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.

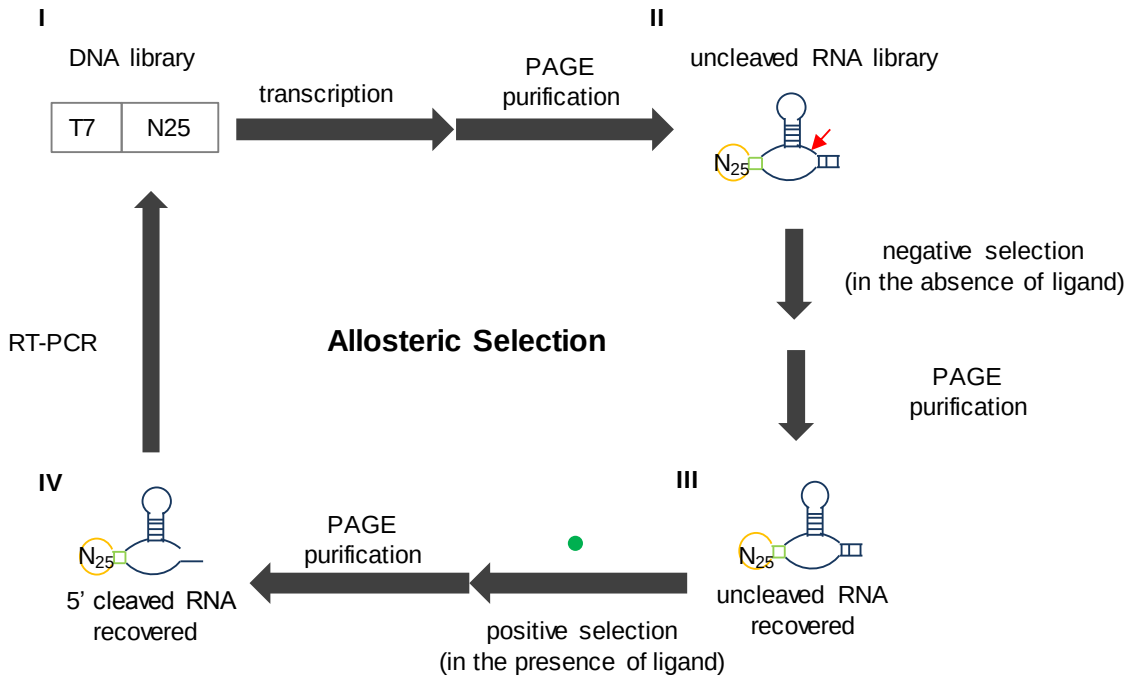


Figure 4.1 Schematic of an *in vitro* allosteric ribozyme selection strategy (11). The RNA library is generated from the DNA library by transcription reactions (I). The full-length RNA is recovered by PAGE and is (II) subject to a negative selection in the absence of ligand. The red arrow indicates the 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.

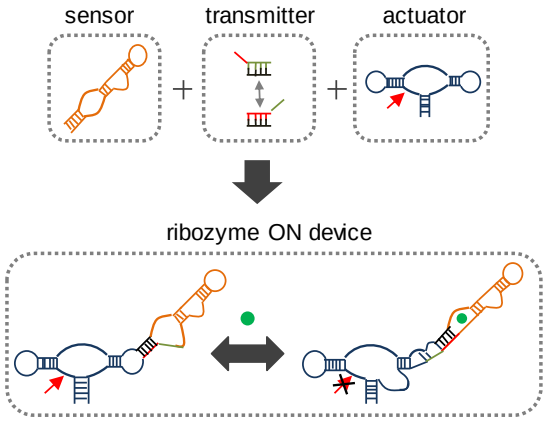


Figure 4.2 A modular ribozyme device platform allowing construction of devices that are amenable to cellular engineering applications. A ribozyme device is comprised of three RNA components: sensor (aptamer), actuator (ribozyme), and transmitter (sequence capable of a strand-displacement event). The 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

