, as compared to proteins[22, 23]. Incremental progress towards overcoming this challenge would allow many new generic devices to be engineered, which operate reliably inside living cells, provide access to otherwise inaccessible information of cellular state, and allow sophisticated exogenous and embedded control of cellular functions.

4.2. Results

4.2.1. Functional RNA device composition framework and general signal integration schemes

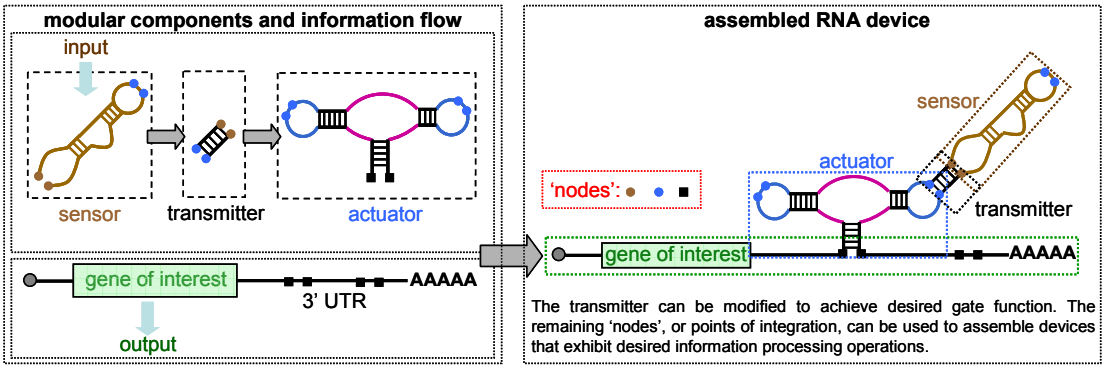
We recently described a framework for the construction of single input/output RNA devices[24] that is based on the modular assembly of three functional components: a sensor component, comprised of an RNA aptamer; an actuator component, comprised of a

hammerhead ribozyme[25]; and a transmitter component, comprised of a sequence that couples the sensor and actuator components. The transmitter component utilizes competitive hybridization events to enable design of conformational changes that are linked to functional states of the molecule. The proposed framework is also based on the modular coupling of the RNA device and the target genetic construct through the 3' untranslated region (UTR), where self-cleavage inactivates the transcript independent of cell-specific machinery. From this early framework, we demonstrated simple RNA devices that function as single-input gene expression ON and OFF switches (here referred to as Buffer and Inverter gates, respectively), which convert both cellular and exogenous molecular inputs to regulated gene expression via input-dependent regulation of ribozyme activity[24].

In engineering design, the utility of a proposed composition framework depends, in part, on the extensibility of the framework itself. We hoped to demonstrate that our careful specification of defined points of integration, or 'nodes', could be used to facilitate the assembly of putatively modular RNA components into more sophisticated cellular information processing devices (Figure 4.1A). Thus, here, we describe an extended framework for engineering higher-order RNA devices, based on three signal integration (SI) schemes that correspond to different modes of assembly for device components (Figure 4.1B). The first signal integration scheme (SI 1) is used to construct RNA devices that perform logic (AND, NOR gates) and bandpass signal filter operations through the assembly of independent single-input gates in the 3' UTR. The second integration scheme (SI 2) is used to construct devices that perform other logic operations (NAND, OR gates) through the assembly of two individual sensor-transmitter components linked to both stems of the ribozyme. The third scheme (SI 3) is used to construct devices that perform logic (AND, OR

gates) and signal gain (cooperativity) operations through the assembly of two individual sensor-transmitter components linked to a single ribozyme stem.

A **Functional composition of an RNA device**



B **Signal integration (SI) schemes**

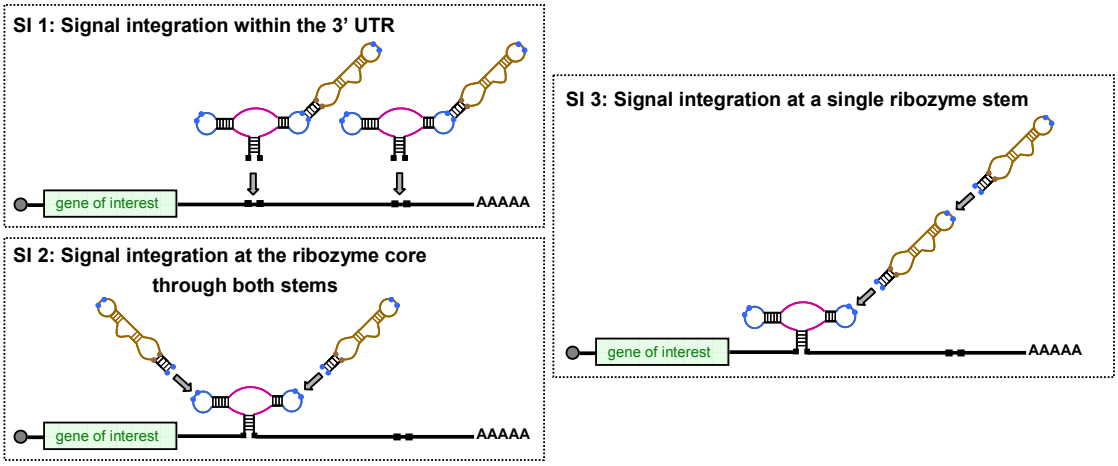
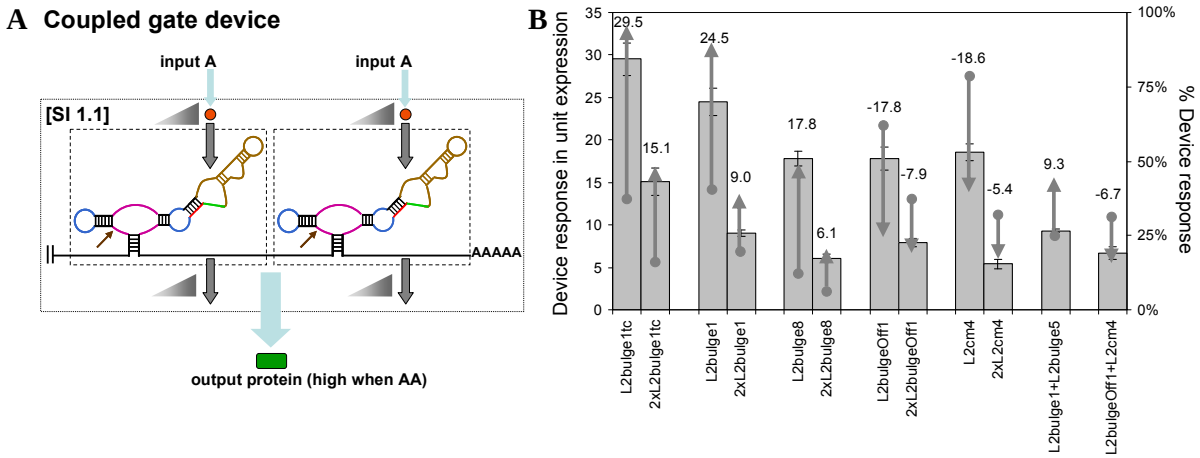


Figure 4.1. Functional RNA device composition framework. Color schemes follow those previously described[24]: brown, aptamer or sensor component; purple, catalytic core of the ribozyme or actuator component; blue, loop regions of the actuator component; green and red, competing and switching strands of the transmitter component, respectively. (A) Schematics of the functional composition framework for assembling RNA devices. An RNA device is composed of three modular components: a sensor, a transmitter, and an actuator. Information in the form of a molecular input is received by the sensor and transmitted by the transmitter to a regulated activity of the actuator, which in turn controls the target expression level as an output. Nodes specify physical points of integration between components through which devices are assembled. (B) Schematics of three primary signal integration schemes representing different component assembly strategies to build higher-order RNA devices. The RNA device in SI 1 involves multiple actuator components controlled by single sensor-transmitter components, whereas those in SI 2 and 3 involve multiple sensor-transmitter components controlling a single actuator component. The mode of assembly determines the mechanism of signal integration as highlighted by the coupled nodes.

4.2.2. Higher-order RNA device based on SI 1 (signal integration within the 3' UTR)

Different information processing operations are implemented through SI 1 by altering the function (Buffer, Inverter) and input responsiveness of the coupled single-input gates. The single-input gates act independently and therefore computation is performed through the integration of individual gate actions in the 3' UTR of the target transcript. We constructed a set of higher-order RNA devices by coupling representative Buffer or Inverter gates[24] responsive to either theophylline or tetracycline (SI 1.1; Figure 4.2A). The coupled same-input gate device has a naturally-occurring functional counterpart composed of two distinct riboswitches responsive to the same metabolite, thiamine pyrophosphate (TPP)[26]. By coupling two functionally identical single-input gates that are responsive to the same molecular input, a signal shift in the device response, or output swing, was observed from the coupled device compared to that of the single-input gate response, confirming the independent action of each gate (Figure 4.2B, Supplementary Text 4.1 and 4.2, Supplementary Table 4.1). This information processing operation can be used to program the output swing and basal output signal of a given single-input device to match the desired threshold values for a particular application.



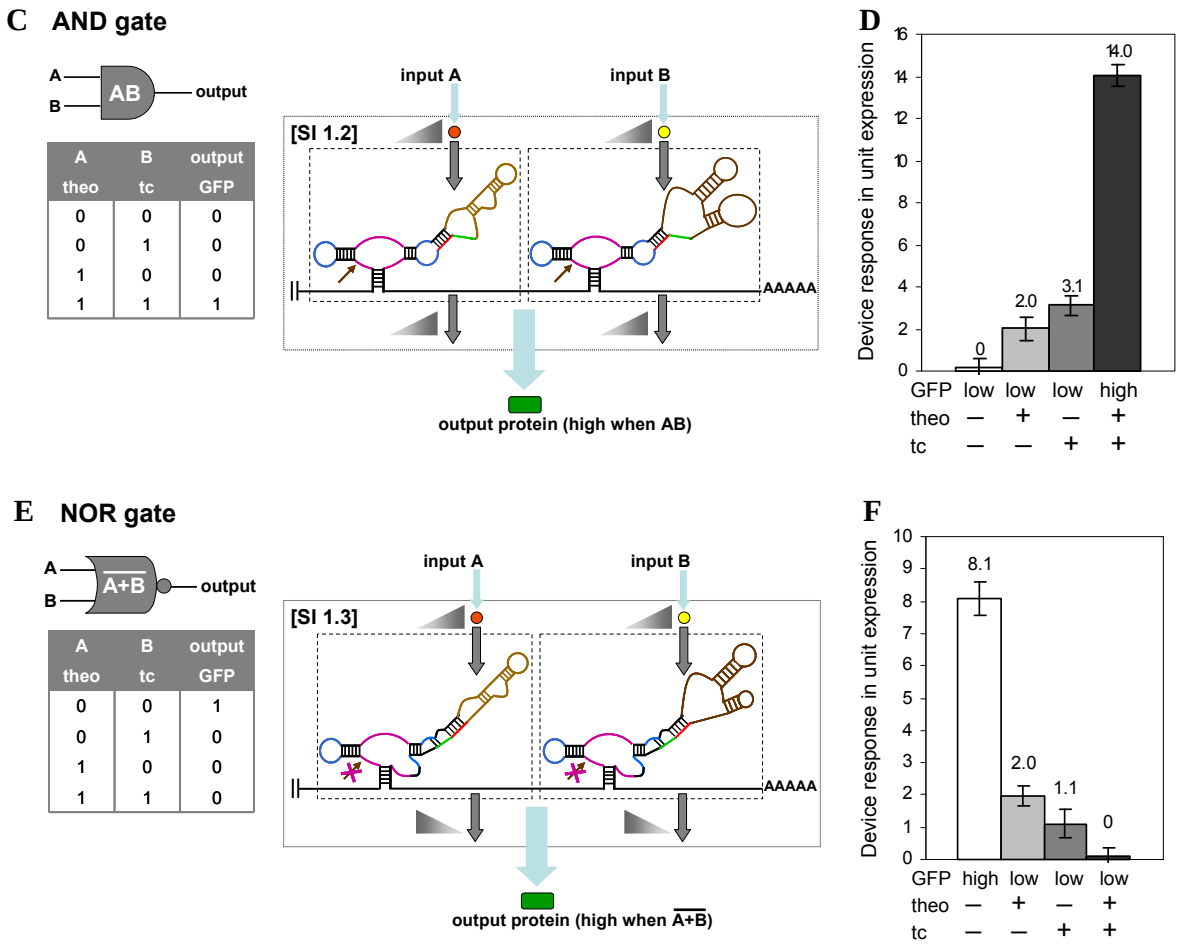


Figure 4.2. Higher-order RNA devices based on signal integration within the 3' UTR (SI 1). Color schemes follow those described in Figure 4.1. Each single-input gate is indicated in a boxed region, and triangles indicate relationships between associated gate inputs and outputs. (A) Schematic representation of an RNA device comprised of two Buffer gates responsive to the same input molecule. The RNA device functions to shift the output swing from that of the single-input gate. (B) The device response of RNA devices comprised of two single-input gates and their single-input gate counterparts. Device response (bars) is reported as the output swing in units of expression as described in Materials and Methods and the corresponding percent device response (arrows) is reported over the full transcriptional range of the employed promoter system. Output swings are reported from 0 mM to 10 mM theophylline and 0 mM to 1 mM tetracycline. The negative sign indicates the down-regulation of target gene expression by the Inverter gates. (C) Schematic representation of an RNA device that performs an AND gate operation by coupling two Buffer gates responsive to different input molecules and the associated truth table. (D) The device response of an AND gate operator (L2bulge1+L2bulge1tc). Device response under different input conditions (theo or tc (-), 0 mM; theo (+), 5 mM; tc (+), 0.25 mM) is reported as the output swing in units of expression relative to the absence of both inputs as described in Materials and Methods. (E) Schematic representation of an RNA device that performs a NOR gate operation by coupling two Inverter gates responsive to different input molecules and the associated truth table. (F) The

device response of a NOR gate operator (L2bulgeOff1+L2bulgeOff1tc). Device response under different input conditions (theo or tc (-), 0 mM; theo (+), 10 mM; tc (+), 0.5 mM) is reported as in (D), except that output swings are reported relative to the presence of both inputs.

We next constructed a higher-order RNA device that performs an AND gate operation by coupling a theophylline-responsive Buffer gate (L2bulge1[24]) and a tetracycline-responsive Buffer gate (L2bulge1tc[24]) in the 3' UTR (SI 1.2; Figure 4.2C). In the absence of the molecular inputs (theophylline or tetracycline), both Buffer gates favor the 'ribozyme-active' state, a conformation that results in transcript cleavage and low device output (low gene expression levels). In the presence of either input, one of the single-input gates remains in the ribozyme-active state and device output remains low. Device output is substantially increased only when both molecular inputs are present (Figure 4.2D). Similar to other molecular systems that perform cellular logic operations[27], the RNA devices reported here exhibit non-digital logic. We constructed a second RNA device that performs an AND gate operation by coupling L2bulge1tc and a different theophylline-responsive Buffer gate (L2bulge9[24]) to demonstrate the generality of SI 1 for constructing AND gate operators with different single-input gates (Supplementary Figure 4.1).

We constructed another higher-order RNA device that performs a NOR gate operation by coupling a theophylline-responsive Inverter gate (L2bulgeOff1[24]) and a tetracycline-responsive Inverter gate (L2bulgeOff1tc; Supplementary Figure 4.2) in the 3' UTR (SI 1.3; Figure 4.2E). The coupled different-input Inverter gate device has a naturally-occurring functional counterpart composed of two distinct riboswitches, responsive to respective metabolites coenzyme B_{12} and S-adenosylmethionine, in which the regulated gene expression output is low in the presence of either metabolite or both, thereby functioning as a

NOR gate operator[27]. Similarly, in the absence of the molecular inputs, both Inverter gates in our engineered device favor the 'ribozyme-inactive' state, a conformation that results in reduced transcript cleavage and high device output. In the presence of either input, one of the single-input gates favors the ribozyme-active state and device output is lowered. Device output is more effectively lowered when both inputs are present, as both single-input gates favor ribozyme-active states (Figure 4.2F). We also constructed a second NOR gate operation by coupling L2bulgeOff1 to a different tetracycline-responsive Inverter gate (L2bulgeOff2tc; Supplementary Figure 4.2) in order to demonstrate the generality of SI 1 for constructing NOR gate operators with different single-input gates (Supplementary Figure 4.3).

We next constructed an RNA device that performs a bandpass filter operation by coupling theophylline-responsive Buffer and Inverter gates (L2bulge1 and L2bulgeOff1) in the 3' UTR (SI 1.4; Supplementary Figure 4.4). In the absence of the molecular input, the Buffer gate favors the ribozyme-active state, resulting in low device output. However, in the presence of the input, the Inverter gate favors the ribozyme-active state, also resulting in low device output. Only over intermediate input concentration ranges do both single-input gates favor a ribozyme-inactive state, resulting in higher device output. Therefore, diverse cellular computation and logic operations can be constructed through SI 1, where layering strategies may be used to extend device designs to other information processing operations (Supplementary Text 4.3).

4.2.3. Higher-order RNA devices based on SI 2 (signal integration at the ribozyme core)

Different information processing operations are implemented through SI 2 by altering the function and input responsiveness of the coupled sensor-transmitter components (Figure

4.1B). The sensor-transmitter components act independently through the linked ribozyme stems and therefore computation is performed through the integration of individual sensor-transmitter actions in the ribozyme core of the RNA device. An independent sensor-transmitter component is indicated as an internal Inverter gate if the presence of input results in activation of the coupled component, such as an actuator or another internal gate. Similarly, an internal Buffer gate indicates a sensor-transmitter component that results in inactivation of the coupled component in the presence of input.

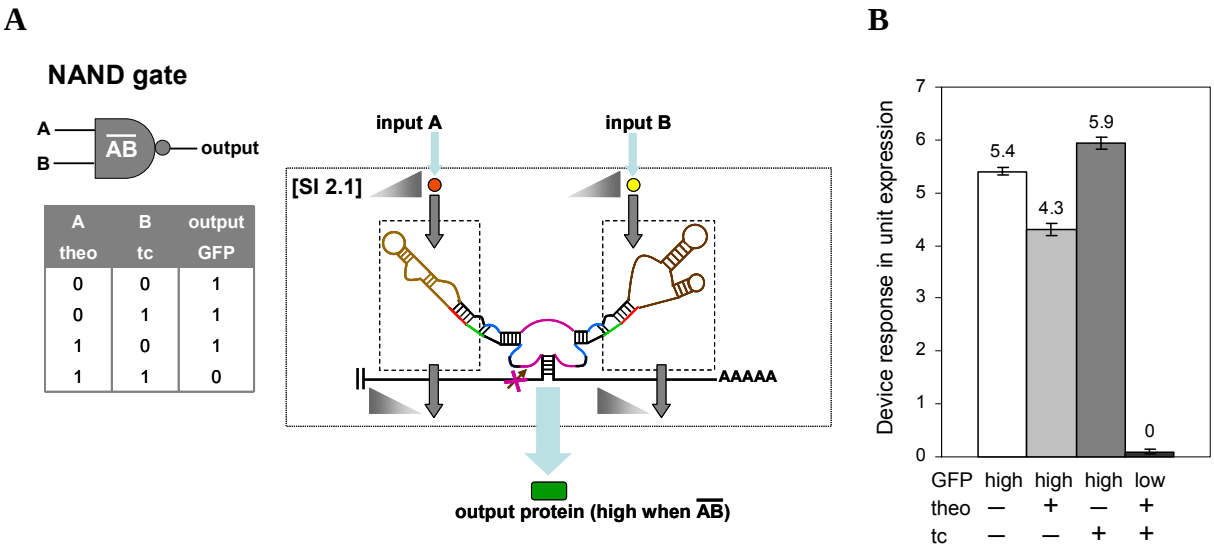


Figure 4.3. Higher-order RNA devices based on signal integration at the ribozyme core (SI 2). Color schemes follow those described in Figure 4.1. Each internal gate, comprised of a sensor-transmitter component, is indicated in a boxed region, and triangles indicate relationships between associated internal gate inputs and outputs. (A) Schematic representation of an RNA device that performs a NAND gate operation by coupling two internal Inverter gates responsive to different input molecules to different ribozyme stems and the associated truth table. (B) The device response of a NAND gate operator (L1cm10-L2bulgeOff3tc). Device response under different input conditions (theo or tc (-), 0 mM; theo (+), 10 mM; tc (+), 1 mM) is reported as in Figure 4.2F.

We constructed a higher-order RNA device that performs a NAND gate operation by coupling a theophylline-responsive internal Inverter gate (L1cm10[24]) through stem I and a tetracycline-responsive internal Inverter gate (L2bulgeOff3tc; Supplementary Figure 4.2) through stem II (SI 2.1; Figure 4.3A). In the absence of the molecular inputs, both internal Inverter gates and hence the RNA device, favor the ribozyme-inactive state, resulting in high device output. In the presence of either input, one of the internal Inverter gates remains in the ribozyme-inactive state and the device output remains high. The RNA device only favors the ribozyme-active state, resulting in low device output, when both molecular inputs are present (Figure 4.3B). We constructed a second RNA device that performs a NAND gate operation by coupling L1cm10 and a different tetracycline-responsive internal Inverter gate (L2bulgeOff1tc; Supplementary Figure 4.2) in order to demonstrate the generality of SI 2 for constructing NAND gate operators with different sensor-transmitter components (Supplementary Figure 4.5). By altering the function of the coupled sensor-transmitter components, other logic operations can be constructed through SI 2, such as an OR gate operation through the coupling of two internal Buffer gates (Supplementary Text 4.4).

