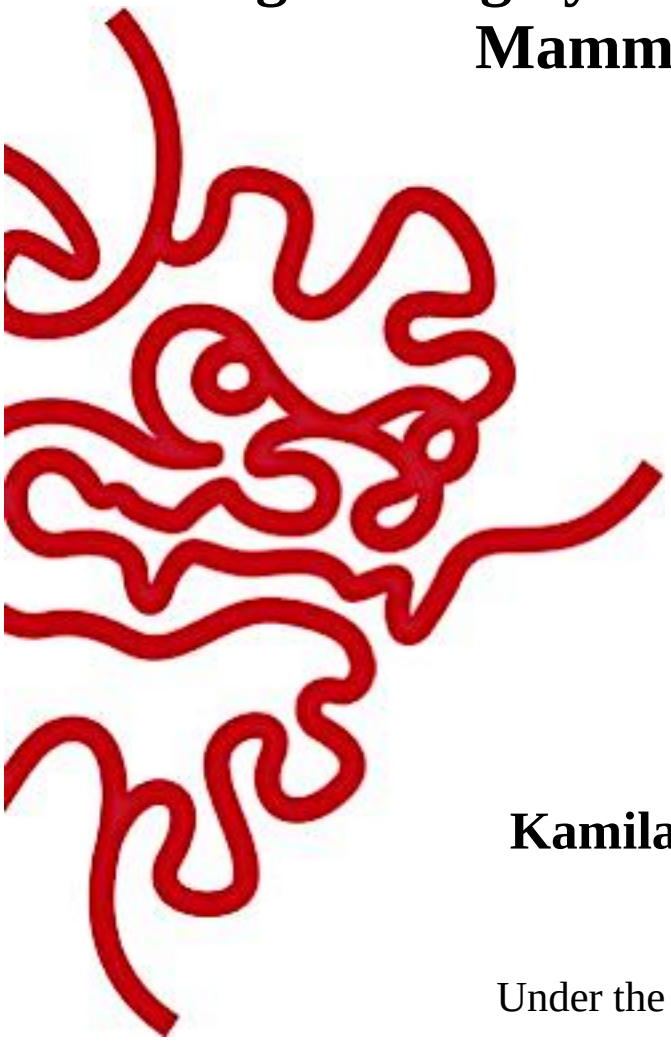
Okinawa Institute of Science and Technology Graduate University

Thesis submitted for the degree

Doctor of Philosophy

Engineering Synthetic Riboswitches for Mammalian Cells



by

Kamila Mustafina

Under the Supervision of:

Prof. Dr. Yohei Yokobayashi

February 2021

Declaration of Original and Sole Authorship

I, Kamila Mustafina, declare that this thesis entitled “Engineering Synthetic Riboswitches for Mammalian Cells” and the data presented in it are original and my own work.

I confirm that:

- No part of this work has previously been submitted for a degree at this or any other university.
- References to the work of others have been clearly acknowledged. Quotations from the work of others have been clearly indicated and attributed to them.
- In cases where others have contributed to part of this work, such contribution has been clearly acknowledged and distinguished from my own work.
- None of this work has been previously published elsewhere, with the exception of the following:

Mustafina, K., Fukunaga, K., & Yokobayashi, Y. (2020). Design of Mammalian ON-Riboswitches Based on Tandemly Fused Aptamer and Ribozyme. *ACS Synthetic Biology*, *9*(1), 19-25. doi: 10.1021/acssynbio.9b00371

Date: February 17, 2021

Signature: 

