UNIVERSITY OF CALIFORNIA,
IRVINE

Discovery of functional RNAs through laboratory evolution

DISSERTATION

submitted in partial satisfaction of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in Biological Sciences

by

Michael Moris Abdelsayed

Dissertation Committee:
Professor Andrej Lupták, Chair
Professor Paul Gershon
Professor Klemens Hertel

2016

DEDICATION

To

My parents, brother, and wife.

TABLE OF CONTENTS

LIST OF FIGURES

ACKNOWLEDGMENTS

I would like thank my committee chair, Andrej Lupták, for giving me the opportunity to join the lab and for all of his support. Andrej helped me to grow as a scientist by giving me the flexibility and options to develop and pursue my ideas, often against his better judgement. The guidance and freedom provided by Andrej further encouraged my enthusiasm and appreciation of science.

I would like to thank my committee members, Professor Paul Gershon and Professor Klemens Hertel, for their guidance and support. I am very grateful for the time and experience I gained in the Hertel lab during my rotation.

I would like to thank Professor Robert Spitale for his collaboration and work presented in chapter 4 of this thesis.

All the members of the Lupták lab throughout the years have played a role in how I view and conduct science. I would like to thank them for making the research experience fun and engaging, but more importantly I have to thank them on updating me on when my gels were done running.

I would like to thank my parents, Moris and Bernarda Abdelsayed, for all their words of encouragement. Thanks for leading by example and always supporting my interests. Thanks to my brother, Jesse Abdelsayed, for always having my back.

I would like to acknowledge my wife Kaitlin Hovanec for keeping me sane throughout my time at graduate school. Thanks for the patience, encouragement, and cupcakes. Finally I would to thank all my pets at home: Ollie, Ginger, and Pepper. Ollie, it's time to sit in the sun.

CURRICULUM VITAE

Michael Abdelsayed

EDUCATION

2010-2016 University of California, Irvine
Department of Molecular Biology and Biochemistry Graduate Program

2008 B.S. in Biochemistry, California State University, Channel Islands

RESEARCH EXPERIENCE

January 2011 – present
University of California, Irvine
Advisor: Andrej Lupták, Ph.D. Department of Pharmaceutical Sciences, Department of Chemistry, Department of Molecular Biology and Biochemistry

September – December 2010
University of California, Irvine
Advisor: Klemens Hertel, Ph.D. UCI School of Medicine Department of Microbiology & Molecular Genetics

August 2009 – May 2010
California State University, Channel Islands
Advisor: Blake Gillespie, Ph.D. Department of Chemistry

TEACHING EXPERIENCE

Teaching Assistant University of California, Irvine
Department of Molecular Biology & Biochemistry
- Molecular Biology M116L lab
- Biochemistry 98 discussion
- Microbiology 188L lab

Department of Pharmaceutical Sciences
- Topics in Pharm SCI
- Biopharmaceutics & Nanomedicine 174L lab
- Molecular Pharmacology 2 discussion
- Medicinal Chemistry 177L lab

Outreach Teacher Santa Ana High School
- Biology aid (2012-2013)

AWARDS AND HONORS

2012-2013 NSF GK-12 Fellowship (NSF GrantDGE-0638751)
2016 Ayala School of Biological Sciences Teaching Award

PUBLICATIONS AND PRESENTATIONS

Abdelsayed M, Ho B, Vu M, Spitale RC, Lupták A. High-throughput aptamer discovery through Apta-Seq and its application to genomic SELEX of human ATP aptamers (submitted).

Advancement to Ph.D. candidacy examination

2013 *In vitro* selections for a lanthanide aptamer. UC Irvine, Department of Molecular Biology & Biochemistry
Committee members: Andrej Lupták, Paul Gershon, Thomas Poulos, Michael Mulligan, Klemens Hertel

Posters

2013 *In vitro* selections for a lanthanide aptamer.
Irvine, Department of Molecular Biology & Biochemistry, department retreat, Lake Arrowhead, CA

2013 *In vitro* selections for a lanthanide aptamer. AAAS Student Poster Competition, Boston, MA.

ABSTRACT OF THE DISSERTATION

Discovery of functional RNAs through laboratory evolution

By

Michael Moris Abdelsayed

Doctor of Philosophy in Biological Sciences

University of California, Irvine, 2016

Professor Andrej Lupták, Chair

RNA is capable of storing genetic information and facilitating catalysis, two fundamental requirements of life. Present day cells use DNA as the main receptacle of genetic information and proteins to enable a majority of functions in the cell. RNA serves as the intermediate between DNA and proteins and possesses the capabilities of both. The vast functionality of RNA has led to the hypothesis that RNA predates both DNA and proteins as one of the earliest molecules necessary and responsible for the evolution of cells. RNA is involved in many important biological processes required for gene expression including mRNA splicing and translation which ultimately lead to protein production. The extent to which RNA is implicated in biological processes remains open to discovery. Many well characterized RNAs capable of catalysis (ribozymes) and metabolite binding (aptamers) have been discovered throughout nature, but their biological significance has not been fully elucidated. New tools to probe and tune expression of RNA would be beneficial in uncovering the roles and capabilities of RNA *in vivo* and *in vitro*.

The scope of the work herein focuses on designing novel tools by laboratory evolution of RNA. The main objectives are to develop RNA based tools capable of tracing RNA by

luminescence, tuning expression of RNA (by a riboswitch), and improving methods for *in vitro* RNA discovery. Improving methods for RNA discovery is crucial to expanding the molecular tools available to study RNA and its biological targets. A luminescent RNA based tracer will provide direct analytic feedback of expression as well as a method to follow the localization of RNAs of interest. Riboswitches are RNA-based expression systems that alter expression based on binding to a target; a riboswitch can be coupled with an RNA to tune and study the expression of the target RNA. These novel RNA tools will aid in uncovering the informational and functional significance of the RNA they are coupled with, providing new methods to discover the biological roles of RNA. Such advancements will result in not only a better understanding of RNA biology *in vivo,* but also produce important insights both in elucidating the possible roles of RNA in life, and in employing RNA in a multitude of reactions *in vitro*.

Chapter 1

Evolution of functional RNA and their applications as molecular probes

1.1 Introduction

RNA plays a crucial role in the central dogma of biology, serving as the liaison between DNA and proteins. The importance of RNA expands far beyond its role as an information carrier; it aids in many crucial cellular processes ranging from gene expression to regulating cellular development. The functionality of RNA molecules can largely be attributed to their abilities to form unique structures. For example, single stranded RNA (ssRNA) is capable of folding into stable structures via intramolecular base-pairing and intermolecular (RNA-protein or RNA-RNA) interactions. There are several common RNA secondary structural motifs that are well understood and characterized, such includes: stems, hairpins, internal loops, bulges, pseudoknots, and kissing loop hairpins (Fig. 1) [1, 2]. A stem is comprised of a RNA that forms a base-paired double helix. A stem-loop, or hairpin, is a RNA molecule that forms intramolecular base-pairings via complementary nucleotide sequences in a pattern of a double helix. An internal loop is a RNA motif that consists of consecutive mismatched or unpaired nucleotides that occur during intramolecular base-pairing, resulting in the formation of a loop in the middle of a double helix RNA. A bulge is a special type of internal loop motif that consist of unpaired nucleotides in only one strand of a double stranded region of the RNA molecule; thus, resulting in one or several bases that are rather stacked into the helix or residing outside of the helix[1]. A pseudoknot is comprised of two helical regions that are connected to each other by a single stranded region or loop. Pseudoknots can form a variety of structures and they have been found to help stabilize the fold and function of ribozymes, self-splicing introns, and telomerase

RNA[2]. Junctions are three or more stems that intersect with each other, and the point at which the stems interacts is known as the junction[1, 3]. Finally, a kissing loop hairpin motif occurs when loops of two different hairpins are base-paired with each other[1]. These and many other interactions can stabilize RNA conformations and can create complex global structures consisting of a combination of these motifs.

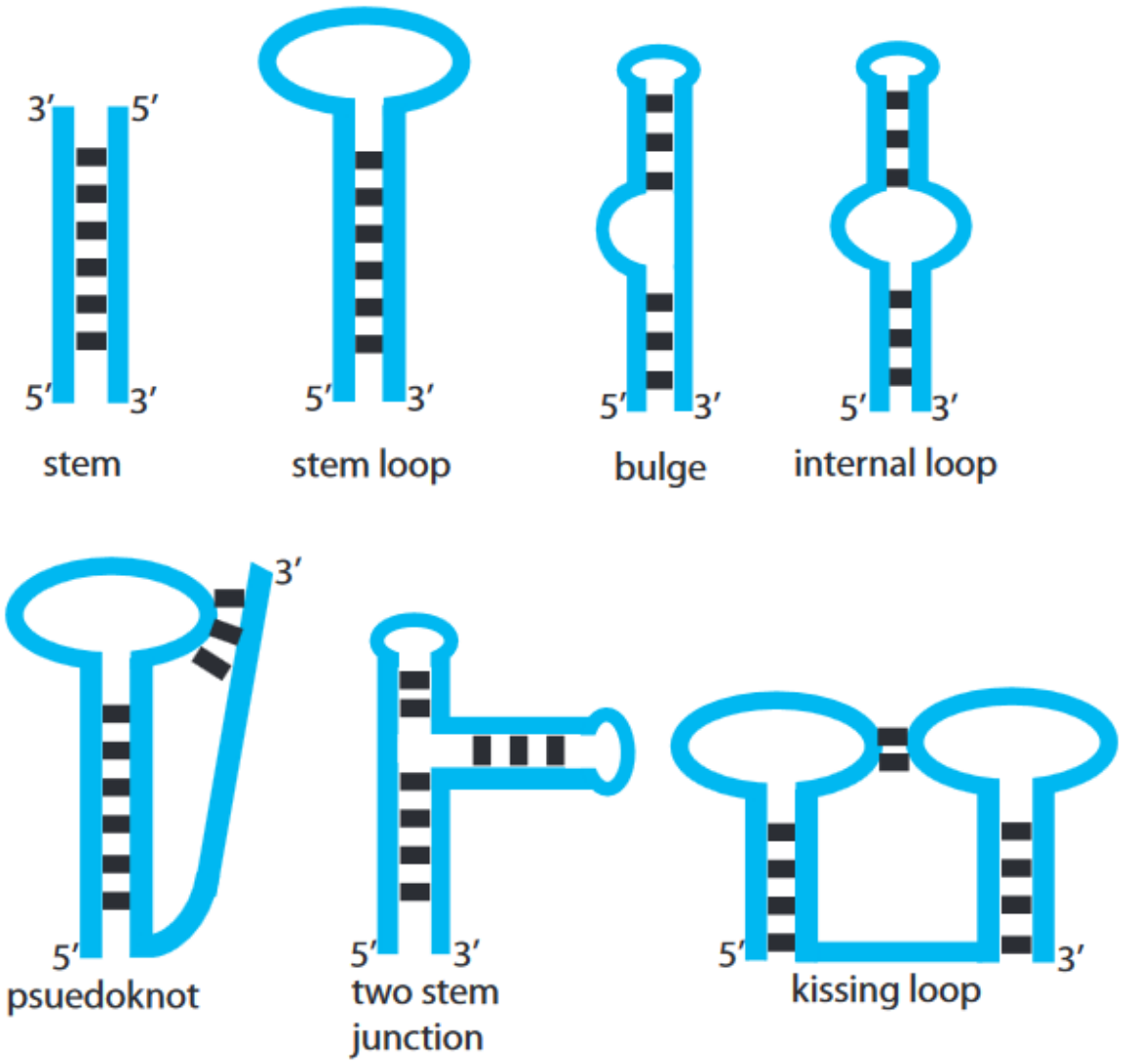


Figure 1.1 RNA structures. Several common structural motifs of RNAs that stabilize RNA and help facilitate function of RNA. RNA (in blue) can fold into a combination of motifs, stabilized by base pairing (in black).

RNA can fold inter higher order tertiary structures, the stability of the RNA is highly dependent on environmental conditions and sequence. A single nucleotide polymorphism (SNP), or a mutation in a given population, can result in a different structural arrangement of RNA structures which can lead to variation in RNA functions[4]. The ability to form a wide variety of structures has aided in the evolution of RNA in biological systems allowing RNA molecules to perform many important biological functions, such as: (1) promoting peptide bond formation in the ribosome, (2) performing the catalytic steps in splicing, (3) initiating steps necessary for translation, and (4) RNA cleavage and ligation by ribozymes[1, 5, 6]. In addition to biological importance, many laboratories have also taken advantage of the malleability or RNA to evolve novel structures and functions.

1.2 Evolution of RNA *in vitro*

The strategies in developing and engineering novel RNA functions are still rapidly evolving; although, many sophisticated functional RNAs have already been discovered using current methods. The challenge of finding an RNA molecule that is selective for binding to a target molecule or performing a specific function resides in the rational prediction of the sequence necessary to achieve the desired structure and function. To overcome this hurdle, the Szostak, Gold, and Joyce laboratories designed a strategy based on evolution and selection, by reasoning that RNA can undergo selection from a large population. Gerald Joyce developed the initial strategies to select for catalytic RNA in 1989 by mutation and amplification of RNA molecules to discover RNAs that are capable of catalyzing trans-splicing reactions[7]. Further refinement of *in vitro* selection procedures was developed in 1990 by the Szostak and Gold laboratories to select for functional RNA molecules that bind to a target molecule. The Szostak laboratory designed a RNA library with of 10^{10} sequence diversity to select for RNAs capable of binding to dyes[8]. Simultaneously, the Gold laboratory also selected RNAs that bind to

bacteriophage T4 DNA polymerase and coined the process Systematic Evolution of Ligands by Exponential Enrichment (SELEX)[9]. Both experiments developed functional RNAs that bound to desired target molecules with high specificity through serial rounds of affinity binding and amplification. These functional RNAs capable of binding target molecules are known as RNA aptamers[8]. Since those initial experiments, new selection strategies have been developed to discover novel RNA functions beyond target binding. Many of these new approaches focused on selection for catalytically competent RNA molecules and regulatory RNAs.

These selection strategies have resulted in the discovery of aptamers (RNA molecules capable of recognizing and binding to a target molecule with high specificity), ribozymes (RNA molecules with catalytic functionality)[10, 11], and riboswitches (RNAs that can modulate gene expression in response to specific ligand)[12]. Recently, a selection for a thiazole orange (TO) binding aptamer resulted in an RNA that increases fluorescence emission from TO upon RNA binding. This resulted in an aptamer-fluorophore pair that can be utilized for direct RNA detection *in vivo*[13]. The selection scheme was performed based on affinity. After several rounds, the selection stringency was increased to preferentially isolate fast on-rate and slow off-rate binders. This is an example of a classical selection scheme resulting in a highly functional RNA that can be utilized in expression studies *in vivo*.

In order to obtain complex functional RNA molecules, novel and more sophisticated selection schemes are used. For example, an RNA polymerase ribozyme capable of polymerizing and elongating RNA molecule up to 20 nucleotides was selected and engineered[14]. This functionality was obtained only after utilizing a combination of several techniques. *In vitro* compartmentalization techniques that couple genotype and phenotype with water-in-oil[15] droplets was combined with classical selection steps, resulting in the discovery of an RNA molecule capable of one of the most complex and important processes in

biology. Only elongated products were selected by annealing to streptavidin labeled oligo with the template for elongation. This selection method deviates far from classical selection schemes; however, many of the important aspects of SELEX remained (library diversity and serial rounds of evolution). As the above examples demonstrate, the key to obtaining functional RNA molecules lies in the design of the starting library and the selection scheme.

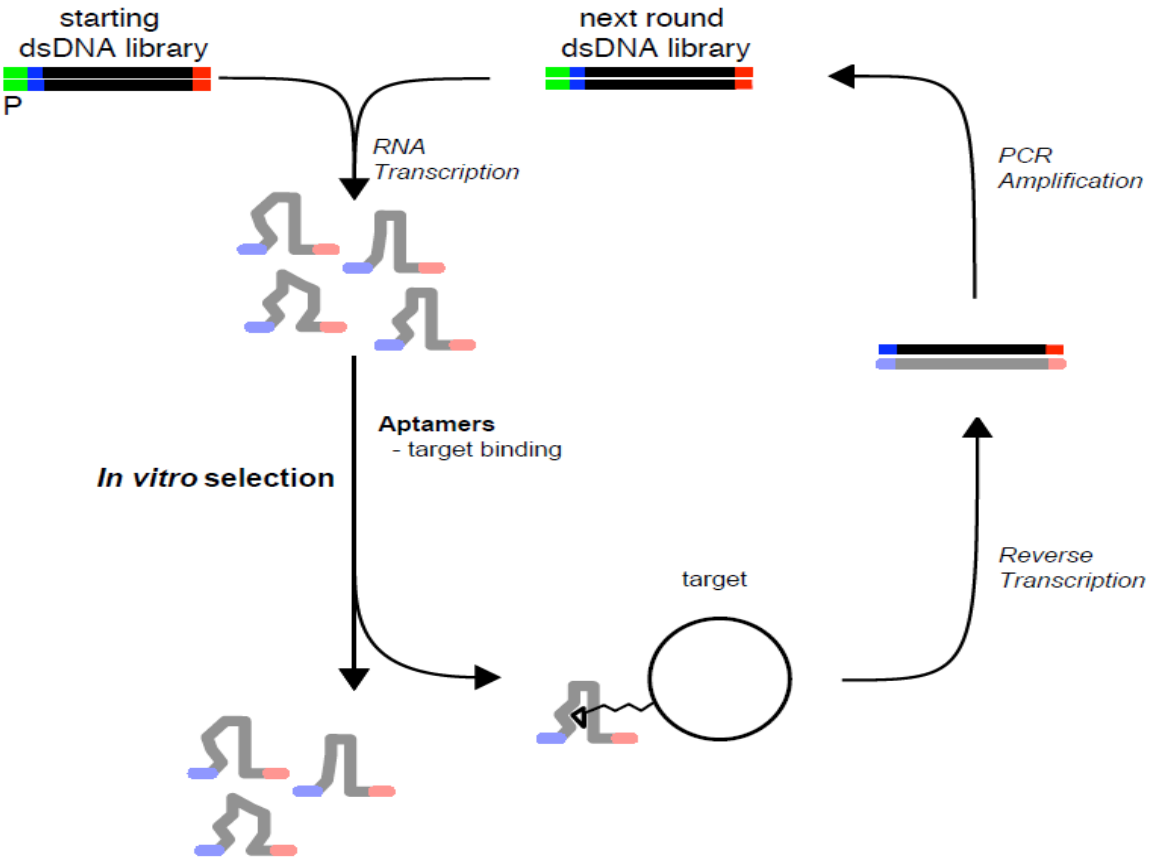


Figure 1.2 *In vitro* selection. A synthetic or genomic pool is incubated with the target, the unbound RNA is washed away, and then bound RNA is competitively eluted off. RNA is reverse transcribed to DNA followed by amplification and transcription for the next round of selection.

1.3 Importance of diversity and pool design

The starting population is a critical component in any system of selection and evolution. The initial diversity of a population is one of the main determinants of the outcome of a selection experiment. In a given *in vitro* selection experiment, the starting library can comprise of a synthetic or genomic pool. A synthetic pool is useful for engineering novel functions or predetermined aptamer domains and has a vast potential for diversity in design. The theoretical range and diversity of such library is 4^N, where N represents the sequence length of the random region. Although a pool designed with 10^{20} diversity (random region composed of 35 nucleotides, which is well within the range of a pool designs) is achievable, pools are generally limited to an upper limit of 10^{16} due to practical restraints. These limitations include the cost of DNA synthesis and the reagents needed for transcription to achieve such a large reaction. Therefore, most starting pools are restricted to a diversity of 10^{14} to 10^{16} molecules[16].

A genomic pool is necessary for the selection-based discovery of regulatory RNA elements *in vivo*. The diversity of a genomic pool is constrained to the diversity of the genome itself. Genomic pools can consist of a single-species genome or a metagenome, which comprises many species genomes combined together[17]. A human genomic pool was created with a diversity of 10^9 and used for several selections, resulting in the discovery of first human ribozymes and ATP aptamers[18]. Although RNA from genomic pools may have biological implications, diversity of the pools is constrained by the size of the genome. Diversity and design of the starting library plays an important role in the potential successes in a given selection.

Important considerations in designing a pool include: (1) adding conserved, primer-binding regions used for serial amplification, (2) the overall length of pool, (3) the length of

random region, (4) the inclusion of structural features (secondary structures), and (5) any covalent modifications to the DNA or RNA pool. Primer design is crucial to the success of a selection, since the primer-binding sequences are the major conserved elements in a given pool. Common considerations for primer design include: (1) lack of mutual complementarity of primers to enhance or prevent the primers from interacting, (2) inclusion of adapters and primers for high throughput sequencing, (3) promoter sequences for transcription, and (4) incorporating structural motifs that influence the stability of the pool[19].

In a selection aimed to obtain catalytic RNAs, the starting library may contain built-in structural motifs, such as stem loops, to help increase the probabilities of catalytic or binding activities. This is done because most known aptamers contain structural motifs such as stem loops that either directly bind to target molecules or assist in stabilizing the binding site. For example, in an experiment done by Szostak and Davis, a pool was designed to contain an equal representation of two sub-pools. One pool was composed of completely random regions, while the other contained a stem loop flanked by random regions. From this, they found most of the high-affinity aptamers obtained from the pool contained the hairpin motif[20]. This demonstrated that the inclusion of a hairpin enhanced the affinity for GTP compared to an entirely random pool.

A selection that aims to improve a known function or to discover a new function can often be achieved via partial randomization of an existing aptamer or ribozyme. For examples, novel ribozymes were discovered from partially randomized pool of spliceosome U2 and U6, small nuclear RNA that acts catalytically in splicing. Interestingly none of the newly selected ribozymes achieved the initial function of the original ribozyme. However, several new functions were isolated from selection, such as ribozymes capable of RNA-processing reactions

(transesterification)[21]. These newly discovered ribozymes illustrated the potential of using well-characterized functional RNAs as starting libraries to achieve catalytic function.

1.4 Selection schemes

The original selections performed by the Szostak and Gold labs were affinity selections done using immobilized surfaces. Ellington and Szostak selected for dyes that became chemically coupled to agarose beads[8], while Tuerk and Gold immobilized T4 DNA polymerase (gp43) on nitrocellulose filters for their selection[9]. In both selections, non-binding RNA molecules were washed away, followed by elution and collection of the target-bound RNAs. Generally, the wash conditions mimic the environments in which the RNA function is desired. Since many selection schemes are designed to discover biologically relevant RNAs, the buffers used typically mimic physiological conditions. Often the target-bound RNA molecules are competitively eluted using increasing concentrations of free-flowing target molecules. The number of washes and elutions performed in a given selection experiment vary. Typically, washes are performed until the amount of RNA washing off is minimal or at background. This is done to ensure that any nonspecific RNA molecules are discarded. Elution steps are performed to collect potential aptamers. Any residual RNA still bound to the immobilized targets can either be denatured off (with urea) or used directly in reverse transcription. The elutions can either be pooled for the next round of selection or used in separate parallel selections. From that, the RNAs are reverse transcribed, amplified, and transcribed for the next round of selection. Over multiple rounds, the affinity for a given target is expected to increase for each round of selection.

Using this classical selection scheme, RNA aptamers for nucleotides[20, 22], fluorophores[13], amino acids[23], cocaine[24], and secondary metabolites[25] have been generated. The plasticity of the selection work flow has also allowed for the generation of aptamers for more complex

targets that cannot be immobilized to surfaces. This is because a selection scheme can be highly malleable and can often be tailored towards specific applications, thus, making it flexible to select for almost any target. For example, Nakamura developed a method to isolate aptamers that recognize Growth Factor-B Type III Receptors on cells by performing a direct selection using hamster ovary cells with overexpressed cell surface receptors[26]. This allowed for the selected aptamers to be directly applied to the target. Based on the aforementioned, a bacterial selection scheme was designed to isolate aptamer molecules that target clinically relevant pathogens; and from the selection, aptamers with nanomolar affinity specific for *S. pyongenes* were isolated[27]. These specific selection protocols are known as cell-based selections. These strategies have resulted in aptamers that are capable of detecting tumor biomarkers, such as tenascin C, a glycoprotein that is overexpressed in leukemia cells[28]. In addition to directed selection of a cancer targer, cell-based selections have been used to elucidate novel cancer biomarkers. Luekemia cells lines were used as the target of selection in order to produce aptamers for highly expressed biomarkers, aptamers for protein Tyrosine Kinase 7 were discovered, and revealed the kinase as a potential biomarker[29]. The development of cell-based selection methods has diversified the potential targets of aptamer molecules facilitating their use and relevance as clinical diagnostic and therapeutic tools.

Selections of functional RNAs, such as ribozymes and riboswitches, often require additional steps to be implemented in the selection scheme to screen for RNAs that demonstrate high target affinity as well as the desired functionality. Ribozyme selections were designed to improve existing ribozymes and to discover novel ones. Szostak and Bartel selected ribozymes that are capable of ligating a substrate oligonucleotide to their own 5'-termini. A novel scheme was implemented specifically for the discovery of ligase functionality. This was done by partially hybridizing a fixed substrate to a pool. A successful ligation results

in a nucleophilic attack of the 3′-hydroxyl group on substrate RNA to an adjacent 5′-triphosphate on the desired ribozyme sequence in the pool.[10] To ensure that only the desired function is isolated, only the ligated sequences could be amplified for subsequent rounds of selection. Other impressive selection schemes have yield ribozymes equipped with polymerase activities[14], kinase activities[30], DNA cleavage[31], self-cleaving[18], and various functions that aid in splicing[21]. These selections have demonstrated that RNAs are evolvable to adapt to vast selection conditions.

Another class of selectable functional RNAs are riboswitches. Riboswitches are generally composed of an aptamer domain coupled to an expression platform. The events of aptamer-ligand interaction will lead to disruption of downstream structural motif of the expression platform; thus, tuning expression on or off[32]. Selection schemes for riboswitches must achieve both ligand affinity and regulation of downstream or upstream expression. Nomura and Yokobayashi designed a dual selection strategy for the discovery of riboswitches. Their protocol required the riboswitch pool to be cloned in a plasmid upstream of a tetracycline resistance gene. Tetracycline resistance is used as a selection marker of "ON" (high expression) and "OFF" (low expression) selections. Termination of expression by the riboswitch upstream will result in "OFF" or low expression. To isolate high expression switches in the presence of a ligand, cells are gown in the presence of tetracycline ligand for the "ON" selection. To eliminate any ligand-independent tetracycline resistance expression activities, cells are grown in the presence of $NiCl_2$, which is toxic to cells expressing tetracycline resistance. Utilizing this novel strategy, a thiamine pyrophosphate riboswitch with 58-fold activation of expression was identified[33]. Another strategy in riboswitch discovery utilizes flow cytometry as a means of isolating functional RNAs. A riboswitch pool is engineered upstream of a fluorescent protein reporter and the selection occurs by cell sorting according to fluorescence intensity[34]. These

strategies can be used to improve current riboswitches or select for novel synthetic riboswitches for a specific target.

1.5 Aptamer probes

In vitro selection techniques have evolved to facilitate the specific needs for the various functions of aptamers. These selection techniques have engendered many aptamers that serve as biological probes or tools that help further fundamental research. Specific binding and target recognition make aptamers ideal molecular probes in biological studies. Aptamers can easily and readily be modified and engineered with fluorophores or markers. These chemical modifications can help enhance or diminish feedback upon aptamer-target interactions. For example, an RNA beacon was developed via the chemical attachments of fluorophore-quencher pair at complementary ends of the RNA molecule. The complementary ends formed a stem structure that put the fluorophore and quencher pair in close proximity, which led to quenching of the fluorophore. Upon binding to the target molecule, the complementary ends were released and fluorescence signal could be detected[35]. This concept was applied to a thrombin aptamer, where a fluorophore-quencher pair was engineered into the aptamer sequence. In the absence of the target molecule (thrombin) the fluorophore was quenched and the presence of thrombin leads to the release of the fluorophore-quencher pair, resulting in an enhanced fluorescence emission[36], resulting in the rapid detection of thrombin by aptamer binding. This simple and elegant design showed how aptamers could be used as molecular probes for protein recognition and detection.

Aptamer "beacons" have also been developed using complementary antisense oligonucleotides coupled with an aptamer sequence. Structure-switching signaling aptamers were developed by engineering a fluorophore and quencher pair that are on the antisense sequences of the template sequence. A DNA-based ATP aptamer was used to demonstrate this

assay. In the absence of ATP, both fluorophore and quencher oligonucleotides were hybridized to the aptamer sequence resulting in fluorescence quenching. With ATP present, it led to a conformational change in the aptamer and prevented the antisense quencher oligonucleotide from base-pairing with the aptamer[37]. This resulted in fluorescence signal to detect the presence of ATP. The concept of aptamer beacons makes it possible to develop any established aptamer and further engineer it to become a fluorescent probe for assay development of the aptamers target.

The development of aptamers capable of fluorescence feedback has broadened the scope of aptamer utilities and applications in research. An enzyme-linked oligonucleotide assay (ELONA) was developed to monitor the levels of Vesicular Endothelial Growth Factor (VEGF) using fluorescent-labeled aptamer that binds to VEGF[38]. ELONA is similar in concept to an enzyme-linked immunosorbent assay (ELISA), which uses antibodies to detect analytes, except the ELONA assay utilizes aptamer instead. The VEGF aptamer was capable of detecting serum VEGF at picomolar concentrations demonstrating the effectiveness of using aptamers as analytical tools[38]. Signaling aptamers have been designed by incorporating fluorescence labels within the body of the aptamer to detect conformational changes upon binding to a ligand. Ellington et al. used the ATP aptamer with known binding conformations to create an aptamer sensor for ATP concentration by incorporating several fluorophore nucleotide analogs to the known binding motif of the ATP aptamer. This resulted in an increased fluorescence signal in response to high ATP concentrations due the combination of the fluorophore nucleotides and the conformational rearrangement of the aptamer sequence[39]. Although the modified aptamer provided increase fluorescence feedback in response to target, this modification was impractical for *in vivo* usage due to the design requirements.

RNA-fluorophore complexes have the potential to be utilized as reporters for RNA

expression analogous to that of fluorescent proteins, such as green fluorescent protein (GFP), that serves as a reporter for protein expression and localization. Developing such a complex for *in vivo* usage will require a system that is dependent on the transcription of the aptamer and does not require any covalent modifications for signal production. Recent advances have resulted in RNA-fluorophore complexes that can be used as fluorescent reporters in living cells. The first major aptamer-fluorophore complex was discovered though a selection for 3,5-dimethoxy-4-hydroxybenzylidene imidazolinone (DMHBI) and 3,5-difluoro-4-hydroxybenzylidene imidazolinone (DHFBI), derivatives of the 4-hydroxybenzlidene imidazolinone (HBI) core responsible for the fluorescence in GFP. The resulting aptamer, termed Spinach, exhibits a strong green fluorescence when bound to DMHBI and DFHBI. The expression of the aptamer and addition of DMHBI in live cells enabled expression and localization studies[40]. The aptamer could be imaged and tracked, leaving the nucleus and migrating to the cytosol of the cell. This demonstrated that the Spinach aptamer-fluorophore complex could provide dynamic RNA tracking in live cells as well as quantification of expression.

Enhancement of emission from HBI is dependent on the stability of the molecule, suppression of movement is necessary for fluorescence[40]. Free DMHBI and DFHBI are not fluorescent, binding of the spinach aptamer results in restriction in the degrees of freedom for DMHBI and DFHBI, resulting in enhanced emission[41]. Crystal structure of the spinach aptamer bound to DFHBI illustrates the mechanism of binding and basis of fluorescence. Fluorescence is enhanced by the aptamer due to immobilization of DFHBI between a base triple, a G-quadraplex and an unpaired G[41] (Fig 1.3). The original selection for Spinach resulted in many RNAs containing a G-quadraplex that are capable of fluorescence when bound to DFHBI, the crystal structure revealed the function of the G-quadraplex. DFHBI is sandwiched between two

Gs of a G-quartet indicating that the quartet plays a direct role in stabilizing the ligand[41]. Elucidation of the crystal structure guided the design of small DFHBI aptamers (baby spinach) and revealed the binding mechanism of the original aptamer.

Aptamer-fluorophore complexes for molecular labeling and dynamic tracking of cellular activities should exhibit high affinity and strongly enhanced signal emission. An aptamer for thiazole orange (TO) was identified with nanomolar affinity that enhanced the fluorescence emission by a 1000-fold. Strong binding affinity also makes it possible to tag and track low abundance or expressed RNA molecules in reporter studies[13]. Due to the nature of the strong fluorescence signal of the aptamer-fluorophore complex, the aptamer (termed Mango) was used in both *C. elegans* and bacterial cells for imaging and detection of RNA expression. All in all, these aptamer-fluorophore complexes proved to have great potential in tagging and monitoring of RNA expression in dynamic cellular situations. The advances in these aptamer-fluorophore complexes can broaden the scope in the studies of the ever changing cellular RNA activities.

