High-Throughput Strategies for the Scalable Generation of RNA Component Functions

Thesis by

Joe Chih Yao Liang

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

California Institute of Technology

Pasadena, California

2012

(Defended January 24th, 2012)

ACKNOWLEDGMENTS

Words and sentences cannot simply describe the transformation that I have gone through in the past years, so I would like to be brief. First, I would like to thank Christina for being a wonderful mentor. Her dedication to students and Twilight movies has influenced me on many levels. I am also inspired to buy more Camper shoes in the near future. Second, I would like to thank all my labmates. Together, we survived the mysterious lab fire that destroyed the entire tissue-culture room. I would certainly miss all the bad movies we watched together (remember Thor?) and, of course, the yummy Hobees brunches. I would like to especially thank DK for being a human Wikipedia of pop culture and showing me the Canadian pride. I would also like to thank Kate for being my best lunch buddy and a wonderful cook. Third, I would like to thank my buds all over California. I would like to thank Frank, Joyce, and Jonathan for being wonderful friends to me. I enjoy beating you all in tennis. I would like to thank the Crazies (in no particular order): Melisa, Jae, Nik, Ant, Jeff, Nate, Chi, Van, and Nhi. I would not forget the infamous SF Lime brunch fiasco and numerous memorable Vegas trips. I would like to also thank the Carpio family for treating me like part of the Carpios. I would not forget all the fun Napa and SF outings we had, and I would make a French Laundry reservation for all in no time. Last but not least, family is what keeps me going. I would like to thank my mom, dad, cousins (Christie, Michelle, Simone, Kailin, Daphne, Jerry, and Ariel), uncles, aunts, grandparents, and my doggie Max. It has been a tough road, but you are what make this experience enjoyable. With all the love and support from you all, I am gladly starting a new and exciting chapter of my life. Thank you.

ABSTRACT

Engineered biological systems hold great promise in providing solutions to many global challenges, including environmental remediation, sustainability, scalable manufacturing, and health and medicine. Synthetic biology is an emerging research field with a primary goal of making the engineering of biology more streamlined and reliable. Recent advances in synthetic RNA biology have led to design of RNA-based gene-regulatory devices from assembly of functional RNA components that encode more basic functions, including sensing, information transmission, and actuation functions. These synthetic RNA control devices allow access and control information on cellular state, thereby advancing our ability to interact with and program biology.

A modular ribozyme device platform was recently described to link an aptamer (sensor) to a hammerhead ribozyme (actuator) through a distinct sequence (information transmitter) capable of a strand-displacement event. The utilization of ribozyme as the actuator in the platform, whose mechanism of action is independent of cell-specific machinery, allows transport of the resultant devices to *in vitro* or different cellular environments. The broad implementation of these devices requires enabling technologies to support efficient generation of new functional RNA components and quantitative tailoring of device regulatory performance for specific cellular applications. Current component generation and device tailoring strategies are limited in their throughputs and efficiencies, and thus have hampered our ability to generate new ribozyme devices for cellular engineering applications.

To support scalable generation and tailoring of ribozyme devices, we have described high-throughput *in vitro* selection and *in vivo* screening strategies based on the

modular ribozyme device platform. We proposed a high-throughput solution-based *in vitro* selection strategy to generate new sensing functions within the device platform. A high-throughput and quantitative two-color FACS-based screening strategy was developed to complement the *in vitro* selection strategy by allowing efficient tailoring of device regulatory activities in the cellular environments. We further developed quantitative assays based on the surface plasmon resonance (SPR) technology to allow rapid measurements of the device and component activities. Together, these enabling strategies will offer a scalable and integrated process for the construction and programming of RNA control devices for broad cellular engineering applications, thus laying an important foundation for engineering more complex biological systems.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1.

Introduction*

*Sections 1.1-1.7: Reproduced/adapted with permission from: Liang, J.C., Bloom, R.J. and Smolke, C.D. (2011) Engineering biological systems with synthetic RNA molecules. *Mol Cell*, **43**, 915-926.

Engineered biological systems have exciting potential in developing solutions to many global challenges, including environmental remediation, sustainability, scalable manufacturing, and health and medicine (1-4). Our ability to design and build synthetic biological systems is a key technology to improving the human condition. In addition, the redesign of biological systems can be used as an effective strategy to test, and thereby strengthen, our understanding of natural systems. Synthetic biology is an emerging research field with a primary goal of making the engineering of biology faster, less expensive, and more reliable. As such, core activities in synthetic biology have been focused on the development of foundational tools and technologies that assist in the design, construction, and characterization of biological systems (5,6). Recent advances in construction and fabrication technologies are supporting synthesis of large pieces of DNA including entire pathways and genomes (7). While progress has been made in the design of complex genetic circuits (8), current capabilities for constructing large genetic systems surpass our ability to design such systems. This growing 'design gap' has highlighted the need to develop methods that support the generation of new functional biological components and scalable design strategies for complex genetic circuits that will lay the foundation for integrated biological devices and systems.

The vast majority of genetic systems engineered to-date have utilized protein-based transcriptional control strategies (8). However, as the examples of functional RNA molecules playing key roles in the behavior of natural biological systems have grown over the past decade, there has been growing interest in the design and implementation of synthetic counterparts. Researchers have taken advantage of the relative ease with which RNA molecules can be modeled and designed to engineer functional RNA molecules that

act as diverse components including sensors, regulators, controllers (ligand-responsive RNA regulators), and scaffolds. More recently, researchers have begun to move beyond molecular design and integrate these synthetic RNA molecules as key elements in genetic circuits to program cellular behavior, highlighting the relevance and advantages of RNA-based control strategies.

1.1 RNA as a natural regulatory molecule

The growing interest in using RNA to build synthetic controllers is due in large part to the steadily increasing examples of natural RNA regulators that control gene expression through diverse mechanisms in different organisms. One of the earliest examples is the regulation of gene expression through RNA secondary structure. The study of differential expression of genes in phage genomes led to the discovery that secondary structure of a mRNA transcript can restrict access to the ribosome binding site (RBS), thereby inhibiting translation (9). Similarly, bacteria utilize the formation of tight hairpins in mRNA transcripts to stall and attenuate translation in the regulation of amino acid biosynthesis (10). In addition, RNA structure is highly temperature sensitive, such that in certain cases hairpin structures that inhibit translation and can be modulated by temperature have been found to have functional roles in the heat and cold shock responses of several bacteria (9). Finally, it has been shown that strong secondary structures on the 5’ and 3’ ends of a mRNA strand can protect the transcript from degradation by exoribonucleases and endoribonucleases (11). The resulting extended half-lives of the transcripts can significantly increase protein production and have functional roles in processes such as photosynthesis and bacterial cell adhesion (11,12).

In addition to structural mechanisms, the discovery that RNA can exhibit catalytic activity opened the door to a wider array of regulatory functions (13). These catalytic RNAs, or ribozymes, typically catalyze cleavage or ligation of the RNA backbone through a reversible phosphodiester cleavage reaction (14). Ribozymes have functional roles in alternative splicing, RNA replication, translation, and transcript stability and function in both prokaryotes and eukaryotes (14). Furthermore, the discovery that ribozyme cleavage of the glmS transcript in bacteria is inhibited by binding of the metabolite GlcN6P has led to several discoveries of ribozymes acting as key components in riboswitches, a class of RNA regulators that respond to cellular metabolites and cofactors to modulate enzyme levels in related biosynthesis (15). Finally, RNase P is a catalytic RNA that functions in trans and can carry out multiple turnover cleavage events in the processing of 5' leader sequences from tRNA (14). The discovery of ribozymes with natural gene regulatory activity in trans presents an intriguing proof of principle that a single catalytic RNA can be used to regulate several different genes in a biological system.

The last major mechanism that RNA uses to regulate protein synthesis is through antisense-mediated regulation of translation. Trans-acting small RNAs in bacteria are generally transcribed from their own independent transcripts and can range from 50-514 nucleotides (nts) in length (16). These RNAs often act in concert with the RNA binding protein Hfq and can promote or inhibit translation of their target mRNAs by relieving secondary structural elements or inhibiting ribosome initiation or processivity (17). Similarly, in higher eukaryotes, RNA interference (RNAi) pathways use small RNAs, siRNA and microRNAs, to guide protein complexes to complementary mRNAs, leading

to silencing of those targets (18). MicroRNAs have diverse roles in almost every cellular process and they are currently thought to regulate up to a third of human genes (19). RNAi, particularly the siRNA pathway, is widely used as a tool in biological research for genetic loss of function studies and is currently being explored for therapeutic and biotechnological uses (20).

Finally, RNA is involved in gene regulation in several other less widespread or well-studied mechanisms. For example, noncoding RNAs have been characterized that have roles in quality control of translation (16), binding to and inhibiting proteins involved in protein synthesis, and epigenetic DNA modification (21). Furthermore, the recent discovery of the widespread regulatory activity and conservation of long noncoding RNAs represents an exciting area for further research in RNA-based gene regulation (22).

1.2 RNA as a programmable and efficient substrate for engineering biological controllers

In addition to the diversity of mechanisms by which RNA can act as a regulatory molecule in nature, RNA exhibits several properties that make it an attractive design substrate in synthetic systems. RNA is composed of four building blocks that interact through well-characterized hydrogen-bond, base-stacking, and electrostatic interactions. The folding of RNA is primarily dictated by its secondary structure, in contrast to the folding of proteins, which involves a large degree of tertiary interactions. Models that predict RNA secondary structure have been developed based on the optimization of energies contributed by the Watson-Crick AU and GC base pairs as well as the GU

wobble pair (23-25). Progress in RNA three-dimensional structural studies have revealed a set of non-canonical base pairing interactions, or Hoogsteen base pairs, that are key in the formation of the RNA tertiary structure. Early modeling frameworks that predict RNA tertiary structure based on a primary sequence have been developed and demonstrated to generate native-like structure predictions (26,27). In addition to structure prediction based on thermodynamics, RNA kinetic folding program has also been developed to capture the stochastic nature of the RNA folding process (28). These computational tools can be utilized to access structural information encoded in the primary RNA sequence, thereby aiding the rational design of genetic controllers based on hybridization schemes or structural elements.

RNA-based controllers exhibit additional advantages as functional elements in synthetic biological systems as the field moves toward more complex genetic systems design. RNA controllers generally exhibit more compact genetic footprints than their protein counterparts. In addition, RNA controllers generally place less of an energetic and resource load on the host cell as functional RNA molecules do not require the translation process to synthesize the functional elements. Another important consideration in genetic system design is timing of coupled control processes, where RNA-based posttranscriptional control strategies will generally act at faster time scales than transcription-based control strategies. The energetic, resource, and space efficiencies of RNA-based control strategies present important features supporting scaling to large-scale genetic system designs.

1.3 Approaches to generate functional RNA components

Functional RNA components are the basic building blocks for constructing genetic devices that encode human-defined functions. A fundamental challenge in the field then is the efficient generation of new component functions, such as sensing, information transmission, and actuation functions, that can be assembled into more complex devices. Three general approaches are taken to engineer new functional RNA components: harvesting from nature, computational design, and molecular evolution (Figure 1.1).

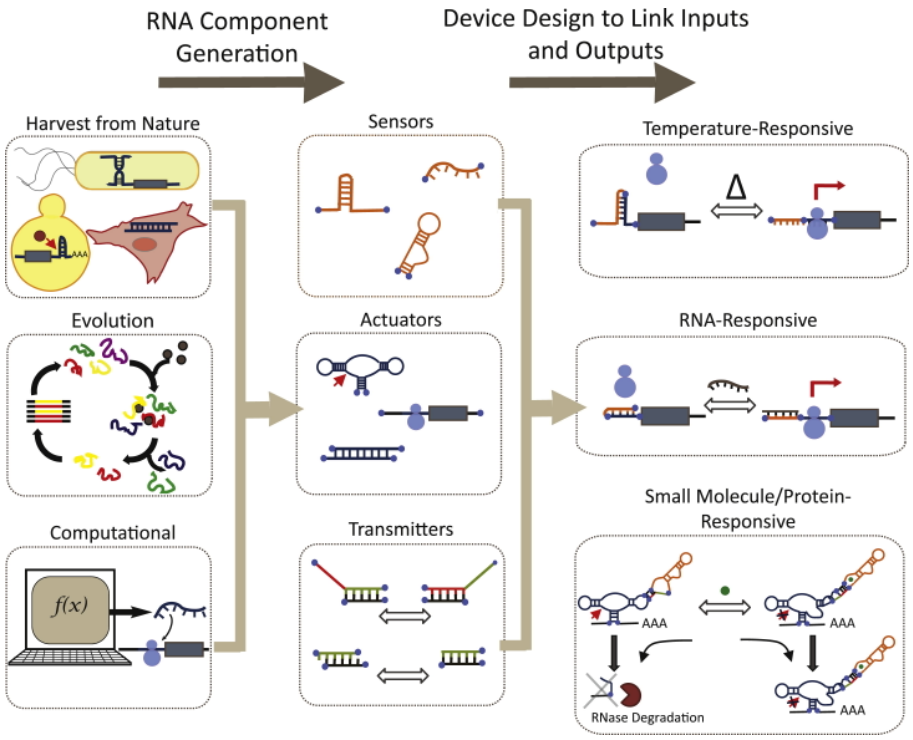


Figure 1.1 RNA components used to engineer synthetic control functions can be harvested from natural systems or generated using molecular evolution and computational approaches. These components, encoding sensing, actuation, and information transmission activities, can then be assembled into RNA devices using various molecular engineering strategies to link one or more inputs of interest, such as temperature, RNA, small molecules, or proteins, to desired regulatory activities.

1.3.1 Harvesting and refining RNA components from nature

Many functional RNA components are derived from naturally-occurring elements, as advances in RNA biology have led to discoveries of natural RNA molecules that exhibit diverse functional activities. In certain instances the RNA component can be 'harvested' from its native context and used in a synthetic genetic context where it will exhibit the desired activity. In one example, researchers isolated a ligand-binding RNA sequence to thiamine pyrophosphate (TPP) from a natural riboswitch and linked this natural RNA ligand-binding component to a hammerhead ribozyme to construct a synthetic RNA-based sensing-actuation element that responded to TPP in *E. coli* (29). The native function of a naturally-occurring RNA element can also be altered to generate new functions or 'refined' to make the element compatible with modular integration into broader genetic device or system platforms. Combined rational and evolutionary strategies have been applied to alter the native function of natural RNA elements. In one example, the native activity of an RNase III hairpin substrate was altered in yeast by modifying key sequences that are essential to protein-binding and cleavage activities within the regulatory element (30,31). The resulting hairpin libraries were inserted in the 3' UTR of the target transcript, and an *in vivo* screen was performed to identify a set of new hairpin sequences exhibiting a wide-range of regulatory activities. Similar approaches have been applied to alter the native activities of diverse RNA regulatory elements, including ribozymes (32), RNase cleavage sites (33), stabilizing elements (34), RBS sequences (35) , and riboswitches (36).

1.3.2 Computational tools for the design of RNA components

Computational tools have been developed to aid researchers in designing functional RNA components. For example, a computational method for designing synthetic RBS sequences was recently described based on predicting the energies of secondary structures around and including the RBS and the energies associated with the RBS:rRNA interaction (37). A two state thermodynamic model was devised, the initial state being the free 30S ribosomal subunit and folded mRNA transcript and the final state being the 30S complex bound to the transcript. The difference in free energy between these two states was used to predict a translation initiation rate for a given mRNA sequence. The computational method was applied to generate RBS sequences in *E. coli* that resulted in fluorescent reporter levels that spanned several orders of magnitude. This tool was then applied to design and optimize the expression of an input to a genetic AND gate. As another example, design tools for microRNA and shRNA elements have been developed by companies such as Life Technologies and Thermo Scientific. Given the sequence of a target gene of interest, these RNAi design tools can predict targeting sequences that will allow for efficient RNAi-mediated knockdown of that gene. These programs will design a shRNA or miRNA that incorporate the predicted targeting sequences, including the structural elements necessary for correct biogenesis and silencing efficiency, and primer sequences for construction and cloning of the regulatory element.

Forward engineering with current computational tools is not precise, and strategies for most effectively using these tools require some level of generating multiple functional RNA elements and then screening for those sequences that exhibit desired

activities. For example, the RBS design tool has a probability of 0.47 of achieving a target protein expression level within two-fold. Further development of design tools that more accurately predict the precise DNA sequence needed to achieve a quantitative functional activity for different RNA components will allow for the efficient design, construction, and implementation of functional components tailored for various networks and systems.

1.3.3 Evolving new RNA component functions

While many functional RNA components can be harvested from natural biological systems and then refined and enhanced through rational design and evolutionary methods, such strategies can be limited to generating functions that are closely-related to the native activities. To generate new RNA component functions *de novo,* researchers have turned to *in vitro* selection strategies. The ability to readily interconvert between genetic information (DNA) and functional molecule (RNA) through transcription and reverse transcription (RT) processes and the ability to readily amplify DNA through polymerase chain reaction (PCR) processes allows efficient sampling of large RNA sequence space *in vitro*. Thus, large RNA libraries can be searched for rare functional sequences, which can subsequently be recovered, amplified, and searched again.

RNA presents a unique advantage over protein as a design substrate for control devices, as new sensing functions can be generated *de novo* through an *in vitro* selection strategy, systematic evolution of ligands by exponential enrichment (SELEX) (38,39). SELEX typically starts with an initial RNA library of ~10^{14}-10^{15} molecules, each

composed of a randomized region spanning ~30-70 nts flanked by constant sequences. Ligand-binding sequences are isolated by partitioning the RNA library through any of a number of different strategies, although affinity chromatography-based methods are most commonly used. The recovered sequences are subsequently amplified through RT-PCR and used as the starting library for the next round of selection. Iterative rounds of selection are performed, and the selection stringency and counterselections can be tailored to enrich for RNA sequences with high affinities and specificities to the target ligand. As one example, by incorporating counterselections against caffeine in the later selection rounds, an aptamer was selected to theophylline that exhibits a 10,000-fold lower affinity for caffeine, which differs from theophylline by a single methyl group (40). *In vitro* selection strategies have also been applied to the generation of novel ribozymes capable of RNA ligation (41).

When utilizing RNA components generated through *in vitro* selection strategies for cellular applications, the activities of the *in vitro* optimized components may not translate directly to the complex cellular environment. Cell-based selection and screening strategies have been used to perform a secondary screen on *in vitro* enriched RNA component libraries by the component function to a measurable gene expression output. In one example, an *in vitro* enriched aptamer pool to a small molecule ligand, atrazine, was screened in *E. coli* through a cell-based motility assay (2). The RNA aptamer library was coupled to an RBS through a randomized linker region, where the resulting device library was linked to a gene that controls cell motility, such that functional aptamer sequences could be recovered from cells exhibiting greatest mobility in the presence of atrazine on solid medium. A similar approach has also been taken by directly inserting an

in vitro enriched aptamer pool to neomycin in the 5’ UTR of a GFP reporter gene in *E. coli* and screening for *in vivo* functional neomycin aptamers (42). The combined *in vitro* and *in vivo* approaches provide a powerful strategy for generating and tailoring new component functions for the cellular environment.

1.4 Design strategies for small molecule-responsive synthetic RNA devices

Construction of RNA control devices generally starts by functionally linking sensor and actuator components to support transmission of information detected by the sensor into regulated activity of the actuator. Two strategies are generally taken in the design of RNA control devices: (i) direct linkage of sensor and actuator components; (ii) linkage of sensor and actuator components through a distinct information transmission component. In the first strategy, the regulatory effect is imparted by the resulting conformational change in the sensor component in the presence of the input signal, which directly affects the activity of the actuator component. In the second strategy, the transmitter component guides secondary structure changes in the sensor and actuator components, which direct these components between active and inactive conformations. Both modular and non-modular design strategies have been adopted in the construction of RNA controllers. Modular device design strategies introduce standardized communication interfaces between the sensor and actuator components to insulate the specific sequences of these components from one another, resulting in design platforms that support the interchange of functional components (i.e., sensor, actuator) without significant device redesign. Utilizing these design approaches researchers have constructed small molecule-responsive RNA control devices by coupling RNA aptamers

(sensors) to various gene-regulatory (actuator) components, whose mechanism of action is generally specific to the gene expression machinery associated with the host cell.

1.4.1 Prokaryotes

Small molecule-responsive control devices have been engineered in bacteria by modifying or adding component functions in natural riboswitch or regulatory elements. In one example, researchers modified the information transmission function encoded in a natural TPP-responsive riboswitch, which exhibits TPP-dependent gene repression, by randomizing a sequence within the riboswitch and screening the resulting library for new synthetic devices that exhibit TPP-dependent gene activation (43). In another example, a theophylline aptamer was coupled to a natural riboswitch through a randomized linker sequence and the resulting library was screened for devices that exhibit logic operations (AND, NAND) by modulating ribosome access to the RBS (44). RNA control devices have also been built by direct integration of RNA sensors into natural regulatory elements. A theophylline aptamer was integrated into a stem in the group I self-splicing intron from the bacteriophage T4 thymidylate synthase gene to construct a theophylline-dependent RNA splicing device that regulated *E. coli* growth (45).

The majority of small molecule-responsive RNA control devices in bacteria have been designed to control translation initiation, due to the relative ease with which ribosome loading can be modulated through structural rearrangement of the RBS. For example, RBS-based devices were built in *E. coli* by coupling a theophylline aptamer to the RBS through a linker sequence capable of structural rearrangement through strand-displacement (46) or helix-slipping (47) mechanisms. The accessibility of the RBS

(related to its single-stranded state) is altered by ligand-dependent changes in base pairing interactions between the sensor and linker (strand-displacement) or local nucleotide shifts within the linker (helix-slipping). In the device design strategy based on strand displacement, part of the aptamer sequence directly interacts with the RBS sequence, such that independent modification of either component may disrupt these interactions and thus require redesign of the linker sequence to maintain device function. Therefore, cell-based screening strategies have been developed to support the identification of new linker sequences through colorimetric (48), flow cytometry (49), and motility assays (50). Other examples of RBS-based devices were designed by coupling a theophylline aptamer to an indirect actuator, a hammerhead ribozyme, which in turn was coupled to a direct actuator, a RBS (51,52). In this device design, the RBS is sequestered within the ribozyme stem, such that ligand-induced ribozyme cleavage results in unwinding of the ribozyme stem, thereby increasing ribosomal access to the RBS and thus gene expression levels.

1.4.2 Yeast

Different design approaches have been taken to engineer small molecule-responsive RNA control devices in eukaryotes, as fewer examples of natural riboswitch elements are available to build from and more sophisticated gene-regulatory mechanisms are available than in prokaryotes. RNA control devices have been built by integrating the RNA sensor directly into part of the actuator component, such that the conformational change associated with ligand binding to the aptamer affects the activity of the actuator. In one example, a tetramethylrosamine (TMR) aptamer and a synthetic RNA

transcriptional activator were joined through the stem of the transcription activator, where part of the stem was randomized to allow screening for TMR-responsive gene-regulatory activity in *Saccharomyces cerevisiae* (53). In another example, a theophylline aptamer was integrated into the core cleavage region of a synthetic RNase III hairpin, which when placed in the 3' UTR of a target transcript, directed cleavage and subsequent inactivation of that transcript in yeast (31). However, the binding of theophylline to the integrated aptamer restricts cleavage activity of the RNase III enzyme, thereby activating gene expression. Other small molecule-responsive RNA control devices have been built by integrating an aptamer in proximity to a 5' splice site (54) or a branch point (55). As an example, a tetracycline aptamer in close proximity to a 5' splice site, where the consensus sequence of the splice site is integrated into the aptamer stem (54). The binding of tetracycline to the aptamer results in a conformation that potentially prevents access to the splice site, thereby achieving ligand-dependent regulation of splicing.

As a second design approach, composition frameworks have been developed to support the modular assembly of RNA control devices from underlying functional components in yeast (56). In the proposed framework, ribozyme-based devices were constructed by linking an aptamer to a hammerhead ribozyme through a distinct transmitter sequence designed to both encode the information processing function of the device and insulate the sensor and actuator components. This design strategy was demonstrated to support independent swapping of the sensor and transmitter components and thus tailoring of the encoded sensing and information processing functions without significant device redesign. Rational design strategies have been developed that utilize RNA structure prediction programs to guide sequence changes in the transmitter

component to alter the energetic values associated with alternate device conformations, thereby changing the energetic partitioning between functional states of the device and the resulting device regulatory activities (56,57). The composition framework also supported extension to the modular assembly of multiple sensor, transmitter, and actuator domains to construct devices that exhibit higher-order information processing functions, including logic operations, signal and bandpass filters, and programmed cooperativity (57). The utilization of ribozymes as actuators in this RNA device platform also allows for transportability of the device across different organisms, because the ribozyme activity is independent of cell-specific gene-regulatory machinery (3).

1.4.3 Mammalian cells

RNA control devices in mammalian cells have been designed through numerous strategies ranging from directly adapting regulatory mechanisms that were developed in simpler organisms to taking advantage of the rich complexity of RNA processing pathways unique to higher organisms. Several RNA control devices have been developed through design strategies that directly integrate an RNA aptamer to a gene-regulatory element. These approaches have been applied to the design of small molecule-responsive RNA control devices that conditionally silence target genes through RNAi-mediated silencing mechanisms (58,59). These devices have been designed by integrating an RNA aptamer directly within the basal segments of a miRNA or in the loop region of a shRNA, such that binding of the input molecule prevents correct biogenesis of the RNAi substrate and thus results in increased target gene expression levels. Conditional RNAi silencing in response to a wide range of ligands, including theophylline, tetracycline, and xanthine,

have been demonstrated with these RNAi-based devices. While direct integration strategies generally limit the tuning of device regulatory activity that one can achieve through molecular alteration strategies, researchers demonstrated a tuning strategy for the miRNA-based device based on rational design of device clusters (59).

Other mammalian RNA control devices have been developed through design strategies that couple the sensor and actuator components through a distinct transmitter component. In one example, a transmitter and RNA aptamer were integrated into the loop region of a shRNA (60). In the active ligand-unbound conformation, the aptamer was incorrectly formed and the RNAi machinery correctly processed the device, resulting in silencing of the target gene. The inactive conformation of the device coincided with correct formation of the aptamer, such that binding of the small molecule input to the aptamer stabilized this conformation. The RNAi machinery did not correctly process this inactive conformation, thus resulting in increased target protein levels in the presence of input. As another example, the previously-described ribozyme-based devices, originally demonstrated in yeast, were shown to retain small molecule-responsive gene-regulatory activity in mammalian cell culture and *in vivo* (3). RNA devices that incorporate a distinct transmitter component can generally be tuned by altering the sequence of the transmitter component to optimize the thermodynamics of the competing hybridization reactions underlying the strand-displacement mechanism. For both the shRNA- and ribozyme-based devices, sequence alterations within the transmitter components were used to optimize the regulatory performance of the resulting devices in mammalian cells (3,60).

1.5 Broad applicability of the small molecule-responsive synthetic RNA devices

Synthetic RNA control devices hold much promise for leading to transformative advances in how we interact with and program biology, providing access to otherwise inaccessible information on cellular state, and allowing sophisticated exogenous and embedded control over cellular functions. Small molecule-responsive synthetic RNA devices have been integrated into engineered biological systems for applications spanning biosynthesis, bioremediation, to health and medicine (Figure 1.2).

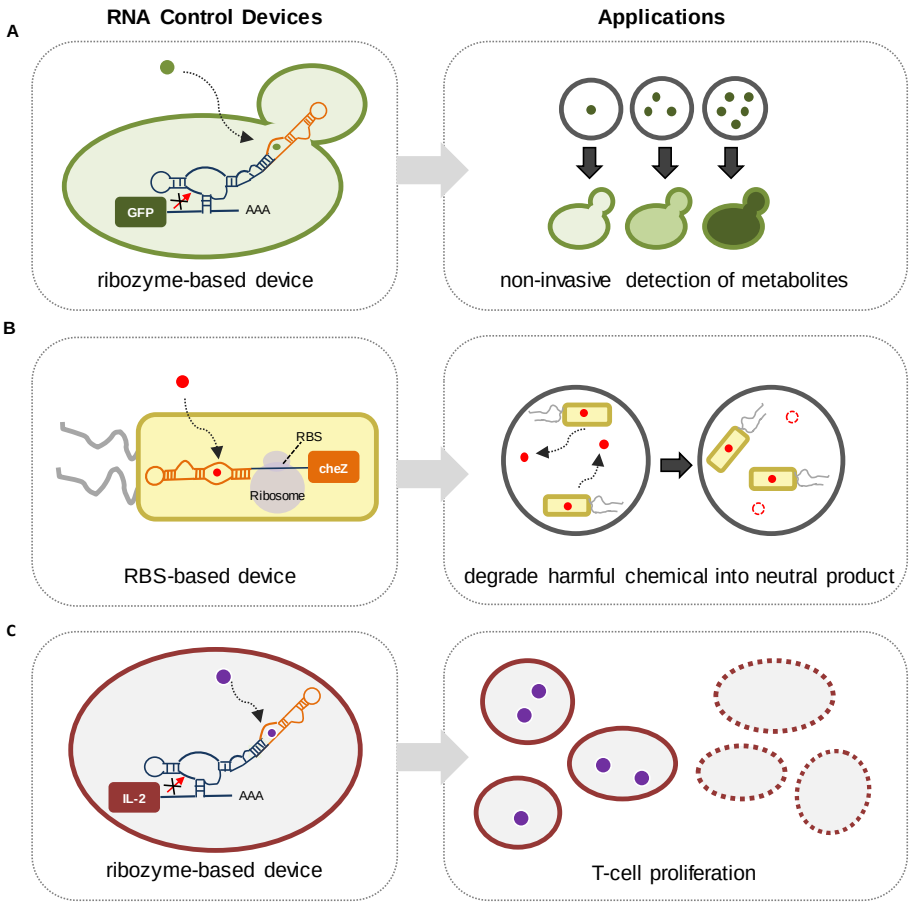


Figure 1.2 RNA-based controllers have been integrated into engineered biological systems for applications spanning biosynthesis, bioremediation, to health and medicine. (A) A metabolite-responsive ribozyme-based device linked to a fluorescent reporter

output was demonstrated in yeast as a noninvasive sensor of metabolite concentration. (B) A pollutant-responsive RBS-based device linked to a motility gene output was demonstrated in bacteria to program the cells to move along a gradient of the pollutant. (C) A small molecule-responsive ribozyme-based device linked to a proliferative cytokine gene output was demonstrated in T cells to control T-cell survival.

1.5.1 Implementation in genetic networks directed to biomanufacturing/biosynthesis applications

Small molecule-responsive synthetic RNA devices have important applications in biosynthesis processes, where they can be implemented as noninvasive sensors of metabolite accumulation and controllers for optimizing flux and product yield. In one example, an RNA control device was utilized for noninvasive detection of metabolite accumulation in yeast (56) (Figure 1.2A). Cells were engineered to express a construct harboring a xanthine-responsive ribozyme-based device regulating a GFP reporter gene. Xanthosine was fed to the yeast cells, which converted this fed substrate to xanthine through an endogenous enzyme activity. The conversion of xanthosine to xanthine was monitored indirectly as an increase in GFP levels, where increases in fluorescence correlated with increased product accumulation as measured by LC-MS.

1.5.2 Implementation in genetic networks directed to environmental/agricultural applications

Small molecule-responsive synthetic RNA devices also have important applications in agricultural biotechnology and environmental remediation. In the latter case, engineered organisms that may be released into the environment will need precise control over designed functions and safety mechanisms in place to prevent uncontrolled release. In one example, an RBS-based RNA control device was implemented to detect a

toxic environmental pollutant, atrazine (2) (Figure 1.2B). The atrazine-responsive device was coupled to the *cheZ* gene to control the motility of an *E. coli* strain engineered to express an atrazine-catabolizing enzyme activity. The presence of atrazine activates the expression the *cheZ* gene, which allows cells to move along the source of the pollutant and convert atrazine into less harmful product, hydroxyatrazine.

1.5.3 Implementation in genetic networks directed to health and medicine application

Finally, small molecule-responsive synthetic RNA devices can be applied to improve the safety or efficacy of cell-based therapeutics. Recent progress has been made in the application of RNA devices to adoptive T-cell therapy, an approach currently in clinical trials, in which a patient's own T cells are harvested and trained against tumor-specific antigens or engineered to express chimeric antigen receptors before being injected back into the body. A major challenge associated with this strategy is ensuring that the T cells can survive and proliferate sufficiently when engrafted into the host to eradicate the diseased cells. To overcome this problem, small molecule-responsive RNA devices were applied to control T-cell proliferation in response to the application of drug molecules (3) (Figure 1.2C). Researchers placed ribozyme-based devices in the 3' UTR of transcripts encoding the proliferative cytokines, IL-15 and IL-2, such that in the absence of the small molecule input the transcript was destabilized through ribozyme cleavage, resulting in reduced T-cell survival. In the presence of the drug input, IL-2 or IL-15 was produced, resulting in increased T-cell proliferation and survival. As a proof of

concept, this platform was shown to be modular in that aptamers to different small molecules (theophylline, tetracycline) were used to control the expression of different proliferative cytokines and hence proliferation of the T cells. The regulatory performance of the RNA-based system was tailored rationally by integrating multiple devices in series to improve the regulatory stringency. Currently, RNA sensors to clinically relevant molecules are needed to advance this strategy forward for therapeutic applications.

1.6 Challenges for the scalable construction and programming of RNA devices

The broad implementation of synthetic RNA control devices in genetic networks will require enabling technologies that support scalable generation of new functional RNA components for the construction of devices and quantitative tailoring of device regulatory performance. Currently, the scalable construction of RNA devices is limited by a small number of existing functional RNA components, in particular components that exhibit sensing functions to desired molecular inputs for cellular engineering applications. While advances have been made in the generation of protein aptamers by utilizing high-throughput, high-efficiency, solution-based partitioning methods, such as capillary electrophoresis (61), the generation of small molecule aptamers still relies on low-throughput, inefficient, matrix-based partitioning methods (38,39), thus limiting our ability to generate new aptamers to diverse molecular ligands.

Another challenge in the field is the reliable construction of devices from components to achieve desired regulatory performance for specific cellular engineering applications. The *in vitro*-generated components need to be linked to other functional component in order to construct synthetic control devices. The regulatory performance of

the resultant devices depends both on the component linkage strategy and the activities of individual components within the host cell. Therefore, the component activities may require further optimization in the context of the device and cellular environment to achieve desired gene-regulatory activities for the targeted applications. Cell-based selection and screening strategies have been developed by linking the gene-regulatory activities of the control devices to a single measurable gene expression output (48-50,62). Although these strategies have been demonstrated for tailoring device regulatory performance, the efficiency of these single-output methods can be negatively impacted by the stochastic nature of gene expression, thereby making these methods less quantitative and efficient.

1.7 Interrelationship among the thesis projects

To allow scalable construction and programming of RNA control devices, this thesis focuses on developing a scalable framework for generating new RNA component functions and quantitative tailoring of the resultant device functions for broad applications within cellular environments. Chapter 1 provides an overview of strategies in generating and optimizing RNA functional components and diverse approaches to the construction of small molecule-responsive synthetic RNA control devices.

We focus primarily on the construction and programming of ribozyme-based devices, as devices utilizing ribozymes as actuator component have the advantage of being portable across different environments (i.e., *in vitro*, *in vivo*). Chapter 2 describes the development of a high-throughput and quantitative two-color FACS-based screen for efficient tailoring of RNA device activity. This screening strategy enables direct

generation and optimization of RNA component functions within the device platform in the cellular environment, which aids in the broader integration of these devices within biological networks. Chapter 3 describes work on an efficient cis-blocking strategy for *in vitro* ribozyme characterization, as our work demonstrated that the *in vitro* cleavage rates associated with these ribozyme-based devices are closely correlated to the corresponding gene-regulatory activities. Chapter 4 describes progress towards the development of a high-throughput, solution-based *in vitro* aptamer selection strategy based on the ribozyme device platform. The successful development of this method will enable the generation of new RNA aptamers to diverse small molecule ligands without the need of conjugation to a solid matrix. Ultimately, this method can be coupled to the two-color FACS-based screen, such that an *in vitro* enriched library can be subjected to a secondary screen for *in vivo* functional activity. Chapter 5 describes the development of a label-free and quantitative method supporting rapid characterization of RNA aptamers to small molecules. Chapter 6 describes the immediate efforts that need to be directed to integrating these enabling technologies into a scalable framework for the engineering of new synthetic devices. Collectively, this thesis describes powerful strategies for tailoring RNA control devices to respond to application-specific small molecule inputs with tailored regulatory properties, thereby enabling their effective implementation within biological networks.

References

1. Ro, D.-K., Paradise, E.M., Ouellet, M., Fisher, K.J., Newman, K.L., Ndungu, J.M., Ho, K.A., Eachus, R.A., Ham, T.S., Kirby, J. *et al.* (2006) Production of the antimalarial drug precursor artemisinic acid in engineered yeast. *Nature*, **440**, 940-943.
2. Sinha, J., Reyes, S.J. and Gallivan, J.P. (2010) Reprogramming bacteria to seek and destroy an herbicide. *Nat Chem Biol*, **6**, 464-470.
3. Chen, Y.Y., Jensen, M.C. and Smolke, C.D. (2010) Genetic control of mammalian T-cell proliferation with synthetic RNA regulatory systems. *Proc Natl Acad Sci U S A*, **107**, 8531-8536.
4. Culler, S.J., Hoff, K.G. and Smolke, C.D. (2010) Reprogramming cellular behavior with RNA controllers responsive to endogenous proteins. *Science*, **330**, 1251-1255.
5. Endy, D. (2005) Foundations for engineering biology. *Nature*, **438**, 449-453.
6. Smolke, C.D. and Silver, P.A. (2011) Informing biological design by integration of systems and synthetic biology. *Cell*, **144**, 855-859.
7. Carr, P.A. and Church, G.M. (2009) Genome engineering. *Nat Biotechnol*, **27**, 1151-1162.
8. Purnick, P.E. and Weiss, R. (2009) The second wave of synthetic biology: from modules to systems. *Nat Rev Mol Cell Biol*, **10**, 410-422.
9. Kozak, M. (2005) Regulation of translation via mRNA structure in prokaryotes and eukaryotes. *Gene*, **361**, 13-37.

10. Yanofsky, C. (1981) Attenuation in the control of expression of bacterial operons. *Nature*, **289**, 751-758.

11. Alifano, P., Bruni, C.B. and Carlomagno, M.S. (1994) Control of mRNA processing and decay in prokaryotes. *Genetica*, **94**, 157-172.

12. Ehretsmann, C.P., Carpousis, A.J. and Krisch, H.M. (1992) mRNA degradation in procaryotes. *FASEB J*, **6**, 3186-3192.

13. Kruger, K., Grabowski, P.J., Zaug, A.J., Sands, J., Gottschling, D.E. and Cech, T.R. (1982) Self-splicing RNA: autoexcision and autocyclization of the ribosomal RNA intervening sequence of Tetrahymena. *Cell*, **31**, 147-157.

14. Serganov, A. and Patel, D.J. (2007) Ribozymes, riboswitches and beyond: regulation of gene expression without proteins. *Nat Rev Genet*, **8**, 776-790.

15. Mandal, M. and Breaker, R.R. (2004) Gene regulation by riboswitches. *Nat Rev Mol Cell Biol*, **5**, 451-463.

16. Gottesman, S. (2004) The small RNA regulators of Escherichia coli: roles and mechanisms*. *Annu Rev Microbiol*, **58**, 303-328.

17. Schluter, J.P., Reinkensmeier, J., Daschkey, S., Evguenieva-Hackenberg, E., Janssen, S., Janicke, S., Becker, J.D., Giegerich, R. and Becker, A. (2010) A genome-wide survey of sRNAs in the symbiotic nitrogen-fixing alpha-proteobacterium Sinorhizobium meliloti. *BMC Genomics*, **11**, 245.

18. Carthew, R.W. and Sontheimer, E.J. (2009) Origins and Mechanisms of miRNAs and siRNAs. *Cell*, **136**, 642-655.

19. Bartel, D.P. (2009) MicroRNAs: target recognition and regulatory functions. *Cell*, **136**, 215-233.

20. Nicoloso, M.S., Spizzo, R., Shimizu, M., Rossi, S. and Calin, G.A. (2009) MicroRNAs--the micro steering wheel of tumour metastases. *Nat Rev Cancer*, **9**, 293-302.

21. Mattick, J.S., Amaral, P.P., Dinger, M.E., Mercer, T.R. and Mehler, M.F. (2009) RNA regulation of epigenetic processes. *Bioessays*, **31**, 51-59.

22. Mercer, T.R., Dinger, M.E. and Mattick, J.S. (2009) Long non-coding RNAs: insights into functions. *Nat Rev Genet*, **10**, 155-159.

23. Hofacker, I.L. (2003) Vienna RNA secondary structure server. *Nucleic Acids Res*, **31**, 3429-3431.

24. Zadeh, J.N., Steenberg, C.D., Bois, J.S., Wolfe, B.R., Pierce, M.B., Khan, A.R., Dirks, R.M. and Pierce, N.A. (2011) NUPACK: Analysis and design of nucleic acid systems. *Journal of computational chemistry*, **32**, 170-173.

25. Mathews, D.H., Disney, M.D., Childs, J.L., Schroeder, S.J., Zuker, M. and Turner, D.H. (2004) Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure. *Proc Natl Acad Sci U S A*, **101**, 7287-7292.

26. Das, R. and Baker, D. (2007) Automated de novo prediction of native-like RNA tertiary structures. *Proc Natl Acad Sci U S A*, **104**, 14664-14669.

27. Parisien, M. and Major, F. (2008) The MC-Fold and MC-Sym pipeline infers RNA structure from sequence data. *Nature*, **452**, 51-55.

28. Flamm, C., Fontana, W., Hofacker, I.L. and Schuster, P. (2000) RNA folding at elementary step resolution. *RNA*, **6**, 325-338.

29. Wieland, M., Benz, A., Klauser, B. and Hartig, J.S. (2009) Artificial ribozyme switches containing natural riboswitch aptamer domains. *Angew Chem Int Ed Engl*, **48**, 2715-2718.

30. Babiskin, A.H. and Smolke, C.D. (2011) A synthetic library of RNA control modules for predictable tuning of gene expression in yeast. *Mol Syst Biol*, **7**, 471.

31. Babiskin, A.H. and Smolke, C.D. (2011) Engineering ligand-responsive RNA controllers in yeast through the assembly of RNase III tuning modules. *Nucleic Acids Res*, **39**, 5299-5311.

32. Chen, X., Denison, L., Levy, M. and Ellington, A.D. (2009) Direct selection for ribozyme cleavage activity in cells. *RNA*, **15**, 2035-2045.

33. Pfleger, B.F., Pitera, D.J., Smolke, C.D. and Keasling, J.D. (2006) Combinatorial engineering of intergenic regions in operons tunes expression of multiple genes. *Nat Biotechnol*, **24**, 1027-1032.

34. Smolke, C.D., Carrier, T.A. and Keasling, J.D. (2000) Coordinated, differential expression of two genes through directed mRNA cleavage and stabilization by secondary structures. *Appl Environ Microbiol*, **66**, 5399-5405.

35. Anderson, J.C., Clarke, E.J., Arkin, A.P. and Voigt, C.A. (2006) Environmentally controlled invasion of cancer cells by engineered bacteria. *J Mol Biol*, **355**, 619-627.

36. Dixon, N., Duncan, J.N., Geerlings, T., Dunstan, M.S., McCarthy, J.E., Leys, D. and Micklefield, J. (2010) Reengineering orthogonally selective riboswitches. *Proc Natl Acad Sci U S A*, **107**, 2830-2835.

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

38. Ellington, A.D. and Szostak, J.W. (1990) In vitro selection of RNA molecules that bind specific ligands. *Nature*, **346**, 818-822.

39. Tuerk, C. and Gold, L. (1990) Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science*, **249**, 505-510.

40. Jenison, R.D., Gill, S.C., Pardi, A. and Polisky, B. (1994) High-resolution molecular discrimination by RNA. *Science*, **263**, 1425-1429.

41. Bartel, D.P. and Szostak, J.W. (1993) Isolation of new ribozymes from a large pool of random sequences [see comment]. *Science*, **261**, 1411-1418.

42. Weigand, J.E., Sanchez, M., Gunnesch, E.B., Zeiher, S., Schroeder, R. and Suess, B. (2008) Screening for engineered neomycin riboswitches that control translation initiation. *RNA*, **14**, 89-97.

43. Nomura, Y. and Yokobayashi, Y. (2007) Reengineering a natural riboswitch by dual genetic selection. *J Am Chem Soc*, **129**, 13814-13815.

44. Sharma, V., Nomura, Y. and Yokobayashi, Y. (2008) Engineering complex riboswitch regulation by dual genetic selection. *J Am Chem Soc*, **130**, 16310-16315.

45. Thompson, K.M., Syrett, H.A., Knudsen, S.M. and Ellington, A.D. (2002) Group I aptazymes as genetic regulatory switches. *BMC Biotechnol*, **2**, 21.

46. Desai, S.K. and Gallivan, J.P. (2004) Genetic screens and selections for small molecules based on a synthetic riboswitch that activates protein translation. *J Am Chem Soc*, **126**, 13247-13254.

47. Suess, B., Fink, B., Berens, C., Stentz, R. and Hillen, W. (2004) A theophylline responsive riboswitch based on helix slipping controls gene expression in vivo. *Nucleic Acids Res*, **32**, 1610-1614.

48. Lynch, S.A., Desai, S.K., Sajja, H.K. and Gallivan, J.P. (2007) A high-throughput screen for synthetic riboswitches reveals mechanistic insights into their function. *Chemistry & biology*, **14**, 173-184.

49. Lynch, S.A. and Gallivan, J.P. (2009) A flow cytometry-based screen for synthetic riboswitches. *Nucleic Acids Res*, **37**, 184-192.

50. Topp, S. and Gallivan, J.P. (2008) Random walks to synthetic riboswitches--a high-throughput selection based on cell motility. *Chembiochem : a European journal of chemical biology*, **9**, 210-213.

51. Wieland, M. and Hartig, J.S. (2008) Improved aptazyme design and in vivo screening enable riboswitching in bacteria. *Angew Chem Int Ed Engl*, **47**, 2604-2607.

52. Ogawa, A. and Maeda, M. (2008) An artificial aptazyme-based riboswitch and its cascading system in E. coli. *Chembiochem : a European journal of chemical biology*, **9**, 206-209.

53. Buskirk, A.R., Landrigan, A. and Liu, D.R. (2004) Engineering a ligand-dependent RNA transcriptional activator. *Chemistry & biology*, **11**, 1157-1163.

54. Weigand, J.E. and Suess, B. (2007) Tetracycline aptamer-controlled regulation of pre-mRNA splicing in yeast. *Nucleic Acids Res*, **35**, 4179-4185.

55. Kim, D.S., Gusti, V., Dery, K.J. and Gaur, R.K. (2008) Ligand-induced sequestering of branchpoint sequence allows conditional control of splicing. *BMC Mol Biol*, **9**, 23.

56. Win, M.N. and Smolke, C.D. (2007) A modular and extensible RNA-based gene-regulatory platform for engineering cellular function. *Proc Natl Acad Sci U S A*, **104**, 14283-14288.

57. Win, M.N. and Smolke, C.D. (2008) Higher-order cellular information processing with synthetic RNA devices. *Science*, **322**, 456-460.

58. An, C.I., Trinh, V.B. and Yokobayashi, Y. (2006) Artificial control of gene expression in mammalian cells by modulating RNA interference through aptamer-small molecule interaction. *RNA*, **12**, 710-716.

59. Beisel, C.L., Chen, Y.Y., Culler, S.J., Hoff, K.G. and Smolke, C.D. (2011) Design of small molecule-responsive microRNAs based on structural requirements for Drosha processing. *Nucleic Acids Res*, **39**, 2981-2994.

60. Beisel, C.L., Bayer, T.S., Hoff, K.G. and Smolke, C.D. (2008) Model-guided design of ligand-regulated RNAi for programmable control of gene expression. *Mol Syst Biol*, **4**, 224.

61. Berezovski, M.V., Musheev, M.U., Drabovich, A.P., Jitkova, J.V. and Krylov, S.N. (2006) Non-SELEX: selection of aptamers without intermediate amplification of candidate oligonucleotides. *Nat Protoc*, **1**, 1359-1369.

62. Fowler, C.C., Brown, E.D. and Li, Y. (2008) A FACS-based approach to engineering artificial riboswitches. *Chembiochem : a European journal of chemical biology*, **9**, 1906-1911.

Chapter 2.

A high-throughput, quantitative cell-based screen for efficient tailoring of RNA device activity

Abstract

Recent advances have demonstrated the use of RNA-based control devices to program sophisticated cellular functions; however, the efficiency with which these devices can be quantitatively tailored has limited their broader implementation in cellular networks. Here, we developed a high-efficiency, high-throughput, and quantitative two-color FACS-based screening strategy to support the rapid generation of ribozyme-based control devices with user-specified regulatory activities. The high-efficiency of this screening strategy enabled the isolation of a single functional sequence from a library of over 10^6 variants within two sorting cycles. We demonstrated the versatility of our approach by screening large libraries generated from randomizing individual components within the ribozyme device platform to efficiently isolate new device sequences that exhibit increased catalytic activity up to 6.8-fold and increased activation ratios up to 160%. We also established a correlation between *in vitro* cleavage rates and *in vivo* gene-regulatory activities, supporting the importance of optimizing RNA device activity directly in the cellular environment. We further demonstrated the utility of the generated devices to regulate cell fate determination through a MAPK pathway associated with the mating response in yeast, highlighting our two-color FACS-based screen as an important strategy in quantitatively tailoring genetic control devices for broader integration within biological networks.

2.1 Introduction

The engineering of biological systems that exhibit complex functions has significant potential to develop solutions addressing global challenges spanning health and medicine, sustainability, and the environment (1-4). Core to the engineering of biological systems is the ability to process information within cellular networks and link this information to new cellular behaviors. Synthetic biology is a rapidly growing interdisciplinary field that involves the application of engineering principles to support the scalable construction and design of complex biological systems. One key focus of synthetic biology is to develop engineering frameworks for the reliable construction of genetic control devices that process and transmit information within living systems. Such information processing capabilities will allow researchers to implement diverse cellular control strategies, thus laying a critical foundation for designing genetic systems that exhibit sophisticated biological functions.

RNA control devices represent an important class of synthetic genetic control devices, whose development has been fueled by advances in RNA biology, engineering, and nucleic acid computing (5-9). We recently described a modular ligand-responsive ribozyme-based platform that supports efficient tailoring of RNA device function (10). The ribozyme device platform specifies physical linkages between three functional RNA components: a sensor encoded by an RNA aptamer, an actuator encoded by a hammerhead ribozyme, and a transmitter encoded by a sequence that undergoes a strand-displacement event (Figure 2.1). The dynamic range of an RNA control device, defined as the difference between gene expression levels in the presence and absence of ligand, is commonly described by two key parameters: basal activity and activation ratio. The basal

activity is a measure of regulatory stringency, which is set by the gene expression level in the absence of the input ligand, whereas the activation ratio is a measure of input responsiveness, which is set by the ratio of expression level in the presence to that in the absence of input ligand. The dynamic range exhibited by an RNA control device is important for its effective integration within biological systems. We have previously demonstrated a rational device tuning strategy guided by secondary structure predictions from an RNA folding program to direct the partitioning between the two functional device conformations (10). While we demonstrated the generation of a series of devices that span wide regulatory ranges, this strategy relies on predictions made by RNA folding programs, which base their computations on parameters acquired *in vitro*, and thus do not fully capture the complex RNA folding process *in vivo*. Therefore, an alternative strategy that enables more direct engineering of device dynamic ranges in the cellular environment is desired to increase the efficiency with which RNA devices can be tailored to cellular networks.