Abstract

Riboswitches are natural and artificial noncoding RNA elements capable of controlling gene expression in response to chemical signals without direct involvement of protein factors. One strategy for engineering synthetic riboswitches is to combine an aptamer –a short RNA sequence that specifically binds to a ligand– and a self-cleaving ribozyme to create an aptazyme whose self-cleavage activity is regulated by the aptamer ligand. These aptazymes can be embedded in the 3'UTR of mRNAs to chemically control gene expression in mammalian cells. This property of riboswitches opens a wide area of applications in biology and medicine. However, engineering riboswitches that function efficiently in mammalian cells remains challenging, partly due to the difficulties associated with generating and screening aptamers and aptazymes that function in the cellular environment. In this thesis, I introduce two new ribozyme scaffolds for aptazyme engineering in mammalian cells. First, I identified highly active variants in mammalian cells from the twister and pistol ribozyme families. Then I used them as scaffolds for a new aptazyme architecture, where the aptamer is placed immediately upstream of the ribozyme in a tandem configuration. I optimized this design in mammalian cells, and then generated randomized libraries of 4096 aptazyme variants for high-throughput *in vitro* screening to identify switches with high on-off ratios. Although the method allowed characterization of a large number of variants, their activities were not always reproducible when tested in cells. Therefore, in addition to *in vitro* screening, I explored rational design approaches for the same tandem architecture. I fine-tuned the activity of the aptazyme by systematically varying the length of the inserted competing stem and introducing single-nucleotide mismatches and spacers. Using this method, I developed mammalian riboswitches with on-off ratios greater than 6.0 for the twister scaffold, and greater than 5.0 for the circularly permuted pistol scaffold. Lastly, learning from the experience of high-throughput *in vitro* screening and rational design in cells, I used high-throughput sequencing to directly screen for functional aptazymes in mammalian cells by quantifying the uncleaved fractions of aptazyme-embedded mRNAs. I verified this method with a small twister ribozyme library containing 256 variants and then applied it for a larger circularly permuted pistol ribozyme library consisting of 1024 variants. This work expands both the tools and the methods available in the field of RNA engineering. Rational and high-throughput design strategies developed in this thesis can be applied to generate other RNA devices for biomedical and synthetic biology applications.

Acknowledgements

I want to thank my supervisor Yohei Yokobayashi for guiding me and always providing constructive and timely feedback. His thoughtful ideas and eagerness to cooperate kept me motivated throughout this PhD. Thanks to his professionalism and organizational skills I had no other things to worry about but my research.

I want to thank all members of Nucleic Acid Chemistry and Engineering Unit for their help and support throughout this PhD. Thank you to Yoko-san for setting up circularly permuted pistol project and helping me with cell culture experiments. Thank you to Gear, Keisuke Morita, Sasha and Kinuko for their help with the circularly permuted pistol project. Thank you to Keisuke Fukunaga (2) for his help with ITC measurements for twister-guanine project. Thank you to Shungo for all the time and effort with setting up *in vitro* screening experiments and teaching me experimental design. Thank you to Kei for introducing me to the field of Synthetic Biology, to Mohammed for experimental tips and sharing his wisdom on academic life and to Narae for teaching me how to keep cells happy. Thank you to Samuel and Takeshi for helpful comments in our weekly meetings and at other times. Thank you to all members of the lab for support and healthy and productive atmosphere at work.
I want to thank all members of Synthetic Biology Journal club for exciting scientific discussions with a good share of humour and criticism. Scientific debates of Mohammed, Laura, Takeshi and Gear were some of the most fascinating moments in my academic life.

I want to acknowledge help from the OIST DNA sequencing section for performing next-generation sequencing for this work.
I am grateful to the Student Support section at OIST for taking care of all my paperwork and travels to let me focus on research.

Thank you to Sasha and Ainash for their discussion and feedback on Chapter 3 and continuous support that cheered me up at difficult times.
Thank you to my family for endless support, sensible advice and optimism about everything I do. Thank you to all my friends for their help, kind words, jokes and patience at all times.

And lastly, thank you to my uncle Zhumash, who supported me in every way possible. His strong will and kindness will always be my biggest inspiration.