4.2.4. Higher-order RNA devices based on SI 3 (signal integration through a single ribozyme stem)

In SI 3 different information processing operations are constructed through the coupling of multiple sensor-transmitter components through one stem of the ribozyme (Figure 4.1B). Here, the actions of the sensor-transmitter components are coupled, with computation occurring via the integrated action of the internal gates within a single ribozyme stem. Internal gates are linked through the aptamer loop of the lower gate, IG(n), to the transmitter sequence of the higher gate, IG(n+1), where the state of the internal gate adjacent

to the ribozyme (IG1) determines the state of the RNA device. We constructed a second type of higher-order RNA device that performs an AND gate operation by coupling a theophylline-responsive internal Buffer gate (IG1) and a tetracycline-responsive internal Inverter gate (IG2) at stem II (SI 3.1; Figure 4.4A). In the absence of both molecular inputs or the presence of input to IG1, the RNA device favors the ribozyme-active state, resulting in low device output. In the presence of the molecular input to IG2, although the state of IG2 changes, the RNA device remains in the ribozyme-active state and the device output remains low. Only in the presence of both molecular inputs do the states of both internal gates change and the RNA device favors the ribozyme-inactive state, resulting in high device output (Figure 4.4B). We constructed two other RNA devices through SI 3 that perform AND gate operations to demonstrate the generality of the assembly scheme for constructing AND gate operators with different internal gates (Supplementary Figure 4.6). We also constructed RNA devices that perform OR gate operations through SI 3 (Supplementary Text 4.4).

We used SI 3 to examine the design strategies for RNA devices that perform signal gain operations, or programmed cooperativity. We constructed RNA devices that perform signal gain by coupling theophylline-responsive internal Buffer (IG1) and Inverter (IG2) gates (SI 3.2; Figure 4.4C). The sensor-transmitter components of the RNA device can be programmed to bind inputs in a cooperative manner by manipulating the relative energies required to switch the device between different states (programmed through the transmitter components). Functional characteristics of cooperative ligand-binding regulatory systems typically involve a larger change in the response properties transiting from a low-affinity state to a high-affinity state as more ligands occupy the available binding sites, and are quantitatively represented by Hill coefficients (n_H) greater than one[28].

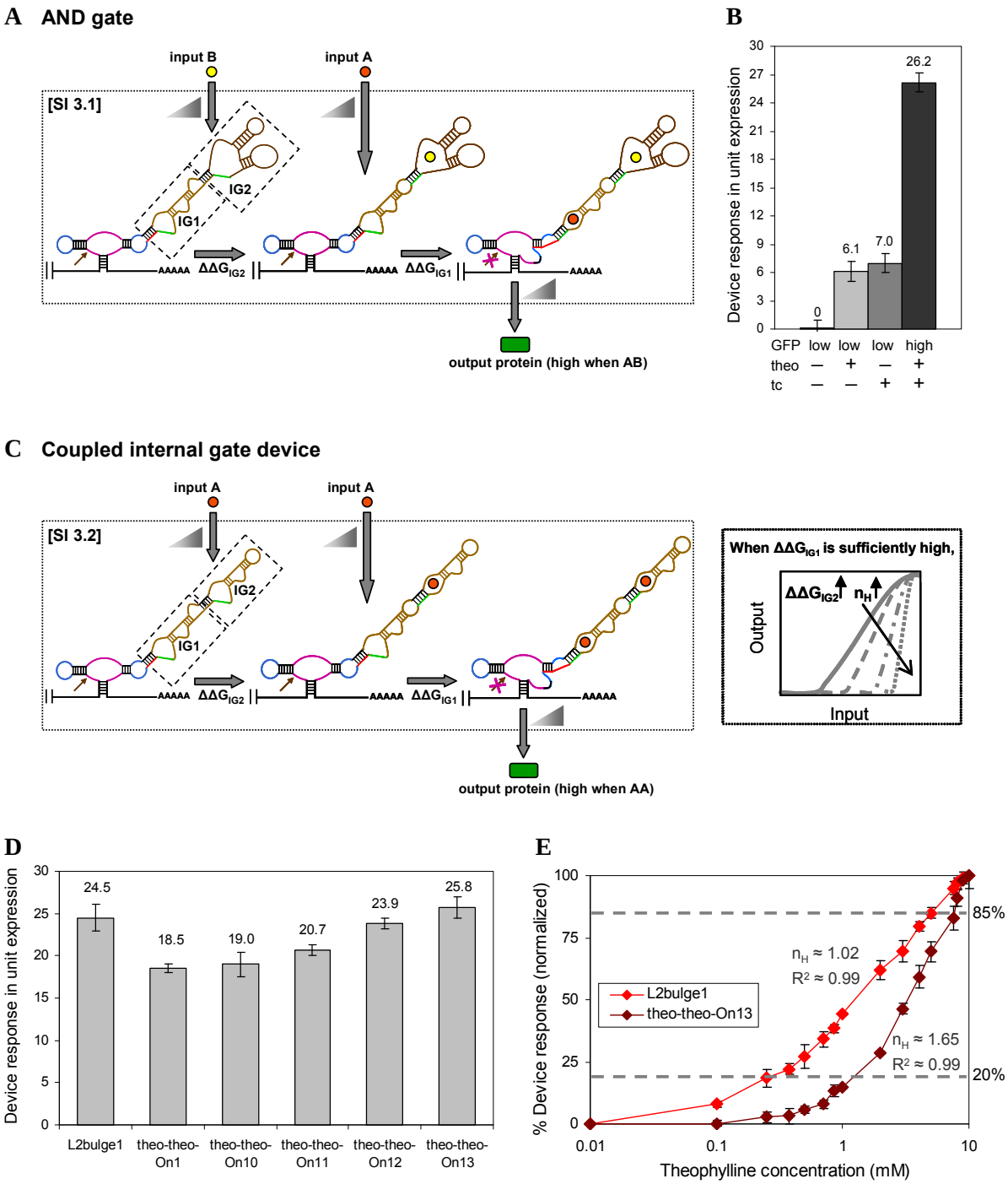


Figure 4.4. Higher-order RNA devices based on signal integration at a single ribozyme stem (SI 3). Color schemes follow those described in Figure 4.1. Each internal gate (IGn), comprised of a sensor-transmitter component, is indicated in a boxed region, and triangles indicate relationships between associated internal gate inputs and device outputs. The three states that the device can adopt and associated free energy changes between each state are illustrated. (A) Schematic representation of an RNA device that performs an AND gate operation by coupling internal Buffer (IG1) and Inverter (IG2) gates responsive to different

input molecules to a single ribozyme stem. (B) The device response of an AND gate operator (tc-theo-On1). Device response under different input conditions (theo or tc (-), 0 mM; theo (+), 2.5 mM; tc (+), 0.5 mM) is reported as in Figure 4.2D. (C) Schematic representation of an RNA device comprised of internal Buffer (IG1) and Inverter (IG2) gates responsive to the same input molecule coupled to a single ribozyme stem. (D) The device response of RNA devices comprised of internal Buffer and Inverter gates and their single-internal gate device counterpart (L2bulge1). Device response is reported as in Figure 4.2B. Theo-theo-On10 – 13 are devices that exhibit varying levels of signal gain. (E) The device response of theo-theo-On13 shows a high degree of programmed cooperativity compared to that of L2bulge1. The percent device response is plotted by normalizing corresponding dynamic switching ranges between the absence and presence of 10 mM theophylline to 0-100% as described in Materials and Methods.

We first engineered a series of nine RNA devices, in which the sequences within the IG2 transmitter component were altered to modify the energetic differences between the first and second states ($\Delta\Delta G_{IG2}$), while keeping the energetic differences between the second and third states constant ($\Delta\Delta G_{IG1}$ = 0.3 kcal/mol) (Supplementary Table 4.2). All nine RNA devices function as Buffer gates similar to L2bulge1 (Supplementary Figure 4.7); however, none of the devices exhibited gain in their output response ($n_H \approx 1$) (Supplementary Figure 4.8). This suggests that the energy required to switch between the second and third states ($\Delta\Delta G_{IG1}$; programmed into IG1 transmitter component) may play a critical role in simulating different effective binding affinities between IG1 and IG2 that will result in a signal gain operation (Supplementary Text 4.5). Therefore, we constructed a second series of RNA devices in which the energetic difference between the second and third states is increased ($\Delta\Delta G_{IG1}$ = 1 kcal/mol; Supplementary Table 4.2). All four devices function as Buffer gates (Figure 4.4D) and exhibited substantial levels of gain in the output response as indicated by Hill coefficients greater than 1 (Supplementary Figure 4.9), where theo-theo-On13 exhibited the highest signal gain ($n_H \approx 1.65$) (Figure 4.4E) and a similar degree of cooperativity as the naturally-occurring glycine riboswitch[29]. In general, the degree of programmed cooperativity (n_H) was positively correlated with $\Delta\Delta G_{IG2}$ values. We also constructed a series of RNA

devices in which internal Inverter gates were placed into IG1. All eight devices function as Inverter gates similar to L2bulgeOff1 (Supplementary Figure 4.10), and one of these devices (theo-theo-Off6) exhibited slight signal gain ($n_H \approx 1.2$) (Supplementary Figure 4.11). Mutational studies validated that the device response and signal gain operation is achieved through binding of the molecular input to both sensor components (Supplementary Figures 4.12-4.15).

4.3. Discussion

Functional composition frameworks that support the programming of complex devices through the modular assembly of distinct components are important foundations to engineering design[1]. Such frameworks support the efficient and reliable engineering of diverse device functions from well-characterized components without complex device redesign. We have developed a first-generation composition framework for constructing higher-order RNA devices. Functional modularity is a critical element of any composition framework and achieved in part here through the separation of device functions (sensing, actuation, and information transmission) into distinct components. Rational modular assembly is achieved by controlling information transmission between the sensing and actuation components through hybridization interactions. Therefore, while the functions of sensing and actuation frequently rely on more complex tertiary interactions, which are not accounted for in this first-generation framework, the integration of these functions into an RNA device is simplified via a transmitter that acts to both insulate component functions and control the interactions between components through predictive secondary structure interactions.

Besides the utility of the devices themselves, the variety of cellular information processing operations demonstrated here contributes important validation for our modular assembly and rational design approach. In addition, the framework may be further extended to more complex devices by combining the proposed signal integration schemes (SI 1, 2, and 3) within an RNA device. Future efforts leading to new device designs that enable other computation, such as signal restoration and amplification, will be critical to the extension to more complex information processing schemes.

The integration of future scientific and technological advances with the design approaches presented here, should lead to improved, next-generation frameworks for more reliable and robust assembly of RNA devices. For example, scientific advances that lend further insight into RNA structure-function relationships[30], and improve predictions of RNA secondary and tertiary structures relevant to *in vivo* folding environments[23], will allow for the development of improved modular assembly schemes, where the insulation of device functions across distinct components and controlled interaction between these components remains a design challenge. As a second example, the development of modeling tools that can predict both thermodynamic and kinetic properties of RNA folding *in vivo*[22], incorporate tertiary interactions, and link those properties to functional states and gene expression pathways will support future design tools that efficiently optimize and program device properties *in silico*. As a third example, technological advances that allow for the efficient generation of well-characterized libraries of sensor components, RNA aptamers[31, 32], that recognize biologically-relevant molecules, function in the cellular environment, and are compatible with the composition frameworks will be critical to the broader implementation of RNA devices toward user-specified processing of environmental and intracellular signals.

Taken together, the thoughtful combination of scientific research and engineering theory will allow still more sophisticated RNA devices to be developed. The resulting improvements in our ability to transmit information to and from living systems, and implement control within cells themselves, will transform how we interact with and program biology.

4.4. Materials and Methods

4.4.1. Plasmid construction, cloning, and cell strains

Using standard molecular biology techniques[33], the plasmid pRzS, harboring the yeast-enhanced green fluorescence protein (yEGFP)[34] under the control of a GAL1-10 promoter, was constructed as previously described[24] and employed as a universal vector for the characterization of all higher-order RNA devices. All RNA device constructs were generated by PCR amplification using the appropriate oligonucleotide templates and primers. All oligonucleotides were synthesized by Integrated DNA Technologies (IDT). Single ribozyme devices (SI 2 and 3) were cloned into two unique restriction sites, *Avr*II and *Xho*I, 3 nucleotides downstream of the stop codon of yEGFP and upstream of an ADH1 terminator sequence. For dual ribozyme devices (SI 1), the second single-input gate including spacer sequences was cloned immediately downstream of the first single-input gate in the second restriction site (*Xho*I). Sequences of all devices are available in Supplementary Text 4.6. Representative secondary structures and sequences are illustrated in Supplementary Figures 4.16 and 4.17. Cloned plasmids were transformed into an electrocompetent *Escherichia coli* strain, DH10B (Invitrogen) and all ribozyme constructs were confirmed by subsequent sequencing (Laragen, Inc). Confirmed plasmid constructs were transformed into a

Saccharomyces cerevisiae strain (W303 *MATα his3-11,15 trp1-1 leu2-3 ura3-1 ade2-1*) using standard lithium acetate procedures[35].

4.4.2. RNA secondary structure prediction, free energy calculation, and corresponding proposed mechanism

RNAstructure 4.2 (http://rna.urmc.rochester.edu/rnastructure.html) was used to predict the secondary structures of all RNA devices and their corresponding thermodynamic properties as previously described[24]. Prediction of the secondary structures of the RNA devices based on SI 1 and 2 have been previously described[24]. RNA sequences that are predicted to adopt at least two stable equilibrium conformations (ribozyme active and inactive) were constructed and characterized for their functional activity. Our design strategy is based on the conformational dynamics characteristic of RNA molecules that enables them to distribute between these two different conformations: one in which the competing strand is not base-paired or base-paired such that the ligand-binding pocket is not formed, and the other in which the competing strand is base-paired with the aptamer (sensor) base stem, displacing the switching strand and thus allowing the formation of the ligand-binding pocket. Strand displacement results in the disruption (Buffer gate) or restoration (Inverter gate) of the catalytic core of the actuator ribozyme. Binding of input to the latter conformation shifts the equilibrium distribution to favor the input-bound form as a function of increasing input concentration. For RNA devices comprised of two internal gates (SI 3), RNA sequences that are predicted to adopt generally at least three stable equilibrium conformations of interest, as illustrated in Figure 4.4A and 4.4C, were constructed and characterized for their functional

activity. The device design strategies and their regulatory mechanisms closely follow those described above.

4.4.3. In vivo assays for characterization of RNA device properties and fluorescence quantification

As previously described[24], *S. cerevisiae* cells harboring plasmids carrying appropriate RNA devices were grown in synthetic complete medium supplemented with an appropriate amino acid dropout solution and sugar (2% raffinose, 1% sucrose) overnight at 30^{o}C. The overnight cell cultures were back-diluted into fresh medium to an OD_{600} of approximately 0.1. At the time of back-dilution, an appropriate volume of galactose (2% final concentration) or an equivalent volume of water were added to the cultures for the induced and non-induced controls, respectively. In addition, an appropriate volume of concentrated input stock dissolved in medium, or an equivalent volume of the medium (no input control) was added to the cultures (to the appropriate final concentration of theophylline, tetracycline, or both inputs, as described in the figure legends). The back-diluted cells were then grown to an OD_{600} of 0.8-1.0 or for a period of approximately 6 hours before measuring output GFP levels on a Cell Lab Quanta SC flow cytometer (Beckman Coulter). Output GFP expression level distributions within the cell populations were measured using the following settings: 488 nm laser line, 525 nm bandpass filter, and a PMT setting of 5.83. Fluorescence data were collected from 10,000 viable cell counts of each culture sample under low flow rates. A non-induced cell population was used to set a background level, and cells exhibiting fluorescence above this background level are defined as the GFP-expressing cell population.

4.4.4. Characterization of device higher-order information processing properties

Device responses are reported as output swings, or dynamic ranges of gene expression, in fluorescence units of expression in the presence of both inputs relative to the levels in the absence of inputs. To better represent the functional behaviors of the NOR and NAND operators, the output swings are reported as levels in the absence of inputs relative to the levels in the presence of both inputs. Output swings represent arithmetic differences between the expression levels in the absence and presence of appropriate molecular inputs. As previously described, 1 unit expression is defined as the gene expression level of the construct carrying the parental active ribozyme sTRSV relative to the background fluorescence level[24]. The expression level of the sTRSV construct is ~2% of that of the construct carrying the inactive ribozyme control sTRSV Contl or the full transcriptional range of 50 units of expression. Percent device response represents the expression level of an RNA device in the absence or presence of appropriate molecular inputs normalized to the expression level of the inactive ribozyme control sTRSV Contl.

Cooperative binding activities of RNA devices were determined using the Hill equation: $y = y_{\max} x^{n_H} / \left(x^{n_H} + K^{n_H} \right)$ where y is the gene expression response at an input concentration x, y_{max} is the maximum gene expression response or saturation level, and n_H and K represent the Hill coefficient and the ligand concentration at the half maximal response, respectively. Experiments demonstrate that the device responses begin to saturate at 10 mM theophylline, such that Hill coefficients were determined by normalizing dynamic switching ranges or device output swings between the absence and presence of 10 mM theophylline to 0-100% and plotting log [fraction expressed (or repressed) / (1 - fraction expressed (or repressed))] versus log [input concentration], where the slope represents the

Hill coefficient (n_H). All fluorescence data and mean ±s.d. are reported from at least three independent experiments.

4.5. Supplementary Information

Supplementary Text 4.1: *RNA device response properties and standards in data presentation*

There has been significant effort directed to the characterization of natural and engineered RNA devices. These efforts have resulted in important descriptions and demonstrations of RNA devices; however, the work is often reported through different metrics and standards. Standard means of reporting the characterized device properties are needed to accurately evaluate, compare, and appreciate the functional properties of the diverse RNA devices that have been developed or will be developed.

The RNA device properties that characterize the performance of a device include output swing (absolute difference of the dynamic range; here reported as device response), output fold induction or repression (ratio of the dynamic range: [signal in the presence (absence) of input]/[signal in the absence (presence) of input]), baseline expression (expression level in the absence of ligand; here reported as output basal signal), and input swing (input concentration over which device output changes). In order to fully characterize the dynamic range of an RNA device, either the baseline expression and the output swing or the baseline expression and the output fold induction (repression) should be reported. However, such dynamic range data cannot be compared across different genetic constructs and systems which can alter the observed response of an RNA device. For example, different organisms will have different transcriptional capacities; different regulated genes will have

different fold expression/activity levels (e.g., enzyme-based reporters exhibit turnover of a substrate and an amplified fold induction range relative to fluorescent protein-based reporters); and different promoters will have different fold transcriptional ranges. Therefore, reporting device response properties relative to standards are critical to enabling comparison of the performance of different devices within the context of different genetic constructs and systems.

Here, we propose the use of two standards in RNA device characterization: (i) the level of gene expression from the genetic construct (including promoter, gene, etc.) in the absence of the RNA device (100%; signal standard), and (ii) the level of gene expression in the absence of the genetic construct (0%; background standard). The proposed standards allow researchers to determine the performance of the RNA device across the full transcriptional range of a specified promoter, without any non-specific effects that an inactive RNA device might exhibit due to its location relative to other components in the genetic construct and its secondary structure. The use of reference standards is important because the RNA device (and therefore its performance) is coupled to other components in the genetic construct, including a promoter. Therefore, components can be changed to alter the baseline expression level relative to the signal standard as appropriate for a given application.

A device architecture that enables modification of baseline expression levels of single-input gates is shown in Figure 4.2A, where multiple single-input gate devices are coupled to alter both the baseline expression and output swing. We selected single-input gates with varying baseline expression levels to demonstrate the effects of gate coupling on baseline expression from the device (Figure 4.2B; Supplementary Text 4.2). We have

previously reported on a tuning strategy targeted to the transmitter component that can be used to build single-input gates with lower baseline expression levels (L2Bulge8; ~12%)[24]. Therefore, the combination of these two strategies (transmitter tuning and gate coupling) results in devices that exhibit much lower baseline expression levels (2xL2Bulge8; ~7%). We report output swing and baseline expression in Figure 4.2B to demonstrate the tuning of baseline expression. To simplify data presentation and focus on the response of the RNA devices to inputs, we report only output swing for most of the other devices in the main figures, and report baseline expression levels in the Supplementary Information (Supplementary Table 4.1). In addition, another straightforward way to alter the baseline expression from an RNA device is to alter the promoter that it is coupled to. For example, in the systems reported here all devices are coupled to a very strong promoter (GAL1-10). If we replaced that promoter with a weaker promoter, the baseline expression level would be much lower relative to the signal standard.

With the goal of integrating RNA devices into different genetic circuits (comprised of various biological components), such standardized characterization information is critical to match properties of the components in the circuit to achieve the desired system response. RNA devices do not necessarily need to exhibit output swings that span the full transcriptional range of a very strong promoter in order to be biologically relevant. Many endogenous proteins and enzymes are expressed at levels much lower than that obtained from the stronger promoters commonly used in recombinant work. In addition, proteins can exhibit very different thresholds of titratable function depending on their activities, such that a very low baseline expression level is not always necessary. Even natural riboswitches may not be used to titrate enzyme concentrations across their full response curves, as that would

require cells to regulate input metabolite concentrations to these regulators over a ~10^4-10^5-fold range. As such, an important property of RNA devices is their ability to be tuned to exhibit different device response properties using (1) energetic tuning strategies targeted to the transmitter component[24]; (2) coupled single-input gates (Figure 4.2B); and (3) component matching[36, 37]. These strategies provide important flexibility in tuning RNA device response to fit applications with different performance requirements. We have demonstrated previously that the output swings and baseline expression levels exhibited by RNA devices are biologically relevant, specifically in the application of intracellular detection of metabolic concentrations (where an output swing outside the noise in gene expression is important) and the regulation of cell growth/death (where the ability to titrate the output swing across a threshold concentration of the regulated protein is important)[24]. In addition, there are many other examples where non-coding RNAs play key regulatory roles in controlling biological function without exhibiting regulatory ranges across the full transcriptional range of the promoter system of the genetic construct[38-41].

Supplementary Text 4.2: *Predicted and observed response properties of coupled single-input gates*

Coupled single-input gate devices (SI 1) are comprised of single-input gates that are expected to act independently. Independent function of the single-input gates results in several predictions, regarding the response properties of such coupled gate devices relative to the single-input gates, previously described by Welz and Breaker in a tandem riboswitch system composed of two independent riboswitches[26]. However, the predicted changes in the

device response properties were not shown to be exhibited by the naturally-occurring functional counterpart[26], and are examined here for the synthetic devices.