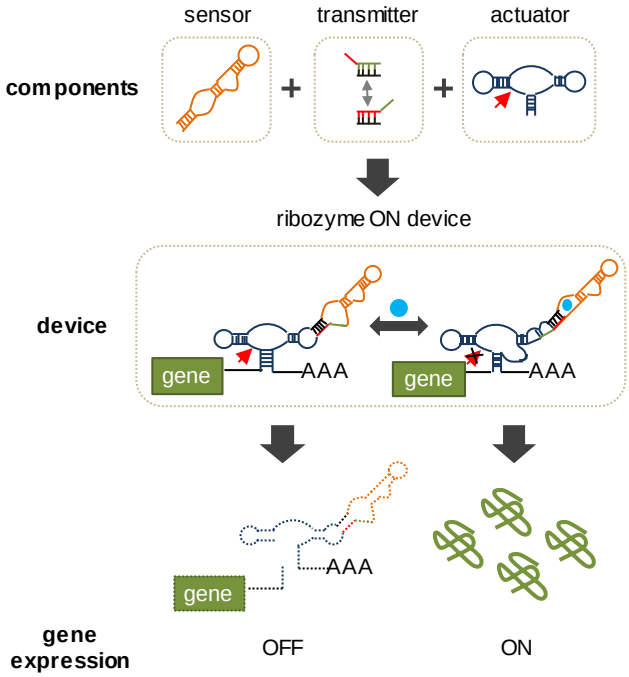


Figure 2.1 Schematic representation of modular assembly and mechanism of an RNA control device based on a ribozyme actuator. Ribozyme-based devices are constructed by modular assembly of three functional RNA components. A sensor (RNA aptamer) is linked to an actuator (hammerhead ribozyme) through a distinct information transmitter sequence (which directs a strand-displacement event and insulates the sensor and actuator components). Ribozyme-based devices are integrated into the 3' untranslated region (UTR) of the target gene and can adopt at least two functional device conformations, where each conformation is associated with different actuator and sensor activities. In the depicted example, a ribozyme ON device (up-regulation of gene expression in response to increased input ligand concentration) adopts a ribozyme-active conformation associated with an aptamer ligand-unbound state, where ribozyme cleavage results in an unprotected transcript that is subject to rapid degradation by ribonucleases, thereby leading to a decrease in gene expression. The ribozyme-inactive conformation is associated with an aptamer ligand-bound state, such that ligand binding to the aptamer stabilizes the ribozyme-inactive conformation, thereby leading to an increase in gene expression in response to ligand.

Significant research efforts have been directed toward developing high-throughput *in vitro* and *in vivo* selection and screening strategies to generate RNA control devices with user-specified regulatory activities. *In vitro* selection strategies were applied to allosteric ribozymes to select for variants that exhibit enhanced ligand-responsive

cleavage activities (11,12). However, in one case the selected variants did not retain their enhanced activities *in vivo*, suggesting that the observed activities were sensitive to the environment in which the selection was performed (11). To address potential loss of activities in transitioning between *in vitro* and *in vivo* environments, researchers have developed cell-based selection and screening strategies to support direct generation of RNA control devices and regulatory components in the relevant environment.

A cell-based selection strategy was developed to select for more active ribozyme variants by directly recovering cleaved ribozyme fragments from cells (13). However, this *in vivo* method decoupled cleavage activity from gene-regulatory activity, resulting in a less quantitative selection strategy that did not support effective identification of sequence variants with varying gene-regulatory activities. Cell-based screening and selection strategies have been developed for RNA control devices by introducing a randomized region into the device and linking the regulatory function of the resulting device library to a single measurable gene expression output, such as colorimetric and fluorescence levels (14-16), motility (17), or viability (18). However, the majority of these methods are most effective at enriching devices from the population exhibiting extreme (either low or high) gene expression levels, limiting the recovery of devices with specific quantitative activities. In addition, the enrichment efficiencies (or, alternatively, the throughput) of screening methods based on a single output can be negatively impacted by the stochastic nature of gene expression. In particular, when utilizing high-throughput methods that measure activities within single cells, such as fluorescence-activated cell sorting (FACS), the inability to resolve changes in gene expression activity

as a result of the genetic control device from cellular noise make single-output strategies less efficient and quantitative.

In this work, we developed a new screening strategy that utilizes two measurable outputs to resolve device regulatory activity from cellular noise. A high-efficiency, high-throughput, and quantitative two-color FACS-based screening strategy was developed to support rapid generation of RNA control devices with user-specified regulatory activities. We validated the high-efficiency of our screening strategy by performing a control sort to isolate the original functional sequence from a library over 10^6 variants within just two sorting cycles. We demonstrated the versatility of our two-color FACS-based approach by screening large libraries generated from randomizing individual components within the ribozyme device platform to efficiently isolate devices with improved regulatory properties. Using this screening strategy we effectively isolated new ribozyme actuator sequences that increase catalytic activity up to 6.8-fold within the device platform and new transmitter sequences that increase activation ratios up to 160%. By generating variants of RNA components with varying quantitative activities, the *in vivo* device screening approach was used to elucidate sequence-function relationships of RNA components and relationships between ribozyme cleavage rates and gene-regulatory activities. Last, we applied a newly generated device from our screens to control the expression of a key component in a MAPK pathway associated with the mating response in yeast. The new device with improved regulatory stringency allowed for robust rerouting of the mating phenotype in response to a small molecule input. These results highlight the utility of our two-color screen in tailoring the quantitative regulatory

properties of RNA devices to support more effective programming of complex cellular behaviors.

2.2 Results

2.2.1 A two-color reporter construct supports a high-efficiency, high-throughput, quantitative, cell-based screening strategy

Genetic control devices with desired regulatory activities are often generated through cell-based screening strategies by coupling the regulatory function to a measurable output. However, the efficiencies of high-throughput cell-based screening strategies based on measuring activities in single cells are negatively impacted by gene expression noise that arises from both extrinsic (e.g., cell size and shape, cell cycle stage, plasmid copy number) and intrinsic (e.g., fluctuation in numbers of DNA, RNA, transcription factors, environmental stimuli) factors (19,20). To develop a more effective and quantitative *in vivo* screening strategy for gene-regulatory devices, we constructed a screening plasmid composed of two independent functional modules (Figure 2.2A). The first module, the device activity reporter, utilizes a GFP reporter to measure the gene expression activity associated with the regulatory device, whereas the second module, the noise reporter, utilizes a mCherry reporter to provide a measure of cellular gene expression activity that is independent of the regulatory device. We used the same constitutive promoter (TEF1) in each module, although other promoter pairs may be substituted into the construct as previous studies in yeast suggest that extrinsic noise, which is the predominant source of gene expression noise, is not promoter specific (20).

Different terminators for each module were used to decrease the frequency of homologous recombination between the two modules.

We first used the two-color screening construct to examine the basal activities of three previously characterized theophylline-responsive ribozyme-based devices (10) based on the output from the device activity reporter (GFP): L2bulge1 (L2b1), 40%; L2bulge5 (L2b5), 70%; and L2bulge8 (L2b8), 11%. The reported basal activities refer to the GFP expression levels relative to that of the inactive ribozyme control in the absence of theophylline. While these three devices exhibit a wide range of regulatory activities, cellular noise results in a broad distribution of GFP fluorescence levels, making it challenging to cleanly isolate members with target activities from a library through a FACS-based sorting strategy based solely on one output (Figure 2.2B, C). To provide a filter for changes in gene expression activity due to cellular noise, we used the noise reporter in the two-color screening construct to assess activity that is independent of the regulatory device. By using the output from the noise reporter (mCherry), we were able to normalize the gene expression activities of individual cells by correlating the outputs between the two reporters and cleanly resolve the cell populations harboring three different devices (Figure 2.2D). For each device, the cell population exhibits a tight linear relationship between the two outputs, such that the device regulatory activity is associated with a distinct slope (Supplementary Figure S2.1). The increased resolution between cell populations harboring devices with different regulatory activities enables the development of a high-efficiency, quantitative, two-color FACS-based screening strategy based on the output correlation between the two reporter modules (Figure 2.2E).

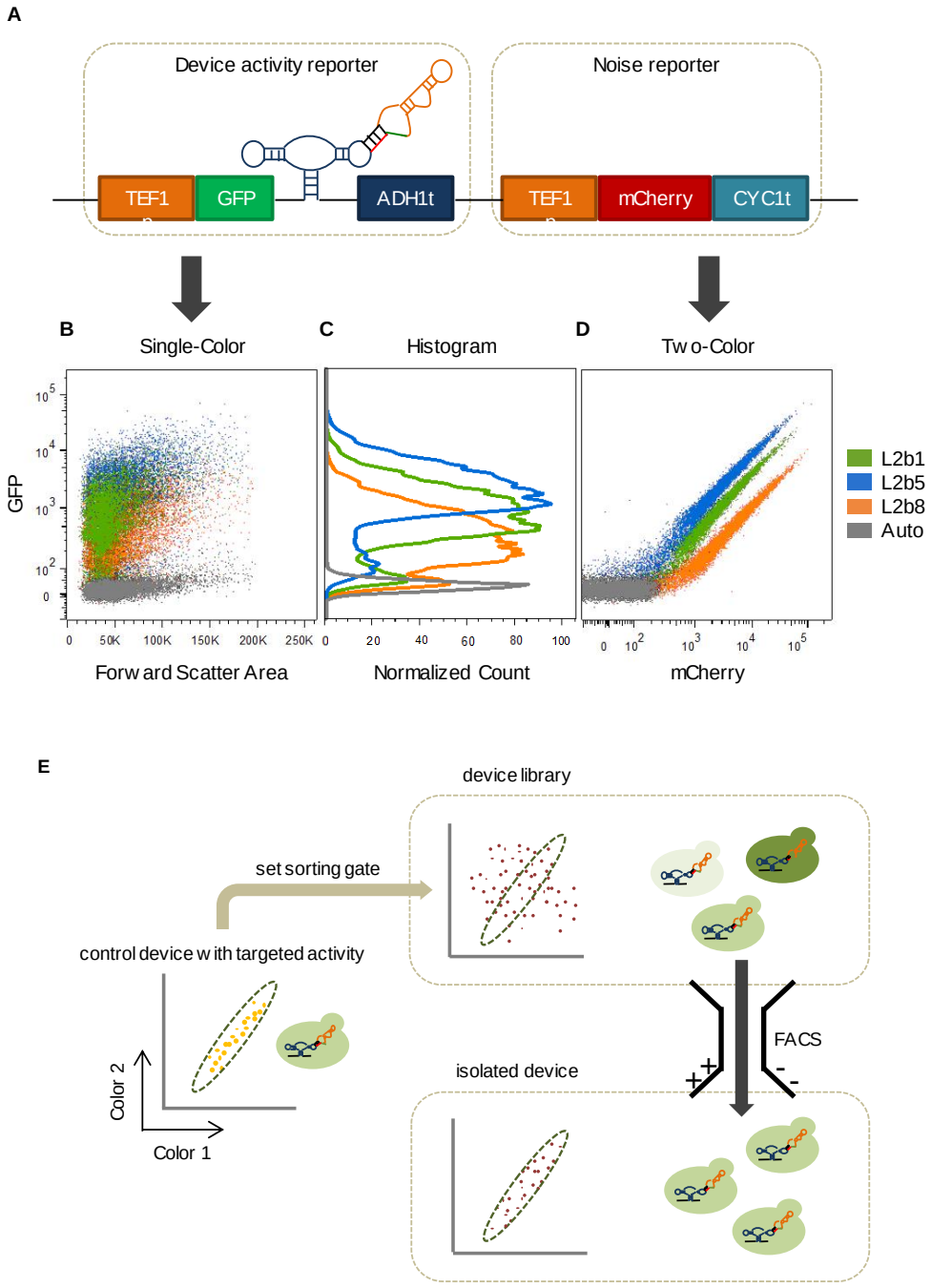


Figure 2.2 A high-efficiency, quantitative cell-based screening strategy for genetic devices based on a two-color screening construct. (A) The two-color screening construct is composed of two independent activity reporters. The device activity reporter measures the gene expression activity associated with the regulatory device from GFP fluorescence, whereas the noise reporter measures the cellular gene expression activity that is independent of the regulatory device from mCherry fluorescence. (B) Single-color (GFP)

scatter plots of three ribozyme-based devices that span a wide range of activities, as measured by their mean values, and cellular autofluorescence from a construct containing no fluorescence reporter gene exhibit significant overlap due to noise associated with gene expression. (C) Single-color (GFP) histograms illustrate that isolation of a device with a specific regulatory activity based on a single reporter output is inefficient due to overlapping population distributions. (D) The gene expression activities of individual cells can be normalized by correlating the device and noise reporter outputs from the two-color screening construct. Cell populations harboring the three ribozyme-based devices in (B) can be cleanly resolved on a two-color scatter plot, where each population exhibits a tight linear relationship between the two outputs. (E) The two-color screening strategy is based on the output correlation between the two reporter modules. A library of control devices can be integrated in the two-color screening construct and transformed into the target cell host. The sorting gate is set by the two-color correlation (slope) associated with a control device that exhibits desired activity and applied to the library to specifically isolate a cell population that exhibits similar activity (slope).

We demonstrated the high enrichment efficiency of our FACS-based two-color screening strategy by performing a control screen on a sensor library (sN10). The sN10 sensor library was generated by randomizing 10 nt positions in the theophylline aptamer binding core (sensor component) within the L2b8 device (~$1x10^6$ variants) (Figure 2.3A). We first assessed the basal activity of the sN10 sensor library through two-color flow cytometry analysis and observed that the majority of the library exhibits higher activity (greater slope) than the parent device (Figure 2.3B). We set a sorting gate based on the basal activity of the parent device (negative gate) and performed a negative sort to enrich cells (~10% of the initial library) that exhibit similar basal activities (same slope). We calculated the fold enrichment based on the percentage of the cell population isolated from each sort. In the initial negative sort, we collected 9.53% of the cell population and therefore this population was enriched ~11 times (100/9.53, or ~11-fold) for the subsequent screen. We performed a positive sort on the recovered cells, in which the lower bound of the sorting gate was set by the activity of the L2b8 device in the presence of 5 mM theophylline. The positive sorting gate was expanded further to enrich members

of the library (~7% of the enriched library) that exhibit equal or increased activities in response to theophylline relative to the parent device by ~15-fold. After one sorting cycle, a distinct small sub-population of cells that exhibit comparable basal activity to the parent device was observed. We then performed another round of sorting to specifically enrich this sub-population (~1% of the enriched library) through a negative sort by ~143-fold, followed by a positive sort on the recovered cells to enrich a clearly distinct cell population that exhibits increased gene expression activity (~4% of the enriched library) in response to theophylline by ~25-fold. After two sorting cycles, ~80% of the enriched library exhibits a clear population shift in response to theophylline (Figure 2.3C).

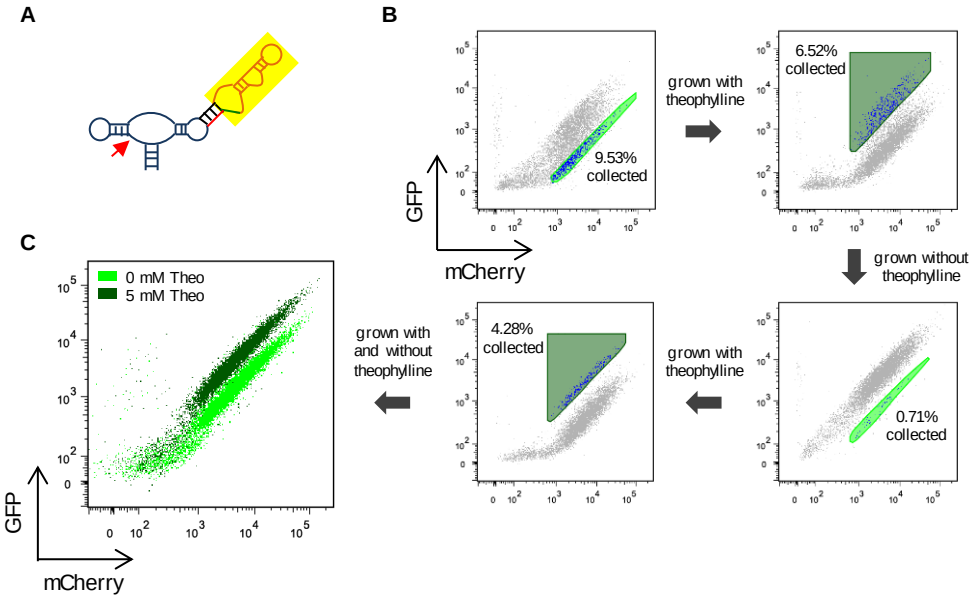


Figure 2.3 Screening of a sensor library within the device platform demonstrates the high enrichment efficiency of the two-color sorting strategy. (A) A sensor library, sN10, is generated by randomizing 10 nts at key positions within the aptamer component in a previously engineered theophylline-responsive ribozyme-based device, L2b8. (B) The sN10 library is subjected to two sorting rounds. Each round consists of one negative sort in the absence of theophylline (light green gate set by the activity of the parent L2b8 device in the absence of theophylline), followed by one positive sort in the presence of theophylline (dark green gate set by the activity of the parent L2b8 device in the presence

of 5 mM theophylline). Percentage of cells collected in the sorting gate is indicated on each plot. (C) After two sorting rounds, ~80% of the enriched sensor library exhibits a clear population shift in response to theophylline (theo).

Since the two rounds of sorting yield an overall ~6×10^5-fold enrichment, given an initial library diversity of ~10^6 variants we expected to recover the original parent device from the enriched library by screening a small number of individual clones. We characterized 48 individual colonies from the enriched library for increased theophylline-responsive activities through flow cytometry. All 48 colonies exhibit similar increases in GFP expression levels in response to theophylline as the parent L2b8 device. We sequenced the recovered devices from 10 colonies, and all were verified to be the parent sequence. These results demonstrate the high efficiency of our two-color FACS-based screening strategy, which can enrich a single sequence from a large ~10^6 library to close to pure isolation in as few as two sorting rounds. The efficiency of our screening strategy is a direct result of the high resolving power of our two-color screening construct. For example, when examining the enriched library after one round of sorting, we could observe a distinct cell population that exhibits comparable basal activity to the parent L2b8 device on a two-color scatter plot (Supplementary Figure S2.2). However, when examining the cell population on a single-color (GFP) histogram, noise associated with the gene expression activity results in poor resolution between the device regulatory activity and cellular noise, thereby significantly decreasing the enrichment efficiency of a screening strategy based on a single output.

2.2.2 Screening of an actuator library results in RNA control devices with improved regulatory stringencies

The implementation of RNA control devices in biological systems requires flexible tailoring of regulatory functions, where regulatory stringency of a device can be an important property for crossing phenotypic thresholds (1). The basal activity of a single ribozyme-based device depends on both the actuator and transmitter components. The catalytic activity of the hammerhead ribozyme (actuator) within the context of the device platform sets a lower limit to the minimal gene expression level a device exhibits in the absence of ligand, whereas the transmitter directs the partitioning between the functional conformations, which in turn impacts the basal activity (10). The device platform specifies the integration of a sensor-transmitter element through loop II (or loop I) of the satellite tobacco ringspot virus (sTRSV) ribozyme actuator. Previous sTRSV ribozyme characterization studies have demonstrated that a tertiary interaction between loop I and loop II is essential for the catalytic activity of the ribozyme in cellular environments (21). The integration of additional structural elements through the ribozyme loops may negatively impact the required loop-loop interaction, thus resulting in a less active ribozyme within the device platform. The previously described theophylline-responsive ribozyme devices (L2b1, 40%; L2b5, 70%; L2b8, 11%) exhibit a basal expression activity higher than that exhibited by the ribozyme alone (1%) (Supplementary Figure S2.3), suggesting that the current natural ribozyme sequence in the device platform may limit our ability to generate RNA control devices that exhibit tighter regulatory stringencies.

We focused initially on optimizing the sTRSV ribozyme sequence in the L2b8 device as it has the lowest basal activity among the series of previously engineered theophylline-responsive devices (10). As prior sTRSV ribozyme characterization studies

have demonstrated that mutations to either loop I or II sequences can enhance or hinder ribozyme cleavage activity (21), we designed a device library by randomizing loop sequences and applied our two-color FACS-based screening approach to search for improved loop-loop interactions that support lower basal activity. We generated an actuator library (aN7) by randomizing 7 nts in loop I within the L2b8 device (~$2x10^4$ variants) (Figure 2.4A). Loop I was targeted for randomization as it is isolated from the conformational change that is facilitated by the transmitter component, which is integrated via loop II. We analyzed the initial aN7 actuator library by two-color flow cytometry analysis and observed that ~99% of the library exhibits a higher basal activity (greater slope) than the parent L2b8 device, indicating that the majority of loop I library sequences have deleterious effects on ribozyme catalytic activity (Figure 2.4B). The two-color approach also enables us to clearly distinguish cells with low gene expression activity from cells that have lost or mutated the plasmid, whereas the two cell populations are almost indistinguishable if GFP is used as the only measure of device regulatory activity (Supplementary Figure S2.3). We set a sorting gate based on cells harboring the parent L2b8 device and collected cells that exhibit a decreased slope relative to the parent device (~0.4% of the initial library) to enrich library members with enhanced loop-loop interactions that support lower device basal activity.

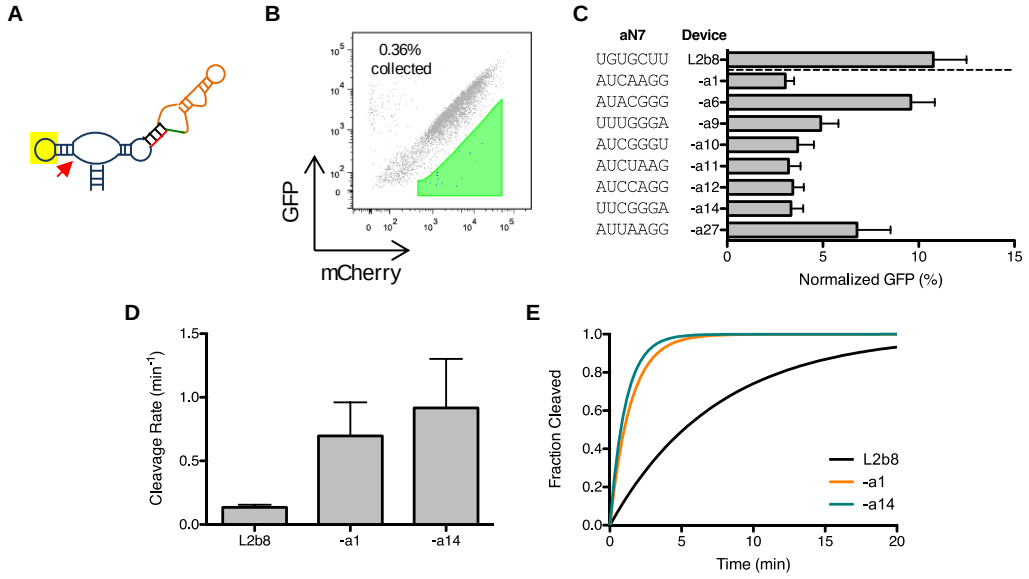


Figure 2.4 Screening of an actuator library within the device platform results in ribozyme variants that exhibit improved gene-regulatory stringencies and cleavage rates. (A) An actuator library, aN7, is generated by randomizing 7 nts at key positions within the loop I region of the ribozyme actuator in the L2b8 device. (B) The aN7 library is subjected to a single sort to enrich for devices that exhibit lower basal gene expression levels than the parent L2b8 device. The majority (~99%) of the aN7 library exhibits a greater slope than that of the parent L2b8 device, such that one sort is sufficient to isolate members that exhibit improved regulatory stringency. (C) Ribozyme variants isolated from the aN7 library screen exhibit lower basal activities relative to the parent L2b8 device. Gene-regulatory activities are reported as the geometric mean of the GFP fluorescence of the indicated sample normalized to that of a positive control (sTRSV Contl, a noncleaving sTRSV ribozyme with a scrambled core) that is grown under identical ligand conditions and is set to 100%. Reported values are the mean and standard deviation of at least three independent experiments. (D) The recovered ribozyme variants (L2b8-al, -a14) exhibit faster cleavage rates than the parent device (L2b8). Cleavage assays were performed at 37˚C with 500 μM $MgCl_2$, 100 mM NaCl, 50 mM Tris-HCl (pH 7.0). Cleavage rate constants (k) and errors are reported as the mean and standard deviation from at least three independent assays. (E) *In vitro* cleavage kinetics of the ribozyme variants (L2b8-al, -a14) and the parent device (L2b8). The projected cleavage kinetics are generated from the single-exponential equation $F_t = F_0 + (F_\infty - F_0)\times(1 - e^{-kt})$, setting the fraction cleaved before the start of the reaction (F_0) and at reaction endpoint (F_∞) to 0 and 1, respectively, and k to the experimentally determined value for each RNA device.

Due to the relatively small sequence space of the library and high-efficiency of our screen, we directly plated the sorted cells to obtain single colonies for further

characterization. We recovered a total of 22 clones and identified 22 unique device sequences from the recovered clones. However, only eight of the recovered devices maintained low basal activities upon re-cloning (Figure 2.4C), while the remaining devices likely had mutations within the recovered plasmid backbone. The basal activities exhibited by the recovered devices are up to 3.5-fold lower than that of the parent device, while five of the devices retained functional switching (i.e., responsiveness to ligand) (Supplementary Figure S2.4).

We performed a sequence analysis to identify the loop I sequence requirements that support high ribozyme catalytic activity within the device platform. A series of point mutations to the recovered loop I sequences were designed, and the activities of these loop I sequences were characterized through flow cytometry analysis (Supplementary Figure S2.5). Two distinct consensus sequences were identified that support lower basal activities than the parent L2b8 device. RNA devices with a loop I heptaloop adhering to the consensus sequence AUNNRRG, where N is any nucleotide base and R is a purine base, exhibit basal activities that are less than or equal to that of the parent (11%). A subset of these devices with the loop I consensus sequence AU(C/U/G)NARG exhibit basal activities less than 80% of that of the L2b8 parent device. Further restricting the loop I consensus sequence to AUCNARG results in additional improvements to the basal activity, to 40% of that of the parent. RNA devices with a second, distinct loop I consensus sequence predicted to form a triloop $N_1UN_2GGGN_1^{|}$, where N_1 and $N_1^{|}$ are any Watson-Crick base pair, exhibit improved basal activities less than 70% of that of the L2b8 parent.

2.2.3 Engineered actuator components result in optimized ribozyme-based devices exhibiting faster cleavage rates

In vitro kinetic analysis of select ribozyme-based devices harboring loop I sequence modifications was performed to assess whether the improved *in vivo* basal activities were a result of increased catalytic rate relative to the parent device. Cleavage rates (*k*) were determined at physiologically-relevant reaction conditions (500 μM $MgCl_2$, 100 mM NaCl, and 50 mM Tris-HCl (pH 7.0)) at 37°C, where the submillimolar magnesium concentration is within the range estimated for intracellular magnesium concentration (22). Cleavage rate constants were obtained for the L2b8 parent device and the L2b8-a1 and -a14 engineered devices (Figure 2.4D, E) by quantifying relative levels of full-length and cleaved radiolabeled transcripts through polyacrylamide gel electrophoresis analysis (Supplementary Figure S2.6). Compared to the cleavage rate constant of L2b8 (0.14 min^{-1}), the cleavage rates associated with the engineered devices are increased 5- and 6.8-fold for L2b8-a1 (0.70 min^{-1}) and -a14 (0.92 min^{-1}), respectively. The cleavage rate of the sTRSV hammerhead ribozyme (4.3 min^{-1}) was determined at these conditions to be 45-fold greater than L2b8 (Supplementary Figure S2.6), in agreement with prior analyses on this ribozyme (21) and other well characterized natural ribozymes under physiologically relevant conditions (23). The results indicate that the modified loop I sequences produce ribozyme-based devices exhibiting faster cleavage rates and thus lower basal expression levels.

2.2.4 Screening of transmitter libraries results in RNA devices exhibiting improved activation ratios

The dynamic range of an RNA control device depends on many parameters, including irreversible rate (e.g., ribozyme cleavage activity), intracellular ligand concentration, and mechanism by which binding information at the sensor is transmitted to an activity change in the actuator (e.g., transmitter design). We applied our two-color FACS-based screening strategy to explore a greater functional space to search for new transmitter sequences that support improved activation ratios given the designated sensor and actuator component pairs. We generated two transmitter libraries, tN11 and a1-tN11, based on two theophylline-responsive ribozyme-based devices, L2b8 and L2b8-a1, respectively, with ribozyme components that exhibit varying catalytic activities. Both libraries were generated by randomizing 11 nts in the transmitter component within the corresponding parent device (~$4x10^6$ variants) (Figure 2.5A). We limited the number of randomized nucleotides in the transmitter components to result in libraries that could be completely searched based on the efficiency of our yeast library transformation method (see Materials and Methods). Both libraries were screened to identify devices with comparable basal activities but with greater activation ratios relative to the corresponding parent devices (Figure 2.5B, D). We first performed a negative sort on both libraries by setting the sorting gates based on the basal activity of the respective parent device. The negative sort enriched members of the library (~3% of the initial tN11 library; ~5% of the initial a1-tN11 library) that exhibit lower or comparable basal activities as the parent device. A subsequent positive sort was performed on the recovered cell populations by setting sorting gates based on the gene expression activity of the respective parent device

in the presence of 5 mM theophylline. To enrich members of the library that exhibit improved activation ratios, cells that exhibit equal or higher gene expression activities (~0.3% of the enriched tN11 library; ~0.4% of the enriched a1-tN11 library) were isolated. One additional positive sort was performed on the enriched tN11 library to further enrich for cells that exhibit activation ratios greater than or equal to the parent device (~24% of the enriched library).

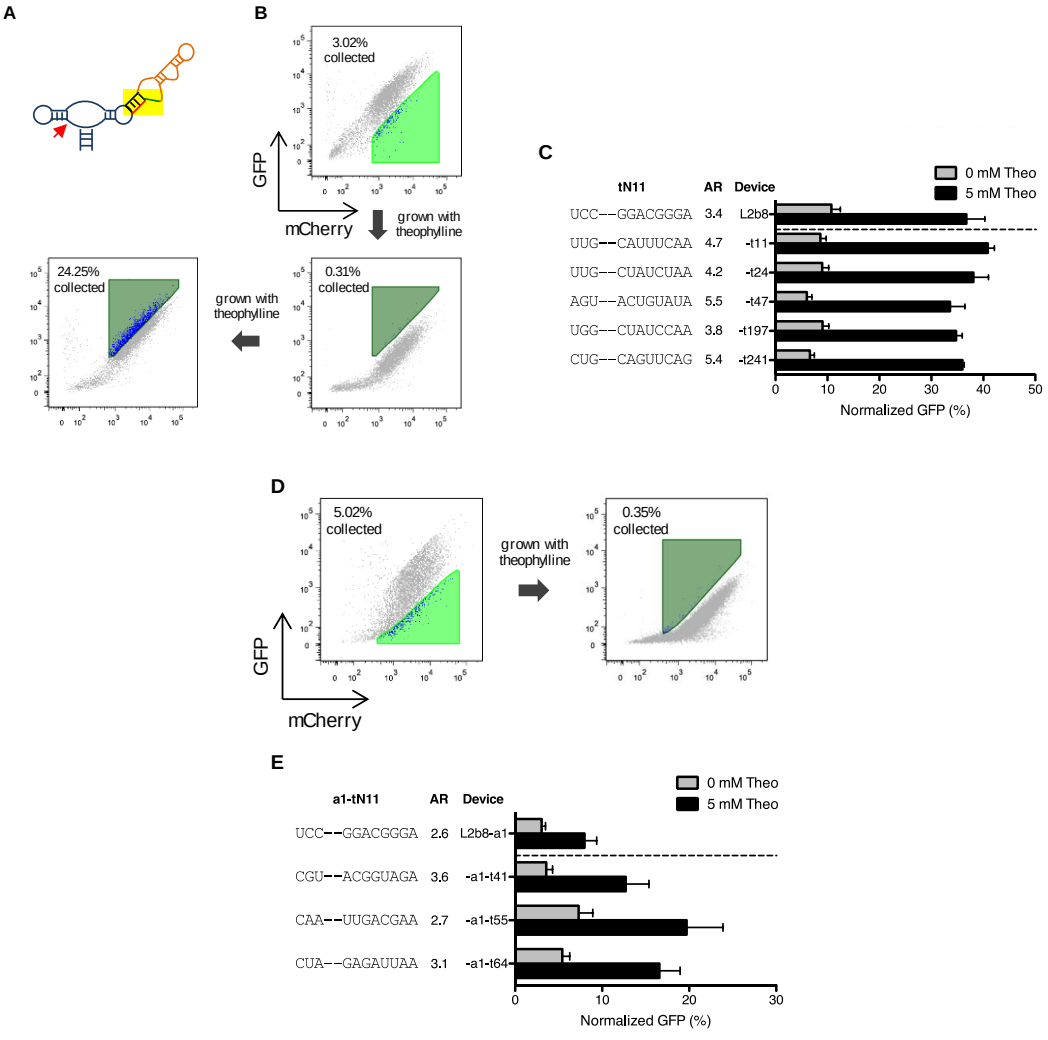


Figure 2.5 Screening of transmitter libraries within the device platform results in transmitter variants that exhibit improved activation ratios. (A) Two transmitter libraries, tN11 and a1-tN11, are generated by randomizing 11 nts within the transmitter components in the L2b8 (wild-type ribozyme actuator) and L2b8-a1 (enhanced ribozyme actuator) devices, respectively. (B) The tN11 library is subjected to one sorting round (negative and positive sort), followed by an additional positive sort to further enrich members of the library that exhibit equal or greater increases in gene-regulatory activities in response to theophylline. The negative (light green) and positive (dark green) sorting gates are set based on the activity of the parent L2b8 device in the absence and presence of 5 mM theophylline, respectively. (C) Transmitter variants isolated from the tN11 library screen exhibit improved activation ratios. Gene-regulatory activities are reported as described in Figure 2.4C. Reported values are the mean and standard deviation of at least three independent experiments. The activation ratio (AR) is determined as the ratio of gene expression levels in the presence and absence of theophylline. (D) The a1-tN11 library is subjected to one sorting round to enrich members of the library that exhibit equal or greater increased in gene-regulatory activities in response to theophylline. The negative (light green) and positive (dark green) sorting gates are set based on the activity of the parent L2b8-a1 device in the absence and presence of 5 mM theophylline, respectively. (E) Transmitter variants isolated from the a1-tN11 library screen exhibit improved activation ratios. Gene-regulatory activities are reported as described in Figure 2.4C. Reported values are the mean and standard deviation of at least three independent experiments.

We plated the recovered cells from both final enriched libraries and screened 287 and 207 individual colonies from the tN11 and a1-tN11 libraries, respectively, through flow cytometry in the absence and presence of 5 mM theophylline and verified improved device activation ratios by re-cloning. We identified a total of five (t11, t24, t47, t197, t241) and three (a1-t41, a1-t55, a1-t64) transmitter variants from the tN11 and a1-tN11 library screen, respectively, that exhibit moderately improved activation ratios up to 160% relative to the parent devices (Figure 2.5C, E). We speculated that the intracellular theophylline concentration in yeast might limit the maximum activation ratio achievable by a device, as previous studies have estimated an 1000-fold drop in theophylline concentration across the *Escherichia coli* cell membrane (24). We tested this hypothesis by characterizing the activation ratios of the transmitter variants at a higher theophylline

concentration (40 mM) and observed devices with activation ratios up to 10.7-fold, corresponding to up to a 290% improvement in activation ratios relative to 5 mM theophylline (Supplementary Figure S2.7, S2.10). No difference in cell viability was observed under all theophylline conditions tested (Supplementary Figure S2.8A), although the higher theophylline concentration resulted in slower cell growth, possibly due to the ligand cytotoxicity, thus potentially rendering the implementation of these devices at higher ligand concentrations less effective for cellular applications.

2.2.5 Modular assembly of optimized actuator components results in devices exhibiting improved regulatory stringencies

The modular composition framework of our ribozyme device platform supports modular assembly of functional RNA components to generate new device functions without significant redesign (10,25). We examined the ability of the optimized ribozyme sequences to be used to build new device regulatory functions, by coupling two of the recovered ribozyme variants (a1, a14) with different sensor and transmitter components. We coupled the new actuator sequences with the transmitter and sensor components from two theophylline-responsive ribozyme ON devices (L2b1, L2b5), one theophylline-responsive OFF device (L2bOFF1), and one tetracycline-responsive ON device (L2b8tc) to generate eight new devices. All newly generated devices exhibit lower basal activities relative to their respective parent, and the majority of them retain switching activity (Supplementary Figure S2.9A-D). The ribozyme variants were also coupled with selected transmitter variants to generate new functional devices with varying regulatory stringencies and activation ratios (Supplementary Figure S2.10). The results indicate that

the improved ribozyme variants are relatively insensitive to the sequence identities of the other device components (transmitter, sensor) and highlight that the modular device platform can be used in combination with improved components to program RNA devices exhibiting desired activities. Finally, we examined the activities of the loop sequences outside of the device context by replacing the wild-type loop I sequence of sTRSV to generate two ribozyme variants (Supplementary Figure S2.9E). The gene expression activities of the resulting ribozyme variants were substantially higher than that of wild-type sTRSV, indicating that the improvement in regulatory stringencies associated with these loop sequences were specific to the context of the device platform, thus supporting the importance of optimizing component functions directly within the device platform.

2.2.6 *In vitro* cleavage kinetics reflect *in vivo* RNA device gene-regulatory activities

The observation that loop I modified ribozyme-based devices exhibit improved *in vivo* basal activities and *in vitro* cleavage rates relative to the L2b8 parent device suggests a correlation between these two measures of activity. However, previous studies had shown that allosteric ribozymes exhibiting enhanced *in vitro* cleavage rates through *in vitro* selections failed to exhibit enhanced *in vivo* gene-regulatory activities (11). Therefore, we examined the relationship between *in vitro* cleavage rates and *in vivo* gene-regulatory activities for a set of RNA devices generated through the two-color screening and rational tuning strategies. Cleavage assays were performed under the previously described physiologically-relevant reaction conditions in the absence and presence of 5 mM theophylline (Supplementary Figure S2.6). The ribozyme-based devices exhibit

cleavage rates that span approximately two orders of magnitude from 0.01 min^{-1} (L2b5) to 1.0 min^{-1} (L2b8-a1, a14) (Figure 2.6A, B). Nearly all of the examined RNA devices exhibit reduced cleavage rates in the presence of theophylline, except for L2b8-a14, which did not exhibit changes in gene-regulatory activity in response to theophylline (Supplementary Figure S2.4). The rank ordering of the examined RNA devices and the sTRSV ribozyme control by *in vitro*-determined cleavage rates matches the inverse ranking by *in vivo* gene-regulatory activities (Figure 2.6C). A correlation analysis between these two measures of activity indicates a strong negative correlation (Pearson coefficient, r = -0.9018) between the cleavage rates 0.007 and 0.16 min^{-1}, where at a cleavage rate of 0.16 min^{-1} near background levels of gene expression are reached such that further decreases in gene-regulatory activity are not observed with faster cleavage rates.

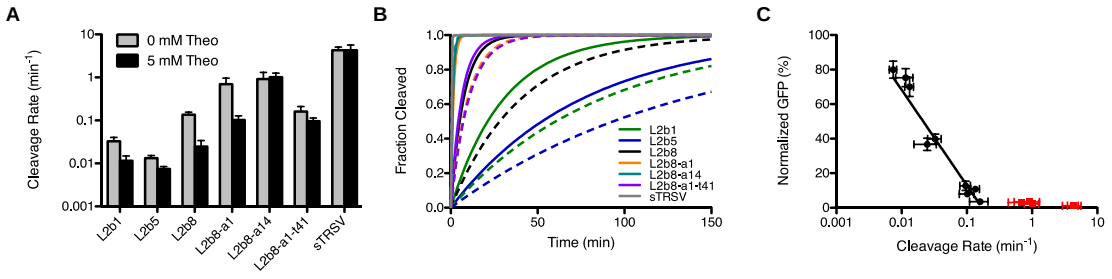


Figure 2.6 *In vitro* cleavage kinetics of selected ribozyme-based devices and controls. Cleavage assays were performed at 37˚C with 500 μM $MgCl_2$, 100 mM NaCl and 50 mM Tris-HCl (pH 7.0) and in the presence of 5 mM theophylline when indicated. (A) Ribozyme-based devices exhibit decreased cleavage rates in the presence of ligand. Cleavage assays were performed as described in Figure 2.4D in the absence or presence of 5 mM theophylline. (B) *In vitro* cleavage kinetics of ribozyme-based devices in the absence and presence of ligand. Projected cleavage kinetics are generated as described in Figure 2.4E. Solid lines: 0 mM theophylline; dashed lines: 5 mM theophylline. (C) Correlation analysis of normalized gene expression levels and cleavage rate constants indicates a strong correlation between the *in vivo* gene-regulatory activity and *in vitro* cleavage rate; Pearson correlation coefficient (r) of -0.9018. A semi-log line is well-fit (R^2 = 0.94) for cleavage rates less than or equal to 0.16 min^{-1} (black data points). Devices excluded from this analysis are indicated in red.

2.2.7 RNA control devices with improved regulatory stringency allow for robust redirecting of yeast mating fates

To achieve robust control of cell fate, the dynamic range of a genetic device must cross the threshold activity of the gene associated with the phenotypic change in the absence and presence of input molecule. RNA control devices have been integrated with biological systems to program complex cellular behaviors, such as proliferation and apoptosis (1,2). In these and other systems, devices with more stringent regulatory activities are necessary to sufficiently reduce background expression to achieve the desired control over the targeted phenotype. From our two-color FACS-based screens, we isolated new device variants that exhibit more than 3.5-fold lower basal activities relative to the parent device. To demonstrate the utility of devices with improved regulatory stringency, we applied these devices to control cell fate determination through a MAPK pathway associated with the mating response in yeast (Figure 2.7A).

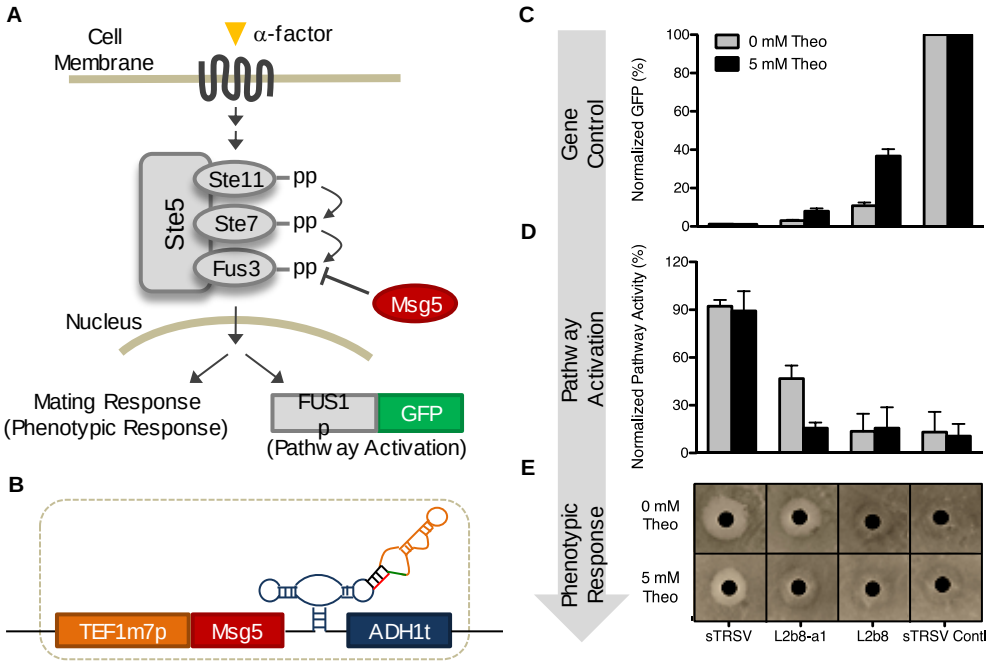


Figure 2.7 Fate routing in the yeast mating pathway with RNA devices. (A) Signaling through the yeast mating pathway starts with pheromone (α-factor) binding to a transmembrane receptor. Receptor binding is transmitted to the scaffold-bound canonical three-tiered MAPK cascade via a phosphorylation relay. Msg5 antagonizes signaling by dephosphorylating Fus3. Translocation of the phosphorylated MAPK, Fus3, to the nucleus initiates transcription at mating genes. Pathway activation is indicated by increased GFP levels (from a transcriptional fusion to the mating-responsive promoter FUS1p) and cell cycle arrest (observed via halo assays). (B) Schematic representation of fate routing constructs, where ribozyme-based devices and controls are placed downstream of a TEF1m7p-Msg5 cassette. (C) Gene-regulatory activities of RNA devices and controls utilized to regulate the yeast mating pathway. Gene-regulatory activities are reported as described in Figure 2.4C. Reported values are the mean and standard deviation of at least three independent experiments. (D) Stringent control of a negative regulator component in the yeast mating pathway is necessary to regulate pathway activation. Normalized pathway activity is reported as the geometric mean of three biological replicates normalized to FUS1p-yEGFP3 expression in the absence of MSG5 expression, and error bars represent the standard deviation of the replicates. (E) Stringent control of a negative regulatory component in the yeast mating pathway is necessary to route cellular fate. Halo formation is a direct read-out of pheromone-induced cell cycle arrest, where reduced halo formation in the presence of pheromone indicates pathway inhibition. Notably, halo formation is inhibited for the systems implementing L2b8 and sTRSV Contl as controllers in the absence or presence of theophylline. In contrast, the system implementing L2b8-a1 exhibits normal halo formation in the absence of theophylline (comparable to sTRSV) and significantly reduces halo formation in the

presence of theophylline. Representative images from three biological replicates are shown.

We targeted a key negative regulator in the mating pathway, Msg5, which is a MAPK phosphatase that inactivates signaling through this pathway by dephosphorylating the MAPK Fus3 (26). We first examined the impact on cell fate by ectopically expressing Msg5 regulated by either the wild-type ribozyme (sTRSV) or the inactive ribozyme (sTRSV Contl) controls (Figure 2.7B, C). These two controls represent the minimal and maximal expression levels of Msg5 in our synthetic system, and the corresponding pathway activation was analyzed through a fluorescent reporter assay that measures transcriptional activation of the pathway and a phenotypic assay that measures cell cycle arrest in the presence of pheromone (Figure 2.7D, E). The low expression level of Msg5 associated with the sTRSV construct allowed pathway activation in the presence of the pheromone, thereby leading to high GFP levels and wild-type halo formation. In contrast, the high expression level of Msg5 associated with the sTRSV Contl construct inhibited pheromone-induced pathway activation, thereby eliminating cell cycle arrest and resulting in low GFP levels.

To redirect yeast mating behavior in response to an exogenously applied molecular input, we implemented two theophylline-responsive ribozyme ON devices, L2b8 and L2b8-a1, to conditionally regulate expression of Msg5 (Figure 2.7B, C). The L2b8 device exhibits the lowest basal expression activity (11%) among the devices previously engineered through a rational design strategy (10). However, the basal expression level of Msg5 associated with the L2b8 device inhibited pathway signaling sufficiently to eliminate cell cycle arrest in the presence of pheromone even in the

absence of theophylline (Figure 2.7D, E). In contrast, the newly selected L2b8-a1 device exhibits a more stringent control profile than that of the L2b8 device and when applied to the regulation of Msg5 achieved rerouting of cell fate in response to the small molecule input. In the absence of theophylline, the basal expression level of Msg5 associated with the L2b8-a1 device was sufficiently low to permit elevated pathway activity in response to pheromone and characteristic halo formation as is observed for the wild-type ribozyme control. In the presence of theophylline, the increased expression level of Msg5 inhibited pathway activation sufficiently to reduce pheromone-induced cell cycle arrest, diminishing the appearance of the halo as is observed for the inactive ribozyme control. These results highlight the importance of tunable controllers, including those exhibiting stringent profiles, in targeting thresholds of key pathway components to achieve desired phenotypic switching. Although the L2b8 device exhibits a greater activation ratio relative to that of the L2b8-a1 device (Supplementary Figure S2.4), the higher basal activity of the L2b8 device rendered it ineffective in redirecting cell fate in this system.

2.3 Discussion

We developed a two-color FACS-based screening approach that enables the efficient isolation of devices with targeted regulatory activities (enriching a single sequence from a large ~10^6 library to close to pure isolation in two sorting rounds), providing a substantial improvement over the enrichment efficiencies of more commonly used single output screening strategies that can be negatively impacted by noise associated with gene expression. We applied our two-color screening approach to optimize the actuator and transmitter components directly within the context of the

ribozyme device platform in the relevant cellular environment, thereby addressing challenges previously encountered with optimizing component activities through *in vitro* strategies (11). The highly quantitative nature of our screening approach allowed us to efficiently isolate ribozyme and transmitter variants that exhibit specified regulatory activities. Our two-color screening approach is also fully complemented by the modular composition framework of the ribozyme-based platform. We demonstrated a plug-and-play device construction strategy by linking components with varying activities to generate new RNA devices that span wide regulatory ranges. In comparison to the previously demonstrated computational-guided device tuning strategy, which is restricted to experimentally sampling a smaller sequence space, our two-color screening approach offers greater flexibility and higher-throughput in quantitatively tailoring device regulatory properties.

We further characterized the isolated ribozyme and transmitter variants in addition to the devices generated previously by rational design through both *in vivo* and *in vitro* assays and found a strong negative correlation between gene-regulatory activities and cleavage rates, for cleavage rates within the range of 0.007 and 0.16 min^{-1} (measured at the specified reaction condition of 500 μM $MgCl_2$). Gene expression levels reach background values at ~0.16 min^{-1}, such that that increases in cleavage rate above this value have minimal impact on expression levels. In addition, our analysis (Figure 2.6C) indicates that below cleavage rates of ~0.007 min^{-1} gene expression levels reach saturating (fully ON) levels, such that further decreases in cleavage rate below this value do not significantly impact gene expression. Our work demonstrates that there is a narrow window of *in vitro* cleavage rates (spanning less than two orders of magnitude) that

correspond to titratable *in vivo* gene-regulatory activities. This observation may explain previous work in the field that observed no changes in gene-regulatory activities between ribozyme variants exhibiting different cleavage rates, if the changes in cleavage rates were occurring outside of the titratable window (13).

Our results also highlight the importance of optimizing individual components within the context of device platforms. The parent actuator component (sTRSV ribozyme) is a type-III hammerhead ribozyme, in which the 5' and 3' ends of viroid RNA extend from stem III. The tertiary interactions essential for cleavage activity at physiological conditions are formed by a Watson-Crick base triple between the 5' U of loop I, next-to-last U of loop I, and the 3' A of loop II. Recent structural analyses have shown that these essential tertiary interactions are conserved across a significant fraction of natural hammerhead ribozymes (27). As the loop I sequence of the originally designed ribozyme-based devices, including the parent device L2b8, is unchanged from the native sTRSV ribozyme (UGUGCUU), it was assumed that aptamer integration into loop II would maintain the same tertiary interactions required for cleavage activity (10). However, both the heptaloop (AUCNARG) and triloop ($N_1UN_2GGGN_1'$) consensus sequences identified through the actuator library screen that result in devices exhibiting lower basal activities than the L2b8 parent device do not follow the common tertiary motif of the sTRSV ribozyme. The divergence of the improved loop I sequences from the L2b8 and sTRSV parents are not necessarily surprising, given recent work identifying thousands of new hammerhead ribozymes in all domains of life with unique tertiary motifs (27-29). In particular, the integration of the transmitter-aptamer element into loop II of the sTRSV hammerhead ribozyme in the device platform makes the overall structure of the

ribozyme-based devices more similar to other, less well-characterized hammerhead ribozymes (Supplementary Figure S2.11). Interestingly, the improved loop I consensus sequences identified in our screen share similarities to these other natural ribozymes and do not exhibit improved activities in the context of the sTRSV ribozyme alone.