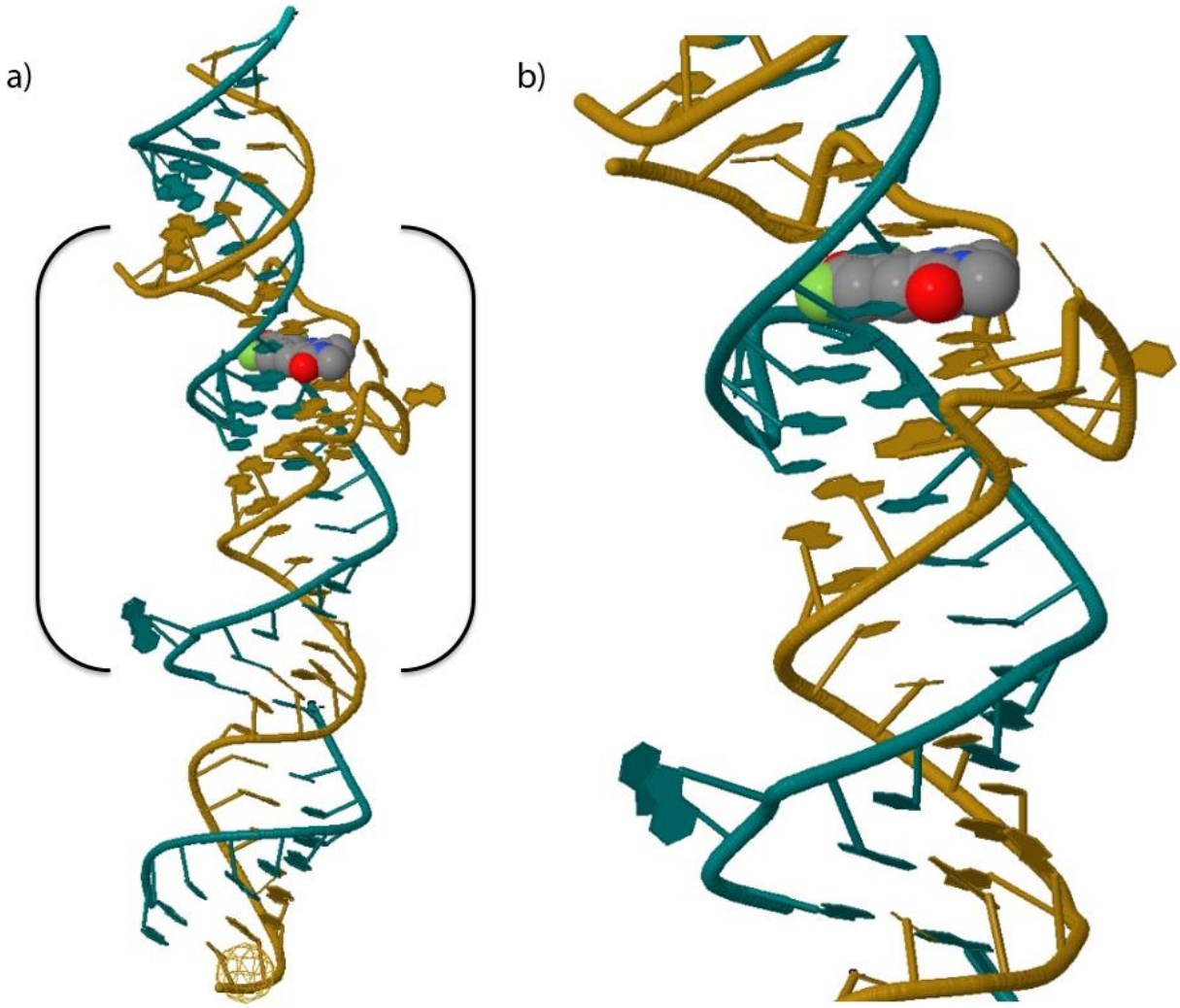


Figure 1.3 Structure of Spinach aptamer with DFHBI. (a) Crystal structure of Spinach bound to DFHBI, binding site indicated by brackets. (b) Close up of the binding site of DFHBI (carbon in grey, oxygen in red, nitrogen blue), DFHBI is stacked between to Gs of a G quartet.

1.6 Photoactive RNA

Optical-based approaches have been applied to tune and regulate genetics in biological systems. Optogenetics is the photoregulation of genetics by optical methods to change the state of a biological process[42]. Often light is employed as the activator or inhibitor of the biological process, light is used to stimulate a biologically relevant light-responsive target. Influence of biological processes by light has the advantage of: 1) being flexible to a wide range of targets by use of appropriate wavelength, 2) quickly delivered to its target, and 3) the ability to tune activation or inhibition by the intensity of light[42, 43]. Optogenetic procedures have been widely adapted to neuroscience research by influencing the membrane potential of neuronal cells in behavioral studies[43]. Optogenetic tools often target ion channels to change polarization state of proteins that are sensitive to light[43]. Engineering new targets and procedures for optogenetic modification will facilitate broader use of optogenetics in other fields of science.

Light can be employed to regulate the state of photo-responsive molecules. Application of light allows for quick tuning of its target, allowing studies that require high temporal control. Attempts to control the state of RNA was achieved by blocking or caging RNA. This keeps RNA in one fixed state until the RNA is uncaged and available to change its state, usually caging is done to impede structure of the RNA. The hammerhead ribozyme was used as a model to demonstrate the ability to cage RNA molecules[44]. The 2′ hydroxyl of RNA often acts as a mediator of catalysis. The 2′ OH acts as a nucleophile for catalytic cleavage in hammerhead ribozymes. A photo-dissociable group, 2′-nitro-benzyladenosine, was coupled to the 2′ OH that acts as a nucleophile for cleavage of the ribozyme, rendering the ribozyme incapable of self-cleavage[44]. Upon photolysis of, 2′-nitro-benzyladenosine, the ribozyme becomes uncaged and capable of cleavage. Caging of the ribozyme proved useful in measuring kinetics for the

ribozyme, as caging can aid in controlling the activation and initiation of cleavage. Similar approaches can be successfully applied to many functional RNA studies; however, they require covalent modification of the substrate or target, which changes the composition the modified RNA and is not easily applicable to *in vivo* studies.

Covalent modification of RNA is another strategy employed to engineer photoresponsive RNA. Photoresponsive nucleotides can be incorporated into RNA to create photoreversible aptamer probes. This strategy uses nucleotides modified with a photoswitch in order to change the conformation of the aptamer to bind or release its target. Azobenzene, a well-characterized photoswitch of stilbene molecules capable of reversible *cis* to *trans* isomerization was coupled to ATP to create a photoresponsive ATP and incorporated into a random pool for selection of hemin as the target molecule. Aptamers from the selection were capable of binding to hemin, and upon irradiation with light, the azobenzene isomerized releasing the aptamer from the hemin[45].

Silbenes have been incorporated into other molecules to take advantage of their optical properties. Azobenzene has also been incorporated into peptides to create photoresponsive peptides. Aptamers were selected for the photoresponsive peptides, isomerization of the peptide from *trans* to *cis* resulted in release of the aptamer, and the binding could be reversed by photoisomerization from *cis* to *trans*[46]. Molecules capable of photoisomerization are an excellent system for modulating structure and binding of a target RNA; however, this technology is not easily applicable for *in vivo* applications due to the necessity of modified nucleotides or peptides. An ideal photoreversible RNA system would be an aptamer that binds to a photoreversible probe directly, thereby controlling photoactivity based on the expression of the aptamer.

1.8 Conclusions

Methods to discover functional RNA have rapidly improved and evolved since the advent of the original *in vitro* selection experiments. Selection schemes have resulted in the discovery of a wide array of functional RNAs including many aptamers, ribozymes, and riboswitches. The RNAs from these selections have been incorporated into many assays both *in vivo* and *in vitro* for a wide variety of applications such as analyte recognition, expression tuning, and fluorescent feedback of expression or function. The common themes in these discoveries are the ability of RNA to bind to a specific target or to achieve a specific catalytic activity. The makeup of the pool and application of selective pressure for the target (either a bound molecule or catalytic function) are the main considerations when designing a selection scheme. The availability of many successful selections makes it possible to improve upon current selection schemes or implement aspects of those selections to a new scheme to discover aptamers for a wide variety of applications.

Molecular probes are extensively used in research for labeling, detection, and feedback (or feedforward) regulation purposes. Many probes are limited in their scope of application as they cannot be easily evolved for recognition or function. Combining probes with aptamer technology allows for selectivity of both target and function of the probe. Several RNA probes have been described herein that demonstrate the ability of RNA that have been evolved by selection to be used in assays for analyte detection, expression feedback, photoresponsive switching, and cellular imaging. The discovery of fluorescent protein reporters such as GFP has revolutionized biomedical research and are continually used in fundamental research techniques that require feedback of expression or labeling of an important target. Selections of RNA have resulted in probes and tracers for many assays and applications as described herein. The ability to evolve RNA will continue to result in more RNA tracers and probes, as the potential for their applications are only bound by the selection scheme used to produce them.

1. Nowakowski, J. & Tinoco, I. RNA Structure and Stability. *Seminars in Virology* **8**, 153-165 (1997).
2. Staple, D.W. & Butcher, S.E. Pseudoknots: RNA Structures with Diverse Functions. *PLoS Biol* **3**, e213 (2005).
3. Stoddard, C.D., Gilbert, S.D. & Batey, R.T. Ligand-dependent folding of the three-way junction in the purine riboswitch. *RNA* **14**, 675-684 (2008).
4. Ritz, J., Martin, J.S. & Laederach, A. Evaluating our ability to predict the structural disruption of RNA by SNPs. *BMC Genomics* **13**, S6-S6 (2012).
5. Jijakli, K. et al. The in vitro selection world. *Methods*.
6. Sharp, P.A. The Centrality of RNA. *Cell* **136**, 577-580 (2009).
7. Joyce, G.F. Amplification, mutation and selection of catalytic RNA. *Gene* **82**, 83-87 (1989).
8. Ellington, A.D. & Szostak, J.W. In vitro selection of RNA molecules that bind specific ligands. *Nature* **346**, 818-822 (1990).
9. Tuerk, C. & Gold, L. Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science* **249**, 505-510 (1990).
10. Bartel, D. & Szostak, J. Isolation of new ribozymes from a large pool of random sequences [see comment]. *Science* **261**, 1411-1418 (1993).
11. Martini, L., Ellington, A.D. & Mansy, S.S. An in vitro selection for small molecule induced switching RNA molecules. *Methods*.
12. Wachsmuth, M., Findeiß, S., Weissheimer, N., Stadler, P.F. & Mörl, M. De novo design of a synthetic riboswitch that regulates transcription termination. *Nucleic Acids Research* **41**, 2541-2551 (2013).
13. Dolgosheina, E.V. et al. RNA Mango Aptamer-Fluorophore: A Bright, High-Affinity Complex for RNA Labeling and Tracking. *ACS Chemical Biology* **9**, 2412-2420 (2014).
14. Zaher, H.S. & Unrau, P.J. Selection of an improved RNA polymerase ribozyme with superior extension and fidelity. *RNA* **13**, 1017-1026 (2007).
15. Levy, M., Griswold, K.E. & Ellington, A.D. Direct selection of trans-acting ligase ribozymes by in vitro compartmentalization. *RNA* **11**, 1555-1562 (2005).
16. Pobanz, K. & Lupták, A. Improving the odds: Influence of starting pools on in vitro selection outcomes. *Methods*.
17. Ho, B., Polanco, J., Jimenez, R. & Lupták, A. in Methods in Enzymology, Vol. Volume 549. (ed. H.B.-A. Donald) 29-46 (Academic Press, 2014).
18. Salehi-Ashtiani, K., Lupták, A., Litovchick, A. & Szostak, J.W. A Genomewide Search for Ribozymes Reveals an HDV-Like Sequence in the Human CPEB3 Gene. *Science* **313**, 1788-1792 (2006).
19. LEGIEWICZ, M., LOZUPONE, C., KNIGHT, R. & YARUS, M. Size, constant sequences, and optimal selection. *RNA* **11**, 1701-1709 (2005).
20. Davis, J.H. & Szostak, J.W. Isolation of high-affinity GTP aptamers from partially structured RNA libraries. *Proceedings of the National Academy of Sciences* **99**, 11616-11621 (2002).
21. Tuschl, T., Sharp, P.A. & Bartel, D.P. Selection in vitro of novel ribozymes from a partially randomized U2 and U6 snRNA library. *The EMBO Journal* **17**, 2637-2650 (1998).
22. Vu, Michael M.K. et al. Convergent Evolution of Adenosine Aptamers Spanning Bacterial, Human, and Random Sequences Revealed by Structure-Based Bioinformatics and Genomic SELEX. *Chemistry & Biology* **19**, 1247-1254 (2012).
23. Ames, T.D. & Breaker, R.R. Bacterial aptamers that selectively bind glutamine. *RNA Biology* **8**, 82-89 (2011).
24. Stojanovic, M.N., de Prada, P. & Landry, D.W. Aptamer-Based Folding Fluorescent Sensor for Cocaine. *Journal of the American Chemical Society* **123**, 4928-4931 (2001).
25. Hwang, C. & Carothers, J.M. Label-free selection of RNA aptamers for metabolic engineering. *Methods*.
26. Ohuchi, S.P., Ohtsu, T. & Nakamura, Y. Selection of RNA aptamers against recombinant transforming growth factor-β type III receptor displayed on cell surface. *Biochimie* **88**, 897-904 (2006).
27. Hamula, C.L.A. et al. The Effects of SELEX Conditions on the Resultant Aptamer Pools in the Selection of Aptamers Binding to Bacterial Cells. *Journal of Molecular Evolution* **81**, 194-209 (2015).
28. Hicke, B.J. et al. Tumor Targeting by an Aptamer. *Journal of Nuclear Medicine* **47**, 668-678 (2006).

29. Shangguan, D. et al. Cell-Specific Aptamer Probes for Membrane Protein Elucidation in Cancer Cells. *Journal of Proteome Research* **7**, 2133-2139 (2008).
30. Curtis, E.A. & Bartel, D.P. Synthetic shuffling and in vitro selection reveal the rugged adaptive fitness landscape of a kinase ribozyme. *RNA* **19**, 1116-1128 (2013).
31. Beaudry, A. & Joyce, G. Directed evolution of an RNA enzyme. *Science* **257**, 635-641 (1992).
32. Breaker, R.R. Riboswitches and the RNA World. *Cold Spring Harbor Perspectives in Biology* **4** (2012).
33. Nomura, Y. & Yokobayashi, Y. in Artificial Riboswitches: Methods and Protocols. (ed. A. Ogawa) 131-140 (Humana Press, Totowa, NJ; 2014).
34. Lynch, S.A. & Gallivan, J.P. A flow cytometry-based screen for synthetic riboswitches. *Nucleic Acids Research* **37**, 184-192 (2009).
35. Tyagi, S. & Kramer, F.R. Molecular Beacons: Probes that Fluoresce upon Hybridization. *Nat Biotech* **14**, 303-308 (1996).
36. Pakkila, H., Blom, S., Kopra, K. & Soukka, T. Aptamer-directed lanthanide chelate self-assembly for rapid thrombin detection. *Analyst* **138**, 5107-5112 (2013).
37. Nutiu, R. & Li, Y. Structure-Switching Signaling Aptamers. *Journal of the American Chemical Society* **125**, 4771-4778 (2003).
38. Drolet, D.W., Moon-McDermott, L. & Romig, T.S. An enzyme-linked oligonucleotide assay. *Nat Biotech* **14**, 1021-1025 (1996).
39. Jhaveri, S., Rajendran, M. & Ellington, A.D. In vitro selection of signaling aptamers. *Nat Biotech* **18**, 1293-1297 (2000).
40. Paige, J.S., Wu, K.Y. & Jaffrey, S.R. RNA Mimics of Green Fluorescent Protein. *Science* **333**, 642-646 (2011).
41. Warner, K.D. et al. Structural basis for activity of highly efficient RNA mimics of green fluorescent protein. *Nat Struct Mol Biol* **21**, 658-663 (2014).
42. Deisseroth, K. Optogenetics. *Nat Meth* **8**, 26-29 (2011).
43. Fenno, L., Yizhar, O. & Deisseroth, K. The Development and Application of Optogenetics. *Annual Review of Neuroscience* **34**, 389-412 (2011).
44. Chaulk, S.G. & MacMillan, A.M. Caged RNA: Photo-control of a ribozyme reaction. *Nucleic Acids Research* **26**, 3173-3178 (1998).
45. Liu, M., Jinmei, H., Abe, H. & Ito, Y. In vitro selection of a photoresponsive RNA aptamer to hemin. *Bioorganic & Medicinal Chemistry Letters* **20**, 2964-2967 (2010).
46. Hayashi, G., Hagihara, M. & Nakatani, K. RNA aptamers that reversibly bind to photoresponsive peptide. *Nucleic Acids Symposium Series* **52**, 703-704 (2008).

Chapter 2

In vitro selection for a lanthanide aptamer

2.1 Specific aim

Lanthanides are f-block transition metals capable of luminescence[1]. Lanthanides luminesce poorly on their own, emission is dependent on an energy transfer from an antenna molecule[1]. Emission of is increased when a luminescent capable lanthanide is bound to an antenna molecule with the ability to transfer energy to the lanthanide. Our goal is to develop a luminescent tracer of RNA by means of lanthanide luminescence. Selection of an aptamer for a lanthanide complex can be employed to achieve a lanthanide-dependent luminescent tracer of RNA. The aptamer can be use as the source of energy transfer in order for luminescence enhancement to occur only in the presence of the aptamer (Fig 2.1).

2.2 Luminescent lanthanides

Molecular probes are extensively used in research as a means to trace and quantify important molecules. Fluorescent proteins are one of the most widely used tools in research, these proteins have expanded our understanding of gene expression and have been employed as probes for many targets in cells and *in vitro*. RNA is a fundamental molecule of life, serving as an informational carrier of genetic information in addition to facilitating many processes necessary for cell development. There are limited options available to trace or tag RNA based on emission of light that do not require covalent modifications. Recent developments have led to RNAs that provide fluorescent feedback when bound to a target, serving as tool to quantify and trace RNA[2, 3] (see chapter 1). As more probes are discovered they will add to the repertoire of techniques available to study the dynamics of RNA. Lanthanides are elements with luminescent properties that differ from the emission

properties of most common fluorophores. A lanthanide based RNA probe can be used to take advantage of these unique properties and create new techniques to trace RNA.

Lanthanides are f-block transition metals capable of luminescence and have been studied extensively for their unique spectroscopic and coordination properties. They occur mainly as hard trivalent cations, capable of coordinating oxygen and nitrogen donor ligands. The coordination sphere of some lanthanides can accept up to nine ligands, allowing for formation of very strong complexes with multivalent ligands such as diethyelene triamine pentaacetic acid (DTPA) and ethylenediaminetetraacetic acid (EDTA). Emission of some lanthanides occurs in the visible region, making them valuable luminescent tools for many biological applications. Luminescent lanthanide bioprobes have been used for immunoassays, protein staining assays, nucleic acid probing, and enzyme activity assays[4].

Lanthanides have several advantages that make them ideal luminescent probes, including: 1) cell permeability of many complexes, 2) large Stokes shifts, 3) low risk of photo bleaching, and 4) long emission lifetimes – on the order of milliseconds[1, 4]. Long emission lifetimes allow tracking of molecules without overly long exposure to light and large Stokes shifts allow for flexibility in selecting a diverse range of antennas or Förster resonance energy transfer (FRET) partners with excitation maxima far from emission maxima. Using an antenna with a large Stokes shift has the advantage of minimizing background signal due to limited auto fluorescence of an excitation source. Energy transfer occurs by the absorption of energy by an antenna molecule, resulting in a singlet excited state that undergoes intersystem crossing to a triplet state of the ligand, and the energy from the triplet excited state of the ligand is transferred to one of the excited states of the target lanthanide[1, 4]. Lanthanides exhibit minimal auto fluorescence on their own; therefore, the reliance on an antenna molecule to sensitize lanthanide excitation makes it possible for high resolution kinetic and

localization studies with a low background compared to many fluorophore probes that exhibit auto fluorescence.

There are several features necessary for an antenna to act as a sensitizer for lanthanides: the antenna molecule must be able to coordinate or chelate with the lanthanide, absorb energy in the form of light, and transfer enough energy to the lanthanide to induce an excited state of the lanthanide[1, 5]. Energy from the triplet state of the antenna is transferred through an intermolecular energy transfer to the lanthanide, resulting in lanthanide luminescence. Commonly used antennas (sensitizers) include derivatives of coumarins[6], picolinic acids[7], and tetracycline[8]. There are several examples of coumarin-based sensitizers for lanthanides; however, not every coumarin capable of chelating lanthanides results in an energy transfer[6]. Even if a potential ligand has many of the known characteristics necessary of an antenna an energy transfer may not occur, it is not possible to predict if the molecule is capable of an energy transfer, thus antennas are screened and discovered experimentally and have been found to possess varying degrees of sensitization capability.

The unique luminescent properties of lanthanides have resulted in their extensive use in many molecular bioimaging techniques. Lanthanides possess many advantages over fluorescent probes that can be exploited for imaging applications in which fluorophores are not suitable. Long emission lifetimes are crucial for observing localization in cells and developing images with high spatial resolution. The millisecond lifetime of lanthanide emission is appropriate to observe many biochemical processes without long exposure of light to the target. The emission properties of lanthanides have made them an excellent candidate as a biomarker of cellular processes, Parker et al. developed several cell permeable lanthanide complexes capable of luminescence inside cells[9]. These lanthanide probes are capable of highly visible luminescence in carcinoma (HeLa) cells, resulting in a probe to

observe ligand uptake and localization by cell imaging. The luminescence was demonstrated to be powerful enough to result in tracers for different compartments of the cell. These complexes were used to image and trace development of cells by tracing cell growth and stability[9].

In addition to long emission lifetimes, cellular imaging techniques require low background noise in order to precisely measure and observe the target being probed. Analyte detection in cells requires high resolution in order to detect and quantify the amount of analyte, lanthanides meet the requirements necessary for sensitive detection. A lanthanide-based zinc sensor was injected into HeLa cells to monitor levels of zinc in cells[10]. At physiological levels of zinc, only background luminescence could be detected, addition of zinc caused emission to increase eight-fold and the difference could be observed by cell imaging. This assay demonstrated the advantage of antenna formation to trigger luminescence: background emission is very low until a complex with the antenna is formed.

2.3 Lanthanide probes for nucleic acids

Nucleic acids have previously been shown to be capable of transferring energy to lanthanides, resulting in an enhancement of emission[1, 5]. Lanthanide luminescence was used to probe the hammerhead ribozyme with both terbium (Tb) and europium (Eu) by the Uhlenbeck lab. Divalent metals are cofactors for the hammerhead ribozyme and facilitate both structure formation and catalysis of self-cleavage. Eu (III) and Tb (III) are capable of binding to magnesium-binding sites on the hammerhead ribozyme, resulting in inhibition of cleavage. The hammerhead ribozyme is capable of transferring energy to Eu (III) and Tb (III), resulting in an enhancement of luminescence from the lanthanide[5]. Emission of Eu (III) and Tb(III) were measured to detect lanthanide binding to the ribozyme. Steady state luminescence was measured of Eu (III) and Tb (III) in the presence and absence of the

ribozyme. UV light at 260 nm, where RNA absorbs light, was used as the excitation source and resulted in emission from the lanthanides only in the presence of the ribozyme, making it possible to monitor lanthanide binding to the ribozyme by luminescence. Binding affinity as well as steady state kinetics were measured by lanthanide luminescence. The findings of the Uhlenbeck group demonstrated that lanthanides could be used to probe binding to nucleic acids, however the extent of their application remained largely unknown as they only tested one target RNA with a high affinity for multivalent cations.

Many different types of nucleic acids were tested in order to understand the specificity and limits of lanthanides as a general nucleic acid probe. Greenbaum and Yuan developed and tested the antenna effect of several RNA constructs with metal binding sites in presence of lanthanides to determine the efficacy of RNA as a lanthanide sensitizer[11]. Lanthanide emission increased with increasing concentration of RNA for all constructs reported, although the sensitivity and enhancement was not as robust as most commonly used lanthanide sensitizers. These experiments demonstrated the capability of RNA with strong metal binding sites to transfer energy to the lanthanide. The highly ordered structure of aptamers creates opportunities for the lanthanide to be directly chelated to the RNA, resulting in an energy transfer from the aptamer to the lanthanide. The results from Greenbaum and Yuan did not produce enough luminescence for practical uses, although their constructs were not selected or specifically developed for direct sensitization.

Lanthanides have also been studied for their binding and optical potential with nucleotides. Several groups have used lanthanides to detect nucleotides by either luminescence or quenching assays. Tang et al. developed a probe that distinguishes ATP from ADP and AMP by luminescence; several lanthanide chelate complexes were developed capable of luminescence in the presence of ATP in water while discriminating against ADP

and AMP. Other phosphate based anions were tested as controls to ensure sensitization of the lanthanide is specific for ATP rather than an ionic strength sensor, these anions produced negligible emission[12]. Wolfbeis et al. utilized the quenching ability of nucleotides to probe for kinase activity of ATPase by lanthanide-induced luminescence quenching. Several phosphate series of ATP and GTP were tested, and each produced a unique quenching profile, making it possible to test the phosphorylation state of each analog and kinase activity of ATPase[8]. These two experiments serve as examples of nucleotides used as both enhancers and quenchers for lanthanide luminesce by finding the correct chelate complexes and conditions to achieve the desired effect on the lanthanide probe.

Aptamers are oligonucleotides capable of binding a target with high specificity[13]. Many fluorescent based biosensors have been discovered, selected, or engineered using aptamer technology (see chapter 1). Pierre and Wickramaratne designed a lanthanide aptamer probe that resulted in 21-fold increase in luminescence upon binding to its target, using a europium chelate complex bioconjugated to the 5′ end of an Hg^{2+} aptamer and used as a turn-on probe for target binding. In the absence of Hg^{2+} the probe intercalated into the aptamer and lanthanide luminescence was quenched, but when the aptamer was bound to Hg^{2+} the probe was displaced from the aptamer and luminescence can be observed[14]. Luminescence was quenched by purines when intercalated into the aptamer. The aptamer used in this assay demonstrates the ability of an aptamer to bind to and alter the emission of a bound lanthanide.

Aptamers are often utilized as molecular beacons by using fluorophore quencher or FRET pairs. These pairs can be conjugated to a single aptamer or an aptamer and its target to observe the interactions and state of the aptamer[15] (see chapter 1). Soukka et al. developed an assay for thrombin detection based on an aptamer-direct lanthanide chelate assembly.

Emission of lanthanide was based on the binding of two thrombin aptamers, the first conjugated with a lanthanide chelate and the second conjugated with the antenna for the lanthanide. When both aptamers were bound to thrombin the antenna and lanthanide assembled in close proximity to result in lanthanide luminescence[16]. The technique resulted in rapid thrombin detection utilizing the antenna effect of the lanthanide. Although many fluorophore quencher pairs are capable of similar detection assays, most applications are dependent on the quenching of a fluorophore when in proximity to its target causing a decrease in light. Lanthanides have the advantage of being turned on and increasing emission when in close proximity to its antenna pair, with low background luminescence in the absence of the antenna.

2.4 An RNA aptamer as a luminescent tracer

We hypothesized that an RNA aptamer for lanthanides could serve as an antenna, enabling energy transfer from RNA to the metal ion and enhancing emission of otherwise poorly luminescent probes. A luminescent RNA aptamer has the potential to be used as a molecular genetics tool for quantifying expression levels of RNA, as well as observing localization of RNA *in vivo.* The aptamer would serve as an optical probe that can be engineered or expressed upstream/downstream a target RNA of interest by the addition of a lanthanide specific-aptamer domain. In the absence of aptamer (RNA) expression lanthanide luminescence would be negligible, in the presence of expression luminescence would be enhanced upon binding to the aptamer, and intensity of luminescence would give direct feedback of RNA expression for quantification or imaging.

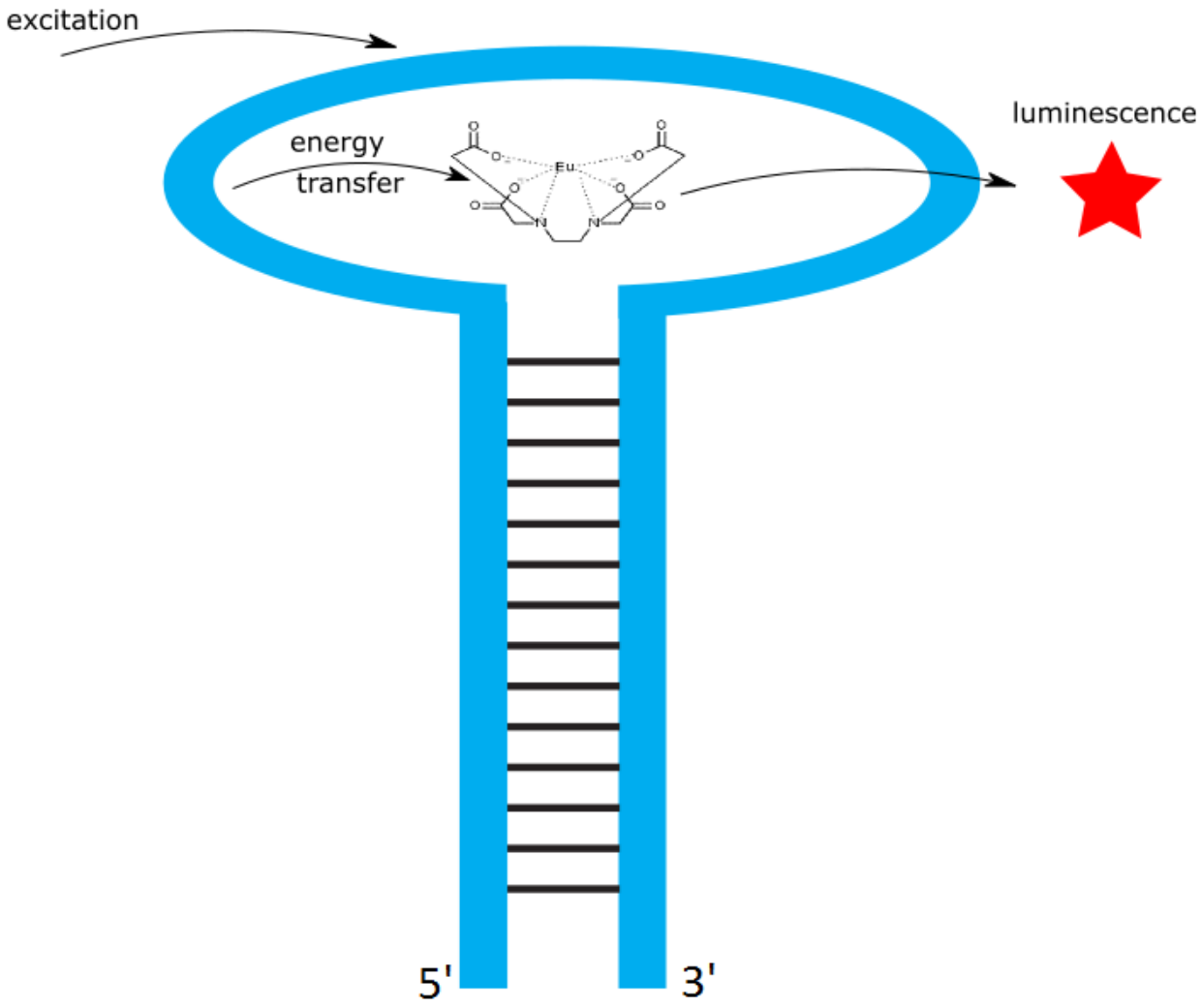


Figure 2.1 Developing an RNA antenna for lanthanide luminescence. An aptamer that specifically binds to a lanthanide chelate can be used as the sensitizer for lanthanide luminescence. The aptamer is excited by light, the excitation energy is transferred to the chelated lanthanide, resulting in luminescence from the metal ion.

2.5 Progress towards a lanthanide aptamer

In vitro selections have been performed with EDTA-Eu and DTPA-Eu chelate complexes as the target ligand (methods 6.5). A random pool of RNA with diversity of ~10^{15} unique sequences was used as the starting population for selection. The starting pool was designed with constant regions consisting of a forward primer containing a T7 promoter for transcription and a reverse primer in order to reverse transcribe and amplify sequences for serial rounds of selection. The primers flank a 44 nucleotide random nucleotide region that gives the pool its complexity and potential aptamer domain for the lanthanide chelate, the total length of the pool is 100 nucleotides. Once an aptamer is found, various constructs can be made of the original aptamer to reduce size and enhance binding.

To isolate the desired property, each RNA pool is incubated with the selection target displayed on beads. The fraction that does not bind to the target is washed away with a buffer mimicking physiological conditions (binding buffer) and discarded. The bound RNA was competitively eluted with mM range of the ligand in a 1:1 ratio of chelator-lanthanide and collected for the next round of selection. RNA was labeled with ^{32}P-α-ATP in order to measure the fraction of RNA that is bound and specifically eluted for each selection round. Selections were performed until pool enrichment was leveled off, as monitored by total RNA eluted as well as the ratio of RNA in the first elution to the last wash.

Ten rounds of selection were performed with DTPA-Eu as the target molecule. After ten rounds, binding of aptamers increased from less than 1% to 25%. Pools from rounds with greater than 20% binding were cloned and tested for potential specific aptamers. Nine rounds of selection were performed on EDTA-Eu and binding of the pool was enriched to 30%. Both selections were successful in enriching pool specificity for the target and individual sequences were further tested for specificity and binding

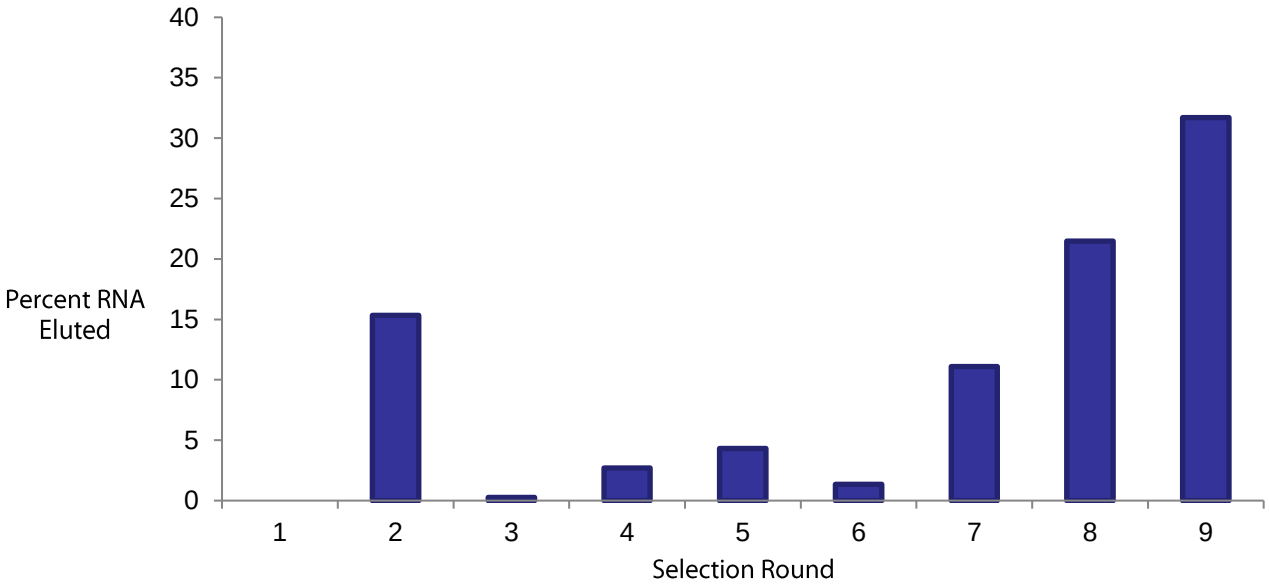


Figure 2.2 Progress of the EDTA-Eu selection. Binding profile for 9 rounds of selection. Several stringency steps were incorporated after round 2 to prevent background binding, including incubating with a pre column and pre eluting with DTPA-Eu. Strong enrichment of eluted RNA can be observed for rounds 7-9.