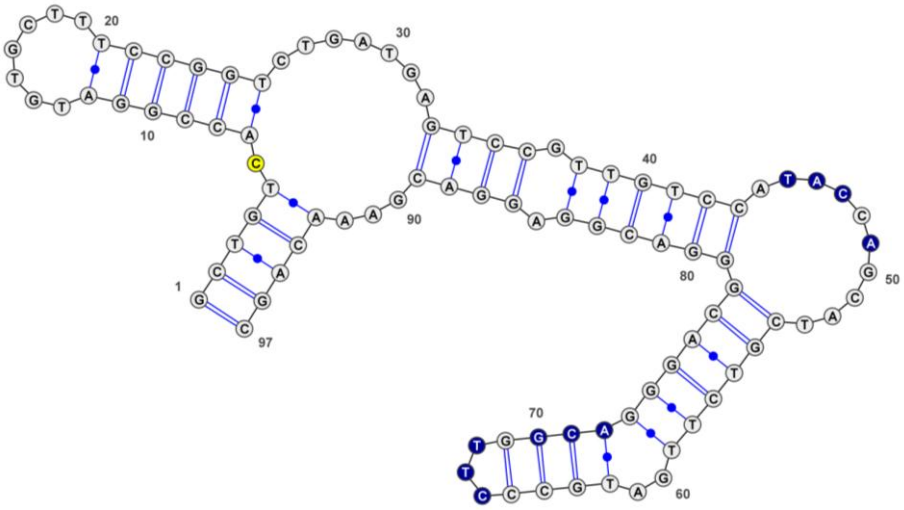
One key challenge in developing a scalable ribozyme-based selection strategy is the efficient generation of the full-length RNA library due to significant ribozyme

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.

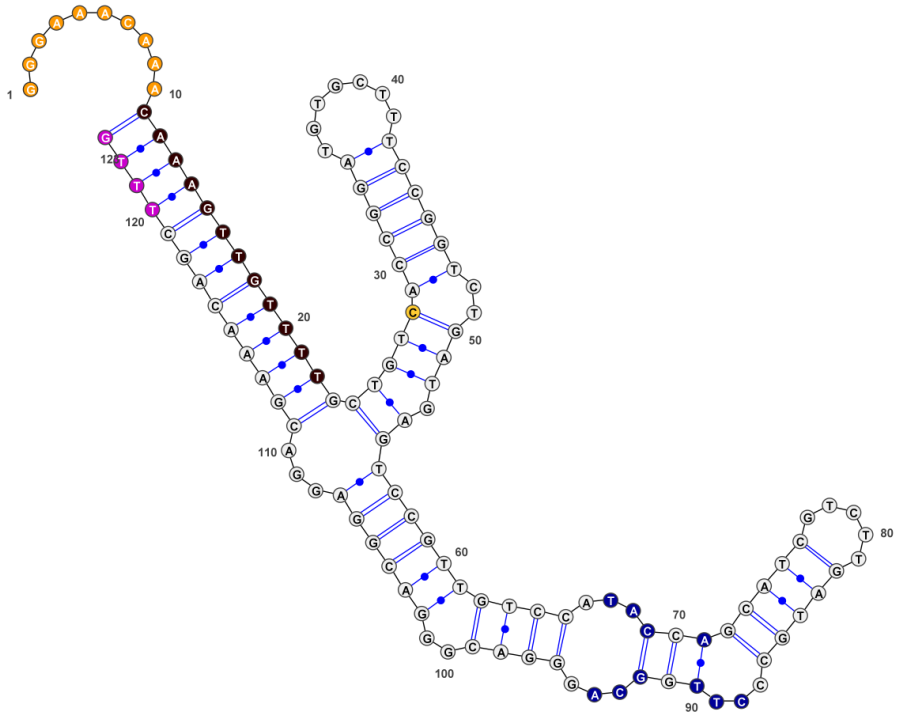
We applied the blocking, activation, and stabilizing sequences previously developed for the ribozyme-based devices to a small control device library (~10^6 variants), L2b8-N10. The L2b8-N10 library was generated by randomizing 10 nt positions in the aptamer binding core (sensor component) of a theophylline-responsive 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

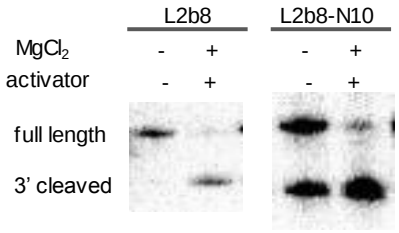


Figure 4.3 The cis-blocking strategy enables efficient generation of full-length ribozyme device library. The predicted secondary structures of the L2b8 device are shown here. (A) The L2b8 ribozyme-active device functional conformation is shown here. The randomized positions in the theophylline aptamer 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.