The first predicted property of a coupled single-input gate device is that it will exhibit decreased basal output signals from the single-input gate. The expected decrease in basal output signal can be predicted from the single-input gate responses and follows a straightforward probability determination that both gates are in the ribozyme-inactive state (requiring AND behavior):

$$p_d = p_1 * p_2$$

where p is the fraction in the ribozyme-inactive state (determined as the percent gene expression relative to the ribozyme-inactive control); subscripts 1, 2, and d indicate single-input gate 1, single-input gate 2, and the coupled single-input gate device, respectively. The predicted and measured basal output signals are shown in Supplementary Table 4.1. For most of the coupled single-input gate devices the predicted and measured basal output signals match well, supporting the independent function of the single-input gates. There are two coupled single-input gate devices, both comprised of L2cm4, for which there is not a strong match between the predicted and measured values. The results indicate that L2cm4 may not function independently when coupled in a higher-order device. L2cm4 has a transmitter component that functions through a different mechanism than the other single-input gates examined here[24], specifically through a helix-slipping mechanism[42]. This information transmission mechanism requires the presence of non-Watson-Crick base pairs within the transmitter component, which may result in weaker device structural stability, potentially allowing non-specific interactions with surrounding sequences and thus interfering with the independent function of this single-input gate.

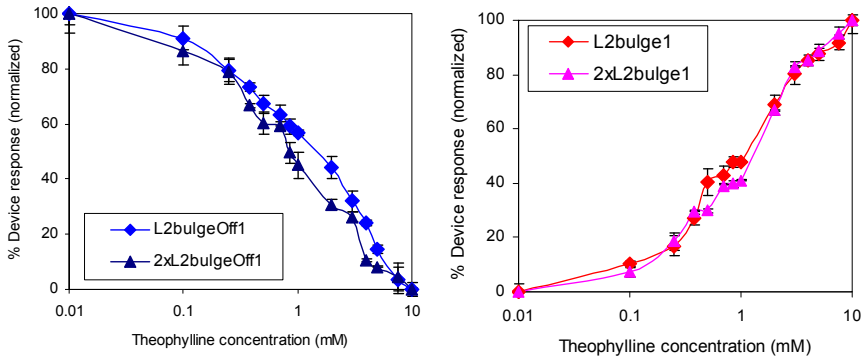
The effect of decreased basal output signal, has also been predicted to result in an increased dynamic range for such systems[26]. This would generally be true under situations in which the input concentration is saturating to the response of the system and irreversible rates do not dominate reversible rates. In the experimental systems examined here, the input ligands may not be at fully saturating concentrations due to transport limitations across the cell membrane and toxicity of the input molecules at high concentrations. In addition, in certain systems the irreversible rate of ribozyme cleavage may compete with the reversible rate of conformational switching.

The second and third predicted properties of coupled single-input gate devices apply to devices that respond to the same inputs (SI 1.1) and apply to the characteristics of the input-response curve. The second property is associated with the sensitivity of the device to input concentration. As previously pointed out, devices that couple Inverter gates (repress gene expression) are predicted to trigger a gene control response at lower input concentrations[26, 43]. This behavior results from such coupled Inverter gate devices functioning essentially through OR behavior, as the independent activation of either single-input gate device through input binding results in the repression of gene expression from a transcript. However, devices that couple Buffer gates (activate gene expression) are expected to trigger a gene control response at higher input concentrations, as the independent activation of both devices through input binding (AND behavior) is required to activate gene expression from a transcript.

The third property is associated with the slope of the response curve over ranges in gene expression. Coupled single-input gate devices are predicted to result in a more 'digital' response curve[26], where the same output dynamic range can be achieved with a lower change

in input concentration. This effect should be true for both coupled Inverter and Buffer gate devices, although the actual increase in the 'digital' nature of the response curve is predicted to be quite low[26]. In addition, this effect would only generally be true under situations in which the input concentration is saturating to the response of the system. For example, at lower input concentrations (i.e., input concentrations lower than the midway point of the input swing), the coupled Inverter gate device is predicted to have a higher slope than the single-input gate, whereas the coupled Buffer gate device is predicted to have a lower slope than the single-input gate. Therefore, the predicted effects on the slope of the response curve are anticipated to be small.

We measured the ligand response curves of two representative coupled single-input gate devices and their single-gate counterparts (Supplementary Text 4.2 Figure 1). The coupled Inverter gate device (2xL2bulgeOff1) exhibits a response at slightly lower concentrations of input than the single Inverter gate (L2bulgeOff1), whereas the coupled Buffer gate device (2xL2bulge1) exhibits a response at slightly higher concentrations of input than the single Buffer gate (L2bulge1). However, the observed changes in the response curves are very slight, such that strong conclusions on the effects of gate coupling on the input-response curves cannot be made.

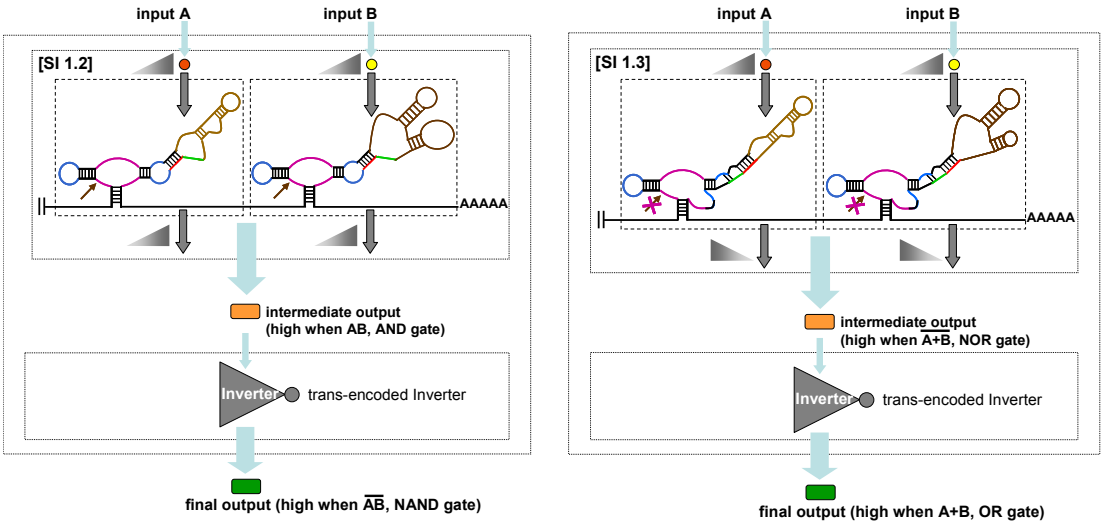


Supplementary Text 4.2 Figure 1. The device response over varying input concentrations of representative coupled gate devices (2xL2bulgeOff1, right; 2xL2bulge1, left) constructed

through SI 1.1 and their corresponding single-gate device counterparts (L2bulgeOff1, L2bulge1). The percent device response is plotted by normalizing corresponding dynamic switching ranges between the absence and presence of 10 mM theophylline to 0-100% as described in Materials and Methods.

Supplementary Text 4.3: *Layered architectures extend the information processing capabilities of SI 1*

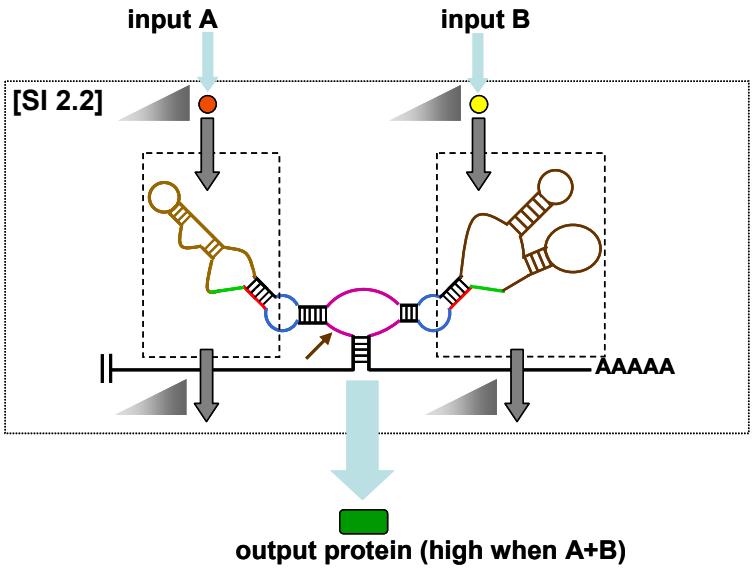
The first assembly scheme based on signal integration within the 3' UTR provides modular composition frameworks for two basic logic operators, AND and NOR. Additional logic operators may be desired, including NAND and OR gates. One way in which to directly obtain these logic operations from the assembled operations in SI 1 is to invert the output from the AND and NOR gate operators, respectively (Supplementary Text 4.3 Figure 1). For example, the resulting output of the AND and NOR gates could be an Inverter device such as a repressor protein[20] or an inhibitory noncoding RNA[44] that acts on a separately encoded gene product resulting in the desired NAND and OR operations, respectively. However, this proposed framework results in a layered architecture, which may have less desirable properties such as loss of signal and longer signal processing times. Alternative assembly strategies for obtaining additional logic operations that result in non-layered architectures are described in the manuscript.



Supplementary Text 4.3 Figure 1. Schematic representation of layered architectures that extend the information processing capabilities of SI 1. Left, schematic illustrating a NAND gate operation by inverting the output of an AND gate. Right, schematic illustrating an OR gate operation by inverting the output of a NOR gate.

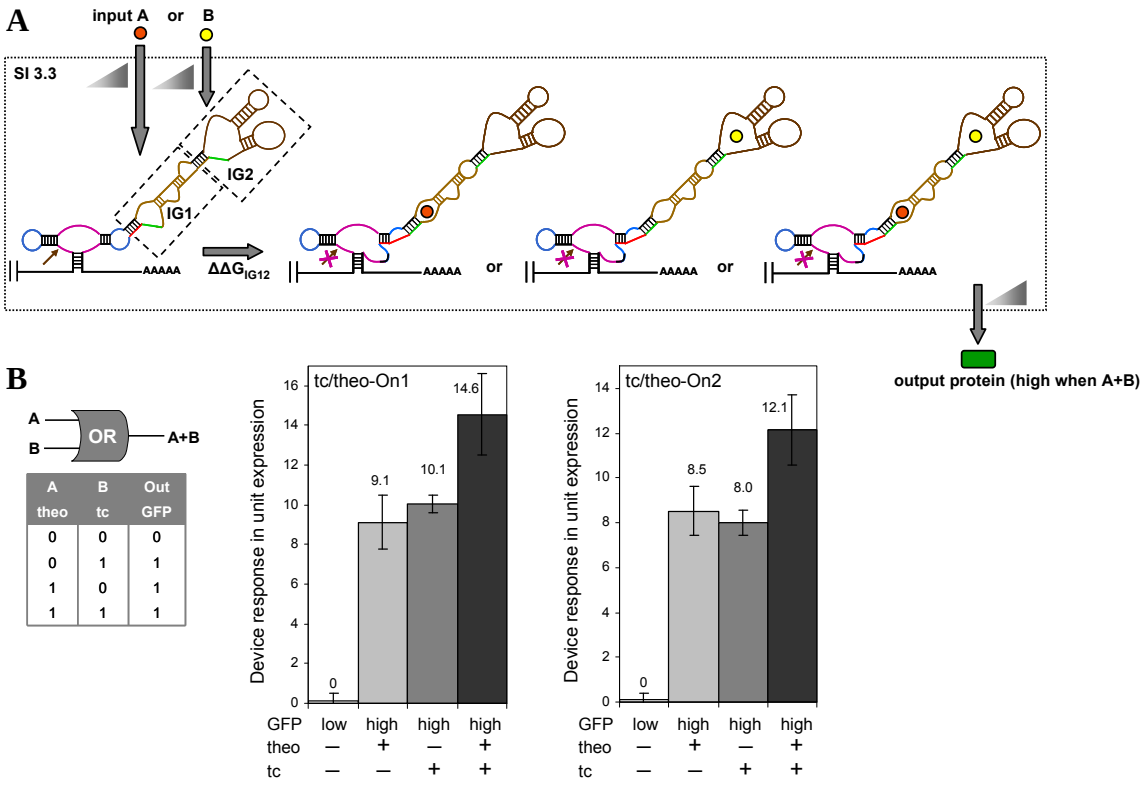
Supplementary Text 4.4: *Non-layered architectures (SI 2, SI 3) for an OR gate operation*

The second assembly scheme based on signal integration at the ribozyme core (SI 2) should be as flexible a composition framework as that specified for integration within the 3' UTR (SI 1). For example, SI 2 can be implemented to construct a higher-order RNA device capable of performing an OR gate operation by coupling internal Buffer gates responsive to different molecular inputs to stems I and II of the ribozyme (SI 2.2, Supplementary Text 4.4 Figure 1). While such a logic operation is theoretically possible, its construction is currently limited by the lack of one necessary component of this device - an internal Buffer gate coupled to stem I. Efforts are currently underway to generate such components. Therefore, SI 2 can provide logic operations that are not obtainable through SI 1 with non-layered architectures.



Supplementary Text 4.4 Figure 1. Schematic representation of an RNA device based on SI 2 that functions as an OR gate operator.

Alternatively, devices that perform an OR gate operation can be constructed through SI 3 (signal integration through a single ribozyme stem) by coupling a theophylline-responsive internal Buffer gate (IG1) and a tetracycline-responsive internal Inverter gate (IG2) at stem II (Supplementary Text 4.4 Figure 2A). The assembly scheme is similar to that used to construct devices that perform an AND gate operation, described in Figure 4.4A, except that the energetic requirements for switching between the conformational states are different. This RNA device (SI 3.3) assumes the conformation in which the binding pockets for both inputs are formed (Supplementary Text 4.4 Figure 2A) with a lower energetic requirement than an AND gate device ($\Delta\Delta G_{IG12}$ in SI 3.3 < $\Delta\Delta G_{IG2}$+ $\Delta\Delta G_{IG1}$ in SI 3.1), effectively allowing either input to bind to its corresponding sensor. The resulting device exhibits low output in the absence of both molecular inputs and high output in the presence of either input or both (Supplementary Text 4.4 Figure 2B). We constructed two OR gate devices, tc/theo-On1 and tc/theo-On2, based on different IG2 transmitter components.



Supplementary Text 4.4 Figure 2. OR gate devices. (A) Schematic representation of an RNA device that performs an OR gate operation by coupling internal Buffer (IG1) and Inverter (IG2) gates responsive to different input molecules to a single ribozyme stem. (B) The device response and truth table of OR gate operators (tc/theo-On1 and tc/theo-On2) based on SI 3.3. Device response under different input conditions (theo or tc (-), 0 mM; theo (+), 10 mM; tc (+), 0.25 mM) is reported as the output swing in units of expression relative to the absence of both inputs as described in Materials and Methods.

Supplementary Text 4.5: *Programming signal gain through multiple sensor-transmitter components*

Cooperativity in biological molecules is often a result of multiple binding sites that transit from a low-affinity state to a high-affinity state as more ligands occupy the available binding sites. In RNA devices comprised of two internal gates to the same input, although the sensor components exhibit similar input affinities (K_{apt}), their effective affinities are a combined effect of the sensor affinity (K_{apt}) and the energetic requirements for the device to switch between two states (K_{IG}), the latter of which can be programmed into the transmitter component ($\Delta\Delta G_{IG}$). Thus, the difference in free energies between states 1 and 3 ($\Delta\Delta G_{IG2}$+ $\Delta\Delta G_{IG1}$) represents an energetic contribution which lowers the effective binding affinity of IG1 to its input. The difference in free energies between states 1 and 2 ($\Delta\Delta G_{IG2}$) represents a lower energetic contribution to the effective binding affinity of IG2 to its input, such that the effective binding affinity of IG2 is higher than that of IG1. However, binding of input to IG2 lowers the energetic contribution to IG1 to the difference in free energies between states 2 and 3 ($\Delta\Delta G_{IG1}$), resulting in an increase in effective binding affinity as a result of input binding to IG2. The RNA device design is expected to result in a larger change in the regulatory response as input concentrations increase and IG1 transits from a lower affinity state to a higher affinity state. By programming the energetic differences between the

different conformational states ($\Delta\Delta G_{IG2}$ and $\Delta\Delta G_{IG1}$), we can program the degree of cooperativity (or level of signal gain) exhibited by the system (Supplementary Table 4.2).

Supplementary Text 4.6: *Device sequences*

The sequences of all devices used in this work are described below. Color schemes in the sequences correspond to those in the schematic device diagrams: purple, catalytic core of the ribozyme or actuator component; blue, loop regions of the actuator component; brown, aptamer or sensor component; green and red, competing and switching strands of the transmitter component, respectively; orange, communication modules of the transmitter component; italicized, spacer sequences; underlined, restriction sites.

Single-input gates

Single-input Buffer gates

L2bugle1

5’CCTAGG*AAACAAACAAA*GCTGTCACCGGATGTGCTTTCCGGTCTGATGAGTCCGT

GTCCATACCAGCATCGTCTTGATGCCCTTGGCAGGGACGGGACGAGGACGAAAC

AGC*AAAAAGAAAAATAAAAA*CTCGAG

L2bulge5

5’CCTAGG*AAACAAACAAA*GCTGTCACCGGATGTGCTTTCCGGTCTGATGAGTCCGT

GTCCAATACCAGCATCGTCTTGATGCCCTTGGCAGTGGACGGGACGAGGACGAA

ACAGC*AAAAAGAAAAATAAAAA*CTCGAG

L2bulge9

5’CCTAGG*AAACAAACAAA*GCTGTCACCGGATGTGCTTTCCGGTCTGATGAGTCCGT

TGTCCAATACCAGCATCGTCTTGATGCCCTTGGCAGTGGATGGGGACGGAGGAC

GAAACAGC*AAAAAGAAAAATAAAAA*CTCGAG

L2bulge1tc

5’CCTAGG*AAACAAACAAA*GCTGTCACCGGATGTGCTTTCCGGTCTGATGAGTCCGT

GTCCAAAACATACCAGATTTCGATCTGGAGAGGTGAAGAATTCGACCACCTGGA

CGGGACGAGGACGAAACAGC*AAAAAGAAAAATAAAAA*CTCGAG

Single-input Inverter gates

L2bulgeOff1

5’CCTAGG*AAACAAACAAA*GCTGTCACCGGATGTGCTTTCCGGTCTGATGAGTCCGT

GTTGCTGATACCAGCATCGTCTTGATGCCCTTGGCAGCAGTGGACGAGGACGAA

ACAGC*AAAAAGAAAAATAAAAA*CTCGAG

L2bulgeOff1tc

5’CCTAGG*AAACAAACAAA*GCTGTCACCGGATGTGCTTTCCGGTCTGATGAGTCCGT

TGTTGAGGAAAACATACCAGATTTCGATCTGGAGAGGTGAAGAATTCGACCACC

TCCTTATGGGAGGACGAAACAGC*AAAAAGAAAAATAAAAA*CTCGAG

L2bulgeOff2tc

5’CCTAGG*AAACAAACAAA*GCTGTCACCGGATGTGCTTTCCGGTCTGATGAGTCCGT

ATGAGGAAAACATACCAGATTTCGATCTGGAGAGGTGAAGAATTCGACCACCTC

CTTAGAGGAGGACGAAACAGC*AAAAAGAAAAATAAAAA*CTCGAG

L2bulgeOff3tc

5’CCTAGG*AAACAAACAAA*GCTGTCACCGGATGTGCTTTCCGGTCTGATGAGTCCGT

TGATGAGGAAAACATACCAGATTTCGATCTGGAGAGGTGAAGAATTCGACCACC

TCCTTAGAGGAGGACGAAACAGC*AAAAAGAAAAATAAAAA*CTCGAG

L2cm4

5’CCTAGG*AAACAAACAAA*GCTGTCACCGGATGTGCTTTCCGGTCTGATGAGTCCGT

CCTGGATACCAGCATCGTCTTGATGCCCTTGGCAGTCATAGAGGACGAAACAGC

*AAAAAGAAAAATAAAAA*CTCGAG

L1cm10

5’CCTAGG*AAACAAACAAA*GCTGTCACCGGATGTAAATGATACCAGCATCGTCTTG

ATGCCCTTGGCAGCTGCGCTTTCCGGTCTGATGAGTCCGTGAGGACGAAACAGC*A*

*AAAAGAAAAATAAAAA*CTCGAG

Higher-order devices (SI 1: signal integration within the 3’ UTR)

Two coupled Buffer or Inverter gates responsive to the same input

2xL2bulge1

5’CCTAGG*AAACAAACAAA*GCTGTCACCGGATGTGCTTTCCGGTCTGATGAGTCCGT

GTCCATACCAGCATCGTCTTGATGCCCTTGGCAGGGACGGGACGAGGACGAAAC

AGCAAAAAGAAAAATAAAAACTCGAGAAACAAACAAAGCTGTCACCGGATGTG

CTTTCCGGTCTGATGAGTCCGTGTCCATACCAGCATCGTCTTGATGCCCTTGGCA

GGGACGGGACGAGGACGAAACAGC*AAAAAGAAAAATAAAAA*CTCGAG

2xL2bulgeOff1

5’CCTAGG*AAACAAACAAA*GCTGTCACCGGATGTGCTTTCCGGTCTGATGAGTCCGT

GTTGCTGATACCAGCATCGTCTTGATGCCCTTGGCAGCAGTGGACGAGGACGAA

ACAGCAAAAAGAAAAATAAAAACTCGAGAAACAAACAAAGCTGTCACCGGATG

TGCTTTCCGGTCTGATGAGTCCGTGTTGCTGATACCAGCATCGTCTTGATGCCCTT

GGCAGCAGTGGACGAGGACGAAACAGC*AAAAAGAAAAATAAAAA*CTCGAG

Note: The sequence assembly of other RNA devices based on SI 1 (2xL2bulge1tc, 2xL2cm4, (L2bulge1+L2bulge9), and (L2bulgeOff1+L2cm4)), and the bandpass filter operator (L2bulge1+L2bulgeOff1) is identical to that of 2xL2bulge1 or 2xL2bulgeOff1, illustrated above as example templates. Sequences of single-input gates are shown above.