List of abbreviations

AAV	Adeno-associated virus
ATP	Adenosine triphosphate
bp	basepair
CPP	Circularly permuted pistol
E. coli	Escherichia coli
EGFP	Enhanced green fluorescent protein
FACS	Fluorescence-activated sorting
FBS	Fetal bovine serum
gRNA	Guide RNA
HDV	Hepatitis delta virus
HHR	Hammerhead ribozyme
iPSC	Induced pluripotent stem cells
IRES	Internal ribosomal entry site
ITC	Isothermal calorimetry
miRNA	microRNA
nt	nucleotide
NTPs	Nucleoside triphosphates
ORF	Open reading frame
PCR	Polymerase chain reaction
PRF	Programmed ribosomal frameshifting
PBS	Phosphate-buffered saline
RBS	Ribosome binding site
RNAi	RNA interference
SELEX	Systematic evolution of ligands by exponential enrichment
sgRNA	Single guide RNA
shRNA	Short hairpin RNA
sTSRV	satellite tobacco ringspot virus
UTR	Untranslated region
VEGF	Vascular endothelial growth factor
VSV	Vesicular stomatitis virus

Table of Contents

Table of Figures

1. Chapter 1. RNA switches for mammalian gene expression control

1.1. Introduction

Technologies to control gene expression in response to chemical signals such as small molecules, proteins, and RNAs in mammalian cells have significant implications in synthetic biology and biomedical applications. Efforts on developing tools to chemically control gene expressions started decades ago and have resulted in many protein- and nucleic acid-based solutions (Auslander & Fussenegger, 2016). Protein-based systems, such as those based on engineered transcription factors (e.g., TetR and CymR) and those based on the protein-RNA hybrid CRISPR-Cas9 platform, are characterized by dynamic ranges over 100-fold and have been used to develop sophisticated gene circuits in mammalian cells (Lienert, Lohmueller, Garg, & Silver, 2014).

While these engineered protein-based approaches are widely used for various applications (Lienert et al., 2014), there is also a rapidly developing field of RNA-based gene switches (riboswitches) that allow activation or inhibition of genes without engineered proteins. Although RNA-based switches typically show lower dynamic ranges compared to the protein counterparts, they offer several important advantages. Since most RNA-based switches consist of at most, several hundred bases, they require smaller genetic footprint compared to the protein-based switches. This is particularly important for applications in viral vectors such as adeno-associated virus (AAV) vectors that have limited capacity for exogenous genes. Additionally, the absence of extra protein-coding genes imposes lower metabolic burden on the host cells. The lack of exogenous proteins also eliminates potential immunogenicity risks. Furthermore, RNA aptamers that recognize various molecules such as proteins and small molecules can be obtained by *in vitro* selection (SELEX) from random RNA sequences (Tuerk & Gold, 1990) (Ellington & Szostak, 1990) (Berens, Groher, & Suess, 2015), therefore, synthetic riboswitches can potentially be engineered to respond to diverse chemical signals. While protein-based switches have been engineered to respond to noncognate chemical signals, such efforts are challenging and limited in the variety of molecules that can be recognized (Kallunki, Barisic, Jaattela, & Liu, 2019). These properties attracted research interest to designing and engineering a variety of RNA-based switches that function in mammalian cells (Yokobayashi, 2019) which are reviewed in this Chapter.

1.2. Types of RNA switches

RNA switches (riboswitches) are gene regulatory elements that function without the help of proteins. In prokaryotes and some eukaryotes, RNA switches occur naturally as compact tools to adjust gene expression to the changing environment. The existence of such a mechanism of gene regulation in mammals remains an unanswered question in RNA biology, as metabolite-binding riboswitches have not been discovered in mammals yet (Venkata Subbaiah, Hedaya, Wu, Jiang, & Yao, 2019). However, synthetic riboswitches have been successfully applied in prokaryotic and eukaryotic organisms, including mammals (Auslander & Fussenegger, 2017) (Sporing, Finke, & Hartig, 2020) (Wurmthaler, Sack, Gense, Hartig, & Gamerdinger, 2019) (Shanidze, Lenkeit, Hartig, & Funck, 2020) (Felletti & Hartig, 2017)

Typically, RNA switches consist of two main components: sensory domain and actuator domain, also referred to as "expression platform". Sensory domain detects input signal and communicates the signal to the expression platform, which affects gene expression. This is usually achieved by a conformational change in the expression platform upon ligand binding, which affects transcription or translation in various ways.