Finally, we demonstrated the utility of our newly selected device variants to control the expression level of a key negative regulator in the yeast mating pathway through an exogenously applied molecular input. In this particular system, the regulated pathway component exhibits a low threshold level, thus requiring a regulatory device with low background activity to permit pathway activation in the absence of the effector molecule. We applied one of the selected device variants with improved regulatory stringency to achieve phenotypic switching in response to the exogenous molecular input, whereas the original device was ineffective in redirecting cell fate. These results highlight the importance of fine-tuning regulatory properties of gene-regulatory elements for their successful integration within biological systems. Recent examples have relied on using low-throughput and inefficient single-color screening strategies on individual clones from libraries of control elements, including promoters and RNase cleavage sites, to identify new elements that exhibit varying activities in different organisms (30-32). In contrast, our two-color FACS-based screening approach is based on the correlation between the gene outputs from two independent functional modules. Similar dual-reporter screening plasmid systems can be constructed in other organisms, including bacteria and mammalian cells, to support the efficient generation of new gene control elements, including those acting through transcriptional and posttranscriptional mechanisms. Our strategy offers a high-throughput and high-efficiency alternative to rapidly screen through

diverse libraries of control elements for members exhibiting specified quantitative activities, thereby addressing a key challenge in quantitatively tailoring these control elements to specific genetic systems.

2.4 Materials and Methods

2.4.1 Plasmid and strain construction

Standard molecular biology cloning techniques were used to construct all plasmids (33). DNA synthesis was performed by Integrated DNA Technologies (Coralville, IA) or the Protein and Nucleic Acid Facility (Stanford, CA). All enzymes, including restriction enzymes and ligases, were obtained through New England Biolabs (Ipswich, MA). Ligation products were electroporated with a GenePulser XCell (Bio-Rad, Hercules, CA) into an *E. coli* DH10B strain (Invitrogen, Carlsbad, CA), where cells harboring cloned plasmids were maintained in Luria-Bertani media containing 50 mg/mL ampicillin (EMD Chemicals, Philadelphia, PA). All cloned constructs were sequence verified by Elim Biopharmaceuticals (Hayward, CA).

The two-color screening plasmid pCS1748 (Supplementary Figure S2.12) was constructed by inserting an open reading frame (ORF) encoding a yeast-optimized mCherry gene (*ymCherry*) flanked by a TEF1 promoter and a CYC1 terminator upstream of the existing ORF encoding a yeast-enhanced GFP gene (*yEGFP3*) flanked by a GAL1-10 promoter and an ADH1 terminator in the pCS321 backbone (10). The TEF1 promoter was PCR amplified from pG2M (34) using forward and reverse primers SacI-TEF1-fwd (5’-GAGAGCTCAAGCTTCAAAATGTTTCTACTCC) and SacII-TEF1-rev (5’-GGCCGCGGCAAAACTTAGATTAGATTGCTATGC), respectively, and inserted into

pCS321 via the unique restriction sites SacI and SacII. The *ymCherry* gene was PCR amplified from BBa_E2060 obtained from the iGEM parts registry (35) using forward and reverse primers SacII-mCherry-fwd (5'-GACCGCGGGAAATAATGTCTATGGTTAGTAAAGGAGAAGAAAATAACATGGC) and NotI-mCherry-rev (5'-GGGCGGCCGCTTATTATTTGTATAGTTCATCCATGCCACCAG), respectively, and inserted downstream of the cloned TEF1 promoter via the unique restriction sites SacII and NotI. The CYC1 terminator was PCR amplified from pCM159 (36) using forward and reverse primers NotI-CYC1t-fwd (5'-GAGCGGCCGCGAGGGCCGCATCATGTAATTAG) and XbaI-CYC1t-rev (5'-GGTCTAGAGGCCGCAAATTAAAGCCTTCG), respectively, and inserted downstream of the cloned *ymCherry* gene via the unique restriction sites NotI and XbaI. A spacer sequence was amplified from pCS745 (gift from M. Jensen, Seattle Children's Research Institute) using forward and reverse primers XbaI-Spacer-fwd (5'-GGTCTAGACGCCTTGAGCCTGGCGAACAGTTC) and Spacer-rev (5'-AGTAAAAAGGAGTAGAAACATTTTGAAGCTATCGATGACAGGATGAGGATCGTTTCGCATG), respectively. A TEF1 promoter fragment was amplified from pG2M using forward and reverse primers Spacer-TEF1-fwd (5'-CATGCGAAACGATCCTCATCCTGTCATCGATAGCTTCAAAATGTTTCTACTCCTTTTTTACT) and BamHI-TEF1-rev (5'-GGGGATCCCAAAACTTAGATTAGATTGCTATGCTTTCTTTC), respectively, and PCR assembled with the spacer fragment. The assembled spacer-TEF1 promoter fragment was inserted into the modified

pCS321 backbone to replace the GAL1-10 promoter via the unique restriction sites XbaI and BamHI.

Two single-color plasmids harboring GFP (pCS1585) and mCherry (pCS1749) were constructed as compensation controls for FACS analysis. A fragment harboring the TEF1 promoter was PCR amplified from pGM2 using forward and reverse primers SacI-TEF1-fwd (5'-GAGAGCTCATAGCTTCAAAATGTTTCTACTCC) and EcoRI-TEF1-rev (5'-GGGAATTCT TTGTAATTAAAACTTAGATTAGA), respectively. The GFP-only plasmid pCS1585 was constructed by inserting the TEF1 promoter fragment into pCS321 to replace the GAL1-10 promoter via the unique cloning sites SacI and EcoRI. The mCherry-only plasmid pCS1749 was constructed by inserting the TEF1p-ymCherry–CYC1t cassette from pCS1748 into a modified version of pRS316 (37) via the unique restriction sites SacI and XbaI. The modified version of pRS316 (pCS4), containing no fluorescence reporter gene, was used as the negative-control construct.

Ribozyme-based devices and appropriate controls were inserted into the 3' untranslated region (UTR) of *yEGFP3* in pCS1748 through appropriate restriction endonuclease and ligation-mediated cloning. DNA fragments encoding the ribozyme-based devices and controls were PCR amplified using forward and reverse primers L1-2-fwd (5'-GACCTAGGAAACAAACAAAGCTGTCACC) and L1-2-rev (5'-GGCTCGAGTTTTTATTTTTCTTTTTGCTGTTTCG), respectively, and inserted into pCS1748 via the unique restriction sites AvrII and XhoI, which are located 3 nts downstream of the *yEGFP3* stop codon. Cloned plasmids were transformed into the budding yeast *Saccharomyces cerevisiae* strain W303α (MAT α his3-11, 15 trp1-1 leu2-3 ura3-1 ade2-1) through a standard lithium acetate method (38). All yeast strains

harboring cloned plasmids were maintained on synthetic complete media with a uracil dropout solution containing 2% dextrose (SC-URA) and grown at 30ºC.

The mating fate routing plasmid pCS2293 was constructed by cloning a TEF1 mutant 7 (TEF1m7) promoter (31), and the coding region for Msg5 into the pCS321 backbone (10). The TEF1m7 promoter was PCR amplified from pCS1142 using forward and reverse primers pTEF7-fwd (5'-AAGAGCTCATAGCTTCAAAATGTCTCT ACTCCTTTTT) and pTEF7-rev (5'-AAAGGATCCAACTTAGATTAGATTGCTATGCTTTCTTTCC), respectively, and inserted into pCS321 via the unique restriction sites SacI and BamHI. The *MSG5* gene was PCR amplified from genomic DNA using forward and reverse primers Msg5.K2-fwd (5'-AAAGGATCCAATTAATAGTGCACATGCAATTTCAC) and Msg5-rev (5'-AAAACCTAGGTTAAGGAAGAAACATCATCTG), respectively, and inserted via the unique restriction sites BamHI and AvrII. Ribozyme-based devices and appropriate controls were inserted into the 3' UTR via the unique restriction sites AvrII and XhoI, located immediately downstream of the *MSG5* stop codon as described previously.

The mating reporter construct (pCS1124) was built by cloning FUS1p, a mating-responsive promoter, into the pCS321 backbone (10). The FUS1 promoter was PCR amplified from pDS71 (39) using forward and reverse primers pDS71.Fus1-fwd (5'-TTTGCGGCCGCCCAATCTCAGAGGCTGAGTCT) and pDS71.Fus1-rev (5'-TTTGGATCCTTTGATTTTCAGAAACTTGATGGC), respectively, and inserted upstream of *yEGFP3* in pCS321 via the unique restriction sites NotI and BamHI. The FUS1p-yEGFP3-ADH1t cassette from pCS1124 was cloned into pCS1391, a loxP integrating vector (40), via the unique restriction sites SacI and KpnI to make pCS2292.

FUS1p-yEGFP3-ADH1t was chromosomally integrated into yeast strain EY1119 (W303a Δsst1 Δkss1::HIS3) (41) via homologous recombination using the gene cassette from pCS2292 to construct yeast strain CSY840 (W303a Δsst1 Δkss1::HIS3 trp1::FUS1p-yEGFP3-ADH1t-loxP-KanR). Briefly, the FUS1p-yEGFP3-loxP-KanR cassette was PCR amplified using Expand High Fidelity PCR system (Roche, Indianapolis, IN) from pCS2292 using forward and reverse primers TRP1.INT.all.fwd (5'-GTATACGTGATTAAGCACACAAAGGCAGCTTGGAGTATGTCTGTTATTAATTTCACAGGAAGATTGTACTGAGAGTGCAC) and TRP1.INT.all.rev (5'-TTGCTTTTCAAAAGGCCTGCAGGCAAGTGCACAAACAATACTTAAATAAATACTACTCAGCGACTCACTATAGGGAGACC), respectively, each carrying 60 nts of homologous sequence to the *TRP1* locus. Yeast strain EY1119 was transformed with 12 μg of gel purified PCR product and plated on G418 plates to build yeast strain CSY840. All plasmids and yeast strains constructed in this study are summarized in Supplementary Table S2.1.

2.4.2 Library-scale yeast transformation

Device libraries (Supplementary Figure S2.13) were amplified using forward and reverse primers GAP-L1-2-fwd (5'-TCCATGGTATGGATGAATTGTACAAATAAAGCCTAGGAAACAAACAAAGCTGTCACC) and GAP-L1-2-rev (5'-AAGAAATTCGCTTATTTAGAAGTGGCGCGCCCTCTCGAGTTTTTATTTTTCTTTTTGCTGTTTCG), respectively. The library was inserted into pCS1748 through homologous recombination-mediated gap-repair during

transformation into yeast strain W303 (42). Briefly, an 800 μL library PCR reaction was performed with 160 pmol of each primer and 16 pmol of the library template. 8 μg of the plasmid pCS1748 was digested with AvrII and XhoI. The digested vector was combined with the library PCR product, extracted with phenol chloroform, and precipitated into a dry pellet with ethanol. A Tris-DTT buffer (2.5 M DTT, 1 M Tris, pH 8.0) was added to a 50 mL yeast culture (OD_{600} 1.3-1.5) and incubated at 30ºC for 10-15 min. The yeast were pelleted, washed with chilled Buffer E (10 mM Tris, pH 7.5, 2 mM $MgCl_2$, 270 mM sucrose), and resuspended in Buffer E to a final volume of 300 μL. The yeast mixture was directly added to the precipitated DNA pellet and 50 μL of the mixture was transferred to a chilled 2 mm gap cuvette for electroporation (540 V, 25 μF, 1000 Ω). Following transformation, the cells were resuspended in 1 mL warmed YPD media and incubated at 30ºC for 1 hr. Multiple transformations (~5) were performed to cover the desired library diversity (~10^6-10^7). Transformation efficiencies were determined by plating serial dilutions of the transformants, and transformed cells were propagated in SC-URA media for FACS.

2.4.3 Two-color FACS-based screen

Cells harboring the RNA device libraries and control constructs were washed, resuspended in FACS buffer (1% BSA in PBS), and stained with a DAPI viability dye (Invitrogen). The cell suspension was filtered through a 40 μM cell strainer (BD Systems, San Jose, CA) prior to analysis on a FACSAria II cell sorter (BD Systems). GFP was excited at 488 nm and measured with a splitter of 505 nm and bandpass filter of 525/50 nm. mCherry was excited at 532 nm and measured with a splitter of 600 nm and a

bandpass filter of 610/20 nm. DAPI was excited at 355 nm and measured with a bandpass filter of 450/50 nm. A scatter gate was set based on the forward and side-scatter area of cells harboring the negative-control plasmid (pCS4) to exclude debris, followed by a DAPI-(-) viability gate to exclude dead cells in the DAPI-(+) gate from the analysis. Cells harboring the single-color control plasmids (pCS1585, pCS1749) were analyzed to compensate spillover from GFP to the mCherry detector. A general sorting strategy was followed in which cells harboring devices with targeted activities were analyzed to set a sorting gate on a two-dimensional scatter plot that correlates GFP and mCherry fluorescence. Cells within this gate were collected into SC-URA media and propagated to sufficient density for further screening or analysis.

2.4.4 Ribozyme-based device characterization through flow cytometry analysis

Enriched device libraries from FACS were directly plated onto SC-URA solid medium. Individual colonies were screened and characterized for gene-regulatory activity of the devices based on flow cytometry analysis. The GFP fluorescence was measured on a Quanta flow cytometer (Beckman Coulter, Fullerton, CA). GFP was excited at 488 nm and measured with a splitter of 488 nm and a bandpass filter of 525/40 nm. Cells harboring the negative-control plasmid (pCS4) were analyzed to set the GFP-(-) and GFP-(+) gate. Activities were determined as the geometric mean of the GFP fluorescence based on the GFP-(+) population using FlowJo (Tree Star), and normalized to the geometric mean of the GFP fluorescence of a positive control (sTRSV Contl, a noncleaving sTRSV ribozyme with a scrambled core) that is grown under identical ligand conditions, run in the same experiment, and set to 100%.

Devices that exhibited desired activities were amplified by colony PCR using forward and reverse primers CS653 (5'-GGTCACAAATTGGAATACAACTATAACTCT) and CS654 (5'-CGGAATTAACCCTCACTAAAGGG), respectively, and sequenced. The recovered devices were resynthesized and recloned into the vector backbone to confirm the observed activity. DNA oligos were synthesized and amplified for insertion into pCS1748 using forward and reverse primers L1-2-fwd (5'-GACCTAGGAAACAAACAAAGCTGTCACC) and L1-2-rev (5'- GGCT CGAGTTTTTATTTTTCTTTTTGCTGTTTCG), respectively. The resynthesized devices were inserted into pCS1748 via the unique restriction sites AvrII and XhoI. The reconstructed device plasmids were transformed into the W303 yeast strain through a standard lithium acetate method (38). Cells harboring the selected devices and appropriate controls were prepared as described above for the sorting experiments and analyzed on an LSRII flow cytometer (BD Systems) to characterize the gene-regulatory activity of each device. GFP was excited at 488 nm and measured with a splitter of 505 nm and a bandpass filter of 525/50. mCherry was excited at 532 nm and measured with a splitter of 600 nm LP and a bandpass filter of 610/20 nm. DAPI was excited at 405 nm and measured with a bandpass filter of 450/50 nm. FlowJo was used to process all flow cytometry data. Cells harboring pCS4 and pCS1749 were analyzed to set the mCherry-(-) and mCherry-(+) gates. Activities in the absence or presence of ligand were determined as the geometric mean of the GFP fluorescence based on the mCherry-(+) population, and normalized to the geometric mean of the GFP fluorescence of a positive control (sTRSV Contl, a noncleaving sTRSV ribozyme with a scrambled core) in the absence or presence

of ligand, respectively, to correct for any non-specific effects of ligand on the measured fluorescence (Supplementary Figure S2.8). Reported device activities are the mean and standard deviation of at least three independent experiments.

2.4.5 *In vitro* transcription and purification of ribozyme-based devices

Selected ribozyme-based devices and ribozyme and noncleaving controls were PCR amplified to include an upstream T7 promoter site and spacer sequence and downstream spacer sequence using forward and reverse primers T7-L1-2-fwd (5'-TTCTAATACGACTCACTATAGGGACCTAGGAAACAAACAAAGCTGTCACC) and L1-2-rev (5'-GGCTCGAGTTTTTATTTTTCTTTTTGCTGTTTCG), respectively. A total of 1-2 μg of PCR product was transcribed in a 25 μl reaction, consisting of the following components: 1×RNA Pol Reaction Buffer (New England Biolabs), 2.5 mM rATP, 2.5 mM rCTP, 2.5 mM rUTP, 0.25 mM rGTP, 1 μl RNaseOUT (Invitrogen), 10 mM $MgCl_2$, 1 μl T7 Polymerase (New England Biolabs), and 0.5 μCi α-^{32}P-GTP (MP Biomedicals, Solon, OH). 400 pmol of antisense DNA oligonucleotide, device-blocking (5'-GGTGACAGCTTTGTTTGTTTCCTAGGTCCCCC) and sTRSV-blocking (5'-GCTGTTTCGTCCTCACG), was added to each reaction to inhibit cleavage of the RNA devices and sTRSV hammerhead ribozyme, respectively, during transcription. After incubation at 37°C for 2 hr, NucAway Spin Columns (Ambion, Austin, TX) were used to remove unincorporated nucleotides from the transcription reactions according to manufacturer's instructions. 10 μl aliquots of the recovered RNA were mixed with 3 volumes of RNA stop/load buffer (95% formamide, 30 mM EDTA, 0.25% bromophenol blue, 0.25% xylene cyanol), heated at 95°C for 5 min, snap cooled on ice for 5 min, and

size-fractionated on a denaturing (8.3 M urea) 10% polyacrylamide gel at 25 W for 45 min. Gels were imaged by phosphorimaging analysis on a FX Molecular Imager (Bio-Rad). Uncleaved transcripts were gel extracted and recovered with the ZR Small-RNA™ PAGE Recovery Kit (Zymo Research, Irvine, CA) according to manufacturer's instructions. Samples were stored in sterile, nuclease-free, deionized water at -80°C to limit the extent of RNA self-cleavage prior to performing the cleavage assays.

2.4.6 *In vitro* ribozyme cleavage assays

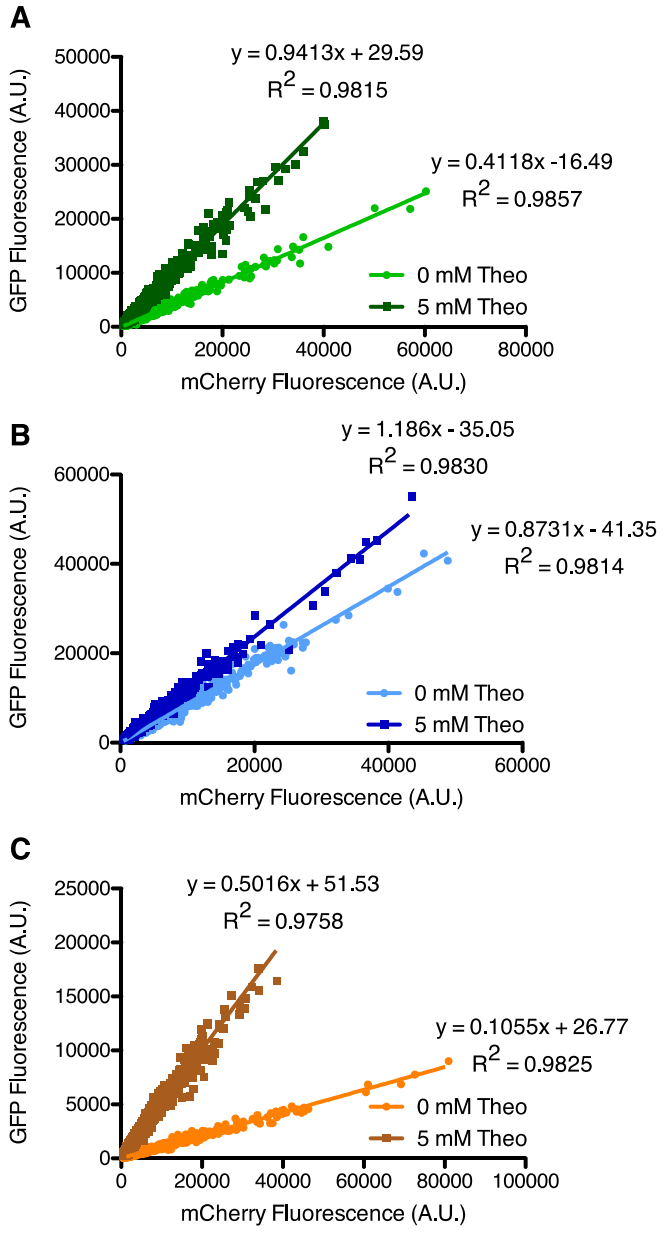
The purified uncleaved transcripts were incubated in 100 mM NaCl, 50 mM Tris-HCl (pH 7.5) at 95°C for 5 min, cooled at a rate of -1.3°C to 37°C, and held there for 10 min to allow equilibration of secondary structure. A zero time-point aliquot was taken prior to initiating the self-cleavage reaction at 37°C with the addition of $MgCl_2$ to a final concentration of 500 μM. Reactions were quenched at specified time points with addition of 3 volumes of RNA stop/load buffer on ice. Samples were heated 95°C for 5 min, snap cooled on ice for 5 min, and size-fractionated on a denaturing (8.3 M urea) 10% polyacrylamide gel at 25 W for 45 to 60 min. Gels were exposed overnight on a phosphor screen and analyzed for relative levels of the full-length transcript and cleaved products by phosphorimaging analysis. The cleaved product fraction at each time point (F_t) was fit to the single exponential equation $F_t = F_0 + (F_\infty - F_0) \times (1 - e^{-kt})$ using Prism 5 (GraphPad), where F_0 and F_∞ are the fractions cleaved before the start of the reaction and at the reaction endpoint, respectively, and k is the first-order rate constant of self-cleavage. Reported values are the mean of at least three independent experiments.

2.4.7 Measuring mating pathway activity via a transcriptional reporter and halo assays

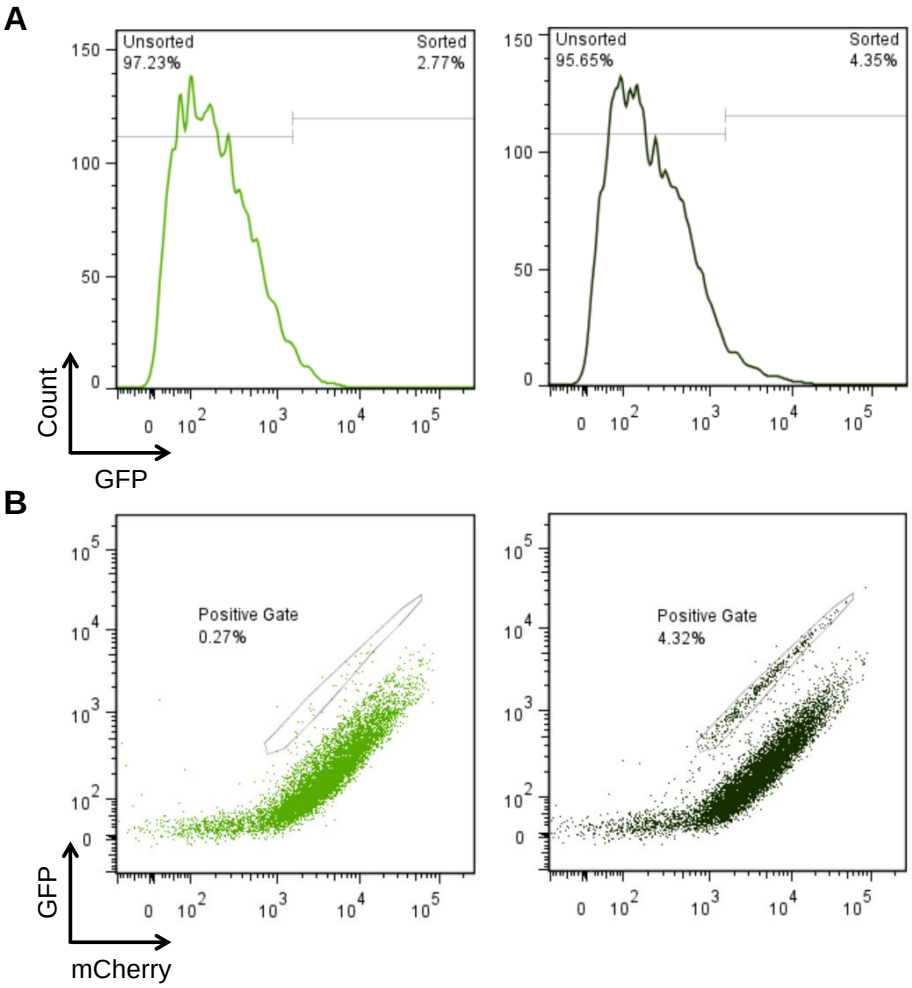
The mating fate routing plasmids with ribozyme-based devices and appropriate controls were transformed into yeast strain CSY840. Cells were inoculated into the appropriate drop-out media, grown overnight at 30°C, and back diluted into fresh media in the presence or absence of 5 mM theophylline to an OD_{600} of <0.1. After growing for 3 hr at 30°C, cells were stimulated with saturating pheromone levels, to a final concentration of 100 nM α mating factor acetate salt (Sigma-Aldrich, St. Louis, MO), to activate the mating pathway. After 3 hr of stimulation, GFP fluorescence levels from the pFus1-yEGFP3 reporter were evaluated via flow cytometry using a Cell Lab Quanta SC flow cytometer (Beckman Coulter, Fullerton, CA). Normalized pathway activity is calculated as the geometric mean of three biological replicates of each sample normalized to the blank plasmid control (no *MSG5*) in the absence of theophylline. Mating associated cell-cycle arrest was evaluated via halo assays (43). Halo assays were performed on cultures grown overnight, back diluted into fresh media in the absence of theophylline, and grown another 6 hr. 200 μl of each replicate was plated on the appropriate drop-out plates in the absence or presence of 5 mM theophylline. A gradient of α mating factor was established by saturating a filter disk (2 mm diameter) of Whatman paper with 9 μl of 0.1 mg/mL α mating factor and placing the disk on the center of the plate. Cells were grown for 18 hr at 30°C and imaged via epi-white illumination with a GelDoc XR+ System (Bio-Rad). To control for any possible differences in growth rate between replicates, plates are compared within one biological replicate ensuring differences in

growth rate arise from divergence following plating in the absence or presence of theophylline.

2.5 Supplementary Information

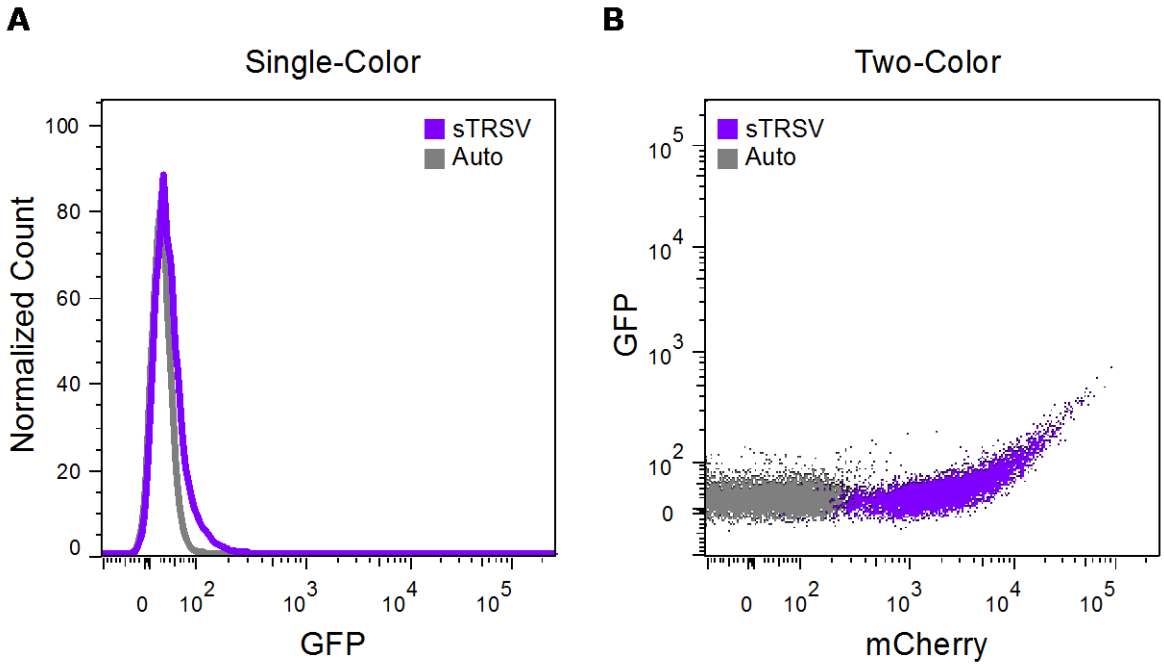


Supplementary Figure S2.1 Linear correlation between mCherry and GFP fluorescence for L2b1 (A), L2b5 (B), and L2b8 (C) devices within the two-color construct. For each device, 2,000 cells are plotted from a single representative experiment at 0 and 5 mM theophylline (theo). Linear regression and R^2 values are reported and support a distinct slope for different device regulatory activities.

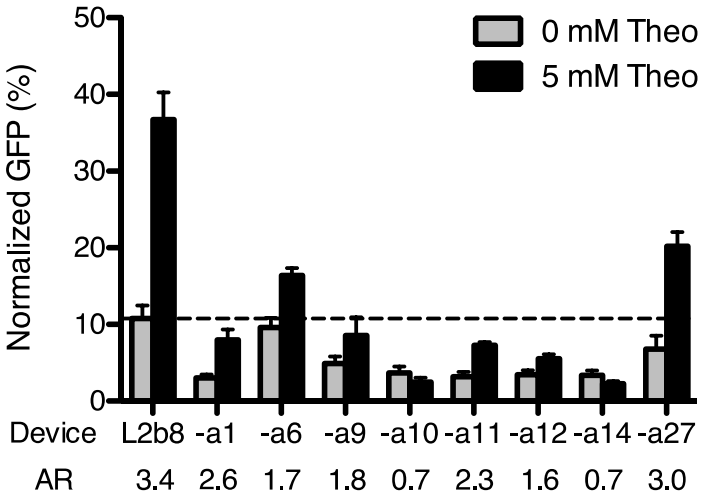


Supplementary Figure S2.2 Comparison of resolution between single-color and two-color screens. Enriched sN10 library after one sorting cycle and an additional negative sort was analyzed through flow cytometry in the presence and absence of 5 mM theophylline. The positive sorting gate is set by the activity of the parent device in the presence of 5 mM theophylline. In the absence of theophylline, the gate contains cells that are false positives; in the presence of theophylline, the gate contains both true and false positives. (A) The left GFP histogram represents the enriched library in the absence of theophylline, while the right GFP histogram represents the enriched library in the presence of theophylline. For the single-color screen, a hypothetical sorting gate was set based on the percentage of cells isolated from the corresponding two-color screen (4.32%). If a single-color screen were performed, ~64% (2.77/4.35) of the isolated cells would be false positives. (B) The left two-color scatter plot represents the enriched library in the absence of theophylline, while the right two-color scatter plot represents the enriched library in the presence of theophylline. The two-color screen exhibits high resolving power, such that only ~6% (0.27/4.32) of the isolated cells were false positives,

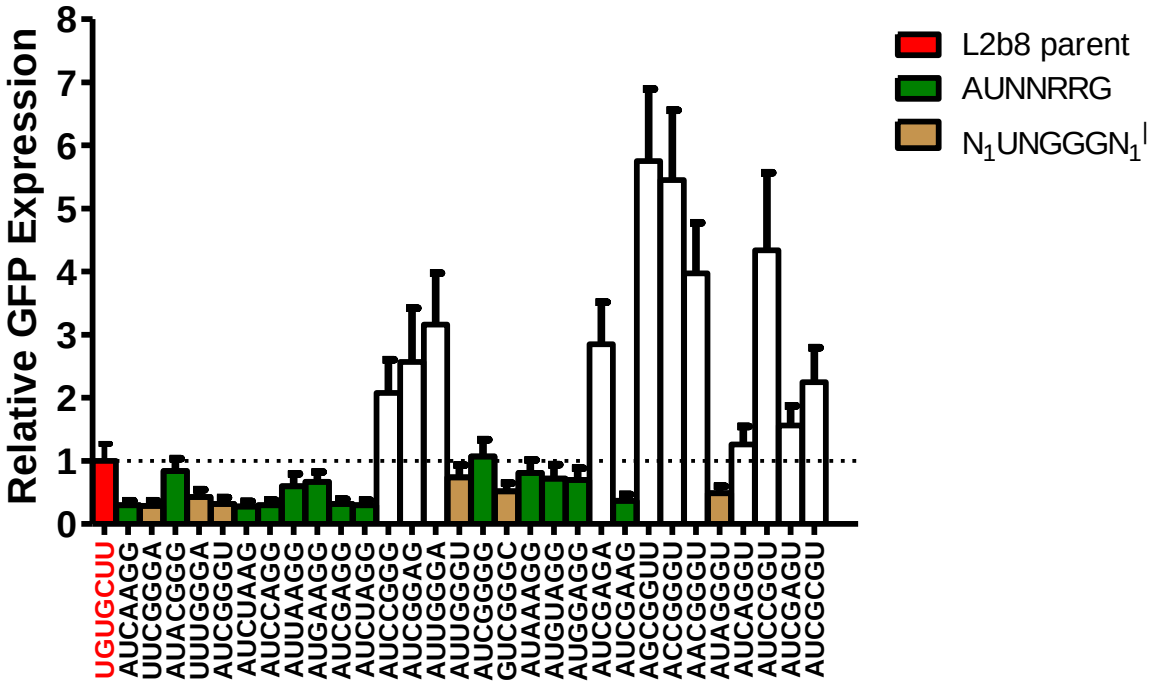
resulting in a significant improvement in enrichment efficiency over the single-color screen.



Supplementary Figure S2.3 The two-color sorting strategy supports separation of autofluorescent and low expression cell populations. sTRSV (the active ribozyme control) exhibits ~1.2% normalized GFP fluorescence, relative to a noncleaving ribozyme positive control (sTRSV Contl) with a scrambled core. (A) A cell population harboring the sTRSV construct exhibits low fluorescence levels that overlap significantly with a cell population harboring no fluorescence reporter gene in the single-color histogram. (B) The sTRSV and autofluorescent populations separate based on mCherry expression in the two-color plot, allowing cells exhibiting low GFP expression due to stringent device gene-regulatory activity to be effectively enriched through two-color sorting.



Supplementary Figure S2.4 An actuator sort preserves switching activity in 6 out of 8 characterized devices. Basal activity of the parent device, L2b8, is indicated by a dashed line. Gene-regulatory activities are reported as the geometric mean of the GFP fluorescence of the indicated sample normalized to that of a positive control (sTRSV Contl, a noncleaving sTRSV ribozyme with a scrambled core) that is grown under identical ligand conditions and is set to 100%. Reported values are the mean and standard deviation of at least three independent experiments. The activation ratio (AR), determined as the ratio of gene expression levels in the presence and absence of theophylline (theo), is indicated for each device.



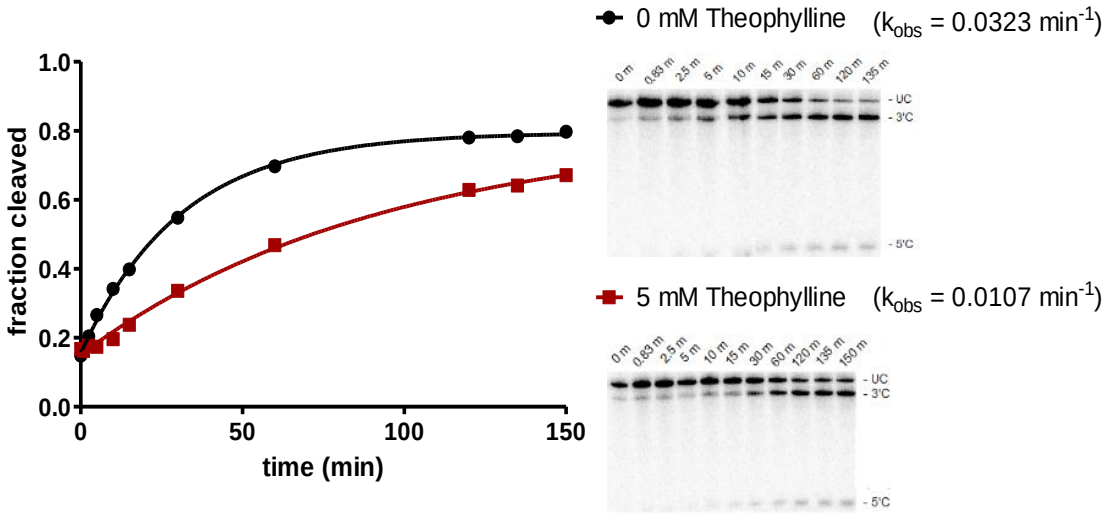
Supplementary Figure S2.5 Point mutation analyses of variant loop I sequences identify consensus sequences of ribozyme variants supporting improved gene-regulatory activities. The impact of point mutations to the recovered loop I sequences on basal activity was measured by flow cytometry. Relative expression levels are reported as the geometric mean of the GFP fluorescence of the indicated sample normalized to that of the L2b8 parent device. Reported values are the mean and standard deviation of at least three independent experiments. Two putative loop I consensus sequences were supported from the sequences recovered from the actuator library screen: a heptaloop (AUYNRRG) and a triloop (NUYGGGN$^{|}$).

Analysis of the point mutants made to interrogate nucleotide base constraints at each position of the heptaloop sequence found that any base (N) is tolerable at the third and fourth positions, but specific identities at the third position result in improved basal activity (C > A/U > G). The fifth and sixth positions accept both purine bases (A, G) and retain activity similar to that of the L2b8 parent; however, an adenine base (A) in the fifth position results in improved basal activity. Point mutants verified the requirement of a guanine base (G) in the seventh position.

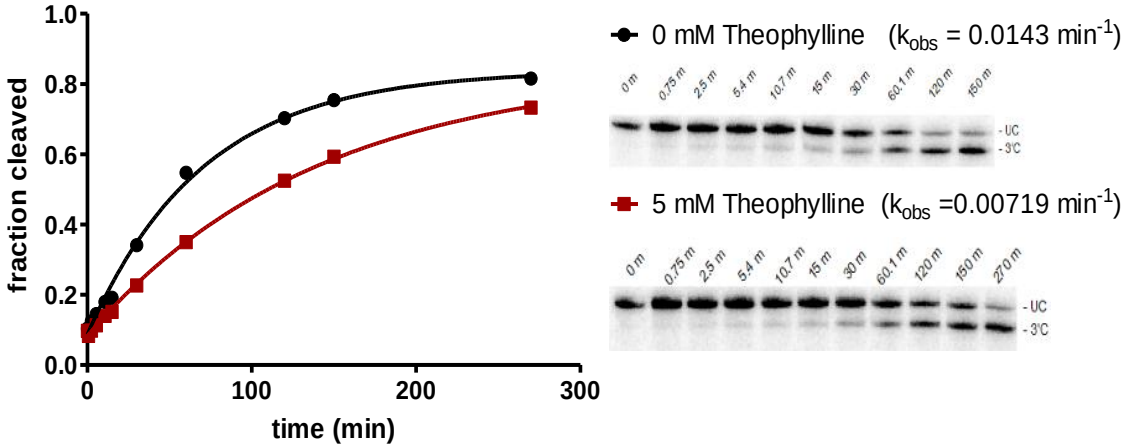
For the triloop sequence, point mutants indicate that any combination of bases in the first and seventh positions resulting in canonical Watson-Crick base pairing (AU, GC; order independent) results in improved basal activity. The bases in the second and sixth positions must result in a Watson-Crick GU wobble pair, where this requirement is sensitive to identity and order (i.e., standard Watson-Crick pairing abolishes improved

activity). The point mutants indicate that the third position is able to accept any nucleotide base (N), whereas a guanine base (G) is required at the fourth and fifth positions to retain improved activity relative to the L2b8 parent device.

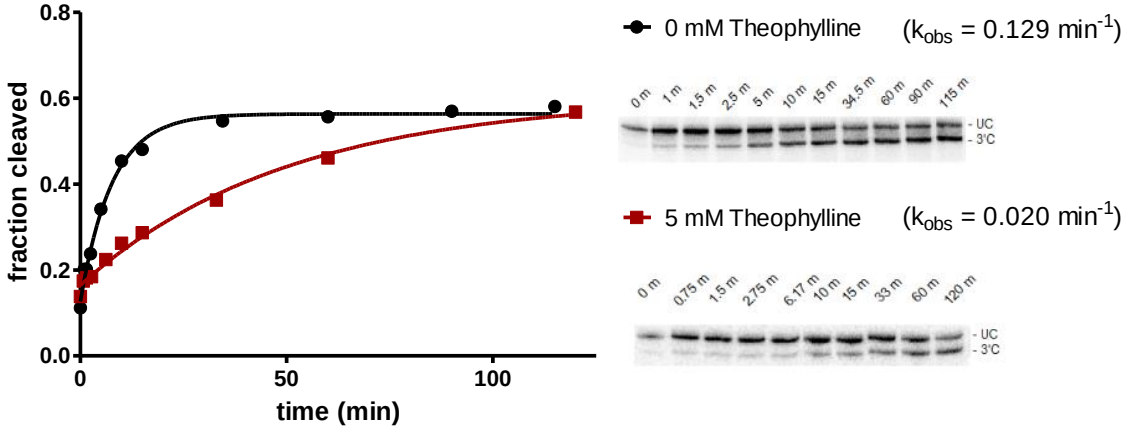
A L2b1:



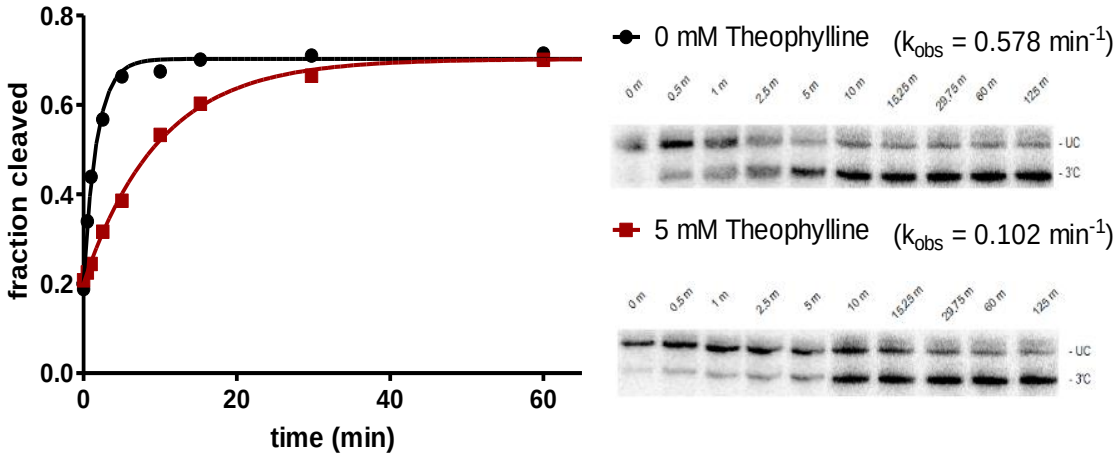
B L2b5:



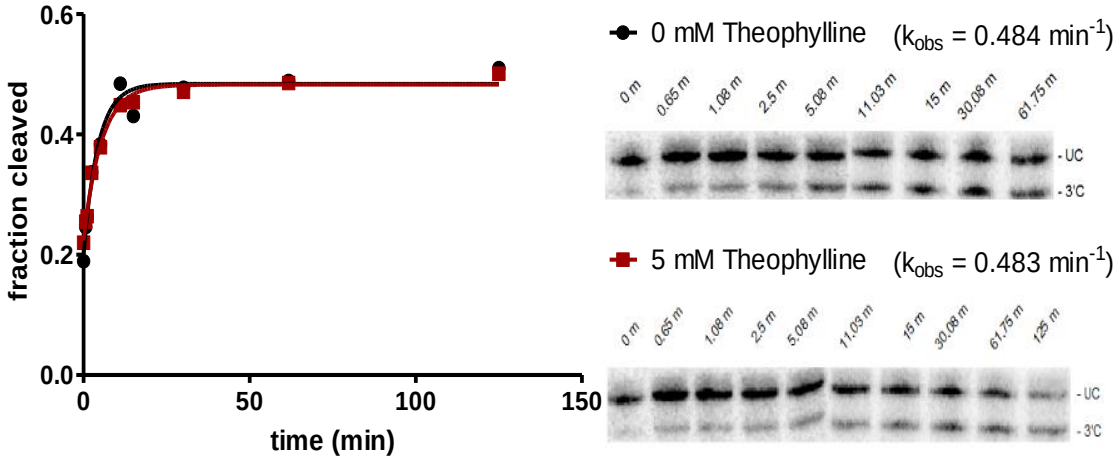
C L2b8:



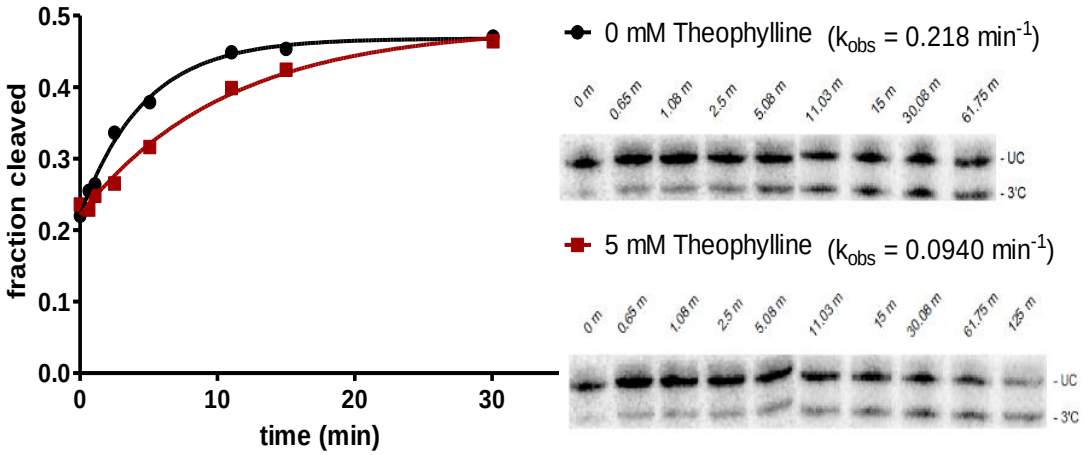
D L2b8-a1:



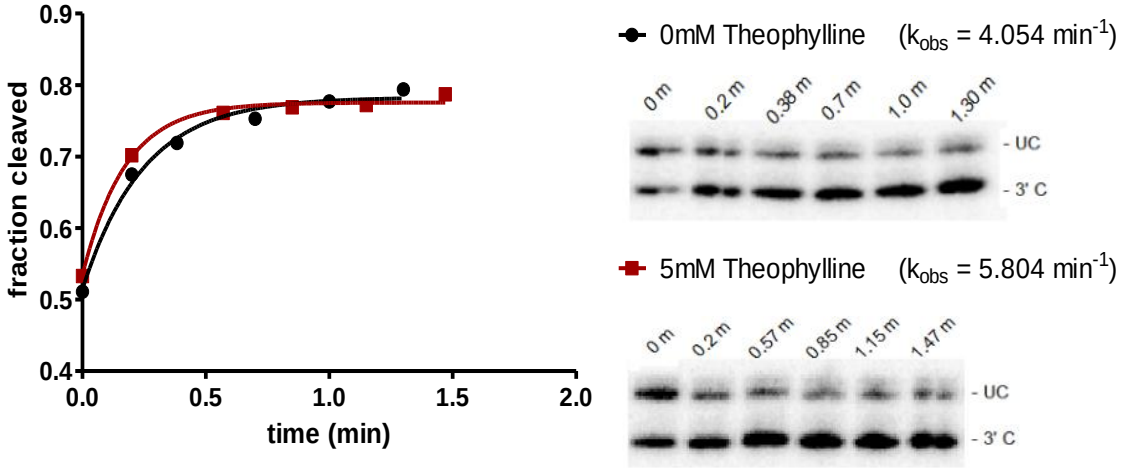
E L2b8-a14:



F L2b8-a1-t41:



G sTRSV ribozyme:



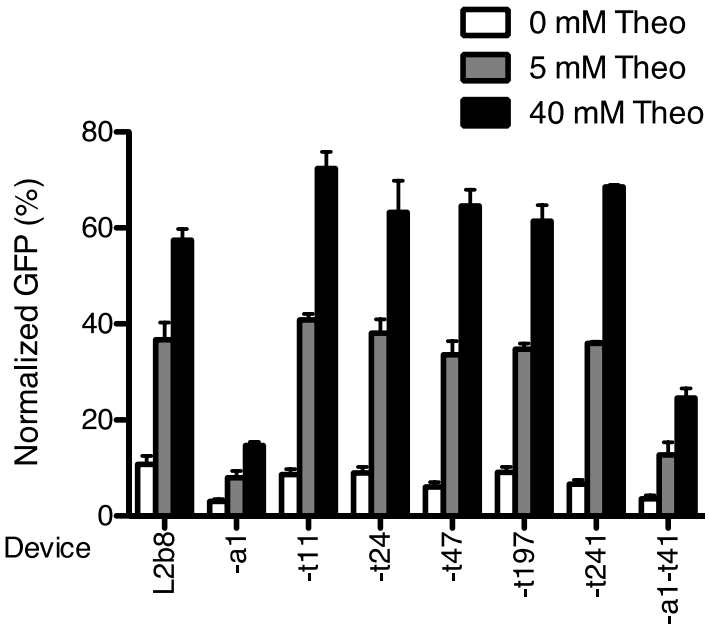
H

RNA Device	Cleavage Rate (k, min^{-1})	
	0 mM theophylline	**5 mM theophylline**
L2b1	0.033 ± 0.008	0.011 ± 0.003
L2b5	0.013 ± 0.002	0.007 ± 0.001
L2b8	0.14 ± 0.02	0.025 ± 0.009
L2b8-a1	0.7 ± 0.3	0.1 ± 0.03
L2b8-a14	0.9 ± 0.4	1.0 ± 0.2
L2b8-a1-t41	0.16 ± 0.05	0.096 ± 0.02
sTRSV ribozyme	4.3 ± 0.8	4.3 ± 1.4

Supplementary Figure S2.6 Representative cleavage assays for measuring cleavage rate constants (k) for ribozyme-based devices and controls. A representative assay is shown for each device in the absence and presence (0 and 5 mM, respectively) of theophylline: L2b1 (A), L2b5 (B), L2b8 (C), L2b8-a1 (D), L2b8-a14 (E), L2b8-a1-t41 (F), and sTRSV ribozyme (G). Bands for both the cleaved products (5'C, 3'C) and the full-length uncleaved substrate (UC) are shown for L2b1. For subsequent devices, the shorter 5'C product is omitted from the inset image for clarity. Methods used to prepare full-length, uncleaved RNA transcripts and conditions of the cleavage assays are detailed in the Materials and Methods section of the main text. Briefly, RNA was heated to 95°C for 5 min and cooled to 37°C in a secondary structure refolding buffer (100 mM NaCl, 50 mM Tris-HCl (pH 7.5)). A zero time point aliquot was removed prior to initiating the reaction with addition of $MgCl_2$ to a final concentration of 500 μM. Reactions were quenched at the indicated time points. At least seven time points were taken in each cleavage assay to capture the cleavage dynamics of RNA devices exhibiting different cleavage kinetics. Phosphorimaging analysis of relative levels of the UC, 5'C, and 3'C bands was used to determine the fraction cleaved at each time point (F_t). The fraction cleaved at the beginning (F_0) and end of reaction (F_∞) varied between assays, but all assays were well-fit to the single exponential equation ($R^2 > 0.95$):

$$F_t = F_0 + (F_\infty - F_0) \times (1 - e^{-kt})$$

The black and red fit lines represent assays performed at 0 and 5 mM theophylline, respectively. The cleavage rate constant value (*k*) was determined for each assay. The reported *k* for each device and theophylline assay condition is the mean and standard deviation of at least three independent experiments (H).