AND gates

AND1 (L2bulge1+L2bulge1tc)

5’CCTAGG*AAACAAACAAA*GCTGTCACCGGATGTGCTTTCCGGTCTGATGAGTCCGT

GTCCATACCAGCATCGTCTTGATGCCCTTGGCAGGGACGGGACGAGGACGAAAC

AGCAAAAAGAAAAATAAAAACTCGAGAAACAAACAAAGCTGTCACCGGATGTG

CTTTCCGGTCTGATGAGTCCGTGTCCAAAACATACCAGATTTCGATCTGGAGAGG

TGAAGAATTCGACCACCTGGACGGGACGAGGACGAAACAGC*AAAAAGAAAAATAA*

*AAA*CTCGAG

AND2 (L2bulge9+L2bulge1tc)

5’CCTAGG*AAACAAACAAA*GCTGTCACCGGATGTGCTTTCCGGTCTGATGAGTCCGT

TGTCCAATACCAGCATCGTCTTGATGCCCTTGGCAGTGGATGGGGACGGAGGAC

GAAACAGCAAAAAGAAAAATAAAAACTCGAGAAACAAACAAAGCTGTCACCGG

ATGTGCTTTCCGGTCTGATGAGTCCGTGTCCAAAACATACCAGATTTCGATCTGG

AGAGGTGAAGAATTCGACCACCTGGACGGGACGAGGACGAAACAGC*AAAAAGAA*

*AAATAAAAA*CTCGAG

NOR gates

NOR1 (L2bulgeOff1+L2bulgeOff1tc)

5’CCTAGG*AAACAAACAAA*GCTGTCACCGGATGTGCTTTCCGGTCTGATGAGTCCGT

GTTGCTGATACCAGCATCGTCTTGATGCCCTTGGCAGCAGTGGACGAGGACGAA

ACAGCAAAAAGAAAAATAAAAACTCGAGAAACAAACAAAGCTGTCACCGGATG

TGCTTTCCGGTCTGATGAGTCCGTTGTTGAGGAAAACATACCAGATTTCGATCTG

GAGAGGTGAAGAATTCGACCACCTCCTTATGGGAGGACGAAACAGC*AAAAAGAAA*

*AATAAAAA*CTCGAG

NOR2 (L2bulgeOff1+L2bulgeOff2tc)

5’CCTAGG*AAACAAACAAA*GCTGTCACCGGATGTGCTTTCCGGTCTGATGAGTCCGT

GTTGCTGATACCAGCATCGTCTTGATGCCCTTGGCAGCAGTGGACGAGGACGAA

ACAGCAAAAAGAAAAATAAAAACTCGAGAAACAAACAAAGCTGTCACCGGATG

TGCTTTCCGGTCTGATGAGTCCGTATGAGGAAAACATACCAGATTTCGATCTGGA

GAGGTGAAGAATTCGACCACCTCCTTAGAGGAGGACGAAACAGC*AAAAAGAAAAA*

*TAAAAA*CTCGAG

Higher-order devices (SI 2: signal integration at the ribozyme core through two stems)

NAND gates

NAND1 (L1cm10-L2bulgeOff1tc)

5’CCTAGG*AAACAAACAAA*GCTGTCACCGGATGTAAATGATACCAGCATCGTCTTG

ATGCCCTTGGCAGCTGCGCTTTCCGGTCTGATGAGTCCGTTGTTGAGGAAAACAT

ACCAGATTTCGATCTGGAGAGGTGAAGAATTCGACCACCTCCTTATGGGAGGAC

GAAACAGC*AAAAAGAAAAATAAAAA*CTCGAG

NAND2 (L1cm10-L2bulgeOff3tc)

5’CCTAGG*AAACAAACAAA*GCTGTCACCGGATGTAAATGATACCAGCATCGTCTTG

ATGCCCTTGGCAGCTGCGCTTTCCGGTCTGATGAGTCCGTTGATGAGGAAAACAT

ACCAGATTTCGATCTGGAGAGGTGAAGAATTCGACCACCTCCTTAGAGGAGGAC

GAAACAGC*AAAAAGAAAAATAAAAA*CTCGAG

Higher-order devices (SI 3: signal integration at a single ribozyme stem)

AND gates

AND1 (tc-theo-On1)

5’CCTAGG*AAACAAACAAA*GCTGTCACCGGATGTGCTTTCCGGTCTGATGAGTCCGT

GTCCATACCAGCATCGCTCAAAACATACCAGATTTCGATCTGGAGAGGTGAAGA

ATTCGACCACCTGAGTCTTGATGCCCTTGGCAGGGACGGGACGAGGACGAAACA

GC*AAAAAGAAAAATAAAAA*CTCGAG

AND2 (tc-theo-On2)

5’CCTAGG*AAACAAACAAA*GCTGTCACCGGATGTGCTTTCCGGTCTGATGAGTCCGT

GTCCATACCAGCATCGCTAAAACATACCAGATTTCGATCTGGAGAGGTGAAGAA

TTCGACCACCTAGTCTTGATGCCCTTGGCAGGGACGGGACGAGGACGAAACAGC

*AAAAAGAAAAATAAAAA*CTCGAG

AND3 (tc-theo-On3)

5’CCTAGG*AAACAAACAAA*GCTGTCACCGGATGTGCTTTCCGGTCTGATGAGTCCGT

GTCCATACCAGCATCGTGTAAAACATACCAGATTTCGATCTGGAGAGGTGAAGA

ATTCGACCACCTACATCTTGATGCCCTTGGCAGGGACGGGACGAGGACGAAACA

GC*AAAAAGAAAAATAAAAA*CTCGAG

OR gates

OR1 (tc/theo-On1)

5’CCTAGG*AAACAAACAAA*GCTGTCACCGGATGTGCTTTCCGGTCTGATGAGTCCGT

GTCCATACCAGCATCGGGCCTAAAACATACCAGATTTCGATCTGGAGAGGTGAA

GAATTCGACCACCTAGGTTTCTTGATGCCCTTGGCAGGGACGGGACGAGGACGA

AACAGC*AAAAAGAAAAATAAAAA*CTCGAG

OR2 (tc/theo-On2)

5’CCTAGG*AAACAAACAAA*GCTGTCACCGGATGTGCTTTCCGGTCTGATGAGTCCGT

GTCCATACCAGCATCGGTGGTAAAACATACCAGATTTCGATCTGGAGAGGTGAA

GAATTCGACCACCTACCATTCTTGATGCCCTTGGCAGGGACGGGACGAGGACGA

AACAGC*AAAAAGAAAAATAAAAA*CTCGAG

Two coupled internal gates responsive to the same input

theo-theo-On1

5’CCTAGG*AAACAAACAAA*GCTGTCACCGGATGTGCTTTCCGGTCTGATGAGTCCGT

GTCCATACCAGCATCGTTTATACCAGCATCGTCTTGATGCCCTTGGCAGAAATCT

TGATGCCCTTGGCAGGGACGGGACGAGGACGAAACAGC*AAAAAGAAAAATAAAAA*

CTCGAG

theo-theo-On2

5’CCTAGG*AAACAAACAAA*GCTGTCACCGGATGTGCTTTCCGGTCTGATGAGTCCGT

GTCCATACCAGCATCGTTGAATACCAGCATCGTCTTGATGCCCTTGGCAGTTGAT

CTTGATGCCCTTGGCAGGGACGGGACGAGGACGAAACAGC*AAAAAGAAAAATAAA*

*AA*CTCGAG

theo-theo-On3

5’CCTAGG*AAACAAACAAA*GCTGTCACCGGATGTGCTTTCCGGTCTGATGAGTCCGT

GTCCATACCAGCATCGATTGATACCAGCATCGTCTTGATGCCCTTGGCAGCAGTT

CTTGATGCCCTTGGCAGGGACGGGACGAGGACGAAACAGC*AAAAAGAAAAATAAA*

*AA*CTCGAG

theo-theo-On4

5’CCTAGG*AAACAAACAAA*GCTGTCACCGGATGTGCTTTCCGGTCTGATGAGTCCGT

GTCCATACCAGCATCGTATGATACCAGCATCGTCTTGATGCCCTTGGCAGCGTAT

CTTGATGCCCTTGGCAGGGACGGGACGAGGACGAAACAGC*AAAAAGAAAAATAAA*

*AA*CTCGAG

theo-theo-On5

5’CCTAGG*AAACAAACAAA*GCTGTCACCGGATGTGCTTTCCGGTCTGATGAGTCCGT

GTCCATACCAGCATCGATCATACCAGCATCGTCTTGATGCCCTTGGCAGGATTCT

TGATGCCCTTGGCAGGGACGGGACGAGGACGAAACAGC*AAAAAGAAAAATAAAAA*

CTCGAG

theo-theo-On6

5’CCTAGG*AAACAAACAAA*GCTGTCACCGGATGTGCTTTCCGGTCTGATGAGTCCGT

GTCCATACCAGCATCGATTGATACCAGCATCGTCTTGATGCCCTTGGCAGCAATT

CTTGATGCCCTTGGCAGGGACGGGACGAGGACGAAACAGC*AAAAAGAAAAATAAA*

*AA*CTCGAG

theo-theo-On7

5’CCTAGG*AAACAAACAAA*GCTGTCACCGGATGTGCTTTCCGGTCTGATGAGTCCGT

GTCCATACCAGCATCGGTAAATACCAGCATCGTCTTGATGCCCTTGGCAGTTGCT

CTTGATGCCCTTGGCAGGGACGGGACGAGGACGAAACAGC*AAAAAGAAAAATAAA*

*AA*CTCGAG

theo-theo-On8

5’CCTAGG*AAACAAACAAA*GCTGTCACCGGATGTGCTTTCCGGTCTGATGAGTCCGT

GTCCATACCAGCATCGTTGAATACCAGCATCGTCTTGATGCCCTTGGCAGTTGAT

CTTGATGCCCTTGGCAGGGACGGGACGAGGACGAAACAGC*AAAAAGAAAAATAAA*

*AA*CTCGAG

theo-theo-On9

5’CCTAGG*AAACAAACAAA*GCTGTCACCGGATGTGCTTTCCGGTCTGATGAGTCCGT

GTCCATACCAGCATCGGTTGAATACCAGCATCGTCTTGATGCCCTTGGCAGTTGA

TTCTTGATGCCCTTGGCAGGGACGGGACGAGGACGAAACAGC*AAAAAGAAAAATA*

*AAAA*CTCGAG

theo-theo-On10 (Cooperative Buffer gate)

5’CCTAGG*AAACAAACAAA*GCTGTCACCGGATGTGCTTTCCGGTCTGATGAGTCCGT

GTCCATACCAGCATCGGTTGAATACCAGCATCGTCTTGATGCCCTTGGCAGTTGA

CTCTTGATGCCCTTGGCAGGGATAGGACGAGGACGAAACAGC*AAAAAGAAAAATA*

*AAAA*CTCGAG

theo-theo-On11 (Cooperative Buffer gate)

5’CCTAGG*AAACAAACAAA*GCTGTCACCGGATGTGCTTTCCGGTCTGATGAGTCCGT

GTCCATACCAGCATCGGTTGAATACCAGCATCGTCTTGATGCCCTTGGCAGTTGA

TTCTTGATGCCCTTGGCAGGGATAGGACGAGGACGAAACAGC*AAAAAGAAAAATA*

*AAAA*CTCGAG

theo-theo-On12 (Cooperative Buffer gate)

5’CCTAGG*AAACAAACAAA*GCTGTCACCGGATGTGCTTTCCGGTCTGATGAGTCCGT

GTCCATACCAGCATCGATTGAATACCAGCATCGTCTTGATGCCCTTGGCAGTTGA

TTCTTGATGCCCTTGGCAGGGATAGGACGAGGACGAAACAGC*AAAAAGAAAAATA*

*AAAA*CTCGAG

theo-theo-On13 (Cooperative Buffer gate)

5’CCTAGG*AAACAAACAAA*GCTGTCACCGGATGTGCTTTCCGGTCTGATGAGTCCGT

GTCCATACCAGCATCGTGTTATACCAGCATCGTCTTGATGCCCTTGGCAGAATGT

CTTGATGCCCTTGGCAGGGATAGGACGAGGACGAAACAGC*AAAAAGAAAAATAAA*

*AA*CTCGAG

Two coupled internal Inverter gates responsive to the same input

theo-theo-Off1

5’CCTAGG*AAACAAACAAA*GCTGTCACCGGATGTGCTTTCCGGTCTGATGAGTCCGT

GTTATGATACCAGCATCGACATACCAGCATCGTCTTGATGCCCTTGGCAGGTTCT

TGATGCCCTTGGCAGCATGGACGAGGACGAAACAGC*AAAAAGAAAAATAAAAA*CTC

GAG

theo-theo-Off2

5’CCTAGG*AAACAAACAAA*GCTGTCACCGGATGTGCTTTCCGGTCTGATGAGTCCGT

GTTGCTGATACCAGCATCGACATACCAGCATCGTCTTGATGCCCTTGGCAGGTTC

TTGATGCCCTTGGCAGCAGTGGACGAGGACGAAACAGC*AAAAAGAAAAATAAAAA*C

TCGAG

theo-theo-Off3

5’CCTAGG*AAACAAACAAA*GCTGTCACCGGATGTGCTTTCCGGTCTGATGAGTCCGT

GTTATGATACCAGCATCGGACATACCAGCATCGTCTTGATGCCCTTGGCAGGTTT

CTTGATGCCCTTGGCAGCATGGACGAGGACGAAACAGC*AAAAAGAAAAATAAAAA*C

TCGAG

theo-theo-Off4

5’CCTAGG*AAACAAACAAA*GCTGTCACCGGATGTGCTTTCCGGTCTGATGAGTCCGT

GTGTCTGATACCAGCATCGACATACCAGCATCGTCTTGATGCCCTTGGCAGGTTC

TTGATGCCCTTGGCAGCAGGGACGAGGACGAAACAGC*AAAAAGAAAAATAAAAA*CT

CGAG

theo-theo-Off5

5’CCTAGG*AAACAAACAAA*GCTGTCACCGGATGTGCTTTCCGGTCTGATGAGTCCGT

GTGTCCTGATACCAGCATCGGACATACCAGCATCGTCTTGATGCCCTTGGCAGGT

TTCTTGATGCCCTTGGCAGCAGGGACGAGGACGAAACAGC*AAAAAGAAAAATAAAA*

*A*CTCGAG

theo-theo-Off6 (Cooperative Inverter gate)

5’CCTAGG*AAACAAACAAA*GCTGTCACCGGATGTGCTTTCCGGTCTGATGAGTCCGT

GTTATGATACCAGCATCGGCATACCAGCATCGTCTTGATGCCCTTGGCAGGTTCT

TGATGCCCTTGGCAGCATGGACGAGGACGAAACAGC*AAAAAGAAAAATAAAAA*CTC

GAG

theo-theo-Off7

5’CCTAGG*AAACAAACAAA*GCTGTCACCGGATGTGCTTTCCGGTCTGATGAGTCCGT

GTTGCTGATACCAGCATCGACATACCAGCATCGTCTTGATGCCCTTGGCAGGTTC

TTGATGCCCTTGGCAGCAGGGACGAGGACGAAACAGC*AAAAAGAAAAATAAAAA*CT

CGAG

theo-theo-Off8

5’CCTAGG*AAACAAACAAA*GCTGTCACCGGATGTGCTTTCCGGTCTGATGAGTCCGT

GTGTTTGATACCAGCATCGACATACCAGCATCGTCTTGATGCCCTTGGCAGGTTC

TTGATGCCCTTGGCAGCAAGGACGAGGACGAAACAGC*AAAAAGAAAAATAAAAA*CT

CGAG

Mutated coupled internal gates

theo-theo-On1M1

5’CCTAGG*AAACAAACAAA*GCTGTCACCGGATGTGCTTTCCGGTCTGATGAGTCCGT

GTCCAGACCAGCATCGTTTATACCAGCATCGTCTTGATGCCCTTGGCAGAAATCT

TGATGCCTATGGCAGGGACGGGACGAGGACGAAACAGC*AAAAAGAAAAATAAAAA*

CTCGAG

theo-theo-On1M2

5’CCTAGG*AAACAAACAAA*GCTGTCACCGGATGTGCTTTCCGGTCTGATGAGTCCGT

GTCCATACCAGCATCGTTTATACCAGCATCGTCTTGATGCCTATGGCAGAAATCT

TGATGCCCTTGGCAGGGACGGGACGAGGACGAAACAGC*AAAAAGAAAAATAAAAA*

CTCGAG

theo-theo-On13M1

5’CCTAGG*AAACAAACAAA*GCTGTCACCGGATGTGCTTTCCGGTCTGATGAGTCCGT

GTCCATACCAGCATCGTGTTATACCAGCATCGTCTTGATGCCCTTGGCAGAATGT

CTTGATGCCTATGGCAGGGATAGGACGAGGACGAAACAGC*AAAAAGAAAAATAAA*

*AA*CTCGAG

theo-theo-On13M2

5’CCTAGG*AAACAAACAAA*GCTGTCACCGGATGTGCTTTCCGGTCTGATGAGTCCGT

GTCCATACCAGCATCGTGTTAGACCAGCATCGTCTTGATGCCTATGGCAGAATGT

CTTGATGCCCTTGGCAGGGATAGGACGAGGACGAAACAGC*AAAAAGAAAAATAAA*

*AA*CTCGAG

theo-theo-Off2M1

5’CCTAGG*AAACAAACAAA*GCTGTCACCGGATGTGCTTTCCGGTCTGATGAGTCCGT

GTTGCTGAGACCAGCATCGACATACCAGCATCGTCTTGATGCCCTTGGCAGGTTC

TTGATGCCTATGGCAGCAGTGGACGAGGACGAAACAGC*AAAAAGAAAAATAAAAA*C

TCGAG

theo-theo-Off2M2

5’CCTAGG*AAACAAACAAA*GCTGTCACCGGATGTGCTTTCCGGTCTGATGAGTCCGT

GTTGCTGATACCAGCATCGACATACCAGCATCGTCTTGATGCCTATGGCAGGTTC

TTGATGCCCTTGGCAGCAGTGGACGAGGACGAAACAGC*AAAAAGAAAAATAAAAA*C

TCGAG

theo-theo-Off6M1

5’CCTAGG*AAACAAACAAA*GCTGTCACCGGATGTGCTTTCCGGTCTGATGAGTCCGT

GTTATGAACCCAGCATCGGCATACCAGCATCGTCTTGATGCCCTTGGCAGGTTCT

TGATGCCTATGGCAGCATGGACGAGGACGAAACAGC*AAAAAGAAAAATAAAAA*CTC

GAG

theo-theo-Off6M2

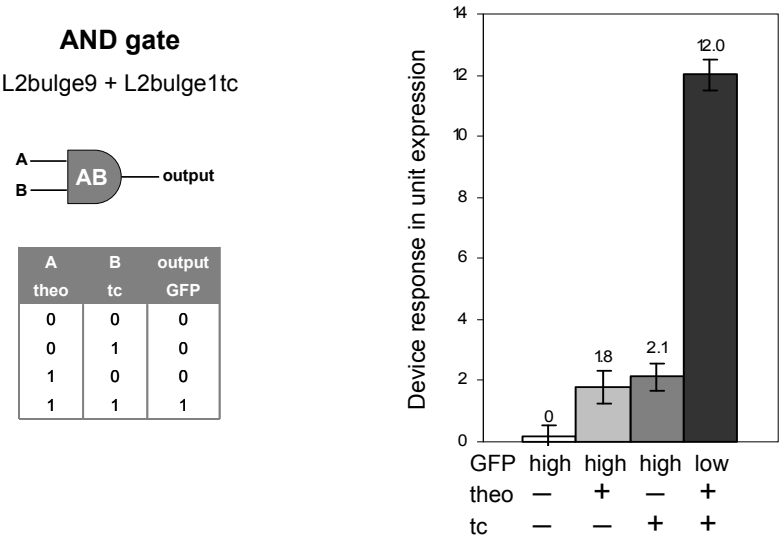
5’CCTAGG*AAACAAACAAA*GCTGTCACCGGATGTGCTTTCCGGTCTGATGAGTCCGT

GTTATGATACCAGCATCGGCATACCAGCATCGTCTTGATGCCTATGGCAGGTTCT

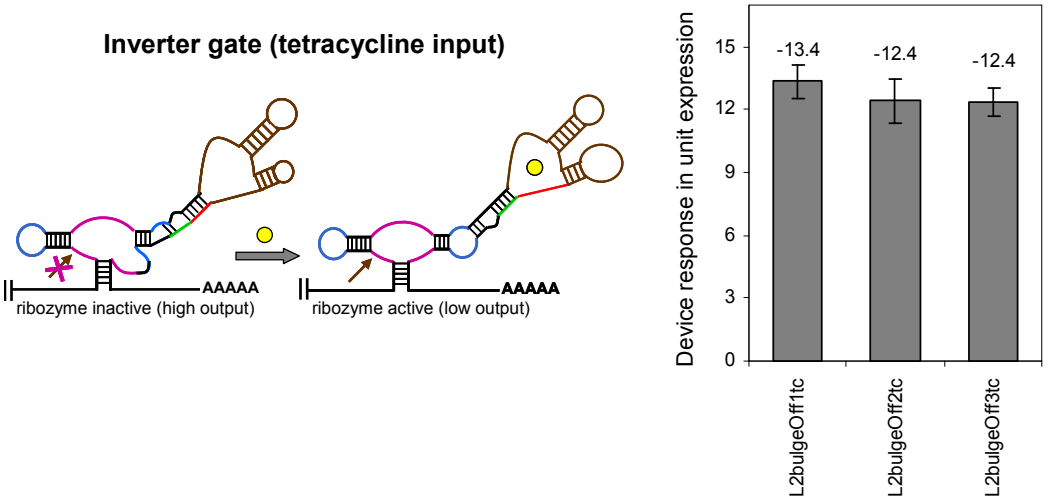
TGATGCCCTTGGCAGCATGGACGAGGACGAAACAGC*AAAAAGAAAAAT*AAAAACTC

GAG

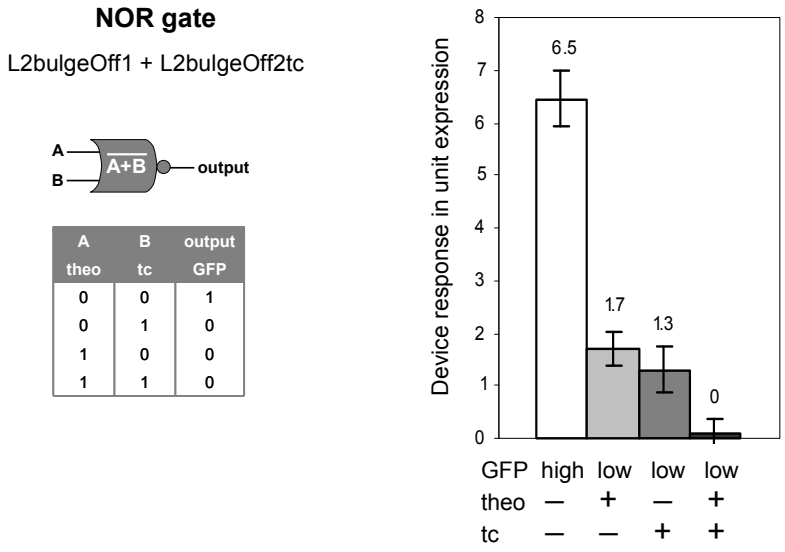
Supplementary Figures



Supplementary Figure 4.1. The device response and truth table of an AND gate operator (L2bulge9+L2bulge1tc) based on SI 1.2. The RNA device is constructed by coupling a theophylline-responsive Buffer gate (L2bulge9) and a tetracycline-responsive Buffer gate (L2bulge1tc) in the 3’ UTR of a target transcript. Device response under different input conditions (theo or tc (-), 0 mM; theo (+), 5 mM; tc (+), 0.5 mM) is reported as the output swing in units of expression relative to the absence of both inputs as described in Materials and Methods.