In synthetic riboswitches, the function of the sensory domain is often carried out by an RNA aptamer, a short oligoribonucleotide capable of specifically binding a certain ligand. Since RNA aptamers can theoretically be generated against any ligand by SELEX (Ellington & Szostak, 1990), it allows tuning riboswitch activity to almost any small molecule or protein. Alternatively, sensory domain can be a miRNA target site or a protein-binding sequence, thus, allowing the switch to respond to intracellular RNA sequences or RNA-binding proteins. Efficient transmission of the signal from the sensory domain to the actuator domain is usually achieved by a specific composition of the connecting sequence, the so-called communication module. It usually communicates input signal by secondary or tertiary interactions with the actuator domain. While the communication module is usually the bridging sequence connecting the two domains, it can sometimes be located in other positions.

1.2.1. miRNA-responsive switches

Presence of miRNA target sites in the untranslated region (UTR) of mRNA leads to downregulation of the corresponding gene by the RNA interference (RNAi) pathway. Since some miRNA expression is cell-type specific (Valencia-Sanchez, Liu, Hannon, & Parker, 2006), a miRNA-responsive device can be used to detect different cell types. Due to small size and simple mode of action, endogenous miRNA detection became an attractive strategy for biomedical applications. The Yoshida and the Saito groups developed a method for sorting heterogeneous population differentiated from induced pluripotent stem cells (iPSCs) into distinct cell types based on their miRNA profiles. For that, they introduced miRNA target site in the untranslated region of a fluorescent protein mRNA and thus linked miRNA abundance to fluorescence levels (Miki et al., 2015). Later, the Saito group presented a higher-resolution method for cells with similar miRNA profiles based on the same principle (Endo, Hayashi, & Saito, 2016). The Benenson and the Weiss groups also used miRNA markers to create multi-input gene circuits that allowed classifying cells into types and selectively killing cancer cells (Xie, Wroblewska, Prochazka, Weiss, & Benenson, 2011).

Detection of short RNAs including miRNAs can also be achieved with toehold switches in cell-free and bacterial contexts (Green, Silver, Collins, & Yin, 2014) (Pardee et al., 2014) (Pardee et al., 2016). Wang et al. developed a toehold switch functioning in mammalian cells by masking the Kozak sequence and the start codon with a miRNA binding site. Upon miRNA binding and resulting toehold strand exchange, the Kozak sequence and the start codon are released, thus activating gene expression (Wang, Emery, & Liu, 2019). This system allowed converting miRNA input into an activating, rather than inhibitory signal. Since programming these miRNA-responsive switches to different miRNAs is straightforward, using RNA-based switches is convenient for sensing and responding to endogenous miRNAs.