Stringency conditions are incorporated into selections to enhance specificity for the selection target. Selections often enrich for sequences that are fit for the environment conditions or excel in an individual step of the selection such as a faster rate of amplification or transcription, resulting in enrichment of sequences that may not be ligand-dependent[17, 18]. Additionally, pools may be selected for binding to another component of the selection, such as the beads that contain the immobilized target or another molecule in the selection. To reduce the population of these non-specific binders, additional stringency steps were implemented in both of the lanthanide chelate selections.

The selection is designed to enrich for aptamers of the specific chelate target displayed on beads, either EDTA-Eu or DTPA-Eu. Specificity is important for broadly applicable use as many molecules are able to chelate and form complexes with lanthanides. To enhance enrichment of a pool for a specific chelate a wash can be incorporated containing a similar chelate target. For example, DTPA-Eu was added to washes for the EDTA-Eu selection to minimize selection of RNA capable of binding a nonspecific lanthanide chelate complex. The beads used in these selections are agarose beads that contain the chelator for the lanthanide (DTPA or EDTA). Any unbound chelator may be present as a target, in addition to the agarose beads they are covalently attached to. To subtract RNA with affinity for these targets the pool was incubated with the beads without any bound lanthanide prior to incubation with the chelated lanthanide. Selection targets may contain sites that have a background affinity for RNA, causing enrichment of nonspecific binding. Non selected RNA was incubated with the target to block these sites and prevent enhancing the pool for low affinity or nonspecific binders. Incubation times can be varied to select for aptamers that contain faster or slower on rates, the incubation time was decreased in later rounds to select for aptamers with faster on rates. Ligand concentration can also be varied to select for higher affinity binders, although

ligand concentration was kept constant for the selections. Conditions are critical in the design of a selection, additional stringency steps aid in enriching the population for the desired outcome.

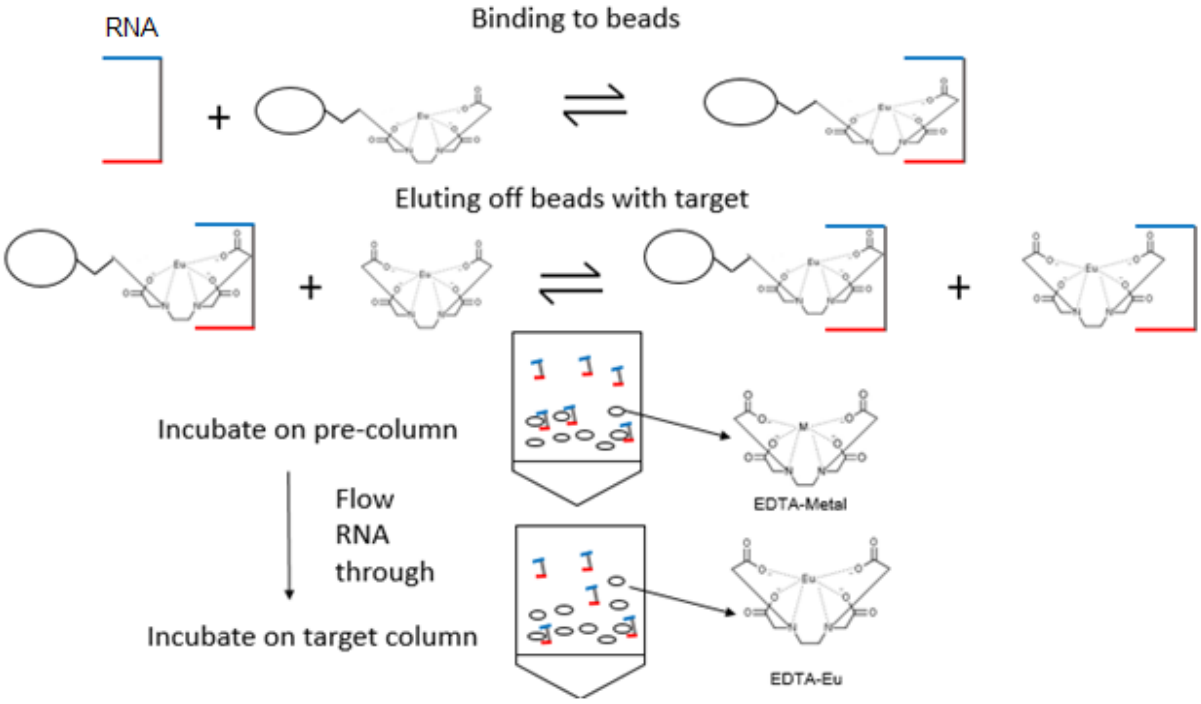


Figure 2.3 Column binding work flow. Pools are incubated on beads and eluted off with free ligand in solution. Additional constraints to enhance specificity include binding to a pre-column with a similar target to enhance specificity of the pool to the correct target and pre-eluting with a structurally similar off-target.

2.6 Screening clones (methods 6.6)

Potential aptamers from the rounds with enriched pools were cloned to isolate single sequences of interest. Individual clones were tested for binding to the ligand on beads to determine the clone's affinity for the ligand. Similar to the selection, clones were incubated on the beads with a lanthanide target, washed with binding buffer, and competitively eluted with an excess of lanthanide chelate complex. If the clone bound specifically, a high percentage of the total RNA will be eluted off with the lanthanide chelate, rather than discarded in prior wash steps. Several clones for both targets indicated a positive binding profile, and were sequenced and tested for luminescence. Individual clones have shown different affinities of binding to the target, from less than 2% to over 13% of total RNA eluted under the same conditions (Fig 2.4).

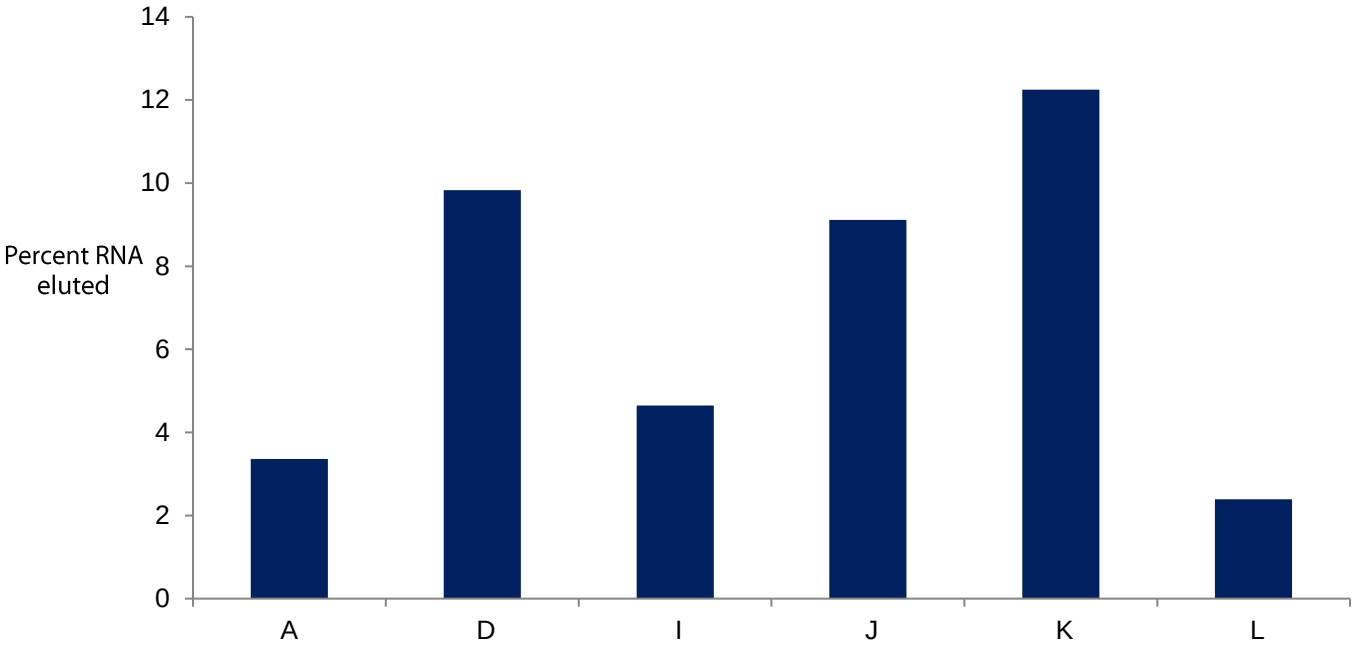


Figure 2.4 Column binding of individual clones. Individual clones (x axis) were tested by column binding steps, analogous to a single round of selection. These clones demonstrate different affinities for the target by amount of RNA eluted, estimated as ~2-13 percent binding.

Clones that demonstrate affinity by column binding were tested for optical activity to investigate whether any of the aptamers were functionally active as a sensitizers for the selected lanthanide complex. Luminescence of lanthanides in the presence of clones was tested to observe emission enhancement. Plasmids from clones were transcribed, RNA was purified on a denaturing polyacrylamide gel (PAGE), and diluted to a concentration of ~100 µM. The RNA of each clone was incubated with a lanthanide chelate complex at a 1:1 ratio in the binding buffer used for the selection. RNA absorbs light at 260 nm and it has been previously demonstrated with the hammerhead ribozyme that RNA can transfer energy to lanthanides when used as an antenna with an excitation of 260 nm[5]. Clones were heated in the presence of the ligand to induce refolding and binding. Selection was performed with purified RNA, however; an aptamer capable of binding co-transcriptionally is more suitable for feedback and localization studies. Clones that exhibited promising results with purified RNA were also tested co-transcriptionally *in vitro* to test if they remain functionally active.

To detect enhancement of emission or novel emission peaks from the lanthanide target in the presence of RNA, 3D emission spectra were measured using an excitation range of 250 nm to 450 nm in steps of 10 nm. For each excitation step, we recorded emission in the range of 500 nm to 800 nm. Although ~ 260 nm is the optimal wavelength for RNA absorbance, a broader range is used in case there is a shift in the absorbance maxima of any clone in the presence of the ligand. Any RNA-ligand complexes exhibiting an increase of intensity or shift in emission from the lanthanide was recorded. Enhanced emission could be observed for some clones (Fig 2.5), although results were not reproducible by this method leading us to test for optical activity directly in cells.

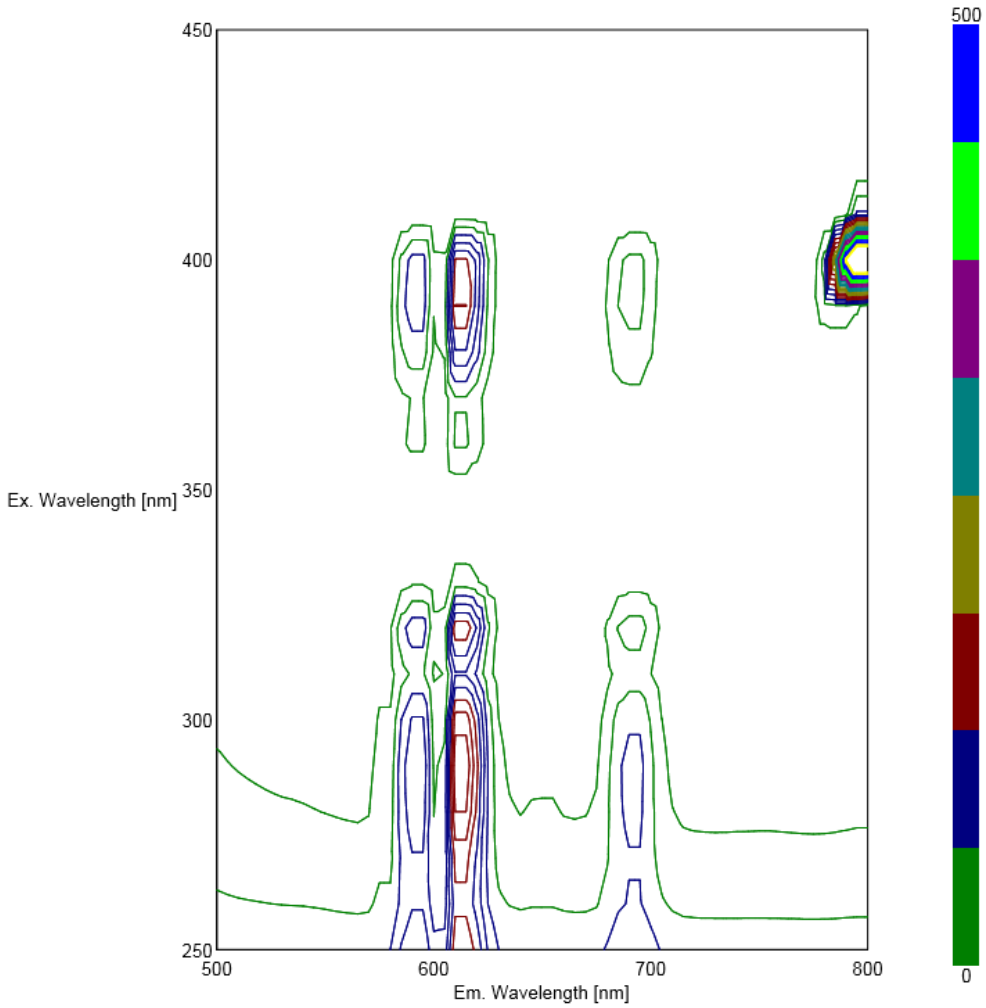


Figure 2.5 3D contour plot of clone 7-5 with EDTA-Eu. Emission maximum of europium is at ~ 615 nm with several additional peaks in the 600 – 750 nm range. Clones are screened to determine intensity of luminescence in the presence of RNA. Enhancement of luminescence was observed up to five-fold compared to EDTA-Eu alone, although results were not reproducible.

Additional luminescence testing was carried directly in bacteria. *In vivo* testing was carried out by fluorescent microscopy as well as LB agar imaging. Droplets of cells containing either selected pools or a high density of a single clone were placed on a microscope slide with 100 μM of the lanthanide chelate. A control RNA from a different selection with another target was used as a control for background luminesce of random RNA with the lanthanide. Clones were compared to the control and each other in order to determine if they had an enhanced luminesce. Selected pools were tested for enhanced luminescence compared to others, although it is not possible physically separate single cells from the pool with this method. Several excitation filters in the range of 200 – 400 nm were used in order to test a wide excitation spectrum while screening clones.

Pools and clones were also tested on bacteria grown on LB plates. High μM to low mM of lanthanide chelate was spread and dried on the surface of LB agar. Colonies were grown on plates containing the lanthanide chelate and screened by luminescence under different excitation wavelengths by fluorescence microscopy or directly under a UV lamp. Any interesting clones were collected, cultured, and further evaluated to determine the identity, optical activity, and affinity for the ligand.

2.7 Competition assay (methods 6.7)

Lanthanide luminescence can be employed to determine the binding state of the lanthanide by using a known quencher or sensitizer. Previously, Felipe-Sotelo et al. developed a competition assay to detect the effect of europium on nickel complexation with picolinic acid by titrating nickel to several Eu (III) chelate complexes[7]. Picolinic acid is a well characterized sensitizer of lanthanide luminescence, lanthanide emission will decrease if another molecule competes with picolinic acid for the lanthanide. To test binding specificity we used lanthanide sensitizer dipicolinate DPA) to compete with clone 7-5 and detect

competition. DPA forms a strong chelate complex with Eu (III) and is also a well characterized sensitizer of lanthanide luminescence *in vitro*[7]. If the RNA binds with the lanthanide complex, the complex will be in equilibrium with both RNA and DPA. As more RNA is added, the equilibrium will shift as it reaches the binding affinity of the aptamer. Luminescence can be used to detect competition using DPA as the excitation source to detect lanthanide not bound to RNA as luminescence will decrease as lanthanide is bound to RNA and unavailable to form a complex with DPA. A dilution series of 0.5 nM to 10 µM RNA was incubated with 1 nM EDTA-Eu in binding buffer for an hour. 10 µM of dipicolinate (DPA) was added to each sample and emission spectra were taken for each dilution. Luminescence intensity was plotted against RNA concentration to create a competition curve. At high nM concentrations of clone 7-5 luminescence decreases due to binding to the clone, the estimated K_D based on competition assay data is 700 nM.

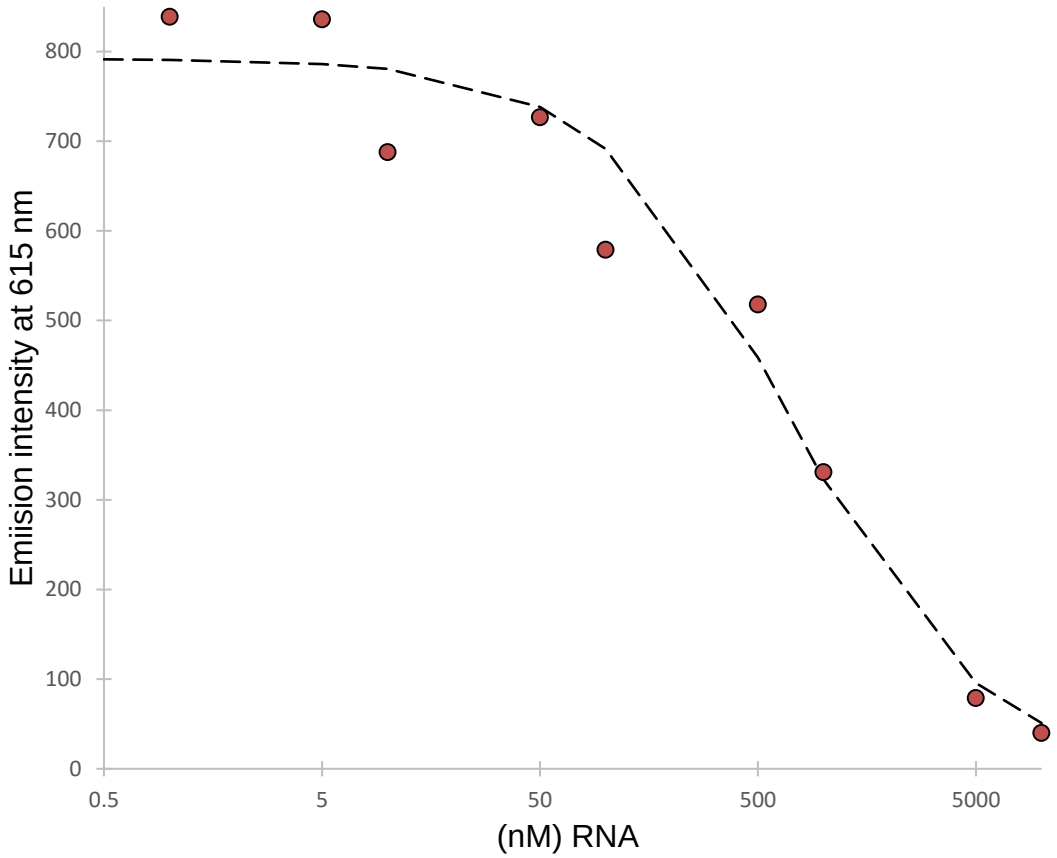


Figure 2.6 Competition assay of clone 7-5. EDTA-Eu is incubated with a lanthanide sensitizer (DPA) and clone 7-5 to compete for ligand binding. At high nM concentrations of clone 7-5 the luminescence of europium decreases, indicating less is bound to DPA due to competition with clone 7-5. Excitation is at 355 nm.

2.8 Identity, structure prediction, and probing of clone 7-5

In-line probing is a technique that can be used to address structural differences in RNA upon binding to a ligand. The technique is based on the propensity of RNA to hydrolyze over time. RNA is susceptible to an in-line attack of the 2′OH of the sugar to the adjacent phosphate that links the 5′ oxygen of the next nucleotide[19]. In the presence of magnesium, the 2′ oxygen becomes a good nucleophile to attack the adjacent phosphodiester linkage, resulting in cleavage of RNA. When an aptamer binds a ligand, it may prevent or enhance an in-line attack directly where the ligand is bound or in-line probing may detect a change in structure elsewhere in the RNA that undergoes a conformational change due to ligand binding. End-labeled RNA is incubated with increasing concentrations of the ligand for two or three days to allow partial cleavage to occur, products are analyzed on a denaturing PAGE gel to observe any differences in structure based on bands that undergo a change in intensity in response to ligand concentration. Structures can be resolved to a single nucleotide resolution with high percentage (12-15%) PAGE gels. Clones 7-3 and 7-5 have shown unique in-line properties in several regions that correlate with ligand concentration (Fig. 2.7, indicated by black brackets). Lanthanides are also capable of catalyzing an in-line attack in a similar fashion to magnesium but at a faster rate, resulting in their use as in-line and footprinting probes[20]. Interestingly clone 7-5 shows several sites that become protected (bands are less intense) with high concentrations of EDTA-Eu, indicating a fully coordinated europium preventing an in-line attack. An aptamer for a lanthanide chelate may bind the lanthanide at a site that prevents it from deprotonating the 2′OH, suggesting specific binding.

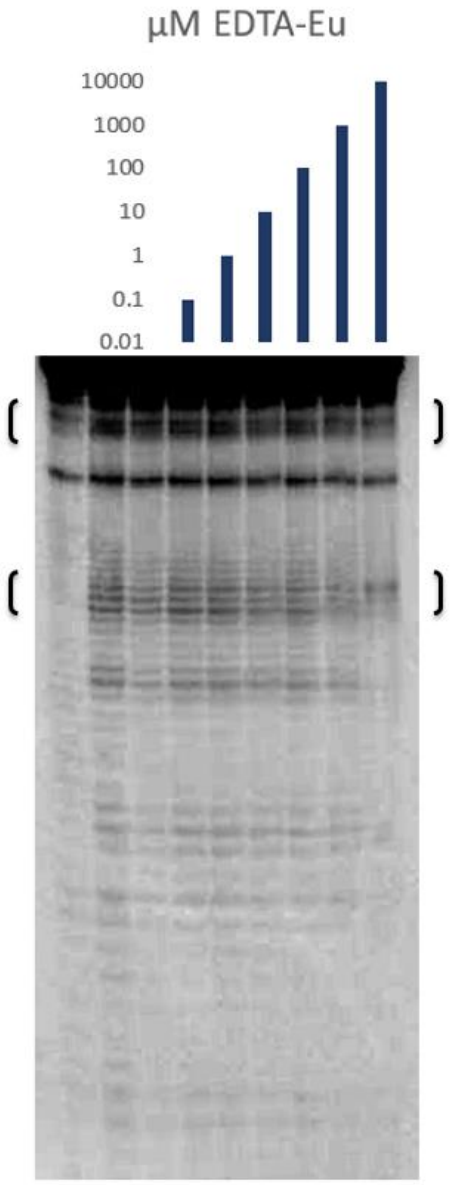


Figure 2.7 In-line probing of clone 7-5 (methods 6.5). Clone 7-5 was incubated with EDTA-Eu for three days to induce an in-line attack. Several areas of the RNA became protected as indicated by the intensities of bands in response to higher concentration of the EDTA-Eu (black brackets). Lanes 1-3 contain controls; lane 1: RNA and water, lane 2: RNA and binding buffer, 3: RNA and 100 µM EDTA-Mg.

Identity and structure prediction for clone 7-5

5'GGGAGAGCGAGCAGAAAUCGUGACAUCUUGAUACGCUUGACUUAACUGCGCAUCGGAUCGU

UACGACUAGCAUCGAUG 3'

A structure for clone 7-5 was predicted using RNAfold WebServer which predicts secondary structures based on minimum free energy. Base pair probabilities are given from 0-1 or a color scale of blue to red (red indicates a high probability of base pairing). Structure predictions are useful in suggesting secondary structure motifs that make up common binding motifs in aptamer domains. A ligand is more likely to be bound in a stable structural motif such as a stem loop rather than an unstructured single stranded region of an RNA, although tertiary contacts can also stabilize a ligand. The structure of clone 7-5 is largely complementary to itself, consisting of a large stem with several bulges and loop. 7-5 has a strong base pairing probability profile based on the prediction.

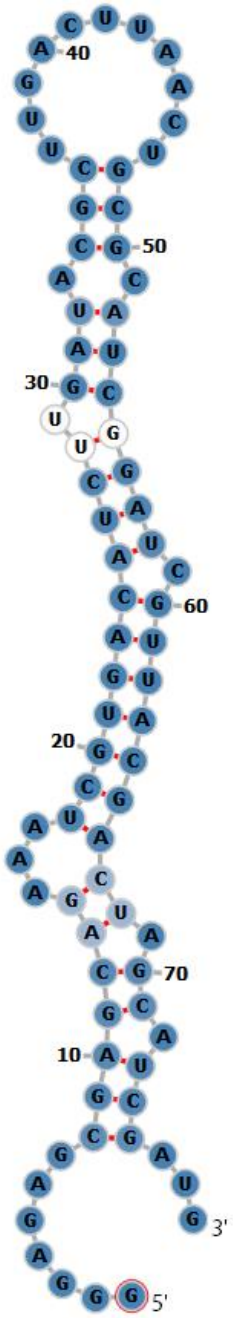


Figure 2.8 Structure prediction of clone 7-5. Clone 7-5 is composed of a long stem loop containing several bulges. Base pair probabilities given by color intensity (blue). Ligand binding is most likely to occur in a bulge or loop of the structure.

2.9 Using known aptamer targets as lanthanide complexes

There are several known lanthanide antenna complexes that enhance the luminescence of lanthanides. Most of these are large aromatic molecules with unique optical properties. Tetracycline (Tc), a widely used antibiotic, is capable of transferring energy to lanthanides, resulting in tetracycline-sensitized luminescence of lanthanides. Lanthanides have been used to determine the presence tetracycline in milk using a polymer based complex consisting of AMP/Eu[21]. The AMP/Eu-Tc complex resulted in lanthanide emission at 615 nm when using tetracycline as sensitizer for europium and upon excitation of tetracycline at 390 nm. Therefore we reasoned than an aptamer binding to Eu-Tc complex could be used to make tetracycline a stronger sensitizer.

An RNA aptamer for tetracycline was previously discovered (cb28), binding the antibiotic with a nanomolar K_D[22]. We used this aptamer to test whether the presence of Eu (III) alters the absorbance or luminescent properties of either tetracycline or the Eu (III) when the aptamer is bound. 1 μM – 20 μM of the cb28 aptamer was incubated with Eu-Tc or Tb-Tc in a 1:1 ratio. The aptamer resulted in an increase in fluorescence of tetracycline (Fig 2.9 black) around 500 nm, however the aptamer quenched lanthanide emission (Fig 2.9 at 615 nm). Although the tested aptamer quenched lanthanide emission, the increased emission from tetracycline itself indicated that aptamer binding changes the excitation or transition state of tetracycline when bound. Alternate binding of Eu-Tc may influence different optical properties and enhance its potential as a sensitizer. Other tetracycline aptamers exist that can be tested as the fold of an aptamer may impact the capacity of sensitization.

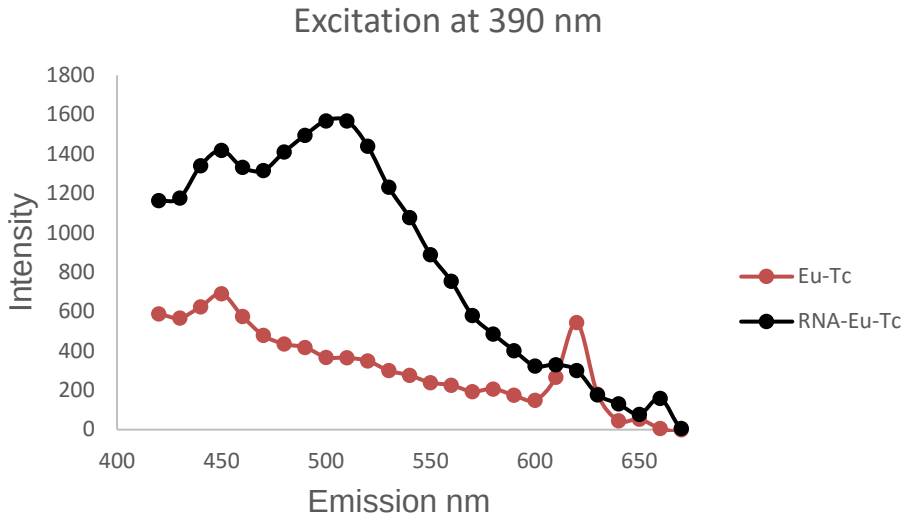


Figure 2.9 Emission spectra of tetracycline aptamer with Eu-Tc (methods 6.8). 1 µM of cb28 was incubated with Eu-Tc in a 1:1 ratio of RNA-ligand (black) and without RNA (orange). The aptamer increased the fluorescence of tetracycline (~ 500 nm) but europium emission at 615 nm was quenched by the aptamer.

2.10 Conclusions

In vitro selections have resulted in many enriched pools with potential aptamers for a lanthanide chelate, with clone 7-5 as the most promising aptamer. Screening by luminescence methods provided no promising results, but it is possible that energy transfer from RNA to a lanthanide chelate may not be strong enough to induce sensitization of the lanthanide for enhanced luminescence. At the same time as our efforts to find an RNA sensitizer, Pierre and Wickramaratne developed an aptamer based light switch probe that used an aptamer to quench luminescence when the lanthanide was bound by the aptamer. Their experiments indicate that purines are effective in quenching lanthanide luminesce[14]. It may be necessary to create a pool that is biased towards pyrimidines in order to influence sensitization from an aptamer. A known tetracycline aptamer, cb28, also resulted in quenching of lanthanide luminesce when bound to Eu-Tc. Although extensive functional screening from our selection did not yield any aptamers which enhanced luminescence there was evidence of aptamers from several enriched selection pools by column binding.

Selection results indicated binding profiles that increased after serial rounds of selection, even after additional stringency conditions were added. Column binding of individual clones displayed a range of affinities for the target. Evidence from in-line probing, filtration assay, and competition assays for clone 7-5 indicated that it is an aptamer for EDTA-Eu, however the aptamer is not functionally active in enhancement of luminescence. In the competition assay with known lanthanide sensitizer DPA (section 2.7) clone 7-5 quenched lanthanide luminescence at high nM concentrations, indicating specific binding by the clone. Although clone 7-5 was not optically active as a sensitizer for luminescence, it may be possible to use it as a quencher in assays that use lanthanide luminescence.

An aptamer that influences lanthanide luminescence upon binding would be advantageous for RNA expression and localization studies both *in vivo* and *in vitro*. Although aptamers that enhance fluorescence upon binding exist, they do not possess many of the advantageous optical properties (such as low background emission) that make lanthanides highly utilized in biological assays. Despite the difficulties of discovering an RNA aptamer that induces lanthanide luminesce, our data indicated discovery of aptamers is achievable. In addition to our own data, recently published data demonstrated that a sensitizing aptamer may be rare and difficult to discover due to quenching upon ligand binding[14] (sections 2.7 and 2.9). Our initial selection efforts focused on affinity, selection schemes or methods can be developed to select for function rather than binding.

A different method for aptamer induced lanthanide luminescence may be to select for an aptamer of a sensitizer rather than utilizing RNA itself as the sensitizer. Several aptamers have demonstrated the ability to change the optical properties of its ligand[2, 3]. Aptamer binding of a sensitizer might result in a change of its excitation profile resulting in enhanced sensitization and luminescence from the lanthanide. There are many known sensitizers for lanthanides that greatly enhance luminescence, while other derivatives are not able to transfer energy. These nonfunctional derivatives are hopeful targets as similar molecules have an aptitude for sensitization, binding of an aptamer can result in a difference of the optical profile for these nonfunctional derivatives. Many aptamers have been discovered for ligands that are capable of chelating lanthanides, these could be screened to discover if these aptamers result in a difference in emission upon aptamer binding. We tried this approach with the tetracycline aptamer (section 2.9); however, there are several other aptamer targets that could be tested such as nucleotides and amino acids capable of chelating lanthanides[4, 12]. There are many known natural RNAs such as tRNA[23] and ribozymes[5] that have binding

pockets or are capable of chelating divalent and trivalent metals, these can be studied as is or reselected and engineered providing another potential avenue for the development of luminescent RNA tracers.