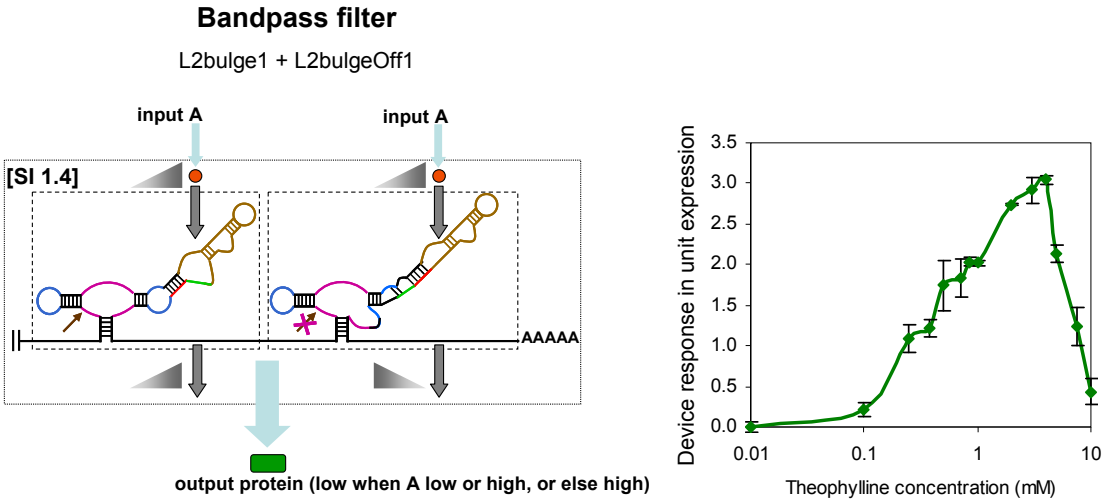


Supplementary Figure 4.2. Schematic representation and device response of tetracycline-responsive Inverter gates. Color schemes follow those described in Figure 4.1. Device response is reported as the output swing in units of expression as described in Materials and Methods. Output swings are reported from 0 mM to 0.5 mM tetracycline. The negative sign indicates the down-regulation of target gene expression by Inverter gates.

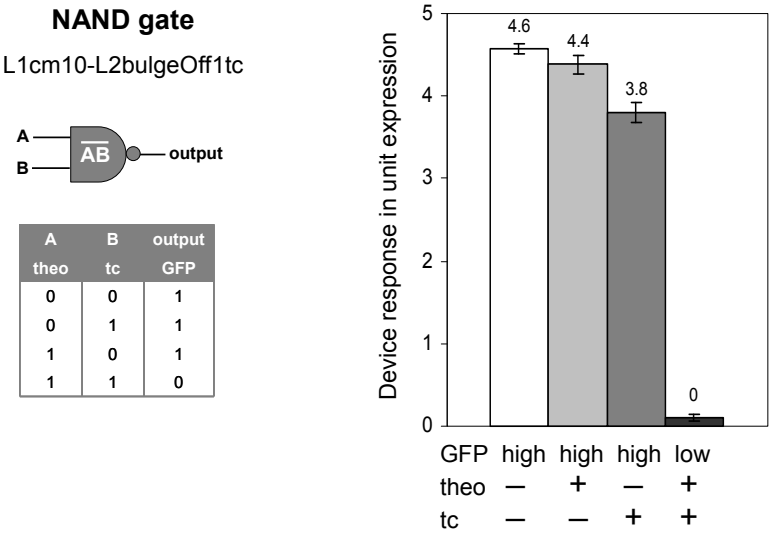


A	B	output
theo	tc	GFP
0	0	1
0	1	0
1	0	0
1	1	0

Supplementary Figure 4.3. The device response and truth table of a NOR gate operator (L2bulgeOff1+L2bulgeOff2tc) based on SI 1.3. The RNA device is constructed by coupling a theophylline-responsive Inverter gate (L2bulgeOff1) and a tetracycline-responsive Inverter gate (L2bulgeOff2tc) in the 3' UTR of a target transcript. Device response under different input conditions (theo or tc (-), 0 mM; theo (+), 10 mM; tc (+), 0.5 mM) is reported as the output swing in units of expression relative to the presence of both inputs as described in Materials and Methods.

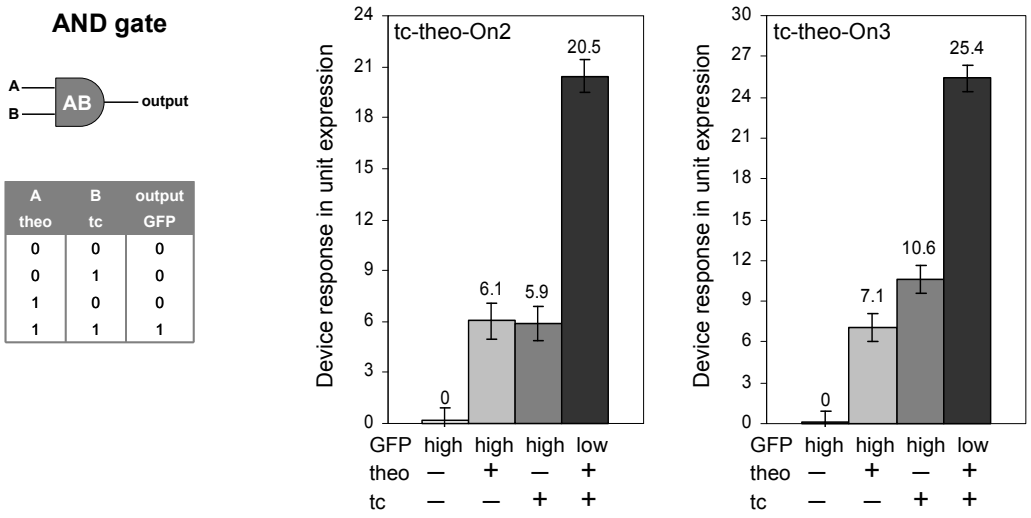


Supplementary Figure 4.4. Schematic representation and device response of a bandpass filter operator (L2bulge1+L2bulgeOff1) based on SI 1.4. Color schemes follow those described in Figure 4.1. Each gate is indicated in a boxed region, and triangles indicate relationships between associated gate inputs and outputs. The RNA device is constructed by coupling a theophylline-responsive Buffer gate (L2bulge1) and a theophylline-responsive Inverter gate (L2bulgeOff1) in the 3' UTR of a target transcript. Device response is reported as the output swing in units of expression as a function of theophylline concentration relative to the absence of theophylline as described in Materials and Methods.

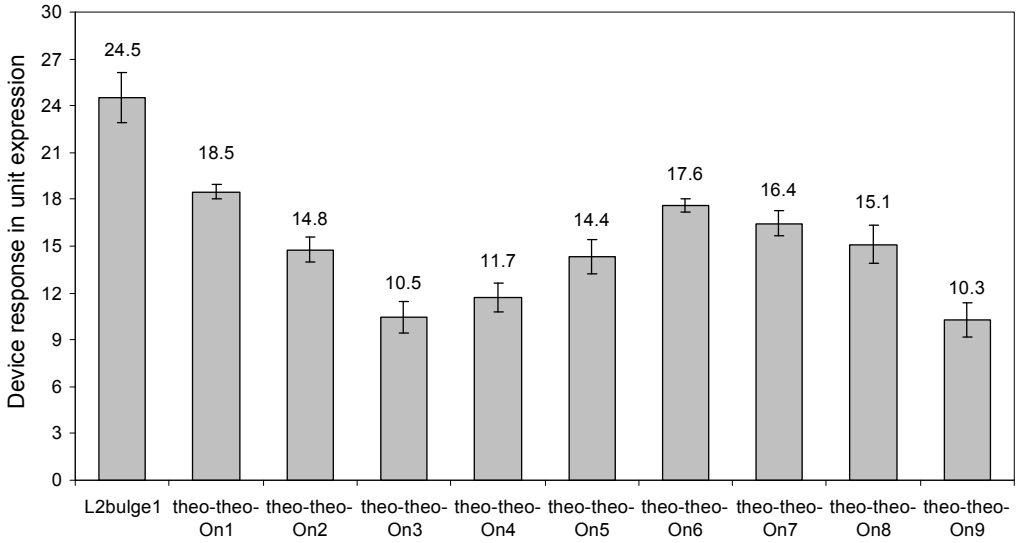


A theo	B tc	output GFP
0	0	1
0	1	1
1	0	1
1	1	0

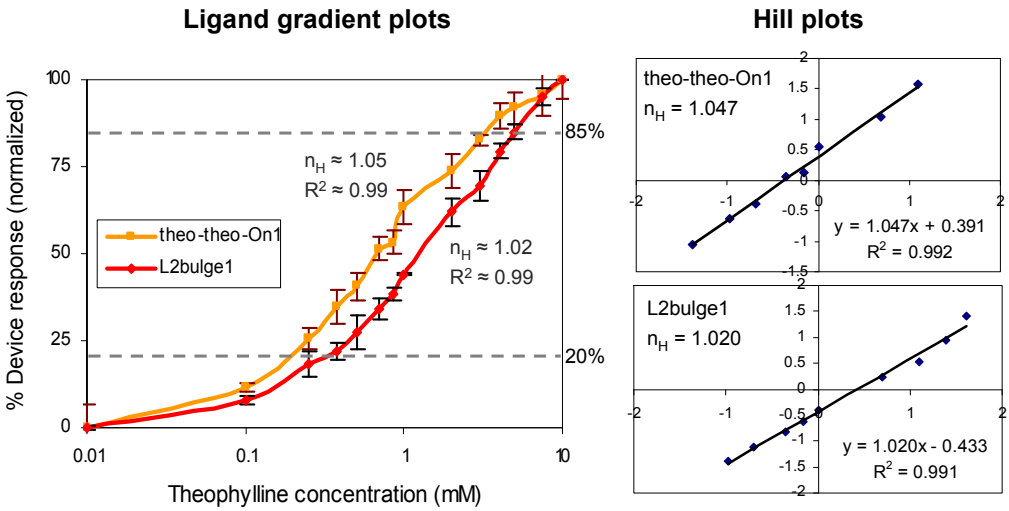
Supplementary Figure 4.5. The device response and truth table of a NAND gate operator (L1cm10-L2bulgeOff1tc) based on SI 2.1. The RNA device is constructed by coupling a theophylline-responsive internal Inverter gate (L1cm10) and a tetracycline-responsive internal Inverter gate (L2bulgeOff1tc) to stems I and II, respectively, of a ribozyme. Device response under different input conditions (theo or tc (-), 0 mM; theo (+), 10 mM; tc (+), 1 mM) is reported as the output swing in units of expression relative to the presence of both inputs as described in Materials and Methods.



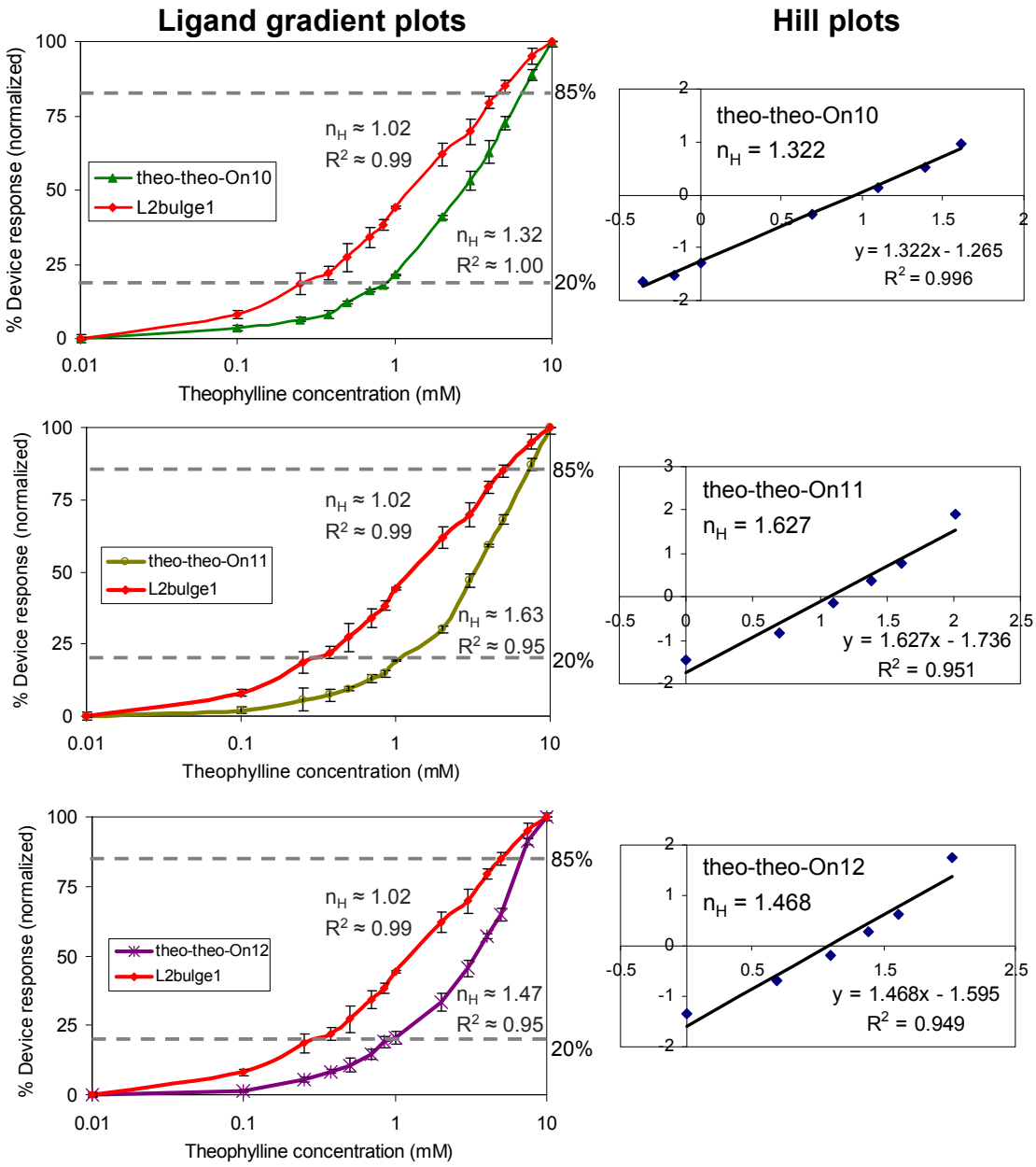
Supplementary Figure 4.6. The device response and truth table of AND gate operators (tc-theo-On2 and tc-theo-On3) based on SI 3.1. The RNA devices are constructed by coupling a theophylline-responsive internal Buffer gate (IG1) and a tetracycline-responsive internal Inverter gate (IG2) to stem II of a ribozyme. Device response under different input conditions (theo or tc (-), 0 mM; theo (+), 2.5 mM; tc (+), 0.5 mM) is reported as the output swing in units of expression relative to the absence of both inputs as described in Materials and Methods.



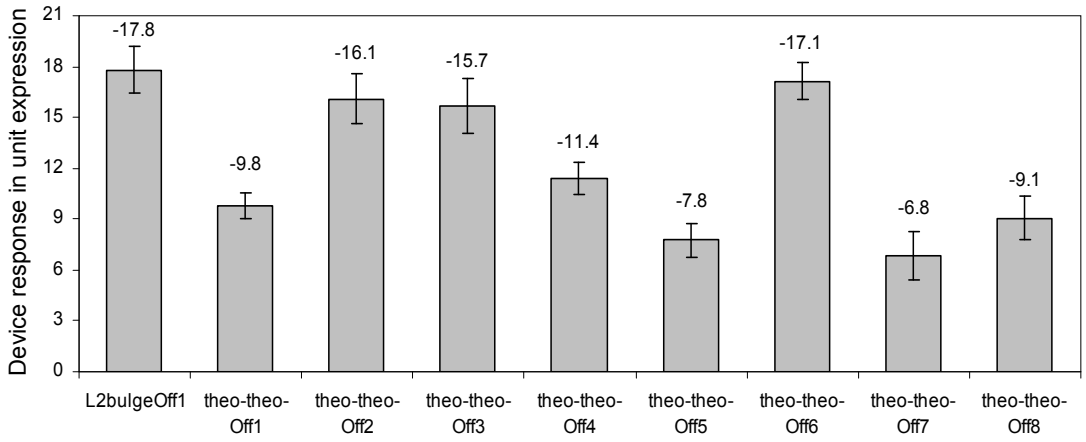
Supplementary Figure 4.7. The device response of RNA devices comprised of internal Buffer and Inverter gates and their single internal gate device counterpart (L2bulge1). The RNA devices are constructed by coupling theophylline-responsive internal Buffer (IG1) and Inverter (IG2) gates to stem II of a ribozyme. Device response is reported as the output swing in units of expression as described in Materials and Methods. Output swings are reported from 0 mM to 10 mM theophylline.



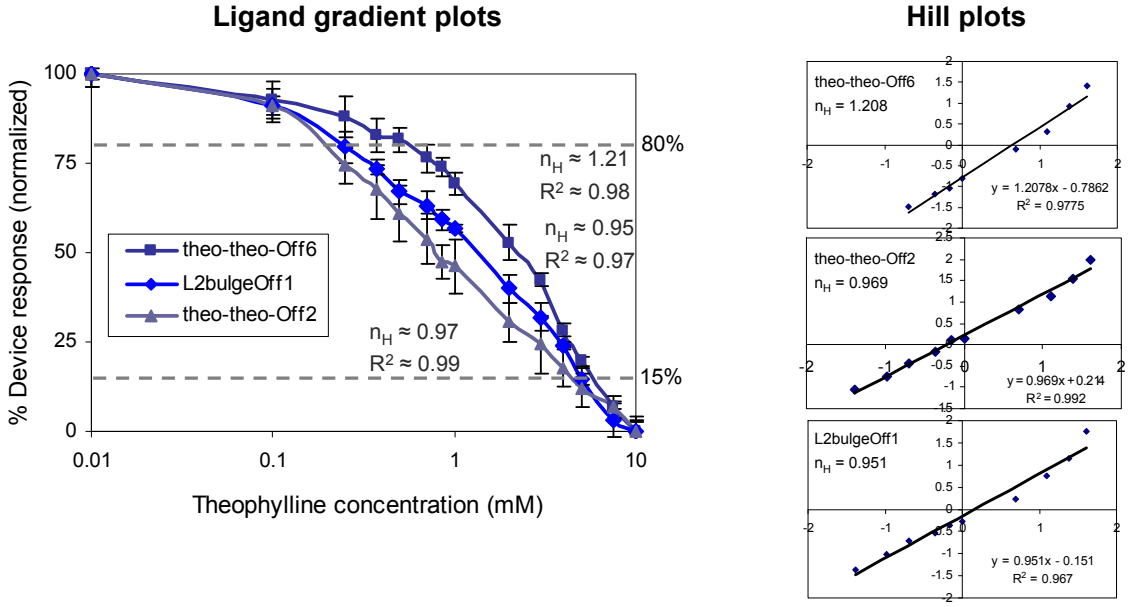
Supplementary Figure 4.8. The device response over varying input concentrations of a representative RNA device comprised of internal Buffer and Inverter gates (theo-theo-On1) and its single internal gate device counterpart (L2bulge1) demonstrates no signal gain ($n_H \approx 1$). The percent device response is plotted by normalizing corresponding dynamic switching ranges between the absence and presence of 10 mM theophylline to 0-100% as described in Materials and Methods. Corresponding Hill plots are constructed for 20-85% of each device switching range by plotting log [fraction expressed / (1 - fraction expressed)] against log [input concentration], where the slope represents the Hill coefficient (n_H).



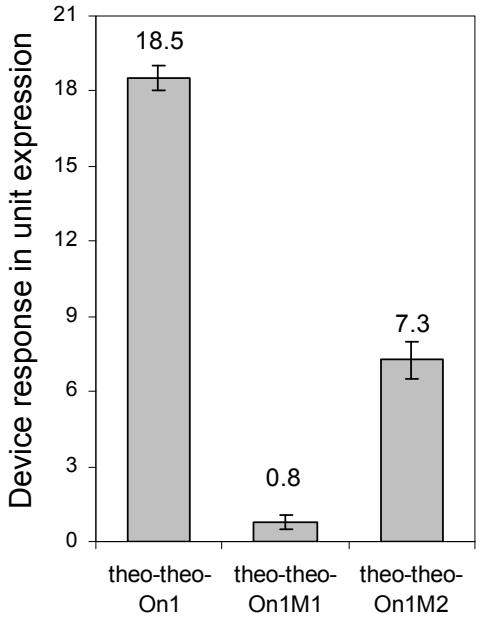
Supplementary Figure 4.9. The device response over varying input concentrations of RNA devices comprised of internal Buffer and Inverter gates (theo-theo-On10–12) and their single internal gate device counterpart (L2bulge1) demonstrates programmed cooperativity. The percent device response is plotted by normalizing corresponding dynamic switching ranges between the absence and presence of 10 mM theophylline to 0-100% as described in Materials and Methods. Corresponding Hill plots are constructed for 20-85% of each device switching range by plotting log [fraction expressed / (1 - fraction expressed)] against log [input concentration], where the slope represents the Hill coefficient (n_H).