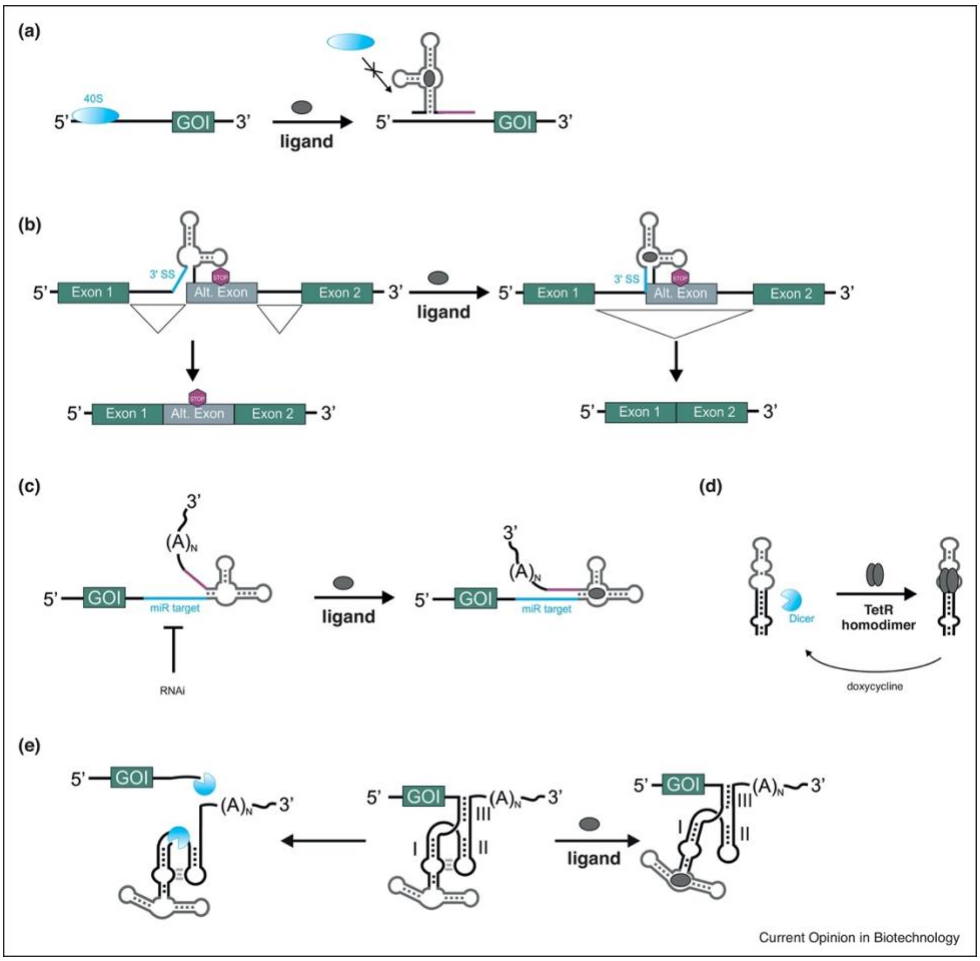


Figure 1-1. Types of aptamer-based switches from Sporing, Finke and Hartig (2020). (A) Control of translation initiation by an aptamer in 5' UTR of mRNA. (B) Alternative splicing regulation by aptamer-ligand binding. (C) Ligand-induced occlusion of miRNA target site. Ligand binding releases antisense strand and allows RNAi. (D) Control of RNAi by aptamer in the miRNA precursor. (E) Aptazyme on-switch. Ligand binding prevents aptazyme cleavage and allows mRNA processing. {Sporing, 2020 #259}

1.2.2. Aptamer-based RNA-switches

i. Aptamers in 5' UTR of mRNA

Protein binding in the 5' UTR of a mammalian mRNA can reduce efficiency of translation. This mode of gene expression control has been demonstrated by several groups by incorporating natural as well as synthetic protein-binding RNA sequences (Figure 1-1A) (Auslander & Fussenegger, 2017) (Kawasaki, Ono, Hirosawa, & Saito, 2020). In one of the earlier works from the Hentze lab, natural interactions of MS2 stem-loop /MS2 coat protein and U1A stem-loop/U1A protein were utilized to make off-switches in mammalian cells (Stripecke, Oliveira, McCarthy, & Hentze, 1994). Similarly, Inoue and co-workers minimized archaeal box C/D RNA motif that naturally binds to L7Ae protein and inserted it upstream of the reporter gene thus inhibiting its expression in response to L7Ae (Saito et al., 2010). Later, Saito's and Inoue's groups developed an inverter for their system to make it function as an on-switch (Endo, Hayashi, Inoue, & Saito, 2013). They introduced an additional RNA module consisting of a bait ORF, tandem premature termination codons,

beta-globin intron and IRES, that was controlled by the original off-switch. Upon addition of L7Ae, bait ORF was blocked, preventing mRNA degradation and allowing IRES-driven expression of the reporter (Endo et al., 2013). Using the same approach, Niles and co-workers developed a synthetic RNA-protein interaction system, using a TetR aptamer and TetR protein, commonly used as a genetic switch (Belmont & Niles, 2010) (Goldfless, Belmont, de Paz, Liu, & Niles, 2012). A similar strategy of upstream aptamer insertion for reducing translation efficiency has been employed for small molecule ligands as well but was less efficient potentially due to the smaller molecule size (Werstuck & Green, 1998) (Groher et al., 2018).