1. Bunzli, J.-C.G. & Piguet, C. Taking advantage of luminescent lanthanide ions. *Chemical Society Reviews* **34**, 1048-1077 (2005).
2. Paige, J.S., Wu, K.Y. & Jaffrey, S.R. RNA Mimics of Green Fluorescent Protein. *Science* **333**, 642-646 (2011).
3. Dolgosheina, E.V. et al. RNA Mango Aptamer-Fluorophore: A Bright, High-Affinity Complex for RNA Labeling and Tracking. *ACS Chemical Biology* **9**, 2412-2420 (2014).
4. Eliseeva, S.V. & Bunzli, J.-C.G. Lanthanide luminescence for functional materials and bio-sciences. *Chemical Society Reviews* **39**, 189-227 (2010).
5. Feig, A.L., Panek, M., Horrocks, W.D., Jr. & Uhlenbeck, O.C. Probing the binding of Tb(III) and Eu(III) to the hammerhead ribozyme using luminescence spectroscopy. *Chem Biol* **6**, 801-810 (1999).
6. Andres, J. & Chauvin, A.-S. Energy transfer in coumarin-sensitised lanthanide luminescence: investigation of the nature of the sensitiser and its distance to the lanthanide ion. *Physical Chemistry Chemical Physics* **15**, 15981-15994 (2013).
7. Evans, N.D.M., Gascón, S.A., Vines, S. & Felipe-Sotelo, M. Effect of competition from other metals on nickel complexation by α-isosaccharinic, gluconic and picolinic acids. *Mineralogical Magazine* **76**, 3425-3434 (2012).
8. Schäferling, M. & Wolfbeis, O.S. Europium Tetracycline as a Luminescent Probe for Nucleoside Phosphates and Its Application to the Determination of Kinase Activity. *Chemistry – A European Journal* **13**, 4342-4349 (2007).
9. Montgomery, C.P., Murray, B.S., New, E.J., Pal, R. & Parker, D. Cell-Penetrating Metal Complex Optical Probes: Targeted and Responsive Systems Based on Lanthanide Luminescence. *Accounts of Chemical Research* **42**, 925-937 (2009).
10. Hanaoka, K., Kikuchi, K., Kojima, H., Urano, Y. & Nagano, T. Development of a Zinc Ion-Selective Luminescent Lanthanide Chemosensor for Biological Applications. *Journal of the American Chemical Society* **126**, 12470-12476 (2004).
11. Yuan, F. & Greenbaum, N.L. Use of RNA-bound Tb3+ as a FRET donor. *Methods* **52**, 173-179 (2010).
12. Liu, X. et al. An ATP-selective, lanthanide complex luminescent probe. *Dalton Transactions* **42**, 9840-9846 (2013).
13. Ellington, A.D. & Szostak, J.W. In vitro selection of RNA molecules that bind specific ligands. *Nature* **346**, 818-822 (1990).
14. Wickramaratne, T.M. & Pierre, V.C. Turning an Aptamer into a Light-Switch Probe with a Single Bioconjugation. *Bioconjugate Chemistry* **26**, 63-70 (2015).
15. Wu, C., Yang, C.J. & Tan, W. in Molecular Beacons. (eds. C.J. Yang & W. Tan) 175-194 (Springer Berlin Heidelberg, Berlin, Heidelberg; 2013).
16. Pakkila, H., Blom, S., Kopra, K. & Soukka, T. Aptamer-directed lanthanide chelate self-assembly for rapid thrombin detection. *Analyst* **138**, 5107-5112 (2013).
17. Jijakli, K. et al. The in vitro selection world. *Methods* **106**, 3-13 (2016).
18. Eulberg, D., Buchner, K., Maasch, C. & Klussmann, S. Development of an automated in vitro selection protocol to obtain RNA-based aptamers: identification of a biostable substance P antagonist. *Nucleic Acids Research* **33**, e45-e45 (2005).
19. Soukup, G.A. & Breaker, R.R. Relationship between internucleotide linkage geometry and the stability of RNA. *RNA* **5**, 1308-1325 (1999).
20. Harris, D.A., Todd, G.C. & Walter, N.G. in Handbook of RNA Biochemistry 255-268 (Wiley-VCH Verlag GmbH & Co. KGaA, 2014).
21. Tan, H. et al. Determination of tetracycline in milk by using nucleotide/lanthanide coordination polymer-based ternary complex. *Biosensors and Bioelectronics* **50**, 447-452 (2013).

22. Berens, C., Thain, A. & Schroeder, R. A tetracycline-binding RNA aptamer. *Bioorganic & Medicinal Chemistry* **9**, 2549-2556 (2001).
23. Tuschl, T., Sharp, P.A. & Bartel, D.P. Selection in vitro of novel ribozymes from a partially randomized U2 and U6 snRNA library. *The EMBO Journal* **17**, 2637-2650 (1998).

Chapter 3

Laboratory selection for a photo-active riboswitch

3.1 Specific aim

Riboswitches are functional RNAs that regulate gene expression upon binding of a ligand[1]. These regulatory RNAs can be employed to change the expression state of a target gene based on ligand-dependent allosteric control of transcription, translation, or splicing elements[1]. Our aim is to develop a riboswitch for a photoreversible ligand in order to obtain a light responsive riboswitch. Photoreversible ligands such as stilbenes are capable of *trans* to *cis* isomerization the influence of light[2]. A riboswitch specific to bind only one isoform of the ligand (such as the *trans* form of a molecule) may be released by light-induced isomerization of the ligand to another isoform (such as *cis*). Upon release by light, the expression state of the riboswitch will reverse to a state corresponding to the absence of the ligand.

3.2 Riboswitch regulation and discovery

Riboswitches are functional RNAs capable of binding a specific target molecule and modulating gene expression in response to binding to its target. There are several known bacterial riboswitches that modulate transcription or translation of genes, mediating regulation of metabolic processes in response to metabolite sensing and binding[1]. Riboswitches are composed entirely of RNA and contain two major domains that facilitate function: an aptamer domain that binds to a ligand, and a nearby domain that mediates gene expression in response to the aptamer domain, usually through conformational changes of the riboswitch. An aptamer is any DNA or RNA sequence capable of binding a selected target with affinity and specificity. Aptamers have been discovered and engineered *in vivo* and *in vitro* [3-5] for a wide variety of targets including metabolites, synthetic molecules, and cells[6]. A

change in the structure upstream or downstream of the aptamer domain can tune or regulate gene expression in response to binding of a ligand.

Aptamers for many targets have been identified experimentally from both genomic and synthetic pools. The propensity of RNA to fold into stable structures allows for selection and identification of aptamers by a method called *in vitro* selection; a process in which populations of DNA or RNA are enriched for a target by multiple rounds of affinity selection[3, 4]. Interestingly, aptamer domains of riboswitches have been discovered *in vivo* for several metabolites. For example, in bacteria to regulate expression in response to binding metabolites, such as thiamine pyrophosphate (coenzyme TPP)[7],*S*-adenosylmethionine (SAM)[8], and coenzyme B_{12}[9]. Although very few examples of eukaryotic riboswitches have been described[10], aptamer domains have been discovered in eukaryotes as well, including humans, by structure-based bioinformatics searches and an *in vitro* selection for ATP-binding RNAs. *In vitro* selection using a human genomic pool led to the discovery of human ATP aptamers[11, 12]. Even though no functional relevance of the human aptamers has been revealed yet, the discovery of aptamers across genomes has engendered interest in the possibility of riboswitch mediated gene expression throughout nature, particularly in higher eukaryotes. Furthermore, the broad occurrence of aptamer domains in nature has led to the investigation of their biological relevance and the discovery of more riboswitches.

Expression platforms have been found and engineered in proximity of important regulatory sequences and genes to regulate gene expression on the level of transcription, translation, and splicing[1, 10, 13]. The Breaker lab discovered the first riboswitch in the 5′ untranslated region of the *E. coli* gene *btuB*[9]. The riboswitch has a high affinity (300 nM) aptamer domain specific to cobalamin (coenzyme B_{12}). Translation of btuB mRNA is greatly reduced upon binding to B_{12}, establishing the capabilities of riboswitches to control gene

expression in response to metabolite sensing[9]. Regulation of the B_{12} riboswitch is achieved by allosteric control of translation initiation by sequestering the ribosome binding site (RBS) of the mRNA[9, 14]. On the other hand, the first transcriptional riboswitch was discovered by the Nudler lab, the TPP riboswitch regulates transcription by allosteric control of a terminator that attenuates transcription upon binding of TPP[7]. Expression domains that mediate control of gene expression on the level of transcription and translation share common motifs and modes of regulation; therefore, understanding these mechanisms is crucial to novel riboswitch discovery.

3.3 Common modes of regulation

Structural rearrangement of the expression domain is common amongst most ligand-dependent riboswitches. The methods of regulation often depend on sequestering a recognition motif necessary for expression, or a change in structure that destabilizes an enzyme necessary for expression. The recognition motif can be sequestered or released in response to a change in the aptamer platform upon ligand binding, with the Shine-Dalgarno (SD) motif and start codon as the most common sequences to control translation[1]. In these cases, regulation of translation relies on the availability of the SD or start codon for ribosome binding and translation initiation, respectively. These sequences undergo structural modification in response to conformational changes of the aptamer domain due to ligand binding, producing a switch of expression through allosteric induced rearrangement. The interaction between domains relies on the riboswitch structure, which is crucial for an active switch. For example the Lafontaine lab illustrated the importance of a kissing loop for the btuB riboswitch by testing variants with mutations that disturb the formation of the loop. Mutations of the loop resulted in loss of function, and the riboswitch no longer regulated translation in response to binding of the ligand[15].

Another form of regulation is achieved by forming a unique structural motif, a terminator, which affects stability of the RNA polymerase during transcription. A terminator is a hairpin structure followed by a run of U residues, which stall or destabilize the RNA polymerase-DNA complex during transcription[1, 7]. A terminator can be formed or prevented based on the presence of an anti-terminator structure, usually a hairpin nearby (either upstream or downstream) of the terminator that contains a sequence complementary to a portion of the terminator, preventing the terminator from forming. The formation of anti-terminator and terminator is dependent on the state of the aptamer domain, resulting in direct regulation in response to the ligand. Understanding these elements has led to understanding of the functions of natural riboswitches and the engineering and selection of synthetic riboswitches.

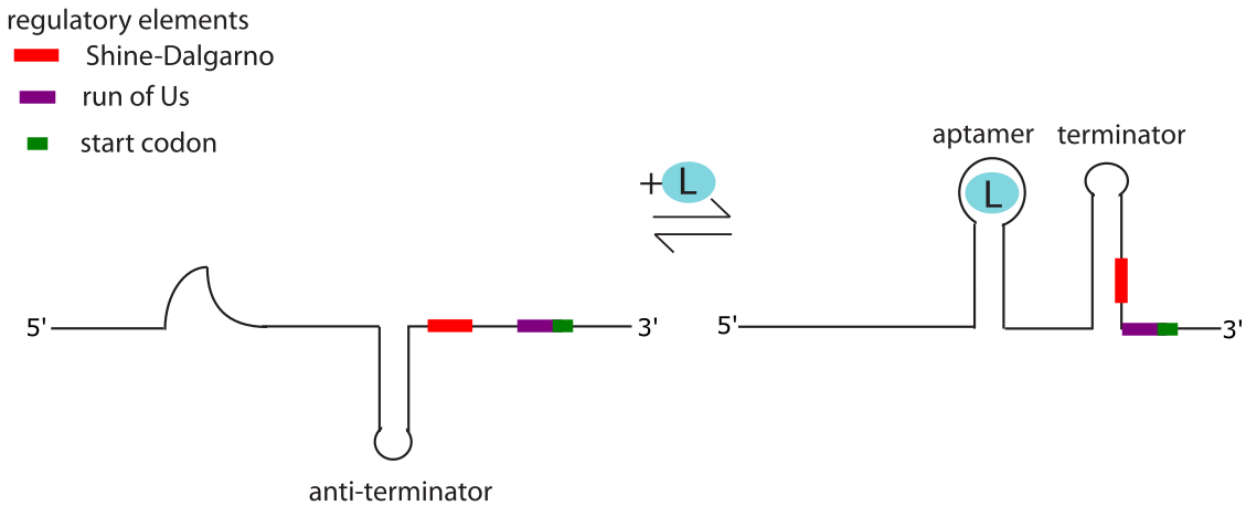


Figure 3.1 Riboswitch mechanisms. Riboswitches can alter levels of transcription or translation based on ligand binding. Switches often depend on the formation of a hairpin that sequesters or releases regulatory elements for expression in response to ligand binding. The most common regulatory elements for translation are dependent on the availability of a Shine-Dalgarno sequence or a start codon. Transcriptional switches in bacteria often contain a run of U residues directly downstream of a terminator hairpin that stalls and releases the polymerase.

3.4 Synthetic riboswitches

A vast array of methods has been employed to evolve and engineer RNA in the lab with the objective of riboswitch discovery. Polisky et al. discovered a theophylline aptamer with a K_D of 0.1 μM capable of discriminating between theophylline and structurally similar caffeine[16]. Interestingly, Gallivan and coworkers successfully integrated the theophylline aptamer into a theophylline-dependent translation riboswitch[17]. This example showed how aptamers can be combined with the aforementioned expression elements to deliver working riboswitches. The theophylline aptamer was cloned 5 base pairs upstream of the ribosome-binding site of a β-galactosidase reporter gene. Galactosidase expression was monitored by blue/white screening of colonies. Expression proved to be dependent on theophylline: presence of caffeine or no ligand resulted in low expression, but presence of theophylline resulted in robust 11-fold increase in activity. Since the discovery of the initial riboswitch, the theophylline riboswitch has been further evolved to utilize different methods of regulation which has improved binding and function. For example, the Mörl lab engineered a theophylline-dependent transcriptional riboswitch by placing the theophylline aptamer upstream of a terminator, resulting in a 6.5 fold increase in expression of reporter genes in *E. coli*[18]. The theophylline aptamer demonstrates that a strong aptamer domain can be engineered into a riboswitch when coupled with an expression domain.

Natural riboswitches have been discovered for a variety of ligands such as adenosylcobalamin[14], thiamine pyrophosphate (TPP)[7], flavin[7], guanine[19], adenine[19], glutamine[20], glycine[21], s-adenosylmethionine[22], lysine[23], and glucosamine-6-phosphate[24] . Binding of these ligands results in the regulation of nearby gene expression, and consequently general metabolism. Some of these riboswitches have been re-engineered and

selected in cells for enhanced binding or functionality. For example Nomura and Yokobayashi reselected the TPP aptamer by a dual genetic selection in plasmids, based on fluorescence and tetracycline resistance selection, to discover a library of artificial riboswitches[25]. Selection was achieved in bacteria by expression of the tetracycline resistant gene (*tetR*) in the "ON" state in the presence of tetracycline. Bacteria that survived the "ON" state were incubated in the presence of Ni^{2+} to screen for the "OFF" state. Ni^{2+} is deleterious to expression of tetracycline resistance. After several rounds of selection, clones were screened by GFP fluorescence, resulting in riboswitches with improvement in both enhanced and suppressed expression in the presence of the ligand. The dual selection design of the plasmid facilitates development of a riboswitch by survival, fluorescence, or both.

In vivo screening of riboswitches is constrained to the specific environment of the organism used, limiting diversity and methods of regulation. *In vitro* assays are advantageous for screening; large starting pools can be used, so evolution cycles occur quicker as they are not reliant on cell growth. In one example Ellington and Mansy developed a strand displacement selection to discover and screen new TPP riboswitches[26]. The method is based on availability of the expression domain and its ability to undergo a conformational change in response to ligand binding. A toehold that is composed of a longer reporter sequence, and a smaller sequence that binds to the reporter, was used to test the availability of an expression domain. If the expression domain is available, it will bind the reporter by displacing the shorter sequence. Screening and selection was achieved by strand displacement in a ligand-dependent fashion. The toehold can be designed to be complementary to any sequence. For riboswitch discovery, it is best to use an aptamer or expression domain. This selection scheme makes it possible to select riboswitches for unique targets, based on synthetic libraries, that

can be coupled with the *in vivo* selection systems to facilitate laboratory evolution of robust synthetic riboswitches

3.5 Stilbene photoisomerization and fluorescence

Targets with unique optical properties are of interest as they add functionality to the riboswitch. Stilbenes are a group of photo-switchable molecules that are capable of undergoing reversible *cis* to *trans* isomerization based upon electronic excitation[27]. Stilbenes, also known as diarylethene compounds, are identified by having flanking aromatic groups across an ethylene double bond. Reversible isomerization can occur under the influence of light, and non-reversible isomerization can be induced by heat. Photo-cyclization can occur by light or oxidation, resulting in an additional ring[27, 28]. The ability of these compounds to easily undergo conformational changes has led to their extensive use as molecular probes and switches. In addition to their capability to act as switches, stilbenes have dynamic absorption, fluorescence, and phosphorescence activity based on the multiple excited and ground states of the *cis* and *trans* isomers, making them desirable tools for biological applications[27, 28]. Recent examples utilizing stilbenes as a photoswitch or fluorescent probe include: incorporation of stilbenes into peptides to isomerize the stilbene and change the conformational dynamics of peptide folding[29], aptamer stilbene-switch assays for identifying novel aptamers[30], and incorporation into antibody assays for identifying new binding targets by stilbene fluoresence[31]. These assays demonstrate the wide range of functionality of stilbenes due to their optical qualities.

Isomerization of stilbenes is possible due to the orientation of the stilbene in its excited state. In its excited state the diarylethene double bond is reduced, resulting in several possible orientations of the groups flanking it. Excitation can result in either excited singlet (S1) or an intersystem crosslink into the triplet state (T1), resulting in fluorescence or

phosphorescence, respectively. Both singlet and triplet excited states can result in isomerization; however, this is a competitive process with radiative emission (fluorescence or phosphorescence)[27]. The energy barrier for the *trans* state is much lower than for the *cis*, and the *trans* state results in higher yield under relaxation to the ground state[28]. Relaxation to the ground state results in a mixture of *cis* and *trans* isomers, typically with a higher yield of *trans* isomers. Tuning between isomers can be achieved by utilizing different excitation methods that favor the desired isomer state. *Cis* and *trans* states have different absorption coefficients, making it possible to favor the isomerization state by choosing a specific wavelength[2]. Slowing down or preventing photoisomerization will result in energy released in the form of luminescence.

Each stilbene has unique isomerization profiles depending on functional groups, solvent conditions, and excitation methods, all of which greatly influence the isomerization potential. Functional groups influence isomerization due to changes in the transition energy and orientation of the molecule[27]. The solvent system can alter isomerization dynamics due to its friction or viscosity, which impact the ability to rotate about the diarylethenebond[27]. *Cis* orientation of stilbenes is sterically hindered, rendering it less stable than the *trans* state[2]. Recovery or switching from *cis* to *trans* due to light occurs with a much greater yield than *trans* to *cis*, and the abundance of switching is highly variable depending on the derivative of stilbene used and wavelength of light used to excite the molecule[2]. In fact, *Cis* to *trans* conversions have been measured at greater than 90 %, meanwhile the maximum observed conversion measured for *trans* to *cis* conversion is 70 % under optimal conditions. We used a stiff stilbene, (E)-6'-(2-aminoethoxy)-2,2',3,3'tetrahydro-[1,1'-biindenylidene]-6-ol (*trans*-**1**), as the target of an *in vitro* selection (Fig 3.2). The *trans* version of the stilbene is more desirable

for systems where reversible switching is necessary, as switching back to *trans* is more accomplishable.

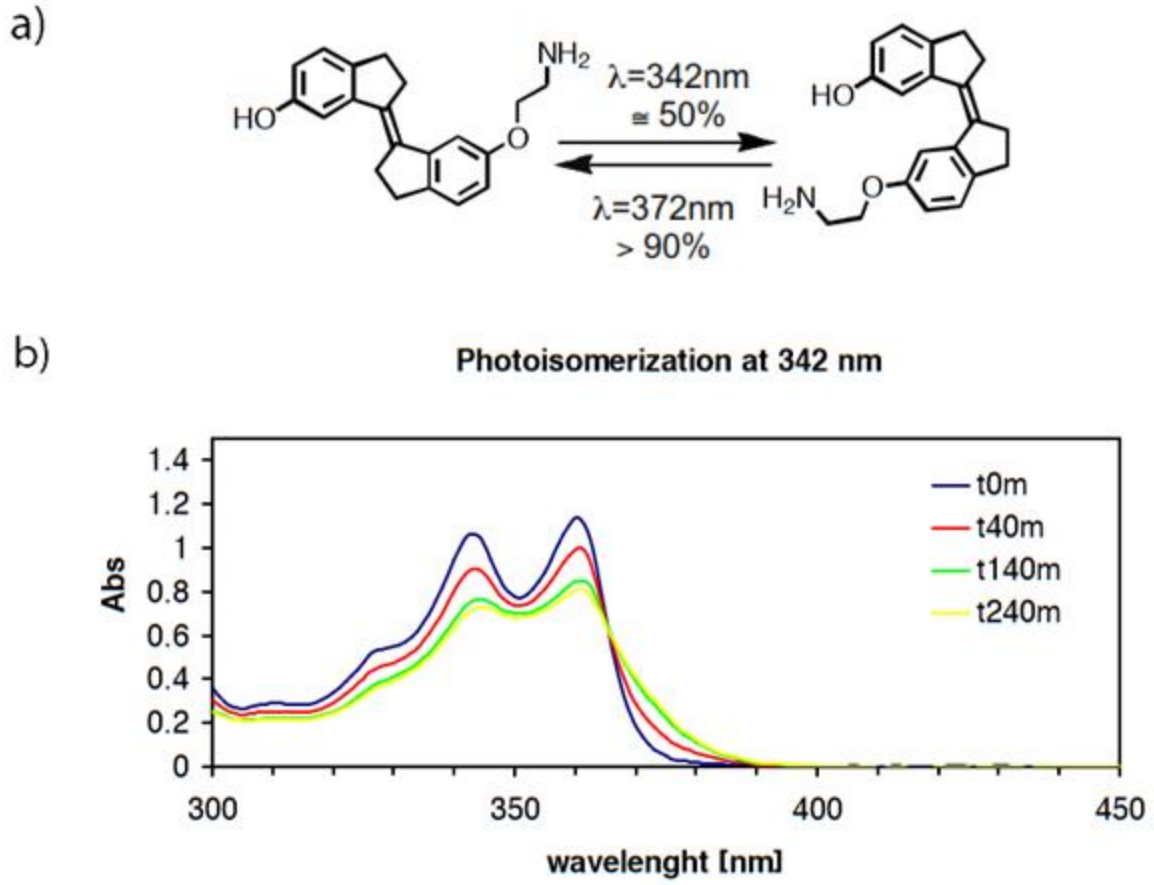


Figure 3.2 Stilbene isomerization. (a) Isomerization of ligand used for selection, the extent (percent) of isomerization is dependent on the wavelength used to excite the stilbene. (b) Rate of absorbance of stilbene undergoing photoisomerization at 342 nm.

3.6 Selection for a stilbene riboswitch

Regulation of gene expression is extensively used in biological research to uncover the roles and mechanisms of genes, and as a result, many techniques have been developed to facilitate expression studies. Gene expression studies often use an inducible or repressible system in order to observe the specific effect of a gene or functional code. Important considerations for expression systems are timescale of activation, specificity of molecules in the system, mode of expression, and bioorthogonality of the system. In bacteria, regulation is often achieved by inducible expressions of operons, with the isopropyl-β-d-thiogalactopyranoside (IPTG)-inducible *lac* operon being perhaps the most widely used tool in genetic studies to induce protein expression[32]. Another widely used inducible system uses the *Tet* operon, which has a dynamic range of expression. In fact, the operon has been demonstrated to increase expression up to 5000-fold[33]. These operators work on the level of transcription to control protein induction. They are not rapid in response, working on a minute timescale as induction is dependent on the amount of time for the inducer to bind to the repressor[33]. The method of induction and expression are fixed in these systems and are not ideal for studies which may require regulation at a specific level of expression (transcription, translation, or splicing), or which require reversible expression. An ideal system of regulation should also be easy to engineer in different types of cells or systems, be able to utilize a cell permeable ligand that does not alter the natural metabolism of the cell, and be rapidly induced. The necessity to tightly regulate and tune expression has resulted in new efforts to create or engineer applicable genetic expression systems.

The ability to fine tune expression with a ligand makes riboswitches a lucrative tool in developing new molecular methods for expression and feedback. Tuneable expression systems are being developed to to improve and expand method of gene regulation. Logic

gates and cell circuits tune expression by responding to multiple inputs to flip or reverse the state of expression. Siuti, Yazkek, and Lu have developed circuits based on recombinase enzymes reversing the state of a DNA terminator and coupled them with GFP feedback in order to adjust and record expression[34]. Caged molecules serve as another mechanism to control expression based being able to "uncage," or release of a molecule by changing the state of the molecule, often by chemical modification or removal of a protecting group. Caged molecules are often activated by light resulting in quick activation of expression; however, the uncaged ligand is usually unable to revert back to its caged state[35, 36]. Riboswitches have the advantage of being selected or engineered for specific ligands and control expression on the level of either/both transcription and translation, tunable expression can be achieved by the target of the aptamer platform.

Many aptamers have been developed to induce fluorescence upon binding, demonstrating that aptamers can provide direct feedback of expression. The Spinach aptamer was discovered by selecting against several derivatives of 4-hydroxybenzlidene imidazolinone (HBI), mimicking the catalytic core of GFP. Fluorescence is greatly enhanced in the presence of the aptamer[37]. The Mango aptamer binds thiazole orange and enhances the emission up to 1000-fold, and has been successfully used for RNA labeling[38]. These aptamers have been used to detect RNA *in vitro*, observe localization in cells, and tag RNA expression with nanomolar affinity[37, 38]. The enhancement of fluorescence by aptamers has been proven to be functional in cells, and discovery of aptamers for additional optically interesting and active probes can be engineered or selected for riboswitch activity resulting in feedback in response to modulating expression.

Another challenge in tuning gene expression is the ability to reverse it. Riboswitches are in a state of equilibrium which favors one state of expression when bound to a ligand, and

they depend on ligand concentration to determine their expression state. In order to reverse expression, a second switching mechanism is necessary for changing the state of the ligand or riboswitch. pH and temperature riboswitch sensors could be developed, but may not be feasible for many *in vivo* experiments given that they are restricted to a specific physiological environment. To reverse the state of a riboswitch once the ligand is bound, it would be beneficial for the ligand to carry out the functionality of switching. This gives the option of tuning expression with the ligand directly, then reversing the bound state by switching the ligand. Stilbenes are largely utilized in switching mechanisms based on the photo-reversibility between *cis* and *trans* states. A stilbene riboswitch could serve as a molecular sensor by exploiting any change in fluorescence due to RNA binding. Isomerization of the stilbene could be used to release the riboswitch and the state of expression.

3.7 Pool design

The higher the diversity in a pool, the greater the likelihood of positive outcomes for a selection experiment[39]. A synthetic pool can be designed to maximize positive outcomes based on diversity of the pool. The diversity of a synthetic pool can theoretically be as high as 4^N, where N represents the length of random region in a pool. Due to experimental constraints (amount of reagents, cost of synthesis, and size of reactions), a synthetic pool was created with an approximate diversity of 10^{15}. The design of the pool was based on the *S*-adenosylmethionine (SAM) riboswitch discovered in *Bacillus subtillis,* in order for the pool to have a fixed expression domain.

The SAM-1 riboswitch regulates transcription by inducing termination when SAM is bound to the aptamer domain. The riboswitch consists of an aptamer domain specific to SAM, and a downstream expression platform that forms a terminator containing a run of Us that destabilizes the polymerase and terminates transcription[22]. In between the aptamer

domain and the terminator is an anti-terminator, a sequence that is complementary to the terminator and prevents the terminator from forming in the absence of SAM. The SAM riboswitch was an ideal candidate *in vivo* and *in vitro*, as it was demonstrated to terminate in bacteria with native *B. subtilis* polymerase *in vivo* and T7 polymerase *in vitro.*

In order to transcribe the pool *in vitro*, the pool has a T7 promoter followed by a random region of 45 nucleotides that makes up the potential aptamer region. Directly downstream of the aptamer region is an anti-terminator followed by a terminator loosely based on the SAM-1 riboswitch, which provides the pool with the ability to switch. The terminator contained a Shine-Dalgarno (SD) ribosome-binding site, followed by a UUG start codon in order to regulate translation initiation, features not included in the original SAM-1 riboswitch. Additional differences to the SAM-1 design include randomization of the loop of the terminator to allow the formation of potential psuedoknots and a partially randomized terminator (10% per position) to select for variant switches. The terminator also contains a run of Us, which typically stall the RNA polymerase and terminate transcription if the RNA of interest is immediately downstream of the termination helix. The aptamer platform was designed to directly precede the anti-terminator, in order to be able to change the availability of the anti-terminator depending on the state of the aptamer.

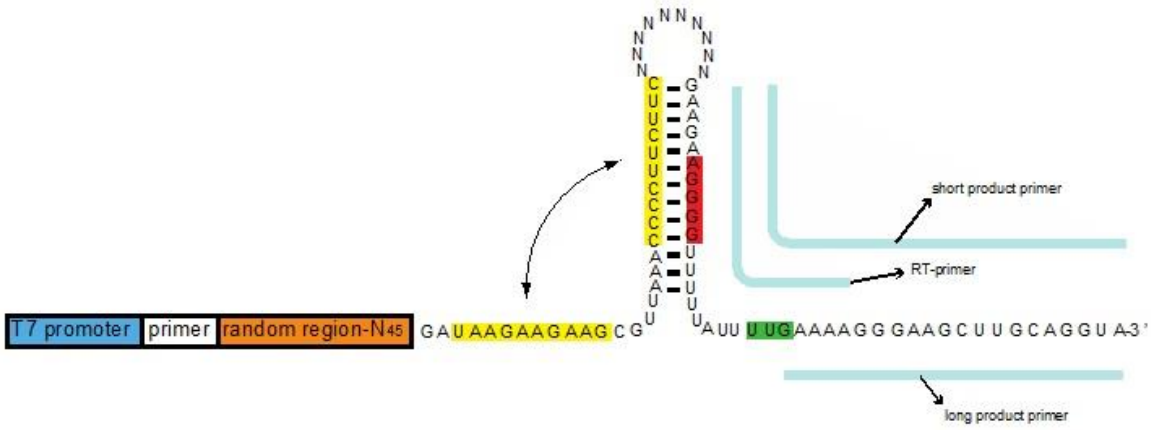


Figure 3.3 Design of pool. Pool used for riboswitch selection is loosely based on SAM-1 riboswitch. A 45-nucleotide region (orange) makes up the potential aptamer domain. Downstream of the aptamer domain is a terminator containing several regulatory elements: Shine-Dalgarno sequence (red), start codon (green), and set of Us. Between the aptamer domain and terminator is a linker complementary to the terminator, potentially forming an anti-terminator.

3.8 Methods for selecting a riboswitch (methods 6.9)

Affinity selection

Selecting a synthetic riboswitch for a new ligand poses two distinct objectives that must be achieved: selecting for affinity (ligand binding) and selecting for function (conformational switching and altering of the expression domain). Selecting for affinity and designing a potential switchable expression platform are not difficult on their own; however, coupling the functions with it is more challenging. Many methods are currently being developed for the selection and genetic screening of riboswitches; however, no current methods exist to discover an aptamer and riboswitch in a single experiment. Several selection schemes have been developed and adapted during the course of the overall selection in order to select for an expression platform that is responsive to a stilbene aptamer. The initial *in vitro* selection experiments were affinity selections developed by incubating a random pool with the ligand on an immobile surface (generally beads), and performing serial rounds of amplification to select for aptamers specific to the ligand[3, 4, 6].

An affinity selection satisfies the aptamer portion of a riboswitch selection, but may not have any implications on the state of the expression platform. An affinity selection was performed with the *trans* stilbene on beads. Pools were incubated on beads with the *trans* stilbene, washed several times with buffers mimicking physiological conditions, and collected by denaturing the pool off the beads with urea followed by reverse transcription and amplification. Binding was enhanced from 2 % to ~ 14 % after six rounds of selection.

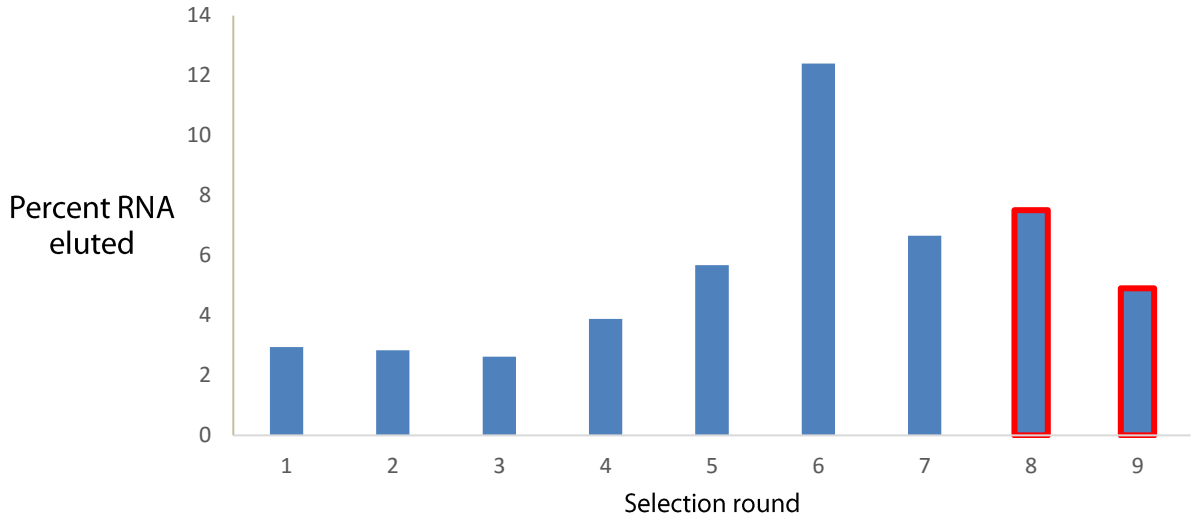


Figure 3.4 Trans-1 affinity selection results. Enrichment is observed after round 4 of selection, peaking at round 6. Red outlines indicate increased stringency by using half the amount of targets on beads. Rounds 6-9 were cloned and screened for potential aptamer or riboswitch activity.

Selection by photo-switching

A selection was performed by binding the RNA pool to the *cis*-**1** ligand on beads followed by photoisomerization of the *cis*-**1** ligand to *trans* in order to release the RNA. The selection scheme was designed to achieve two goals: select for binding and affinity for one state of the ligand (*cis*), and select for an RNA that is released when the ligand is isomerized to the opposite state (*trans*) on beads. To achieve this, the RNA was incubated on beads and washed with buffer to remove non-binders. To select for the function of the release of the pool off the ligand, the pool was eluted with light, rather than introducing free-ligand or using denaturing conditions. RNA was eluted off the *cis* beads with 365 nm of light for 30 seconds. Then, RNA that came off the beads and into solution was collected for serial rounds of selection.

In vitro selection by ligand-dependent transcription termination (methods 6.11)

In order to find a riboswitch based on transcription termination we created a selection method of co-transcription in the presence of *trans*-**1** *in vitro*. The pool consists of a T7 promoter and could be transcribed by T7 polymerase *in vitro*. The experiment was designed based on the hypothesis that if the riboswitch functions as a regulator of transcription, there should be a difference in the abundance of the pool transcribed according to length of transcripts in the presence of ligand. Transcriptional termination in riboswitches generally occurs after a run of Us following a stem that destabilizes the polymerase. With this in mind, the pool was designed with 20 nucleotides following the run of Us of the terminator; this would result in a construct that was shorter by ~ 20 nucleotides if the polymerase stalled or dissociated from the terminator motif. Two scenarios are possible: a ligand-based "ON" selection that produces a full transcript or a ligand-based "OFF" selection that results in termination. For example, if riboswitch expression is turned on by the ligand then in the

presence of ligand the product should be longer, as the polymerase should read through the potential terminator. Terminated and non-terminated products can be separated and purified by denaturing PAGE analysis, allowing for a selection based on the product size of transcription products from the pool.