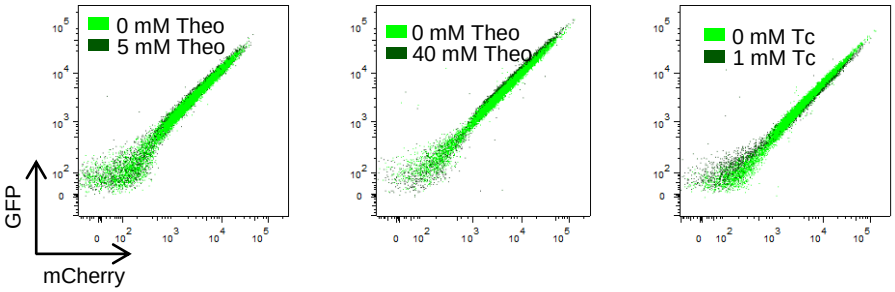
AR at 5 mM Theophylline	3.4	2.6	4.7	4.2	5.5	3.8	5.4	3.6
AR at 40 mM Theophylline	5.3	4.8	8.4	7.0	10.7	6.8	10.4	6.9

Supplementary Figure S2.7 Increased input ligand concentration results in higher device ON states. Increasing ligand concentration to 40 mM theophylline (theo) increases device activation ratio (AR) compared to 5 mM theophylline. Gene-regulatory activities are reported as the geometric mean of the GFP fluorescence of the indicated sample normalized to that of a positive control (sTRSV Contl, a noncleaving sTRSV ribozyme with a scrambled core) that is grown under identical ligand conditions and is set to 100%. Reported values are the mean and standard deviation of at least three independent experiments. The activation ratio (AR), determined as the ratio of gene expression levels in the presence and absence of theophylline (theo), is indicated for each device.

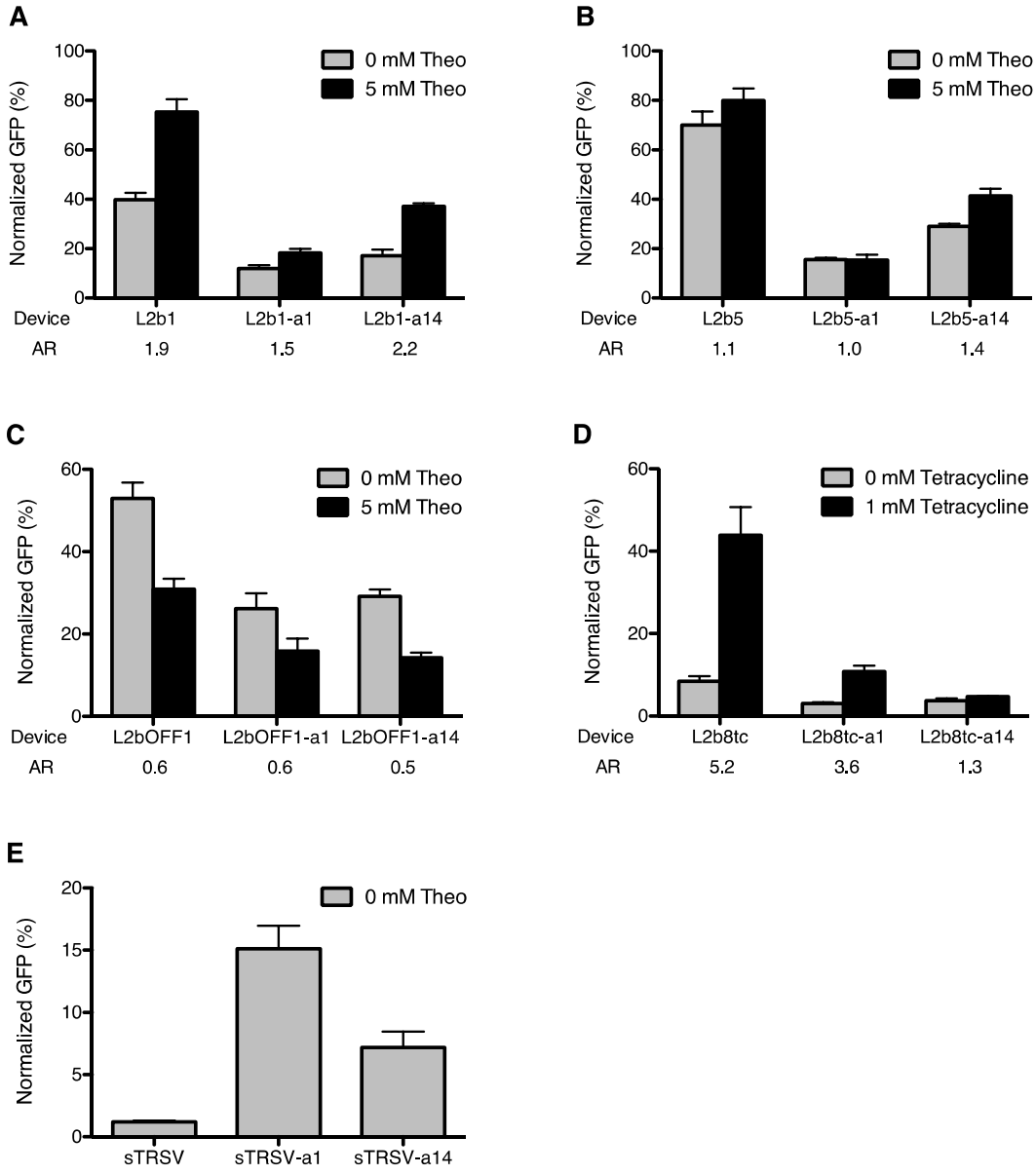
A

	GFP (A.U.)			mCherry (A.U.)			Viability (%)		
	0mM Theo	**5mM Theo**	**Ratio**	**0mM Theo**	**5mM Theo**	**Ratio**	**0mM Theo**	**5mM Theo**	**Ratio**
Replicate 1	3969.31	4532.12	1.14	2097.92	2339.26	1.12	98.24	97.16	0.99
Replicate 2	6282.93	6868.85	1.09	2699.57	2589.77	0.96	97.51	96.60	0.99
Replicate 3	2960.27	3129.06	1.06	2505.69	2496.95	1.00	96.47	96.73	1.00
Mean ± S.D.			1.10 ± 0.04			1.02 ± 0.08	97.41 ± 0.89	96.83 ± 0.29	0.99 ± 0.01
	0mM Theo	**40mM Theo**	**Ratio**	**0mM Theo**	**40mM Theo**	**Ratio**	**0mM Theo**	**40mM Theo**	**Ratio**
Replicate 1	4687.84	6380.01	1.36	5653.65	5504.26	0.97	97.63	96.62	0.99
Replicate 2	4777.71	6249.37	1.31	5137.39	5502.14	1.07	98.39	96.59	0.98
Replicate 3	4878.88	5977.69	1.23	4722.92	4921.71	1.04	97.99	94.83	0.97
Mean ± S.D.			1.30 ± 0.07			1.03 ± 0.05	98.00 ± 0.38	96.01 ± 1.02	0.98 ± 0.01
	0mM Tc	**1mM Tc**	**Ratio**	**0mM Tc**	**1mM Tc**	**Ratio**	**0mM Tc**	**1mM Tc**	**Ratio**
Replicate 1	2585.21	2382.28	0.92	2791.13	2278.18	0.82	98.05	98.72	1.01
Replicate 2	2310.21	2036.61	0.88	2896.15	2628.35	0.91	97.73	98.74	1.01
Replicate 3	2326.70	2203.89	0.95	3694.26	3691.82	1.00	97.75	98.22	1.00
Mean ± S.D.			0.92 ± 0.03			0.91 ± 0.09	97.84 ± 0.18	98.56 ± 0.29	1.01 ± 0.00

B

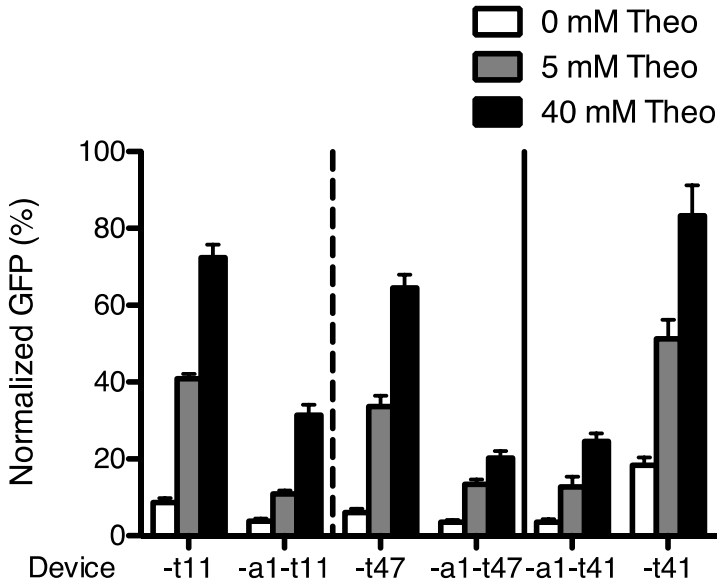


Supplementary Figure S2.8 Quantification of nonspecific ligand effect on fluorescence intensity and cell viability. (A) Unnormalized data for GFP fluorescence, mCherry fluorescence, and cell viability for the positive control construct (sTRSV Contl, a noncleaving sTRSV ribozyme with a scrambled core) in all ligand conditions tested. Fluorescence values are reported as the geometric mean of the GFP or mCherry fluorescence of sTRSV Contl. Cell viability is reported as the percentage of cells included in the DAPI-(-) gate. Mean and standard deviation of three independent experiments are indicated. The ratio of fluorescence levels or cell viability percentages in the presence and absence of theophylline (theo) or tetracycline (tc) is indicated for each ligand condition. Variability in absolute GFP or mCherry signal among replicates for an individual ligand condition is due to differences in instrument calibration between experiments. Values for each replicate pair in the absence and presence of ligand were obtained in the same experiment with identical instrument calibration. (B) Scatter plots for sTRSV Contl under each ligand condition from a single representative experiment.



Supplementary Figure S2.9 Component swapping demonstrates modularity of the optimized actuator components. Replacement of the loop I sequence with the a1 or a14 sequence decreases basal activity in all devices tested, including theophylline (theo)-responsive ON (A, B) and OFF (C) switches and a tetracycline-responsive ON switch (D). Switching activity is maintained in most devices. Notably, replacement of the loop I sequence in the sTRSV ribozyme (E) significantly increases basal activity. Gene-regulatory activities are reported as the geometric mean of the GFP fluorescence of the indicated sample normalized to that of a positive control (sTRSV Contl, a noncleaving sTRSV ribozyme with a scrambled core) that is grown under identical ligand conditions and is set to 100%. Reported values are the mean and standard deviation of at least three

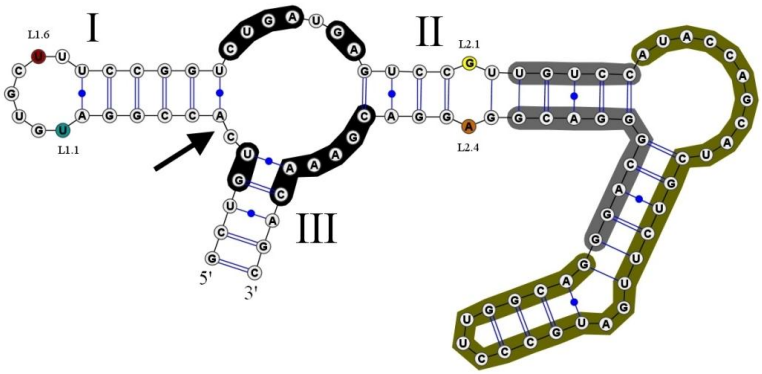
independent experiments. The activation ratio (AR), determined as the ratio of gene expression levels in the presence and absence of theophylline (theo) (A, B, C) or of tetracycline (D), is indicated for each device.



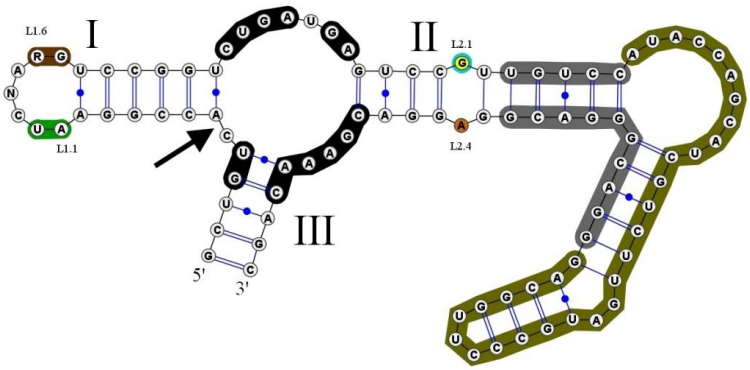
Device	-t11	-a1-t11	-t47	-a1-t47	-a1-t41	-t41
AR at 5 mM Theophylline	4.7	2.8	5.5	3.8	3.6	2.8
AR at 40 mM Theophylline	8.4	8.2	10.7	5.7	6.9	4.5

Supplementary Figure S2.10 The activities of the optimized actuator components are maintained in the context of the newly selected transmitter components. For devices isolated from the tN11 sort (L2b8-t11 and L2b8-t47), replacement of the actuator loop I sequence with the a1 sequence results in lower basal activity. For L2b8-a1-t41, which was isolated from the a1-tN11 sort, replacement of the a1 sequence with the loop I sequence of L2b8 results in higher basal activity. Gene-regulatory activities are reported as the geometric mean of the GFP fluorescence of the indicated sample normalized to that of a positive control (sTRSV Contl, a noncleaving sTRSV ribozyme with a scrambled core) that is grown under identical ligand conditions and is set to 100%. Reported values are the mean and standard deviation of at least three independent experiments. The activation ratio (AR), determined as the ratio of gene expression levels in the presence and absence of theophylline (theo), is indicated for each device.

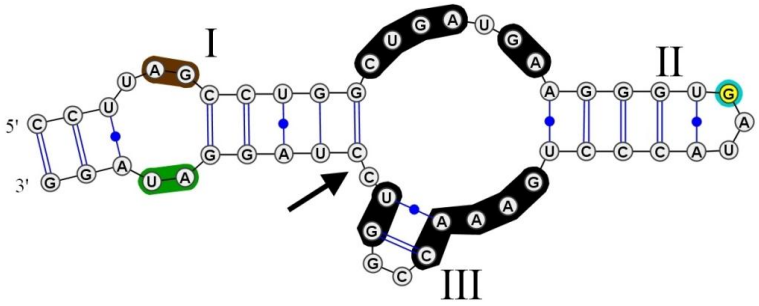
A L2b8 parent (Loop I: UGUGCUU)



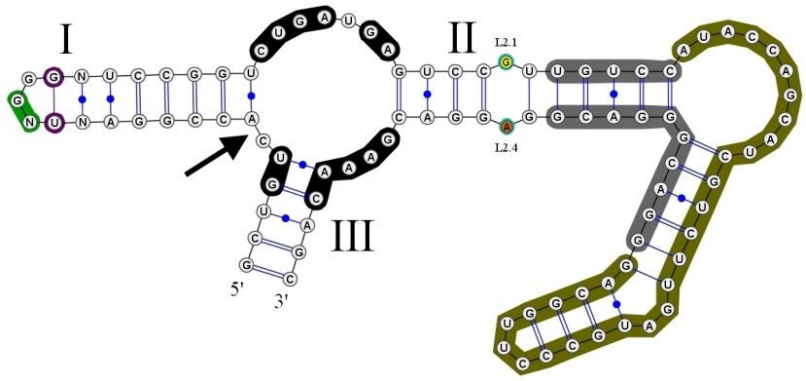
B aN7 heptaloop consensus (AUCNARG)



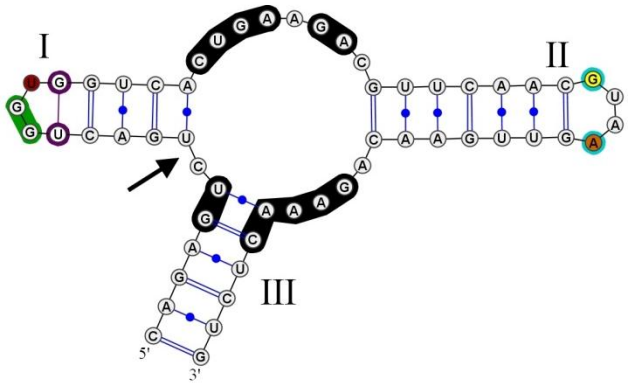
C Newt hammerhead ribozyme



D aN7 triloop consensus ($N_1UN_2GGGN_1$)

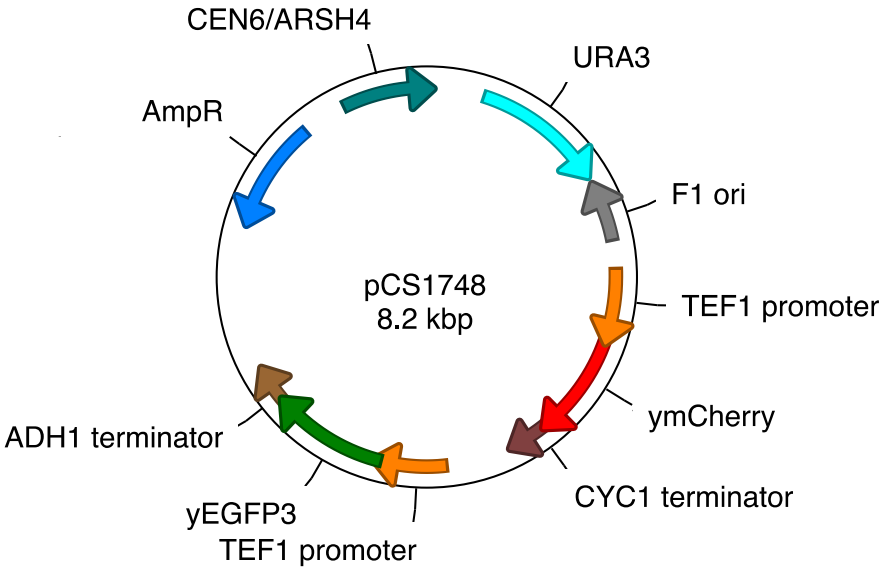


E CarSVRNA(-) hammerhead ribozyme

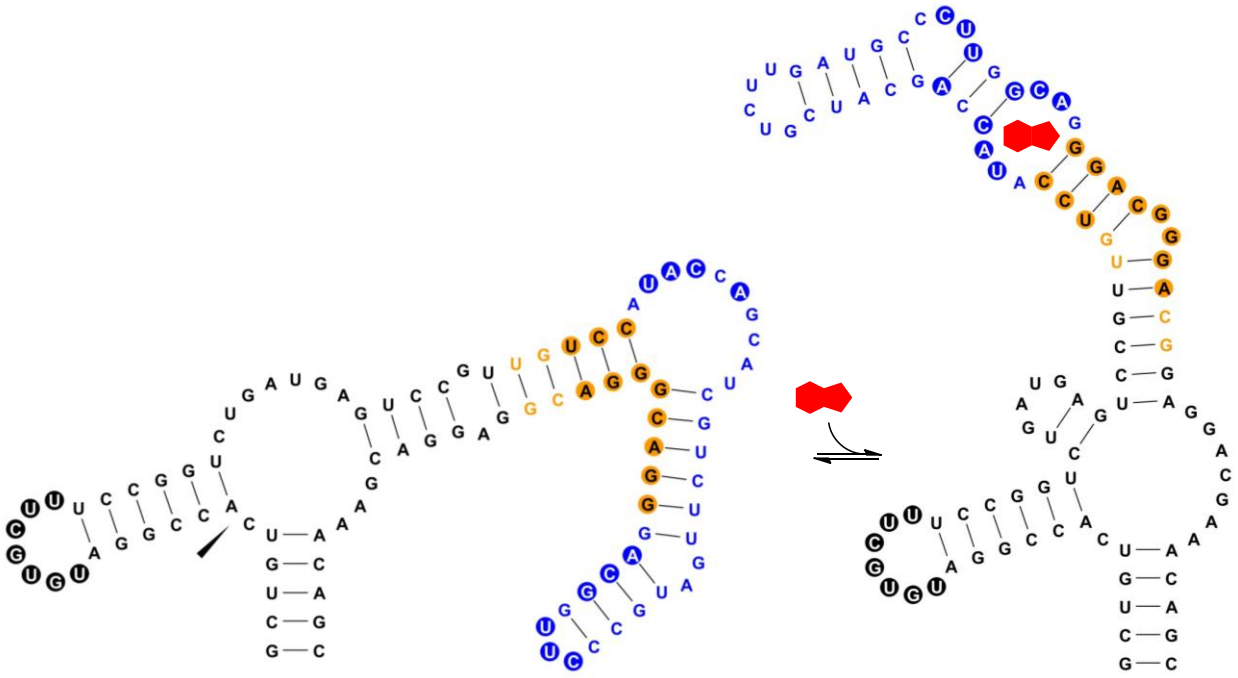


Supplementary Figure S2.11 Secondary structure analysis indicates similarities between engineered RNA devices and natural hammerhead ribozymes (HHRzs). The HHRz invariant nucleotide bases are boxed in black, and the cleavage site is denoted by an arrow in all structures. In RNA device structures (A, B, D), nucleotide bases of the transmitter and aptamer domains are boxed in grey and olive green respectively. Using the common numbering system for HHRz with apical loops (44), nucleotide bases highlighted in cyan (L1.1 U) and red (L1.6 U) are those conforming to the common HHRz loop I motif, UN_mYN, where m is typically 3 or 4 bases. Common loop II motif nucleotide bases, (RN_nA), are highlighted in yellow (L2.1 G) and orange (L2.1 A), where n is usually 2 or 4 bases (27). R and Y denote purine and pyrimidine bases respectively. (A) The HHRz core of the L2b8 parent device forms a predicted GU base pair at L2.2-L2.3 due to transmitter-sensor integration into loop II when analyzed by RNAstructure folding software (45). HHRz loop sequence similarities exist between a device containing the variant heptaloop consensus sequence selected in this study (B) and the newt HHRz (C). In loop I, both HHRz structures have AU in the first two positions 5' from the

catalytic core (boxed in green) and RG in last two positions (boxed in brown). In loop II, both structures have a G nucleotide not constrained in a Watson-Crick base pair (circled in cyan). Loop sequence similarities also exist between a device containing the variant triloop consensus sequence (D) and the CarSVRNA(-) HHRz (E). Both loop I sequences in these HHRzs contain a triloop with a GU base pair closing the loop (circled in purple) and accommodate GG in the first two positions 5’ from the catalytic core (boxed in green). In loop II, both HHRz structures have L2.1 G and L2.4 A nucleotides not constrained in Watson-Crick base pairs, available for loop I-loop II interaction, circled in cyan (27). Secondary structures rendered using VARNA software (46).



Supplementary Figure S2.12 Plasmid map of pCS1748. RNA devices are inserted in the 3' UTR of yEGFP3.



Supplementary Figure S2.13 Library design for sN10 sensor, aN7 actuator, and tN11 and a1-tN11 transmitter libraries. Predicted secondary structures of the active ribozyme conformation with unbound theophylline aptamer (left) and the inactive ribozyme conformation with bound theophylline aptamer (right) are shown. The sequence for the L2b8 parent device is shown. Nucleotides that comprise the sensor, actuator, and transmitter components are colored blue, black, and yellow, respectively. Theophylline is represented by the red polygon. Filled circles within each component indicate nucleotides randomized to generate respective device libraries. Secondary structures were predicted by RNAstructure folding software (45) and rendered using VARNA software (46).

Supplementary Table S2.1 Summary of plasmids and yeast strains constructed.

Plasmid	Color Controls
pCS1748	two-color screening plasmid: TEF1p-GFP-ADH1t and TEF1p-mCherry-CYC1t
pCS1585	single-color plasmid: TEF1p-GFP-ADH1t
pCS1749	single-color plasmid: TEF1p-mCherry-CYC1t
	Ribozyme-based Devices and Controls
pCS1750	pCS1748+sTRSV
pCS1751	pCS1748+sTRSV Contl
pCS1752	pCS1748+L2b1
pCS1753	pCS1748+L2b8
pCS2260	pCS1748+L2b5
pCS2261	pCS1748+L2b8-a1
pCS2262	pCS1748+L2b8-a6
pCS2263	pCS1748+L2b8-a9
pCS2264	pCS1748+L2b8-a10
pCS2265	pCS1748+L2b8-a11
pCS2266	pCS1748+L2b8-a12
pCS2267	pCS1748+L2b8-a14
pCS2268	pCS1748+L2b8-a27
pCS2269	pCS1748+L2b8-t11
pCS2270	pCS1748+L2b8-t24
pCS2271	pCS1748+L2b8-t47
pCS2272	pCS1748+L2b8-t197
pCS2273	pCS1748+L2b8-t241
pCS2274	pCS1748+L2b8-a1-t41
pCS2275	pCS1748+L2b8-a1-t55
pCS2276	pCS1748+L2b8-a1-t64
pCS2277	pCS1748+L2b1-a1
pCS2278	pCS1748+L2b1-a14
pCS2279	pCS1748+L2b5-a1
pCS2280	pCS1748+L2b5-a14
pCS2281	pCS1748+L2bOFF1
pCS2282	pCS1748+L2bOFF1-a1
pCS2283	pCS1748+L2bOFF1-a14
pCS2284	pCS1748+L2b8tc
pCS2285	pCS1748+L2b8tc-a1
pCS2286	pCS1748+L2b8tc-a14

pCS2287	pCS1748+sTRSV-a1
pCS2288	pCS1748+sTRSV-a14
pCS2289	pCS1748+L2b8-a1-t11
pCS2290	pCS1748+L2b8-a1-t47
pCS2291	pCS1748+L2b8-t41
	Yeast Mating Pathway
pCS1124	FUS1p-GFP-ADH1t
pCS2292	FUS1p-GFP-ADH1t in LoxP integrating vector
pCS2293	TEF1m7p-Msg5-ADH1t
pCS2294	pCS2293+sTRSV
pCS2295	pCS2293+L2b8-a1
pCS2297	pCS2293+L2b8
pCS2299	pCS2293+sTRSV Contl
Strain	
CSY840	W303a Δsst1 Δkss1::HIS3 trp1::FUS1p-yEGFP3-ADH1t-loxP-KanR

Acknowledgements

We thank C Crumpton and M Bigos of the Stanford Shared FACS facility. This work was supported by the National Institutes of Health (R01GM086663), the National Science Foundation (CBET-0917638, CCF-0943269), the Defense Advanced Research Projects Agency (HR0011-11-2-0002), the National Sciences and Engineering Research Council of Canada (fellowship to ABK), and the Alfred P. Sloan Foundation (fellowship to CDS).

Contributions

JCL designed and performed research and wrote the paper; ALC designed and performed research associated with the two-color FACS-based screen and wrote the paper; ABK designed and performed research associated with the *in vitro* characterization of ribozyme cleavage rates and wrote the paper; KEG designed and performed research associated with the MAPK pathway and wrote the paper; CDS designed research and wrote the paper.

References

1. Chen, Y.Y., Jensen, M.C. and Smolke, C.D. (2010) Genetic control of mammalian T-cell proliferation with synthetic RNA regulatory systems. *Proc Natl Acad Sci U S A*, **107**, 8531-8536.
2. Culler, S.J., Hoff, K.G. and Smolke, C.D. (2010) Reprogramming cellular behavior with RNA controllers responsive to endogenous proteins. *Science*, **330**, 1251-1255.
3. Ro, D.-K., Paradise, E.M., Ouellet, M., Fisher, K.J., Newman, K.L., Ndungu, J.M., Ho, K.A., Eachus, R.A., Ham, T.S., Kirby, J. *et al.* (2006) Production of the antimalarial drug precursor artemisinic acid in engineered yeast. *Nature*, **440**, 940-943.
4. Sinha, J., Reyes, S.J. and Gallivan, J.P. (2010) Reprogramming bacteria to seek and destroy an herbicide. *Nat Chem Biol*, **6**, 464-470.
5. Joyce, G.F. (2007) Forty years of in vitro evolution. *Angew Chem Int Ed Engl*, **46**, 6420-6436.
6. Seeman, N.C. (2005) From genes to machines: DNA nanomechanical devices. *Trends Biochem Sci*, **30**, 119-125.
7. Win, M.N., Liang, J.C. and Smolke, C.D. (2009) Frameworks for programming biological function through RNA parts and devices. *Chem Biol*, **16**, 298-310.
8. Fedor, M.J. and Williamson, J.R. (2005) The catalytic diversity of RNAs. *Nat Rev Mol Cell Biol*, **6**, 399-412.
9. Bartel, D.P. (2009) MicroRNAs: target recognition and regulatory functions. *Cell*, **136**, 215-233.

10. Win, M.N. and Smolke, C.D. (2007) A modular and extensible RNA-based gene-regulatory platform for engineering cellular function. *Proceedings of the National Academy of Sciences*, **104**, 14283-14288.

11. Link, K.H., Guo, L., Ames, T.D., Yen, L., Mulligan, R.C. and Breaker, R.R. (2007) Engineering high-speed allosteric hammerhead ribozymes. *Biol Chem*, **388**, 779-786.

12. Wittmann, A. and Suess, B. (2011) Selection of tetracycline inducible self-cleaving ribozymes as synthetic devices for gene regulation in yeast. *Mol Biosyst*, **7**, 2419-2427.

13. Chen, X., Denison, L., Levy, M. and Ellington, A.D. (2009) Direct selection for ribozyme cleavage activity in cells. *RNA*, **15**, 2035-2045.

14. Lynch, S.A. and Gallivan, J.P. (2009) A flow cytometry-based screen for synthetic riboswitches. *Nucleic Acids Res*, **37**, 184-192.

15. Fowler, C.C., Brown, E.D. and Li, Y. (2008) A FACS-based approach to engineering artificial riboswitches. *Chembiochem*, **9**, 1906-1911.

16. Desai, S.K. and Gallivan, J.P. (2004) Genetic screens and selections for small molecules based on a synthetic riboswitch that activates protein translation. *J Am Chem Soc*, **126**, 13247-13254.

17. Topp, S. and Gallivan, J.P. (2008) Random walks to synthetic riboswitches--a high-throughput selection based on cell motility. *Chembiochem*, **9**, 210-213.

18. Nomura, Y. and Yokobayashi, Y. (2007) Dual selection of a genetic switch by a single selection marker. *Biosystems*, **90**, 115-120.

19. Elowitz, M.B., Levine, A.J., Siggia, E.D. and Swain, P.S. (2002) Stochastic gene expression in a single cell. *Science*, **297**, 1183-1186.

20. Raser, J.M. and O'Shea, E.K. (2004) Control of stochasticity in eukaryotic gene expression. *Science*, **304**, 1811-1814.

21. Khvorova, A., Lescoute, A., Westhof, E. and Jayasena, S.D. (2003) Sequence elements outside the hammerhead ribozyme catalytic core enable intracellular activity. *Nat Struct Biol*, **10**, 708-712.

22. London, R.E. (1991) Methods for measurement of intracellular magnesium: NMR and fluorescence. *Annu Rev Physiol*, **53**, 241-258.

23. Canny, M.D., Jucker, F.M., Kellogg, E., Khvorova, A., Jayasena, S.D. and Pardi, A. (2004) Fast cleavage kinetics of a natural hammerhead ribozyme. *J Am Chem Soc*, **126**, 10848-10849.

24. Koch, A.L. and Lamont, W.A. (1956) The metabolism of methylpurines by Escherichia coli. II. Enzymatic studies. *J Biol Chem*, **219**, 189-201.

25. Win, M.N. and Smolke, C.D. (2008) Higher-order cellular information processing with synthetic RNA devices. *Science*, **322**, 456-460.

26. Andersson, J., Simpson, D.M., Qi, M., Wang, Y. and Elion, E.A. (2004) Differential input by Ste5 scaffold and Msg5 phosphatase route a MAPK cascade to multiple outcomes. *EMBO J*, **23**, 2564-2576.

27. Dufour, D., de la Pena, M., Gago, S., Flores, R. and Gallego, J. (2009) Structure-function analysis of the ribozymes of chrysanthemum chlorotic mottle viroid: a loop-loop interaction motif conserved in most natural hammerheads. *Nucleic Acids Res*, **37**, 368-381.

28. Perreault, J., Weinberg, Z., Roth, A., Popescu, O., Chartrand, P., Ferbeyre, G. and Breaker, R.R. (2011) Identification of hammerhead ribozymes in all domains of life reveals novel structural variations. *PLoS Comput Biol*, **7**, e1002031.

29. Seehafer, C., Kalweit, A., Steger, G., Graf, S. and Hammann, C. (2011) From alpaca to zebrafish: hammerhead ribozymes wherever you look. *RNA*, **17**, 21-26.

30. Alper, H., Fischer, C., Nevoigt, E. and Stephanopoulos, G. (2005) Tuning genetic control through promoter engineering. *Proc Natl Acad Sci U S A*, **102**, 12678-12683.

31. Nevoigt, E., Kohnke, J., Fischer, C.R., Alper, H., Stahl, U. and Stephanopoulos, G. (2006) Engineering of promoter replacement cassettes for fine-tuning of gene expression in Saccharomyces cerevisiae. *Appl Environ Microbiol*, **72**, 5266-5273.

32. Babiskin, A.H. and Smolke, C.D. (2011) A synthetic library of RNA control modules for predictable tuning of gene expression in yeast. *Mol Syst Biol*, **7**, 471.

33. Sambrook J, R.D. (2001) *Molecular Cloning: A Laboratory Manual, 3rd edn*. Cold Spring Harbor Lab Press, Cold Spring Harbor, NY.

34. Uppaluri, R. and Towle, H.C. (1995) Genetic dissection of thyroid hormone receptor beta: identification of mutations that separate hormone binding and transcriptional activation. *Mol Cell Biol*, **15**, 1499-1512.

35. Shaner, N.C., Campbell, R.E., Steinbach, P.A., Giepmans, B.N., Palmer, A.E. and Tsien, R.Y. (2004) Improved monomeric red, orange and yellow fluorescent proteins derived from Discosoma sp. red fluorescent protein. *Nat Biotechnol*, **22**, 1567-1572.

36. Gari, E., Piedrafita, L., Aldea, M. and Herrero, E. (1997) A set of vectors with a tetracycline-regulatable promoter system for modulated gene expression in Saccharomyces cerevisiae. *Yeast*, **13**, 837-848.

37. Sikorski, R.S. and Hieter, P. (1989) A system of shuttle vectors and yeast host strains designed for efficient manipulation of DNA in Saccharomyces cerevisiae. *Genetics*, **122**, 19-27.

38. Gietz, R.D. and Woods, R.A. (2002) Transformation of yeast by lithium acetate/single-stranded carrier DNA/polyethylene glycol method. *Methods Enzymol*, **350**, 87-96.

39. Siekhaus, D.E. and Drubin, D.G. (2003) Spontaneous receptor-independent heterotrimeric G-protein signalling in an RGS mutant. *Nat Cell Biol*, **5**, 231-235.

40. Guldener, U., Heck, S., Fielder, T., Beinhauer, J. and Hegemann, J.H. (1996) A new efficient gene disruption cassette for repeated use in budding yeast. *Nucleic Acids Res*, **24**, 2519-2524.

41. Flotho, A., Simpson, D.M., Qi, M. and Elion, E.A. (2004) Localized feedback phosphorylation of Ste5p scaffold by associated MAPK cascade. *J Biol Chem*, **279**, 47391-47401.

42. Chao, G., Lau, W.L., Hackel, B.J., Sazinsky, S.L., Lippow, S.M. and Wittrup, K.D. (2006) Isolating and engineering human antibodies using yeast surface display. *Nat Protoc*, **1**, 755-768.

43. Sprague, G.F., Jr. (1991) Assay of yeast mating reaction. *Methods Enzymol*, **194**, 77-93.

44. Hertel, K.J., Pardi, A., Uhlenbeck, O.C., Koizumi, M., Ohtsuka, E., Uesugi, S., Cedergren, R., Eckstein, F., Gerlach, W.L., Hodgson, R. *et al.* (1992) Numbering system for the hammerhead. *Nucleic Acids Res*, **20**, 3252.

45. Mathews, D.H., Disney, M.D., Childs, J.L., Schroeder, S.J., Zuker, M. and Turner, D.H. (2004) Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure. *Proc Natl Acad Sci U S A*, **101**, 7287-7292.

46. Darty, K., Denise, A. and Ponty, Y. (2009) VARNA: Interactive drawing and editing of the RNA secondary structure. *Bioinformatics*, **25**, 1974-1975.

Chapter 3.

A versatile cis-blocking strategy for ribozyme characterization

Abstract

Synthetic RNA control devices that utilize ribozyme-based actuator components have been demonstrated in diverse organisms to regulate cellular behaviors in response to environmental signals. The quantitative measurements of the *in vitro* cleavage rates associated with these synthetic devices are essential for advancing our understanding of the corresponding *in vivo* gene-regulatory activities. One of the key challenges in the ribozyme characterization is the efficient generation of full-length RNA through transcription reactions. Current methods of full-length RNA generation require a laborious gel purification step. We developed a cis-blocking strategy that supports efficient generation of full-length RNA by a simple gel-free process. We demonstrated that our cis-blocking strategy is applicable to both natural ribozymes and synthetic ribozyme devices. The cleavage rates obtained on the full-length RNA generated from the cis-blocking strategy were characterized and closely correlated to those from the trans-blocking strategy. We further developed a rapid, label-free cleavage assay based on surface plasmon resonance (SPR). The SPR-based assay allows rapid monitoring of ribozyme cleavage under varying reaction conditions. With further optimization and development, our cis-blocking and SPR-based characterization strategies will provide a powerful method for ribozyme characterization.

3.1 INTRODUCTION

Synthetic biology is advancing the design of genetic circuits encoding desired functions (1). As the proper functioning of synthetic genetic circuits often relies on the precise control and tuning of key protein component levels, much effort in the field has focused on developing programmable gene-regulatory devices. One class of gene-regulatory devices, so called RNA control devices, utilize RNA molecules to link changes in molecular signals to gene expression events. Progress in the fields of RNA biology and engineering have allowed construction of RNA control devices through the assembly of natural or synthetic components that encode more basic functions, such as sensing, information transmission, and actuation (2). Because hammerhead ribozymes can exhibit activity independent of cell-specific machinery, RNA control devices that utilize ribozyme-based actuator components have been demonstrated in diverse organisms spanning from bacterial, yeast, and mammalian systems to regulate cellular behaviors in response to one or more environmental signals (3-6).

We recently described a modular ribozyme device platform based on the functional coupling of three distinct components: an actuator encoded by a hammerhead ribozyme from satellite RNA of tobacco ringspot virus (sTRSV), a sensor encoded by an RNA aptamer, and an information transmitter encoded by a sequence capable of a programmed strand-displacement event (3) (Figure 3.1). The transmitter component directs the device to partition between two primary functional conformations, where the aptamer-folded conformation can be associated with the disruption or restoration of the ribozyme catalytic core to construct ON (increased gene expression) or OFF (decreased gene expression) devices, respectively, in response to ligand binding to the aptamer. The

quantitative tuning of device regulatory dynamic range, which is set by the gene expression activity in the absence and presence of input ligand, has been demonstrated to be important for achieving effective control over phenotypic behaviors (see Chapter 2). Computational models of ribozyme devices have predicted that the values of the ribozyme cleavage rate can have substantial impact on the device gene-regulatory activities (7). Therefore, quantitative measurements of the ribozyme cleavage rates are essential to advance our understanding of the relationship between the *in vitro* cleavage rates and the corresponding *in vivo* gene-regulatory activities.

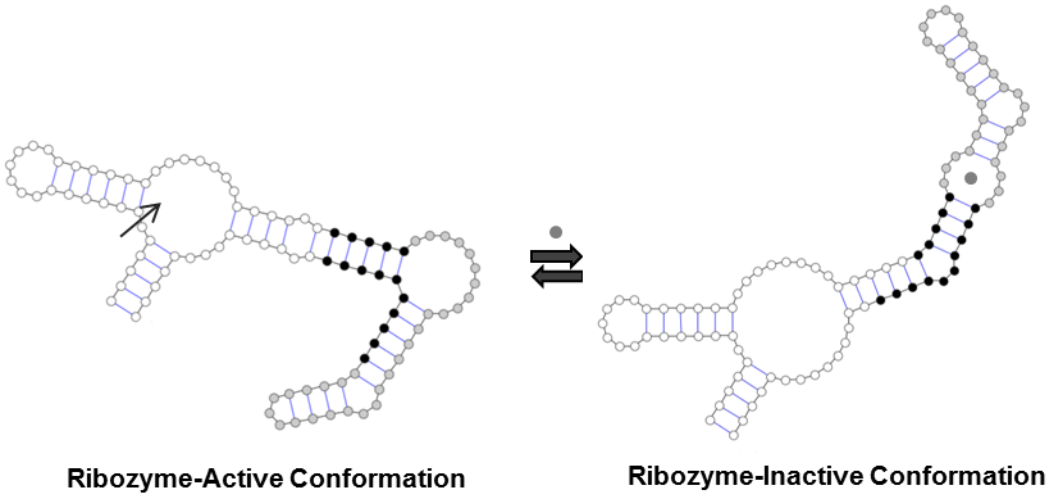


Figure 3.1 Modular composition of a synthetic ribozyme device. The ribozyme device is constructed by linking a RNA aptamer (indicated in grey) to a hammerhead ribozyme (indicated in white) through a programmed sequence (indicated in black) that directs a strand-displacement event. The device can adopt two primary functional conformations, in which the ribozyme-active conformation is associated with either aptamer-unformed or aptamer–formed conformation to construct a ribozyme ON or OFF device, respectively. In the case of a ribozyme ON device (depicted here), the device adopts a ribozyme-active conformation in the absence of the input ligand, thereby leading to self-cleavage. In the presence of the input ligand, binding of the ligand to the aptamer stabilizes a device conformation in which the ribozyme catalytic core is disrupted, thus resulting in no cleavage. Secondary structures were rendered using VARNA software (8).

A key challenge in characterizing ribozyme cleavage rates is in the efficient generation of full-length RNA for ribozyme constructs, including natural ribozymes and synthetic ribozyme devices, by *in vitro* T7 transcription reactions. Typical transcription reactions require high $MgCl_2$ (~mM) to achieve sufficient yield, conditions at which most ribozyme constructs exhibit high cleavage activity, thereby resulting in low yield of full-length RNA for the downstream kinetic analysis. In early characterization studies of natural ribozymes, researchers split the whole ribozyme sequence into an enzyme and a substrate strand (9,10). The two strands were transcribed separately and annealed to generate the functional two-stranded ribozyme complex. One disadvantage of this strategy is that experimental conditions need to be optimized to ensure that the measured cleavage rate is not affected by the association step between the two strands (11). Later, researchers developed a trans-blocking strategy to generate full-length single-stranded ribozyme transcripts by transcription reactions in the presence of an antisense DNA strand complementary to the ribozyme catalytic core. Full-length trans-blocked RNA was purified through polyacrylamide gel electrophoresis (PAGE) under denaturing conditions and renatured prior to cleavage assays (9,12). Although this strategy allows characterization of a ribozyme in its native conformation, previous studies have shown that the gel-purified full-length RNA was still capable of cleaving during sample handling, resulting in an ~30% loss prior to cleavage assays, and ~50% of the remaining full-length RNA does not exhibit cleavage activity (9). Therefore, a more efficient ribozyme-blocking strategy is desired.

As the kinetics of ribozyme cleavage are slow relative to many protein enzymes, ribozyme cleavage rates are typically measured through a gel electrophoretic separation

method (11,12). The cleavage reaction is initiated by the addition of $MgCl_2$ to the radiolabelled full-length RNA and quenched at different time points. The quenched samples are resolved by denaturing PAGE to determine the fraction RNA cleaved as a function of time, which are fit to a first-order exponential equation to obtain the cleavage rate. Because the gel-based ribozyme cleavage assay is discontinuous, several time points are required in the initial as well as the final phase of the reaction for proper analysis, thereby making the gel-based assay a time-consuming process. A continuous ribozyme cleavage assay based on the intramolecular fluorescence resonance energy transfer (FRET) has been developed and applied to characterize ribozyme cleavage rates (13,14). However, the FRET-based cleavage assay requires labeling of RNA molecules with appropriate fluorophores and is currently limited to the characterization of two-stranded ribozyme complexes only.

Advances in biosensor technologies have led to the development of the Biacore platform based on surface plasmon resonance (SPR) for real-time detection of biomolecular interactions (15). The Biacore platform utilizes a sensor surface modified with one molecule and measures the refractive index change, which is associated with the local mass density change, due to the association or dissociation of another molecule in solution. The refractive index change is converted into an SPR signal, which is expressed in resonance units (RU), and fitted to a mathematical model to acquire thermodynamic and kinetic parameters. The Biacore platform requires no labeling of molecules, consumes little sample, and is highly-automated, thus making it an attractive tool for developing novel ribozyme characterization strategies.

In this work, we developed a simple two-step gel-free process that allows efficient generation of full-length transcription products for diverse ribozyme constructs. We designed an RNA sequence that was directly incorporated into the target transcript to co-transcriptionally inhibit the formation of the ribozyme-active conformations. The catalytic activity of the RNA was restored by the use of a DNA competing strand to activate the blocked RNA through a fast toehold-mediated strand-displacement reaction. We demonstrated our cis-blocking strategy to generate full-length RNA transcripts for various natural ribozymes and synthetic ribozyme devices, with a blocking efficiency of up to ~90%. We performed gel-based cleavage assays to measure the cleavage rates of RNA generated from the cis-blocking strategy and compared these rates to those obtained from the more traditional trans-blocking strategy. The two rates were closely correlated, but further optimization of gel-based assays on RNA generated from the cis-blocking strategy is needed. Lastly, we described a label-free and real-time SPR-based ribozyme cleavage assay. Initial results suggested that the SPR-based cleavage assay can provide a rapid and quantitative alternative to the traditional gel-based assay.

3.2 RESULTS

3.2.1 A cis-acting RNA blocking sequence supports inhibition of ribozyme cleavage during transcription

To prevent ribozyme constructs from cleaving during transcription, we developed a new strategy to inhibit the formation of ribozyme-active conformation co-transcriptionally. We illustrated our strategy using a hammerhead ribozyme from satellite RNA of tobacco ringspot virus (sTRSV) as an example (Figure 3.2A). We rationally designed a 12 nt RNA blocking sequence that is complementary to part of the sTRSV

ribozyme transcript (Figure 3.2A). The RNA blocking sequence was inserted in the 5' end of the sTRSV ribozyme transcript to generate a cis-blocked sTRSV construct. As the RNA is being synthesized, the blocking sequence is capable of competitively hybridizing to the targeted ribozyme sequence, thus preventing the transcribed RNA from adopting a ribozyme-active conformation (Figure 3.2A).

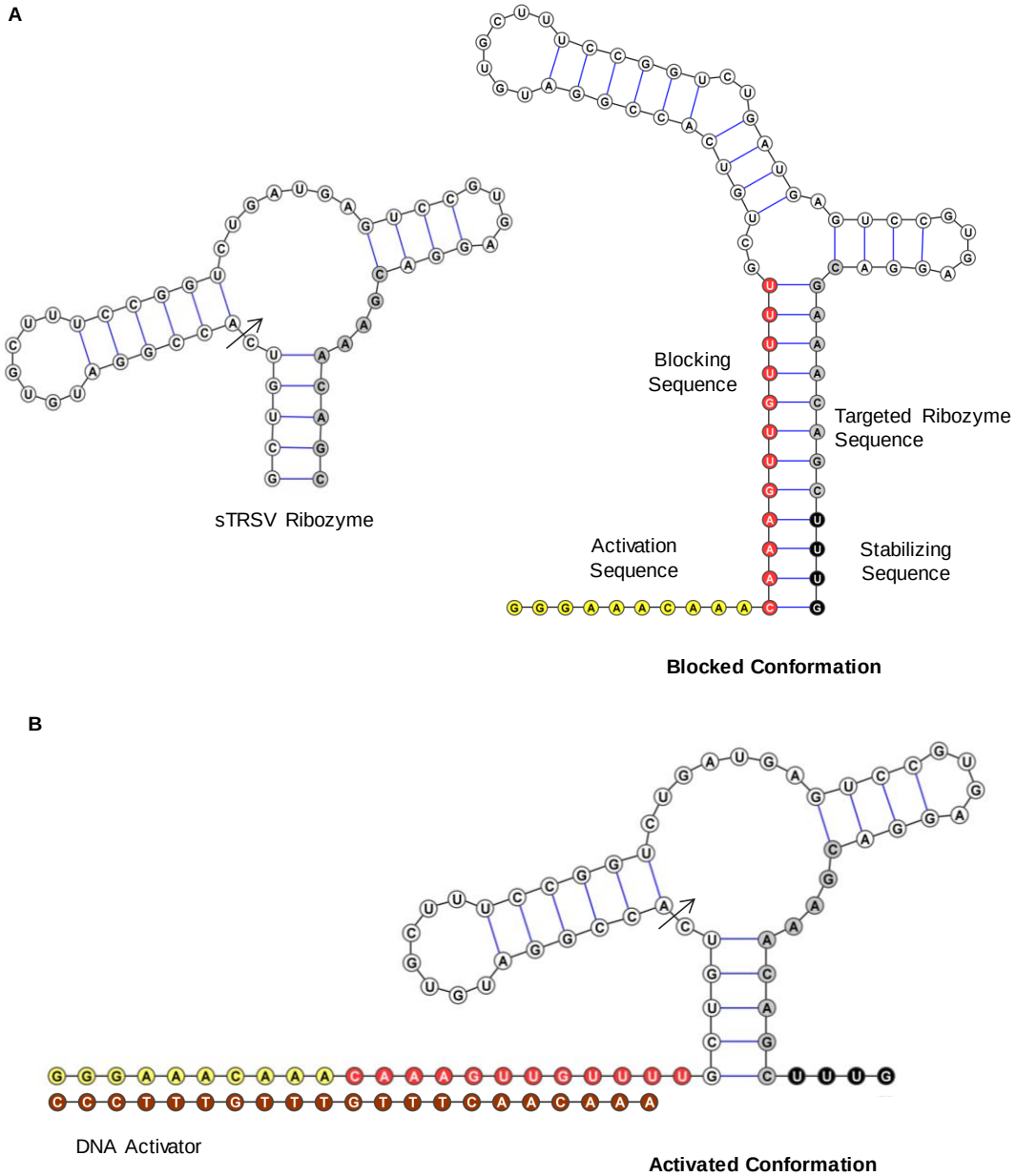


Figure 3.2 Development of the cis-blocking strategy to generate full-length RNA through transcription reactions. The cis-blocking strategy is demonstrated on a sTRSV ribozyme construct. The arrow indicates the cleavage site. (A) RNA blocking (indicated in red) and activation sequences (indicated in yellow) complementary to part of the ribozyme (indicated in grey) are directly incorporated in the 5' end of the ribozyme transcript and can inhibit the ribozyme construct from folding into the active conformation during the transcription reaction. (B) The cis-blocked ribozyme construct can be relieved by the addition of a DNA activator strand (indicated in brown), which competes with the blocking sequence to bind to the ribozyme sequence through a toehold-mediated strand-displacement reaction. As shown here, the fully hybridized ribozyme construct with the activator results in the ribozyme-active conformation. Secondary structures were predicted by RNAstructure folding software (16) and rendered using VARNA software (8).