While riboswitches responsive to nucleic acids or proteins listed above allow detecting cellular state or rewiring intracellular pathways, they are not ideal for exogenous gene expression control. In case of exogenous control of a transgene, switches that respond to small molecules are preferable. The trigger molecule for exogenous gene control should be soluble in physiological conditions and be able to diffuse through cell membrane. Furthermore, it should have little or no biological activity to minimize effect on other processes in the cell. Hence, one of the major challenges in riboswitch engineering is to design devices responsive to small molecules that could eventually be applied *in vivo.* With a few exceptions, the rest of this chapter will review small molecule-responsive riboswitches in mammalian cells.

ii. Aptamer-controlled RNA interference

Aside from miRNA detection described earlier, several groups linked different stages of the RNAi pathway to ligand-aptamer binding. For example, An et al. controlled Dicer-mediated cleavage of shRNA by a theophylline aptamer, whereas Beisel et al. inhibited Drosha-mediated cleavage in ligand-dependent manner by inserting a theophylline aptamer in the miRNA basal segment (An, Trinh, & Yokobayashi, 2006) (Beisel, Chen, Culler, Hoff, & Smolke, 2011). Later Smolke's and Saito's groups did comprehensive studies on regulation of gene expression by endogenous protein-binding aptamers in pre-miRNA (Bloom, Winkler, & Smolke, 2014) and shRNA (Kashida, Inoue, & Saito, 2012).

In a different approach, Farzan's group took advantage of miRNA-mediated mRNA degradation to make tetracycline-responsive on-switches (Mou et al., 2018). They inserted a miRNA sequence and an aptamer in the UTR of mRNA. Upon addition of aptamer ligand, the miRNA target sequence was occluded and prevented miRNA-mediated decay, thus upregulating gene expression (Figure 1-1C). This method provides a way to exogenously control gene expression in cells with certain miRNA profiles.

iii. Other mRNA-targeting switches

In addition to simple aptamer insertion in the 5' UTR of the target mRNA, more sophisticated ways to control gene expression by aptamer-ligand binding have been explored, such as alternative splicing (Figure 1-1B) and programmed ribosomal frameshifting (PRF). For example, Smolke's group placed aptamers that bind to endogenous proteins in the intronic sequence of mRNA to control splicing by addition of the ligand (Culler, Hoff, & Smolke, 2010). In their system, protein-aptamer binding decreased accessibility of an exon located upstream of the reporter gene and containing a stop codon. This resulted in an alternatively spliced mRNA where the exon with a stop codon was excluded and, therefore,

the reporter gene was expressed. They demonstrated the system with four different protein-aptamer pairs (Culler et al., 2010). Similarly, the Suess group controlled splicing pattern of a mammalian mRNA with tetracycline (Vogel, Weigand, Kluge, Grez, & Suess, 2018). They designed a synthetic suicide exon, which was controlled by a tetracycline aptamer in its 3' splice site (Figure 1-1B).

Another way to control translation of mammalian mRNA is -1 programmed ribosomal frameshifting (-1PRF) with X XXY YYZ motif followed by a tertiary structure element such as pseudoknot or stem-loop. Controlling the structural element folding with an aptamer allows control of gene expression through the reading frame shift. This strategy has been employed by the Chang group with different structural elements and aptamers (Hsu, Lin, & Chang, 2014) (Lin & Chang, 2016). The Nakatani group also utilized -1PRF phenomenon to control gene expression (Matsumoto, Caliskan, Rodnina, Murata, & Nakatani, 2018). However, instead of controlling the formation of the structural element with an aptamer, they used a rationally designed small molecule naphthyridine carbamate tetramer (NCTn) to stabilize the pseudoknot structure.