The selection was designed to promote termination in the presence of the ligand and produce full length transcripts in the absence of the ligand. A single round consists of co-transcription with the ligand to cut out terminated products, followed by transcription without the ligand to collect full length products. Products are purified by denaturing polyacrylamide (PAGE) after every cycle of transcription to cut out full length product (~145 nucleotides) or terminated product (~ 120 nucleotides). After several rounds of selection, enhancement of termination was observed by a higher ratio of terminated products, but ligand dependence has not been accomplished by this method. Later rounds were cloned, with individual clones checked for ligand based termination by PAGE analysis.

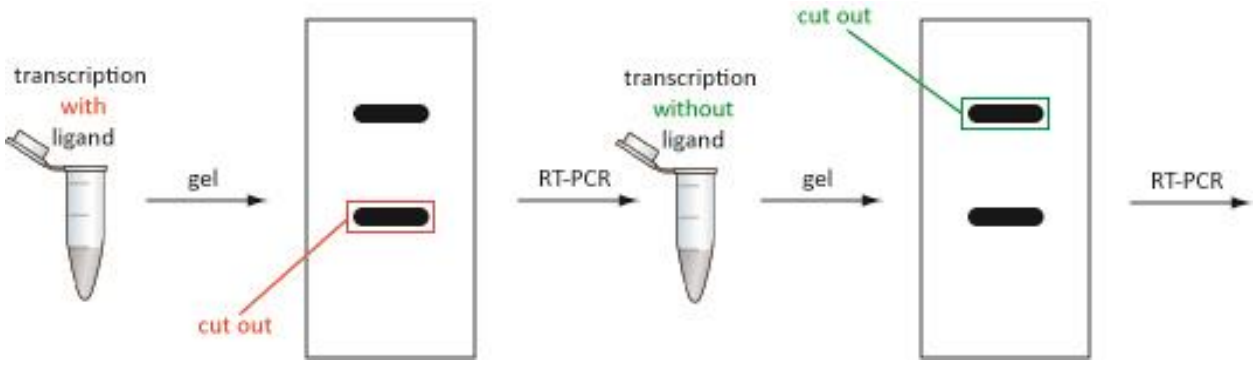


Figure 3.5 Termination based selection scheme. Pool products are cut out of denaturing PAGE gel after transcription according to the size of the pool. In the presence of ligand a terminated or shorter product is collected, followed by a subsequent round without ligand where the larger product is cut out.

Riboswitch selection by strand displacement (methods 6.16)

In order to discover riboswitches by selection and screening *in vitro* the Mansy lab created a selection scheme based on strand displacement[26]. The scheme takes advantage of the structure=al rearrangement of the expression domains found in riboswitches. They used a library based off a known riboswitch for thiamine pyrophosphate (TPP) to select for TPP riboswitches with enhanced activity. A double stranded DNA reporter containing a sequence that is complementary to the terminator was used to capture the riboswitch in a toehold, in the presence or absence of ligand. The reporter contains a partially double stranded toehold that must be displaced by the RNA pool itself in order to bind. The double stranded toehold is made up of a longer reporter strand that is modified with biotin or a fluorophore and a shorter strand that is displaced by the pool or clone of interest. Affinity to the reporter must be strong enough to displace the shorter strand and bind the toehold. The displacement is dependent on the availability of the expression platform. If the expression platform is released, the riboswitch can bind to the reporter while the supernatant contains the RNA that does not have the terminator released[26]. One can select for either the RNA bound to the reporter or the RNA in the supernatant based on the "ON" and 'OFF" state with respect to binding of the ligand. For example, to select for an "ON" riboswitch when the ligand is bound, the terminator should be released and the portion of the pool bound to the DNA reporter should be collected. The biotinylated reporter and can be collected using streptavidin magnetic beads. This scheme was used to yield TPP riboswitches that were active *in vitro* after only three rounds of selection with a 7.3 fold increase in expression over the original riboswitch[26].

The strand displacement method was adopted to select for a *trans*-**1**riboswitch. Because this method selects for function, and not binding, we used the pool that was

previously selected for affinity of *trans*-**1**. A biotinylated DNA reporter complementary to the terminator of the pool was designed to create a toe-hold when the terminator is available or released. The pool is incubated with the DNA reporter, the portion of the pool capable of strand displacement, and binding to the reporter is collected with streptavidin beads and discarded. This step is repeated several times to ensure that the RNA remaining has low affinity for the toehold or is in a terminated state when no ligand is around. *Trans*-**1** is added to select for RNA that switches in the presence of the ligand and is now able to bind the reporter. Sequences bound to the reporter are collected for the next round of selection. 8 rounds of this selection scheme were performed starting with 10 μM of ligand for 6 rounds followed by lowering the concentration of ligand to 1 μM for the final two rounds of selection. Displacement enhancement was monitored by radioactivity and after 6 rounds with 10 μM selections final binding to the reporter increased from 3 % to 50 % binding.

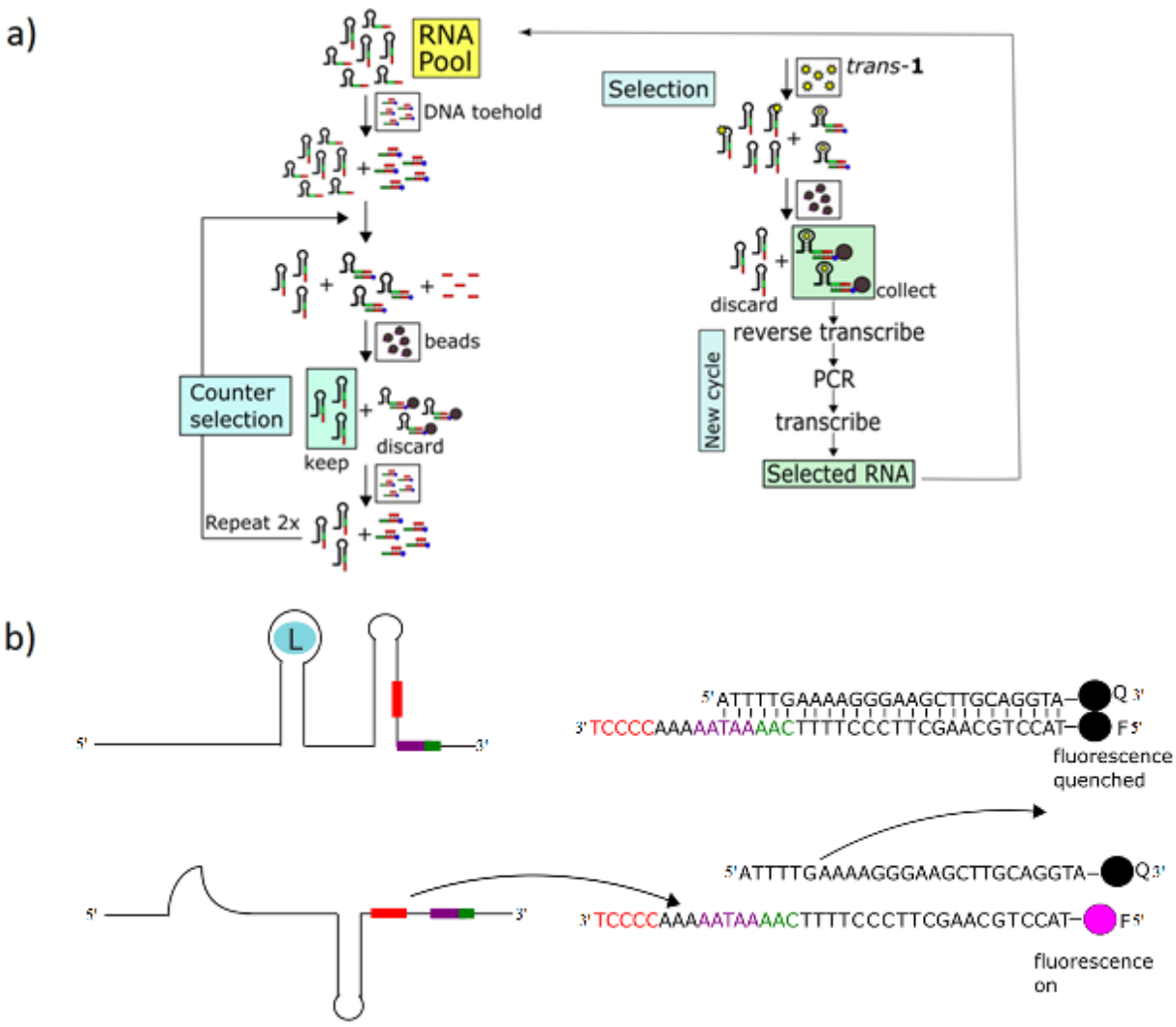


Figure 3.6 Strand displacement scheme. (a)Strand displacement based on availability of expression platform to displace a toehold. RNAs capable of binding the toehold without ligand are discarded to remove pools with available expression platforms in absence of ligand. Toehold is collected in the presence of ligand to select for ligand-dependent strand displacement. (b) Strand displacement can be used to screen potential clones by fluorescence. The toehold reporter strand contains a fluorophore and the commentary strand contains a quencher. Displacement of quencher by binding of the pool or clone results in enhance fluorescence.

3.9 Methods for screening potential aptamers and riboswitches

Screening clones by *in vitro* transcription termination

Pools from all previously mentioned selections were cloned in order to identify and screen unique sequences. A riboswitch that terminates transcription should produce fewer full length transcripts when in a terminated state. This would correspond to higher abundance of premature stops and short transcripts being produced during transcription (as described in 3.7). *In vitro* transcription followed by PAGE analysis was performed in order to characterize and visualize transcription products from individual clones. Clones were transcribed with radiolabeled ^{32}P-α-ATP in the presence and absence of ligand. Transcripts were fractionated on a 15 % denaturing PAGE gel to determine the overall yield of full-length product in comparison to terminated products. A typical run consisted of transcription for 30 minutes at 37 ˚C. Differences in total and terminated products were quantified and plotted in order to measure differences in riboswitch activity in response to the ligand.

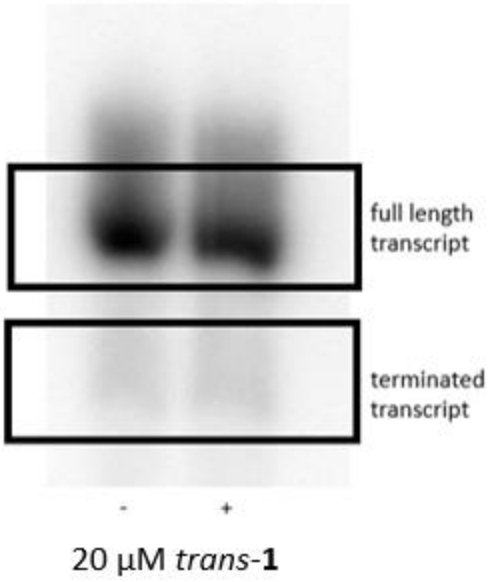


Figure 3.7 Termination assay of clone 36. Individual clones are tested for termination by *in vitro* transcription in the presence or absence of ligand. Full length and terminated products can be observed on denaturing PAGE gel and analyzed for percentage of termination.

In order to observe the optimal conditions for riboswitch activity, temperature and kinetic analyses were performed for clones that had promising activity, and clone 36 was the most extensively studied. Clones were transcribed at various temperatures ranging from 16-42 ˚C in the presence and absence of *trans*-**1**. Additionally, kinetic studies were carried out on a 10-300 second time scale to observe how transcription initiation, elongation, and termination were affected when co-transcribed with *trans*-**1**. Yield of products (full length and premature stops) was plotted against time to determine transcription rates. Slight differences in terminated products could be observed for clone 36 in the presence and absence of *trans*-**1** from both temperature and kinetic studies; however, the overall effect was not substantial enough for applications in gene expression regulation.

SHAPE (selective 2′-hydroxyl acylation analyzed by primer extension) analysis (methods 6.12)

Selective 2′-hydroxyl acylation analyzed by primer extension (SHAPE) is a structural probing technique that can provide insight on more flexible and solvent accessible regions of RNA. SHAPE reagents acylate the backbone of more structurally flexible and accessible regions of RNA, thus providing structural insight on a single nucleotide level with respect to changes of RNA confirmation[40, 41]. Reverse transcription with a radioactive end labeled primer is performed to analyze stops of primer extension due to acylation. Extension reactions can be franctionated by denaturing PAGE to determine differences in flexible regions upon ligand binding. SHAPE of individual clones in the presence of a dilution series of *trans*-**1**were completed for several clones. The clone with the most promising binding activity based on SHAPE analysis was clone 36.

Clone 36 was incubated with 20 nM to 20 μM *trans*-**1**. Changes in acylation can be observed in several regions of the RNA in a ligand-dependent fashion, corresponding to

binding. Clone 36 has an alternate UUG start codon. The UU of the start codon shows reactivity trends in response to the ligand that may change the availability of the start codon. The band for the first U becomes faint with increasing concentration of stilbene, correlating with the nucleotide becoming less available. The band for the second U grows in intensity, corresponding to the nucleotide becoming more flexible. Portions of the terminator become both more and less accessible as the RNA structure undergoes rearrangement in response to the ligand (Fig 3.8.). This is a strong indicator that the RNA binds to the ligand but has no bearing on the RNAs capability to function as a riboswitch. SHAPE analysis is useful for finding individual RNAs with affinity for the ligand but further analysis must be carried out to test or select for function, the data are sufficient to suggest that the sequence is an aptamer for the stilbene.

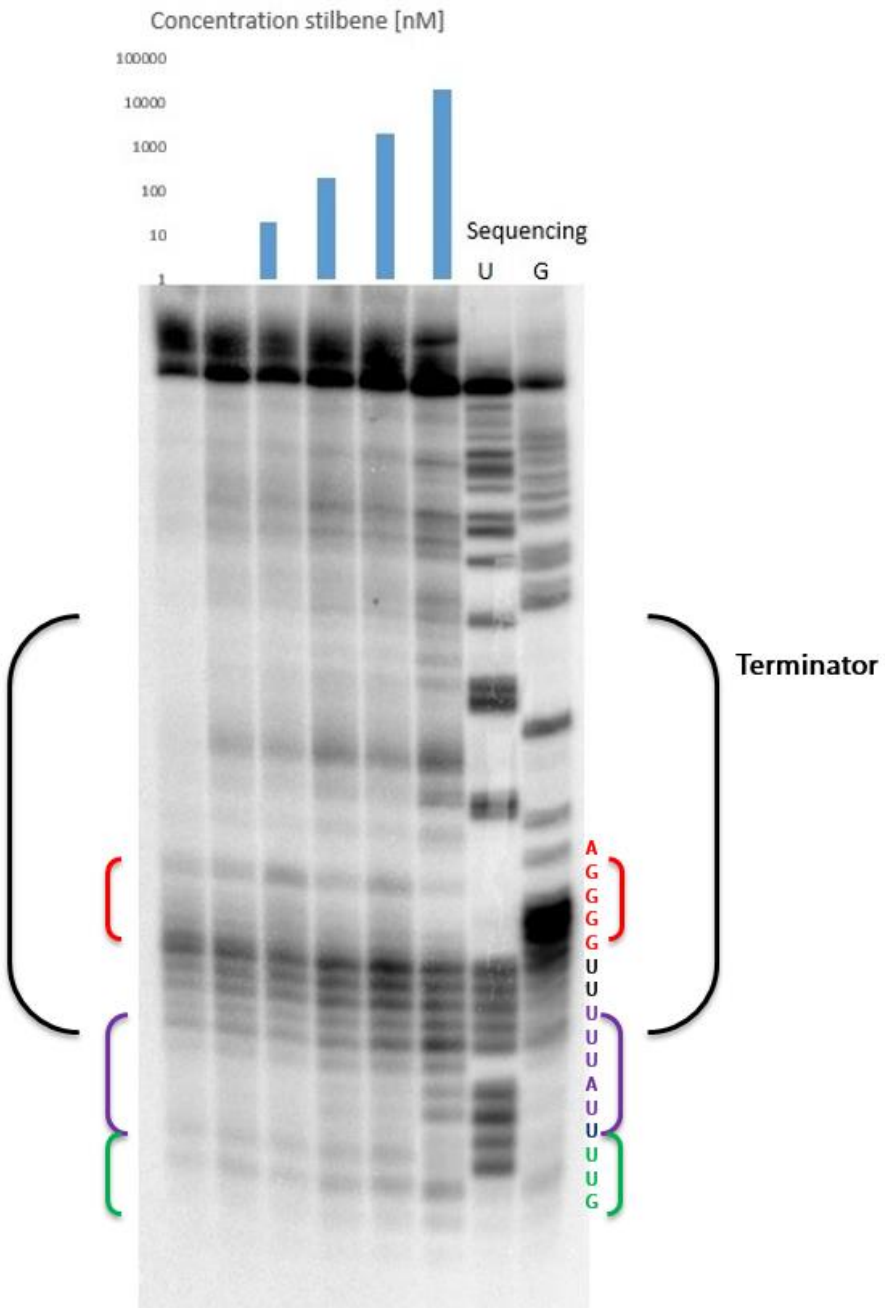


Figure 3.8 SHAPE analysis of clone 36 (method 6.12). Lane 1 control without ligand or SHAPE reagent, lane 2-6 titration of stilbene in the presence of SHAPE reagent for 30 minutes. Brackets and sequence indicate regulatory elements. Red is a Shine-Dalgarno sequence, purple is run of Us in the terminator stem, and green indicates (an alternate UUG) start codon, black bracket is the predicted terminator.

Clone 36 alternate structure design

Positive SHAPE profiles of clone 36 led us to further screen and analyze the sequence (section 3.8). Alternate structures of clone 36 were created in order to enhance function or binding from the sequence that came out of the selection. Variants were created by making several truncated versions of the clone or using primers to introduce mutations to different regions of the aptamers. These variants were tested by the previously mentioned screening methods; however, no functionality was gained.

Structure prediction for clones

A necessary component for a successful selection of a transcriptional riboswitch is the ability for the riboswitch to form a terminator. Structures of potential riboswitches can be predicted using RNAfold based on minimum free energy. The affinity selection provided several sequences that underwent fold prediction. Most important to our search is clone 36 due to positive results. Fold predictions for clone 36 suggested that it has a large portion of the terminator unformed and base-pairs with the 5′ end of the RNA rather than forming the termination stem loop (Fig 3.9. left). A large portion of the anti-terminator is structured in two different stem loops, one of which partially hybridized with a portion of the terminator, preventing the terminator from forming. If the anti-terminator is released in response to the binding *trans*-**1,** the terminator may be available to form, resulting in a conformational switch. To increase the reliability of the structure predictions, we also included structural constraints derived from the SHAPE data. Predictions were also investigated using constraints based on SHAPE data for clone 36. Predictions with SHAPE constraints still contain a large portion of the terminator unformed, although the anti-terminator is less strongly predicted (Fig 3.9. right). A weaker anti-terminator and other structural rearrangement predictions in the presence of the ligand indicate the potential for a switch in structure upon aptamer binding.

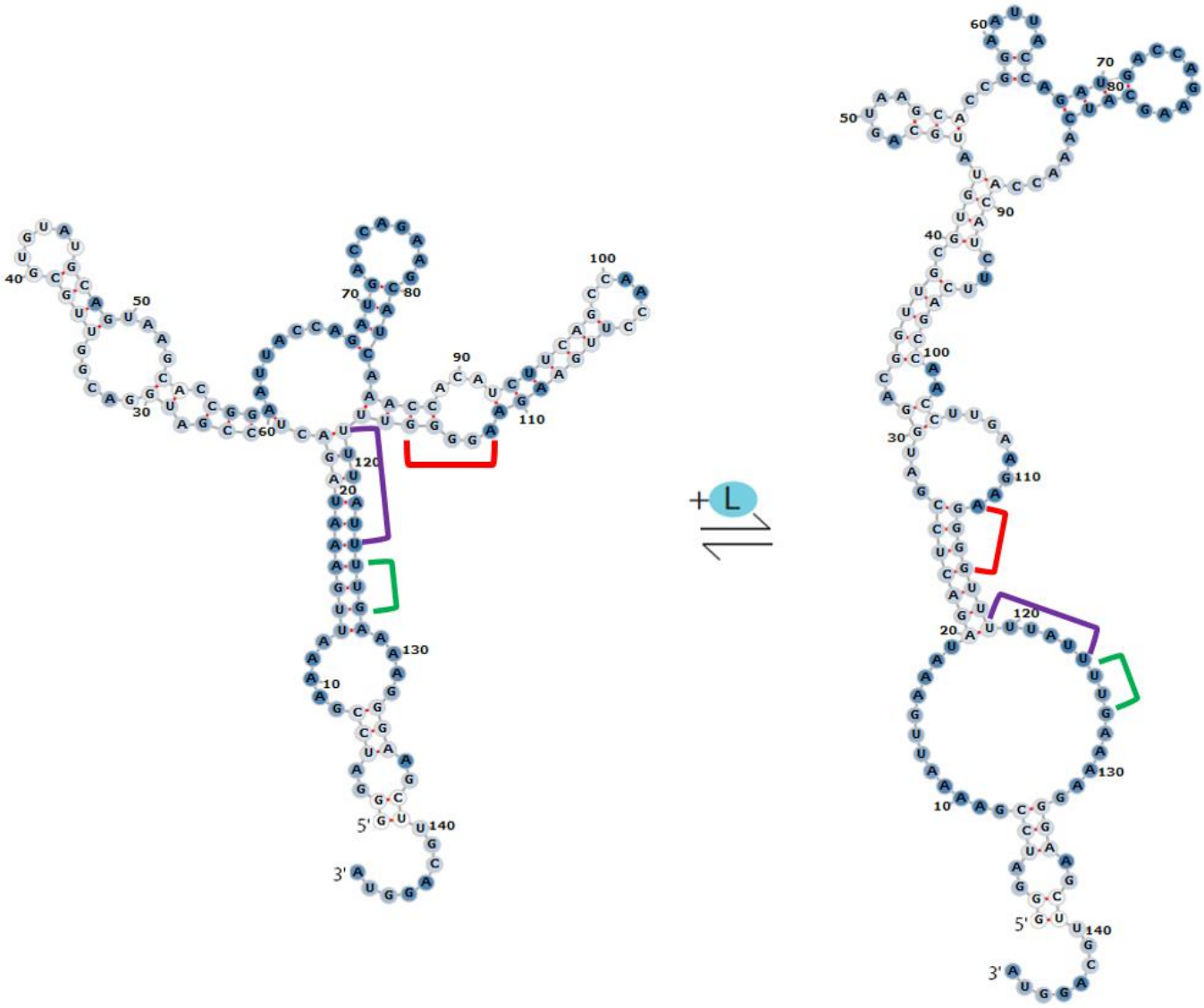


Figure 3.9 Structure predictions of clone 36 (methods 6.13)**.** Prediction of clone 36 with no ligand (left) and prediction of clone 36 with 10 µM ligand according to constraints based on SHAPE analysis (right). Base-pair probabilities are given by intensity of color (blue). Brackets indicate regulatory elements. Red is a Shine-Dalgarno sequence, purple is a run of Us in the terminator stem, and green indicates (an alternate UUG) start codon.

Screening based on optical activity (methods 6.15)

Excitation under UV light causes both isomerization and fluorescence of stilbenes. The *trans*-**1** stilbene used in our selection exhibits fluorescence in range of 400-450 nm, giving off a blue light when excited with UV light. RNA binding to the stilbene may change its electronic transition to its excited or ground state due to restricting or enhancing isomerization. Specific binding of RNA could alter the isomerization rate or change the fluorescence properties of the stilbene. Radiative and nonradiative competition of the excited state are susceptible to alteration based on the environment and interactions of the stilbene. Change in optical or switching properties due to binding can be directly monitored by absorbance and fluorescence, each state of the ligand has a unique absorption, excitation and emission spectrum. To analyze results of the selection, excitation and emission spectra were taken with ligand in the presence of individual clones to observe any change in optical activity from ligand alone. Over 60 clones were tested for optical activity, and surprisingly, every clone resulted in enhancement of emission from the stilbene corresponding to fluorescence simply based on RNA concentration rather than specificity.

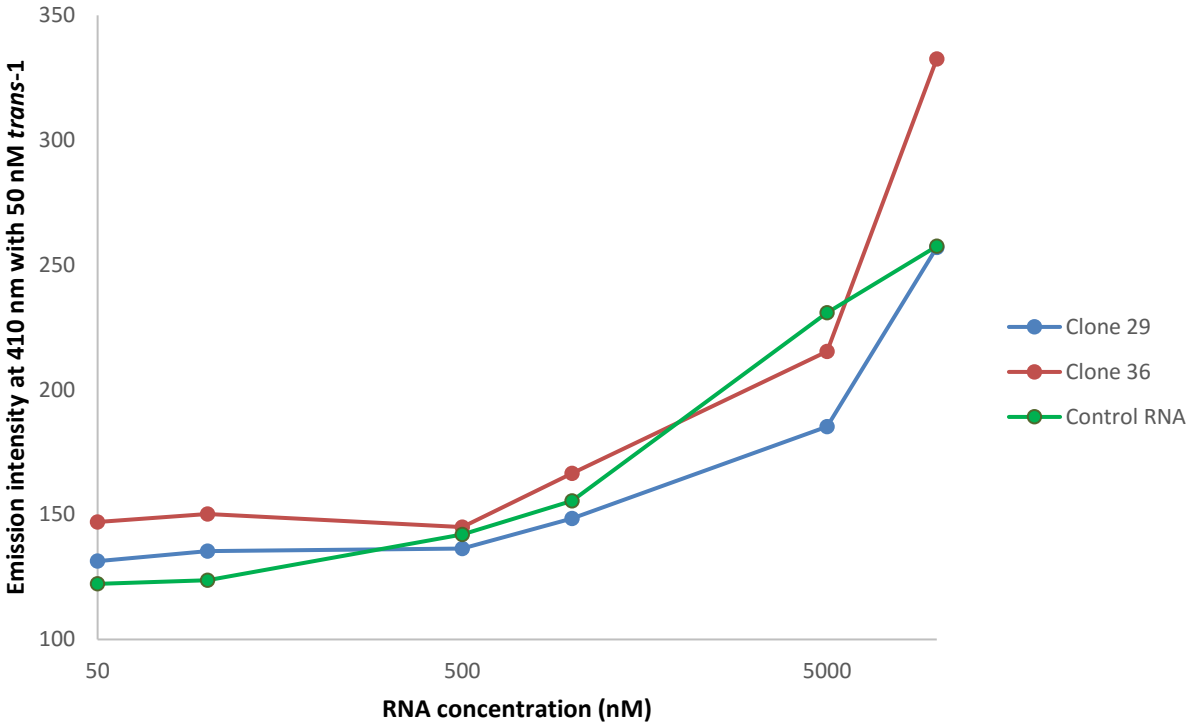


Figure 3.10 Emission of clones 36 and 29 in the presence of stilbene (methods 6.15). Several clones were tested for optical properties. The top two clones from initial fluorescence screening, 36 (red) and 29 (blue) were incubated with a range of concentrations of RNA from 50 nM to 10 μM and tested for emission with 50 nM *trans*-**1** (excitation 355 nm). A control RNA (green) from a selection for another target also showed enhancement of emission.

Some clones showed a much higher enhancement of fluorescence compared to other clones under the same conditions, indicating these clones might have a higher affinity or specificity for *trans*-**1**. Clones were tested over a nM - μM concentration range to determine whether the enhancement of fluorescence was due to binding specificity. A titration series of RNA for each clone ranging from 50 nM to 1 μM was tested with 50 nM of *trans*-**1** (Fig 3.10.). Fluorescence enhancement of *trans*-**1** increased with greater concentrations of RNA, confirming RNA-dependent fluorescence of the stilbene. Some clones exhibited enhanced emission response compared to others; however, subsequent tests have yielded inconsistent results. Another interesting observation from fluorescence screening was a shift in the emission maxima for several clones. It is unknown if this corresponds to isomerization, decaying, or binding of the stilbene.

Screening by plasmid expression (methods 6.14)

The main function of the riboswitch is to be able to reversibly turn on or off gene expression downstream of the aptamer domain in response to *trans*-**1**. Pools that were affinity selected and selected by strand displacement were cloned in plasmids that express RFP or GFP as reporters. Selected pools were cloned directly upstream of the reporter. Termination of transcription or translation would result in less expression, and therefore less emission from the reporter. Plasmid pBbB5a-GFP was ordered from Addgene for use as a GFP reporter. This plasmid has a *lac* promoter and is inducible by IPTG[42]. Shine-Dalgarno and start codon were removed from the plasmids of interest in order to ensure translation would be dependent on the availability of the Shine-Dalgarno and start codon of the pool alone.

Various levels of GFP expression can be observed for different clones of the pool, verifying the potential for clones to perturb expression of downstream reporters (Fig 3.11).

Pools that were selected from the strand displacement displayed the most variation in expression when cloned with a GFP reporter, in comparison to those selected for affinity. Individual clones from round 6 and round 8 of the strand displacement selection were most studied. These clones were selected to have the terminator unavailable for binding to the toehold, corresponding to a potential terminated state, in the absence of ligand. In the presence of ligand, the terminator is released and available to bind to and displace the toehold. Clones with low GFP expression were collected by plating, under the assumption that without the ligand clones selected from this pool should have decreased expression. Colonies displaying a decreased expression compared to the "ON" control (plasmid without an insert) were selected and tested for function. Selected colonies were grown overnight, sequenced, and regrown the next day for 6 hours in a 100-fold dilution of LB broth with 1 mM IPTG and ampicillin in the presence and absence of ligand 20 μM ligand. After six hours, cells were pelleted and resuspended in PBS buffer. Fluorescence was measured using a plate reader.

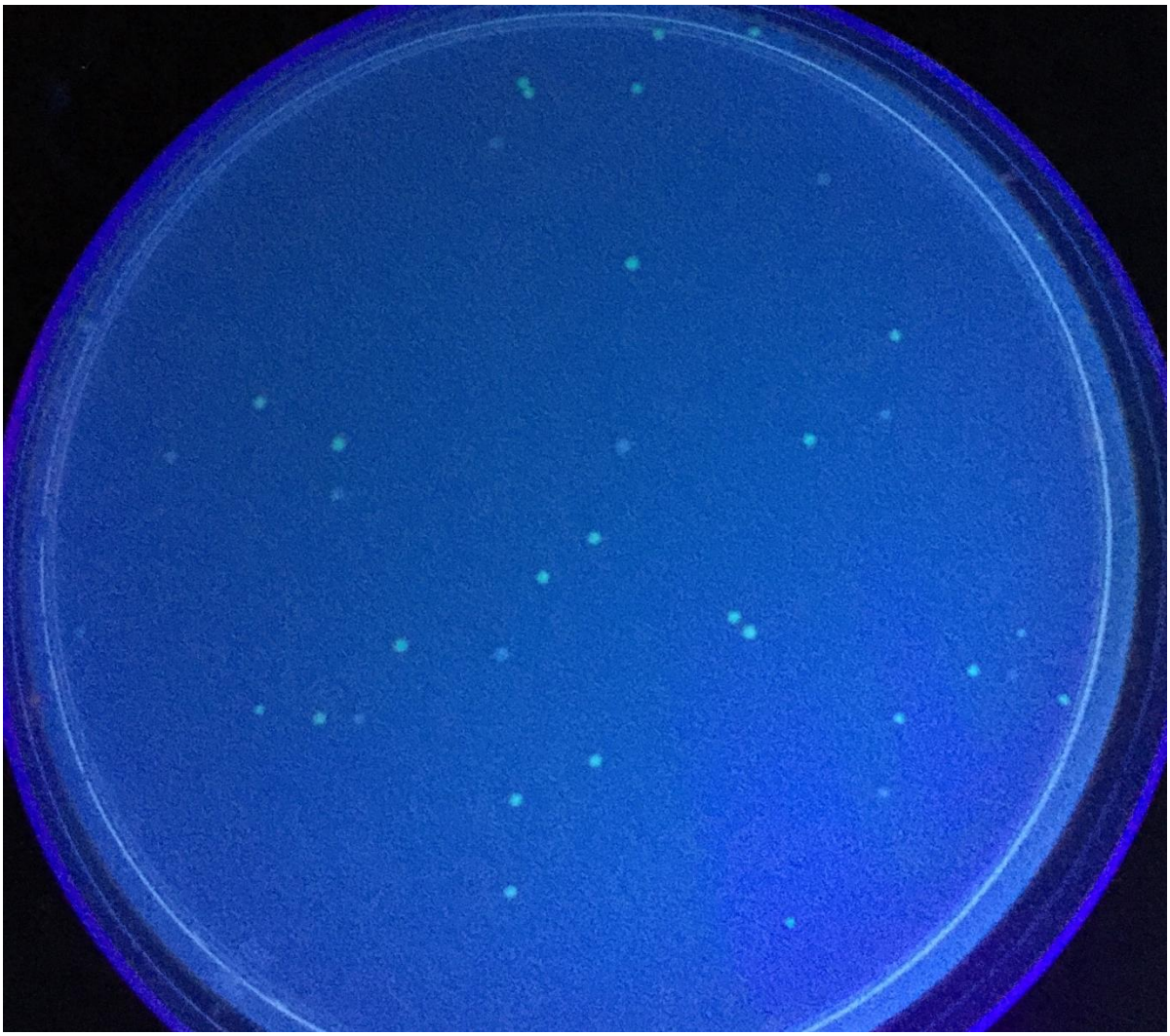


Figure 3.11 Selected pool transformed into *E. coli* for expression studies (methods 6.14)**.** Pools were cloned into *E. coli* using the pBbB5a expression plasmid. Colonies display various levels of GFP expression likely corresponding to the strength of termination of individual sequences.

Cell sorting

As described above, GFP reporter plasmids containing a selected pool from round 6 and round 8 of strand displacement selection were sorted according to fluorescence intensity. Cells were grown overnight in the absence of ligand. On the day of sorting, cells were diluted 1000-fold and grown for six hours in the absence of ligand, diluted 1000 fold in PBS, and sorted according to an "ON" control (plasmid without insert) with high GFP expression (+ GFP) and an "OFF" control with little GFP expression (- GFP). Two separate aliquots were collected from each pool after sorting cells, corresponding to the two gates set for the "ON" and "OFF" controls. Aliquots were diluted into 1 mL of LB broth with 1 mM IPTG and ampicillin then grown overnight for the next round of sorting. Overnight incubations were diluted 100 fold and grown for six hours in the presence of the ligand. The + GFP aliquots from the previous round were sorted and now only the – GFP cells were collected in order to select for the function of a ligand-dependent termination switch. The – GFP aliquots from the previous round were sorted and only the + GFP cells were collected in order to select for the function of a ligand-dependent anti-termination switch. Individual clones were tested for expression after four rounds of sorting. While expression of clones varied widely due to termination strength, there was no difference in expression due to the ligand.