3.2.2 A trans-acting DNA activator allows release of blocked RNA through a strand-displacement reaction

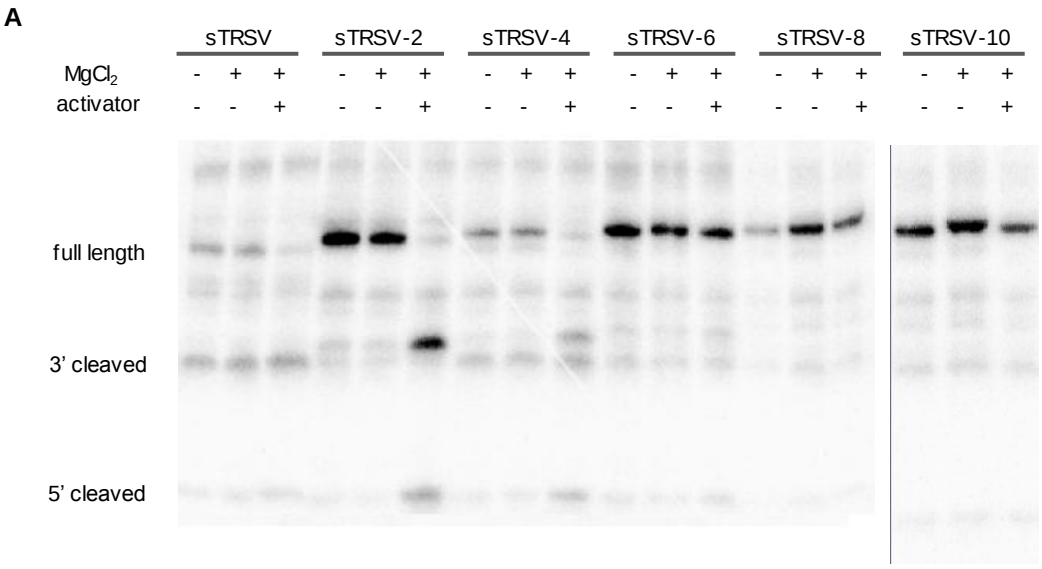
To activate the blocked sTRSV RNA for subsequent cleavage assays, we rationally designed a 10 nt RNA activation sequence as a single-stranded domain directly attached to the 5' end of the RNA blocking sequence (Figure 3.2A). The RNA activation sequence serves as a toehold to allow for the design of a DNA activator complementary to the RNA blocking sequence (Figure 3.2B). A toehold length between 5-10 nt has been demonstrated to enhance the rate of strand-displacement reaction by ~10^6 fold (17). The DNA activator can sequester the RNA blocking sequence through a fast toehold-mediated strand-displacement reaction, thus activating the blocked RNA for subsequent cleavage assays.

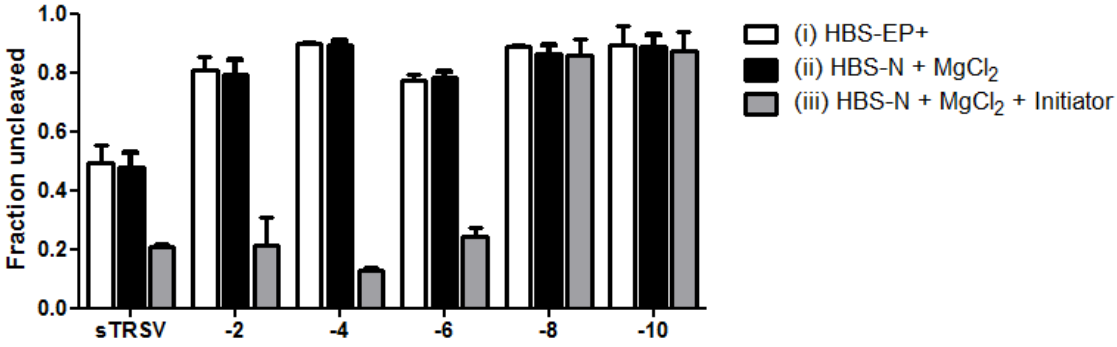
3.2.3 A RNA stabilizing sequence allows tuning of blocking efficiency

We tuned the strength of the competitive hybridization between the RNA blocking and targeted ribozyme sequences by introducing a stabilizing sequence at the 3' end of the sTRSV construct (Figure 3.2A). The stabilizing sequence is designed to

hybridize to the RNA activation sequence, such that the thermal stability of the blocked conformation can be tuned by varying the length of the stabilizing sequence. We increased the length of the stabilizing sequence from 4 to 14 nt to generate a variety of cis-blocked sTRSV constructs (Supplementary Figure S3.1). The blocking efficiency of the resultant sTRSV construct variants were determined by analyzing the transcription products by PAGE analysis (Figure 3.3A). By adding 4 nt to the originally designed stabilizing sequence (sTRSV-4 construct), we increased the uncleaved RNA percentage to ~90% from the transcription reaction. These results support that the cis-blocking strategy is efficient in inhibiting ribozyme cleavage during transcription.

We examined the efficacy of the toehold-mediated activation strategy by incubating the cis-blocked sTRSV construct variants at 5 mM $MgCl_2$ in the absence and presence of the activator strand for 30 minutes at 25ºC (Figure 3.3A). In the absence of the activator strand, no cleavage was observed even at high $MgCl_2$, indicating that the RNA was incapable of spontaneously releasing itself from the blocked conformation. In the presence of the activator strand, however, up to ~80% of the constructs with a toehold length between 4–6 nt cleaved. Little cleavage was observed for constructs with a toehold length between 0-2 nt, suggesting that the rate of strand-displacement reaction is too slow to allow effective activation of the blocked RNA (17). These results highlight that our cis-blocking strategy is efficient in generating full-length functional RNA through two simple steps: transcription and activation.


A
sTRSV
sTRSV-2
sTRSV-4
sTRSV-6
sTRSV-8
sTRSV-10
MgCl2
activator
full length
3' cleaved
5' cleaved


Fraction uncleaved
1.0
0.8
0.6
0.4
0.2
0.0
sTRSV
-2
-4
-6
-8
-10
(i) HBS-EP+
(ii) HBS-N + MgCl2
(iii) HBS-N + MgCl2 + Initiator

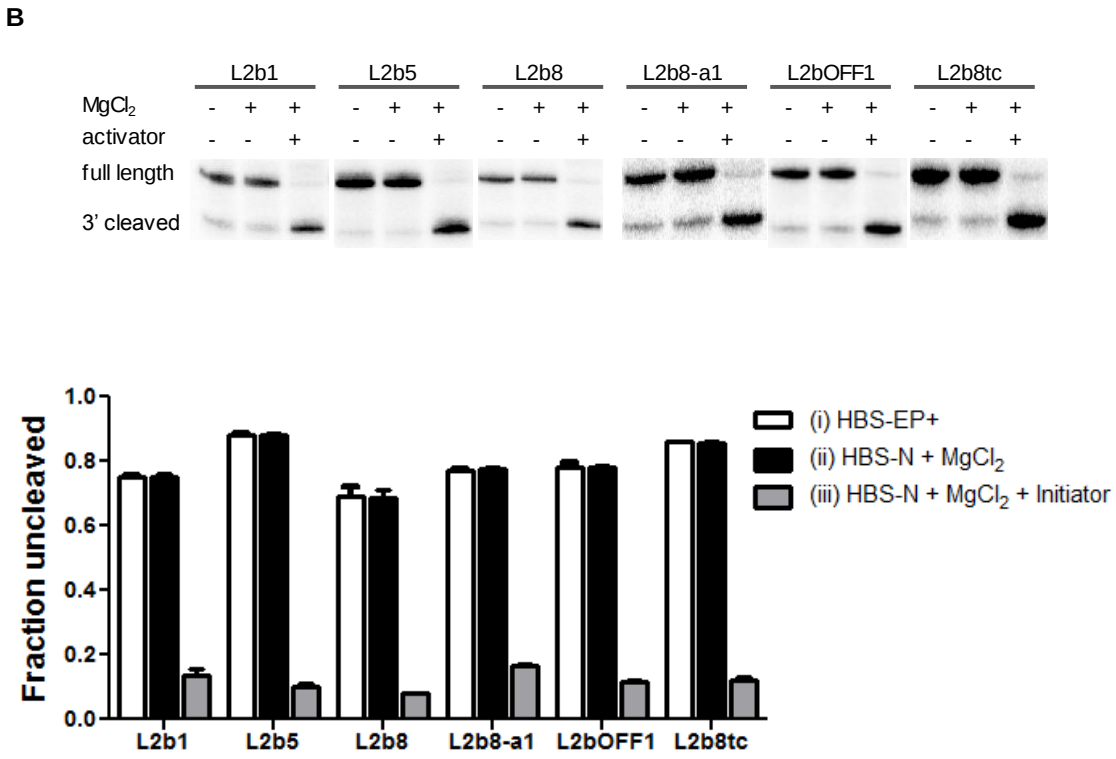


Figure 3.3 Quantification of the blocking efficiencies of the cis-blocking strategy on cis-blocked ribozyme constructs. (A) The cis-blocked sTRSV construct variants (Supplementary Figure S1) were incubated in: (i) HBS-EP buffer (10 mM HEPES, 150 mM NaCl, 3 mM EDTA, 0.05% P20, pH 7.4) (first lane); (ii) HBS-N buffer (10 mM HEPES and 150 mM NaCl, pH 7.4) in the presence of 5 mM $MgCl_2$ for 30 minutes at 25ºC (second lane); (iii) HBS-N buffer in the presence of 5 mM $MgCl_2$ and 6 μM activator for 30 minutes at 25ºC (third lane). The fraction of uncleaved fragment is determined through phosphorimaging analysis. A black vertical bar is used to denote samples run on different gels. (B) The blocking and activation sequences designed for the sTRSV ribozyme were directly incorporated into different synthetic ribozyme devices (L2b1, L2b5, L2b8, L2bOFF1, and L2b8tc) (Supplementary Figure S3.2). The transcription products of these constructs were analyzed as described for the cis-blocked sTRSV constructs. For clarity, only the full-length and 3' cleaved products are shown.

3.2.4 The sTRSV-blocking sequence provides a modular strategy for inhibiting cleavage of sTRSV ribozyme-based devices

Because the designed RNA blocking sequence targets only the sTRSV ribozyme, it is expected that the cis-blocking strategy can be applied to generate full-length RNA

for the sTRSV-based ribozyme devices. We inserted the RNA blocking and activation sequences in several previously characterized ribozyme devices composed of different sensors, transmitters, and actuators to generate the corresponding cis-blocked L2b1, L2b5, L2b8, L2b8-a1, L2bOFF1, and L2b8tc constructs (see Chapter 2) (Supplementary Figure S3.2). The transcription products of these constructs were diluted in HBS-EP buffer and analyzed on the gel to determine the blocking efficiencies (Figure 3.3B). Up to 90% of the total transcribed RNA was uncleaved, supporting that the RNA blocking sequence is modular to constructs based on the same ribozyme. In contrast, performing transcription reactions on a ribozyme construct without the blocking sequence resulted in significant cleavage (Supplementary Figure S3.3). The cis-blocked constructs were further incubated at 5 mM $MgCl_2$ in the absence and presence of the activator strand for 30 minutes at 25ºC (Figure 3.3B). Similarly, the blocked constructs required both the presence of $MgCl_2$ and activator to cleave. Up to ~80% of the full-length blocked RNA was released by the activator and cleaved, supporting that the activation sequence optimized for the sTRSV ribozyme is modular to the corresponding ribozyme devices as well.

3.2.5 Full-length RNA generated through cis- and trans-blocking strategies exhibits comparable cleavage rates

We next examined the cleavage rates (*k*) measured for the full-length RNA generated from the cis-blocking strategies. The gel-based cleavage assays were performed at physiologically-relevant reaction conditions (500 μM $MgCl_2$, 100 mM NaCl, and 50 mM Tris-HCl (pH 7.0)) at 37°C (Supplementary Figure S3.4). The blocked RNA was activated by incubating with the DNA activator strands prior to addition of $MgCl_2$. The cleavage rate constants were obtained for the L2b8 (0.063, 0.008 min^{-1}) and L2b8-a1

(0.351, 0.043 min^{-1}) constructs in the absence and presence of 5 mM theophylline, respectively (Figure 3.4A). These rates were compared to those previously obtained on the full-length RNA generated from standard trans-blocking strategies requiring PAGE purification of the blocked RNA strands (see Chapter 2) (Figure 3.4B). A correlation analysis between the cleavage rates determined through these two methods indicates a strong positive correlation (Pearson coefficient, r = 0.9999). However, the cleavage rates associated with the cis-blocking strategy are slower than those associated with the trans-blocking strategy. We postulate that the hybridizing reaction between the cis-blocked RNA and the activator strand may result in a slower apparent cleavage rate, thus requiring a longer incubation prior to the initiation of the cleavage reaction to allow complete activation of the RNA.

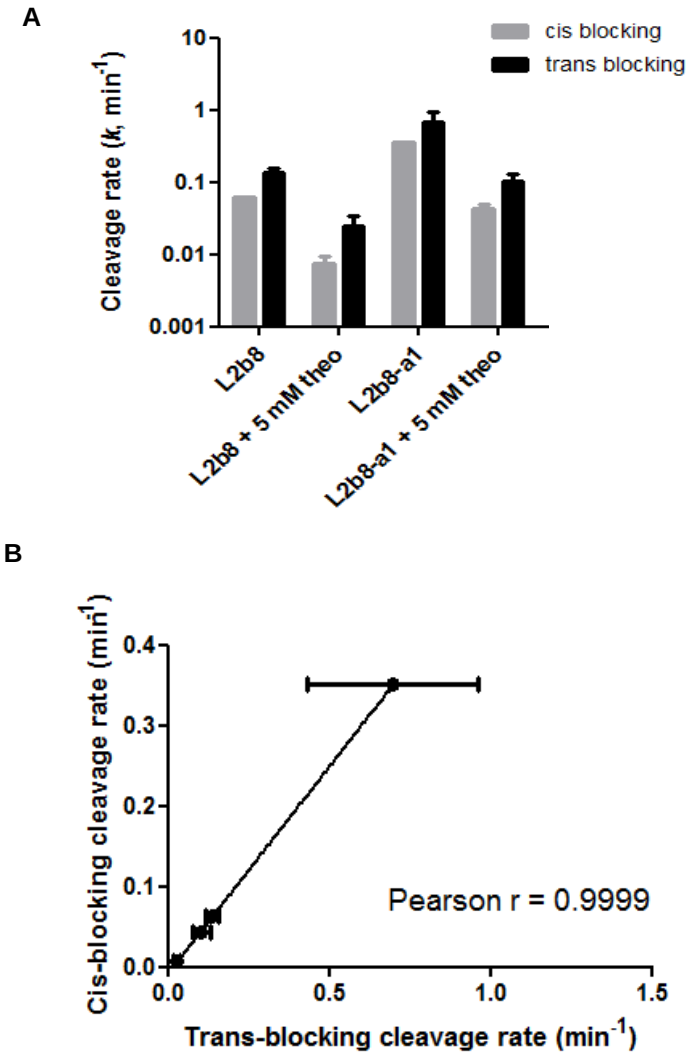


Figure 3.4 Comparison of the ribozyme cleavage rates on full-length RNA generated from different blocking strategies. (A) Gel-based cleavage assays were performed on the cis-blocked L2b8 and L2b8-a1 constructs at 37˚C with 500 μM $MgCl_2$, 100 mM NaCl, 50 mM Tris-HCl (pH 7.0). Cleavage rate constants (*k*) and errors are reported as the mean and standard deviation from at least three independent assays. The cleavage rate constants for the trans-blocked L2b8 (0.14, 0.025 min^{-1}) and L2b8-a1 (0.7, 0.1 min^{-1}) constructs were characterized previously at the same reaction conditions in the absence and presence of 5 mM theophylline, respectively (see Chapter 2). (B) Correlation analysis of cleavage rates of full-length RNA generated from the cis- and trans-blocking strategies indicates a strong correlation between these two rates; Pearson correlation coefficient (r) of 0.9999.

3.2.6 A label-free SPR-based assay to support rapid monitoring of ribozyme cleavage

Our cis-blocking and trans-activating strategies allow the development of a label-free and continuous cleavage assay based on the Biacore sensor platform (Figure 3.5). We generated a reaction sensor surface by covalently linking the DNA activator strand to the sensor surface (Figure 3.5A). The reaction sensor surface then allows both the capture and activation of the cis-blocked RNA through programmed hybridization interactions (Figure 3.5B). The capture of RNA is monitored in real time and represented as an increase in the SPR signal due to increased mass density near the sensor surface. The cleavage reaction is initiated by the injection of a buffer containing $MgCl_2$ (Figure 3.5C). Cleavage of the RNA results in a 3' cleaved fragment that is weakly hybridized to the immobilized DNA activator strand through five bases (14). Therefore, the 3' cleavage fragment rapidly dissociates, resulting in a decrease in the SPR signal due to decreased mass density near the sensor surface. The observed RNA dissociation rate, which reflects how fast the ribozyme construct cleaves, is obtained by fitting the sensorgram to a first-order exponential decay equation. Because the DNA-based reaction sensor surface is highly stable, the surface can be regenerated by an injection of nuclei acid denaturants (i.e., NaOH) to remove residual RNA and reused for the next cleavage assay (Figure 3.5D).

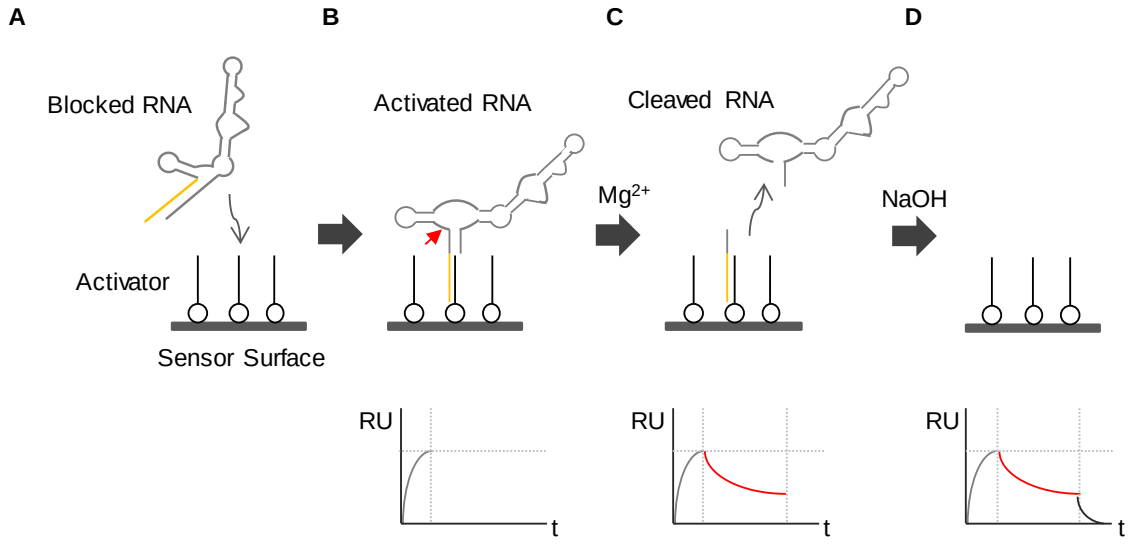


Figure 3.5 Schematic of the SPR-based ribozyme cleavage assay. (A) A sensor surface for ribozyme cleavage characterization is generated by conjugating the DNA activator strand directly to the chip surface. (B) The blocked ribozyme construct is injected over the activator-coated surface, and the hybridization interaction between the ribozyme construct and the activator, resulting in an increase in SPR signal, represented by resonance units (RU). The hybridization of the ribozyme constructs allows the activated of the blocked RNA. (C) The cleavage reaction is initiated by the injection of buffers containing $MgCl_2$ over the sensor surface. Ribozyme cleavage results in a quick dissociation of the 3' cleaved fragment, resulting in a decrease in SPR signal. The observed RNA dissociation rate is obtained by fitting the injection portion of the sensorgram to a first-order exponential decay equation (see Materials and Methods). (D) Finally, the surface is regenerated by an injection of 25 mM NaOH to remove residual RNA for the next assay.

3.2.7 The observed RNA dissociation rates are closely correlated to the cleavage rates obtained through the gel-based assay

We applied the SPR-based cleavage assays to monitor the cleavage of cis-blocked sTRSV, L2b8, and L2b8-a1 constructs under similar reaction conditions used for the gel-based cleavage assays (Supplementary Figure S3.5). The reactions were initiated by injecting buffers (500 μM $MgCl_2$, 100 mM NaCl, and 50 mM Tris-HCl (pH 7.0)) containing: (i) no $MgCl_2$; (ii) 500 μM $MgCl_2$; and (iii) 500 μM $MgCl_2$ and 5 mM

theophylline (Figure 3.6A). Previous ribozyme kinetic studies have demonstrated that much higher NaCl concentrations (>4 M) are required for ribozymes to exhibit cleavage activities (18). Thus, the ribozyme constructs are expected to exhibit little cleavage activity in the absence of $MgCl_2$ (reaction condition (i)), and the decrease in SPR signal over the course of injection is solely contributed by spontaneous RNA dissociation (full-length and 5' cleavage fragments) from the activator strands.

The RNA dissociations rates were obtained through model fitting for each construct under different reaction condition (Figure 3.6B). The measured spontaneous RNA dissociation rates (reaction condition (i)) ranged from 0.05 to 0.18 min^{-1}. The presence of theophylline has little nonspecific effect on the observed dissociation rates by examining the rates obtained for the sTRSV construct in the absence (1.66 min^{-1}) and presence (1.54 min^{-1}) of 5 mM theophylline. The observed RNA dissociation rates obtained for L2b8 (0.13, 0.09 min^{-1}) and L2b8-a1 (0.55, 0.17 min^{-1}) in the absence and presence of 5 mM theophylline, respectively, were generally in close agreement with those obtained through the trans-blocking gel-based assays (Figure 3.4B). However, the measured RNA dissociation rate (0.09 min^{-1}) for the L2b8 construct in the presence of theophylline was significantly greater than the corresponding gel-based cleavage rate (0.025 min^{-1}) obtained from the trans-blocking strategy, and was comparable to the spontaneous RNA dissociation rate (0.05 min^{-1}) measured in the absence of $MgCl_2$. This result suggests that our current SPR-based assay is limited by the spontaneous dissociation of RNA from the sensor surface, and thus may not be applicable to characterize ribozyme constructs that exhibit slow cleavage rates.

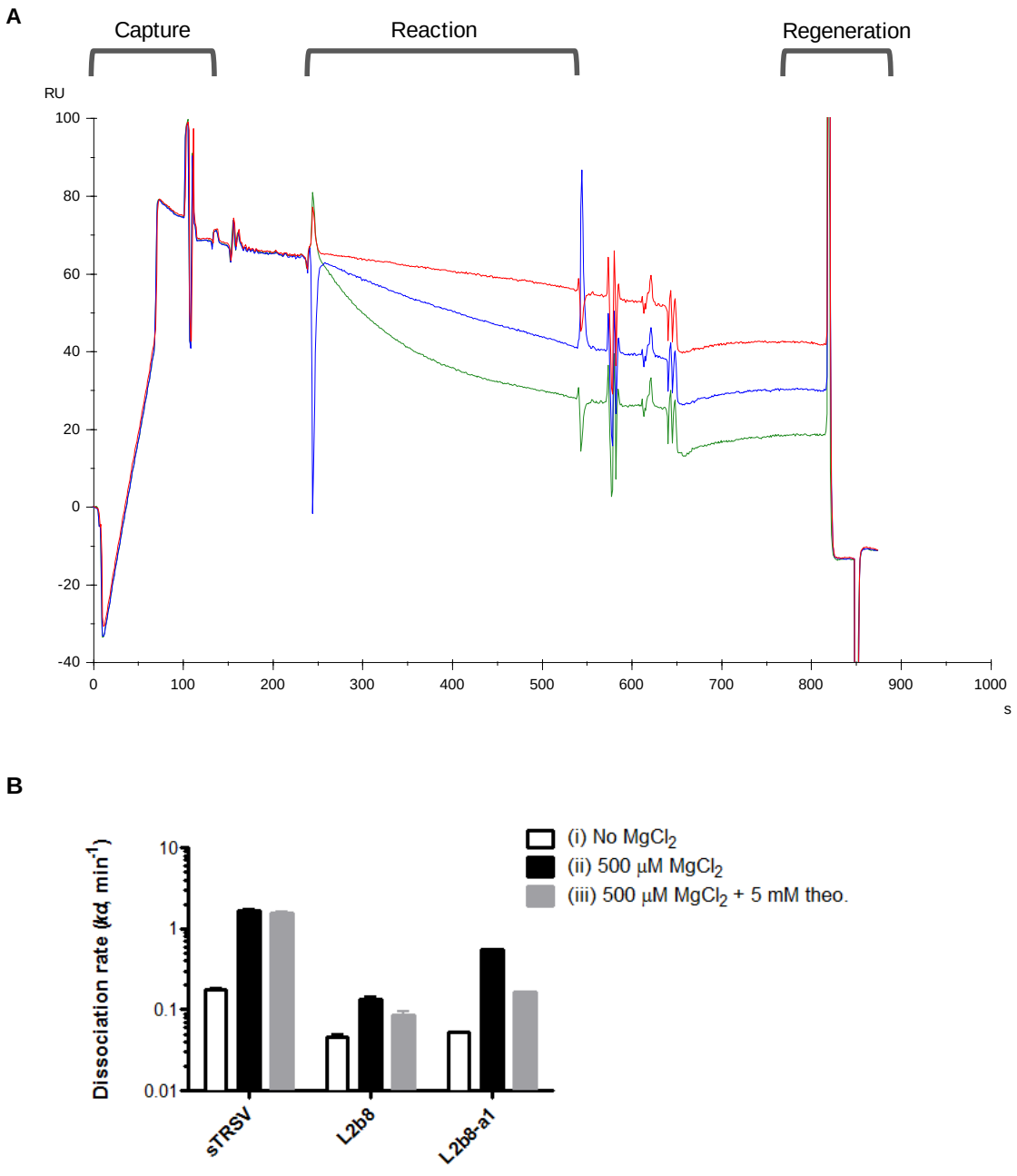


Figure 3.6 The observed RNA dissociation rate are closely correlated to the corresponding cleavage rates. (A) The SPR-based assays were performed on the cis-blocked sTRSV, L2b8, and L2b8-a1 constructs. The reactions were initiated by injecting buffers (500 µM $MgCl_2$, 100 mM NaCl, and 50 mM Tris-HCl (pH 7.0)) containing: (i) no $MgCl_2$; (ii) 500 µM $MgCl_2$; and (iii) 500 µM $MgCl_2$ and 5 mM theophylline. A representative sensorgram for the characterization of the L2b8-a1 construct at three different assay conditions is shown here. The detailed assay steps are indicated. (B) The injection portions of the sensorgrams were fit to a first-order decay equation to obtain the RNA dissociation rate constants (*kd*) (Supplementary Figure S3.5). RNA dissociation rate

constants and errors are reported as the mean and standard deviation from at least three independent assays.

3.3 DISCUSSION

We described a ribozyme-blocking strategy to support efficient generation of full-length RNA transcripts containing ribozyme constructs through *in vitro* transcription reactions. In contrast to the traditional trans-blocking strategy that requires laborious PAGE purification process for the recovery of full-length transcription products, our cis-blocking strategy offers a rapid two-step process (transcription and activation) to yield large quantities of full-length RNA. We demonstrated the cis-blocking strategy on a variety of different ribozyme elements, including natural ribozymes and synthetic ribozyme devices. The cleavage rate constants associated with the full-length RNA generated from the cis- and trans- blocking strategies were shown to be in close agreement and correlate with one another, thus supporting our cis-blocking strategy as an efficient method for ribozyme characterization.

We employed a robust toehold-mediated strand-displacement reaction to facilitate the activation of RNA. A similar toehold-mediated activation mechanism has been applied to the kinetic study of the hepatitis delta virus (HDV) ribozyme through the use of antisense oligomers that target the upstream natural cis-blocking sequence (19). We rationally designed the blocking, activation, and stabilizing sequences to achieve an optimal blocking efficiency up to ~90% for the sTRSV ribozyme. Previous kinetic studies on the sTRSV ribozyme have shown that ~50% of the full-length RNA generated by the trans-blocking strategy remained uncleaved at the end of reaction (9). In contrast, the full-length RNA generated by our cis-blocking strategy resulted in only ~16%

uncleaved products at the end of reaction, suggesting that full-length RNA prepared by our method is less prone to misfolding.

Our cis-blocking strategy utilizes a trans-acting DNA strand to activate the blocked RNA, allowing the development of an 'on-chip' activation strategy by generating a reaction surface immobilized with DNA activator strands. We developed a rapid, label-free cleavage assay based on SPR, which allows real-time monitoring of ribozyme cleavage. Our SPR-based assay offers several unique advantages over the more traditional gel-based assay in that: (i) it requires no radioactivity; (ii) it requires little material (microfluidics-based flow system); and (iii) it is highly automatable. However, to broaden the utility of our SPR-based strategy for the quantification of ribozyme cleavage rates, efforts need to be made to address several immediate challenges. Our strategy is currently based on a capture-based sensor surface. Although such strategy allows the reuse of the surface by regeneration, the measured observed RNA dissociation rate is contributed to both cleavage-dependent (3' cleaved fragments) and spontaneous (full-length and 5' cleaved fragments) dissociation rates. Thus, the spontaneous RNA dissociation rate prevents the quantitative measurement of cleavage rates solely based on the SPR-based assays. In addition, our current SPR-based assay is limited by the maximum injection volume allowed for the given instrument, which presents a challenge for the characterization of ribozyme constructs exhibiting low cleavage activities (thus requiring a longer injection).

To allow direct measurement of cleavage rates through the SPR-based assays, efforts can be directed toward enhancing the stability of the hybridization reaction between the RNA and the DNA activator strand by extending the length of the RNA

activation sequence through its 5’ end. Further optimization of the blocking and stabilizing sequences may be necessary to ensure that a longer activation sequence does not interact nonspecifically with neighboring sequences, thus impacting the blocking efficiency. Alternative activator backbone, such as peptide nucleic acid (PNA), may also be considered to enhance the stability of the RNA-captured sensor surface, as the PNA-RNA duplex has been shown to exhibit greater thermal stability than the corresponding PNA-DNA and DNA-DNA duplexes (20). To expand the range of cleavage activities that can be analyzed by the SPR-based assays, strategies that allow a longer reaction have been described by either directly using the characterization buffer as the instrument running buffer (thus requiring no injection of the characterization buffer) or modifying the SPR instrument with an external peristaltic pump (21). With further development and optimization, the SPR-based cleavage assay will offer a powerful strategy for rapidly quantifying cleavage rates for a wide variety of ribozyme constructs, thereby supporting the functional design of RNA for diverse applications.

3.4 MATERIALS AND METHODS

3.4.1 Preparation of DNA templates for T7 transcription reaction

DNA synthesis was performed by Integrated DNA Technologies (Coralville, IA) or the Protein and Nucleic Acid Facility (Stanford, CA). PCR products were used as DNA templates for *in vitro* T7 transcription reaction. For the cis-blocking strategy, PCR products were amplified from templates using forward primer T7-fwd (5’-TTCTAATACGACTCACTATAGG) and corresponding reverse primers that are specific to each construct (Supplementary Table S3.1), respectively.

3.4.2 Generation of full-length RNA through cis-blocking strategy

A total of 1-2 μg of PCR product was transcribed in a 25 μl reaction, consisting of the following components: 1×RNA Pol Reaction Buffer (New England Biolabs, Ipswich, MA), 2.5 mM of each rNTP, 1 μl RNaseOUT (Invitrogen, Carlsbad, CA), 10 mM $MgCl_2$, 1 μl T7 Polymerase (New England Biolabs), and 0.5 μCi α-^{32}P-GTP (MP Biomedicals, Solon, OH). After incubation at 37°C for 2 hr, NucAway Spin Columns (Ambion, Austin, TX) were used to remove unincorporated nucleotides from the transcription reactions according to manufacturer's instructions. The transcription products were purified with the RNA Clean & Concentrator^{TM}-5 (Zymo Research, Irvine, CA) kit according to manufacturer's instructions.

3.4.3 Gel-based ribozyme cleavage assays

All gel-based ribozyme cleavage assays were performed in a physiologically-relevant reaction buffer composed of 500 μM $MgCl_2$, 100 mM NaCl, and 50 mM Tris-HCl (pH 7.0) at 37°C. Full-length RNA generated from the cis-blocking strategy were first incubated with 60 pmol DNA activator strand (5'- AAACAACTTTGTTTGTTTCC CCC) for 2 min to activate the blocked RNA. Full-length RNA generated from the trans-blocking strategy were incubated at 95°C for 5 min, cooled at a rate of -1.3°C to 37°C, and held there for 10 min to allow equilibration of secondary structure. A zero time-point aliquot was taken prior to initiating the self-cleavage reaction with the addition of $MgCl_2$. Reactions were quenched at specified time points with addition of 3 volumes of RNA stop/load buffer (95% formamide, 30 mM EDTA, 0.25% bromophenol blue, 0.25% xylene cyanol) on ice. Samples were heated to 95°C for 5 min, snap cooled on ice for 5

min, and size-fractionated on a denaturing (8.3 M Urea) 10% polyacrylamide gel at 25 W for 45 to 60 min. Gels were exposed overnight on a phosphor screen and imaged on a FX Molecular Imager (Bio-Rad, Hercules, CA). The relative levels of the full-length transcript and cleaved products were determined by phosphorimaging analysis. The cleaved product fraction at each time point (F_t) was fit to the single exponential equation $F_t = F_0 + (F_\infty - F_0)\times(1 - e^{-kt})$ using Prism 5 (GraphPad), where F_0 and F_∞ are the fractions cleaved before the start of the reaction and at the reaction endpoint, respectively, and *k* is the first-order rate constant of self-cleavage. All reported cleavage rates are the mean of at least three independent experiments.

3.4.4 Biacore sensor chip surface generation

Biosensor experiments were performed on a Biacore X100 instrument (Biacore, Uppsala, Sweden). A CM5 sensor chip (Biacore) was docked in the Biacore X100 and equilibrated with HBS-N buffer at 25ºC. The DNA activator strand (5'-AAACAACTTTGTTTGTTTCCCCC-/AmMO/), with an amino modification on its 3' end, was immobilized to the chip surface via standard amine-coupling chemistry. Briefly, the carboxymethyl surface of the CM5 chip was activated for 7 min at a flow rate of 5 μl/min using a 1:1 volume ratio of 0.4 M 1-ethyl-3-(3-dimethylaminopropyl) carbodiimide (Biacore) and 0.1 M N-hydroxysuccinimide (Biacore). A molar ratio of 1:60 of DNA activator to cetyl trimethylammonium bromide (Amresco, Solon, OH) was diluted in 10 mM HEPES buffer (Sigma, St. Louis, MO) to a final concentration of 10 μM and 0.6 mM, respectively, and injected over the activated surface for 10 min at a flow rate of 5 μl/min. Excess activated groups were blocked by an injection of 1 M

ethanolamine (Biacore), pH 8.5, for 7 min at a flow rate of 5 μl/min. The immobilization reaction was performed on both flow cells (FC1, FC2), where FC1 was used as a reference cell to correct for bulk refractive index changes, matrix effects, nonspecific binding, injection noise and baseline drift (22). Approximately 1600 RU of the activator strand was immobilized using this protocol.

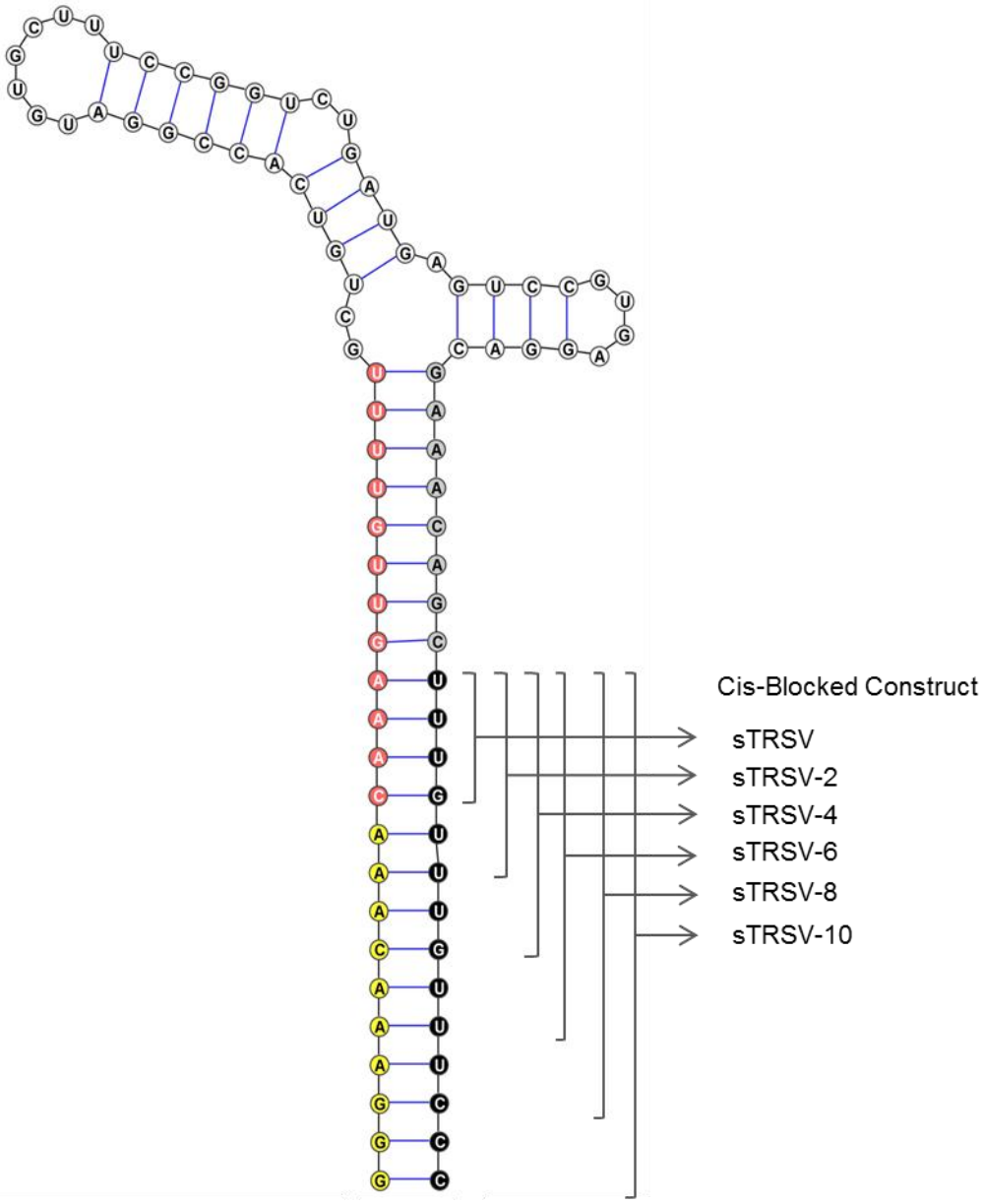
3.4.5 Label-free SPR-based ribozyme cleavage assays

Full-length RNA was prepared by the cis-blocking strategy without the radiolabelled nucleotides. The Biacore X100 instrument was equilibrated with the physiologically-relevant reaction buffer at 37ºC prior to all ribozyme cleavage assays. The SPR baseline was stabilized by performing 2-5 startup cycles, each cycle includes a capture and a regeneration step. The capture step was performed by an injection of a total of 10-25 ng transcribed RNA diluted in the reaction buffer over the reaction flow cell (FC2) for 1 min at a flow rate of 10 μl/min. The capture step typically yielded ~50-300 RU of the SPR signal for the described constructs. The regeneration step was performed by an injection of 25 mM NaOH over both flow cells for 30 sec at a flow rate of 30 μl/min.

Following the startup cycles, assay cycles were performed. Each assay cycle includes a capture, a reaction, and a regeneration step. The capture and regeneration steps were performed the same as those in the startup cycle. The reaction step was performed by an injection of the running buffer or running buffers containing 500 μM $MgCl_2$ with or without 5 mM theophylline over both FCs for 5 min at a flow rate of 10 μl/min. Biacore sensorgram processing and analysis were performed using Biacore X100 Evaluation

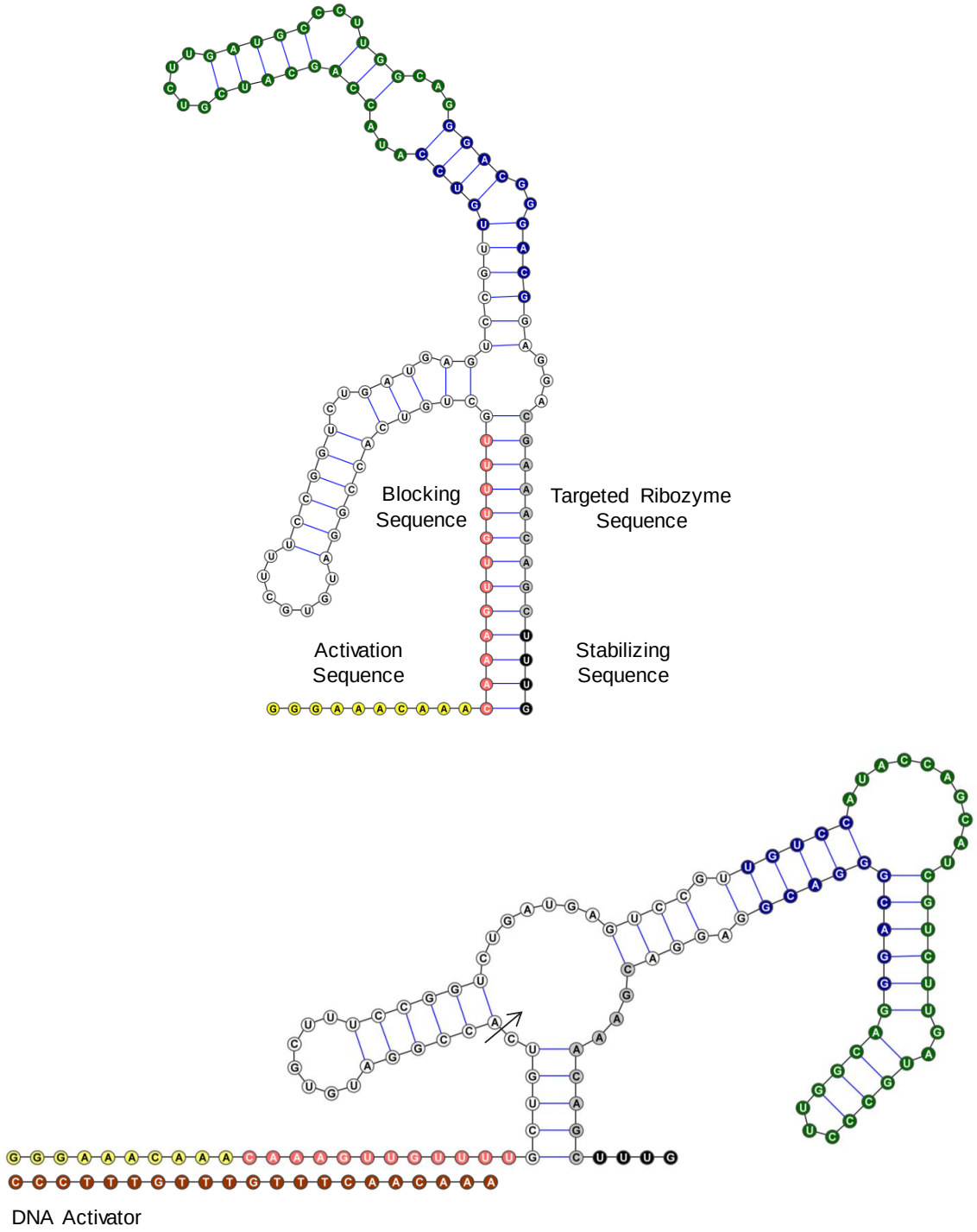
Software v2.0 (Biacore). Due to the slight time delay at which injected analyte reaches the respective flow cells, the resultant sharp spikes at the beginning and the end of injection were excluded from the analysis (23). The processed sensorgram (R) was fit to a simple exponential equation $R = R_0 \times (e^{-kdt} - 1) + \text{offset}$, where R_0 (fit globally for a given replicate) is the initial SPR signal before the cleavage reaction, offset (fit locally for a given replicate) is the residual response at the end of the cleavage reaction, and *kd* is the first-order RNA dissociation rate constant. Reported values are the mean of at least three independent experiments.

3.5 Supplementary Information

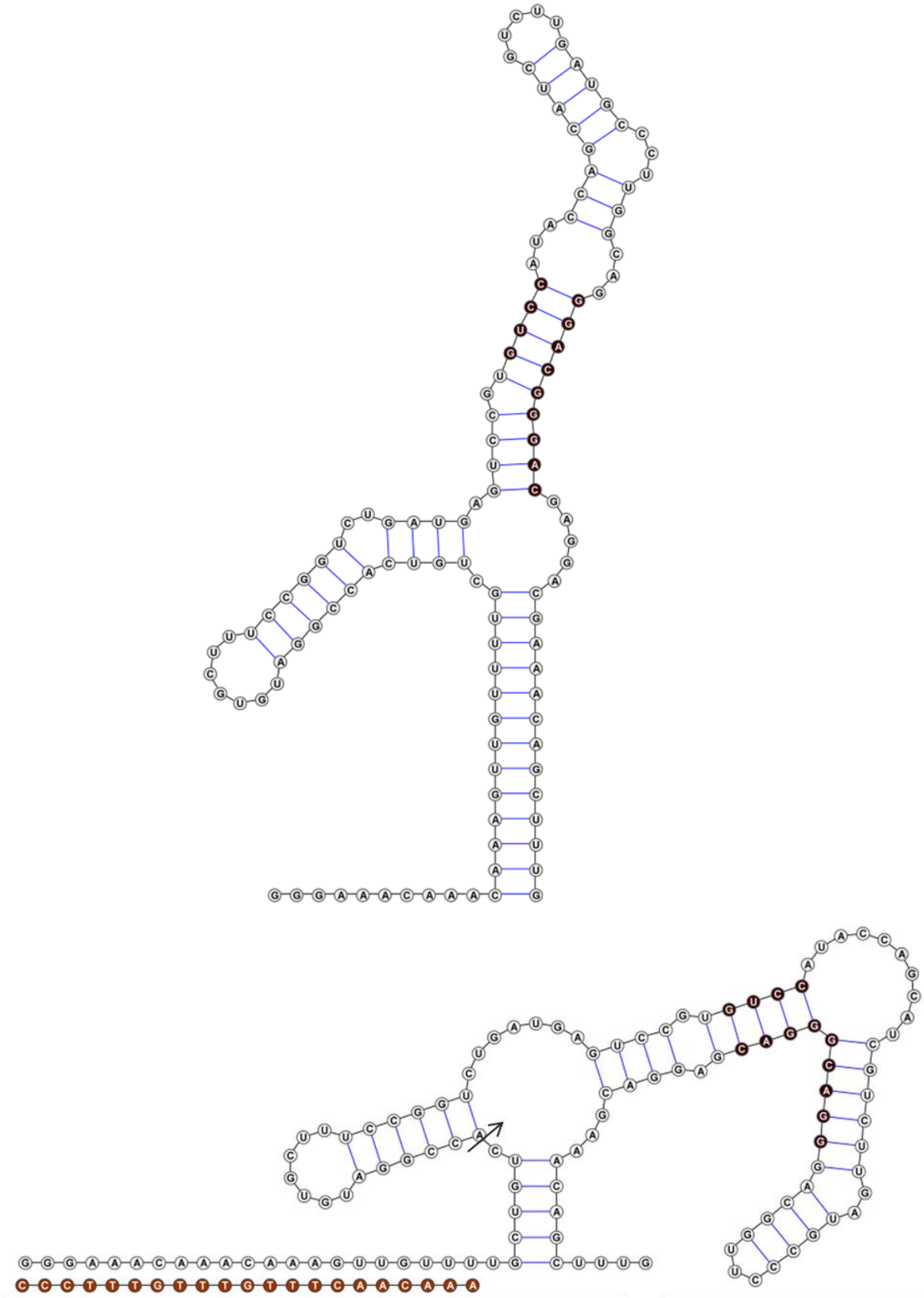


Supplementary Figure S3.1 Cis-blocked sTRSV constructs with varying stabilizing sequence length. The stabilizing sequence in the cis-blocked sTRSV construct (Figure 3.2) is increased in 2 nt increment to generate sTRSV-2, -4, -6, -8, and -10 constructs (Supplementary Table S3.1). The targeted ribozyme sequence, and the RNA blocking, activation, and stabilizing sequences are indicated in grey, red, yellow, and black, respectively. The additional 2, 4, 6, 8, and 10 nt to the stabilizing sequence in the original cis-blocked sTRSV construct results in a toehold length of 8, 6, 4, 2, and 0 nt, respectively.