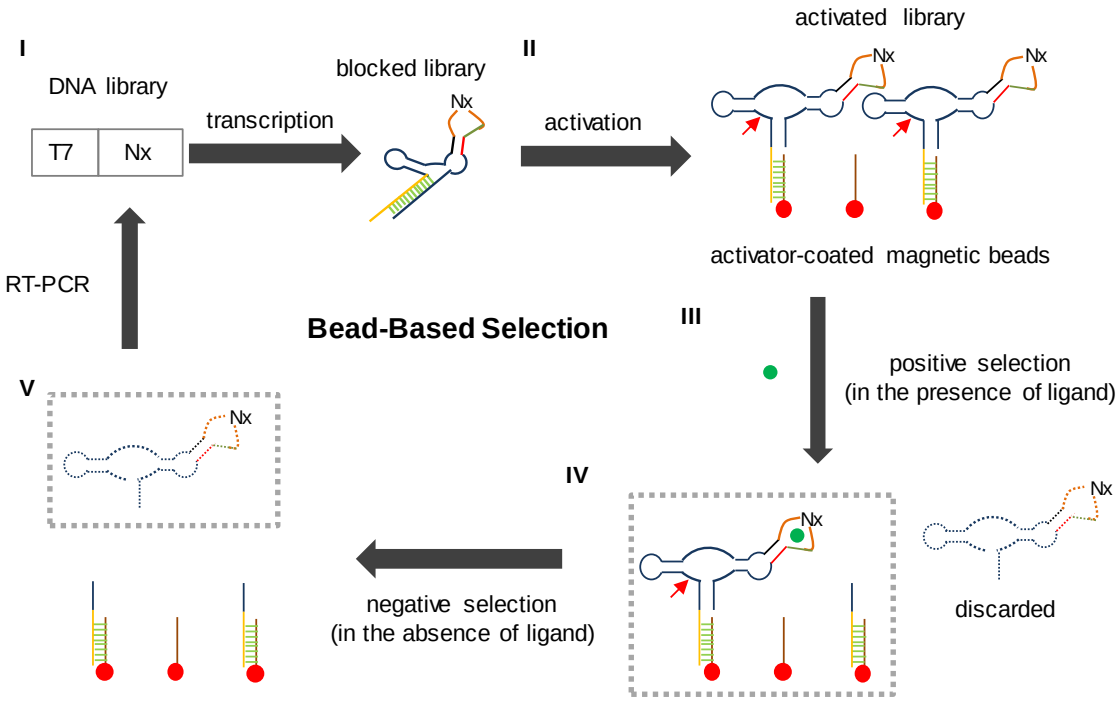


Figure 4.4 Schematic of the dual selection strategy facilitated by magnetic bead-based separation. The cis-blocked RNA device library is transcribed from DNA template (I). The blocked RNA is activated by incubating with the DNA activator strands immobilized to magnetic beads (II). The red arrow indicates 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

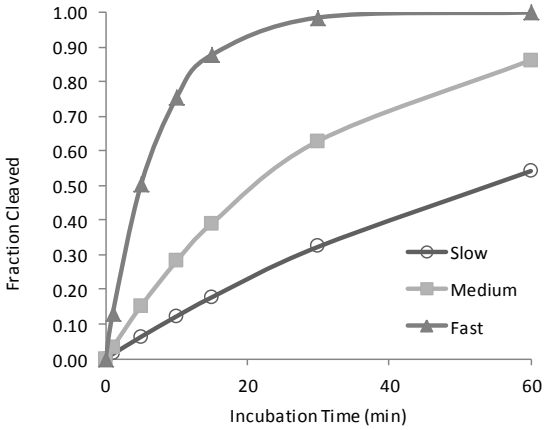
The negative selection step in the dual selection process enriches sequences that are capable of cleaving in the absence of ligand, and the cleaved sequences are recovered to be amplified through RT-PCR. We performed calculations based on a device library of 10^{15} diversity (one functional sequence 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,