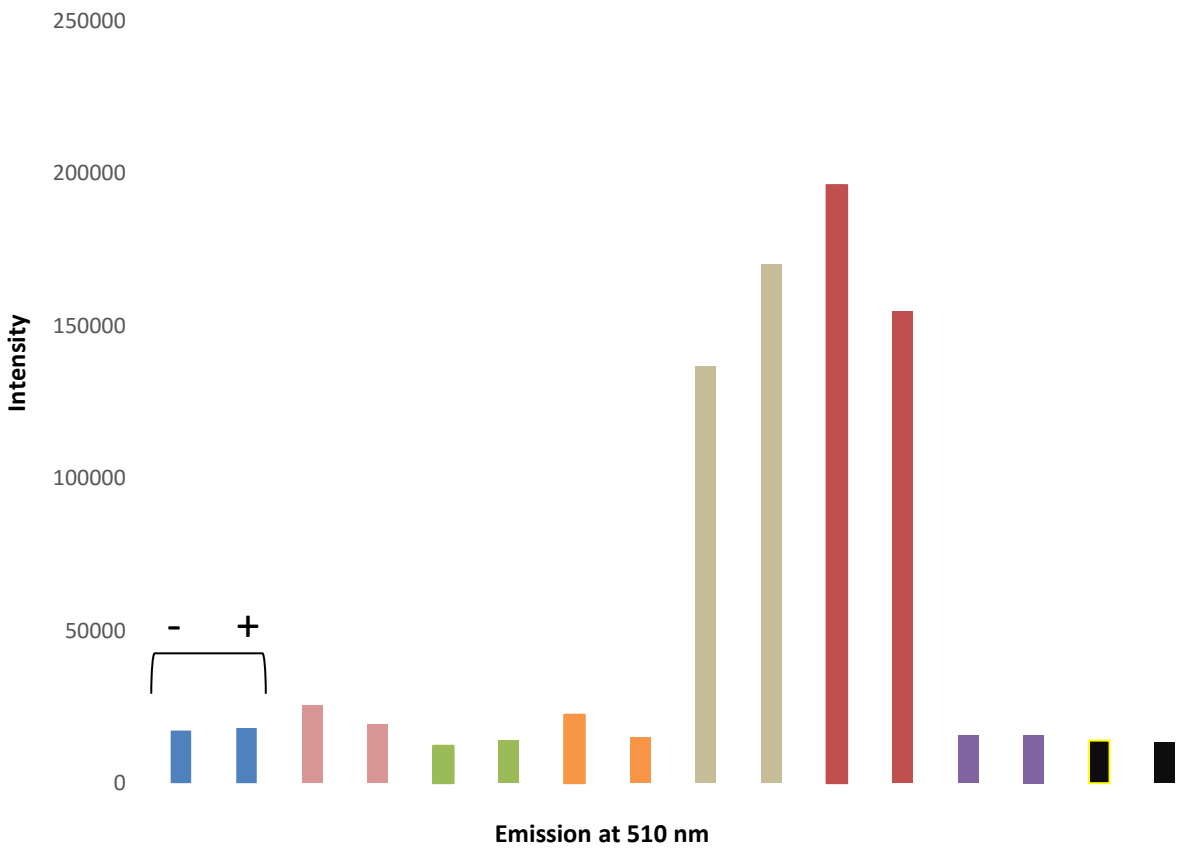


Figure 3.12 Emission of clones after cell sorting. Clones were tested for GFP expression in the absence (left) and presence (right) of ligand, individual clones are indicated by color. Excitation at 395 nm.

Fluorescent toehold reporter (methods 6.17)

A toehold reporter assay was developed by the Mansy lab analogous to the previously described selection by strand displacement (section 3.7.). As previously described, the toehold contains a duplex complimentary to the terminator of the pool. The reporter strand is the longer strand containing fluorescein for fluorescent feedback, and the shorter strand, which can be displaced by the pool, contains a quencher[26]. Upon displacement by RNA the quencher is released and fluorescence can be measured according to the affinity of the clone or pool for the reporter strand. Several pools and clones were tested in the presence of 10 μM ligand. Clone 36 demonstrated a strong ligand dependence in the μM range, as observed by a decrease in fluorescence corresponding to a less displacement compared to no ligand. SHAPE analysis of clone 36 previously indicated structural rearrangement at and near the terminator, formation of a terminator is expected to be less available to bind to the reporter due to the hairpin fold. Although binding of *trans*-**1** results in less strand displacement signifying structural rearrangement of the expression domain, it does not guarantee formation or function of the terminator. Clone 36 was tested with a titration series of the ligand ranging from 100 nM to 30 μM to determine the activity of the sequence. Strand displacement decreases with increasing concentration of ligand, the data was fit to measure the dissociation constant (K_D) of clone 36. Two separate experiments using fresh stocks of RNA and reporter revealed a K_D of 21 and 25 μM (Fig 3.13 a grey and black). We tested a c*is* stilbene, (*Z*)-1,2-Diphenylethene, as a control for specificity (Fig 3.13 a red), the *cis* stilbene had no consequence on clone 36 ability to bind to the toehold.

The fluorescence reporter was used to determine if the photo-isomerization of *trans*-**1** resulted in a difference of affinity of clone 36 to the toehold. *Trans*-**1** inhibits strand displacement of clone 36, and isomerization of *trans*-**1** may release the bound aptamer,

resulting in a recovery of strand displacement and enhanced fluorescence from the reporter. Three samples of clone 36 from the same stock were incubated with 20 μM of *trans*-**1** and toehold measurements were taken for 5 minutes, resulting in a similar amount of displacement for all three samples. After 5 minutes one sample was excited at 342 nm to isomerize *trans*-**1** while the other two control samples were shielded from light and toehold measurements commenced for an additional hour. A sharp increase in toehold binding was observed for the sample excited at 342 nm for 1 minute, indicating an aptamer capable of switching its conformation upon isomerization (Fig 3.13 b). Results were confirmed by repeating the experiment using a new stock of RNA with 30 μM of *trans*-**1**. In order to observe if longer exposure to light led to a greater affect in toehold switching three samples were taken at 1, 2, and 3 minutes exposure to 342 nm, respectively. These samples exhibited negligible difference in switching with prolonged exposure (Fig 3.13 c). To our knowledge, this is the first aptamer switch (apta-switch) to bind directly to a stilbene analog.

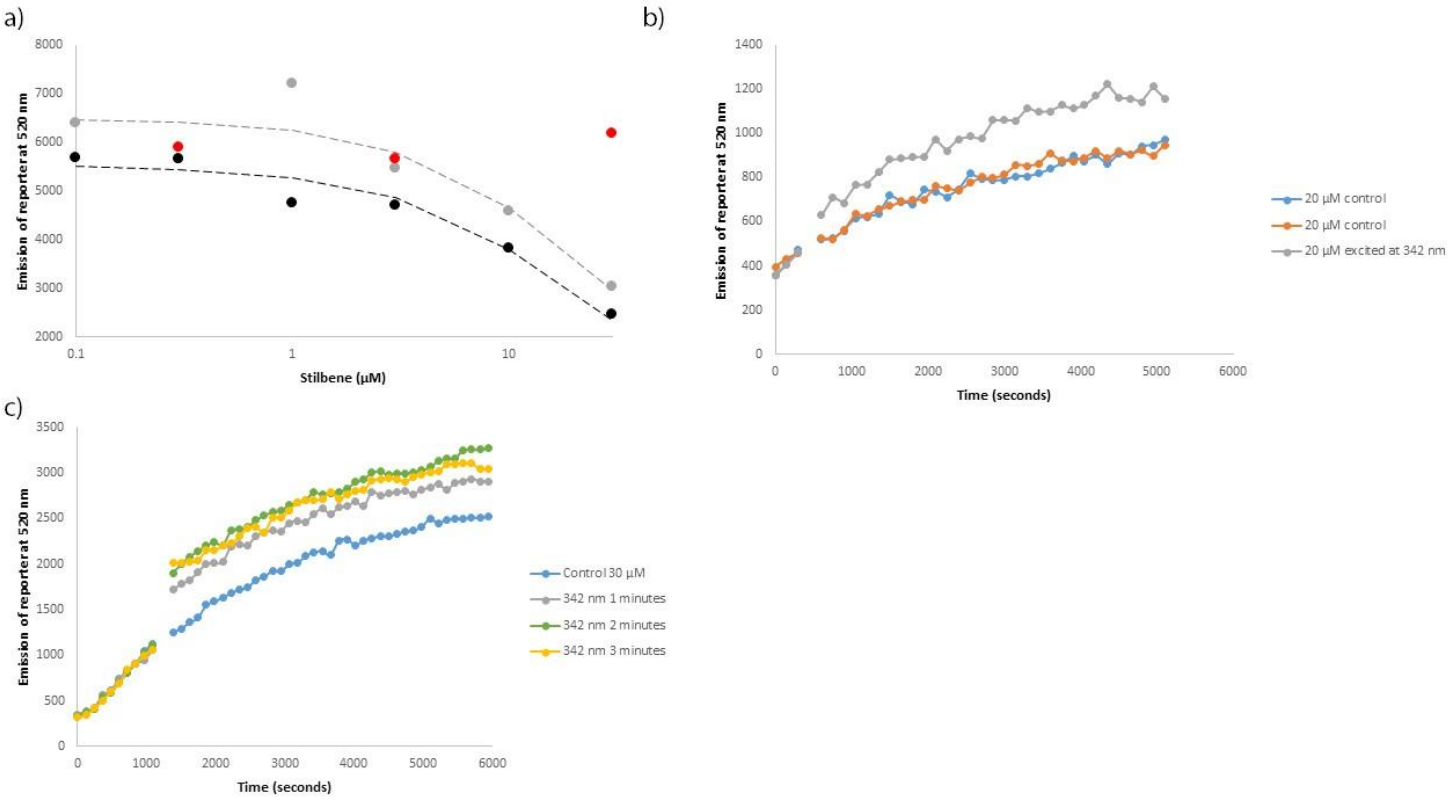


Figure 3.13. Fluorescence toehold assays of clone 36. (a) Binding curve based on fluorescence intensity of strand displacement with increasing amount of *trans*-**1**. Displacement is inhibited with increasing concentrations of *trans*-**1**. Two sets of data (black and grey) set up on two different days result in apparent K_Ds of 21 μM and 25 μM respectively. Red is a control using *cis* stilbene. (b) Clone 36 was incubated with *trans*-**1** and irradiated at 342 nm to isomerize *trans*-**1** and release bound RNA. Control samples (blue and orange) contain clone 36 with 20 μM of *trans*-**1** that were not irradiated. Photo-isomerization of *trans*-**1** was achieved by irradiation at 342 nm for 1 minute (gray). A break in toehold measurements occurs (~4 minutes) to pause, excite the sample, and continue the strand displacement measurements. Upon resuming measurements the excited sample has an increased affinity for the toehold compared to two controls. (c) Clone 36 was incubated with 30 μM of *trans*-**1**

and irradiated for different periods of time. Control sample contains RNA with 20 μM of *trans*-**1** that were not irradiated. Photo-isomerization of *trans*-**1** was achieved by irradiation 342 nm for 1, 2, or 3 minutes (grey, green, and yellow, respectively). Prolonged exposure to light did not result in higher recovery of toehold binding.

Conclusion

An essential component of a riboswitch is a strong aptamer with a high binding affinity domain. Several selections have been performed containing a built-in aptamer domain and selecting for an expression platform, resulting in synthetic riboswitches for theophylline[18] (see section 3.3). However, to the best of our knowledge, selection protocols for completely novel aptamers with a built in expression platforms have not been established yet. Additionally, aptamers have been utilized upstream or downstream of fixed expression platforms to regulate expression, illustrating the importance of a defined aptamer domain to mediate regulation. The Suess lab engineered the tetracycline aptamer in the 5′ UTR of yeast, proximal to the cap structure and the start codon, resulting in a tetracycline-dependent decrease in translation expression[43]. The pool design was dependent on the discovery of an aptamer that facilitates a switch with a fixed expression domain. The pool design may be too stringent to select for both aptamer and expression functions by the methods described. A ligand-dependent expression component has not been discovered; however, we have successfully achieved enrichment of an aptamer domain.

Clone 36 exhibits affinity for*trans*-**1** by both SHAPE and strand displacement analysis. Results of strand displacement indicate a K_D of ~ 25 μM, similar to the binding capacity of the beads used for affinity selection. This aptamer was also demonstrated to switch upon isomerizing the *trans*-**1** stilbene, resulting in a stilbene-dependent aptaswitch. The aptaswitch can be used in applications by engineering an RNA of interest upstream or downstream of the switch, reversing availability of the RNA by isomerization of *trans*-**1**.

Clone 36 can be used as a template for selections by making variants of the sequence to reselect for a riboswitch or a stronger aptamer switch effect. Reselections of riboswitches have resulted in variants that have a much larger ligand-dependent difference in expression. The aptamer can be incorporated up or downstream of a new random region to select for an expression domain that is complementary or dependent on the aptamer region. Half of the requirements of a riboswitch have been achieved by an aptamer domain with the additional function being released upon isomerization of the *trans*-**1** stilbene. A scheme to accomplish both selection of binding and function of a novel RNA has yet to be described; however, the efforts and methods described herein provide on the benefits and pitfalls of the methods employed.

1. Breaker, R.R. Riboswitches and the RNA World. *Cold Spring Harbor Perspectives in Biology* **4** (2012).
2. Birks, J.B. The photo-isomerization of stilbene. *Chemical Physics Letters* **38**, 437-440 (1976).
3. Tuerk, C. & Gold, L. Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science* **249**, 505-510 (1990).
4. Ellington, A.D. & Szostak, J.W. In vitro selection of RNA molecules that bind specific ligands. *Nature* **346**, 818-822 (1990).
5. Wu, Y.X. & Kwon, Y.J. Aptamers: The "evolution" of SELEX. *Methods*.
6. Jijakli, K. et al. The in vitro selection world. *Methods*.
7. Mironov, A.S. et al. Sensing Small Molecules by Nascent RNA: A Mechanism to Control Transcription in Bacteria. *Cell* **111**, 747-756 (2002).
8. Rodionov, D.A., Vitreschak, A.G., Mironov, A.A. & Gelfand, M.S. Comparative genomics of the methionine metabolism in Gram-positive bacteria: a variety of regulatory systems. *Nucleic Acids Research* **32**, 3340-3353 (2004).
9. Nahvi, A. et al. Genetic Control by a Metabolite Binding mRNA. *Chemistry & Biology* **9**, 1043-1049 (2002).
10. SUDARSAN, N., BARRICK, J.E. & BREAKER, R.R. Metabolite-binding RNA domains are present in the genes of eukaryotes. *RNA* **9**, 644-647 (2003).
11. Vu, Michael M.K. et al. Convergent Evolution of Adenosine Aptamers Spanning Bacterial, Human, and Random Sequences Revealed by Structure-Based Bioinformatics and Genomic SELEX. *Chemistry & Biology* **19**, 1247-1254 (2012).
12. Salehi-Ashtiani, K., Lupták, A., Litovchick, A. & Szostak, J.W. A Genomewide Search for Ribozymes Reveals an HDV-Like Sequence in the Human CPEB3 Gene. *Science* **313**, 1788-1792 (2006).
13. Li, S. & Breaker, R.R. Eukaryotic TPP riboswitch regulation of alternative splicing involving long-distance base pairing. *Nucleic Acids Research* **41**, 3022-3031 (2013).
14. Mandal, M. & Breaker, R.R. Gene regulation by riboswitches. *Nat Rev Mol Cell Biol* **5**, 451-463 (2004).
15. Lussier, A., Bastet, L., Chauvier, A. & Lafontaine, D.A. A Kissing Loop Is Important for btuB Riboswitch Ligand Sensing and Regulatory Control. *Journal of Biological Chemistry* **290**, 26739-26751 (2015).
16. Jenison, R., Gill, S., Pardi, A. & Polisky, B. High-resolution molecular discrimination by RNA. *Science* **263**, 1425-1429 (1994).

17. Desai, S.K. & Gallivan, J.P. Genetic Screens and Selections for Small Molecules Based on a Synthetic Riboswitch That Activates Protein Translation. *Journal of the American Chemical Society* **126**, 13247-13254 (2004).
18. Wachsmuth, M., Findeiß, S., Weissheimer, N., Stadler, P.F. & Mörl, M. De novo design of a synthetic riboswitch that regulates transcription termination. *Nucleic Acids Research* **41**, 2541-2551 (2013).
19. Mandal, M., Boese, B., Barrick, J.E., Winkler, W.C. & Breaker, R.R. Riboswitches Control Fundamental Biochemical Pathways in Bacillus subtilis and Other Bacteria. *Cell* **113**, 577-586 (2003).
20. Ames, T.D. & Breaker, R.R. Bacterial aptamers that selectively bind glutamine. *RNA Biology* **8**, 82-89 (2011).
21. Mandal, M. et al. A Glycine-Dependent Riboswitch That Uses Cooperative Binding to Control Gene Expression. *Science* **306**, 275-279 (2004).
22. Winkler, W.C., Nahvi, A., Sudarsan, N., Barrick, J.E. & Breaker, R.R. An mRNA structure that controls gene expression by binding S-adenosylmethionine. *Nat Struct Mol Biol* **10**, 701-707 (2003).
23. Grundy, F.J., Lehman, S.C. & Henkin, T.M. The L box regulon: Lysine sensing by leader RNAs of bacterial lysine biosynthesis genes. *Proceedings of the National Academy of Sciences of the United States of America* **100**, 12057-12062 (2003).
24. Lau, M.W.L. & Ferré-D'Amaré, A.R. In vitro evolution of coenzyme-independent variants from the glmS ribozyme structural scaffold. *Methods*.
25. Muranaka, N., Sharma, V., Nomura, Y. & Yokobayashi, Y. An efficient platform for genetic selection and screening of gene switches in Escherichia coli. *Nucleic Acids Research* **37**, e39-e39 (2009).
26. Martini, L., Ellington, A.D. & Mansy, S.S. An in vitro selection for small molecule induced switching RNA molecules. *Methods*.
27. Waldeck, D.H. Photoisomerization dynamics of stilbenes. *Chemical Reviews* **91**, 415-436 (1991).
28. Syage, J.A., Felker, P.M. & Zewail, A.H. Picosecond dynamics and photoisomerization of stilbene in supersonic beams. II. Reaction rates and potential energy surface. *The Journal of Chemical Physics* **81**, 4706-4723 (1984).
29. Erdélyi, M., Karlén, A. & Gogoll, A. A New Tool in Peptide Engineering: A Photoswitchable Stilbene-type β-Hairpin Mimetic. *Chemistry – A European Journal* **12**, 403-412 (2006).
30. Papper, V. et al. Novel Photochrome Aptamer Switch Assay (PHASA) for Adaptive Binding to Aptamers. *Journal of Fluorescence* **24**, 1581-1591 (2014).
31. Debler, E.W. et al. Deeply Inverted Electron-Hole Recombination in a Luminescent Antibody-Stilbene Complex. *Science* **319**, 1232 (2008).
32. Jacob, F. & Monod, J. Genetic regulatory mechanisms in the synthesis of proteins. *Journal of Molecular Biology* **3**, 318-356 (1961).
33. Lutz, R. & Bujard, H. Independent and Tight Regulation of Transcriptional Units in Escherichia Coli Via the LacR/O, the TetR/O and AraC/I1-I2 Regulatory Elements. *Nucleic Acids Research* **25**, 1203-1210 (1997).
34. Siuti, P., Yazbek, J. & Lu, T.K. Synthetic circuits integrating logic and memory in living cells. *Nat Biotech* **31**, 448-452 (2013).
35. Kao, J.P.Y. in Current Protocols in Neuroscience (John Wiley & Sons, Inc., 2001).
36. Deiters, A. Light Activation as a Method of Regulating and Studying Gene Expression. *Current opinion in chemical biology* **13**, 678-686 (2009).
37. Paige, J.S., Wu, K.Y. & Jaffrey, S.R. RNA Mimics of Green Fluorescent Protein. *Science* **333**, 642-646 (2011).
38. Dolgosheina, E.V. et al. RNA Mango Aptamer-Fluorophore: A Bright, High-Affinity Complex for RNA Labeling and Tracking. *ACS Chemical Biology* **9**, 2412-2420 (2014).
39. (!!! INVALID CITATION !!! {}).
40. Merino, E.J., Wilkinson, K.A., Coughlan, J.L. & Weeks, K.M. RNA Structure Analysis at Single Nucleotide Resolution by Selective 2'-Hydroxyl Acylation and Primer Extension (SHAPE). *Journal of the American Chemical Society* **127**, 4223-4231 (2005).
41. McGinnis, J.L., Dunkle, J.A., Cate, J.H.D. & Weeks, K.M. The Mechanisms of RNA SHAPE Chemistry. *Journal of the American Chemical Society* **134**, 6617-6624 (2012).
42. Lee, T.S. et al. BglBrick vectors and datasheets: A synthetic biology platform for gene expression. *Journal of Biological Engineering* **5**, 1-14 (2011).
43. Hanson, S., Berthelot, K., Fink, B., McCarthy, J.E.G. & Suess, B. Tetracycline-aptamer-mediated translational regulation in yeast. *Molecular Microbiology* **49**, 1627-1637 (2003).

Chapter 4

High-throughput aptamer discovery through Apta-Seq and its application to genomic SELEX of human ATP aptamers. (Submitted).

Michael M. Abdelsayed[1], Bao Ho[2], Michael M. K. Vu[3], Robert C. Spitale[2,3], and Andrej Lupták[1,2,3*]

[1]Department of Molecular Biology and Biochemistry, University of California–Irvine, Irvine, California 92697, USA

[2]Department of Pharmaceutical Sciences, University of California–Irvine, Irvine, California 92697, USA

[3]Department of Chemistry, University of California–Irvine, Irvine, California 92697, USA

Abstract

Laboratory-evolved RNAs serve highly diverse functions, including as diagnostic and therapeutic aptamers, binding a wide variety of targets. The majority of aptamers have been identified using *in vitro* selection (SELEX), a molecular evolution technique based on selecting target-binding RNAs from highly diverse pools through serial rounds of enrichment and amplification. *In vitro* selection often yields multiple distinct motifs of highly variable abundance and target-binding affinities, and the discovery of new aptamers is often limited by the difficulty characterizing all selected motifs, because testing of individual sequences tends to be a tedious process. We developed Apta-Seq for a simple, high-throughput analysis of in *vitro* selected pools, revealing the identity, structural features, and target dissociation constants for aptamers present even at low abundance. Apta-Seq of a human genomic pool enriched for ATP-binding RNAs yielded three new aptamers, suggesting that ligand-binding RNAs may be common in mammals.

Functional RNAs play central roles in regulating gene expression and catalyzing essential cellular reactions.[1] RNAs evolved in the laboratory serve equally diverse functions, including as diagnostic and therapeutic aptamers.[2] Aptamers bind a wide variety of targets and are either *in vitro* selected or occur naturally as the ligand-binding domains of cellular riboswitches, which regulate gene expression in bacteria and some eukaryotes by modulating transcription, translation, splicing, or RNA stability.[3] Whereas riboswitches have been discovered through genetic and bioinformatics approaches, the vast majority of aptamers have been identified using *in vitro* selection (or SELEX), a molecular evolution technique based on selecting target-binding RNAs from highly diverse pools through serial rounds of enrichment and amplification.[4, 5] The RNA pools are transcribed from either synthetic (typically random) or genomic DNAs and selections often yield multiple distinct motifs of highly variable abundance, and target-binding affinities and specificities.

The discovery of new aptamers is often hampered by the difficulty of identifying and characterizing all aptamer motifs that result from the selection process, because testing of individual sequences identified in selected pools tends to be a tedious process. High-throughput sequencing can be applied to measure sequence diversity of selected pools and identify potential ligand-binding RNAs, but their structural and binding characteristics have to be established individually for each sequence, often making this the limiting step in the discovery of novel aptamers. Indeed, many functional aptamers may go uncharacterized due to the challenge associated with fully characterizing every sequence within a selected pool. This handicaps the full description of selected RNAs and is a key hurdle to overcome for the efficient isolation and characterization of functional aptamers that may have diverse and important functions.

To overcome these challenging limitations, we herein present a high-throughput approach to couple RNA selection with structural and binding characterization of individual sequences within the selected pools. We marry selection with chemical probing of RNA structure to reveal the sequence, structural features, and ligand affinities of both dominant and minor species from the same pool. We use this technique, Apta-Seq, to discover and characterize both known and novel adenosine aptamers in the human genome. Our novel methodology not only increases the rate of novel aptamers for our studied ligand (ATP), but also has the potential to be applied to any ligand-pool pair, therefore greatly enhancing the speed of aptamer discovery and structural characterization.

Information about structure and ligand-binding sites in RNAs can be extracted from experiments based on partial hydrolysis (in-line probing) and chemical modification of RNAs, using selective 2′-hydroxyl acylation (SHAPE) or base modification (e.g. by dimethyl sulfate.[6, 7] The SHAPE method detects 2′-OH accessibility and reactivity to acylation, thereby informing which aptamer segments are more flexible and reactive in response to a ligand.[8] Changes in structure can be probed under different environmental conditions or in the presence of a ligand to determine an apparent K_D.[9] Several methods combine SHAPE with high-throughput sequencing (SHAPE-seq) to achieve single-nucleotide resolution of acylation reactivity on a diverse set of sequences simultaneously and couple the output to computational modules developed to yield genomic locations, intrinsic reverse transcriptase (RT) stops, SHAPE reactivates, and secondary structure models for each transcript.[10-12] Although useful, most of these efforts have been largely descriptive with few examples of their use for novel biological discovery. Our results herein provide the first example, to our knowledge, of demonstrating the power of high-throughput chemical probing of RNA structure toward the discovery of novel potential biological functions associated with RNA-ligand interactions.

In order to establish the identity, secondary structure, as well as binding properties of aptamers, we combined *in vitro* selection, SHAPE-seq, and StructureFold into a pipeline (Fig. 4.1), providing all the information required for aptamer discovery in a single experiment. We applied this method to an *in vitro* selected pool derived from the human genome[13] and enriched for ATP-binding aptamers, as described previously.[14] Adenosine aptamers with a conserved binding motif consisting of an 11 nucleotide binding loop and an opposing bulging guanosine, flanked by two helices (the Sassanfar-Szostak motif), were initially discovered by *in vitro* selections from random pools[15] and a genomic SELEX experiment revealed the same motif in two distinct loci in the human genome:[14] the FGD3 aptamer resides in an intron of the *FGD3* gene, which codes for a guanine nucleotide exchange factor for a cell division control protein (Cdc42), and the ERV1 aptamer maps antisense to a junction between an ERV1 LTR repeat and its 3' insertion site. These two clones were the only ones revealed using traditional cloning; however, this approach tests only a small number of sequences, we therefore re-analyzed the pool using high-throughput sequencing and uncovered other genomic sequences. To test whether these are bona fide aptamers, we adapted the SHAPE-Seq analysis coupled to StructureFold to the selected RNA pool at varying ATP concentrations (Figs. 4.1 and 4.2).

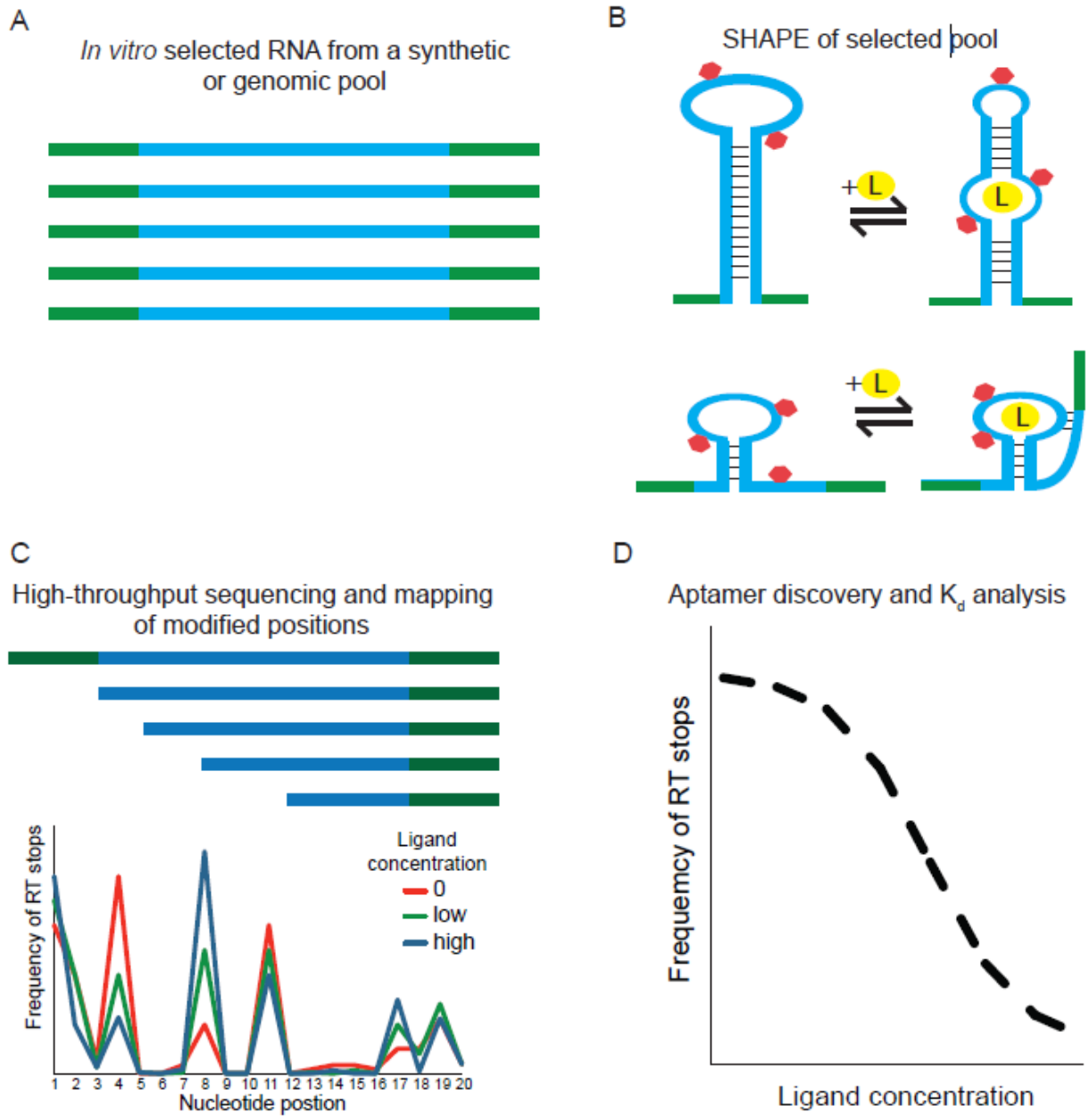


Figure 4.1 Apta-Seq scheme. The pipeline consists of (A) an *in vitro* selection, (B) SHAPE analysis of selected pool, and (C) high-throughput sequencing for determination of sequence identity, secondary structure, and (D) binding isotherms of individual aptamers.

Frequency of RT stops, which mark the nucleotide downstream of positions modified by 2-(azidomethyl)nicotinic acid acyl imidazole (NAI-N_3),[16] was used to obtain SHAPE profiles in the presence of increasing concentration of ATP-Mg (Fig. 4.2A, methods 6.26). The RT stops were resolved with a single-nucleotide resolution and the ATP dependence of SHAPE profiles allowed us to determine the apparent K_D for each aptamer at several positions. For the FGD3 aptamer, two positions (G67 and A16) yielded a K_D ~ 700 μM (Fig. 4.2B). The Apta-Seq data show that the adenosine-binding loop of the Sassanfar-Szostak motif becomes more accessible at high concentrations of ATP at the third guanosine (G 60) of the binding loop (Figs. 4.2A and 4.2E). Previous in-line probing data for the FGD3 and other ATP-binding aptamers have demonstrated that the adenosine-binding loop becomes more susceptible to in-line attack at the third adenosine (equivalent to A61), which is directly downstream of G60.[6, 14] In the solution structures of the *in vitro* selected aptamers bound to AMP, the nucleotides equivalent to G60 and A61 make direct contacts with the ligand through stacking and hydrogen bonding, respectively[17, 18] and the 2′ OH of G60 equivalent appears partially solvent-exposed and hydrogen-bonded to the adjacent phosphate, which likely activates it for acylation.[19] The sugar-phosphate backbone of this nucleotide in the adenosine-bound conformation is thus highly sensitive to modification, and in the case of the FDG3 aptamer, the binding loop must undergo a significant conformational change upon ligand binding because the same position is weakly acylated in the absence of the ligand.

To validate the results obtained from Apta-Seq of the pool with the reactivity of the purified FGD3 aptamer, we performed a SHAPE analysis on an isolated clone of the aptamer. One of the positions with the most prominent ATP-dependent change in SHAPE reactivity is C73, which maps to a domain adjacent to the adenosine binding loop (Fig. 4.2C). PAGE-based analysis of the aptamer clone also revealed strong, ATP-dependent increase in reverse

transcription termination at position C73 (Fig. 4.2D), confirming that the high-throughput method reveals comparable data for the same aptamer within a highly heterogeneous pool of sequences. These results importantly demonstrate that our high-throughput sequencing approach parallels more traditional structural analysis normally reserved to single clone analyses.