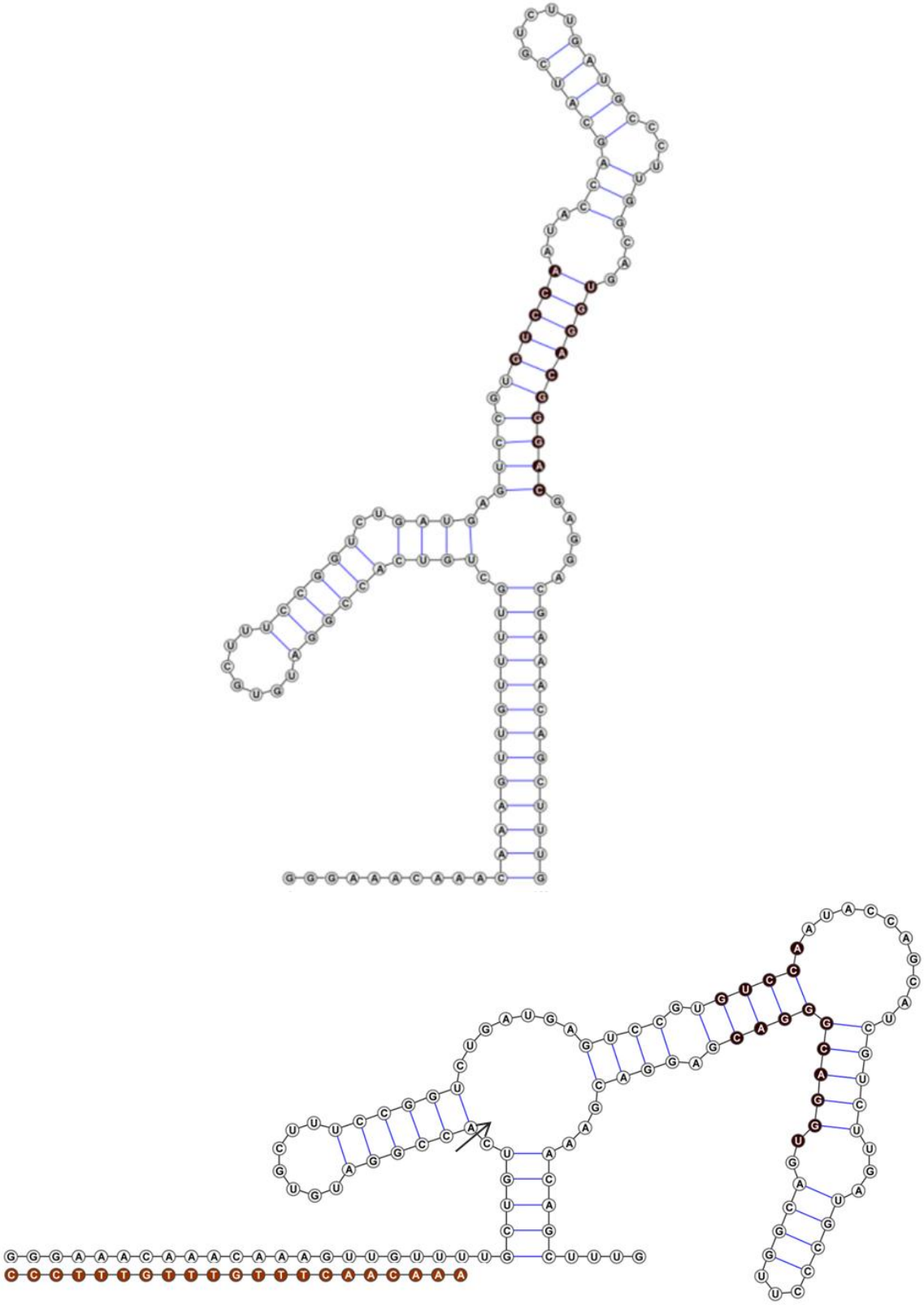
A L2b8


Blocking Sequence
Targeted Ribozyme Sequence
Activation Sequence
Stabilizing Sequence
DNA Activator

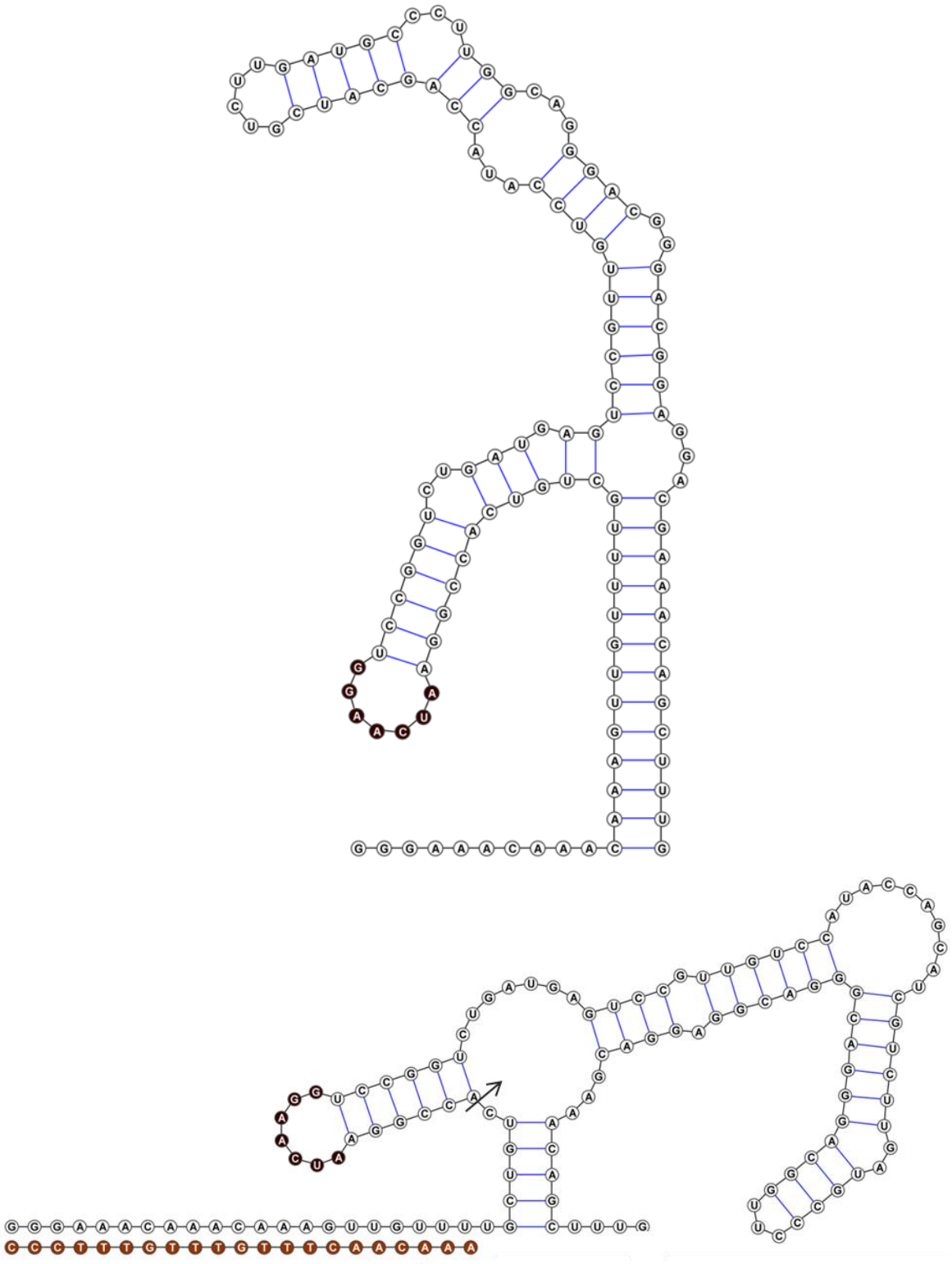
B L2b1



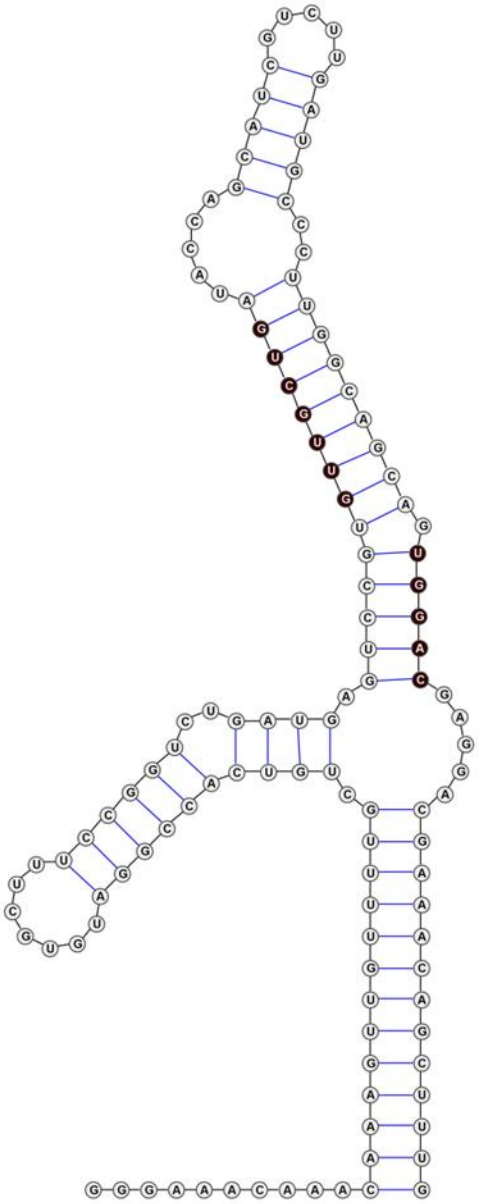
C L2b5

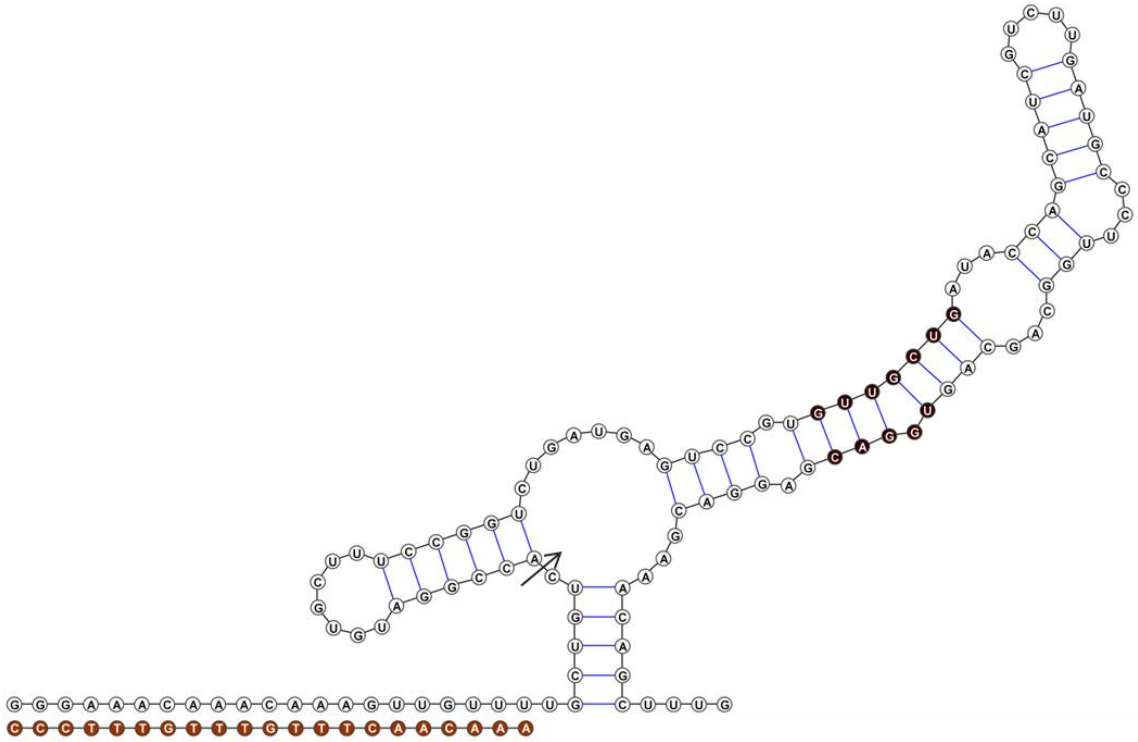


D L2b8-a1

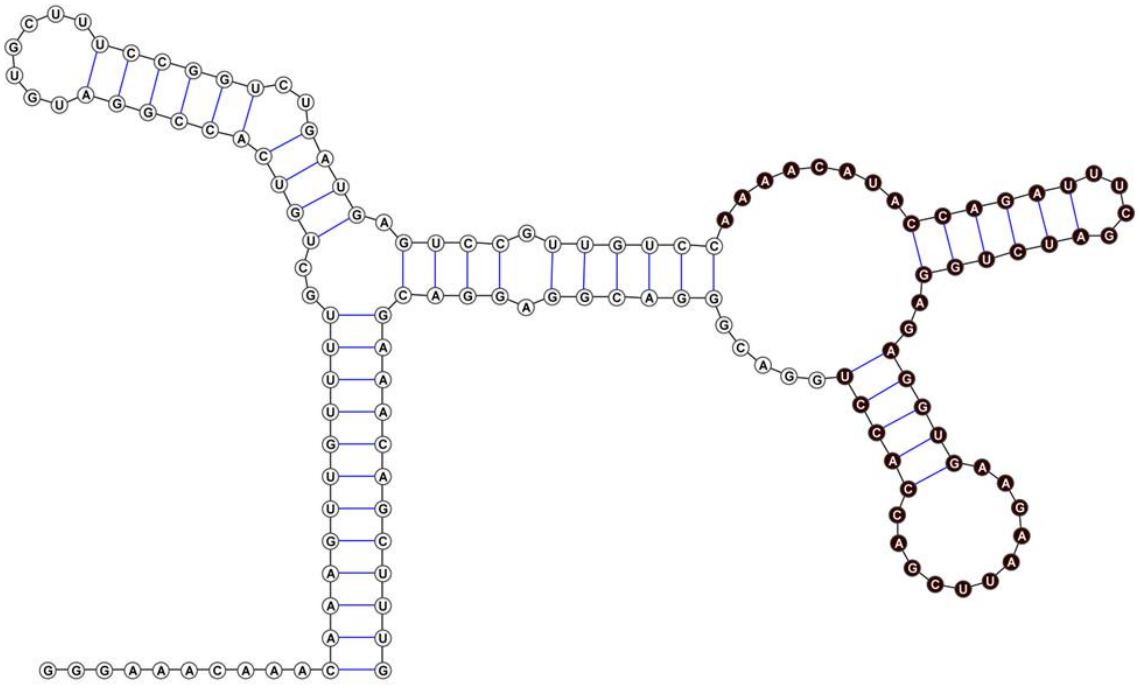


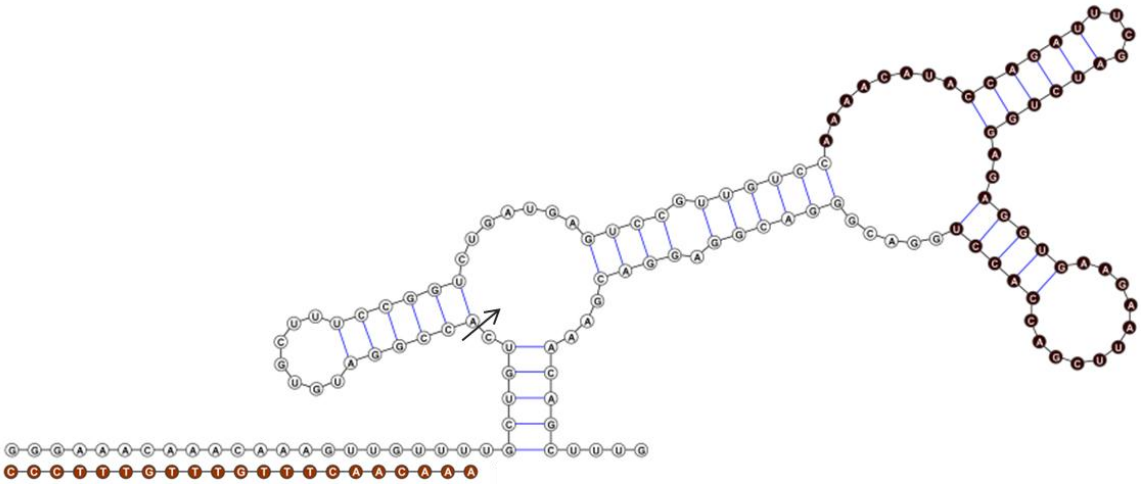
E L2bOFF1



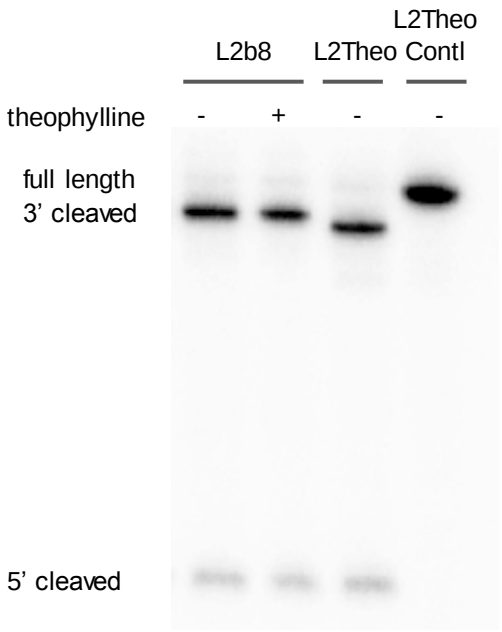


F L2b8tc



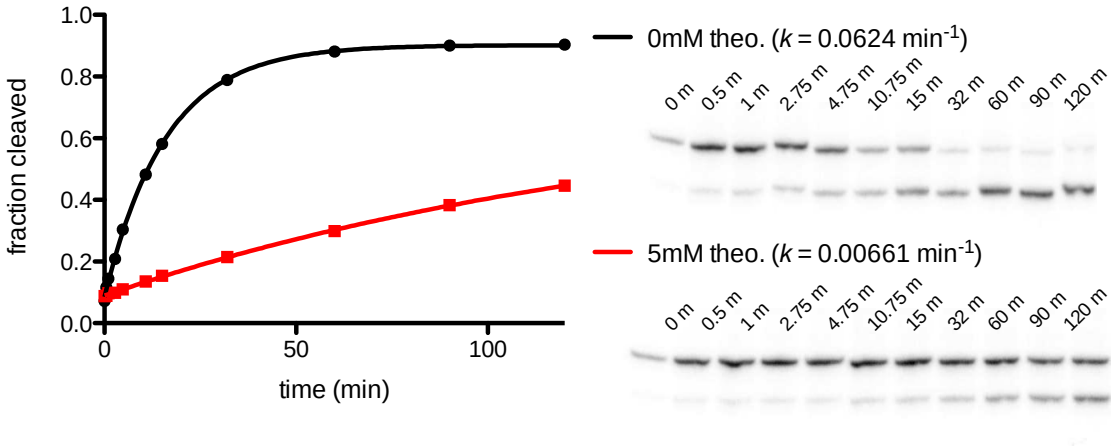


Supplementary Figure S3.2 Cis-blocked constructs for synthetic ribozyme devices composed of different functional RNA components. The same blocking (targeting the same sTRSV ribozyme sequence), activation, and stabilizing sequences shown in Figure 2 were directly incorporated into the synthetic ribozyme devices to generate the corresponding cis-blocked ribozyme constructs (Supplementary Table S3.1). The targeted ribozyme sequence, RNA blocking, activation, stabilizing sequences, and DNA activator are indicated in grey, red, yellow, black, and brown, respectively, for the (A) L2b8 construct. The sensor (RNA aptamer), transmitter, and actuator (sTRSV ribozyme) in the L2b8 construct are indicated in green, blue, and white, respectively. For the subsequent constructs, only components that are different from those in the L2b8 device are indicated in brown. The blocked (top) and activated (bottom) conformations are shown for each construct. The (A) L2b8, (B) L2b1, and (C) L2b5 devices are theophylline-responsive ribozyme ON devices that differ slightly in the transmitter sequences. The (D) L2b8-a1 device was modified from the L2b8 device by incorporating a ribozyme variant that exhibits improved cleavage activity (see Chapter 2). The (E) L2bOFF1 device is a theophylline-responsive ribozyme OFF device that is composed of a transmitter sequence that is different from those in the ribozyme ON devices. The (F) L2b8tc device is a tetracycline-responsive ribozyme ON device that was generated by directly replacing the theophylline aptamer in the L2b8 device with the tetracycline aptamer (see Chapter 2). Secondary structures were predicted by RNAstructure folding software (16) and rendered using VARNA software (8).

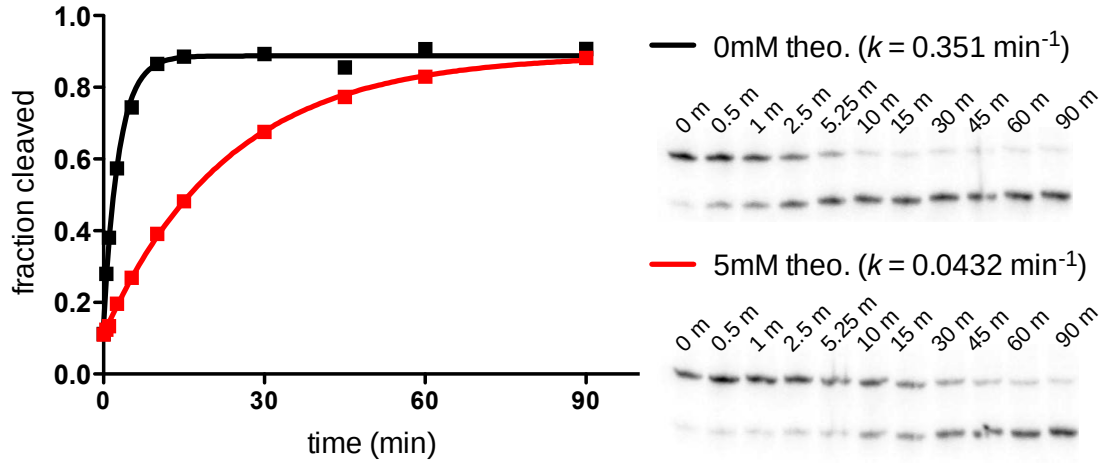


Supplementary Figure S3.3 Gel analysis of the transcription products for the L2b8 construct without the cis-blocking sequence. PCR products, which were used as DNA templates for *in vitro* T7 transcription reactions, were amplified using forward and reverse primers T7-L1-2-fwd (5'- TTCTAATACGACTCACTATAGGGACCTAGGAA ACAAACAAAGCTGTCACC) and L1-2-rev (5'-GGCTCGAGTTTTTATTTTTCTTTT TGCTGTTTCG), respectively. The transcription reactions were performed in the absence and presence of 5 mM theophylline at the same reaction conditions as described for the generation of cis-blocked RNA (see Materials and Methods). A previously described non-switch control, L2Theo (3), which lacks the transmitter sequence in the L2b8 device (thus only adopting the ribozyme-inactive conformation and cleaves) was transcribed, resulting in 5' (25 nt) and 3' (109 nt) cleaved fragments. In addition, a ribozyme-inactive control, L2Theo Contl, generated by randomizing the ribozyme catalytic core in the L2Theo construct (thus abolishing the ribozyme cleavage activity) was transcribed, resulting in a full-length fragment (134 nt). The transcription products were size-fractionated on a denaturing (7 M urea) 10% polyacrylamide gel at 25 W for 45 min and imaged by phosphorimaging analysis. The transcription of the L2b8 construct without the cis-blocking sequence resulted in 5' (25 nt) and 3' (116 nt) cleavage fragments, and little full-length (141 nt) fragment was observed.

A L2b8:



B L2b8-a1:



C

RNA Device	**Cleavage Rate (*k*, min^{-1})**	
	0 mM theophylline	**5 mM theophylline**
L2b8	0.063 ± 0.001	0.008 ± 0.002
L2b8-a1	0.351 ± 0.007	0.043 ± 0.007

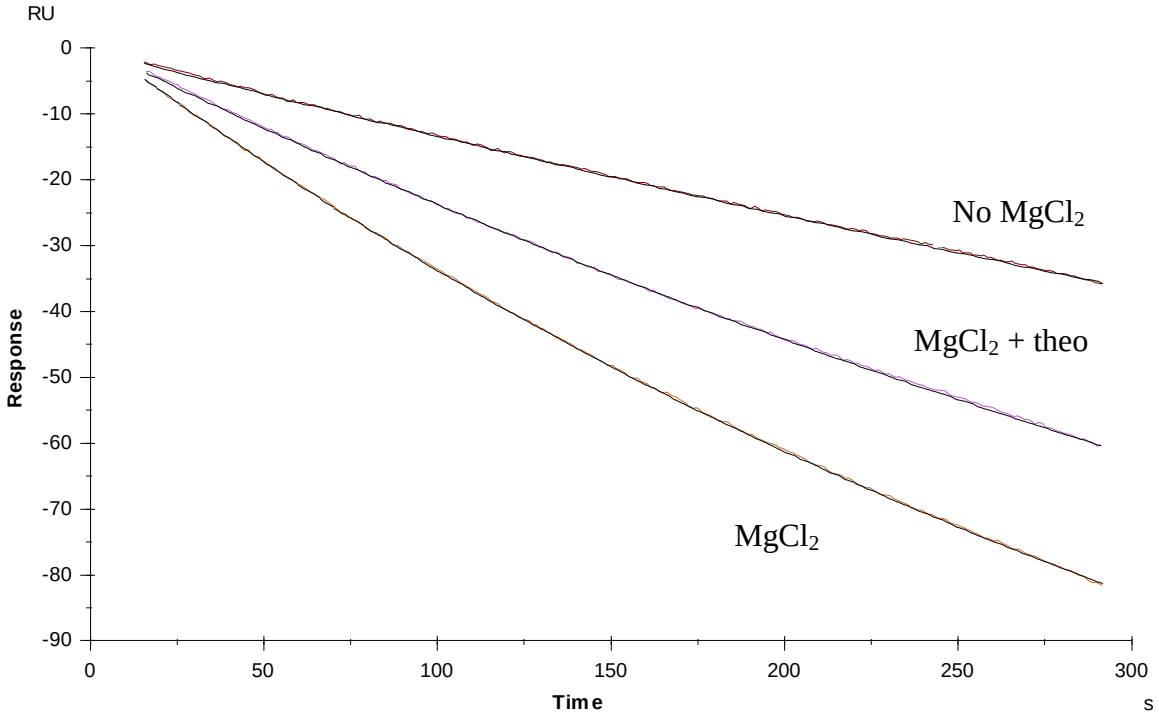
Supplementary Figure S3.4 Representative gel-based cleavage assays for measuring cleavage rate constants (*k*) for RNA generated from the cis-blocking strategy. A representative assay is shown for each cis-blocked construct in the absence and presence (0 and 5 mM, respectively) of theophylline: (A) L2b8 and (B) L2b8-a1. Bands for the full-length uncleaved substrate (UC) and longer cleaved product (3'C) are shown, the shorter 5'C product is omitted from the inset image for clarity. Methods used to prepare

full-length, uncleaved RNA transcripts and conditions of the cleavage assays are detailed in the Materials and Methods section of the main text. Briefly, cis-blocked RNA was incubated with the DNA activator in a buffer (100 mM NaCl, 50 mM Tris-HCl (pH 7.0)) for 2 min to activate the blocked RNA. A zero time point aliquot was removed prior to initiating the reaction with addition of $MgCl_2$ to a final concentration of 500 μM. Reactions were quenched at the indicated time points. Phosphorimaging analysis of relative levels of the UC, 5'C, and 3'C bands was used to determine the fraction cleaved at each time point (F_t). The fraction cleaved at the beginning (F_0) and end of reaction (F_∞) varied between assays, but all assays were well-fit to the single exponential equation ($R^2 > 0.95$):

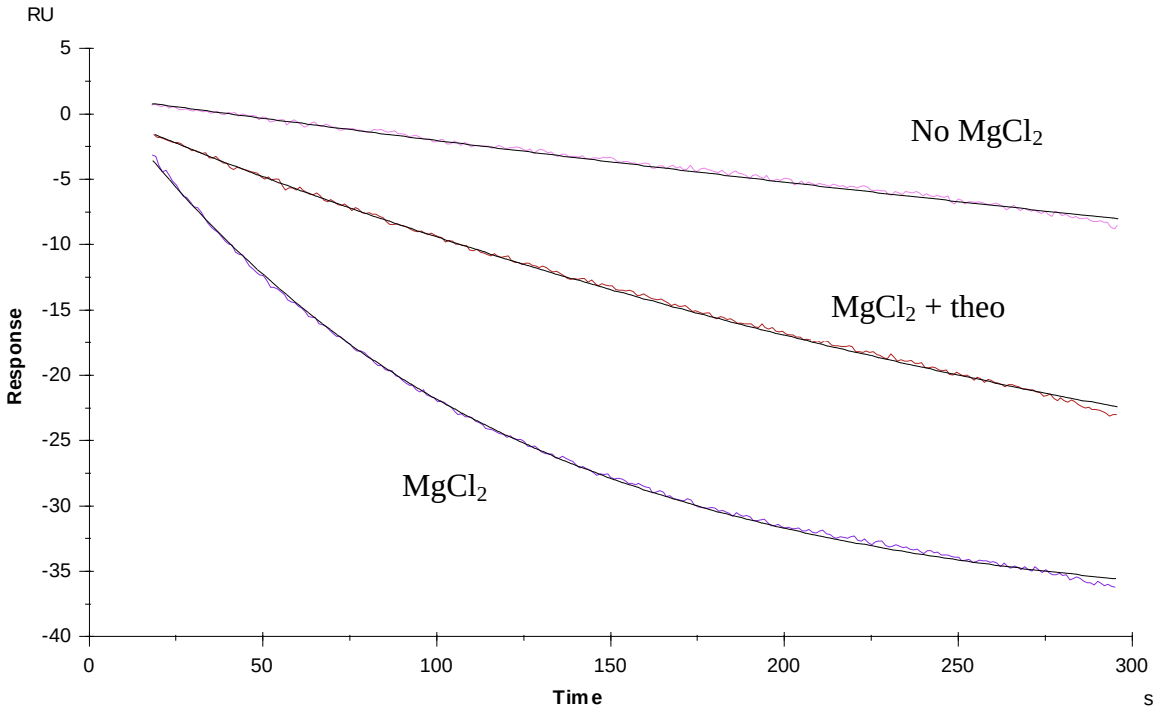
$$F_t = F_0 + (F_\infty - F_0) \times (1 - e^{-kt})$$

The black and red fit lines represent assays performed at 0 and 5 mM theophylline, respectively. (C) The cleavage rate constant value (*k*) was determined for each assay. The reported *k* for each device and theophylline assay condition is the mean and standard deviation of at least three independent experiments.

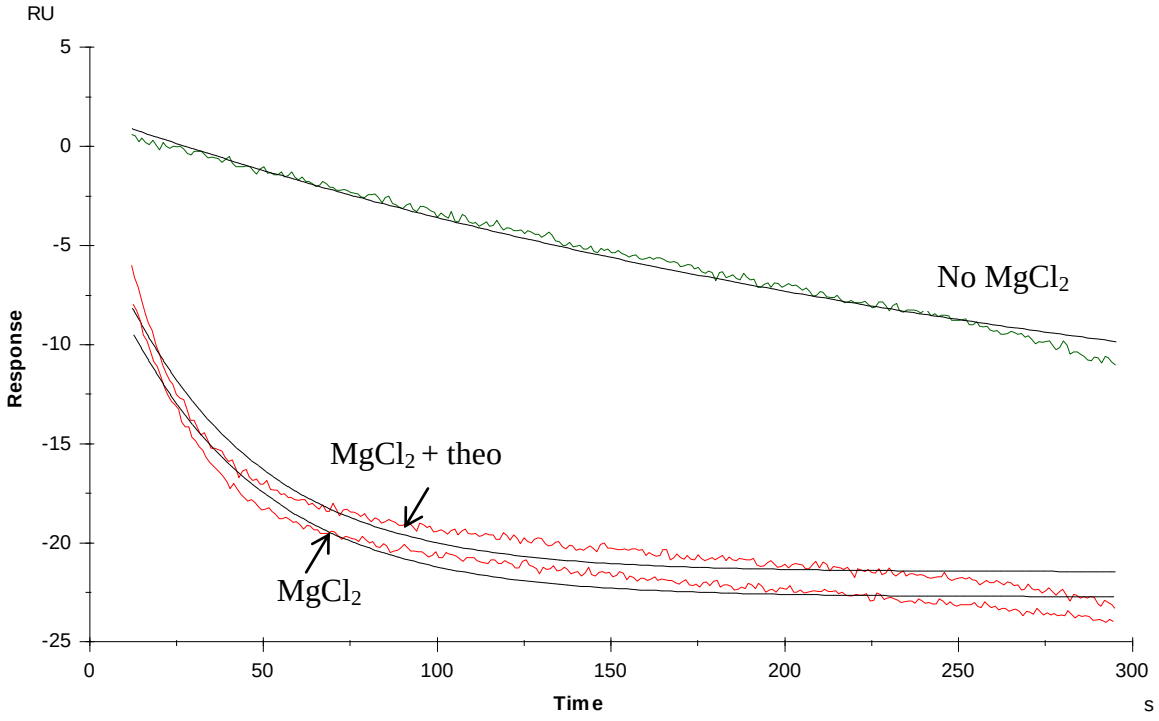
A L2b8 (Chi^2: 0.0255):



B L2b8-a1(Chi^2: 0.0395):



C sTRSV(Chi^2: 0.33):



D

RNA Device	Dissociation Rate (*k*, min^{-1})		
	No $MgCl_2$	**$MgCl_2$**	**$MgCl_2$ + theo**
L2b8	0.047± 0.003	0.133 ± 0.011	0.087 ± 0.008
L2b8-a1	0.052 ± 0.001	0.552 ± 0.006	0.166 ± 0.001
sTRSV ribozyme	0.18 ± 0.01	1.66 ± 0.04	1.54 ± 0.10

Supplementary Figure S3.5 Representative SPR-based cleavage assays for measuring RNA dissociation rate constants (*kd*) for RNA generated from the cis-blocking strategy. A representative Biacore sensorgram is shown for each cis-blocked construct: (A) L2b8, (B) L2b8-a1, (C) sTRSV. The assays were performed in reactions buffers (100 mM NaCl, 50 mM Tris-HCl (pH 7.0)) containing: (i) no $MgCl_2$; (ii) 500 μM $MgCl_2$; and (iii) 500 μM $MgCl_2$ and 5 mM theophylline at 37ºC. The capture, reaction, and regeneration steps in the assay cycle are indicated for the L2b8 sensorgrams. Methods used to prepare full-length RNA transcripts and conditions of the cleavage assays are detailed in the Materials and Methods section of the main text. The reaction step of the sensorgram was well-fit to the single exponential equation (Chi^2 < 0.34):

$R_t = R_0 \times (e^{-kdt} - 1) + \text{offset}$

(D) The RNA dissociation rate (*kd*) was determined for each assay. The reported *kd* for each construct are the mean and standard deviation of at least three independent experiments.

Supplementary Table S3.1 Sequences of primers and templates for all described cis-blocked constructs in this work. The T7 promoter region in the template sequence is indicated in bold.

Primer	**Sequence**
T7-fwd	TTCTAATACGACTCACTATAGG
sTRSV-rev	CAAAGCTGTTTCGTCCTCAC
sTRSV-2-rev	AACAAAGCTGTTTCGTCCTCAC
sTRSV-4-rev	CAAACAAAGCTGTTTCGTCCTCAC
sTRSV-6-rev	AACAAACAAAGCTGTTTCGTCCTCAC
sTRSV-8-rev	GAAACAAACAAAGCTGTTTCGTCCTCAC
sTRSV-10-rev	GGGAAACAAACAAAGCTGTTTCGTCCTCAC
L2b1-rev	CAAAGCTGTTTCGTCCTCGTCCCGTCC
L2b5-rev	Same as b-L2b1-rev
L2b8-rev	CAAAGCTGTTTCGTCCTCCGT
L2b8-a1-rev	Same as b-L2b8-rev
L2bOFF1-rev	CAAAGCTGTTTCGTCCTCGTCCACTGC
L2b8tc-rev	Same as b-L2b8-rev
Template	**Sequence**
sTRSV	**TTCTAATACGACTCACTATAGGG**AAACAAACAAAGTTGTT TTGCTGTCACCGGATGTGCTTTCCGGTCTGATGAGTCCGTGA GGACGAAACAGCTTTG
sTRSV-2	**TTCTAATACGACTCACTATAGGG**AAACAAACAAAGTTGTT TTGCTGTCACCGGATGTGCTTTCCGGTCTGATGAGTCCGTGA GGACGAAACAGCTTTGTT
sTRSV-4	**TTCTAATACGACTCACTATAGGG**AAACAAACAAAGTTGTT TTGCTGTCACCGGATGTGCTTTCCGGTCTGATGAGTCCGTGA GGACGAAACAGCTTTGTTTG
sTRSV-6	**TTCTAATACGACTCACTATAGGG**AAACAAACAAAGTTGTT TTGCTGTCACCGGATGTGCTTTCCGGTCTGATGAGTCCGTGA GGACGAAACAGCTTTGTTTGTT
sTRSV-8	**TTCTAATACGACTCACTATAGGG**AAACAAACAAAGTTGTT TTGCTGTCACCGGATGTGCTTTCCGGTCTGATGAGTCCGTGA GGACGAAACAGCTTTGTTTGTTTC
sTRSV-10	**TTCTAATACGACTCACTATAGGG**AAACAAACAAAGTTGTT TTGCTGTCACCGGATGTGCTTTCCGGTCTGATGAGTCCGTGA GGACGAAACAGCTTTGTTTGTTTCCC
L2b1	**TTCTAATACGACTCACTATAGGG**AAACAAACAAAGTTGTT TTGCTGTCACCGGATGTGCTTTCCGGTCTGATGAGTCCGTGT CCATACCAGCATCGTCTTGATGCCCTTGGCAGGGACGGGAC GAGGACGAAACAGCTTTG
L2b5	**TTCTAATACGACTCACTATAGGG**AAACAAACAAAGTTGTT

	TTGCTGTCACCGGATGTGCTTTCCGGTCTGATGAGTCCGTGT CCAATACCAGCATCGTCTTGATGCCCTTGGCAGTGGACGGG ACGAGGACGAAACAGCTTTG
L2b8	**TTCTAATACGACTCACTATAGGG**AAACAAACAAAGTTGTT TTGCTGTCACCGGATGTGCTTTCCGGTCTGATGAGTCCGTTG TCCATACCAGCATCGTCTTGATGCCCTTGGCAGGGACGGGA CGGAGGACGAAACAGCTTTG
L2b8-a1	**TTCTAATACGACTCACTATAGGG**AAACAAACAAAGTTGTT TTGCTGTCACCGGAATCAAGGTCCGGTCTGATGAGTCCGTT GTCCATACCAGCATCGTCTTGATGCCCTTGGCAGGGACGGG ACGGAGGACGAAACAGCTTTG
L2bOFF1	**TTCTAATACGACTCACTATAGGG**AAACAAACAAAGTTGTT TTGCTGTCACCGGATGTGCTTTCCGGTCTGATGAGTCCGTGT TGCTGATACCAGCATCGTCTTGATGCCCTTGGCAGCAGTGG ACGAGGACGAAACAGCTTTG
L2b8tc	**TTCTAATACGACTCACTATAGGG**AAACAAACAAAGTTGTT TTGCTGTCACCGGATGTGCTTTCCGGTCTGATGAGTCCGTTG TCCAAAACATACCAGATTTCGATCTGGAGAGGTGAAGAATT CGACCACCTGGACGGGACGGAGGACGAAACAGCTTTG

Acknowledgements

We thank J Schuman of GE and M Eckhart of the Stanford PAN facility. This work was supported by the National Institutes of Health (R01GM086663), the National Science Foundation (CBET-0917638, CCF-0943269), the Defense Advanced Research Projects Agency (HR0011-11-2-0002), and the National Sciences and Engineering Research Council of Canada (fellowship to ABK).

Contributions

JCL designed research, performed research, and wrote the paper; ABK designed and performed research associated with the *in vitro* characterization of ribozyme cleavage rates and wrote the paper; CDS designed research and wrote the paper.

References

1. Nandagopal, N. and Elowitz, M.B. (2011) Synthetic biology: integrated gene circuits. *Science*, **333**, 1244-1248.
2. Liang, J.C., Bloom, R.J. and Smolke, C.D. (2011) Engineering biological systems with synthetic RNA molecules. *Mol Cell*, **43**, 915-926.
3. Win, M.N. and Smolke, C.D. (2007) A modular and extensible RNA-based gene-regulatory platform for engineering cellular function. *Proceedings of the National Academy of Sciences*, **104**, 14283-14288.
4. Chen, Y.Y., Jensen, M.C. and Smolke, C.D. (2010) Genetic control of mammalian T-cell proliferation with synthetic RNA regulatory systems. *Proc Natl Acad Sci U S A*, **107**, 8531-8536.
5. Wieland, M. and Hartig, J.S. (2008) Improved aptazyme design and in vivo screening enable riboswitching in bacteria. *Angew Chem Int Ed Engl*, **47**, 2604-2607.
6. Win, M.N. and Smolke, C.D. (2008) Higher-order cellular information processing with synthetic RNA devices. *Science*, **322**, 456-460.
7. Beisel, C.L. and Smolke, C.D. (2009) Design principles for riboswitch function. *PLoS Comput Biol*, **5**, e1000363.
8. Darty, K., Denise, A. and Ponty, Y. (2009) VARNA: Interactive drawing and editing of the RNA secondary structure. *Bioinformatics*, **25**, 1974-1975.
9. Nelson, J.A., Shepotinovskaya, I. and Uhlenbeck, O.C. (2005) Hammerheads derived from sTRSV show enhanced cleavage and ligation rate constants. *Biochemistry*, **44**, 14577-14585.

10. Mercure, S., Lafontaine, D., Ananvoranich, S. and Perreault, J.P. (1998) Kinetic analysis of delta ribozyme cleavage. *Biochemistry*, **37**, 16975-16982.

11. Stage-Zimmermann, T.K. and Uhlenbeck, O.C. (1998) Hammerhead ribozyme kinetics. *RNA*, **4**, 875-889.

12. Khvorova, A., Lescoute, A., Westhof, E. and Jayasena, S.D. (2003) Sequence elements outside the hammerhead ribozyme catalytic core enable intracellular activity. *Nat Struct Biol*, **10**, 708-712.

13. Singh, K.K., Parwaresch, R. and Krupp, G. (1999) Rapid kinetic characterization of hammerhead ribozymes by real-time monitoring of fluorescence resonance energy transfer (FRET). *RNA*, **5**, 1348-1356.

14. Penedo, J.C., Wilson, T.J., Jayasena, S.D., Khvorova, A. and Lilley, D.M. (2004) Folding of the natural hammerhead ribozyme is enhanced by interaction of auxiliary elements. *RNA*, **10**, 880-888.

15. Rich, R.L., Papalia, G.A., Flynn, P.J., Furneisen, J., Quinn, J., Klein, J.S., Katsamba, P.S., Waddell, M.B., Scott, M., Thompson, J. *et al.* (2009) A global benchmark study using affinity-based biosensors. *Anal Biochem*, **386**, 194-216.

16. Mathews, D.H., Disney, M.D., Childs, J.L., Schroeder, S.J., Zuker, M. and Turner, D.H. (2004) Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure. *Proc Natl Acad Sci U S A*, **101**, 7287-7292.

17. Zhang, D.Y. and Seelig, G. (2011) Dynamic DNA nanotechnology using strand-displacement reactions. *Nature Chemistry*, **3**, 103-113.

18. Curtis, E.A. and Bartel, D.P. (2001) The hammerhead cleavage reaction in monovalent cations. *RNA*, **7**, 546-552.

19. Chadalavada, D.M., Knudsen, S.M., Nakano, S. and Bevilacqua, P.C. (2000) A role for upstream RNA structure in facilitating the catalytic fold of the genomic hepatitis delta virus ribozyme. *J Mol Biol*, **301**, 349-367.

20. Jensen, K.K., Orum, H., Nielsen, P.E. and Norden, B. (1997) Kinetics for hybridization of peptide nucleic acids (PNA) with DNA and RNA studied with the BIAcore technique. *Biochemistry*, **36**, 5072-5077.

21. Navratilova, I., Eisenstien, E. and Myszka, D.G. (2005) Measuring long association phases using Biacore. *Anal Biochem*, **344**, 295-297.

22. Myszka, D.G. (2000) Kinetic, equilibrium, and thermodynamic analysis of macromolecular interactions with BIACORE. *Methods Enzymol*, **323**, 325-340.

23. Myszka, D.G. (1999) Improving biosensor analysis. *J Mol Recognit*, **12**, 279-284.

Chapter 4.

A high-throughput *in vitro* ribozyme-based selection strategy for generating new RNA sensing functions*

Abstract

In vitro selection strategies have been developed to support *de novo* generation of RNA aptamers, allowing the potential to generate sensors to diverse ligands of interest. The scalability of these strategies to small molecule ligands is primarily limited by the partitioning method, which utilizes column-based affinity chromatography and therefore requires chemical modification of the target ligands. We propose a solution-based selection strategy that enables generation of new RNA sensing functions directly within a modular ribozyme device platform. In the proposed strategy, a large device library is generated by randomizing the sensor component within the device platform, and the resulting library is partitioned through ligand-dependent cleavage activities. We addressed challenges associated with the efficient generation of full-length ribozyme device libraries through *in vitro* transcription reactions by employing a cis-blocking strategy. We further developed a dual selection process based on magnetic bead-based partitioning, which supports rapid gel-free recovery of the desired library members. The dual selection process was optimized by modeling the effects of incubation times on the fold enrichment of an active device. Under optimal incubation times for positive and negative selections, our model predicted a ~13-fold enrichment per selection round. With further optimization and development, our ribozyme-based selection strategy may address current challenges in scaling the generation of synthetic RNA control devices exhibiting new sensing functions to small molecule ligands for broad cellular engineering applications.

4.1 Introduction

Recent discovery of the widespread regulatory activity and conservation of noncoding RNAs represents an exciting area in RNA-based gene regulation (1). Significant research efforts have focused on the design of RNA-based sensing-actuation systems that link molecular binding information to programmed gene regulation (2). A unique advantage of RNA-based sensing systems is the potential to tailor the sensing function to detect a diverse array of small molecules through the *de novo* generation of RNA aptamers (3,4). *In vitro* selection strategies have been applied to generate RNA aptamers to many small molecule ligands, including alkaloids (5), amino acids (6), and oligosaccharides (7,8). Partitioning methods based on column affinity chromatography are generally used in selecting aptamers to small molecule ligands, where the target small molecule is conjugated to a solid support through an appropriate functional group.

Small molecule aptamers are generated by applying iterative rounds of screening for sequences exhibiting desired binding affinities from very large libraries *in vitro*. The aptamer selection process begins with incubation of an initial RNA library (~10^{14}-10^{15} molecules) in a column modified with the target small molecule. Unbound or weak-binding sequences are removed from the column by washing with the selection buffer. The washing step can be repeated to increase the selection stringency and therefore isolate rare sequences that exhibit high affinities to the target ligand. The bound sequences are recovered by a competitive elution process in which high concentration of free ligand is applied to the column to compete with the immobilized ligand for the binding of RNA. The recovered RNA is amplified through reverse transcription (RT) and polymerase chain reaction (PCR), and used as the input library for the next selection

round. In general, the selection process is performed iteratively 10-15 times to reduce library diversity sufficiently for characterization. Aptamers with affinities ranging from 10 nM to 100 μM have been generated through the described selection process (9). By incorporating appropriate counterselection cycles in the later rounds of the selection process, aptamers can be selected to discriminate between structurally similar metabolites. In one example, an aptamer selected to theophylline exhibits a 10,000 fold lower affinity for caffeine, which differs from theophylline by a single methyl group (5).

One key challenge to the broad generation of RNA aptamers capable of detecting diverse small molecule ligands is the limited scalability of current *in vitro* selection methods. In particular, current strategies require that the small molecule ligands have functional groups exhibiting chemistries that are appropriate for conjugation to solid supports. In many cases such functional groups are not readily available, necessitating the development of chemical modification strategies to append appropriate functional groups to the desired ligands. The need for coupling strategies have added disadvantages associated with (i) longer, more cumbersome, less efficient partitioning processes based on column chromatography; (ii) presentation of the ligand in an environment that does not closely mimic that of the cellular environment (i.e., ligand in solution versus ligand coupled to solid support); and (iii) removal of functional groups through chemical coupling that could play important roles in interacting with the RNA sequence to increase binding affinities. Improved solution-based partitioning methods, such as capillary electrophoresis, have been developed for the selection of aptamers to protein ligands (10). However, these methods are often based on the large size of the protein ligand and are not applicable to small molecule ligands. Therefore, the development of rapid, solution-

based partitioning methods for small molecule ligands is critical to scaling aptamer selection to diverse molecular targets.

An *in vitro* solution-based selection strategy to facilitate the isolation of allosteric ribozymes that exhibit new ligand-activated ribozyme cleavage activities from large nucleic acid libraries was previously described (11) (Figure 4.1). The selection strategy utilized an allosteric ribozyme comprised of a minimal hammerhead ribozyme (HHRz) and a randomized sensor domain, coupled through a previously selected communication module (12). An initial DNA library was generated by coupling a randomized 25-mer sensor domain (N25) to the 5' end of the minimal HHRz through the communication module. The DNA library was transcribed and purified by denaturing polyacrylamide gel electrophoresis (PAGE). The full-length allosteric ribozyme library was recovered from the gel and subjected to two sequential selections through which library members were partitioned based on ligand-activated ribozyme cleavage activity. An initial negative selection round was performed by incubating the library in the absence of ligand. Library members that resisted cleavage in the absence of ligand were recovered by PAGE purification. A subsequent positive selection round was performed by incubating the recovered full-length library with the ligand of interest. Library members that cleaved in the presence of ligand, or the 5' cleaved library fragments, were recovered by PAGE purification and amplified through RT-PCR to generate the library pool for the next selection cycle.

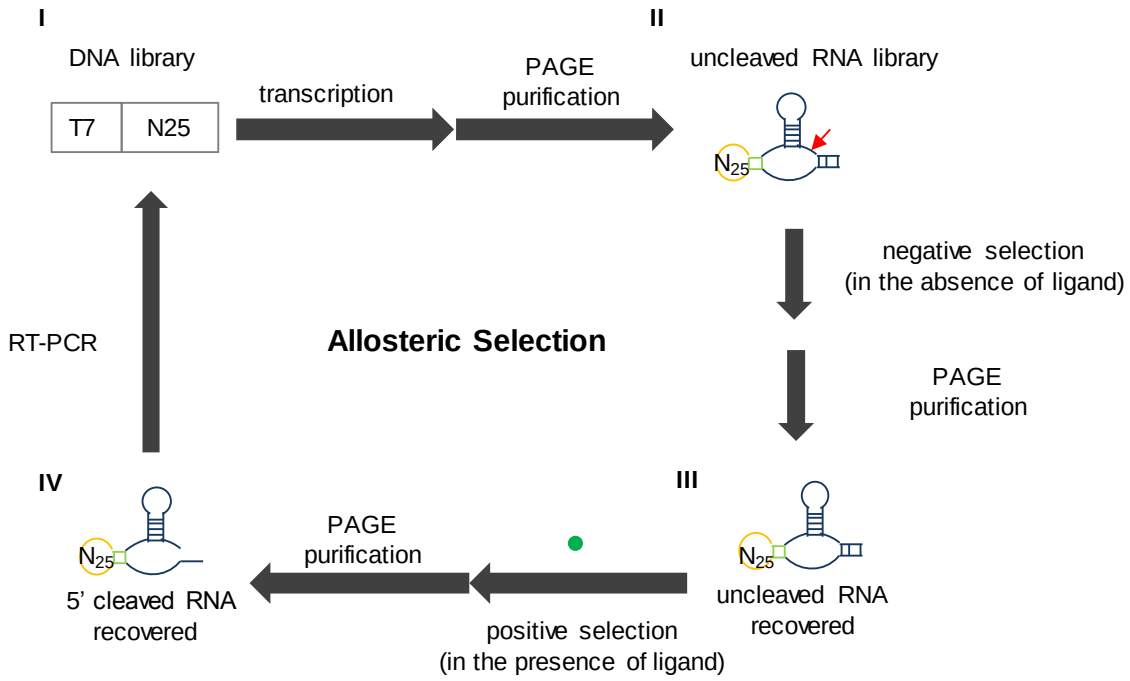


Figure 4.1 Schematic of an *in vitro* allosteric ribozyme selection strategy (11). The RNA library is generated from the DNA library by transcription reactions (I). The full-length RNA is recovered by PAGE and is (II) subject to a negative selection in the absence of ligand. The red arrow indicates the cleavage site. Uncleaved library members are recovered by PAGE and subject to a positive selection in the presence of ligand (III). The 5' cleaved library members are recovered by PAGE and (IV) amplified through RT/PCR to generate the double-stranded DNA templates for the next selection cycle. Each round of selection requires three PAGE purifications, thus making the allosteric selection strategy laborious and time-consuming.

The described allosteric selection strategy was applied to generate allosteric ribozymes that respond to secondary messengers cGMP and cAMP through a total of 22 selection rounds, resulting in allosteric ribozymes with binding affinities ranging from ~200 μM to 4 mM (11,13). However, these ribozymes are based on a truncated form of HHRz, which requires high $MgCl_2$ concentration (>10 mM) to exhibit activity and thus do not exhibit activity in cellular environments, where $MgCl_2$ is in the submillimolar

range (14).

We recently described a modular ligand-responsive ribozyme device platform that specifies physical linkage of three RNA components (15) (Figure 4.2). A sensor encoded by an RNA aptamer is linked to a HHRz from the satellite RNA of tobacco ringspot virus (sTRSV) through a transmitter sequence capable of facilitating a strand-displacement event. The device platform specifies the integration of the sensor-transmitter element through the ribozyme loops, which have been shown to provide a stabilizing tertiary interaction essential for activity under physiological concentrations of $MgCl_2$ (14). We have demonstrated the gene-regulatory function of the ribozyme devices in yeast (15) and mammalian cells (16) and verified their ligand-dependent cleavage activities through *in vitro* characterization (see Chapter 2). Therefore, the potential exists to adapt the described allosteric selection strategy to our platform to generate new ligand-responsive ribozyme devices that are amenable to cellular engineering applications.

The efficiency and throughput of the described selection strategy based on differential ribozyme cleavage activities are currently limited by the inefficient generation of uncleaved RNA through *in vitro* transcription reactions and the low-throughput gel separation method to recover desired RNA fragments. Large RNA libraries are generated by *in vitro* T7 transcription, which requires high $MgCl_2$ concentrations (~mM) to achieve sufficient yields. Because the described allosteric ribozyme selection process is based on the truncated form of the HHRz, which only exhibits activity at high $MgCl_2$ concentrations (>10 mM), sufficient yields of uncleaved RNA can be generated by transcription reactions to cover the desired sequence space. However, libraries based on our ribozyme device platform, which maintains the

stabilizing loop-loop interactions, can exhibit cleavage activity at submillimolar concentrations of $MgCl_2$, such that library members that can adopt a ribozyme-active conformation will cleave during transcription, resulting in low yields of uncleaved RNA. Further, the cleavage-based partitioning strategy requires isolation of desired RNA fragments of different lengths. PAGE purification, a low-throughput and cumbersome process, is typically performed to recover the desired RNA fragments, which increases both the time and labor required for each selection round. Therefore, more efficient ribozyme generation and high-throughput separation strategies are desired.