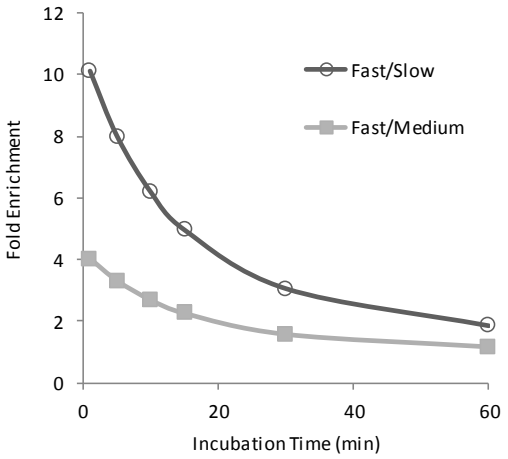
the fractions of cleaved slow, medium, and fast-cleaving populations after incubating the library for 1 minute are 0.01, 0.03, and 0.13, respectively (Supplementary Table S4.1). These fractions correspond to 4.31×10^{13}, 1.08×10^{14}, and 4.35×10^{14} cleaved members at the end of incubation, resulting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

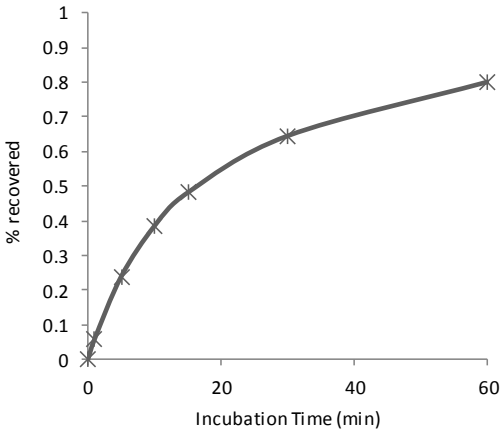


Figure 4.5 Optimization of the negative selection conditions. The model is based on an initial device library composed of 3.3×10^{15} members in each slow, medium, and fast-cleaving population (a total of 10^{16} variants) (Supplementary Table S4.1). (A) Fraction of cleaved members within each population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

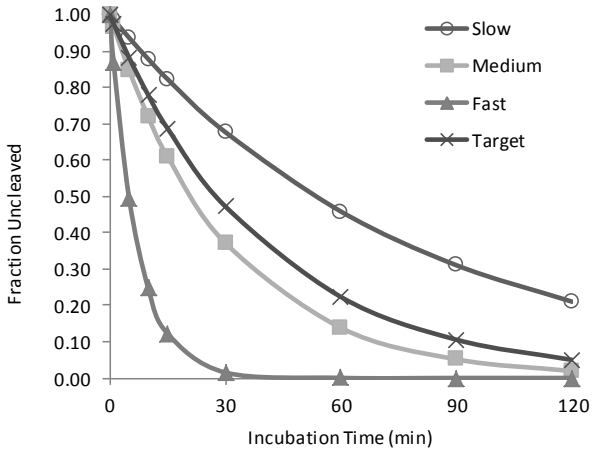
Optimization of the positive selection conditions

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

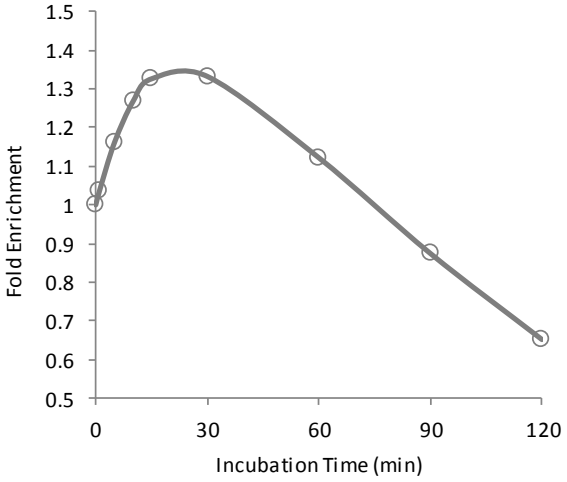


Figure 4.6 Optimization of the positive selection conditions. The model is based on the same initial device library as described in Figure 4.5 (Supplementary Table S4.2). (A) Fraction of uncleaved members within each population as a function of incubation time. The fraction of the cleaved population 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.