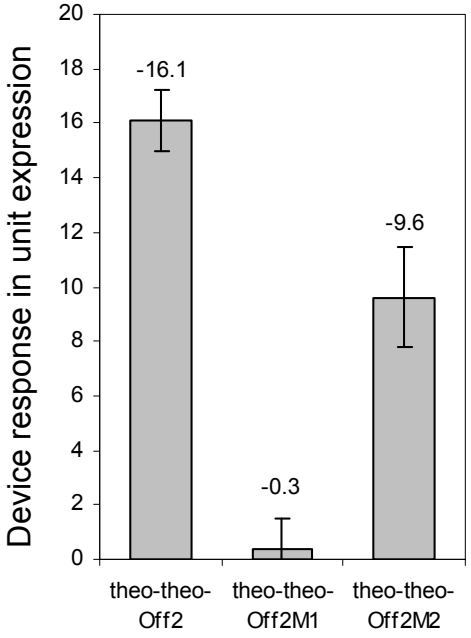
Supplementary Figure 4.10. The device response of RNA devices comprised of two internal Inverter gates and their single internal gate device counterpart (L2bulgeOff1). The RNA devices are constructed by coupling two theophylline-responsive internal Inverter gates (IG1, IG2) to stem II of a ribozyme. Device response is reported as the output swing in units of expression as described in Materials and Methods. Output swings are reported from 0 mM to 10 mM theophylline. The negative sign indicates the down-regulation of target gene expression.



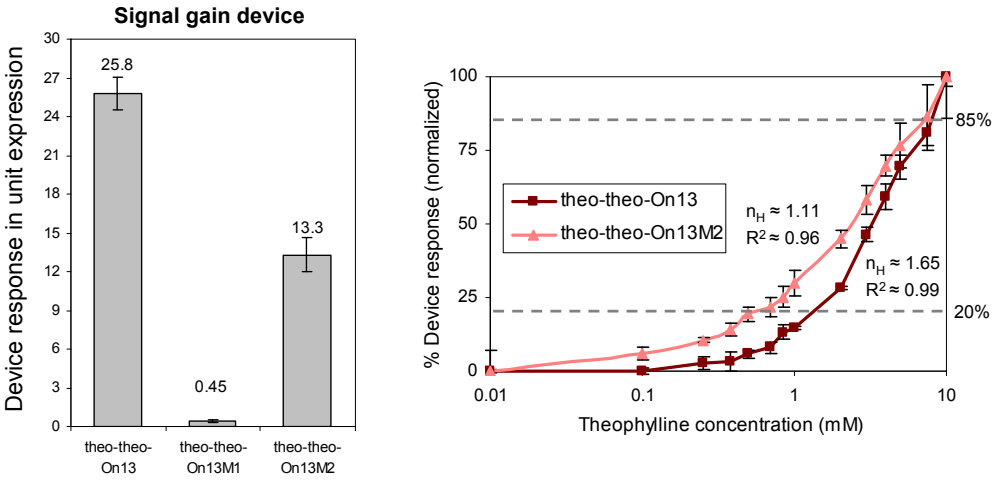
Supplementary Figure 4.11. The device response over varying theophylline concentrations of representative RNA devices comprised of two internal Inverter gates (theo-theo-Off2, $n_H \approx 1$; theo-theo-Off6, $n_H \approx 1.2$), and their single internal gate device counterpart (L2bulgeOff1, $n_H \approx 1$). The percent device response is plotted by normalizing corresponding dynamic switching ranges between the absence and presence of 10 mM theophylline to 0-100% as described in Materials and Methods. Corresponding Hill plots are constructed for 15-80% of each device switching range by plotting log [fraction repressed / (1 - fraction repressed)] against log [input concentration], where the slope represents the Hill coefficient (n_H).



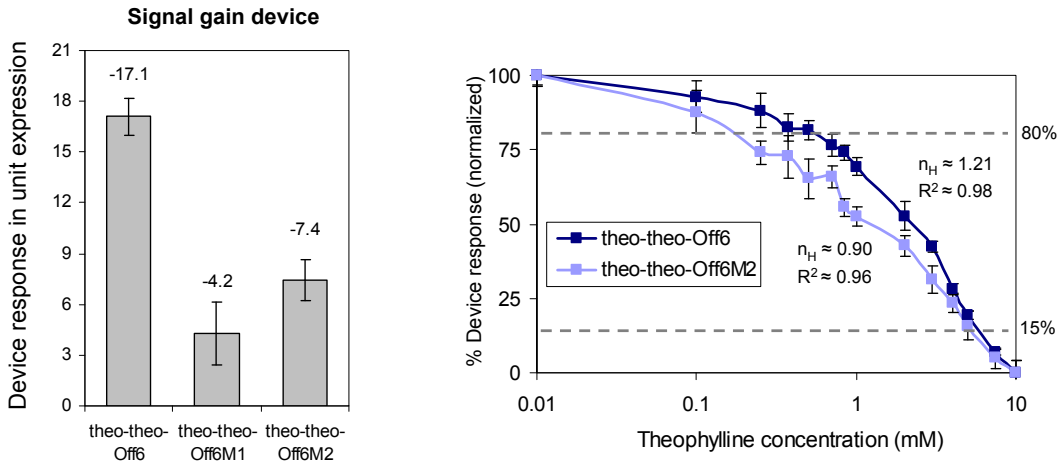
Supplementary Figure 4.12. The device response of a representative RNA device comprised of internal Buffer and Inverter gates (theo-theo-On1) and its mutated sensor variants demonstrates that input binding at both internal gates is responsible for the device response. Theo-theo-On1M1, mutation to the sensor in IG1; theo-theo-On1M2, mutation to sensor in IG2. Device response is reported as the output swing in units of expression as described in Materials and Methods. Output swings are reported from 0 mM to 10 mM theophylline. Individual mutations in both internal gates exhibited considerably lower output levels, supporting that both internal gates contribute to the overall device response. However, it was observed that theo-theo-On1M2 demonstrated less inhibition of device response compared to theo-theo-On1M1. The mutation of IG1 is anticipated to have a more significant impact on device performance as the device response is directly regulated by IG1.



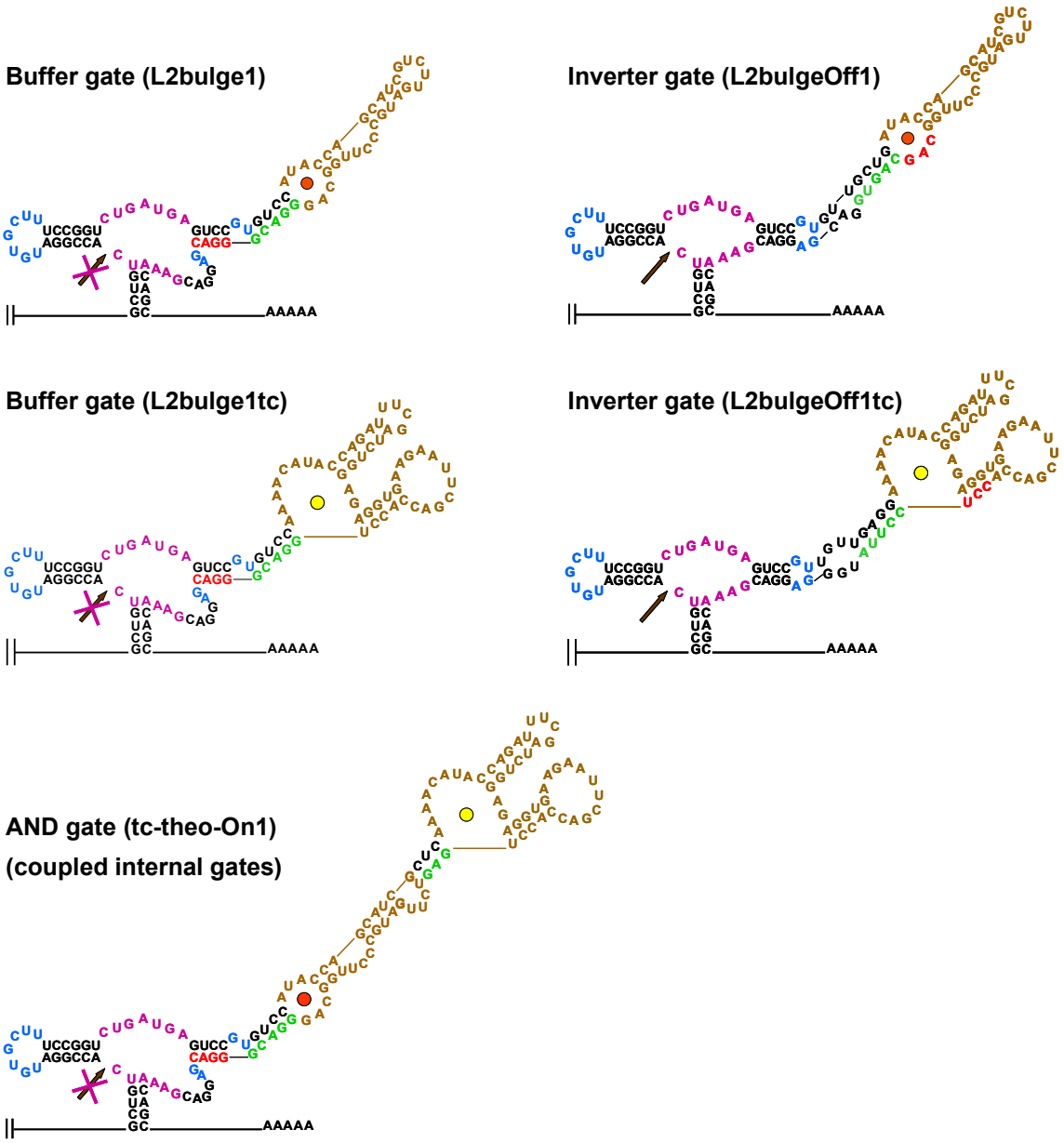
Supplementary Figure 4.13. The device response of a representative RNA device comprised of two internal Inverter gates (theo-theo-Off2) and its mutated sensor variants demonstrates that input binding at both internal gates is responsible for the device response. Theo-theo-Off2M1, mutation to the sensor in IG1; theo-theo-Off2M2, mutation to sensor in IG2. Device response is reported as the output swing in units of expression as described in Materials and Methods. Output swings are reported from 0 mM to 10 mM theophylline. The negative sign indicates the down-regulation of target gene expression. The mutation of IG1 is anticipated to have a more significant impact on device performance as the device response is directly regulated by IG1.



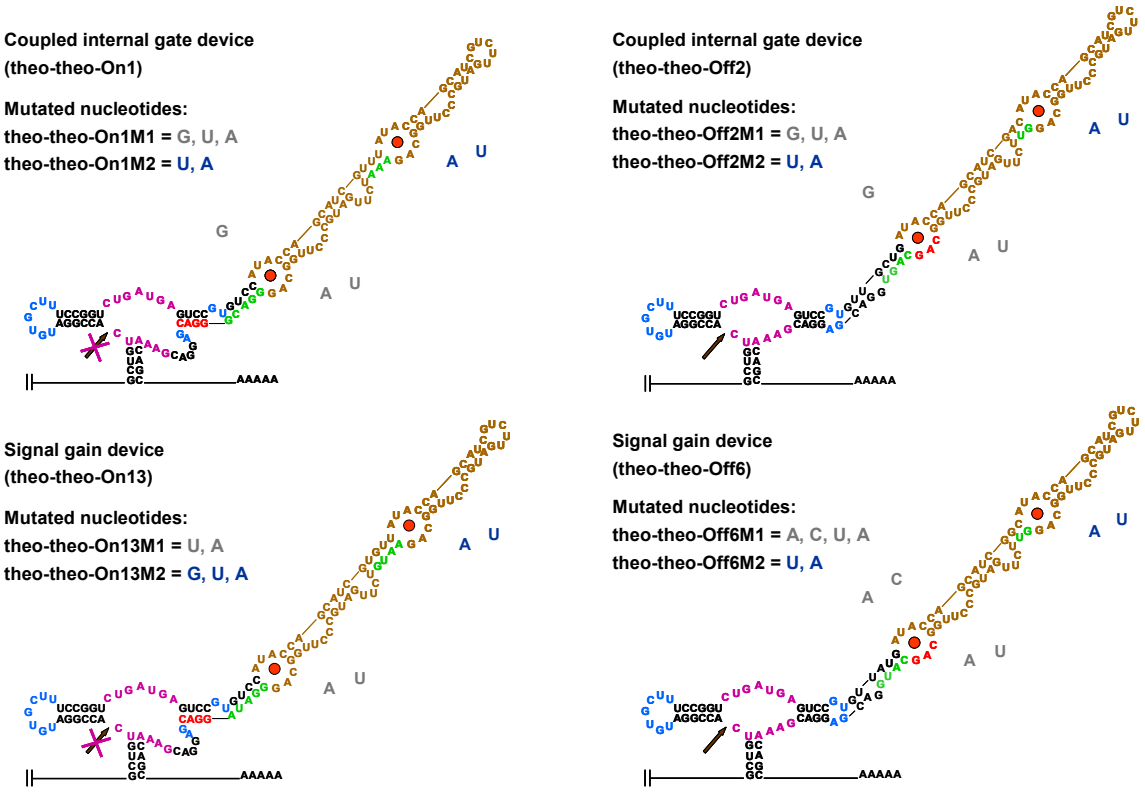
Supplementary Figure 4.14. The device response of a representative RNA device comprised of internal Buffer and Inverter gates that exhibits programmed cooperativity (theo-theo-On13) and its mutated sensor variants demonstrates that input binding at both internal gates is responsible for the device response. Theo-theo-On13M1, mutation to the sensor in IG1; theo-theo-On13M2, mutation to sensor in IG2. Device response is reported as the output swing in units of expression as described in Materials and Methods. Output swings are reported from 0 mM to 10 mM theophylline. The percent device response is plotted as described in Figure. 4.4E. The mutation of IG1 is anticipated to have a more significant impact on device performance as the device response is directly regulated by IG1.



Supplementary Figure 4.15. The device response of a representative RNA device comprised of two internal Inverter gates that exhibits programmed cooperativity (theo-theo-Off6) and its mutated sensor variants demonstrates that input binding at both internal gates is responsible for the device response. Theo-theo-Off6M1, mutation to the sensor in IG1; theo-theo-Off6M2, mutation to sensor in IG2. Device response is reported as the output swing in units of expression as described in Materials and Methods. Output swings are reported from 0 mM to 10 mM theophylline. The negative sign indicates the down-regulation of target gene expression. The percent device response is plotted as described in Materials and Methods. The mutation of IG1 is anticipated to have a more significant impact on device performance as the device response is directly regulated by IG1.



Supplementary Figure 4.16. Secondary structures and sequences of input-bound states of representative RNA devices. Single-input Buffer gates: L2bulge1, L2bulge1tc; single-input Inverter gates: L2bulgeOff1, L2bulgeOff1tc; RNA device comprised of internal Buffer (IG1) and Inverter (IG2) gates responsive to different inputs, illustrating points of coupling of two sensor-transmitter components: tc-theo-On1.



Supplementary Figure 4.17. Secondary structures and sequences of input-bound states of representative RNA device comprised of internal Buffer and Inverter gates responsive to the same input, illustrating points of coupling of two sensor-transmitter components. Nucleotides that were altered in the mutational studies are indicated for the sensors in IG1 and IG2. RNA devices that do not exhibit programmed cooperativity: theo-theo-On1, theo-theo-Off2; RNA devices that exhibit programmed cooperativity: theo-theo-On13, theo-theo-Off6.

Supplementary Table 4.1. The basal output signals and output swings of the RNA devices studied in this work are shown in % device response over the full transcriptional range of the employed promoter. The predicted basal output signals of coupled devices based on the appropriate single-gate response(s) and independent function are also reported. Predicted signals that do not match the measured output signals are indicated in italics.

Device	% Device response (over the full transcriptional range)				predicted basal signal for coupled devices
	theo, tc (-)	theo (+)	tc (+)	theo, tc (+)	
SI 1.1					
L2bulge1tc (Buffer)	0.37		0.96		
2xL2bulge1tc (Buffer)	0.16		0.46		0.14
L2bulge1 (Buffer)	0.40	0.89			
2xL2bulge1 (Buffer)	0.20	0.37			0.16
L2bugle8 (Buffer)	0.12	0.48			
2xL2bulge8 (Buffer)	0.07	0.19			0.01
L2bulge5 (Buffer)	0.82	1.00			
L2bulge1+L2bulge5 (Buffer)	0.25	0.43			0.33
L2bulgeOff1 (Inverter)	0.62	0.26			
2xL2bulgeOff1 (Inverter)	0.37	0.21			0.38
L2cm4 (Inverter)	0.78	0.41			
2xL2cm4 (Inverter)	0.32	0.20			*0.61*
L2bulgeOff1+L2cm4 (Inverter)	0.31	0.17			*0.48*
tc-responsive Inverter gates					
L2bulgeOff1tc (Inverter)	0.39		0.12		
L2bulgeOff2tc (Inverter)	0.42		0.17		
L2bulgeOff3tc (Inverter)	0.42		0.17		
SI 1.2 (AND gate)					
L2bulge1+L2bulge1tc	0.18	0.22	0.24	0.46	0.15
L2bulge9+L2bulge1tc	0.12	0.15	0.16	0.36	0.11
L2bulge9 (single-input Buffer)	0.30	0.72			
SI 1.3 (NOR gate)					
L2bulgeOff1+L2bulgeOff1tc	0.27	0.15	0.13	0.11	0.24
L2bulgeOff1+L2bulgeOff2tc	0.28	0.18	0.17	0.15	0.26
SI 2.1 (NAND gate)					
L1cm10+L2bulgeOff3tc	0.54	0.52	0.55	0.43	
L1cm10+L2bulgeOff1tc	0.51	0.51	0.50	0.42	
SI 3.1 (AND gate)					
tc-theo-On1	0.36	0.48	0.50	0.89	
tc-theo-On2	0.39	0.51	0.51	0.80	
tc-theo-On3	0.39	0.53	0.61	0.90	
SI 3.2 (dual sensor-transmitter)					
Buffer function					
theo-theo-On1	0.36	0.73			
theo-theo-On2	0.41	0.70			
theo-theo-On3	0.54	0.75			
theo-theo-On4	0.66	0.89			
theo-theo-On5	0.69	0.98			
theo-theo-On6	0.46	0.81			
theo-theo-On7	0.42	0.75			
theo-theo-On8	0.31	0.61			
theo-theo-On9	0.23	0.44			
theo-theo-On10 (cooperative)	0.16	0.54			
theo-theo-On11 (cooperative)	0.13	0.55			
theo-theo-On12 (cooperative)	0.12	0.60			
theo-theo-On13 (cooperative)	0.23	0.75			
Inverter function					
theo-theo-Off1	0.34	0.15			
theo-theo-Off2	0.60	0.27			
theo-theo-Off3	0.67	0.36			
theo-theo-Off4	0.47	0.24			
theo-theo-Off5	0.40	0.24			
theo-theo-Off6 (cooperative)	0.58	0.24			
theo-theo-Off7	0.54	0.40			
theo-theo-Off8	0.43	0.24			
SI 3.3 (OR gate)					
tc/theo-On1	0.48	0.65	0.64	0.72	
tc/theo-On2	0.42	0.60	0.62	0.71	

Supplementary Table 4.2. Free energy changes associated with RNA devices comprised of internal Buffer and Inverter gates and associated Hill coefficients. Free energy changes between RNA device states are predicted from a standard RNA folding program, RNAStructure 4.2.

Device	$\Delta\Delta G_{IG2}$ (kcal / mol)	$\Delta\Delta G_{IG1}$ (kcal / mol)	Degree of cooperativity or signal gain
Non-cooperative			
theo-theo-On1	0.3	0.3	none
theo-theo-On2	2.8	0.3	none
theo-theo-On3	1.8	0.3	none
theo-theo-On4	1.9	0.3	none
theo-theo-On5	0.0	0.3	none
theo-theo-On6	0.9	0.3	none
theo-theo-On7	3.0	0.3	none
theo-theo-On8	2.8	0.3	none
theo-theo-On9	2.9	0.0	none
Cooperative			
theo-theo-On10	0.3	1.0	$n_H \approx 1.32$
theo-theo-On11	1	1.0	$n_H \approx 1.63$
theo-theo-On12	1.4	1.0	$n_H \approx 1.47$
theo-theo-On13	2.2	1.0	$n_H \approx 1.65$

Acknowledgements

We thank A. Babiskin for providing pRzS; Y. Chen and J. Liang for constructive suggestions in data analysis and presentation; and D. Endy for critical reading of the manuscript. This work was supported by the Center for Biological Circuit Design at Caltech (fellowship to M.N.W.), the Arnold and Mabel Beckman Foundation (grant to C.D.S.), and the National Institutes of Health (grant to C.D.S.).

References

1. Endy, D. Foundations for engineering biology. Nature 438, 449-53 (2005).

2. Guet, C. C., Elowitz, M. B., Hsing, W. & Leibler, S. Combinatorial synthesis of genetic networks. Science 296, 1466-70 (2002).

3. Kramer, B. P., Fischer, C. & Fussenegger, M. BioLogic gates enable logical transcription control in mammalian cells. Biotechnol Bioeng 87, 478-84 (2004).

4. Cox, R. S., 3rd, Surette, M. G. & Elowitz, M. B. Programming gene expression with combinatorial promoters. Mol Syst Biol 3, 145 (2007).

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

6. Breaker, R. R. Natural and engineered nucleic acids as tools to explore biology. Nature 432, 838-45 (2004).

7. Seelig, G., Soloveichik, D., Zhang, D. Y. & Winfree, E. Enzyme-free nucleic acid logic circuits. Science 314, 1585-8 (2006).