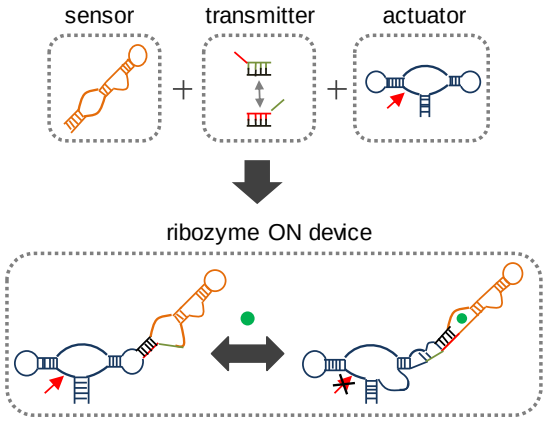


Figure 4.2 A modular ribozyme device platform allowing construction of devices that are amenable to cellular engineering applications. A ribozyme device is comprised of three RNA components: sensor (aptamer), actuator (ribozyme), and transmitter (sequence capable of a strand-displacement event). The device platform specifies the integration of the sensor-transmitter element through the ribozyme loops, which have been shown to provide a stabilizing tertiary interaction essential for activity under physiological concentrations of $MgCl_2$. A ribozyme ON device is depicted here. In the absence of input ligand, the device adopts a ribozyme-active conformation, thereby leading to cleavage or decreased gene expression. In the presence of input ligand, the device adopts a ribozyme-inactive conformation, which is stabilized by the ligand binding to the aptamer, thereby leading to no cleavage or increased gene expression.

We aim to develop a scalable *in vitro* selection strategy based on the ribozyme

device platform to generate new sensing functions. The selection strategy will be applied to large device libraries (10^{14}-10^{15} molecules) generated by randomizing the sensor component (Nx, where x indicates the number of randomized nucleotides) within the ribozyme device platform. A high-throughput and high-efficiency cleavage-based partitioning strategy will be developed to facilitate isolation of new ligand-responsive ribozyme devices. We focused initially on addressing limitations associated with the generation of uncleaved RNA device libraries through *in vitro* transcription reactions by employing an efficient cis-blocking strategy (see Chapter 3). A dual selection strategy based on magnetic beads was proposed to facilitate cleavage-based library partitioning and rapid isolation of desired RNA fragments. We optimized the fold enrichment of the dual selection process by developing a mathematical model based on a simple control device library composed of three populations that exhibit distinct cleavage activities. We examined the effects of incubation times on the fold enrichment of the desired population during the negative and positive selection steps. Our model predicted a ~13-fold enrichment for each round of selection under optimal incubation conditions. Future work will focus on the validation of our strategy on a control library (~10^6 variants) that contains one known functional sequence.

4.2 Results

4.2.1 The cis-blocking strategy allows efficient generation of a full-length ribozyme device library by *in vitro* transcription

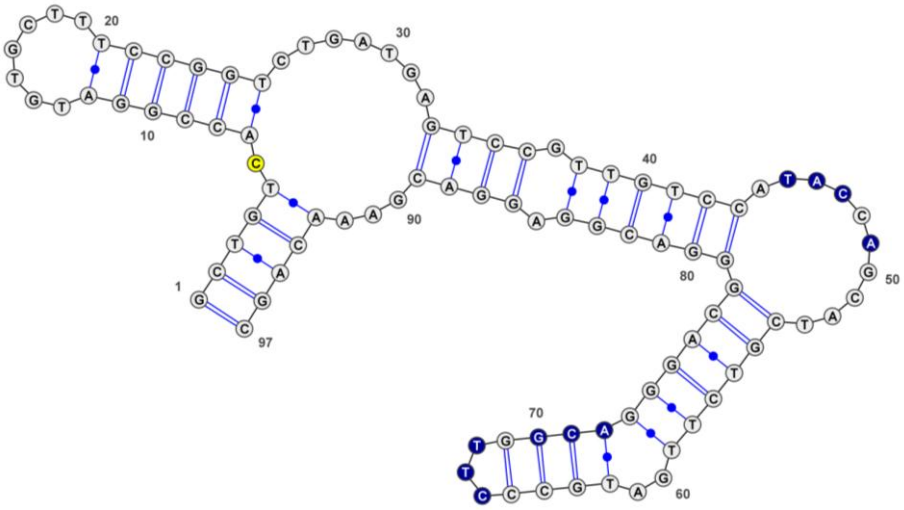
One key challenge in developing a scalable ribozyme-based selection strategy is the efficient generation of the full-length RNA library due to significant ribozyme

cleavage activity at the transcription reaction conditions. Our previous work demonstrated a cis-blocking strategy for the efficient generation of full-length RNA containing a ribozyme, including natural ribozymes and ribozyme devices (see Chapter 3). Specifically, an RNA blocking sequence was designed to inhibit the cleavage of a variety of ribozyme devices composed of different components during *in vitro* transcription, resulting in a blocking efficiency of up to ~90%. In addition, a RNA activation sequence was designed to allow triggered release of the blocked RNA in the presence of a DNA activator strand. We examined whether a similar strategy could be applied to the generation of full-length ribozyme device libraries.

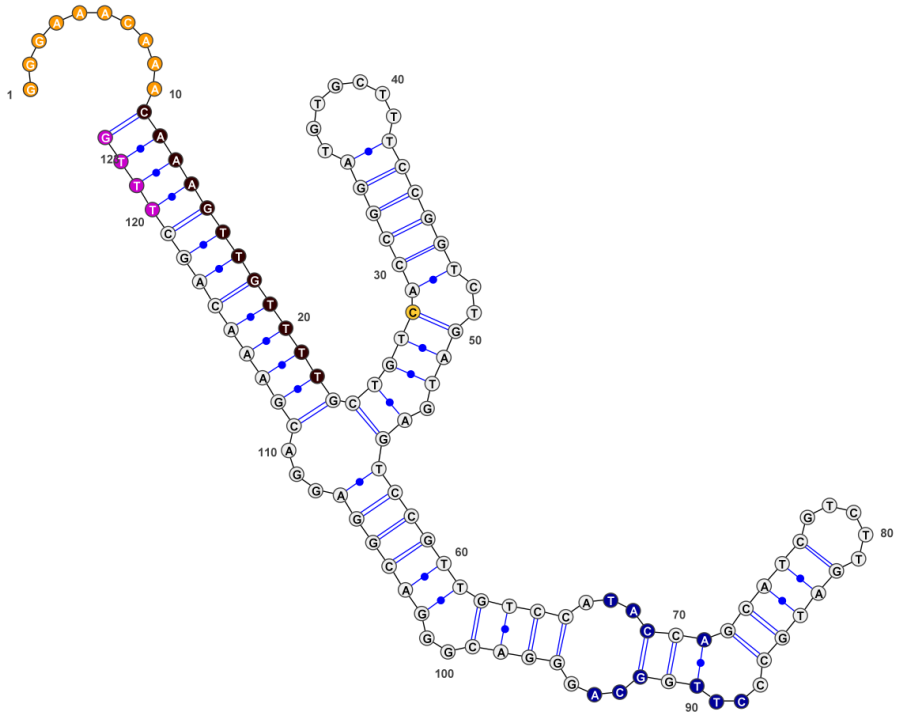
We applied the blocking, activation, and stabilizing sequences previously developed for the ribozyme-based devices to a small control device library (~10^6 variants), L2b8-N10. The L2b8-N10 library was generated by randomizing 10 nt positions in the aptamer binding core (sensor component) of a theophylline-responsive ribozyme device, L2b8 (Figure 4.3A and B). A radiolabeled library was transcribed through a standard T7 *in vitro* reaction with the cis-blocking strategy and the transcription products were analyzed by PAGE to quantify the blocking efficiency (Figure 4.3C). More than 50% of the transcription products were uncleaved, indicating that the cis-blocking sequence inhibited a majority of the library members from adopting a ribozyme-active conformation during transcription. We subsequently incubated the transcribed cis-blocked L2b8-N10 RNA with the DNA activator strand and 5 mM $MgCl_2$ for 15 minutes and verified that the activator was capable of restoring the ribozyme activity such that more than 60% of the full-length blocked RNA cleaved (Figure 4.3C). These results support that the cis-blocking strategy provides a gel-free strategy for

generating large yields of full-length RNA from a ribozyme device library.

A Functional conformation



B Blocked conformation



C

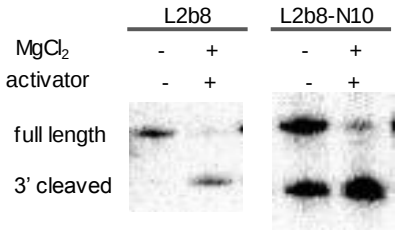


Figure 4.3 The cis-blocking strategy enables efficient generation of full-length ribozyme device library. The predicted secondary structures of the L2b8 device are shown here. (A) The L2b8 ribozyme-active device functional conformation is shown here. The randomized positions in the theophylline aptamer of the L2b8 device (to generate the L2b8-N10 device library) are indicated in blue. The cleavage site is indicated in yellow. (B) The blocked conformation is shown here, where the blocking, activation, and stabilizing sequences are indicated in orange, brown, and magenta, respectively. Secondary structures were predicted by RNAstructure folding software (17) and rendered using VARNA software (18). (B) The transcribed L2b8 and L2b8-N10 products with the cis-acting blocking sequence incorporated in the 5' end of the transcript were characterized by PAGE analysis. More than 50% of the transcribed L2b8-N10 products, compared to ~90% of the transcribed L2b8 products, were uncleaved. When incubating with 5 mM $MgCl_2$ and 10 μM DNA activator, ~60% of the full-length L2b8-N10 products, compared to ~90% of the full-length L2b8 products, were capable of self-cleavage.

4.2.2 Design of a dual selection strategy based on magnetic bead separation

We designed a dual selection strategy that allows partitioning of library members through sequential cleavage reactions (Figure 4.4). A reaction surface is generated by immobilizing biotinylated DNA activator strands to the streptavidin-coated magnetic beads. The transcribed cis-blocked RNA libraries are incubated with the beads to allow activation of the full-length RNA through hybridization interactions between the RNA and the activator strand. The binding capacity of commercially available magnetic beads (Promega) is ~1 nmol/ml, making them scalable for library selection. For example, generation of new sensing functions typically requires searching within a large library of

~10^{14}-10^{15} sequences, corresponding to ~0.2-2 nmol of molecules. Therefore, at least 2 ml of the magnetic beads need to be used to immobilize a sufficient amount of the DNA activator strands (and thus allow sufficient hybridization with RNA device libraries).

The streptavidin-coated beads respond to a magnetic field, allowing the beads to be rapidly partitioned to the surface of a microfuge tube. Thus, RNA bound to the beads can be rapidly separated from unbound species in the supernatant, such as 3' cleaved fragments, which can be removed by aspiration or pipetting. The bound full-length library is partitioned through sequential cleavage reactions. A positive selection is performed first by incubating the library in the presence of ligand and $MgCl_2$. Active library members that resist cleavage in the presence of ligand will remain bound to the DNA activator strand, whereas cleaved library members are weakly bound to the activator strand and thus rapidly dissociate from the bead surface. The beads are collected to the tube surface by the magnetic field, allowing the cleaved library members in the supernatant to be removed by aspiration or pipetting. The remaining bound full-length library is next subjected to a negative selection by incubating in the absence of ligand and $MgCl_2$. The dissociated cleaved library members are recovered from the supernatant and subsequently amplified through RT-PCR to generate the input library for the next selection round. In the proposed scheme, no gel purification step is required to facilitate the separation of RNA fragments.

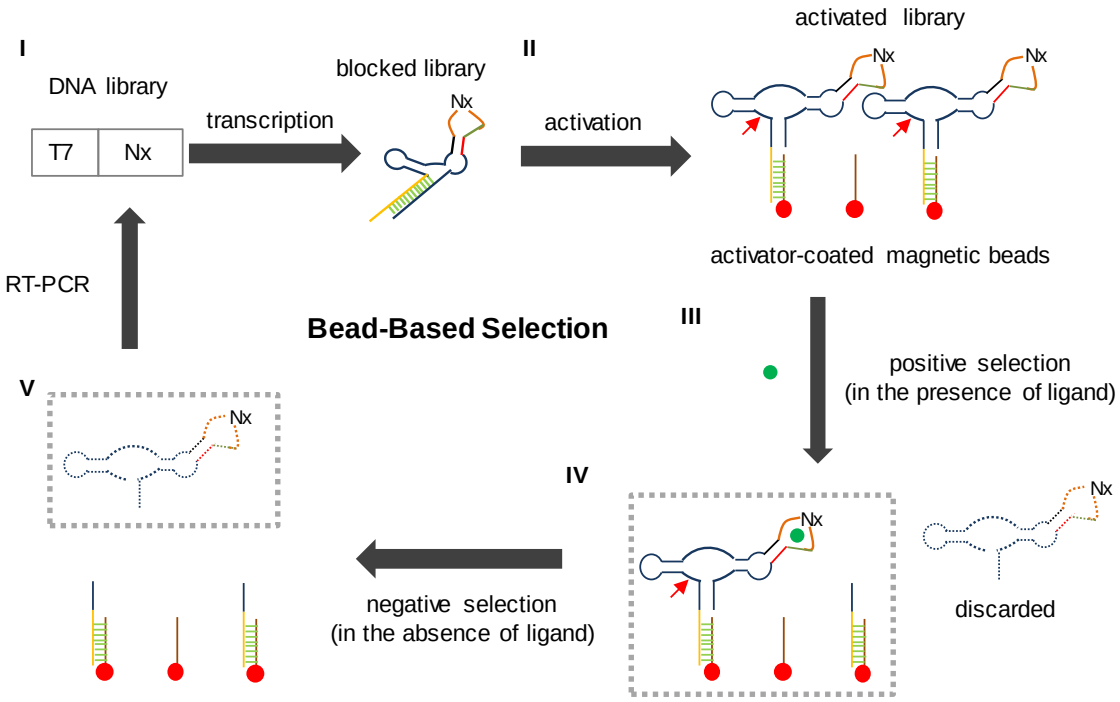


Figure 4.4 Schematic of the dual selection strategy facilitated by magnetic bead-based separation. The cis-blocked RNA device library is transcribed from DNA template (I). The blocked RNA is activated by incubating with the DNA activator strands immobilized to magnetic beads (II). The red arrow indicates the cleavage site. The activated library is first subject to a positive selection (III) by incubating with the target ligand and $MgCl_2$. The library members that resist cleavage in the presence of ligand (boxed) will remain bound to the DNA activator strand, whereas cleaved library members are discarded. The remaining library is subject to a negative selection (IV) by incubating with $MgCl_2$. The cleaved library members (boxed) are recovered to be amplified through RT-PCR (V) to generate the DNA template for the next round of selection.

4.2.3 Optimization of reaction parameters to achieve greater enrichment efficiency

Our ribozyme device platform utilizes a linker sequence capable of a strand-displacement event to facilitate information transmission between the sensor and actuator components. The interaction between the linker sequence and neighboring components dictates the partitioning between functional conformations of a device and thus the associated cleavage rate. When generating a device library by randomizing the sensor component, it is expected that each new sensor sequence will interact with the existing

linker sequence differently, such that the library will exhibit a distribution of cleavage activities.

We have previously characterized three theophylline-responsive ribozyme devices, L2b1, L2b5, and L2b8, that span a wide range of *in vitro* cleavage rates in the absence (0.033, 0.013, and 0.14 min^{-1}) and presence of 5 mM theophylline (0.011, 0.007, and 0.025 min^{-1}) at physiologically-relevant reaction conditions (500 μM $MgCl_2$, 100 mM NaCl, and 50 mM Tris-HCl (pH 7.0) at 37°C) (see Chapter 2). The L2b5 and L2b8 devices were characterized to be the slowest and fastest cleaving devices among the series based on the same ribozyme component, and thus these three devices were selected to represent the broad activities exhibited by a device library.

To model the enrichment patterns of a typical device library, we assumed an initial library composed of three equally represented slow, medium, and fast-cleaving populations that exhibit similar cleavage activities as the L2b5, L2b1, and L2b8 devices, respectively, and do not respond to the ligand. We also assumed that a ligand-responsive target sequence (i.e., theophylline-responsive device) exists in the library and exhibits similar cleavage activities in the absence (0.14 min^{-1}) and presence (0.025 min^{-1}) of 5 mM theophylline as the L2b8 device. The fraction of the cleaved population as a function of time can be modeled by a first-order exponential decay equation: e^{-kt}, where k is the cleavage rate constant. For example, the fast-cleaving population, which exhibits the same cleavage rate (0.14 min^{-1}) as the L2b8 device, will require 15 minutes to achieve ~90% cleavage under the specified conditions, whereas the slow-cleaving population, which exhibits the same cleavage rate (0.013 min^{-1}) as the L2b5 device, will require 180 minutes to achieve the same extent of cleavage. Incubation times that are either too short

or too long will present practical limitations on the selection process, such as ease of liquid sample handling (e.g., short incubation time) and stability of the RNA-DNA activator hybrid and magnetic bead surface (e.g., long incubation time). Therefore, careful optimization of the cleavage reaction conditions is needed for both negative and positive selections.

Optimization of the negative selection conditions

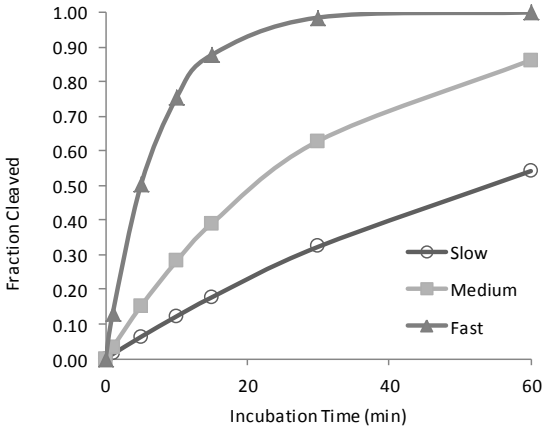
The negative selection step in the dual selection process enriches sequences that are capable of cleaving in the absence of ligand, and the cleaved sequences are recovered to be amplified through RT-PCR. We performed calculations based on a device library of 10^{15} diversity (one functional sequence in 10^{15} variants), where the transcription reaction was set up to generate 10 copies of each sequence in the library (a total of 10^{16} members). As such, the initial library is composed of 3.3×10^{15} members in each population. We examined fractions of cleaved sequences within the three populations as a function of incubation time using our computational model (Figure 4.5A). Predicted fractions of cleaved sequences from each population were calculated based on rates obtained at the physiologically-relevant reaction conditions in the absence of theophylline (Supplementary Table S4.1).

As the target device is part of the fast-cleaving population, we determined the respective fold enrichment of the fast-cleaving population relative to the medium and slow-cleaving populations for the negative selection step (Figure 4.5B). The fold enrichment is determined by calculating the ratios of the number of cleaved fast-cleaving members to that of the medium and slow-cleaving members, respectively. For example,

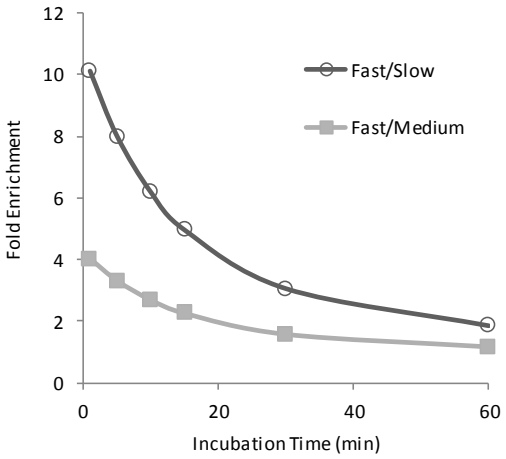
the fractions of cleaved slow, medium, and fast-cleaving populations after incubating the library for 1 minute are 0.01, 0.03, and 0.13, respectively (Supplementary Table S4.1). These fractions correspond to 4.31×10^{13}, 1.08×10^{14}, and 4.35×10^{14} cleaved members at the end of incubation, resulting in 10 and 4-fold enrichment of the fast-cleaving population relative to the slow and medium-cleaving populations, respectively.

We further determined the percent of input library members recovered for the subsequent RT-PCR step (Figure 4.5C). The total number of recovered library members is calculated by summing up the numbers of cleaved members for each population. The percent recovery can then be determined by dividing the number of recovered library members to that of the input library (Supplementary Table S4.1). For example, after 1-minute incubation the sum of the numbers of all cleaved members is 5.87×10^{14}, corresponding to only ~6% of the input initial library (10^{16} members). This modeling result suggested that when there is a large diversity in the library (i.e., in the earlier selection rounds), the functional sequences are rare and therefore require proper scaling of the transcription reactions to ensure that library diversity is not lost in the process. Therefore, a longer incubation time (e.g., 5-10 minutes) is desired for the earlier selection rounds. For the later selection rounds when the functional sequences have been enriched, a shorter incubation time (<1 minute) can be employed to achieve greater fold enrichment for the fast-cleaving population.

A



B



C

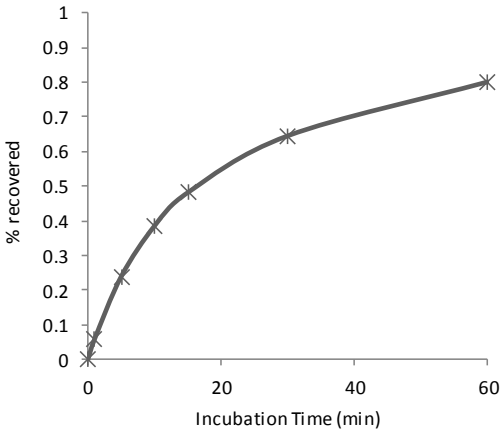


Figure 4.5 Optimization of the negative selection conditions. The model is based on an initial device library composed of 3.3×10^{15} members in each slow, medium, and fast-cleaving population (a total of 10^{16} variants) (Supplementary Table S4.1). (A) Fraction of cleaved members within each population as a function of incubation time. The fraction of

the cleaved population is determined by the first-order exponential decay equation: e^{-kt}, where *k* is the cleavage rate constant. The fractions of the cleaved populations increase at different rates as a function of time. (B) Fold enrichment of the fast-cleaving population relative to the medium and slow-cleaving populations as a function of time. The fold enrichment is determined by calculating the ratios of the number of cleaved fast-cleaving members to that of the medium and slow-cleaving members, respectively. (C) Percent recovered library as a function of incubation time. The percent of the recovered cleaved library members is determined by summing up the numbers of cleaved members from the three populations and dividing the sum by the number of the input library members.

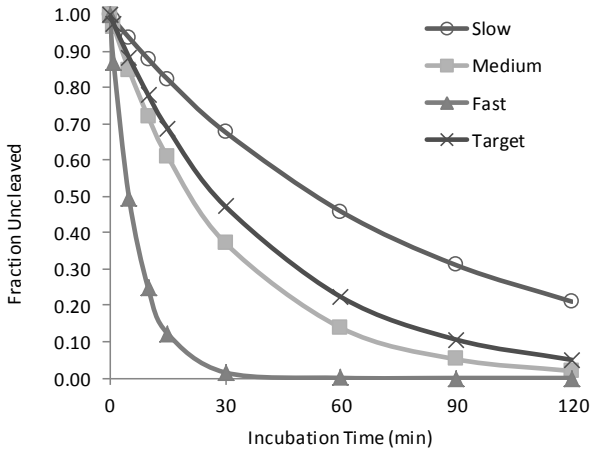
Optimization of the positive selection conditions

The positive selection step in the dual selection process enriches sequences that resist cleaving in the presence of ligand, and the cleaved sequences are discarded. Fractions of uncleaved sequences within the three populations during the positive selection were determined by the first-order exponential decay equation ($1-e^{-kt}$) based on the same rates used for the modeling of negative selection, as these populations do not response to the ligand (theophylline). The fraction of cleaved target sequence was calculated based on the rate for the L2b8 device in the presence of 5 mM theophylline (Figure 4.6A and Supplementary Table S4.2).

The fold enrichment of the target sequence is determined by the ratio of fraction compositions of the uncleaved target sequence in the library before and after the incubation (Figure 4.6B and Supplementary Table S4.2). As an example, the number of the uncleaved target sequence in the initial library of 10^{16} members is 10, corresponding to a fraction composition of 10^{-15}. An incubation of 15 minutes results in the cleavage of ~30% of the target sequences, thus decreasing the number of target sequence from 10 to 7. The fraction composition of the target sequence after the selection becomes 1.33×10^{-15}, corresponding to a 1.3-fold enrichment. The modeling results predicted optimal fold enrichment of 1.3-fold for an incubation time ranged from 15 to 30 minutes, and a longer

incubation time (>30 minutes) has a negative impact on the fold enrichment. Although fold enrichment of the positive selection is modest, when coupled with the negative selection the overall fold enrichment of the target sequence is increased to ~5-fold after one round of dual selection on the initial device library (Supplementary Table S4.3).

A



B

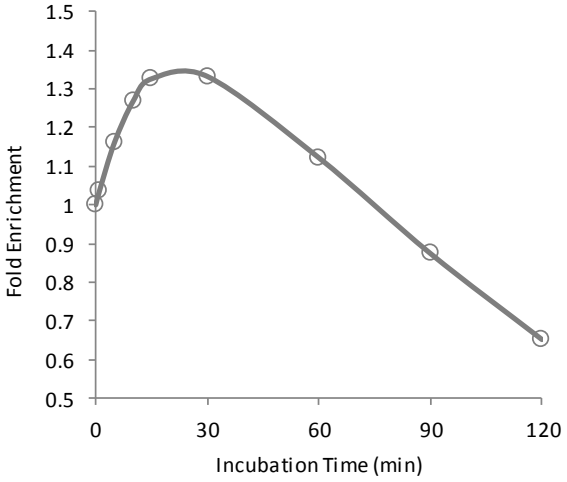


Figure 4.6 Optimization of the positive selection conditions. The model is based on the same initial device library as described in Figure 4.5 (Supplementary Table S4.2). (A) Fraction of uncleaved members within each population as a function of incubation time. The fraction of the cleaved population is determined by the first-order exponential decay equation: $1\text{-}e^{-kt}$, where k is the cleavage rate constant. The fractions of the uncleaved populations decrease at different rates as a function of time. (B) Fold enrichment of the target sequence relative to the medium and slow-cleaving populations as a function of

time. The fold enrichment of the target sequence is determined by the ratio of fraction compositions of the uncleaved target sequence in the library before and after the incubation. The modeling results predicted optimal fold enrichment of 1.3-fold for an incubation time ranged from 15 to 30 minutes.

4.3 Discussion

We proposed a solution-based ribozyme-based selection strategy to facilitate generation of new sensing functions. The solution-based selection strategy allows scaling of the selection process to diverse molecular targets. In contrast to a previously described selection with allosteric ribozymes, our proposed strategy offers several unique advantages by utilizing: (i) a device platform that has been demonstrated for gene-regulatory activities *in vivo*; (ii) an efficient cis-blocking strategy for gel-free, full-length RNA library generation by *in vitro* transcription reactions; and (iii) an efficient gel-free strategy for partitioning RNA fragments. Thus, our proposed strategy may provide a more rapid and efficient method for generating RNA sensing functions.

We demonstrated our cis-blocking strategy on a control L2b8-N10 library (~10^6 variants), which is much smaller than the library diversity (~10^{14}-10^{15} variants) necessary for generating new sensing functions. Because the blocking sequence was optimized based on the L2b8 device, the blocking efficiency decreased by ~40% when being applied to the L2b8-N10 library and may further decrease when applied to a library with even more random sequences introduced. The reduced efficiency of blocking ribozyme cleavage associated with larger device libraries during transcription reactions will require proper scaling of the transcription reactions to cover the desired sequence space or appending additional stabilizing nucleotides to improve the stability of the blocked conformations (see Chapter 3).

We described a dual selection strategy based on separation of cleaved and uncleaved fragments with magnetic beads. Selection strategies based on magnetic bead partitioning have been employed for the generation of both DNA and RNA aptamers to protein targets (19-21). These examples highlight the advantages of utilizing magnetic beads to facilitate the rapid isolation of bound sequences from unbound sequences. However, the removal and recovery of the unbound and bound sequences in the proposed dual selection scheme are performed by techniques (aspiration, pipetting) subject to researcher-dependent variations. Therefore, automatable technologies, such as liquid-handling robots and microfluidics, may be incorporated into our proposed selection strategy to develop a more reliable and consistent process.

A microfluidics-based process offers many advantages over the magnetic bead-based process by allowing: (i) lower fluid volumes (microfluidic channel vs. microfuge tube); (ii) quicker response upon change of fluids (continuous vs. batch process); (iii) better control of reaction conditions (i.e., constant temperature throughout the cleavage reaction); and (iv) higher throughputs by automating and parallelizing the selection processes across multiple microfluidic channels. To adapt our current magnetic bead-based process to microfluidics-based process, we need to develop microfluidic devices that are amenable for library selection. Specifically, efficient strategies need to be developed for modifying the device surface with the DNA activator strands to cover large sequence space. Challenges associated with RNA dissociation from the activator strand in a flow-based system will also need to be addressed to improve the overall fold enrichment of the microfluidics-based selection process (see Chapter 3).

Initial efforts will focus on the validation of our bead-based selection process on a

small control library, such as the described L2b8-N10 device library (~10^6 variants). Control selections will be performed on the library to recover the parent L2b8 device. Furthermore, the fold enrichment of the proposed selection process can be improved by the re-design and optimization of the transmitter sequence. Currently, the fold enrichment per selection round (~13-fold) is limited by the small differences in ribozyme device cleavage rates. Previous efforts on screening new transmitter sequences that support greater ligand-dependent fold change of gene-regulatory activities (and thus cleavage activities) only showed modest improvement (see Chapter 2). Because the regulatory activities of the ribozyme devices depend on the partitioning between two functional device conformations, we postulate that the relatively slow rates of conformational switching may limit the regulatory dynamic range. Thus, different component linkage strategies or alternative transmitter architectures may be considered to engineer an improved device platform that supports greater fold change. Ultimately, the *in vitro* selection will be integrated with the *in vivo* two-color FACS-based screen (see Chapter 2). *In vitro* selections will be performed on large device libraries (~10^{14}-10^{15} variants) to enrich library members that exhibit ligand-dependent cleavage activities. The enriched libraries (~10^6-10^7 variants) will be directly transformed into *Saccharomyces cerevisiae* to isolate library members that exhibit ligand-dependent gene-regulatory activities within the cellular environment. With further optimization and development, our ribozyme-based selection strategy may address current challenges in scaling the generation of small molecule-responsive ribozyme devices for broad cellular engineering applications.

4.4 Materials and Methods

4.4.1 *In vitro* transcription of cis-blocked ribozyme constructs

Sequences of the described L2b8 and L2b8-N10 constructs are provided in Supplementary Table S4.4. DNA synthesis was performed by Integrated DNA Technologies (Coralville, IA) or the Protein and Nucleic Acid Facility (Stanford, CA). The cis-blocked L2b8 construct was prepared through a T7 transcription reaction using PCR products amplified from templates using forward and reverse primers T7-fwd (5'-TTCTAATACGACTCACTATAGG) and b-L2b8-rev (5'- CAAAGCTGTTTCGTCCTC CGT), respectively. The cis-blocked L2b8-N10 construct was prepared through a T7 transcription reaction using PCR products amplified from templates using forward and reverse primers b-library-fwd (5'-TTCTAATACGACTCACTATAGGGAAACAAACA AAGTTGTTTTGCTGTCACCGGATGTGCTTTCCGG) and b-library-rev (5'-CAAAG CTGTTTCGTCCTCCGTCCCG), respectively. A total of 1-2 μg of PCR product was transcribed in a 25 μl reaction, consisting of the following components: 1×RNA Pol Reaction Buffer (New England Biolabs, Ipswich, MA), 2.5 mM of each rNTP, 1 μl RNaseOUT (Invitrogen, Carlsbad, CA), 10 mM $MgCl_2$, and 1 μl T7 Polymerase (New England Biolabs). For radiolabeled transcription reactions, 0.5 μCi α-^{32}P-GTP (MP Biomedicals, Solon, OH) was added to the reaction. After incubation at 37°C for 2 hr, NucAway Spin Columns (Ambion, Austin, TX) were used to remove unincorporated nucleotides from the transcription reactions according to manufacturer's instructions. The transcription products were cleaned up with the RNA Clean & Concentrator-5 (Zymo Research, Irvine, CA) kit according to manufacturer's instructions.

4.4.2 Gel analysis of cis-blocked ribozyme constructs

Radiolabeled cis-blocked transcription products were incubated for 5 min at 25ºC in the HBS-N buffer (Biacore, Uppsala, Sweden), comprising 10 mM HEPES and 150 mM NaCl, pH 7.4, in the absence and presence of 5 mM $MgCl_2$ and 60 pmol DNA initiator strand (5'- AAACAACTTTGTTTGTTTCCCCC). Reactions were quenched after 5 min with addition of 3 volumes of RNA stop/load buffer (95% formamide, 30 mM EDTA, 0.25% bromophenol blue, 0.25% xylene cyanol) on ice. Samples were heated at 95°C for 5 min, snap cooled on ice for 5 min, and size-fractionated on a denaturing (8.3 M Urea) 10% polyacrylamide gel at 25 W for 45 to 60 min. Gels were exposed for 10 min on a phosphor screen and analyzed for relative levels of the full-length transcript and cleaved products by phosphorimaging analysis on a FX Molecular Imager (Bio-Rad, Hercules, CA).

4.5 Supplementary Information

Supplementary Table S4.1 Fraction of each cleaved population, fold enrichment of the fast-cleaving population, and % input library recovered as a function of time during the negative selection.

Population	**Slow**	**Medium**	**Fast**
***k* (min^{-1})**	0.013	0.033	0.14
Time (min)	**Fraction cleaved = exp(-*k*t)**		
0	0	0	0
1	0.01	0.03	0.13
5	0.06	0.15	0.50
10	0.12	0.28	0.75
15	0.18	0.39	0.88
30	0.32	0.63	0.99
60	0.54	0.86	1.00

Number of members before selection		
Slow	**Medium**	**Fast**
3.33E+15	3.33E+15	3.33E+15

	Number of members before selection				
Time (min)	**Slow**	**Medium**	**Fast**	**Sum**	**% recovered**
0	0.00	0.00	0.00	0.00	0.00%
1	4.31E+13	1.08E+14	4.35E+14	5.87E+14	5.87%
5	2.10E+14	5.07E+14	1.68E+15	2.39E+15	23.95%
10	4.06E+14	9.37E+14	2.51E+15	3.85E+15	38.55%
15	5.91E+14	1.30E+15	2.93E+15	4.82E+15	48.17%
30	1.08E+15	2.09E+15	3.28E+15	6.45E+15	64.55%
60	1.81E+15	2.87E+15	3.33E+15	8.01E+15	80.11%

	Fold Enrichment	
Time (min)	**Fast/Slow**	**Fast/Medium**
0		
1	10.11	4.02
5	8.00	3.31
10	6.18	2.68
15	4.95	2.25
30	3.05	1.57
60	1.85	1.16

Supplementary Table S4.2 Fraction of each uncleaved population and fold enrichment of the target sequence as a function of time during the negative selection.

Population	**Slow**	**Medium**	**Fast**	**Target**
k **(min^{-1})**	0.013	0.033	0.14	0.025
Time (min)	**Fraction uncleaved = 1-exp(-*k*t)**			
0	1	1	1	1
1	0.99	0.97	0.87	0.98
5	0.94	0.85	0.50	0.88
10	0.88	0.72	0.25	0.78
15	0.82	0.61	0.12	0.69
30	0.68	0.37	0.01	0.47
60	0.46	0.14	0.0002	0.22
90	0.31	0.05	0.00	0.11
120	0.21	0.02	0.00	0.05

Number of members before selection				
	Slow	**Medium**	**Fast**	**Target**
	3.33E+15	3.33E+15	3.33E+15	10

Number of members after selection					
Time (min)	**Slow**	**Medium**	**Fast**	**Target**	**Sum**
0	3.33E+15	3.33E+15	3.33E+15	10.0	1.00E+16
1	3.29E+15	3.23E+15	2.90E+15	9.8	9.41E+15
5	3.12E+15	2.83E+15	1.66E+15	8.8	7.61E+15
10	2.93E+15	2.40E+15	8.22E+14	7.8	6.15E+15
15	2.74E+15	2.03E+15	4.08E+14	6.9	5.18E+15
30	2.26E+15	1.24E+15	5.00E+13	4.7	3.55E+15
60	1.53E+15	4.60E+14	7.50E+11	2.2	1.99E+15
90	1.03E+15	1.71E+14	1.12E+10	1.1	1.21E+15
120	7.00E+14	6.35E+13	1.69E+08	0.5	7.64E+14

Fraction composition before selection				
Slow	**Medium**	**Fast**	**Target**	
0.33	0.33	0.33	1.00E-15	
Fraction composition after selection				
Time (min)	**Slow**	**Medium**	**Fast**	**Target**	**Fold**
0	0.33	0.33	0.33	1.00E-15	1.00
1	0.35	0.34	0.31	1.04E-15	1.04
5	0.41	0.37	0.22	1.16E-15	1.16
10	0.48	0.39	0.13	1.27E-15	1.27
15	0.53	0.39	0.08	1.33E-15	1.33
30	0.64	0.35	0.01	1.33E-15	1.33
60	0.77	0.23	0.00	1.12E-15	1.12
90	0.86	0.14	0.00	8.74E-16	0.87
120	0.92	0.08	0.00	6.52E-16	0.65

Supplementary Table S4.3 Fold enrichment of the target sequence after one round of dual selection on the initial device library. The calculations are based on the fractions cleaved and uncleaved populations as described in Supplementary Table S4.1 and S4.2. An incubation time of 15 min and 1 min were used for the positive and negative selection, respectively, to determine the fold enrichment of the target sequence after one round of dual selection process (~5-fold).

	Number of members				
	Slow	**Medium**	**Fast**	**Target**	**Sum**
Initial library	3.33E+15	3.33E+15	3.33E+15	10	1.00E+16
Positive selection	2.93E+15	2.40E+15	8.22E+14	7.79	6.15E+15
Negative selection	3.78E+13	7.78E+13	1.07E+14	1.02	2.23E+14

	Fraction composition				
	Slow	**Medium**	**Fast**	**Target**	**Fold**
Initial library	0.33	0.33	0.33	1.00E-15	
Positive selection	0.48	0.39	0.13	1.27E-15	
Negative selection	0.17	0.35	0.48	4.56E-15	4.56

Supplementary Table S4.4 Summary of the cis-blocked ribozyme device sequences used in this work. The T7 promoter region is indicated in bold.

Construct	Sequence
L2b8	**TTCTAATACGACTCACTATAGGG**AAACAAACAAAGTTGTTT TGCTGTCACCGGATGTGCTTTCCGGTCTGATGAGTCCGTTGTC CATACCAGCATCGTCTTGATGCCCTTGGCAGGGACGGGACGG AGGACGAAACAGCTTTG
L2b8-N10	**TTCTAATACGACTCACTATAGGG**AAACAAACAAAGTTGTTT TGCTGTCACCGGATGTGCTTTCCGGTCTGATGAGTCCGTTGTC CANNNCNGCATCGTCTTGATGCCNNNGNNNGGGACGGGACG GAGGACGAAACAGCTTTG

Acknowledgements

This work was supported by the National Institutes of Health (R01GM086663), the National Science Foundation (CBET-0917638, CCF-0943269), the Defense Advanced Research Projects Agency (HR0011-11-2-0002), and the Alfred P. Sloan Foundation (fellowship to CDS).

References

1. Mercer, T.R., Dinger, M.E. and Mattick, J.S. (2009) Long non-coding RNAs: insights into functions. *Nat Rev Genet*, **10**, 155-159.
2. Liang, J.C., Bloom, R.J. and Smolke, C.D. (2011) Engineering biological systems with synthetic RNA molecules. *Mol Cell*, **43**, 915-926.
3. Ellington, A.D. and Szostak, J.W. (1990) In vitro selection of RNA molecules that bind specific ligands. *Nature*, **346**, 818-822.
4. Tuerk, C. and Gold, L. (1990) Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science*, **249**, 505-510.
5. Jenison, R.D., Gill, S.C., Pardi, A. and Polisky, B. (1994) High-resolution molecular discrimination by RNA. *Science*, **263**, 1425-1429.
6. Famulok, M. (1994) Molecular Recognition of Amino Acids by RNA-Aptamers: An L-Citrulline Binding RNA Motif and Its Evolution into an L-Arginine Binder. *Journal of the American Chemical Society*, **116**, 1698-1706.
7. Wang, Y. and Rando, R.R. (1995) Specific binding of aminoglycoside antibiotics to RNA. *Chemistry & biology*, **2**, 281-290.
8. Wallis, M.G., von Ahsen, U., Schroeder, R. and Famulok, M. (1995) A novel RNA motif for neomycin recognition. *Chemistry & biology*, **2**, 543-552.
9. Hermann, T. and Patel, D.J. (2000) Adaptive recognition by nucleic acid aptamers. *Science*, **287**, 820-825.

10. Berezovski, M.V., Musheev, M.U., Drabovich, A.P., Jitkova, J.V. and Krylov, S.N. (2006) Non-SELEX: selection of aptamers without intermediate amplification of candidate oligonucleotides. *Nat Protoc*, **1**, 1359-1369.

11. Koizumi, M., Soukup, G.A., Kerr, J.N. and Breaker, R.R. (1999) Allosteric selection of ribozymes that respond to the second messengers cGMP and cAMP. *Nat Struct Biol*, **6**, 1062-1071.

12. Winkler, W.C., Cohen-Chalamish, S. and Breaker, R.R. (2002) An mRNA structure that controls gene expression by binding FMN. *Proc Natl Acad Sci U S A*, **99**, 15908-15913.

13. Koizumi, M., Kerr, J.N., Soukup, G.A. and Breaker, R.R. (1999) Allosteric ribozymes sensitive to the second messengers cAMP and cGMP. *Nucleic Acids Symp Ser*, 275-276.

14. Khvorova, A., Lescoute, A., Westhof, E. and Jayasena, S.D. (2003) Sequence elements outside the hammerhead ribozyme catalytic core enable intracellular activity. *Nat Struct Biol*, **10**, 708-712.

15. Win, M.N. and Smolke, C.D. (2007) A modular and extensible RNA-based gene-regulatory platform for engineering cellular function. *Proc Natl Acad Sci U S A*, **104**, 14283-14288.

16. Chen, Y.Y., Jensen, M.C. and Smolke, C.D. (2010) Genetic control of mammalian T-cell proliferation with synthetic RNA regulatory systems. *Proc Natl Acad Sci U S A*, **107**, 8531-8536.

17. Mathews, D.H., Disney, M.D., Childs, J.L., Schroeder, S.J., Zuker, M. and Turner, D.H. (2004) Incorporating chemical modification constraints into a dynamic

programming algorithm for prediction of RNA secondary structure. *Proc Natl Acad Sci U S A*, **101**, 7287-7292.

18. Darty, K., Denise, A. and Ponty, Y. (2009) VARNA: Interactive drawing and editing of the RNA secondary structure. *Bioinformatics*, **25**, 1974-1975.

19. Wang, C., Yang, G., Luo, Z. and Ding, H. (2009) In vitro selection of high-affinity DNA aptamers for streptavidin. *Acta Biochimica et Biophysica Sinica*, **41**, 335-340.

20. Stoltenburg, R., Reinemann, C. and Strehlitz, B. (2005) FluMag-SELEX as an advantageous method for DNA aptamer selection. *Analytical and Bioanalytical Chemistry*, **383**, 83-91.

21. Daniels, D.A., Sohal, A.K., Rees, S. and Grisshammer, R. (2002) Generation of RNA aptamers to the G-protein-coupled receptor for neurotensin, NTS-1. *Anal Biochem*, **305**, 214-226.

Chapter 5.

A rapid, label-free, and quantitative method for characterizing binding properties of RNA aptamers

Abstract

Aptamers are nucleic acid sequences that are capable of binding molecular ligands with high affinities and provide important functional components for construction of synthetic genetic devices. RNA aptamers are typically generated *de novo* through iterative *in vitro* selection strategies to isolate sequences that bind small molecule ligands with high affinities and specificities from large nucleic acid libraries. Small molecule RNA aptamers have traditionally been characterized through isocratic elution and equilibrium filtration methods. However, these methods require labeling of molecules, thereby limiting their general applicability to characterize many RNA-small molecule pairs. Newer methods based on isothermal titration calorimetry (ITC) offer the potential of providing a label-free aptamer characterization strategy, but the requirement of large quantities of homogenous RNA makes the ITC-based method difficult to scale for the characterization of many RNA-ligand pairs. Here, we describe a rapid, label-free, and flexible aptamer characterization strategy based on surface plasmon resonance (SPR). We utilized a sensor surface modified with a DNA linker to capture the RNA aptamer that contains a complementary sequence to the DNA linker through hybridization. The small molecule ligand in free solution is injected over the RNA-hybridized surface to monitor binding interactions in real time. The rate of SPR signal change is fit to mathematical models to obtain associated thermodynamic and kinetic parameters. We validated the SPR-based method on several previously characterized RNA aptamers. Our SPR-based strategy requires little sample consumption and can be quickly tailored to qualitatively screen or quantitatively characterize many small molecule-RNA pairs.

5.1 Introduction

Aptamers are nucleic acid sequences that are capable of binding molecular ligands with high affinities (1). RNA aptamers are typically generated *de novo* through an iterative *in vitro* selection strategy or SELEX (*S*ystematic *E*volution of *L*igands by *EX*ponential enrichment) (2,3) or obtained from naturally-occurring sequences (4-6). RNA aptamers have been generated through *in vitro* selection strategies to many small molecule ligands, including alkaloids, amino acids, and oligosaccharides (1). By incorporating appropriate counterselection cycles in the selection process, aptamers have been selected *in vitro* to discriminate between structurally similar molecules. In one example, an aptamer selected to theophylline exhibits a 10,000-fold lower affinity for caffeine, which differs from theophylline by a single methyl group (7).

Binding affinities of small molecule RNA aptamers are typically determined by isocratic elution (8,9) or equilibrium filtration (7,9-11). Both methods require labeling of molecules (RNA or small molecule, respectively) and provide a low-throughput process for obtaining binding affinities. More recently, a label-free analytical method, isothermal titration calorimetry (ITC), has been used to characterize small molecule aptamers by measuring heat generation or absorption associated with the binding interactions (5,12). However, the ITC-based method requires large quantities of homogenous RNA, which limits its ability to be applied to the characterization of many RNA-small molecule pairs. In addition, all of these methods are limited to thermodynamic analysis of binding interactions, whereas for many applications information on binding kinetics are also desired.

Progress in biosensor technologies has led to the development of platforms based on surface plasmon resonance (SPR), which allows for the real-time, label-free detection of biomolecular interactions (13). The SPR-based biosensor platform involves immobilization of one molecule (ligand) on a sensor surface and monitoring its interaction with another molecule (analyte) in solution. The association or dissociation of the ligand-analyte complex results in a change in the refractive index, which is associated with the local mass density, near the sensor surface. The refractive index change is converted into an SPR signal, which is expressed in resonance units (RU). The rate of SPR signal change is fit to a mathematical model to acquire thermodynamic and kinetic parameters, including binding affinities, association rates, and disassociation rates. One of the pioneering SPR platforms, Biacore, has been applied to characterize thermodynamic and kinetics properties of diverse biomolecular interactions, including protein-protein (14), nucleic acid-protein (15), small molecule-nucleic acid (16), and small molecule-protein (17,18) interactions. In addition, the Biacore platform utilizes a microfluidic-based flow system that requires small sample volumes, thereby presenting a significant advantage over the ITC-based method.

Small molecule binding interactions have been challenging for the SPR-based methods to characterize due to the small local mass density change (19). The maximum expected SPR binding response is proportional to the ligand immobilization level on the sensor surface, analyte-to-ligand molecular weight ratio, and number of available binding sites on the ligand (20). Researchers have previously chosen to immobilize small molecule ligands on the sensor surface, such that binding of the corresponding RNA analyte resulted in a large binding response due to the large analyte-ligand molecular

weight ratio (16,21-24). However, this strategy requires the small molecule ligands to have or be modified with appropriate functional groups, such as the amine, thiol, maleimide, or aldehyde groups, for the conjugation chemistry to attach the ligand to the sensor surface and is therefore not scalable to the characterization of many diverse RNA-small molecule pairs. In addition, such conjugation strategies may impact functional groups that play a role in the ligand binding interactions and thus impact measured ligand binding properties.

We describe a rapid, quantitative aptamer characterization strategy based on the Biacore biosensor platform. An efficient strategy was developed for capturing RNA aptamers to the sensor surface, by appending a common binding sequence to the 5' end of each RNA aptamer sequence and covalently coupling a DNA linker sequence to the sensor surface, which serves to capture the modified aptamer sequences through programmed hybridization interactions. Small molecule analytes in solution were injected over the sensor surface to monitor binding interactions in real time. An efficient regeneration strategy was developed to remove the captured RNA from the sensor surface upon completion of the assay to allow for characterization of multiple RNA-small molecule pairs on the same sensor surface. We validated the SPR-based aptamer characterization strategy by measuring the binding properties of theophylline, citrulline, and arginine aptamers and compared the measured values to those determined by isocratic elution or equilibrium filtration. Our SPR-based strategy supports the rapid characterization of binding properties for RNA aptamers to small molecule ligands, thereby providing a powerful characterization tool for functional RNA design.

5.2 Results

5.2.1 Development of a SPR-based binding assay to characterize RNA-small molecule binding interaction

We developed a capture-based binding assay by indirectly immobilizing RNA through hybridization to a DNA linker sequence which is covalently coupled to the sensor surface (Figure 5.1). By capturing RNA to the sensor surface, small molecule analytes can be directly diluted in solution and injected over the sensor surface, thereby avoiding limitations associated with conjugating the small molecule to the sensor surface. The sensor surface for the binding assay is generated by covalently coupling a 24-mer poly-T DNA linker with an amino modification on the 5' end to a dextran-based sensor surface through an EDC-mediated amine coupling reaction. Each sensor surface includes two flow cells (FC1 and FC2), and both FCs were modified with the poly-T DNA linker. FC1 is used as a reference flow cell to correct for bulk refractive index changes, matrix effects, nonspecific binding, injection noise and baseline drift (25). The DNA linker allows the capture of RNA containing a 24-mer poly-A tail, which is appended to the 3' end of the RNA sequence to FC2 through hybridization. It is expected that the poly-A sequence interferes minimally with the native aptamer structure.

The DNA linker is coupled at high density (~4000 RU) to the sensor surface to increase the amount of RNA captured and therefore increase the small molecule binding signal. For a typical small molecule (>150 Da) binding assay, the capture of a 30 kDa RNA to ~1500 RU is sufficient to generate a reliable and measurable binding signal (>5 RU). The small molecule analyte at a particular concentration is subsequently injected over the sensor surface, and the SPR signal change associated with the binding

interaction is monitored. Since the capture of the RNA aptamer to the sensor surface is based on hybridization interactions with the DNA linker, a nucleic acid denaturant, such as 25 mM NaOH, can be injected to remove the RNA from the surface, thereby allowing reuse of the sensor surface to measure the binding signal at varying analyte concentrations.