For the ERV1 aptamer, we derived the SHAPE reactivity using the Reactivity Calculation module of StructureFold, with RT stops from the no-ligand SHAPE dataset as the negative control for the reactivity profile of 10 mM ATP experiment and carried out structure predictions using the RNAprobing server of the Vienna RNA Package (methods 6.27).[20] In the absence of ATP, the ERV1 structure is not predicted to form the ATP binding motif, because the binding loop is constrained within a stem flanked by two bulges; however, in the presence of ATP, the structure rearranges to form the classical ATP binding loop (Fig. 4.3A). The binding loop showed similar acylation trends to the FGD3 aptamer: the third guanosine (G16) of the ATP binding loop becomes more accessible with increasing concentration of ATP, analogous to the change of G60 in the FGD3 aptamer. Altogether, the data are consistent with previous findings and structure predictions for the FGD3 and ERV1 aptamers and the K_Ds we extracted from the Apta-Seq data are within two-fold of the K_Ds determined previously by in-line probing of individual aptamers. The FGD3 and ERV1 aptamers make up a majority of the analyzed pool, with a relative abundance of 30% and 39%, respectively, to all the aptamers presented here, but only a small fraction of all sequences obtained from the Apta-Seq experiment.

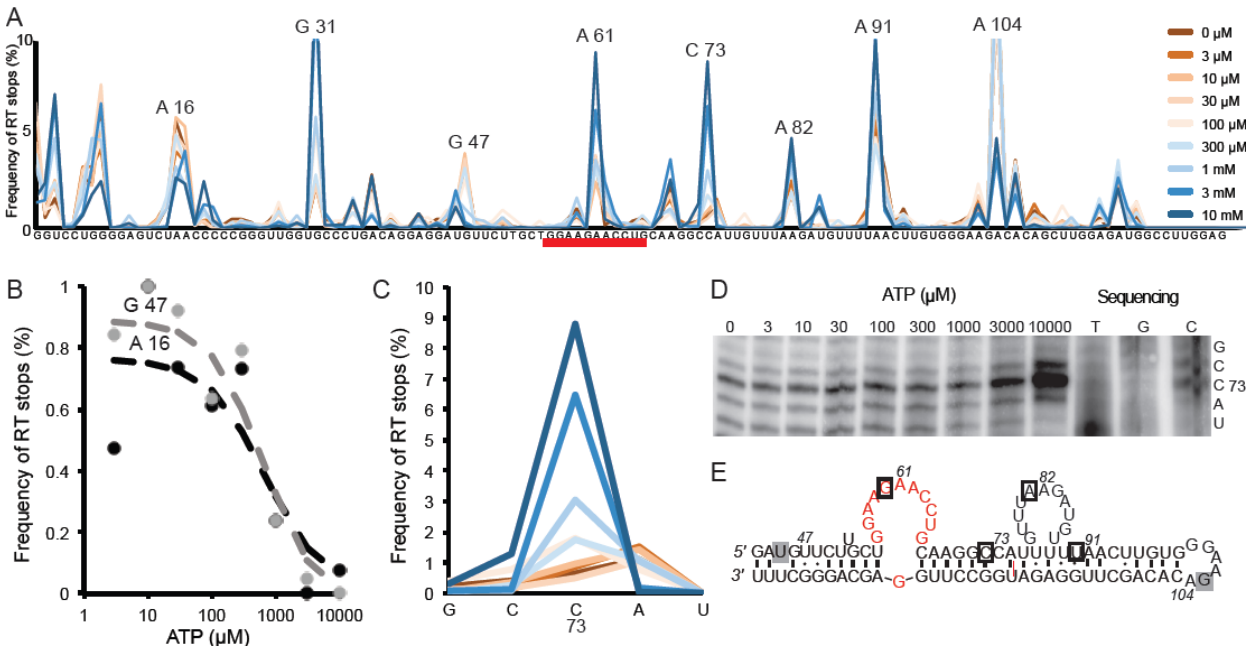


Figure 4.2 Apta-Seq profile of the human FGD3 adenosine aptamer (methods 6.24-6.27) (A) Graph of frequency of reverse transcriptase (RT) stops at every position of the FDG3 aptamer and varying concentrations of ATP. The acylated positions in the aptamer are one nucleotide upstream of the positions indicated on the horizontal axis. The adenosine-binding motif is underlined in red. (B) Binding isotherms of the aptamer extracted from positions A16 (black circles) and G47 (grey circles). Both positions reveal a K_D ~ 700 µM. (C) SHAPE-Seq data for position C73 showing RT stops increasing with ATP concentration. (D) PAGE analysis of a purified FGD3 aptamer clone showing that RT stops for position C73 strongly increase at high concentrations of ATP and validating the SHAPE-Seq data. (E) Predicted secondary structure of the FDG3 aptamer with ATP binding loop in red. Boxed nucleotides indicate aptamer positions that show an increase (black outlines) or decrease (grey) in acylation with increasing ATP and resulting in RT stops in the SHAPE-Seq data shown in (A) and indicated by position numbers.

The Apta-Seq pipeline also yielded novel, less abundant sequences not found in previous analysis. Our previously reported aptamers were discovered by ATP column blinding analysis of individual clones to find potential candidates to undergo structural analysis; Apta-Seq condenses the process to gain a full analysis of the pool in one experiment in solution. Surprisingly, three new aptamers also contained the Sassanfar-Szostak motif. The PRR5 aptamer (Fig. 4.3B) maps to the second intron of the *PRR5* gene or the first intron of the PRR5-ARHGAP8 fusion protein. *PRR5* codes for a protein that is part of the mTORC2 complex[21] and like the FGD3 plays important roles in pathways that regulate cell growth, but the significance of the ATP/adenosine-binding aptamers in their introns remains unknown. The PRR5 aptamer bound to ATP-agarose beads and eluted in the presence of free ATP, and Apta-Seq data revealed a K_D of ~ 950 at a relative abundance of 10% among the analyzed aptamers. Interestingly, mutations (including a 14-nt insertion) of the aptamer sequence throughout the genomes of primates preserve the aptamer structure, suggesting that it may be a functional RNA in primates.

The second aptamer (Fig. 4.3C) was found in a long interspersed element (LINE), L1PA16, and the third (Fig. 4.3D) maps antisense to a repeat element THE1B, which is derived from a subfamily of ERV Mammalian apparent LTR-retrotransposons (ERV-MaLR), and maybe stabilized by fortuitous base-pairing with the sequence derived form the forward primer of the pool (Fig. 4.3D). L1PA15 and THE1B-ERVL aptamers bound to ATP-agarose beads and eluted in the presence of free ATP (Figs. 4.3C and 4.3D). The K_Ds derived from Atpa-Seq were ~940 μM and 980 μM, with a relative abundance among these five aptamers of 7% and 10%, respectively. These aptamers, together with the previously discovered adenosine[14] and GTP[22] aptamers indicate that ligand-binding RNAs are likely common in higher eukaryotes.

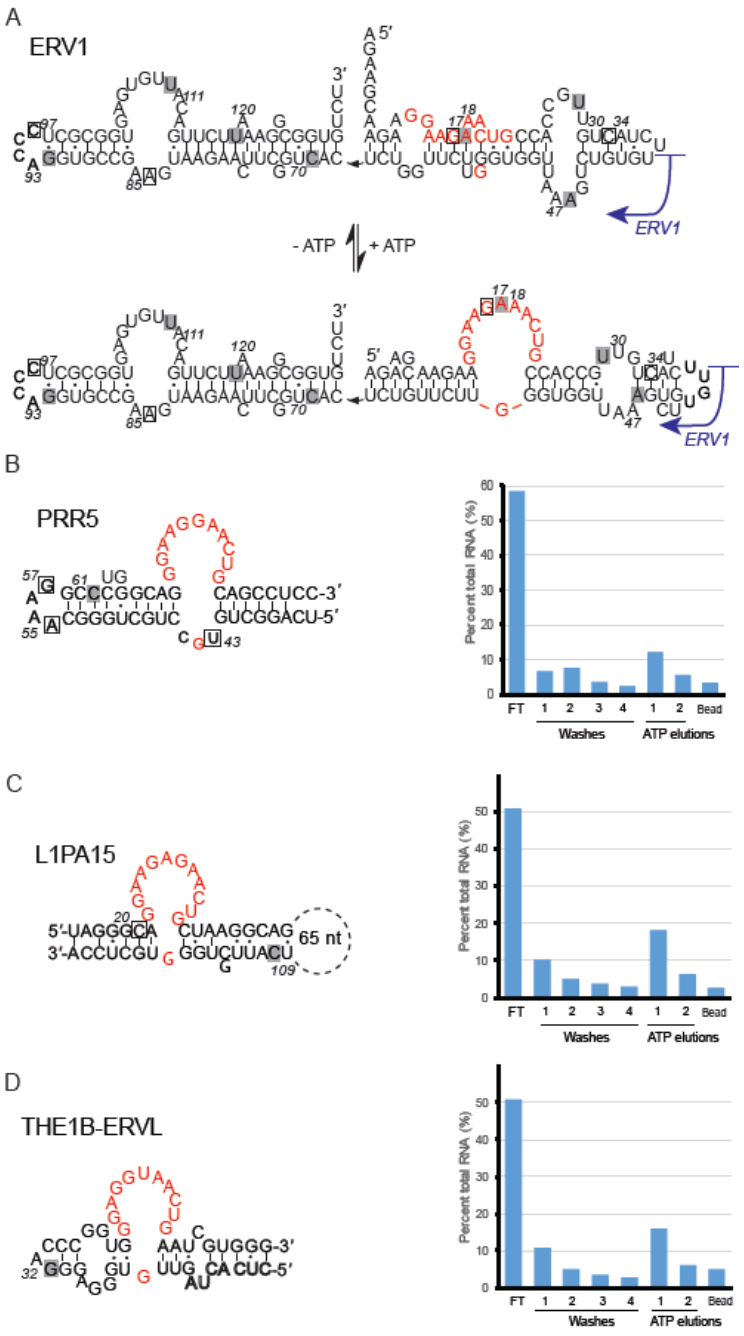


Figure 4.3 Secondary structures and binding profiles of human adenosine aptamers (methods 6.27). (A) Secondary structure models of the ERV1 aptamer in absence (top) and 10 mM ATP (bottom) with` acylation positions sensitive to ATP binding boxed as in Fig. 4.2. Adenosine-binding loop is shown in red. Novel human ATP aptamers revealed by Apta-Seq of

the *in vitro* selected pool and their ATP column binding profiles (FT, flow-through). The aptamers form the Sassanfar-Szostak motif and map to an intron of the PRR5 gene (B), L1PA16 LINE element (C), and antisense to a THE1B retrotransposon (D; part of the 5′ sequence that originated form the forward primer-binding region of the pool is shown in outline).

We describe an efficient process of high-throughput analysis of *in vitro* selected pools. A single *in vitro* selection experiment is combined with SHAPE-Seq and StructureFold analysis to efficiently and quantitatively analyze the selected pools at a single nucleotide resolution. Apta-Seq utilizes these techniques to make it easier to sort and characterize aptamers. Importantly, Apta-Seq is a powerful enabling technological pipeline that is sure to expedite the transition from aptamer selection to the unraveling of their biological or biotechnological relevance.

References

1. Sharp, P.A. The Centrality of RNA. *Cell* **136**, 577-580 (2009).
2. Jijakli, K. et al. The in vitro selection world. *Methods*.
3. Breaker, R.R. Riboswitches and the RNA World. *Cold Spring Harbor Perspectives in Biology* **4** (2012).
4. Ellington, A.D. & Szostak, J.W. In vitro selection of RNA molecules that bind specific ligands. *Nature* **346**, 818-822 (1990).
5. Tuerk, C. & Gold, L. Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science* **249**, 505-510 (1990).
6. Soukup, G.A. & Breaker, R.R. Relationship between internucleotide linkage geometry and the stability of RNA. *RNA* **5**, 1308-1325 (1999).
7. Ding, Y., Kwok, C.K., Tang, Y., Bevilacqua, P.C. & Assmann, S.M. Genome-wide profiling of in vivo RNA structure at single-nucleotide resolution using structure-seq. *Nat. Protocols* **10**, 1050-1066 (2015).
8. Stoddard, C.D., Gilbert, S.D. & Batey, R.T. Ligand-dependent folding of the three-way junction in the purine riboswitch. *RNA* **14**, 675-684 (2008).
9. Steen, K.-A., Siegfried, N.A. & Weeks, K.M. Selective 2[prime]-hydroxyl acylation analyzed by protection from exoribonuclease (RNase-detected SHAPE) for direct analysis of covalent adducts and of nucleotide flexibility in RNA. *Nat. Protocols* **6**, 1683-1694 (2011).
10. Lucks, J.B. et al. Multiplexed RNA structure characterization with selective 2′-hydroxyl acylation analyzed by primer extension sequencing (SHAPE-Seq). *Proceedings of the National Academy of Sciences* **108**, 11063-11068 (2011).
11. Loughrey, D., Watters, K.E., Settle, A.H. & Lucks, J.B. SHAPE-Seq 2.0: systematic optimization and extension of high-throughput chemical probing of RNA secondary structure with next generation sequencing. *Nucleic Acids Research* (2014).
12. Tang, Y. et al. StructureFold: genome-wide RNA secondary structure mapping and reconstruction in vivo. *Bioinformatics* **31**, 2668-2675 (2015).
13. Salehi-Ashtiani, K., Lupták, A., Litovchick, A. & Szostak, J.W. A Genomewide Search for Ribozymes Reveals an HDV-Like Sequence in the Human CPEB3 Gene. *Science* **313**, 1788-1792 (2006).
14. Vu, Michael M.K. et al. Convergent Evolution of Adenosine Aptamers Spanning Bacterial, Human, and Random Sequences Revealed by Structure-Based Bioinformatics and Genomic SELEX. *Chemistry & Biology* **19**, 1247-1254 (2012).
15. Sassanfar, M. & Szostak, J.W. An RNA motif that binds ATP. *Nature* **364**, 550-553 (1993).
16. Spitale, R.C. et al. Structural imprints in vivo decode RNA regulatory mechanisms. *Nature* **519**, 486-490 (2015).
17. Dieckmann, T., Suzuki, E., Nakamura, G.K. & Feigon, J. Solution structure of an ATP-binding RNA aptamer reveals a novel fold. *RNA* **2**, 628-640 (1996).
18. Jiang, F., Kumar, R.A., Jones, R.A. & Patel, D.J. Structural basis of RNA folding and recognition in an AMP-RNA aptamer complex. *Nature* **382**, 183-186 (1996).
19. McGinnis, J.L., Dunkle, J.A., Cate, J.H.D. & Weeks, K.M. The Mechanisms of RNA SHAPE Chemistry. *Journal of the American Chemical Society* **134**, 6617-6624 (2012).
20. Washietl, S., Hofacker, I.L., Stadler, P.F. & Kellis, M. RNA folding with soft constraints: reconciliation of probing data and thermodynamic secondary structure prediction. *Nucleic Acids Research* (2012).
21. Woo, S.-Y. et al. PRR5, a Novel Component of mTOR Complex 2, Regulates Platelet-derived Growth Factor Receptor β Expression and Signaling. *Journal of Biological Chemistry* **282**, 25604-25612 (2007).
22. Curtis, Edward A. & Liu, David R. Discovery of Widespread GTP-Binding Motifs in Genomic DNA and RNA. *Chemistry & Biology* **20**, 521-532 (2013).

Chapter 5

Conclusions and future directions

5.1 Progress towards a luminescent aptamer

A ligand-dependent luminescent aptamer would provide a means to trace and quantify RNA by luminescence. The desirable optical properties of lanthanides (chapter 2) make them an alluring luminescent-capable target. Our strategy to obtain a luminescent aptamer was carried out through *in vitro* selection. Selection resulted in many sequences with affinity for a lanthanide chelate, but a luminescent aptamer was not found. Affinity selections resulted in RNAs capable of binding to the target, but steps to select and screen based on luminescence were never employed. Potential aptamers were selected from a random pool by selection, screened for affinity by column binding, then characterized by in-line probing experiments. Luminescence screening was performed on potential aptamers by testing individual clones that exhibited positive binding profiles.

The potential of an RNA aptamer to act as a sensitizer for lanthanide luminescence remains unknown. The capacity for RNA to act as a lanthanide sensitizer may be limited, requiring very specific environmental and binding conditions. During our studies, it was reported by Pierre et al. that RNA aptamer was developed that can act as a luminescent quencher- the opposite function of our desired aim. It was discovered that purine containing RNA are effective in quenching lanthanide luminescence[1]. Binding of a lanthanide to a purine rich region of an aptamer reduces the probability of an energy transfer. It is expected that biasing a pool with a higher ratio of pyrimidines may increase the potential for positive outcomes. In addition to pool design, methods to select and screen for favorable outcomes are necessary for the evolution of functional RNA. Throughput was restrictive, due to the amount of time and reagents needed to test purified RNAs of interest. High-throughput screening and

analysis (affinity and structural properties) would aid in the discovery of a luminescent lanthanide aptamer by revealing the most appealing sequences.

High-throughput affinity analysis could be accomplished by utilizing Apta-Seq with selected pools. This would identify sequences with the highest affinity to the target. Transfer of energy from RNA to the lanthanide target may be correlated with high affinity for the RNA, as the hammerhead ribozyme was demonstrated to enhance lanthanide emission when chelated with Eu (III) and Tb (III)[2]. Tb (III) was demonstrated to inhibit hammerhead cleavage by competing with magnesium (II), with an apparent K_D of 500 nM for Tb (III)[3]. The hammerhead ribozyme is one of the few examples of a functional RNA to show sensitized luminescence, because the ribozyme already has a metal ion binding site capable of binding a variety of ions[4]. The structured and high affinity metal ion binding site may aid in the ability of the RNA to transfer energy to the lanthanide. It is reasonable to suspect that screening only high affinity binders, based on Apta-Seq profiles, may increase the chance of finding an aptamer capable of enhancing lanthanide luminescence. Apta-Seq would reveal which aptamers have the highest affinity, and understanding the specificity and affinity is valuable in determining which sequences are worth testing.

Other selection schemes may be necessary to enrich for lanthanide-based luminescence enhancement. Screens can be designed to isolate for clones capable of luminescence. In such an experiment, selected pools would be inserted in plasmids and cloned into cells for sorting, based on lanthanide emission. Several EDTA containing lanthanide complexes have been proven to be cell permeable[5, 6]. Selection is achieved by sorting, in the absence of the target lanthanide chelate, and collecting cells that do not show luminescence. These cells can be regrown in the presence of the lanthanide target and collected based on emission intensity. Serial rounds of this procedure can be performed to achieve enrichment.

This method is capable of identifying lanthanide-dependent luminescent aptamers, but does not affect affinity. Cell sorting should be used with affinity-selected pools to ensure both affinity and luminescence function.

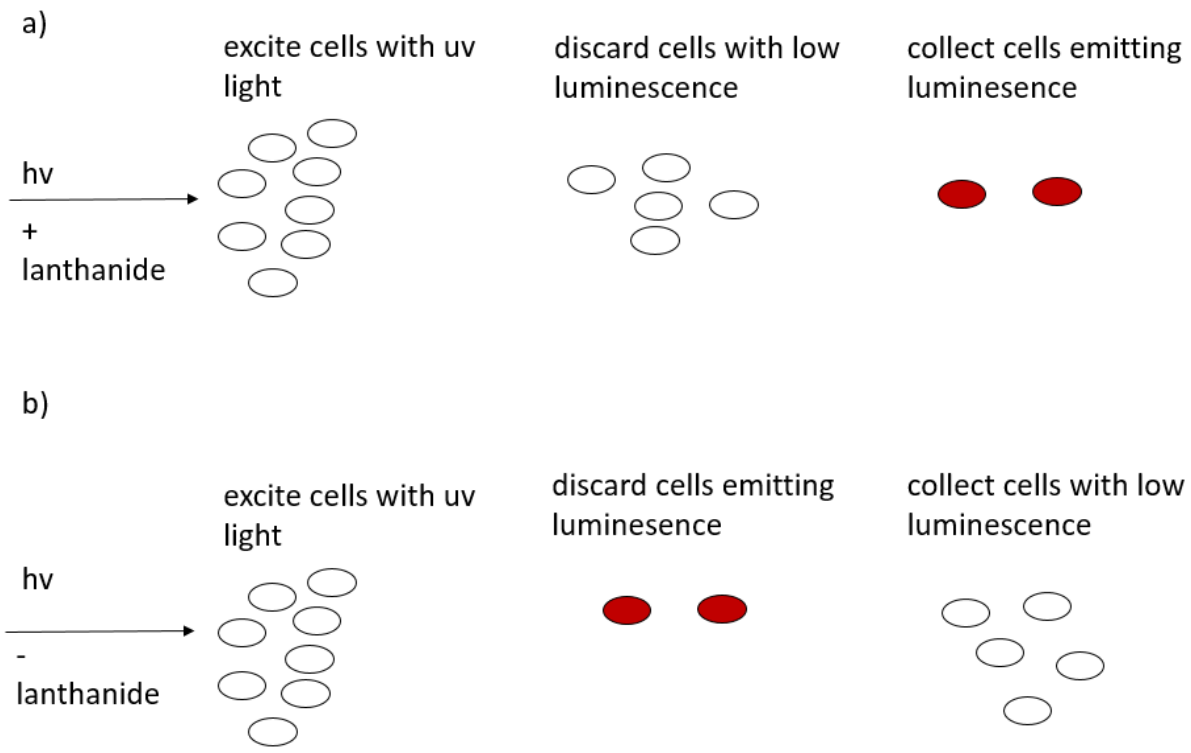


Figure 5.1 **Selection based on cell sorting.** Pools can be cloned and expressed in cells to sort based on luminescence in the presence and absence of a lanthanide target. (a) Cell sorting based on enhanced luminescence in the presence of the lanthanide target. Cells that display strong emission in the presence of the lanthanide are collected and undergo serial rounds of sorting until the pool demonstrates ligand-dependent lanthanide luminescence. (b) Counter selection steps can be taken to assure cells with high background fluorescence are not collected.

A different type of lanthanide chelate may increase the chances of aptamer-induced sensitization by exploiting the optical properties of the chelator bound to the lanthanide. Several classes of molecules have been identified that sensitize lanthanide luminescence. Organic chromophores possess high potential for sensitization, due to their ability to harvest light[7]. An organic chromophore-lanthanide complex can be used as the target of selection to increase the probability of lanthanide sensitization. Chromophores vary in their ability to enhance luminescence enhancement depending on the excitation state and environment of the chelated complex[7, 8].

A new selection can be designed for a chromophore-lanthanide chelate with inherently poor sensitization, in which binding of an aptamer may cause enhancement of sensitization due to changes of the environment or electronic state of the complex when bound by RNA. This is possible, because displacement of water from the coordination sphere of Eu (III) and Tb (III) is directly correlated with increased emission[9]. Binding of an aptamer may reduce the contacts a ligand has with the solvent it is in, so that reducing water contact in an aqueous solution could result in aptamer-dependent lanthanide luminescence.

EDTA and DTPA were used as the chelators for selection because of their strong affinity for multivalent cations. The two chelators contain multivalent coordination properties ideal for binding a lanthanide ion. EDTA and DTPA are often covalently attached to chromophores for the purpose of chelating lanthanides[8, 10]. The amount of different lanthanide chelate targets tested could be considerably increased. In this case, selected pools for EDTA and DTPA lanthanide chelates could be screened for their ability to bind to EDTA and DTPA-containing chromophores and screened for optical activity. Selected pools would undergo more rounds of affinity selection with different EDTA and DTPA chromophore targets to increase affinity for each individual chromophore. The newly selected pools can be screened for luminescence by

methods described in chapter 2. Improvements to our initial selection schemes and targets may thereby produce a lanthanide-dependent luminescent aptamer.

Unexpectedly, selection for a *trans*-**1** stilbene showed that fluorescence of the stilbene is enhanced by RNA. The specificity and characteristics of the fluorescence enhancement was not thoroughly investigated. Stilbenes are capable of radiative and non-radiative release of energy, these processes are influenced by the environment of the stilbene and method of excitation[11, 12]. Affinity-selected pools for *trans*-**1** could undergo screening and reselection for fluorescence enhancement. As described above, affinity-selected pools could be cloned in cells for sorting based on aptamer-dependent fluorescence. Fluorescence screening has already identified many sequences capable of enhancing *trans*-**1** emission. Sequences resulting in the highest level of emission enhancement could be pooled and reselected by error-prone PCR to increase the diversity of potential sequences capable of a large sequence-specific emission enhancement. Our results have established the ability of RNA to induce fluorescence of *trans*-**1** in a concentration-dependent manner. Sequence-specific fluorescence enhancement may be achieved by further screening and selection.

5.2 Progress towards a riboswitch

Several selection methods were developed with the aim of discovering a novel synthetic riboswitch. Two functions are necessary in order to achieve riboswitch function: the ability to sense/bind a ligand, and allosteric control of the expression platform in response to ligand binding[13]. A scheme (3.7) was developed to select for riboswitch function by *in vitro* transcription. The selection pool was based on the SAM-1 riboswitch, Breaker demonstrated termination of the riboswitch could be observed transcription *in vitro*[14]. This was followed by PAGE analysis to compare terminated and full length products[14]. We took advantage of the ability to resolve differences in termination *in vitro* by transcribing in the presence or absence

of *trans*-**1** and purifying the transcript corresponding to the desired function (early terminated transcripts with ligand, and full length product without ligand). Transcription-based selection accomplished selection of expression. Later rounds of selection resulted in pools that had a higher ratio of terminated:unterminated transcripts compared to previous rounds. Due to this enhancement of termination, clones from the selection were screened for ligand dependent termination. Clones showed various levels of termination, although termination was not linked to the ligand. Since we reasoned the lack of ligand dependent termination was due to poor affinity for the ligand, a new selection was performed based on affinity.

A selection based on affinity for the target on beads was carried out to accomplish one of the two functions necessary for a riboswitch, namely ligand binding by the aptamer platform. The affinity selection was successful in enriching the population of potential riboswitches for binding to *trans*-**1** (Fig 3.4) based on the elution profiles. Clones from the selection were tested by *in vitro* transcription in the presence and absence of *trans*-**1** to screen for termination as previously described. Ligand-dependent termination was never observed despite the apparent binding. It remains uncertain if this is due to lack of selection for transcription termination, poor affinity for the target, or the limitations of screening individual clones. Our efforts shifted to identification of sequences with affinity for *trans*-**1** by structure probing to assure sequences tested presented evidence of affinity for the ligand.

SHAPE was used to probe the structures of potential aptamers from the selected pool with a titration series of *trans*-**1**. Clone 36 was the most promising sequence produced by the affinity selection, as SHAPE profiles showed ligand dependent rearrangement of the clone (Fig 3.8). Despite ligand-dependent changes in confirmation, clone 36 did not display any reproducible difference in termination when transcribed in the presence and absence *trans*-**1**. Structure predictions in the presence and absence of *trans*-**1** revealed the stem loop structures

of both anti-terminator and terminator motifs of clone 36 were not formed as intended. Atpa-Seq can be carried out on selected pools with *trans*-**1** to identify high affinity aptamers, as well as predict the secondary structures of individual sequences. Structure predictions would reveal sequences with a high likelihood of forming an anti-terminator or terminator as originally designed, in so far as sequences with a predicted anti-terminator in the absence of *trans*-**1** and a terminator in presence of *trans*-**1** would be of most interest. The SAM-1 based starting pool contains a ribosome binding site and start codon in addition to a potential transcriptional terminator in order to screen for a translational riboswitch. Ligand-dependent availability of these regulatory elements can be investigated by structure predictions.

Due to the difficulties of identifying a riboswitch based on transcription, our focus shifted to identifying potential translational riboswitches. Mansy et al. used a toehold reporter assay to screen the availability of the ribosome binding site[15] (RBS) of potential riboswitches, the assay resulted in the identification of several new synthetic riboswitches[15] (see 3.8). This method was used for our selection by design of a toehold to use for our selection pool. Fluorescence was used to screen for strand displacement, with a toehold containing a fluorophore on one strand and a quencher on a shorter complementary strand. Upon binding of the riboswitch, the quencher-containing oligonucleotide should be displaced for pool members with translational riboswitch activity[15]. Displacement of the toehold by clone 36 in the presence of the *trans*-**1** was greatly inhibited compared to displacement in the absence of *trans*-**1**, demonstrating ligand dependence and the potential of clone 36 to be a functional translation-based riboswitch. An apparent K_D of ~20 μM was derived using the strand displacement reaction for kinetic analysis. Clone 36 remains to be tested for translational gene expression with μM range of *trans*-**1**. The toehold reporter study indicated clone 36 was an aptamer, and thus fulfilled at least one of the two functions necessary for riboswitch activity.

The reason for selecting a stilbene target was to potentially take advantage of the isomerization properties of the stilbene in the future. A stilbene is capable of photo-reversible *trans* to *cis* isomerization under the influence of light[16]. The expression state of a riboswitch bound to *trans*-**1** can be reversed by releasing the riboswitch; which could be accomplished by changing the isomerization state of the stilbene. To test whether clone 36 could be released from *trans*-**1** upon isomerization to *cis*-**1**, we sought to isomerize the stilbene while it was bound to clone 36. Irradiation of the toehold solution containing clone 36 and *trans*-**1** resulted in greater strand displacement compared to non-irradiated samples. Clone 36 DNA was ligated in a plasmid upstream a GFP reporter to use GFP to screen for expression. Future GFP reporter studies will reveal if gene expression is attenuated by fluorescence feedback. Strand displacement was also used to evolve novel synthetic riboswitches, and the basis of selection was analogous to the fluorescence screening by strand displacement[15].

A selection based on strand displacement was performed using pools from the affinity selection (rounds 6-8). The strand displacement selection was designed to select for riboswitches that are capable of strand displacement only in the presence of *trans*-**1**. The selection was carried out in two steps: 1) a counter selection by incubating the pool with the toehold over an hour to subtract the population that binds the toehold in the absence of ligand followed by 2) a selection step by incubating the pool with the toehold and *trans*-**1** to collect the population that binds in the toehold in the presence of ligand. Screening selected pools by the fluorescence toehold assay revealed a population that was enriched for RNAs that have a slow on rate for the toehold, rather than ligand-dependent availability of the RBS. The above selection schemes implemented produced various results. Selection strategies can result in the evolution of undesired functions, as the correct conditions for a selection can be difficult to foresee and predict. Refinement of selection conditions and alternate selection approaches

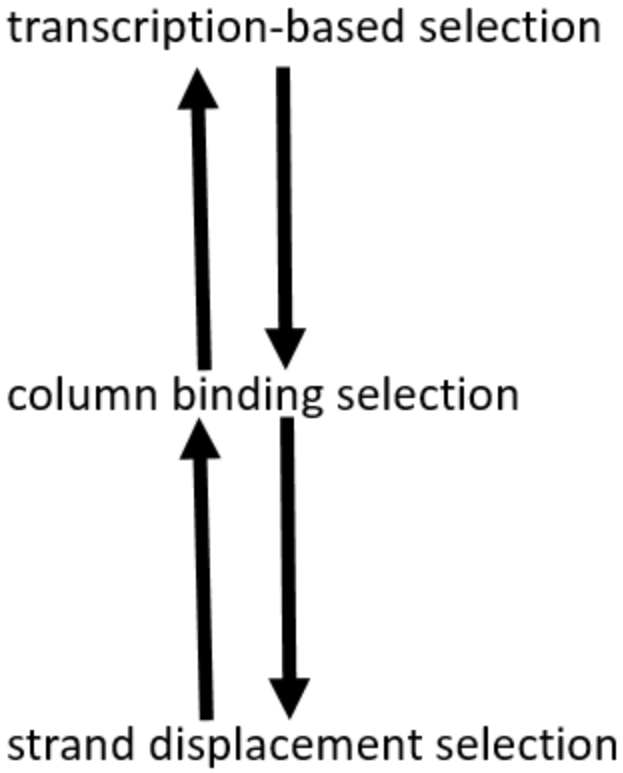
may yield better outcomes.

5.3 Alternate selection schemes

Several selection techniques were developed to discover a stilbene riboswitch (above), each method resulted in the enrichment of one specific function; however, these functions have yet to include a ligand-dependent riboswitch. The *in vitro* transcription selection scheme was successful in isolating RNAs capable of transcriptional termination. The affinity scheme enriched the population for binding *trans*-**1**, without a ligand-dependent expression platform. Results from both affinity and expression selections only produced one of the two components necessary for a ligand dependent riboswitch. Coupling of these two components is necessary for ligand dependent riboswitch activity. In the future the transcription and affinity selection schemes could be combined to select for both functions of the riboswitch.

Serial rounds of a selection results in the enrichment for function with each sequential round. Enhancement of one of the functions (affinity or expression) may result in a reduction of the population that is capable of the other essential function. In the future, transcription and affinity selections could be performed in an oscillating the order of selections to discriminate by both affinity and expression. This method would first consist of a round of transcription-based selection consisting of transcription in the presence of the ligand to collect for terminated product, followed by transcription without the ligand to select the full length product. After a round of transcription-based selection, the selected pool could undergo a round of affinity selection to enforce ligand dependence. This staggered selection scheme is completed when ligand-dependent transcription termination is observed by screening selected pools. Other selection schemes that enhance for either affinity or expression can be combined as well, but it is crucial to follow a round of one function with a round to enhance the second function.

Selection by strand displacement has resulted in the discovery of new synthetic riboswitches, although riboswitches were evolved from a starting pool of a well characterized riboswitch[15]. Our strand displacement selection resulted in RNAs capable of strand displacement with a slow on-rate and displayed no displacement dependence with *trans*-**1**. To improve this method, incubation times could be lowered in order to prevent selection for slow on-rate sequences. Original incubation times were one hour for counter selection and one hour for selection, but the time of the selection step could be reduced to 5-10 minutes to select against slow on-rate binders. Additionally, the RNA pool can be denatured and cooled after the counter selection step, for the pool to remain in a similar state at the start of counter selection. The strand displacement selection method can also be combined with a previous selections method such as the affinity selection or transcription-based selection, however; it is important affinity selections stagger each round of expression selection to maintain ligand dependence.



5.2 Alternate selection schemes. Alternate selection schemes combining multiple methods can be employed to select for both affinity and expression platforms of a riboswitch. Here, it is important that an affinity selection (column binding) takes place between every round of expression selection (transcription-based or strand displacement selections).

5.4 Alternate riboswitch pool designs

The pool was designed with a random region to make a potential aptamer platform followed by a defined expression region consisting of a possible anti-terminator and terminator. Gene expression is often mediated by ligand-dependent allosteric control of two competing stem loops of the riboswitch. In most riboswitches there is some complementarity to the anti-terminator by a sequence (the "anti-anti terminator") upstream of the expression platform. Availability of the anti-anti terminator is dependent on allosteric changes due to ligand binding[14, 17, 18]. The pool, as currently designed, relies on a random region to fulfil both ligand binding and sequestering either the anti-terminator or terminator to mediate allosteric control. Selecting for affinity, or sequestering of the expression platform, greatly reduces the population available for the function yet to be selected for. This can be mediated by staggering selection strategies as described in 5.3 to enhance both functions without losing a large part of the population.