8. Dirks, R. M. & Pierce, N. A. Triggered amplification by hybridization chain reaction. Proc Natl Acad Sci U S A 101, 15275-8 (2004).

9. Stojanovic, M. N., Mitchell, T. E. & Stefanovic, D. Deoxyribozyme-based logic gates. J Am Chem Soc 124, 3555-61 (2002).

10. Stojanovic, M. N. & Stefanovic, D. Deoxyribozyme-based half-adder. J Am Chem Soc 125, 6673-6 (2003).

11. Stojanovic, M. N. & Stefanovic, D. A deoxyribozyme-based molecular automaton. Nat Biotechnol 21, 1069-74 (2003).

12. Benenson, Y. et al. Programmable and autonomous computing machine made of biomolecules. Nature 414, 430-4 (2001).

13. Benenson, Y., Gil, B., Ben-Dor, U., Adar, R. & Shapiro, E. An autonomous molecular computer for logical control of gene expression. Nature 429, 423-9 (2004).

14. Breaker, R. R. Engineered allosteric ribozymes as biosensor components. Curr Opin Biotechnol 13, 31-9 (2002).

15. Jose, A. M., Soukup, G. A. & Breaker, R. R. Cooperative binding of effectors by an allosteric ribozyme. Nucleic Acids Res 29, 1631-7 (2001).

16. Penchovsky, R. & Breaker, R. R. Computational design and experimental validation of oligonucleotide-sensing allosteric ribozymes. Nat Biotechnol 23, 1424-33 (2005).

17. Robertson, M. P. & Ellington, A. D. In vitro selection of an allosteric ribozyme that transduces analytes to amplicons. Nat Biotechnol 17, 62-6 (1999).

18. Isaacs, F. J., Dwyer, D. J. & Collins, J. J. RNA synthetic biology. Nat Biotechnol 24, 545-54 (2006).

19. Suess, B. & Weigand, J. E. Engineered riboswitches - Overview, Problems and Trends. RNA Biol 5 (2008).

20. Rinaudo, K. et al. A universal RNAi-based logic evaluator that operates in mammalian cells. Nat Biotechnol 25, 795-801 (2007).

21. Brown, B. D. et al. Endogenous microRNA can be broadly exploited to regulate transgene expression according to tissue, lineage and differentiation state. Nat Biotechnol 25, 1457-67 (2007).

22. Mathews, D. H. & Turner, D. H. Prediction of RNA secondary structure by free energy minimization. Curr Opin Struct Biol 16, 270-8 (2006).

23. Parisien, M. & Major, F. The MC-Fold and MC-Sym pipeline infers RNA structure from sequence data. Nature 452, 51-5 (2008).

24. Win, M. N. & Smolke, C. D. From the Cover: A modular and extensible RNA-based gene-regulatory platform for engineering cellular function. Proc Natl Acad Sci U S A 104, 14283-8 (2007).

25. Khvorova, A., Lescoute, A., Westhof, E. & Jayasena, S. D. Sequence elements outside the hammerhead ribozyme catalytic core enable intracellular activity. Nat Struct Biol 10, 708-12 (2003).

26. Welz, R. & Breaker, R. R. Ligand binding and gene control characteristics of tandem riboswitches in Bacillus anthracis. Rna 13, 573-82 (2007).

27. Sudarsan, N. et al. Tandem riboswitch architectures exhibit complex gene control functions. Science 314, 300-4 (2006).

28. Nelson, D. L. & Cox, M. M. Lehninger Principles of Biochemistry (W. H. Freeman and Company, New York, 2005).

29. Mandal, M. et al. A glycine-dependent riboswitch that uses cooperative binding to control gene expression. Science 306, 275-9 (2004).

30. Woodside, M. T. et al. Nanomechanical measurements of the sequence-dependent folding landscapes of single nucleic acid hairpins. Proc Natl Acad Sci U S A 103, 6190-5 (2006).

31. Tuerk, C. & Gold, L. Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. Science 249, 505-10 (1990).

32. Ellington, A. D. & Szostak, J. W. In vitro selection of RNA molecules that bind specific ligands. Nature 346, 818-22 (1990).

33. Sambrook, J. & Russell, D. W. Molecular cloning: a laboratory manual (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 2001).

34. Mateus, C. & Avery, S. V. Destabilized green fluorescent protein for monitoring dynamic changes in yeast gene expression with flow cytometry. Yeast 16, 1313-23 (2000).

35. Gietz, R. & Woods, R. in Guide to Yeast Genetics and Molecular and Cell Biology, Part B. (eds. Guthrie, C. & Fink, G.) 87-96 (Academic Press, San Diego, 2002).

36. Yokobayashi, Y., Weiss, R. & Arnold, F. H. Directed evolution of a genetic circuit. Proc Natl Acad Sci U S A 99, 16587-91 (2002).

37. Basu, S., Mehreja, R., Thiberge, S., Chen, M. T. & Weiss, R. Spatiotemporal control of gene expression with pulse-generating networks. Proc Natl Acad Sci U S A 101, 6355-60 (2004).

38. Levine, E., Zhang, Z., Kuhlman, T. & Hwa, T. Quantitative characteristics of gene regulation by small RNA. PLoS Biol 5, e229 (2007).

39. Hebert, S. S. et al. Loss of microRNA cluster miR-29a/b-1 in sporadic Alzheimer's disease correlates with increased BACE1/beta-secretase expression. Proc Natl Acad Sci U S A 105, 6415-20 (2008).

40. Calin, G. A. et al. MiR-15a and miR-16-1 cluster functions in human leukemia. Proc Natl Acad Sci U S A 105, 5166-71 (2008).

41. Ventura, A. et al. Targeted deletion reveals essential and overlapping functions of the miR-17 through 92 family of miRNA clusters. Cell 132, 875-86 (2008).

42. Soukup, G. A., Emilsson, G. A. & Breaker, R. R. Altering molecular recognition of RNA aptamers by allosteric selection. J Mol Biol 298, 623-32 (2000).

43. Rodionov, D. A., Dubchak, I., Arkin, A., Alm, E. & Gelfand, M. S. Reconstruction of regulatory and metabolic pathways in metal-reducing delta-proteobacteria. Genome Biol 5, R90 (2004).

44. Deans, T. L., Cantor, C. R. & Collins, J. J. A tunable genetic switch based on RNAi and repressor proteins for regulating gene expression in mammalian cells. Cell 130, 363-72 (2007).

Chapter V: Engineering protein-responsive gene regulators and cellular biosensors

Abstract

We have previously developed small molecule-responsive RNA devices, called ribozyme switches[1], which are capable of regulating gene expression in response to target ligands and detecting the biosynthesis of target metabolites in a noninvasive manner in the cellular environment. The switch devices were constructed through a modular and portable platform comprised of distinct functional domains: the sensor, the actuator, and the transmitter. In this study, we explore the extensibility of the small molecule-responsive ribozyme-switch platform to respond to a different class of target ligands, proteins, by implementing protein-responsive RNA aptamers within the sensor domain of the device molecule. Preliminary data demonstrates that the resulting devices function as gene regulators/cellular biosensors in response to the presence of the target protein. The engineering of a protein-responsive ribozyme-switch platform may enable the construction of 'designer' gene regulators and cellular biosensors that can be used to respond to, report on, and manipulate the expression levels of a specific target protein of interest, such as those associated with a particular diseased state.

5.1. Introduction

Proteins are fundamental cellular constituents that exhibit a wide variety of functional activities and are encoded within the genomes of organisms. Numerous diseases originate from faulty and/or improperly-regulated protein expression. Genetically-encoded molecular tools are needed for controlled regulation of the expression of targeted genes and probing of corresponding protein expression. Novel gene regulators and cellular biosensors that provide noninvasive dynamic detection, control, and manipulation of target protein expression and activity are highly desired and represent powerful tools for basic and applied biological research.

Many synthetic biosensors have been developed to probe various target proteins. Antibodies are the most popular class of biosensors with a broad range of remarkable target ligand recognition capabilities[2] and have made profound contributions to the elucidation of protein interactions and functions. However, applications of antibodies as biosensors have primarily been targeted to *in vitro* studies[3, 4] or invasive studies in cellular systems[5]. In addition, the employment of antibodies in biosensing applications presents issues associated with structural stability, difficult manipulation, and use of animals for generating new antibodies[6, 7]. As a result, the development of antibody-based biosensors can be labor-intensive, time-consuming, and expensive.

Alternatively, biosensors have been developed based on fluorescent protein fusion and fluorescence resonance energy transfer (FRET) for *in vivo* studies of protein interactions and functions. In the former, proteins of interest are fused with a fluorescent protein and their expression levels and subcellular localization can be visualized using a fluorescent microscope[8]. In the latter, the synthetic genetically-encoded biosensors are composed of the

target proteins fused with appropriate donor-acceptor fluorescent protein pairs suitable for FRET[9, 10]. FRET-based biosensors have become more powerful following the development of fluorescent proteins with enhanced signal response and sensitivity[11]. Although fluorescent protein-fused biosensors are valuable tools in monitoring protein-associated cellular events in living systems, these sensors pose a couple of major challenges[9, 10]. First, many proteins are not amenable to protein fusion strategies, which may perturb the native folding state of the target protein and thus its functional activity, or the intensity of the fluorescent reporter protein. Second, there are a limited number of good FRET pairs, and many classes of protein molecules are incompatible or cannot be properly coupled with any FRET pair based on conformations and distances required for generating detectable FRET signals. Therefore, despite the tremendous value held by fluorescent protein-fused biosensors, their applicability is fairly limited to certain proteins, thus lacking platform universality and potential for broad applications. In addition, while antibody-based or fluorescent protein fusion-based biosensors are capable of probing target proteins, they are limited in application as gene regulators.

We have previously developed modular synthetic RNA devices, called ribozyme switches, based on an extensible gene-regulatory/biosensor platform that exploits RNA aptamers as sensing elements for small-molecule ligands[1]. Aptamers are nucleic acid-based molecular sensing elements that can bind ligands with high affinity and specificity[12, 13] and are well suited as sensing components of biosensors[7]. Other than antibodies, aptamers are the only molecular species that exhibit 'universal' binding activities for a diverse range of ligands[6]. The ribozyme switch platform was demonstrated to be modular such that the sensing component of the biosensor was directly replaced with a different small molecule-binding RNA aptamer, where the resulting biosensor was shown to exhibit ligand

responsiveness to this new small-molecule ligand[1]. The ability of the direct sensor replacement strategy to generate functional switches highlights the platform adaptability to construct new biosensors for various target ligands. In addition to biosensing, ribozyme switches can also function as gene regulators in response to ligand binding. However, the switch platform has not been adapted to probe proteins. Recently, a similar ribozyme-based system was demonstrated *in vitro*, in which the ribozyme cleavage activity was regulated by a protein kinase, ERK2, through molecular recognition of the kinase by its RNA aptamer coupled to the ribozyme[14]. This system, however, does not support portability to the cellular environment, as the sequence elements required for the *in vivo* functional activity of the ribozyme are absent in the design[15].

In this study, we set out to develop an *in vivo*-functional platform for the construction of RNA devices that enable both biosensing of specific protein molecules of interest and regulation of target gene expression in response to those protein molecules in living cells. We extend our small molecule-responsive ribozyme switch platform by implementing protein-binding RNA aptamers within the sensor domain of the switch molecule to generate protein-responsive ribozyme switches. Preliminary data show that the resulting switch devices function as gene regulators/cellular biosensors in response to the presence of a target protein. Additional experiments as described in Discussion will be conducted to further support and confirm the functional activity of these devices currently observed. Successful construction and demonstration of modular protein-responsive gene regulator/biosensor platform will further advance the elucidation of protein interactions and functions. In addition, such tools are essential for the construction of synthetic biological systems, which hold promise in furthering current understanding of natural biological systems.

5.2. Results

5.2.1. General composition framework and construction scheme for protein-responsive ribozyme switches

The ribozyme switch platform is comprised of three distinct functional components: a sensor component, comprised of an RNA aptamer; an actuator component, comprised of a hammerhead ribozyme including the sequence elements required for its *in vivo* cleavage activity[15]; and a transmitter component, comprised of a sequence that couples the sensor and actuator components and translates the binding event in the sensor to a conformation change in the actuator through a competitive hybridization event called strand displacement[1] (Figure 5.1). The sensor, actuator, and transmitter components are modularly coupled and functionally independent of one another.

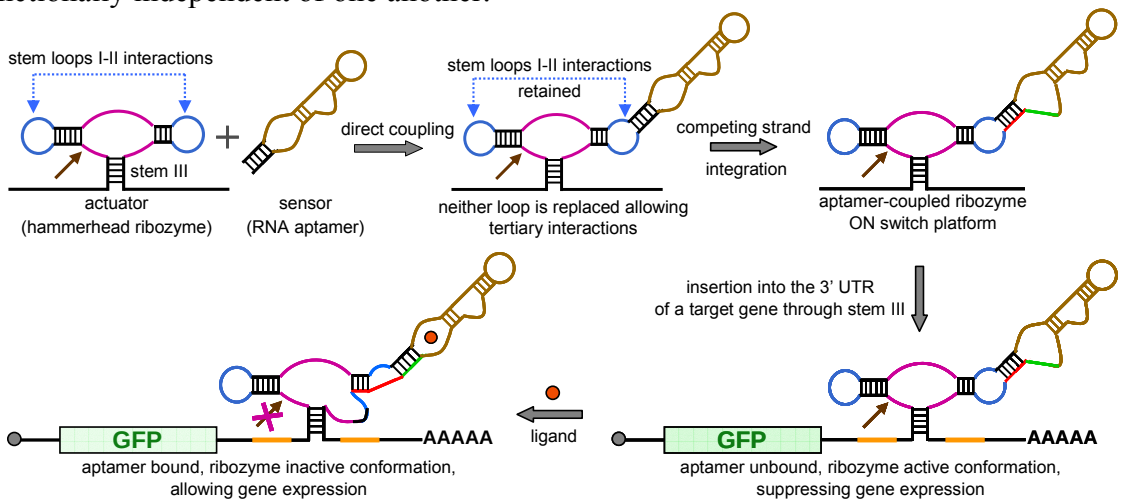


Figure 5.1. General composition framework and modular design strategy for engineering ligand-controlled ribozyme switch-based gene regulatory systems[1]. Color schemes: catalytic core, purple; loop sequences, blue; aptamer sequence, brown; competing strand, green; switching strand, red; spacer sequences, orange; cleavage site, brown arrow. An aptamer is directly attached to the ribozyme through one of its loops without replacing any part of the ribozyme, thereby maintaining loop I-II interactions required for *in vivo* functionality. Spacer sequences are included on both ends of the ribozyme switch to insulate the molecule from non-specific interactions with the surrounding sequences. A competing strand, whose sequence is similar to that of the switching strand, is integrated into the aptamer-coupled ribozyme, which enables the RNA molecule to adopt two primary conformations at equilibrium through the strand displacement mechanism. Ligand binding shifts the

equilibrium distribution towards the conformation in which the ligand-binding pocket is formed.

A general construction scheme for protein-responsive ribozyme switches is illustrated in Figure 5.2 based on a ribozyme ON switch platform, where an RNA aptamer for potentially any target protein of interest is directly replaced within the sensor domain of the switch molecule. In principle, in the absence or little expression of the target protein ligand inside the cell, the ribozyme switch will down-regulate the mRNA level within which the switch is embedded. As the expression level of the protein ligand increases and the ligand binds to its aptamer, the ribozyme switch will begin to up-regulate the level of the encoded transcript. In the case of regulating reporter gene expression, ribozyme switches can serve as cellular biosensors by transmitting a change in the levels of a protein ligand of interest to a change in the reporter gene expression level. In the case of regulating the expression of a particular target gene, ribozyme switches can function as specific protein-responsive gene regulators. The latter regulatory scheme potentially holds a therapeutic value in the treatment of protein-associated diseases.

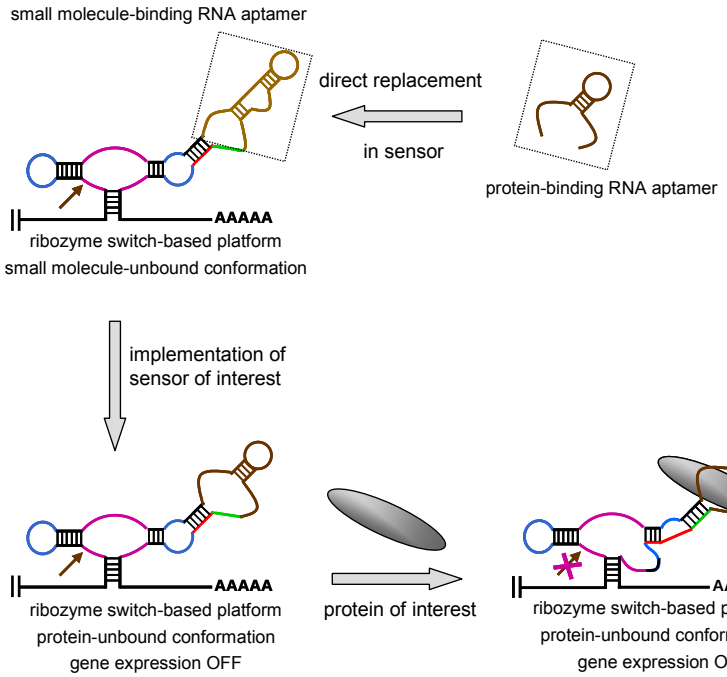


Figure 5.2. A general construction scheme for protein-responsive ribozyme switches. Color schemes follow those described in Figure 5.1.

5.2.2. Development of protein-responsive ribozyme switches

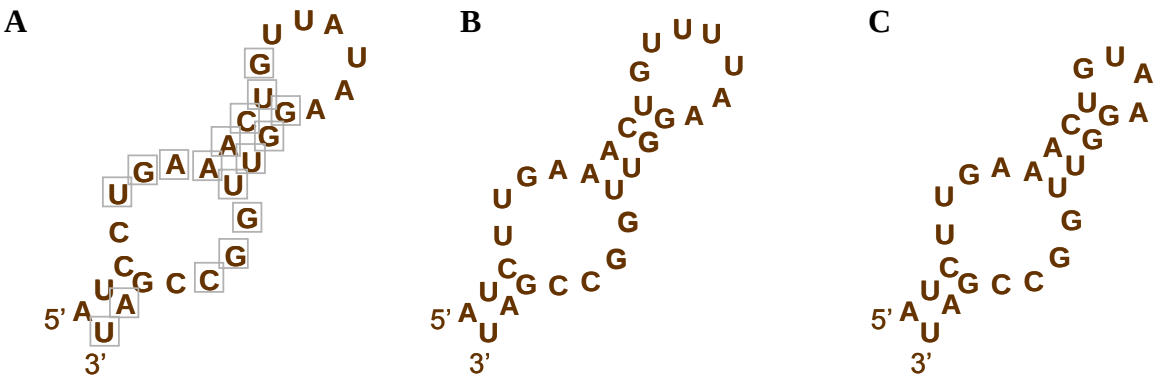


Figure 5.3. Sequences and structures of RNA aptamers with affinity and specificity for NF-κB p50. (A) the original NF-κB p50-binding RNA aptamer[16] (NF-kB1). (B) and (C) the optimized NF-κB p50-binding RNA aptamers[17, 18] (NF-kB4 and NF-kB5, respectively). The conserved nucleotides are indicated in gray boxes.

To develop a protein-responsive ribozyme switch, we employed an existing RNA aptamer with high binding affinity and specificity for the p50 subunit of the human nuclear transcription factor kappa B (NF-κB p50)[16] as a sensor component (NF-κB1, Figure 5.3A). We employed this aptamer primarily because it has been previously shown to bind its target transcription factor in a yeast three-hybrid system in activating the transcriptional event, thus supporting the *in vivo* target recognition capability of this *in vitro* selected aptamer[17, 18]. In addition to this original NF-κB p50 aptamer, we also employed two optimized derivatives of this aptamer[17, 18] (NF-κB4 and NF-κB5, respectively, Figure 5.3, B and C). We implemented these aptamers within the sensor domain of a gene ON ribozyme switch, L2bulge1[1]. As the conserved nucleotides are present in the base stem of these aptamers, we modified the transmitter domain of L2bulge1 to include these nucleotides in constructing an NF-κB p50-

responsive ribozyme switch, L2bulge1NF-κB1 (Figure 5.4). The sequences of the other two switch constructs, L2bulge1NF-κB4 and L2bulge1NF-κB5 are shown in Figure 5.5.