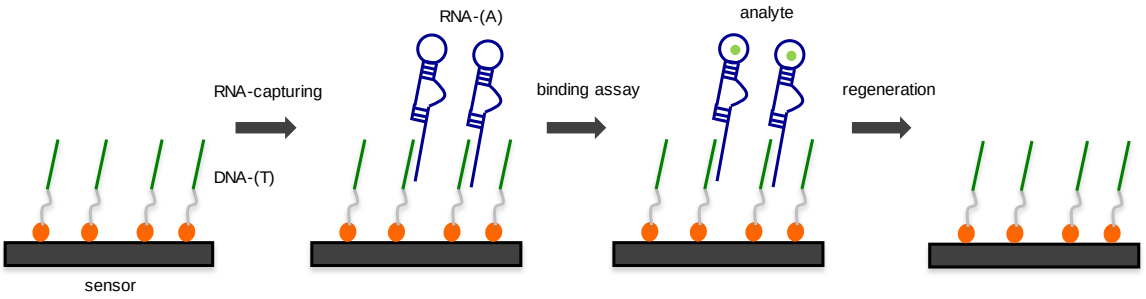


Figure 5.1 Schematic of the SPR-based RNA-small molecule binding assay. The sensor surface is generated by covalently linking a 5' amino-modified poly-T DNA strand to the carboxymethyl dextran sensor surface via an EDC-mediated amine coupling reaction. RNA, with a poly-A sequence at its 3'end, is hybridized to the DNA linker, followed by the injection of small molecule analyte to monitor binding interactions in real time. The surface is regenerated with 25 mM NaOH for the next assay.

5.2.2 Validation of the SPR-based aptamer binding assay

We applied our SPR-based method to measure the binding properties of three aptamer sequences that have been extensively characterized through different methods: the theophylline, citrulline, and arginine aptamers (7,9) (Figure 5.2). The theophylline aptamer has been characterized to have a binding affinity of 0.2-0.4 μM by equilibrium filtration (7) and fluorescence stopped-flow methods (26). The citrulline and arginine aptamers have been characterized by both isocratic elution and equilibration filtration methods to have a binding affinities of 62-68 and 56-76 μM, respectively (9). We performed the SPR-based binding assays on these three aptamer sequences at a series of

concentrations that span ~10-fold above and below the expected binding affinities for proper analysis (Figure 5.3). As proper formation of RNA tertiary structure can be dependent on $MgCl_2$ concentrations, the ligands were directly dissolved in the binding characterization buffer (10 mM HEPES, 150 mM NaCl, 5 mM $MgCl_2$, and 0.05% P20 surfactant, pH7.4) comprising the same $MgCl_2$ concentration as that used in the buffers used in the *in vitro* selection of these aptamers (26). The binding characterization buffer was also used as the Biacore running buffer to equilibrate the instrument.

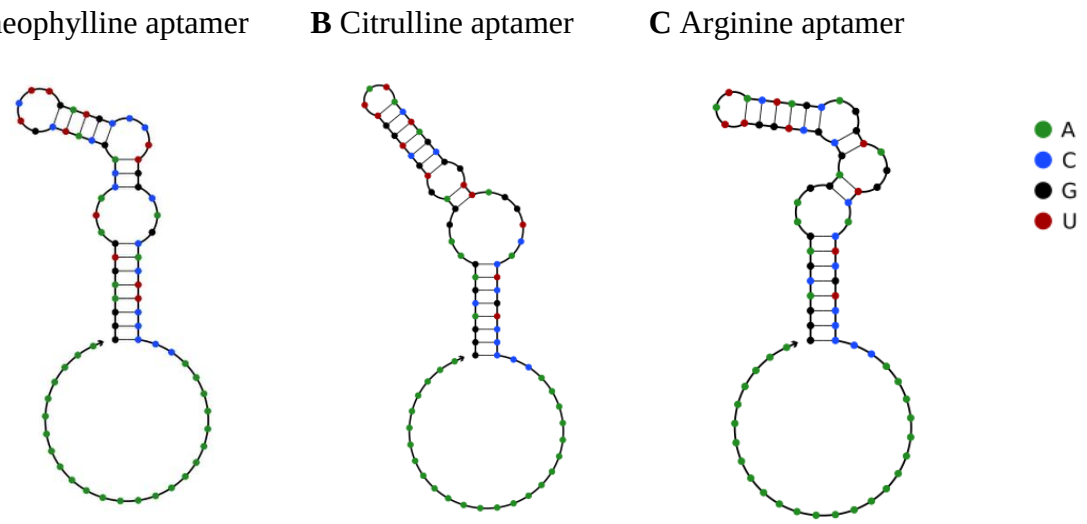


Figure 5.2 Predicted secondary structures of RNA aptamers with a 24-mer poly-A tail appended to the 3' end. The secondary structures of (A) theophylline, (B) citrulline, and (C) arginine aptamers were generated using Nupack by setting the temperature to 25ºC (27). The sequence for each aptamer is provided in Supplementary Table S5.1.

For the characterization of the theophylline aptamer, three independent replicated binding assays were performed. Each binding assay consists of multiple assay cycles, and each cycle includes an RNA-capture, theophylline-injection, and regeneration steps. For each assay cycle, ~10-50 ng/µl of RNA was captured onto the sensor surface for 2 minutes at a flow rate of 2 µl/min, resulting in an ~1700-2000 RU of RNA-capture level.

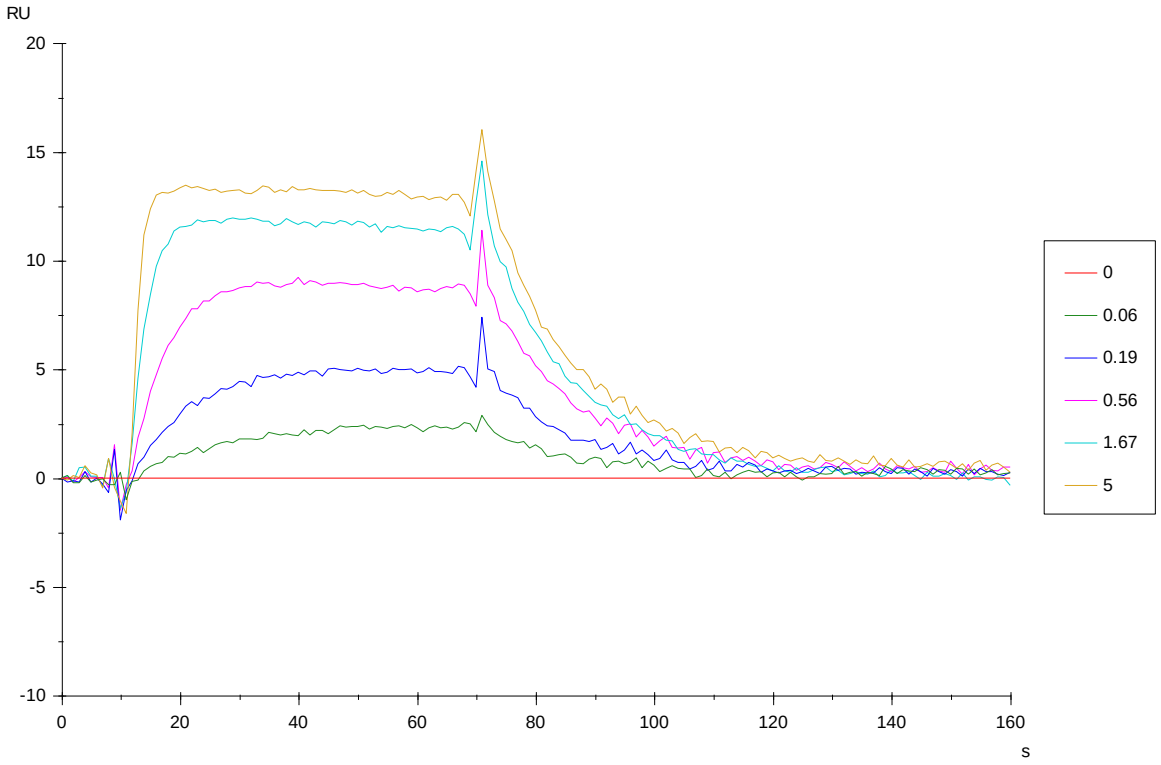
The RNA-capture levels were highly consistent within each replicate, with a coefficient of variation (CV) less than 0.46%. Varying concentrations of theophylline were injected over the sensor surface in random orders (0.56, 1.67, 0.19, 5, 0.06, and 0 μM) (Figure 5.3A). In between each injection of theophylline, the captured RNA was removed by an injection of 25 mM NaOH for 30 seconds at a flow rate of 30 μl/min, and new RNA was re-captured onto the sensor surface prior to the injection of next theophylline sample.

The sensorgrams from FC2 were referenced first by subtracting the corresponding sensorgrams from FC1 to correct for bulk refractive index changes, matrix effects, nonspecific binding, injection noise and baseline drift (25). The reference-subtracted sensorgram (FC2-FC1) was double-referenced by subtracting the sensorgram associated with the blank injection (0 μM) to account for any systematic drift over the course of the injection (Figure 5.3A). The double-referenced sensorgram was fit to a 1:1 kinetic binding model (Supplementary Figure S5.1) and a steady-state affinity model (Supplementary Figure S5.2). The association ($1.46\times10^{5} \pm 7.32\times10^{3}$ $M^{-1}s^{-1}$) and dissociation ($6.31\times10^{-2} \pm 4.72\times10^{-3}$ s^{-1}) rates were determined for the theophylline aptamer. The binding affinity (433 ± 37.8 nM) obtained from calculating the ratio of the dissociation over association constants was in close agreement with that obtained from the steady-state affinity model (324 ± 22.9 nM).

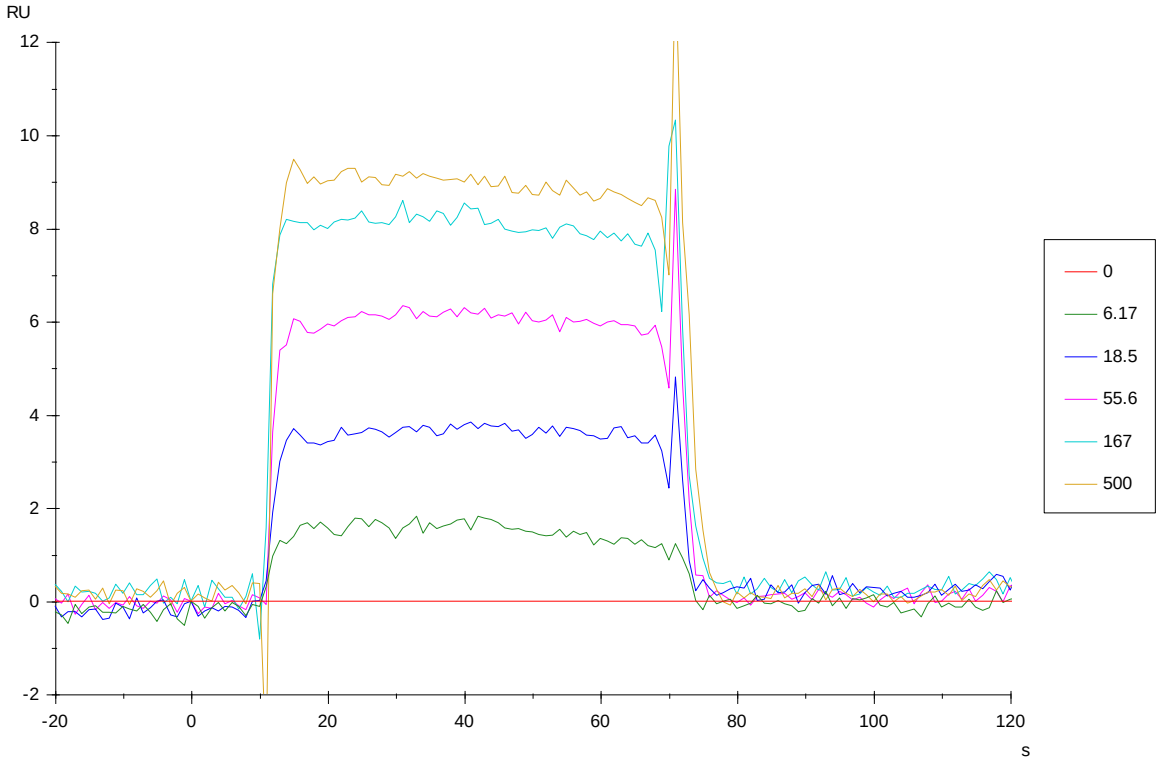
Similar binding assays and analyses were performed on the citrulline (Supplemental Figure S5.3) and arginine (Supplemental Figure S5.4) aptamers. For the characterization of citrulline and arginine aptamers, the RNA-capture level ranged from ~2200-2300 RU and ~1900-2100 RU, respectively. Varying concentrations of citrulline/arginine were injected over the sensor surface in random orders (0.56, 1.67,

0.19, 5, 0.06, and 0 μM) (Figure 5.3B and C). Kinetic analysis for the citrulline and arginine aptamers was not possible because the fast kinetic rates exceeded the machine detection limit (association rate: 10^3-10^7 $M^{-1}s^{-1}$; dissociation rate: 10^{-5}-0.1 s^{-1}). Binding affinities for the citrulline (23.2 ± 0.21 μM) and arginine (15.9 ± 0.32 μM) aptamers were obtained from the steady-state affinity model. For the arginine aptamer, the assay performed at 500 μM arginine was excluded from the analysis due to non-specific binding to the reference flow cell (Figure 5.3C and Supplementary Figure S5.4). The measured affinities for the three aptamers were in close agreement with those obtained by the more traditional methods, thus validating our SPR-based binding assay.

A Theophylline aptamer



B Citrulline aptamer



C Arginine aptamer

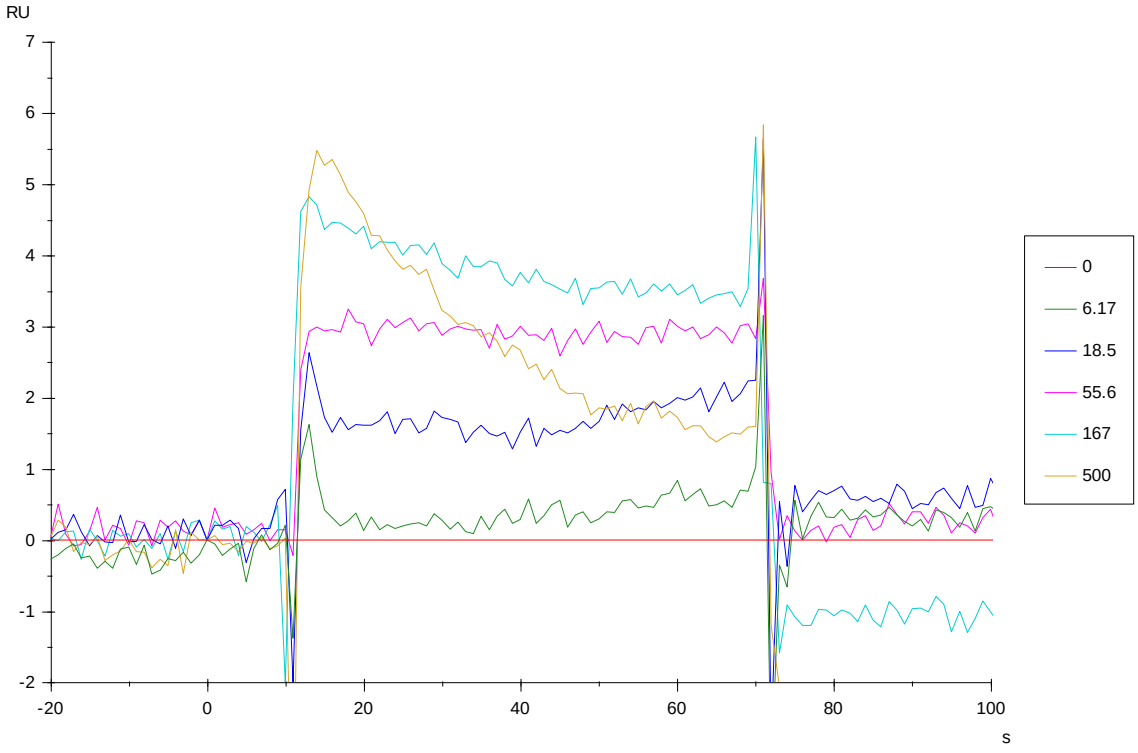


Figure 5.3 Biacore response (RU) as a function of time (s). Representative Biacore sensorgrams for (A) theophylline, (B) citrulline, and (C) arginine aptamers are shown here. The sensorgrams were double-referenced by the reference flow cell (FC1) and the blank injection as described in the main text. The indicated concentrations are in μM.

5.3 Discussion

We developed a rapid, flexible SPR-based aptamer characterization strategy that can be applied for the characterization of many aptamer-small molecule ligand binding interactions. By appending a common hybridization sequence to the RNA aptamer, we can capture different RNA sequences to the same sensor surface, thereby presenting a more cost-effective strategy than direct coupling of the RNA sequence to the sensor surface (23). A similar RNA-capture strategy has been previously employed for the characterization of RNA-protein binding interactions by immobilizing a biotin-modified DNA linker to a streptavidin-based sensor surface (28). Although the streptavidin-biotin binding interaction is essentially irreversible, with a dissociation constant of ~1 fM, the protein-based sensor surface is less resilient to the harsh denaturant (25 mM NaOH) used in our strategy to regenerate the surface. We chose to covalently immobilize the DNA linker to the dextran-based sensor surface, as the dextran is highly resistant to harsh chemicals, thereby increasing the stability of the characterization surface for many binding assays. In addition, for the Biacore system used in our strategy the dextran-based surface offers more reactive sites than the protein-based sensor surface for the immobilization of DNA linker. Therefore, by using a dextran-based sensor surface higher RNA-capture levels can be achieved, allowing a higher small molecule binding signal.

We demonstrated that our method is quantitative and yields similar values for thermodynamic binding properties as those obtained through the more traditional

methods based on equilibrium filtration or isocratic elution. Our SPR-based strategy allows measurements of kinetic parameters in addition to equilibrium parameters and requires little sample volume due to its microfluidic-based flow system, thereby providing a significant advantage over the ITC-based method. One limitation to our assay is in the possibility of non-specific interactions between the small molecule analyte and the sensor surface, particularly at high analyte concentrations. However, the extent of non-specific interactions may be eliminated or reduced by optimizing the characterization buffer.

By developing a flexible and rapid capture-based assay method, our strategy can be readily coupled to *in vitro* aptamer selection strategies to allow for: (i) screening of enriched aptamer libraries for qualitative (yes/no) binding; (ii) ranking of functional aptamer sequences based on their affinities; (iii) identifying consensus sequences through systematic truncation of functional aptamer sequences; and (vi) screening for specificities against a family of molecules that share structural similarities. Furthermore, our strategy can be applied to characterize binding interactions between nucleic acids (DNA, RNA) and diverse molecules (small molecules, DNA, RNA, proteins), or further extended to monitoring binding to more complex RNA molecules, such as synthetic RNA devices (29). Thus, the described SPR-based strategy provides a versatile and powerful tool for advancing functional RNA design.

5.4 Materials and Methods

5.4.1 *In vitro* transcription of RNA aptamers

DNA synthesis was performed by Integrated DNA Technologies (Coralville, IA) or the Protein and Nucleic Acid Facility (Stanford, CA). Aptamers with the poly-A sequence were prepared through a T7 transcription reaction using PCR products amplified from templates using forward and reverse primers Biacore-fwd (5'-TTCTAATACGACTCACTATAGG) and Biacore-rev (5'-TTTTTTTTTTTTTTTTTTTTTTTTTTTGGGG), respectively. All described aptamer template sequences are summarized in Supplementary Table S5.1. A total of 1-2 μg of PCR product was transcribed in a 25 μl reaction, consisting of the following components: 1×RNA Pol Reaction Buffer (New England Biolabs, Ipswich, MA), 2.5 mM of each rNTP, 1 μl RNaseOUT (Invitrogen, Carlsbad, CA), 10 mM $MgCl_2$, and 1 μl T7 Polymerase (New England Biolabs). After incubation at 37°C for 2 hr, NucAway Spin Columns (Ambion, Austin, TX) were used to remove unincorporated nucleotides from the transcription reactions according to the manufacturer's instructions.

5.4.2 Biacore sensor chip surface generation

Biosensor experiments were performed on a Biacore X100 instrument (Biacore, Uppsala, Sweden) at 25ºC. A CM5 sensor chip (Biacore) was docked in the Biacore X100 and equilibrated with the HBS-P buffer (Biacore) (10 mM HEPES, 150 mM NaCl, pH 7.4, 0.05% Surfactant P20) supplemented with 5 mM $MgCl_2$. The DNA linker strand (5'-TTTTTTTTTTTTTTTTTTTTTTTT), with an amino modification on the 5' end, was immobilized to the chip surface via standard amine-coupling chemistry. Briefly, the

carboxymethyl surface of the CM5 chip was activated for 7 min at a flow rate of 5 μl/min using a 1:1 volume ratio of 0.4 M 1-ethyl-3-(3-dimethylaminopropyl) carbodiimide (Biacore) and 0.1 M N-hydroxysuccinimide (Biacore). A molar ratio of 1:60 of DNA strand to cetyl trimethylammonium bromide (Amresco, Solon, OH) was diluted in 10 mM HEPES buffer (Sigma, St. Louis, MO) to a final concentration of 10 μM and 0.6 mM, respectively, and injected over the activated surface for 10 min at a flow rate of 5 μl/min. Excess activated groups were blocked by an injection of 1 M ethanolamine (Biacore), pH 8.5, for 7 min at a flow rate of 5 μl/min. The immobilization reaction was performed sequentially on both flow cells (FC1, FC2), where FC1 was used as a reference cell to correct for bulk refractive index changes, matrix effects, nonspecific binding, injection noise and baseline drift (25). Approximately 3000 RU of the DNA strand was immobilized using this protocol.

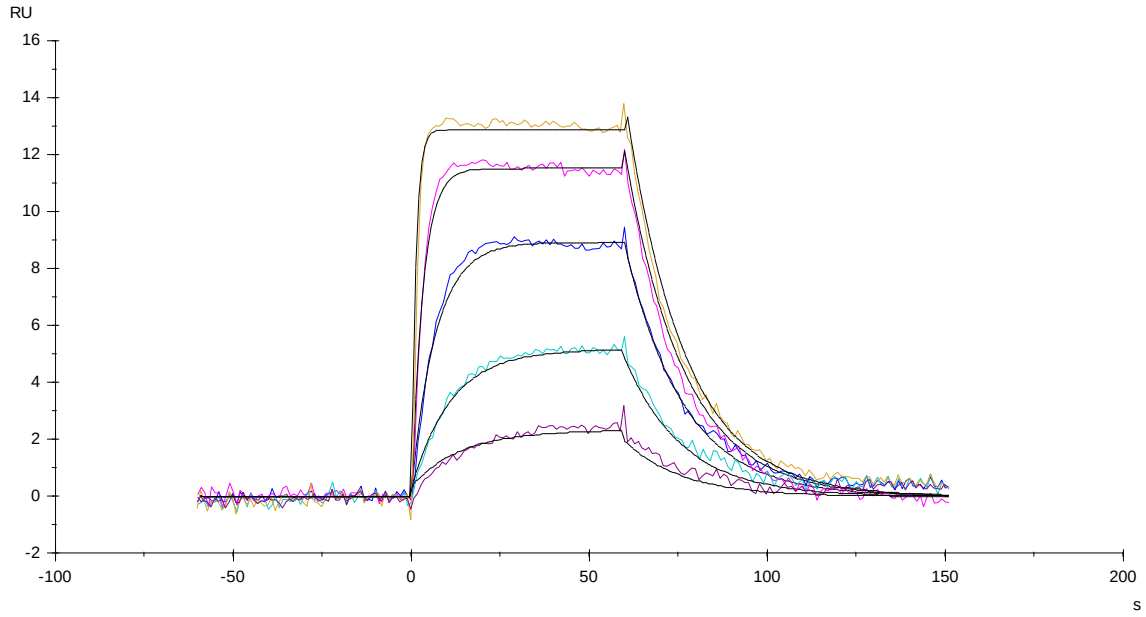
5.4.3 SPR-based aptamer binding assays

The Biacore X100 instrument was primed three times with the running buffer (HBS-P + 5 mM $MgCl_2$) prior to all binding assays. For each assay, five startup cycles were performed to stabilize the sensorgram baseline. For each startup cycle, the transcribed aptamer construct (~10-50 ng/μl) was captured onto the reaction flow cell (FC2) for 2 min at a flow rate of 2 μl/min (~1700~2300 RU), followed by an injection of 25 mM NaOH for 30 sec at a flow rate of 30 μl/min over both flow cells to regenerate the sensor surface. Dilution series of the ligand were made up in the running buffer and injected in random orders. Each concentration sample was injected over both flow cells for 60 sec at a flow rate of 30 μl/min, followed by a 90-sec dissociation phase during

which the running buffer flows over the sensor surface. The remaining RNA was removed by an injection of 25 mM NaOH for 30 sec at a flow rate of 30 μl/min over both flow cells, and RNA can be re-captured onto the sensor surface for the next assay cycle.

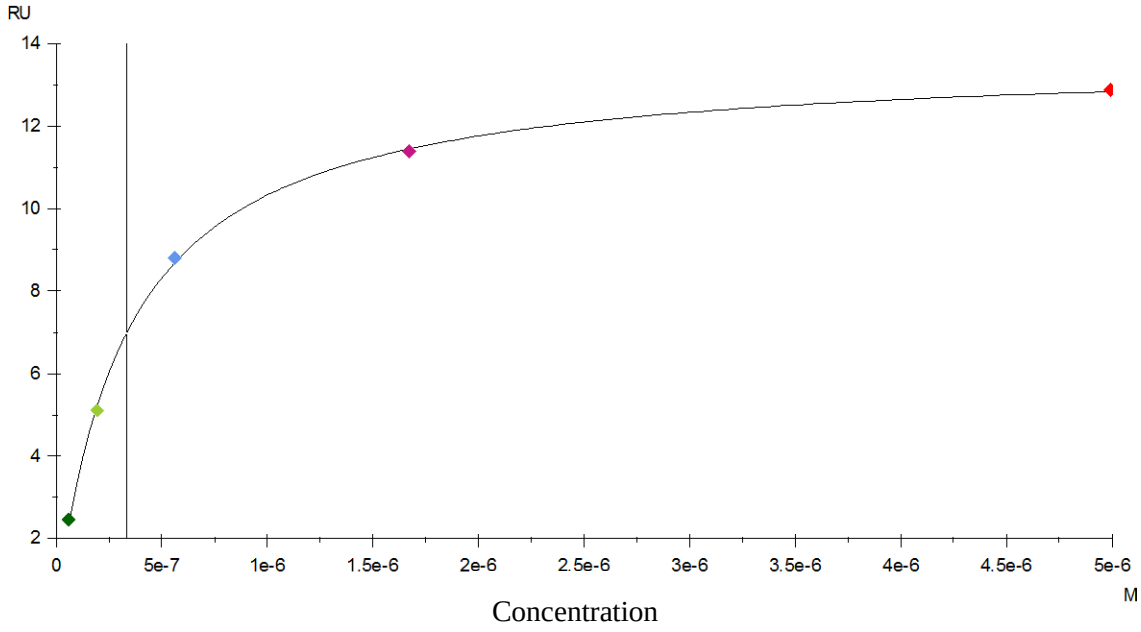
Data processing and analysis were performed using Biacore X100 Evaluation Software v2.0 (Biacore). A double-referencing method was performed to process all datasets (30). In brief, data from the reaction flow cell (FC2) was referenced first by subtracting data from the reference flow cell (FC1) to correct for bulk response due to differences in sample compositions. The reference-subtracted data (FC2-FC1) was referenced again by a blank injection (running buffer) to account for any systematic drift over the course of the injection. The double-referenced data was fit to the default 1:1 binding model for kinetic analysis (Supplementary Figure S5.2) or steady-state affinity model for thermodynamic analysis (Supplementary Figure S5.3). Reported values are the mean of at least three independent experiments.

5.5 Supplementary Information



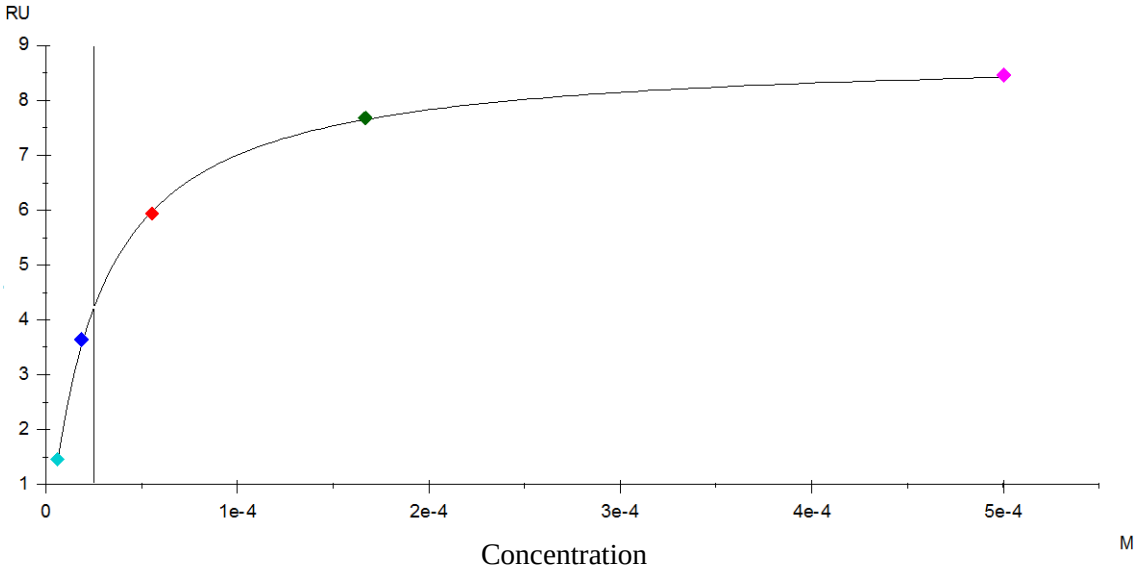
Conc. (µM)	RNA (RU)	ka (1/Ms)	kd (1/s)	KD (M)	RI (RU)
0.56	1994	1.54E+05	0.06077	3.96E-07	0.123
1.67	1993	**Rmax (RU)**	**tc**	**Chi² (RU²)**	-0.6335
0.19	1996	15.05	3.58E+10	0.0938	0.2921
5	1991				-1.048
0.06	1988				0.364
Mean	1992	**CV**	0.14%		

Supplementary Figure S5.1 Biacore 1:1 kinetic binding model for the theophylline aptamer sequence. The double-referenced sensorgrams for theophylline binding assays were fit to a 1:1 kinetic binding model (A+B↔AB): dA/dt = (tc*flow rate^(1/3))*(Conc-A) - (ka*A*B - kd*AB); dB/dt = - (ka*A*B - kd*AB); dAB/dt = (ka*A*B - kd*AB); Total response = AB + RI, with initial values A[0] = 0; B[0] = Rmax (maximum binding signal); AB[0] = 0. A, B, and AB represents the small molecule analyte concentration (Conc), unbound RNA, and small molecule-RNA complex. The model was fit to the sensorgrams to obtained tc (mass transfer constant), ka (association constant), kd (dissociation constant), Rmax, and RI (residuals). One representative fitted replicate is shown here. Small Chi^2 values indicate that the fitted values are close to the observed values.



KD (M)	Rmax (RU)	offset (RU)	Chi² (RU²)
3.37E-07	13.29	0.3887	0.0145

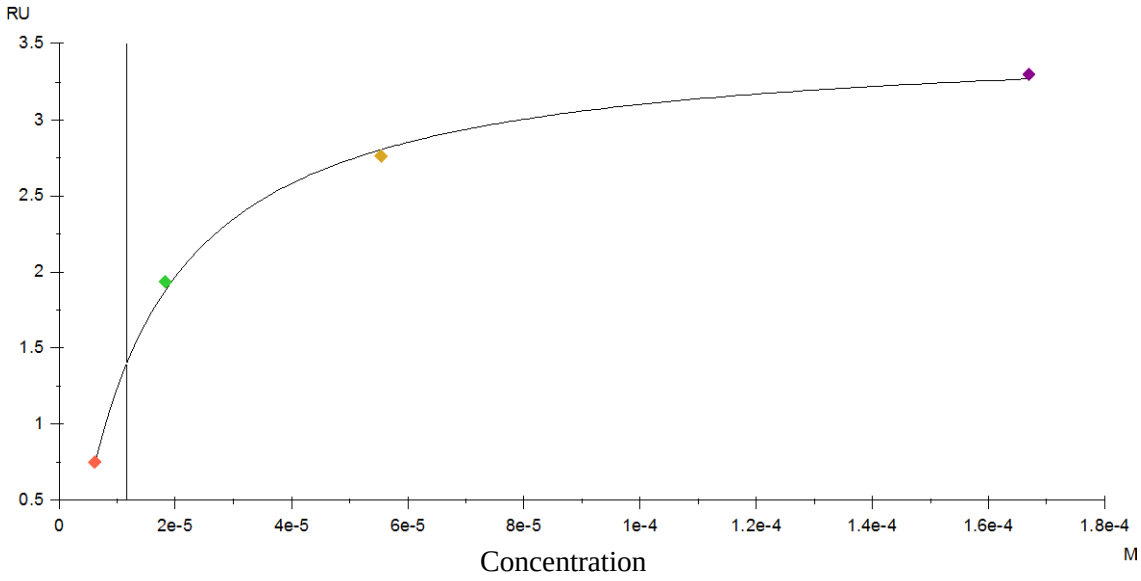
Supplementary Figure S5.2 Biacore steady-state affinity model for the theophylline aptamer sequence. The double-referenced sensorgrams for theophylline binding assays were fit to a steady-state affinity model: Conc*Rmax/(Conc+KD)+offset. The model was fit to the sensorgrams to obtained KD (binding affinity), Rmax (maximum binding signal), and offset (residuals). One representative fitted replicate is shown here. The top sensorgram shows the steady state RU taken for the affinity analysis. The bottom sensorgram shows the fitted curve against the experimental values. Small Chi^2 values indicate that the fitted values are close to the observed values.



KD (M)	Rmax (RU)	offset (RU)	Chi² (RU²)
2.51E-05	9.242	-0.3783	0.00514

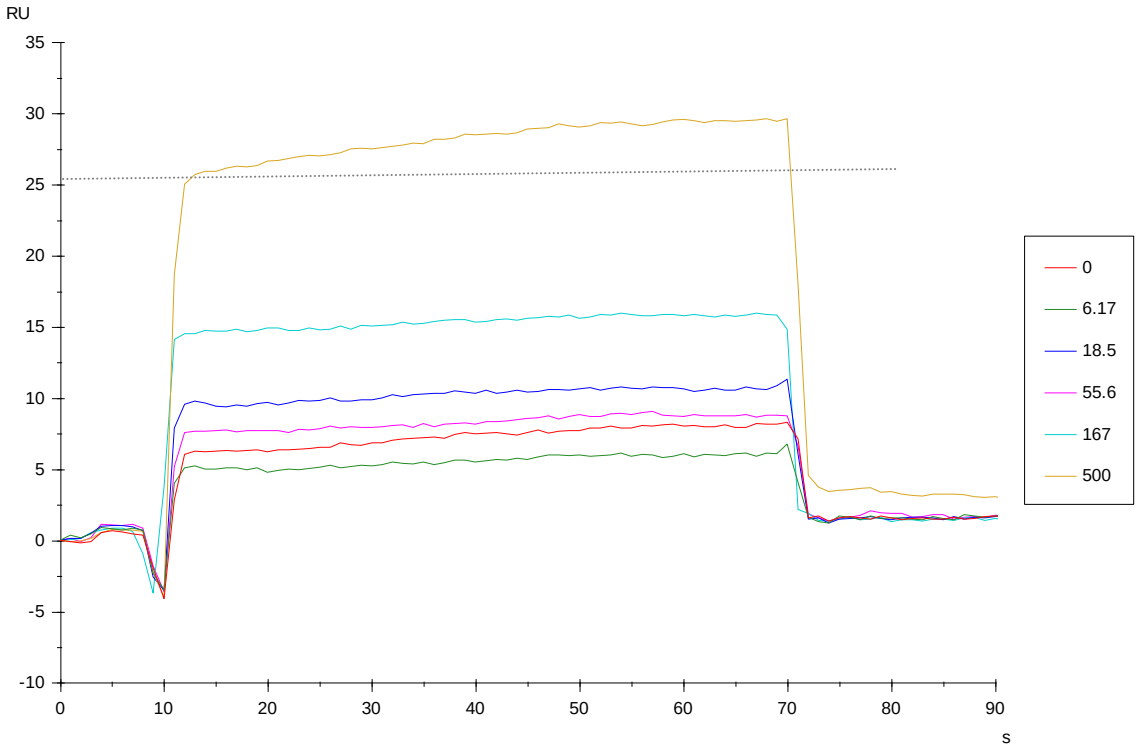
Supplementary Figure S5.3 Biacore steady-state affinity model for the citrulline aptamer sequence. The sensorgrams from the citrulline binding assays were fit to a steady-state affinity model as described in Supplementary Figure S5.2.

A



KD (M)	Rmax (RU)	offset (RU)	Chi² (RU²)
1.17E-05	4.281	-0.7322	0.00444

B



Supplementary Figure S5.4 Biacore steady-state affinity model for the arginine aptamer sequence. (A) The sensorgrams from the arginine binding assays were fit to a steady-state affinity model as described in Supplementary Figure S5.2. (B) The assay performed at 500 μM arginine was excluded from the analysis due to non-specific binding to the reference flow cell. The non-specific binding was observed by the difference between the response curve and the dashed line.

Supplementary Table S5.1 Sequences of aptamers used in this work. The T7 promoter region is indicated in bold and the poly-A linker sequence is indicated in italics.

Aptamer	Sequence
Theophylline	**TTCTAATACGACTCACTATAGGG**AAGTGATACCAGCATCGTCTTGATGCCCTTGGCAGCACTTCCCCC*AAAAAAAAAAAAAAAAAAAAAA*
Citrulline	**TTCTAATACGACTCACTATAGGG**ACGAGAAGGAGTGCTGGTTATACTAGCGGTTAGGTCACTCGTCCCCC*AAAAAAAAAAAAAAAAAAAAAAAA*
Arginine	**TTCTAATACGACTCACTATAGGG**ACGAGAAGGAGCGCTGGTTATACTAGCAGGTAGGTCACTCGTCCCCC*AAAAAAAAAAAAAAAAAAAAAAAA*

Acknowledgements

This work was supported by the National Institutes of Health (R01GM086663), the National Science Foundation (CBET-0917638, CCF-0943269), the Defense Advanced Research Projects Agency (HR0011-11-2-0002), and the Alfred P. Sloan Foundation (fellowship to CDS).

References

1. Hermann, T. and Patel, D.J. (2000) Adaptive recognition by nucleic acid aptamers. *Science*, **287**, 820-825.
2. Ellington, A.D. and Szostak, J.W. (1990) In vitro selection of RNA molecules that bind specific ligands. *Nature*, **346**, 818-822.
3. Tuerk, C. and Gold, L. (1990) Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science*, **249**, 505-510.
4. Wieland, M., Benz, A., Klauser, B. and Hartig, J.S. (2009) Artificial ribozyme switches containing natural riboswitch aptamer domains. *Angew Chem Int Ed Engl*, **48**, 2715-2718.
5. Dixon, N., Duncan, J.N., Geerlings, T., Dunstan, M.S., McCarthy, J.E., Leys, D. and Micklefield, J. (2010) Reengineering orthogonally selective riboswitches. *Proc Natl Acad Sci U S A*, **107**, 2830-2835.
6. Greenleaf, W.J., Frieda, K.L., Foster, D.A., Woodside, M.T. and Block, S.M. (2008) Direct observation of hierarchical folding in single riboswitch aptamers. *Science*, **319**, 630-633.
7. Jenison, R.D., Gill, S.C., Pardi, A. and Polisky, B. (1994) High-resolution molecular discrimination by RNA. *Science*, **263**, 1425-1429.
8. Brockstedt, U., Uzarowska, A., Montpetit, A., Pfau, W. and Labuda, D. (2004) In vitro evolution of RNA aptamers recognizing carcinogenic aromatic amines. *Biochem Biophys Res Commun*, **313**, 1004-1008.

9. Famulok, M. (1994) Molecular Recognition of Amino Acids by RNA-Aptamers: An L-Citrulline Binding RNA Motif and Its Evolution into an L-Arginine Binder. *Journal of the American Chemical Society*, **116**, 1698-1706.

10. Mannironi, C., Di Nardo, A., Fruscoloni, P. and Tocchini-Valentini, G.P. (1997) In vitro selection of dopamine RNA ligands. *Biochemistry*, **36**, 9726-9734.

11. Lorsch, J.R. and Szostak, J.W. (1994) In vitro selection of RNA aptamers specific for cyanocobalamin. *Biochemistry*, **33**, 973-982.

12. Muller, M., Weigand, J.E., Weichenrieder, O. and Suess, B. (2006) Thermodynamic characterization of an engineered tetracycline-binding riboswitch. *Nucleic Acids Res*, **34**, 2607-2617.

13. Rich, R.L., Papalia, G.A., Flynn, P.J., Furneisen, J., Quinn, J., Klein, J.S., Katsamba, P.S., Waddell, M.B., Scott, M., Thompson, J. *et al.* (2009) A global benchmark study using affinity-based biosensors. *Anal Biochem*, **386**, 194-216.

14. Leonard, P., Hearty, S. and O'Kennedy, R. (2011) Measuring protein-protein interactions using Biacore. *Methods Mol Biol*, **681**, 403-418.

15. Katsamba, P.S., Park, S. and Laird-Offringa, I.A. (2002) Kinetic studies of RNA-protein interactions using surface plasmon resonance. *Methods*, **26**, 95-104.

16. Win, M.N., Klein, J.S. and Smolke, C.D. (2006) Codeine-binding RNA aptamers and rapid determination of their binding constants using a direct coupling surface plasmon resonance assay. *Nucleic Acids Res*, **34**, 5670-5682.

17. Myszka, D.G. (2004) Analysis of small-molecule interactions using Biacore S51 technology. *Anal Biochem*, **329**, 316-323.

18. Cannon, M.J., Myszka, D.G., Bagnato, J.D., Alpers, D.H., West, F.G. and Grissom, C.B. (2002) Equilibrium and kinetic analyses of the interactions between vitamin B(12) binding proteins and cobalamins by surface plasmon resonance. *Anal Biochem*, **305**, 1-9.

19. Cannon, M.J., Papalia, G.A., Navratilova, I., Fisher, R.J., Roberts, L.R., Worthy, K.M., Stephen, A.G., Marchesini, G.R., Collins, E.J., Casper, D. *et al.* (2004) Comparative analyses of a small molecule/enzyme interaction by multiple users of Biacore technology. *Anal Biochem*, **330**, 98-113.

20. Murphy, M., Jason-Moller, L. and Bruno, J. (2006) Using Biacore to measure the binding kinetics of an antibody-antigen interaction. *Current protocols in protein science*, **Chapter 19**, Unit 19 14.

21. Kwon, M., Chun, S.M., Jeong, S. and Yu, J. (2001) In vitro selection of RNA against kanamycin B. *Molecules and Cells*, **11**, 303-311.

22. Gebhardt, K., Shokraei, A., Babaie, E. and Lindqvist, B.H. (2000) RNA aptamers to S-adenosylhomocysteine: kinetic properties, divalent cation dependency, and comparison with anti-S-adenosylhomocysteine antibody. *Biochemistry*, **39**, 7255-7265.

23. Davis, J.H. and Szostak, J.W. (2002) Isolation of high-affinity GTP aptamers from partially structured RNA libraries. *Proc Natl Acad Sci U S A*, **99**, 11616-11621.

24. de-los-Santos-Alvarez, N., Lobo-Castanon, M.J., Miranda-Ordieres, A.J. and Tunon-Blanco, P. (2009) SPR sensing of small molecules with modified RNA aptamers: detection of neomycin B. *Biosensors & Bioelectronics*, **24**, 2547-2553.

25. Myszka, D.G. (2000) Kinetic, equilibrium, and thermodynamic analysis of macromolecular interactions with BIACORE. *Methods Enzymol*, **323**, 325-340.

26. Jucker, F.M., Phillips, R.M., McCallum, S.A. and Pardi, A. (2003) Role of a heterogeneous free state in the formation of a specific RNA-theophylline complex. *Biochemistry*, **42**, 2560-2567.

27. Zadeh, J.N., Steenberg, C.D., Bois, J.S., Wolfe, B.R., Pierce, M.B., Khan, A.R., Dirks, R.M. and Pierce, N.A. (2011) NUPACK: Analysis and design of nucleic acid systems. *Journal of Computational Chemistry*, **32**, 170-173.

28. Misono, T.S. and Kumar, P.K. (2005) Selection of RNA aptamers against human influenza virus hemagglutinin using surface plasmon resonance. *Anal Biochem*, **342**, 312-317.

29. Liang, J.C., Bloom, R.J. and Smolke, C.D. (2011) Engineering biological systems with synthetic RNA molecules. *Mol Cell*, **43**, 915-926.

30. Myszka, D.G. (1999) Improving biosensor analysis. *J Mol Recognit*, **12**, 279-284.

Chapter 6.

Conclusions

Living systems rely on the precise control of gene expression to affect appropriate phenotypic responses over different temporal and spatial scales, and therefore maintain proper cellular function. Increased understanding of natural biological systems, in addition to advances in nucleic acid and protein engineering, have inspired researchers to construct synthetic genetic systems and catalyze the emerging field of synthetic biology (1). Recent advances in the field of RNA synthetic biology led to the development of a modular platform from which ligand-responsive ribozyme-based gene-regulatory devices can be assembled (2,3). The ability to construct user-specified ribozyme devices tailored to diverse applications depends on the development and integration of several enabling technologies. Specifically, high throughput technologies that enable the efficient *de novo* generation of new sensor functions and tailoring of device regulatory properties for specific cellular applications are needed to broaden the utility of these genetic control devices.

The ribozyme device platform offers two distinct advantages in advancing our ability to generate new ligand-responsive RNA control devices with tailored regulatory properties. First, the modular composition framework of the device platform supports direct interchange of functional RNA components, including sensors, transmitters, and actuators, to program devices with user-specified regulatory activities. Second, the mechanism of action of the ribozyme component is independent of cell-specific machinery, allowing transport of the resultant ribozyme devices to *in vitro* or diverse cellular environments. These two advantages allow direct generation and optimization of the individual RNA component activities within the device platform through *in vitro* and/or *in vivo* strategies, thereby allowing the development of a scalable framework from

which ribozyme devices with tailored ligand responsiveness and regulatory properties can be efficiently generated.

We developed a high-throughput and quantitative two-color FACS-based strategy that allows rapid generation and optimization of component activities in the cellular environment. Our fluorescence-based screening strategy provides an advantage in that the device output (GFP) can be normalized by an internal standard (mCherry) across cell populations, whereas cell-based screening and selection strategies based on enzyme activities (i.e., motility, colorimetry, viability) are more prone to gene expression noise that is not normalizable (4-6). Our two-color screening approach can be extended beyond RNA device tailoring. As an example, our screening approach can be applied toward direct enzyme evolution in cellular environments. In this instance, an enzyme library is transformed into a cell host, which harbors an integrated biosensor in the form of an RNA control device responsive to the product of the enzyme-catalyzed reaction in the two-color screen construct. The RNA device can respond to the intracellular accumulation of the enzymatic product and translate this information to increased cellular fluorescence levels. As the fluorescence levels are normalized by the internal standard, enzyme variants that exhibit enhanced activities can be efficiently isolated, increasing the sensitivity of the biosensor to small changes in enzyme activity. Thus, our two-color FACS-based screen can provide a powerful strategy that allows programming of diverse biological components.

Scalable generation of biological components is currently a rate-limiting step in the process of building new genetic control devices. Generation of new component functions, such as ones that exhibit new sensing functions, requires searching large

sequence space (~10^{14}-10^{15} variants), exceeding the capacity of any existing cell-based selection or screening strategies. We proposed a solution-based *in vitro* selection strategy based on the modular ribozyme device platform. Large device libraries are generated by randomizing the sensor component within the device platform. The selection strategy links ligand binding to ribozyme cleavage activity and therefore does not require conjugation of target ligand to a solid support (i.e., column). As such, the solution-based selection process can be applied to a complex ligand pool in which mixtures of ligand targets are simultaneously presented to the RNA library to generate multiple sensing functions. We described a dual selection process, optimized by mathematical modeling, based on magnetic bead separation to facilitate partitioning of the library members. Future efforts will be directed to validating our selection strategy initially on a small control library to recover existing functional sequences and ultimately on a large library to generate new sensing functions.

The enrichment efficiencies of our *in vitro* cleavage-based selection and *in vivo* FACS-based screening strategies are closely associated with the device regulatory dynamic range. Our current platform design is based on conformational switching facilitated by a strand-displacing transmitter sequence. Previous modeling studies of riboswitch function have predicted that slow conformational switching rates may limit the device dynamic range, resulting in small changes in the output signal in response to changing concentrations of the target ligand (7). This limitation has a more pronounced impact on the enrichment efficiency of the *in vitro* selection. In contrast to the *in vivo* FACS-based screen, where library members exhibiting desired activities can be specifically sorted, the proposed *in vitro* cleavage-based selection collects all the cleaved

library members in the supernatant. As such, library members exhibiting similar cleavage rates relative to the target member will also be carried over into the next selection round, resulting in poor enrichment of the target member. To address this limitation, early efforts may be directed toward developing a transmitter design that allows for faster conformational switching. One potential strategy is to reduce the random walk step size, a measure of the number of base pairs that need to be broken or formed between two conformations. As an example, a more compact strand displacement-based transmitter sequence (with fewer nucleotides) will be designed through rational or evolutionary strategies to facilitate faster conformational switching. Alternative transmitter architectures, such as design based on a helix-slipping mechanism (8), may also be considered for the development of a next-generation device platform.

By developing enabling technologies based on the same ribozyme-based platform, we will ultimately be able to integrate our *in vitro* solution-based selection and *in vivo* two-color FACS-based screening strategies into one seamless integrated process. The *in vitro* selection strategy can be performed initially on a large device library (~10^{14}-10^{15} variants) to enrich for sequences that exhibit desired cleavage activities. The *in vitro*-enriched libraries can be directly transformed into a cell host (~10^{6}-10^{7} clones for yeast) and further enriched and optimized by the *in vivo* two-color FACS-based screen. To integrate the resulting devices into biological networks, the regulatory activities of the devices can be readily tailored by the same two-color screen to meet application-specific performance requirement. Collectively, progress in this thesis will aid in the scalable construction and programming of RNA control devices for diverse cellular engineering applications and further advance our ability to interact with and program biology.

References

1. Khalil, A.S. and Collins, J.J. (2010) Synthetic biology: applications come of age. *Nat Rev Genet*, **11**, 367-379.
2. Win, M.N., Liang, J.C. and Smolke, C.D. (2009) Frameworks for programming biological function through RNA parts and devices. *Chemistry & Biology*, **16**, 298-310.
3. Win, M.N. and Smolke, C.D. (2007) A modular and extensible RNA-based gene-regulatory platform for engineering cellular function. *Proceedings of the National Academy of Sciences*, **104**, 14283-14288.
4. Lynch, S.A., Desai, S.K., Sajja, H.K. and Gallivan, J.P. (2007) A high-throughput screen for synthetic riboswitches reveals mechanistic insights into their function. *Chemistry & Biology*, **14**, 173-184.
5. Topp, S. and Gallivan, J.P. (2008) Random walks to synthetic riboswitches--a high-throughput selection based on cell motility. *Chembiochem : A European Journal of Chemical Biology*, **9**, 210-213.
6. Nomura, Y. and Yokobayashi, Y. (2007) Dual selection of a genetic switch by a single selection marker. *Biosystems*, **90**, 115-120.
7. Beisel, C.L. and Smolke, C.D. (2009) Design principles for riboswitch function. *PLoS Comput Biol*, **5**, e1000363.
8. Soukup, G.A. and Breaker, R.R. (1999) Engineering precision RNA molecular switches. *Proc Natl Acad Sci U S A*, **96**, 3584-3589.