Pools can be designed to improve the odds of allosteric control in response to the binding of a ligand. A pool could be designed with an anti-anti-terminator, an anti-terminator, and terminator to insure all three components necessary for allosteric control are present (Fig 5.3 a). Including an anti-anti-terminator would alleviate the necessity of the random region to contain complementarity to the anti-terminator or terminator, expanding the diversity of potential outcomes. A random region making up the potential aptamer platform would be flanked between the 3' end with an anti-anti-terminator and 5' end with an anti-terminator, followed by the terminator. Many of the known riboswitches sequester the anti-terminator toward the 5' end of the riboswitch by a P1 that has the anti-anti-terminator[14, 18]. Sequestration of the anti-terminator is often based on long range interactions and tertiary contacts between different regions of an RNA. Placement of a highly structured ligand-dependent aptamer

between regulatory elements could bring about aptamer dependent allosteric control[18]. Additionally, an anti-anti-terminator can be engineered upstream clone 36 and tested for ligand-dependent termination based on previously described methods (5.3 a).

It is possible to use a known aptamer for rational design of a riboswitch without selection of an expression platform. Aptamers have been utilized to make novel riboswitches by engineering the aptamer upstream or downstream of an expression element[19]. An aptamer for theophylline was integrated upstream a RBS of a β-galactosidase reporter gene, and expression of the reporter was increased 11-fold in the presence of theophylline[20]. Our data indicate that clone 36 is an aptamer for *trans*-**1**. In the future, clone 36 could be incorporated upstream of different expression elements in order engineer a synthetic riboswitch. Clone 36 could also be utilized in the design of a new pool. The pool could be engineered with a defined aptamer region (clone 36) and a random region that makes up the potential expression platform. The resulting pool would possess the function of affinity, only selection for expression would be necessary. A pool design consisting of clone 36 toward the 5′ end, followed by a random region to select for a potential expression platform, and regulatory elements (run of U's, RBS, start codon) at the 3′ end (Fig 5.3 b) could be used for expression-based selection. The pool would only have to undergo selection for one function, namely expression. Transcription and translation-based selections described in this text could be used to select for ligand dependent gene expression.

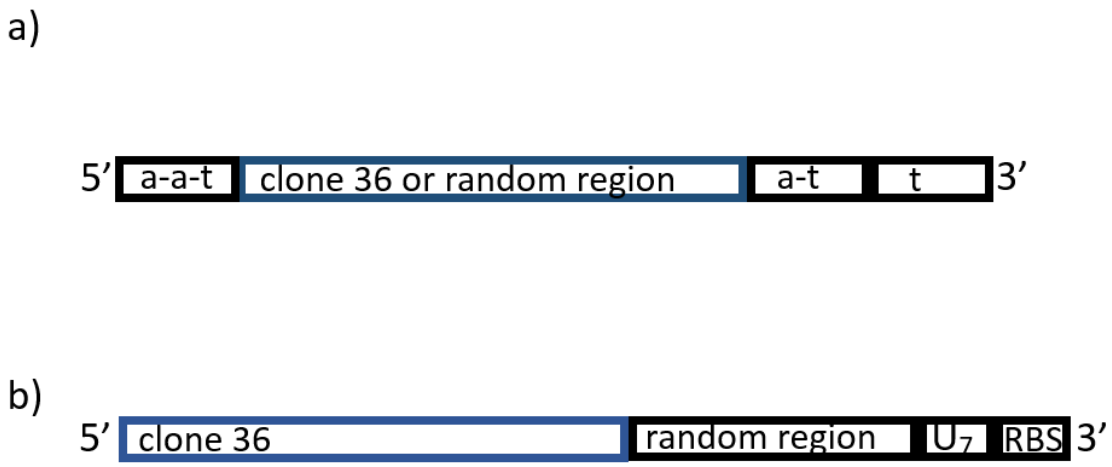


Figure 5.3 Alternate pool designs. Several pools can be used in parallel to increase the odds of discovering a riboswitch. Aptamer regions are outlined in blue and expression regions outlined in black. (a) A pool that contains an anti-anti-terminator (a-a-t), anti-terminator (a-t), and terminator (t). Allosteric control is achieved by sequestering the anti-terminator by the anti-anti-terminator in many riboswitches. Placing an aptamer platform between the anti-anti-terminator and anti-terminator may induce allosteric control based on the ligand-dependent formation of a structured aptamer in between the two elements. (b) A pool containing a defined aptamer and a downstream random region making up a potential expression platform. This pool can undergo selection based on expression alone, as the aptamer requirement is already fulfilled.

5.5 Overall conclusions

The scope of the work described herein was undertaken to discover functional RNAs through laboratory selection and evolution. Many of the methods employed resulted in selection of RNA without the desired properties. *In vitro* selection is an established technique to evolve RNA on the basis of affinity. There have been many successful selections based on iterative rounds of selection and amplification, resulting in aptamers from synthetic[21] and genomic pools[22]. Coupling affinity to a second desired function raises many difficulties. Desired targets may have interesting properties that could be exploited by an aptamer, although ligand binding alone does not insure activation of the wanted function from the ligand.

One of the major goals of our efforts was to accomplish an aptamer capable of luminescence enhancement upon binding to its target. Aptamers capable of enhancing fluorescence of a ligand have been discovered through selection, an RNA with GFP-like properties was selected through 5 rounds of *in vitro* selection[23]. The aptamer, Spinach, exhibits enhanced fluorescence when bound to its target. It was revealed that emission enhancement was largely due to the immobilization of ligand when bound to Spinach[24]. It is difficult to predict or rationalize the requirements necessary for an aptamer to be capable of an energy transfer to a lanthanide chelate that results in luminescence. Our results indicate that binding alone may not be sufficient to induce lanthanide luminescence. On the other hand, we discovered a stilbene molecule, *trans*-**1**, that displays a large increase in emission in the presence of RNA. It may be possible to use*trans*-**1** and lanthanides as aptamer targets for their optical properties, however; more careful considerations must be made when designing their selection and screening methods.

Another major theme of this work describes the development of a synthetic riboswitch by selection. Naturally occurring riboswitches occurred through natural selections, desigining

the correct laboratory conditions for a riboswitch selection has been difficult. Riboswitches contain two essential platforms for function, an aptamer platform and an expression platform[25]. Coupling these platforms by ligand-dependent allosteric control is necessary to mediate gene expression. We demonstrated the ability to evolve either platform independently, but coupling both platforms in a ligand-dependent manner has proven to be a challenge. Coupling of the two platforms may be accomplished by combining the methods capable of developing each function independently. Most importantly, a scheme capable of selecting for both functions must be adopted. A *trans*-**1** dependent aptamer switch, clone 36, was discovered by affinity selection and screening by strand displacement. It remains to be seen if clone 36 is capable of altering gene expression in cells or by *in vitro* translation. This promising sequence could be further investigated in its current state or engineered into a new pool for selection. The selected pools for both lanthanide and *trans*-**1** targets may contain active sequences yet uncharacterized. Many of our challenges arise from the time and work necessary to test individual sequences.

Identifying aptamers is a prominent challenge of selections, selected pools may contain too many sequences for individual testing and active sequences that are not well represented. Apta-Seq was developed to find and analyze all relevant sequences of a selected pool. Apta-Seq data can provide affinity and structural insights of all sequences in a pool, easing the identification processes and aid in focusing on important sequences in a pool. We focused on a previously selected pool to develop and validate the technique, resulting in affirming previous findings in addition to the discovery of three new aptamers. Apta-Seq could be combined with the other selections performed within this text to determine the affinity of sequences and gain structural predictions of sequences with and without the target. Functional capabilities of RNA bound to a lanthanide or *trans*-**1** may have gone undiscovered due to the

limitations of screening individual sequences. Apta-Seq would not uncover RNAs capable of RNA induced emission enhancement or mediating gene expression directly, nevertheless it could provide great insight on which sequences have a higher potential to be functionally active.

Our work illustrates many of the difficulties that a selection can bring about. Using a selection pool capable of the desired outcome, implementing appropriate selection conditions, and methods used to screen sequences are all critical to the discovery of functional RNA. We describe many selection methods and their results. Although many of these methods may not have produced the desired RNA, the results are useful in understanding the capabilities of each method. Selection for a luminescent aptamer and stilbene riboswitch may require a combination of methods or entirely new methods to be employed, our investigations indicate that selection for both affinity and function are required for our desired outcomes.

1. Wickramaratne, T.M. & Pierre, V.C. Turning an Aptamer into a Light-Switch Probe with a Single Bioconjugation. *Bioconjugate Chemistry* **26**, 63-70 (2015).
2. Feig, A.L., Panek, M., Horrocks, W.D., Jr. & Uhlenbeck, O.C. Probing the binding of Tb(III) and Eu(III) to the hammerhead ribozyme using luminescence spectroscopy. *Chem Biol* **6**, 801-810 (1999).
3. Feig, A.L., Scott, W.G. & Uhlenbeck, O.C. Inhibition of the Hammerhead Ribozyme Cleavage Reaction by Site-Specific Binding of Tb(III). *Science* **279**, 81 (1998).
4. Scott, W.G., Horan, L.H. & Martick, M. The Hammerhead Ribozyme: Structure, Catalysis and Gene Regulation. *Progress in molecular biology and translational science* **120**, 1-23 (2013).
5. Bünzli, J.-C.G. Lighting up cells with lanthanide self-assembled helicates. *Interface Focus* **3** (2013).
6. Cable, M.L., Kirby, J.P., Gray, H.B. & Ponce, A. Enhancement of anion binding in lanthanide optical sensors. *Accounts of chemical research* **46**, 2576-2584 (2013).
7. Xu, L.-J., Xu, G.-T. & Chen, Z.-N. Recent advances in lanthanide luminescence with metal-organic chromophores as sensitizers. *Coordination Chemistry Reviews* **273–274**, 47-62 (2014).
8. Bunzli, J.-C.G. & Piguet, C. Taking advantage of luminescent lanthanide ions. *Chemical Society Reviews* **34**, 1048-1077 (2005).
9. Meshkova, S.B., Topilova, Z.M., Lozinskii, M.O. & Bol'shoi, D.V. Luminescence quenching of lanthanides in complexes with β-diketones containing different fluorinated radicals. *Journal of Applied Spectroscopy* **64**, 229-233 (1997).
10. Ruston, L.L., Robertson, G.M. & Pikramenou, Z. Luminescence Screening Assays for the Identification of Sensitizers for Lanthanides Based on the Controlled Formation of Ternary Lanthanide Complexes with DTPA–Bisamide Ligands. *Chemistry – An Asian Journal* **5**, 571-580 (2010).

11. Waldeck, D.H. Photoisomerization dynamics of stilbenes. *Chemical Reviews* **91**, 415-436 (1991).
12. Debler, E.W. et al. Deeply Inverted Electron-Hole Recombination in a Luminescent Antibody-Stilbene Complex. *Science* **319**, 1232 (2008).
13. Mandal, M., Boese, B., Barrick, J.E., Winkler, W.C. & Breaker, R.R. Riboswitches Control Fundamental Biochemical Pathways in Bacillus subtilis and Other Bacteria. *Cell* **113**, 577-586 (2003).
14. Winkler, W.C., Nahvi, A., Sudarsan, N., Barrick, J.E. & Breaker, R.R. An mRNA structure that controls gene expression by binding S-adenosylmethionine. *Nat Struct Mol Biol* **10**, 701-707 (2003).
15. Martini, L., Ellington, A.D. & Mansy, S.S. An in vitro selection for small molecule induced switching RNA molecules. *Methods*.
16. Birks, J.B. The photo-isomerization of stilbene. *Chemical Physics Letters* **38**, 437-440 (1976).
17. Mandal, M. & Breaker, R.R. Gene regulation by riboswitches. *Nat Rev Mol Cell Biol* **5**, 451-463 (2004).
18. Lussier, A., Bastet, L., Chauvier, A. & Lafontaine, D.A. A Kissing Loop Is Important for btuB Riboswitch Ligand Sensing and Regulatory Control. *Journal of Biological Chemistry* **290**, 26739-26751 (2015).
19. Wachsmuth, M., Findeiß, S., Weissheimer, N., Stadler, P.F. & Mörl, M. De novo design of a synthetic riboswitch that regulates transcription termination. *Nucleic Acids Research* **41**, 2541-2551 (2013).
20. Desai, S.K. & Gallivan, J.P. Genetic Screens and Selections for Small Molecules Based on a Synthetic Riboswitch That Activates Protein Translation. *Journal of the American Chemical Society* **126**, 13247-13254 (2004).
21. Ellington, A.D. & Szostak, J.W. In vitro selection of RNA molecules that bind specific ligands. *Nature* **346**, 818-822 (1990).
22. Salehi-Ashtiani, K., Lupták, A., Litovchick, A. & Szostak, J.W. A Genomewide Search for Ribozymes Reveals an HDV-Like Sequence in the Human CPEB3 Gene. *Science* **313**, 1788-1792 (2006).
23. Paige, J.S., Wu, K.Y. & Jaffrey, S.R. RNA Mimics of Green Fluorescent Protein. *Science* **333**, 642-646 (2011).
24. You, M. & Jaffrey, S.R. Structure and Mechanism of RNA Mimics of Green Fluorescent Protein. *Annual Review of Biophysics* **44**, 187-206 (2015).
25. Breaker, R.R. Riboswitches and the RNA World. *Cold Spring Harbor Perspectives in Biology* **4** (2012).

Chapter 6

Materials and methods

6.1 Radiolabeled [α-^{32} P] RNA Transcription

RNA was transcribed at 37° C (unless otherwise noted) containing 40 mM tris chloride, 10 mM dithiothreitol (DTT), 2 mM spermidine, 2.5 mM each CTP, GTP, and UTP, 250 mM ATP, 2.25 µCi [α-^{32}P]-ATP (Perkin Elmer, Waltham, MA, USA), 25 mM $MgCl_2$, one unit of T7 RNA polymerase, and 6.5 µL of DNA template. Transcripts were purified using denaturing PAGE.

6.2 RNA Transcription

RNA was transcribed at 37° C (unless otherwise noted) containing 40 mM tris chloride, 10 mM dithiothreitol (DTT), 2 mM spermidine, 5 mM each CTP, GTP, and UTP, ATP, 25 mM $MgCl_2$, one unit of T7 RNA polymerase, and 6.5 µL of DNA template. Transcripts were purified using denaturing PAGE.

6.3 3′-Terminal Labeling

Purified RNA transcripts were ligated at 37 °C for 3 hours containing RNA ligase buffer (New England Biolabs [NEB] Upswich, MA, USA), one unit of T4 RNA ligase (NEB), and 2 µCi [5′-^{32}P] cytidine 3′, 5′-bisphosphate (Perkin Elmer) and PAGE purified.

6.4 In-line probing[1]

3′-end radiolabeled RNA was incubated with various concentrations of ligand (1:1 chelator:lanthanide) in binding buffer containing containing 140 mM KCl, 10 mM NaCl, 10 mM Tris chloride, pH 7.5, and 5 mM $MgCl_2$ over two nights. Hydrolyzed RNAs were resolved on a

denaturing PAGE, exposed to phosphorimage screens and scanned by GE typhoon phophoimager.

6.5 *In vitro* selection[2] for lanthanide aptamers

The DNA pool with a length of 100 nucleotides containing a random region of 44 nucleotides was synthesized by Keck Oligo Synthesis Resource and transcribed with a diversity of ~ 10^{15}. Purified [α-32 P] RNA transcripts were precipitated, dried, and resuspended in 50 µL binding buffer containing 140 mM KCl, 10 mM NaCl, 10 mM Tris chloride, pH 7.5, and 5 mM $MgCl_2$ and heated to 70° C for 3 min before loading on silica-coated iron oxide magnetic beads grafted with DTPA or EDTA on the surface (Bioclone Inc, San Diego, CA, USA) and an equivalent of Eu (III) or Tb (III). For counter selections, purified RNA was incubated on EDTA or DTPA beads without incubation with the target lanthanide. Flow through was collected after the columns were shaken for 10 min at room temperature. The beads were washed with 50 µL of binding buffer for 5 minutes at room temperature for a total of 8 times. Potential aptamers were eluted with the same buffer supplemented with 5 mM of the target lanthanide chelate complex (1:1 chelator with lanthanide) with 30 min of shaking at room temperature for a total of 4 times. Each fraction was analyzed for radioactivity using a liquid scintillation counter. Elutions were pooled, precipitated, dried, and resuspended in 10 µL H_2O for reverse transcription.

6.6 Luminescence screening for lanthanide aptamers

Various concentrations of purified RNA were incubated with various concentrations of 1:1 chelator:lanthanide in binding buffer and heated to 70 °C for 3 min then cooled to room temperature for 10 minutes. Emission spectra were measured using an excitation range of

250 nm to 450 nm in steps of 10 nm. For each excitation step, emission was recorded in the range of 500 nm to 800 nm.

6.7 Competition assay

Varying amounts of purified RNA were incubated with 1 nM EDTA-Eu in binding buffer for 2 hours. Then 10 µM of dipicolinate (DPA) was added to each sample and emission spectra at 615 nm were taken using an excitation of 355 nm.

6.8 Tetracycline fluorescence assay

DNA containing the sequence of tetracycline aptamer cb28 was ordered from Integrated DNA Technologies (IDT). Varying amounts of tetracycline was incubated 1:1 with Eu (III) or Tb (III) in binding buffer for 10 minutes to form tetracycline-lanthanide complex. Tetracycline-lanthanide complexes were incubated with varying amounts of purified RNA for 10 minutes at room temperature. Emission spectra at 615 were taken using an excitation of 355 nm.

6.9 *In vitro* selection for *trans*-1 stilbene

The DNA pool with a length of 165 nucleotides containing a random region of 45 nucleotides was synthesized by the Keck Oligo Synthesis Resource and transcribed with a diversity of ~ 10^{15}. Purified [α-32 P] RNA transcripts were precipitated, dried, and resuspended in 50 µL binding buffer containing 140 mM KCl, 10 mM NaCl, 10 mM Tris chloride, pH 7.5, and 5 mM $MgCl_2$ and heated to 70° C for 3 min before loading on agarose beads for counter selection. For selection, purified RNA was incubated on agarose labelled with *trans*-**1** on the surface. Flow through was collected after the columns were shaken for 5 min at room temperature. Beads were washed with 50 µL of binding buffer for 5 minutes at room temperature for a total of 6 times. Potential

aptamers were eluted twice with denaturing buffer consisting of 7 M Urea and 5 mM EDTA in 0.5 % TBE and heated for 5 minutes at 95 °C for five minutes. Each fraction was analyzed for radioactivity using a liquid scintillation counter. Elutions were pooled, precipitated, dried, and resuspended in 10 µL H_2O for reverse transcription.

6.10 Selection by photoswitching

Agarose beads with cis-1 were incubated with purified RNA for 20 minutes. Beads were washed with binding buffer 6 times. 6 elutions were performed by irradiation at 365 nm ~ 30 seconds in binding buffer. Elutions were pooled, precipitated, dried, and dissolved in H_2O for reverse transcription.

6.11 *In vitro* selection by ligand-dependent transcription termination

Radiolabeled [α-^{32}P] transcription was prepared in the presence or absence of various concentrations of *trans*-**1**. Transcripts were fractionated on 7-15% denaturing PAGE in order to resolve full-length and terminated products. Full length product was collected for transcription without ligand, terminated product was collected for transcription in the presence of *trans*-**1**. Products were purified and eluted in 300 mM KCl, EtOH precipitated, dried, and dissolved in 10 µL H_2O for reverse transcription.

6.12 Selective 2'-hydroxyl acylation and primer extension (SHAPE) of clone 36

Purified transcripts of clone 36 were heated to 70° C for 5 min. 10 nM of purified RNA was added to a buffer containing 140 mM KCl, 10 mM NaCl, 10 mM Tris chloride, pH 7.5, and 5 mM $MgCl_2$. A dilution series of *trans*-**1**, prepared from 20 nM to 20 µM, was aliquoted to the RNA in buffer and incubated at room temperature (~23 °C) for 1 minute. 50 mM 2-

(azidomethyl)nicotinic acid acyl imidazole was added to the mixture and the reaction was incubated for 45 minutes at room temperature. Reactions containing no SHAPE reagent were substituted with 10 % DMSO. Reactions were precipitated with 10 µL 3M KCl, 1 µL glycoblue, 89 µL H_2O, and 300 µL 98 % ethanol. SHAPE products were raised in 10 µL of H_2O and fractionated on 12-15 % denaturing PAGE to resolve SHAPE modifications.

6.13 Structure prediction of clone 36

Structure predictions were performed using RNAfold to fold the sequence in the presence and absence of *trans*-**1**. Structure predictions were made based on analysis of Fig 3.8. ImageJ was used to determine if positions became more or less highly modified by band intensity with increasing concentration of *trans*-**1**. Positions that became more highly modified were given an input constraint of "x" corresponding to not being base paired. Positions that become less highly modified were given an input constraint of "|" to indicate pairing with another base.

6.14 Plasmid construction for GFP screening

Plasmid pBbB5a-GFP was ordered from addgene (Cambridge, MA). Two restriction sites were inserted into the pool by polymerase chain reaction (PCR). The pool was amplified with a forward primer with an EcoRI restriction site (5′ TGCGATCGATATCGAGAATTCTAATACGACTCACTATAGGGA 3′) and a reverse primer with Nde I restriction site (5′ AGTAGCTGGTCGACCATATGAAATAAAAACCCCTTCTTC 3′), the amplicon was purified using QIAquick PCR purification kit (Qiagen). 1 µg the purified pool with restriction sites was double-digested with 10 units of EcoRI (NEB) and 10 units of NdeI (NEB) in Cutsmart buffer (NEB) then purified using the QIAquick PCR purification kit. 500 ng of pBbB5a-GFP was

double digested with 10 units of Ecor I (NEB) and 10 units of NDE I (NEB) in Cutsmart buffer (NEB) then purified by QIAquick PCR purification kit. 20 ng of the digested pool was combined with 50 ng of digested pBbB5a-GFP and ligated with T4 DNA ligase in 1x reaction buffer and purified with QIAquick PCR purification kit (Qiagen).

6.15 Screening potential *trans*-1 binders based on optical activity

A dilution series of purified RNAs was incubated with 50 nM *trans*-**1** in binding buffer. Emission spectra at 410 nm were collected using an excitation of 355 nm.

6.16 Strand displacement selection

Oligonucleotides for a dsDNA toehold reporter were ordered by IDT. The toehold consists of a longer biotinylated strand extending to the complement of the (alternate) Shine-Dalgarno sequence (CCCCT) with biotin at the 5′ end and a shorter complementary strand. A 10 μM toehold stock was prepared in binding buffer by combining the biotinylated strand with the shorter strand 2:1. The toehold was annealed by incubating at 95° C for 1 minute then cooled to 25° C for 5 minutes.

Biotin reporter: 5′/52-Bio/TAC CTG CAA GCT TCC CTT TTC AAA ATA AAA ACC CCT 3′

Shorter strand: 5′ ATT TTG AAA AGG GAA GCT TGC AGG TA 3′

Purified RNA was incubated with the toehold reporter in binding buffer for 37 ° C for 1 hour.

Counter selection:

4 washes were performed by incubating streptavidin magnetic beads (Life Technologies) to the solution for 20 minutes. Beads are discarded after magnetic separation.

Selection

20 µM *trans*-**1** was added to remaining supernatant and incubated for an hour. Beads were collected after magnetic separation. Beads were resuspended in 10 µL H_2O for reverse transcription.

6.17 Detection of strand displacement

Oligonucleotides for a dsDNA toehold reporter were ordered fpr IDT. The toehold consists of two unequal length oligonucleotides to detect strand displacement by fluorescence. The longer strand extends to the complement the (alternate) Shine-Dalgarno sequence (CCCCT) of the RNA pool. The shorter stand (Rep Q) was tagged with a 3′ Iowa black quencher molecule and was complementary to the reporter strand. The longer strand (Rep F) contained a 5′ fluorescein. A 1 µM toehold stock solution of 2:1 of Rep Q/Rep F was prepared and used for all reactions. 100 nM of the toehold stock solution was annealed in a 50 µL solution in binding buffer. The solution was incubated at 95 °C for 1 min then annealed at 25 °C for 5 min. Strand displacement was initiated in a solution of 100 µL by adding 100 nM of RNA and 50 nM of the annealed toehold stock in binding buffer. When present, various concentrations of *trans*-**1** were added to the mixture to test for ligand dependent displacement.

Rep F: 5′ /5FluorT/TA CCT GCA AGC TTC CCT TTT CAA AAT AAA AAC CCC T 3′

Rep Q: 5′ ATT TTG AAA AGG GAA GCT TGC AGG TA/3IABkFQ/ 3′

Fluorescence was recorded using BioTek Synergy H1 plate reader (BioTek, Winooski, VT). Samples were measured in a Falcon 384 well Optilux Black/Clear Flat Bottom plate (Corning, Corning, NY). The following conditions were used for all experiments:

excitation: 485, emission: 520, optics: Top, Gain :100, lamp energy high.

For kinetic experiments, emission intervals were collected every 2 mins. For strand displacement recovery experiments, kinetics were paused and the plate was ejected to

prevent irradiation of control samples by covering wells with foil. Irradiation was achieved by taking emission spectra using an excitation of 342 nm and emission range of 470-550 nm in steps of 5 nm.

6.18 Transcription for Apta-Seq

RNA was transcribed for 2 hours at 37 °C in 400 µl of 40 mM tris chloride, 10% dimethyl sulfoxide (DMSO), 10 mM dithiothreitol (DTT), 2 mM spermidine, 5 mM each rCTP, rGTP, rUTP, and rATP, 20 mM $MgCl_2$, one unit of T7 RNA polymerase, and ~0.5 µM DNA template. Transcripts were purified by 7 % polyacrylamide gel electrophoresis (PAGE) under denaturing conditions (7M urea). RNA was eluted from gel slices into 400 µL of 400 mM KCl and precipitated by adding 800 µL of 100% ethanol at –20 °C.

6.19 Primer phosphorylation

Primer labeling was prepared in a total volume of 20 µL. 20 µM of primer, 1x T4 Polynucletoide Kinase (PNK) ligase buffer (NEB), 1 unit T4 PNK (NEB), and 0.5 µCi [γ-32P]ATP was incubated at 37 °C for one hour then purified from denaturing PAGE.

6.20 Synthesis of the SHAPE reagent

The SHAPE reagent, 2-(azidomethyl)nicotinic acid acyl imidazole, was synthesized following a previously described protocol.[3]

6.21 Selective 2′-hydroxyl acylation and primer extension (SHAPE) for Apta-Seq

SHAPE reactions were prepared in a total volume of 10 µL. RNAs (pools or individual aptamer sequences) were resuspended in water and heated to 70° C for 3 min. 1 µM of purified RNA

was added to a buffer containing 140 mM KCl, 10 mM NaCl, 10 mM Tris chloride, pH 7.5, and 5 mM $MgCl_2$. A dilution series of 5′-adenosine triphosphate (ATP) was prepared using a 1:1 stock of ATP:Mg^{2+}. 1 µM to 10 mM ATP were aliquoted to the RNA in buffer and incubated at room temperature (~23 °C) for 1 minute. 50 mM 2-(azidomethyl)nicotinic acid acyl imidazole was added to the mixture and the reaction was incubated for 45 minutes at room temperature. Reactions containing no SHAPE reagent were substituted with 10% DMSO. Reactions were precipitated with 10 µL 3M KCl, 1 µL glycoblue, 89 µL H_2O, and 300 µL 98% ethanol.

6.22 Primer extension

RNA pellets were reconstituted in a 10 µL reaction volume containing 0.1 µM 5′-[^{32}P]-radiolabeled reverse transcription DNA primer, 2 µl 5x M-MuLV Reverse Transcriptase Reaction Buffer (New England Biolabs), 1 unit M-MuLV enzyme, and 500 µM each deoxynucleotide triphosphate. Extensions were performed at 42° C for 15 minutes. 400 mM NaOH was added and the reaction was incubated at 95 °C for 5 minutes to hydrolyze the RNA. Reactions were precipitated with 10 µL 3M KCl, 1 µL glycoblue, 89 µL H_2O, and 300 µL 98% ethanol. Complementary DNA (cDNA) was resolved using 12% denaturing PAGE or amplified for high-throughput analysis.

6. 23 K_D analysis

Reverse transcription (RT) stops were analyzed using ImageQuant. Band intensities were analyzed by measuring the total radioactivity intensity of each lane and the intensity of the band of interest. RT stops were normalized by dividing band intensities by the overall

intensity of the respective lane. Ratios were plotted in Excel as a function of ligand concentration and modeled with a dissociation constant equation for each ligand:

$$\text{Fraction Bound} = \frac{\left(\frac{[\text{ligand}]}{([\text{ligand}] + K_D)} - \text{Baseline}\right)}{\text{Range}}.$$

The model was fit to the data using a linear least-squares analysis and the Solver module of Microsoft Excel. The same model was used to extract ATP K_D from the Apta-Seq data. Because some of the apparent binding constants are near the maximum of the titration and in many cases the rising SHAPE signal did not level off at the highest ATP concentration, we used the positions where the SHAPE signal decreased with ATP concentration and approached zero for K_D modeling (see for example Fig. 1B).

6.24 SHAPE-Seq library and primer design

Round 6 of an *in vitro* selection for an ATP aptamers from a human genomic library [4] was used as the library for Apta-Seq. Libraries were given individual barcodes based on concentration of ligand used during SHAPE. Ten libraries in total were made using a ligand titration of 1-10000 µM, no ligand control, and no-SHAPE (only DMSO, as described above) control. Apta-Seq primers contain reverse primer for the pool of interest and flanking Illumina primers in order to barcode and sequence primer extensions by Illumina Sequencing. Apta-Seq primers for reverse transcription primer extension were designed 5′ to 3′ with the following: Illumina forward primer reverse-complement, NotI digestion site, Illumina reverse primer, reverse primer for RNA of interest

5′

AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT GCGGCCGC GTGACTGGAGTTCAGACG

TGTGCT CTTCCGATC CTGAGCTTGACGCA 3′

Primers for amplification were designed 5′ to 3′ with the following:

Forward primer containing Illumina forward adapter and primer.

5′ AATGATACGGCGACCACCGAGATCT ACACTCTTTCCCTACACGACGCTCTTCCGATCT 3′

Reverse primer containing Illumina reverse adapter, barcode, and Illumina reverse primer.

5′ CAAGCAGAAGACGGCATACGAGAT [barcode]

GTGACTGGAGTTCAGACGTGTGCTCTTCCG 3′

6.25 Primer Extension for Apta-Seq

Primer extension was carried out with Apta-Seq primer as described above. cDNA was self-ligated in a 20 µL reaction using CircLigaseII Reaction Buffer (Epicentre), 2.5 mM $MnCl_2$, 50 µM ATP, and 5 units CircLigase ssDNA Ligase. Reactions were precipitated with 10 µL 3M KCl, 1 µL glycoblue, 89 µL H_2O, and 300 µL 98% ethanol and cDNA was reconstituted in 20 µL H_2O. Polymerase chain reaction (PCR) was performed using 1 µM each of forward and reverse primers, cDNA template, DreamTaq Master Mix (Thermo Fisher) and amplified for 16 cycles (denaturing 94 °C, 30 s, annealing 55 °C, 30 s, and elongation 72 °C, 30 s). Amplicons were sequenced on Illumina HiSeq 2500 at the UCI Genomics Facility.

6.26 SHAPE-Seq reactivity mapping

Galaxy (https://usegalaxy.org/) and the Structurefold module[5] were used to map and determine the SHAPE reactivity of aptamers. Forward reads of libraries were used to analyze RT stops. Adapters and primers (CAATGCGTCAAG) were clipped using Clip adapter sequences on galaxy. Default settings were used except for minimum sequence length of 10, and an output of both clipped and non-clipped sequences. Clipped libraries were then

processed using StructureFold, a series of web-based programs to characterize RNA that have undergone a SHAPE modifications. Aptamers were mapped to selected libraries using Iterative Mapping on Galaxy. Default settings for mapping were used except for minimum read length of 12 nucleotides and 3 mismatches allowed (-v 3). RT stops counts were calculated using Get RT Stop Counts on Galaxy. RT stop counts were derived using mapped files from Iterative Mapping and aptamers as the reference sequences. RT stop counts were exported to Excel and normalized to the total stop counts for each aptamer library. Percentage of counts for each position was derived by dividing each position by the total number of RT counts for each library.

6.27 Structure prediction for Atpa-Seq

The Reactivity Calculation module on galaxy was used to get SHAPE reactivities by using the output from the Get RT Stop Counts module. The RT stop counts for 10 mM ligand concertation was used as the (+) library and no ligand for (-) library, both in presence of SHAPE reagent. Default settings were used except for Nucleotide specificity was changed to AUCG. To predict structure RNAprobing web server was used. All default settings were used according to the Washietel et al. SHAPE method[6].

1. Regulski, E.E. & Breaker, R.R. in Post-Transcriptional Gene Regulation. (ed. J. Wilusz) 53-67 (Humana Press, Totowa, NJ; 2008).
2. Ellington, A.D. & Szostak, J.W. In vitro selection of RNA molecules that bind specific ligands. *Nature* **346**, 818-822 (1990).
3. Spitale, R.C. et al. Structural imprints in vivo decode RNA regulatory mechanisms. *Nature* **519**, 486-490 (2015).
4. Vu, Michael M.K. et al. Convergent Evolution of Adenosine Aptamers Spanning Bacterial, Human, and Random Sequences Revealed by Structure-Based Bioinformatics and Genomic SELEX. *Chemistry & Biology* **19**, 1247-1254 (2012).
5. Tang, Y. et al. StructureFold: genome-wide RNA secondary structure mapping and reconstruction in vivo. *Bioinformatics* **31**, 2668-2675 (2015).

6. Washietl, S., Hofacker, I.L., Stadler, P.F. & Kellis, M. RNA folding with soft constraints: reconciliation of probing data and thermodynamic secondary structure prediction. *Nucleic Acids Research* (2012).