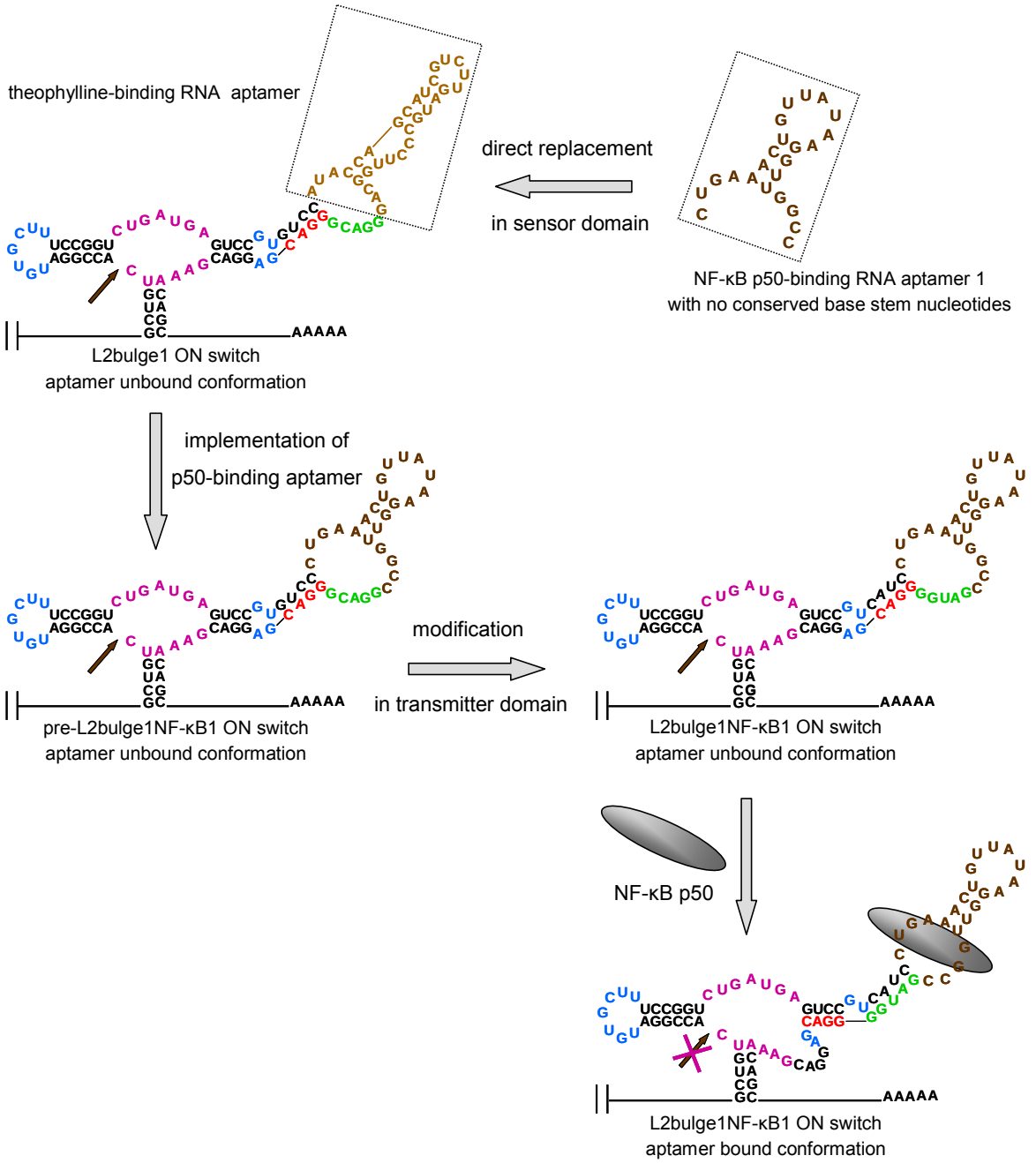


Figure 5.4. Modular design strategies and systematic engineering of an NF-κB p50-responsive ribozyme switch. Color schemes follow those described in Figure 5.1. An ON switch platform, L2bulge1, is used for illustration, where an NF-κB p50-binding aptamer (right dashed box) is directly replaced within the existing sensor domain (left dashed box) to construct a p50-responsive gene regulator/cellular biosensor. Sequence modifications were made within the transmitter domain to include the conserved nucleotides present in the base stem of the p50-binding aptamer.

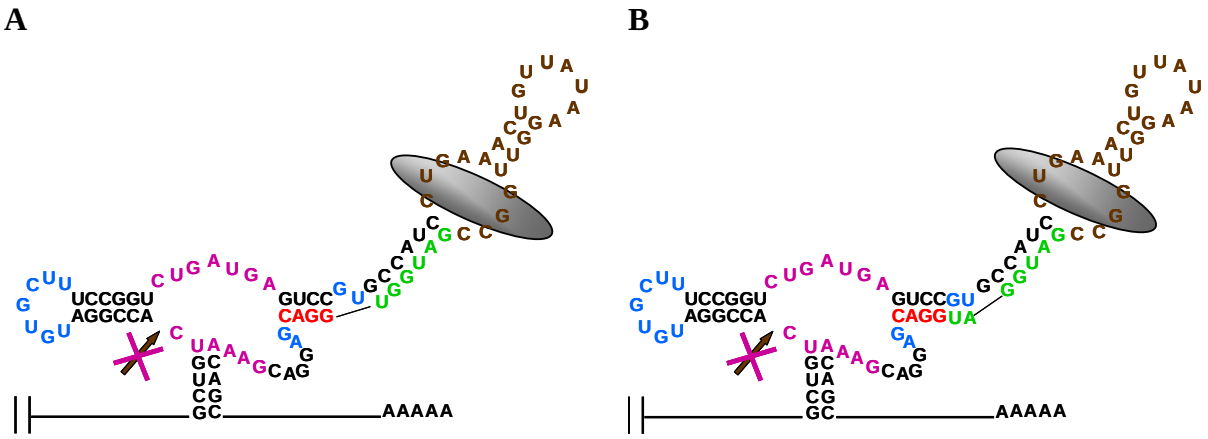


Figure 5.5. Sequences and secondary structures of two p50-responsive ribozyme switches in their ligand-bound conformations: (A) L2bulge1NF-κB4 and (B) L2bulge1NF-κB5. Color schemes follow those described in Figure 5.1. The transmitter domains of these switches are modified from that of L2bulge1 to include the conserved nucleotides present in the aptamer base stem.

5.2.3. In vivo functional activity of p50-responsive ribozyme switches

The resulting ribozyme switches were each integrated into the 3' untranslated region (UTR) of a fluorescent reporter gene and expressed from a ribozyme-switch characterization plasmid. As the plasmid employed in the yeast three hybrid system, which contains the gene encoding NF-κB p50 fused with GAL4 activation domain (GAL4AD) and tagged with a nuclear localization signal[18], was compatible with our expression system, this plasmid was co-transformed with each ribozyme switch to endogenously express the protein ligand in *Saccharomyces cerevisiae* cells. The up-regulation of reporter gene expression was observed in cells harboring both the ribozyme switch and p50 expression constructs in comparison to cells harboring just the ribozyme switch and no p50 expression construct, demonstrating the protein-responsive gene regulatory/biosensor function of these ribozyme switches (Figure 5.6).

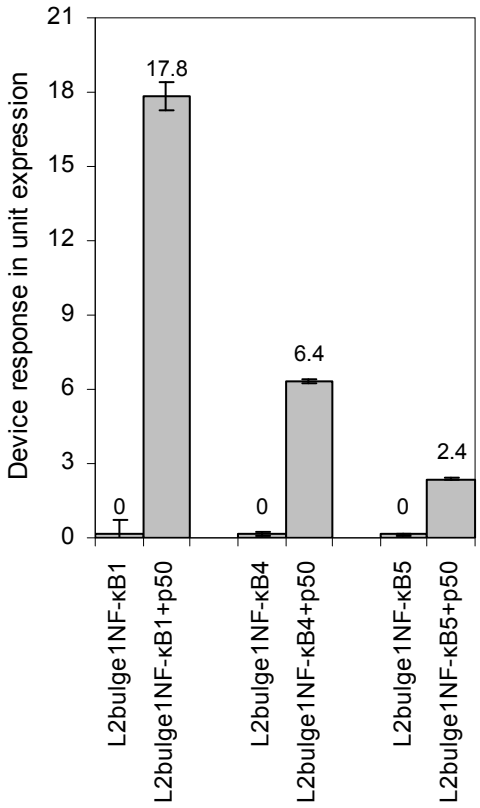


Figure 5.6. The device response of p50-responsive ribozyme switches exhibiting ON switch regulatory responses. Device response is reported in unit expression as described in Materials and Methods.

Among the three p50-responsive ribozyme switches, L2b1NF-κB1 comprised of the original *in vitro*-selected aptamer, exhibits the largest regulatory response (Figure 5.6) compared to the other two switches, L2b1NF-κB4 and L2b1NF-κB5, which are comprised of the optimized aptamer sequences, indicating that the optimization in the yeast three-hybrid system was not translatable to the ribozyme-switch system. This may be explained by the fact that these sequences were optimized in the context of the three hybrid system to more effectively activate transcription and thus might be dependent on the surrounding sequences to which they were coupled in the hybrid system to achieve the optimized activity.

5.3. Discussion and Future Work

The preliminary data show that the ribozyme switches comprised of p50-binding RNA aptamers as sensing elements are capable of regulating target gene expression through molecular recognition of p50 in living cells, thus serving as p50-responsive gene regulators and cellular biosensors. However, additional experiments will be conducted to further support and confirm the p50-responsive ribozyme-switch activity currently observed. First, the p50-directed switch activity will be characterized without GAL4AD fused into p50 to exclude the possible non-specific activity arising from the presence of GAL4AD. In addition, nuclear localization signal will be removed from p50 to examine if the switch activity is still observed when the ligand protein is present only in the cytoplasm. This study will provide insights into the mechanism of ribozyme switch regulation, specifically where the binding-cleavage events occur inside the cell. Second, p50 will be expressed under the control of an inducible promoter system so that a transfer function of device response may be generated across a gradient of p50 expression levels. Third, mutational studies will be conducted by introducing a single or few nucleotide mutations within the p50-binding aptamer sequences that are involved in p50 binding to confirm that the observed regulatory and sensing activities are indeed due to the precise recognition of the target protein. Fourth, quantification of cellular transcript levels through quantitative real-time polymerase chain reaction (qRT-PCR) will be carried out, which will provide information associated with p50-dependent regulation of transcripts through catalytic cleavage. Fifth, protein gel assay using Coomassie brilliant blue[19] or Western blots[20] will be performed to ensure *in vivo* expression of the effector protein p50. Finally, the regulation of gene expression through more advanced devices will be examined by integrating a single-input small molecule-responsive ribozyme switch and a

single-input protein-responsive ribozyme switch within the 3' UTR of a reporter gene. This implementation allows sophisticated control of gene expression in response to two different classes of input ligands, small molecules and proteins, and thus represents a higher-order signal integration scheme. The p50-binding RNA aptamers may also be implemented within the sensor domain of a gene OFF switch, such as L2bulgeOff1[1]. The functional demonstration of a protein-responsive OFF switch will broaden the application areas of these protein-responsive gene regulators/cellular biosensors.

Successful demonstration of ribozyme switch-based protein-responsive gene regulators and cellular biosensors will provide a significant advance in the current technologies available for controlling gene expression and probing protein interactions and functions. For instance, these gene regulators will enable the reprogramming of cellular behavior through regulation of the expression of certain proteins of interest in a specific effector protein-dependent manner, where the regulated protein may be different from the effector protein and the latter may be native to the cellular system employed. In regulating the expression of reporter proteins, ribozyme switches will serve as noninvasive synthetic cellular biosensors to monitor temporal and spatial fluctuations in the expression levels of both exogenous and endogenous proteins of interest. Therefore, the development of protein-responsive gene regulators and cellular biosensors based on a modular and extensible functional platform is critical for broader applications of these tools to a wide range of basic and applied biological research.

5.4. Materials and Methods

5.4.1. Plasmid construction and transformation

A ribozyme-switch characterization low-copy plasmid, pRzS-TEF1, harboring the yeast enhanced green fluorescent protein (yEGFP) under the control of a TEF1 promoter and an ADH1 terminator was used to characterize protein-responsive ribozyme switches. Ribozyme switches comprising the p50-binding RNA aptamer in the sensor domain were cloned into the 3' UTR of the *yegfp* gene, using two unique restriction sites, Avr II and Xho I, 3 nucleotides downstream of the stop codon of yEGFP. The gene encoding p50 under the control of an ADH1 promoter was expressed from a high-copy plasmid[18].

The engineered ribozyme constructs were generated by PCR amplification using the appropriate oligonucleotide templates and primers. Cloned plasmids were transformed into an electrocompetent *Escherichia coli* strain, DH10B (Invitrogen) and all ribozyme constructs were confirmed by subsequent sequencing (Laragen, Inc). Confirmed ribozyme plasmids were co-transformed with the plasmid encoding p50 into a *Saccharomyces cerevisiae* strain (W303 *MATα his3-11,15 trp1-1 leu2-3 ura3-1 ade2-1*) using standard lithium acetate procedures[21]. As a control, each ribozyme plasmid was transformed alone into *S. cerevisiae*, which serves as the basal expression level of the ribozyme construct in the absence of p50 expression.

5.4.2. RNA secondary structure prediction and free energy calculation

RNAstructure 4.2 (http://rna.urmc.rochester.edu/rnastructure.html) was used to predict the secondary structures of all switch constructs and their thermodynamic properties as described previously[1]. RNA sequences that were predicted to adopt at least two stable equilibrium conformations (ribozyme inactive and active) were constructed and examined for functional activity.

5.4.3. Ribozyme characterization assays

Ribozyme switch activities were examined through characterization assays as described previously[1]. *S. cerevisiae* cells harboring the appropriate plasmids were grown in synthetic complete medium supplemented with an appropriate dropout solution and sugar overnight at 30^{o}C. Overnight cultures were back diluted into fresh medium to an optical density at 600 nm (OD_{600}) of approximately 0.1 and grown at 30^{o}C. Back-diluted cells were grown to an OD_{600} of 0.8-1.0 or for a period of approximately 6 h before measuring GFP levels on a Cell Lab Quanta SC flow cytometer (Beckman Coulter).

5.4.4. Fluorescence quantification

Fluorescence measurements were taken following the protocols described previously[1]. The population-averaged fluorescence of each sample was measured on a Safire fluorescence plate reader with the following settings: excitation wavelength of 485 nm, an emission wavelength of 515 nm, and a gain of 100. Fluorescence readings were normalized to cell number by dividing fluorescence units by the OD_{600} of the cell sample and subtracting the background fluorescence level to eliminate autofluorescence.

Fluorescence distributions within the cell populations were measured on a Quanta flow cytometer with the following settings: 488 nm laser line, 525 nm bandpass filter, and PMT setting of 5.83. Fluorescence data was collected under low flow rates for approximately 30,000 cells. Viable cells were selected and fluorescence levels were determined from 10,000 counts in this selected population. Background fluorescent cell population was used to set a ‘negative GFP’ gate. Cells exhibiting fluorescence above this negative gate are defined as the ‘positive GFP’ cell population.

Ligand protein-directed regulatory effects are reported as device response in units of expression normalized to the levels in the absence of ligand where 1 unit expression is defined as the reporter gene expression level of the native ribozyme, sTRSV[15], relative to the background fluoresence level. The expression level of sTRSV is ~2% of that of the inactive ribozyme or the full transcriptional range of 50 units of expression[1]. All fluorescence data and mean ±s.d. are reported from at least three independent experiments.

Acknowledgements

We thank A. Babiskin for providing pRzS and J. Liang and L. d'Espaux for experimental assistance. This work was supported by the Arnold and Mabel Beckman Foundation, the National Institutes of Health, and the Center for Biological Circuit Design at Caltech (fellowship to M.N.W.).

References

1. Win, M. N. & Smolke, C. D. From the Cover: A modular and extensible RNA-based gene-regulatory platform for engineering cellular function. Proc Natl Acad Sci U S A 104, 14283-8 (2007).
2. Angenendt, P. Progress in protein and antibody microarray technology. Drug Discov Today 10, 503-11 (2005).
3. Backmann, N. et al. A label-free immunosensor array using single-chain antibody fragments. Proc Natl Acad Sci U S A 102, 14587-92 (2005).
4. Mattoon, D., Michaud, G., Merkel, J. & Schweitzer, B. Biomarker discovery using protein microarray technology platforms: antibody-antigen complex profiling. Expert Rev Proteomics 2, 879-89 (2005).

5. Ghaemmaghami, S. et al. Global analysis of protein expression in yeast. Nature 425, 737-41 (2003).

6. Jayasena, S. D. Aptamers: an emerging class of molecules that rival antibodies in diagnostics. Clin Chem 45, 1628-50 (1999).

7. O'Sullivan, C. K. Aptasensors--the future of biosensing? Anal Bioanal Chem 372, 44-8 (2002).

8. Huh, W. K. et al. Global analysis of protein localization in budding yeast. Nature 425, 686-91 (2003).

9. Hahn, K. & Toutchkine, A. Live-cell fluorescent biosensors for activated signaling proteins. Curr Opin Cell Biol 14, 167-72 (2002).

10. Li, I. T., Pham, E. & Truong, K. Protein biosensors based on the principle of fluorescence resonance energy transfer for monitoring cellular dynamics. Biotechnol Lett 28, 1971-82 (2006).

11. Nguyen, A. W. & Daugherty, P. S. Evolutionary optimization of fluorescent proteins for intracellular FRET. Nat Biotechnol 23, 355-60 (2005).

12. Tuerk, C. & Gold, L. Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. Science 249, 505-10 (1990).

13. Ellington, A. D. & Szostak, J. W. In vitro selection of RNA molecules that bind specific ligands. Nature 346, 818-22 (1990).

14. Vaish, N. K. et al. Monitoring post-translational modification of proteins with allosteric ribozymes. Nat Biotechnol 20, 810-5 (2002).

15. Khvorova, A., Lescoute, A., Westhof, E. & Jayasena, S. D. Sequence elements outside the hammerhead ribozyme catalytic core enable intracellular activity. Nat Struct Biol 10, 708-12 (2003).

16. Lebruska, L. L. & Maher, L. J., 3rd. Selection and characterization of an RNA decoy for transcription factor NF-kappa B. Biochemistry 38, 3168-74 (1999).

17. Cassiday, L. A. & Maher, L. J., 3rd. In vivo recognition of an RNA aptamer by its transcription factor target. Biochemistry 40, 2433-8 (2001).

18. Cassiday, L. A. & Maher, L. J., 3rd. Yeast genetic selections to optimize RNA decoys for transcription factor NF-kappa B. Proc Natl Acad Sci U S A 100, 3930-5 (2003).

19. Uefuji, H., Ogita, S., Yamaguchi, Y., Koizumi, N. & Sano, H. Molecular cloning and functional characterization of three distinct N-methyltransferases involved in the caffeine biosynthetic pathway in coffee plants. Plant Physiol 132, 372-80 (2003).

20. Mizuno, K. et al. Isolation of a new dual-functional caffeine synthase gene encoding an enzyme for the conversion of 7-methylxanthine to caffeine from coffee (Coffea arabica L.). FEBS Lett 534, 75-81 (2003).

21. Gietz, R. & Woods, R. in Guide to Yeast Genetics and Molecular and Cell Biology, Part B, (eds. Guthrie, C. & Fink, G.) 87-96 (Academic Press, San Diego, 2002).

Conclusions

It is becoming increasingly evident that RNA is a functionally diverse and versatile molecule that plays an essential role in regulating gene expression in various organisms. Regulation can occur at different levels in the gene expression pathways, including transcription and translation, and through diverse mechanisms, including RNA-RNA base-pairing interactions and cleavage via different physical implementations, *in cis* or *in trans*. RNA is a flexible molecule that can adopt different conformations and fold into secondary and tertiary structures, yet still possesses the ability to undergo dynamic conformational changes, which allow it to interact with a wide range of biological molecules such as DNA, protein, and other RNA molecules. RNA molecules exhibit ligand recognition properties, through which they can sense their environment, and allosteric binding properties, through which they can self-regulate their own activity. These properties enable RNA to function as precise molecular sensors and autonomous control systems that require no additional aid from proteins as intermediate sensor or actuator elements.

Due to its unique array of functional properties, RNA is a powerful platform for the design of regulatory molecules for a wide range of biotechnological and medical applications. Unlike larger biomolecules such as proteins, the functional activity of RNA is more directly defined by its secondary structure. This relationship between RNA secondary structure and function, in combination with predictive RNA secondary structure / energetic folding programs, has enabled molecular engineers to construct diverse synthetic regulatory RNA elements and to manipulate biological processes at the molecular level with greater flexibility and reliability.

RNA engineering has rapidly emerged in the last decade, as RNA has proven to be a versatile molecule capable of performing various cellular functions that go beyond passive transfer of genetic information between the genome and the proteome. RNA represents an attractive and excellent substrate for the construction of gene control devices, and thus numerous synthetic RNA devices have been engineered for regulating the expression of various target genes. Supported by technologies that allow the generation of RNA sensors, researchers have integrated sensory elements and developed many examples of RNA devices comprised of integrated sensor and actuator domains that control gene expression in response to specific target molecules. While these examples have made profound contributions in advancing the field of RNA engineering, they either fail to function *in vivo*, or lack one or more key engineering properties, such as portability, scalability, and modularity. In addition, their artisanal designs and construction do not support component reuse and a general device composition framework.

We set out to develop a functional composition framework that supports the reliable design and construction of synthetic molecular switches and sensors from modular functional components to support the generation of many such devices that can be implemented in a broad range of biotechnological applications such as gene expression regulation, biosensors, therapeutic molecule design, metabolic reprogramming, and tools for elucidating cellular function. The RNA devices are based on a modular, extensible, and portable regulatory platform that enables the tailored construction of application-specific switches and sensors responsive to user-defined target ligands without complex device redesign. In addition, the RNA devices can be utilized in concert

with other gene-regulatory components such as inducible promoters, and synergistic and application-specific gene-regulatory responses can be readily obtained through such combinatorial designs. For instance, similar to the strategies described previously, such as energetic tuning of individual devices or signal integration within the 3' UTR by multi-copy device expression, the regulatory dynamic range of an RNA device can be altered to fit particular applications by expressing the device under the control of an inducible promoter at an appropriate induction level. Such strategies will enable the tuning of the device response levels in the absence and presence of a target ligand to match the application-specific threshold levels. Such component matching between the ribozyme switches and other biological components in the genetic system will broaden the utility of this molecular tool. Furthermore, validation of the preliminary results observed with the p50-responsive ribozyme switches is important to demonstrate the extensibility of the sensing platform to a different class of targets, protein ligands, to extend the utility of this platform. In addition, technological advances that enable the efficient generation of well-characterized libraries of sensor components (RNA aptamers) that recognize biologically-relevant molecules, function in the cellular environment, and are compatible with the described composition frameworks will be critical to the broader implementation of these frameworks for the construction of 'designer' RNA devices capable of processing user-specified environmental and intracellular signals.

The functional performance of the RNA devices described in this thesis can also be improved through the integration of future scientific and technological advances. For instance, advances in RNA engineering that provide insights into RNA structure-function relationships and improved predictions of RNA secondary and tertiary structures relevant

to *in vivo* folding environments will enable the development of RNA devices with more robust functional performance. In addition, modeling tools that can predict both thermodynamic and kinetic properties of RNA folding *in vivo*, incorporate tertiary interactions, and link those properties to functional states and gene expression pathways will support future design tools that efficiently optimize and program device properties *in silico*.

Engineered RNA devices have significant potential to transform our ability to interface with biological systems and to fundamentally change the ways in which we manipulate, program, and probe living systems. Many effective solutions to the societal challenges faced today, including ‘renewable’ energy and the ‘green’ environment, sustainability, and health and medicine, will be derived from engineered biological systems, and thus we hope that our engineered RNA devices play a role in achieving these solutions.