1.2.3. Ribozyme-based switches

Self-cleaving ribozymes are a class of functional RNAs capable of site-specific phosphodiester scission. Correct ribozyme folding is critical for its catalytic function. This property has been used to link ribozyme cleavage to various stimuli. For example, Nakatani's group used their "molecular glue" compound for ligand-induced stabilization of RNA structure to make a switch with a self-cleaving ribozyme (Dohno, Kimura, & Nakatani, 2018). The group used naphthyridine carbamate tetramer with a Z-stilbene linker (Z-NCTS) that binds XGG/XGG mismatches in DNA and RNA to stabilize hammerhead ribozyme's active conformation. They introduced the target sequences in the loops I and II of the hammerhead ribozyme so that Z-NCTS binding interferes with the loop-loop interaction essential for the ribozyme cleavage. When the ribozyme was inserted in the UTR of mRNA, its self-cleavage led to the loss of the polyA tail necessary for further mRNA processing, thereby reducing expression of the encoded protein. Thus, engineered hammerhead ribozyme incorporated in the 3'UTR of mRNA was able to reduce luciferase levels 4-fold by addition of the "molecular glue" (Dohno et al., 2018).

Another way to control ribozyme activity is to fuse it with an aptamer and thus making it responsive to the aptamer ligand. Such allosterically controlled ribozymes, or aptazymes, can be used to control stability of a mammalian mRNA (Figure 1-1E). Communication between the aptamer and the ribozyme is achieved by the communication module, a short sequence that communicates aptamer-ligand binding to ribozyme activity. Depending on the design, aptazymes can be engineered to be either on- or off-switches. For example, Nomura et al. developed an off-switch that inhibited gene expression in HEK293 cells by 29.5-fold in response to guanine, while Beilstein et al. achieved 8.7-fold upregulation of a transgene by an on-switch responsive to tetracycline (Nomura, Zhou, Miu, & Yokobayashi, 2013) (Beilstein, Wittmann, Grez, & Suess, 2015).

1.3. History of aptazymes

First allosterically controlled ribozymes were developed *in vitro* as early as 1997 by Breaker's group by fusing hammerhead ribozyme with ATP aptamer at stem II (J. Tang &

Breaker, 1997). It took several proof-of-concept studies before allosterically controlled ribozymes reached applications in mammalian cells. First, in 1998, Werstuck and Green used an aptamer to control gene expression in mammalian cells (Werstuck & Green, 1998). They inserted aptamers specific for the Hoechst dye (H33258) into the 5′-UTR of a reporter gene and thus inhibited gene expression by addition of ligand. Then, in 2004, Yen et al. demonstrated mammalian gene expression control by incorporation of a hammerhead ribozyme from a trematode *Schistosoma mansoni* in the UTR of mRNA (Yen et al., 2004). They optimized the ribozyme by a series of mutations and derived an efficient variant N79, which was later used to develop aptazymes by other groups. In addition to cell culture, the group demonstrated the same effect in mice and also achieved ribozyme inhibition by addition of an antibiotic toyocamycin. Hence, by 2004, both aptamers and ribozymes had been independently demonstrated to be able to control mammalian gene expression. Meanwhile, in 2007, Win and Smolke demonstrated gene expression control by aptazyme in yeast, by incorporating a theophylline-responsive hammerhead ribozyme in the UTR of mRNA (Win & Smolke, 2007). This was followed by Wieland and Hartig who were developing an aptazyme in bacteria (Wieland & Hartig, 2008), which was subsequently used by Kumar et al. in 2009 to control RNAi in mammalian cells (Kumar, An, & Yokobayashi, 2009). These three discoveries prepared the ground for the insertion of an aptazyme in the UTR of a mammalian mRNA by the Smolke and Hartig groups (Auslander, Ketzer, & Hartig, 2010; Y. Y. Chen, Jensen, & Smolke, 2010). Following these pioneering works, many strategies have been employed for aptazyme engineering.

1.4 Aptazyme design strategies

Most of the existing aptazymes for applications in mammalian cells are based on the hammerhead ribozyme, the first self-cleaving ribozyme class to be discovered and characterized. The first aptazyme demonstrated to work in mammalian cells was made by inserting a theophylline aptamer in the stem III of the N79 hammerhead ribozyme variant from *Schistosoma mansoni* (Auslander et al., 2010). Unlike stems I and II, stem III is not involved in tertiary interactions and tolerates modifications, such as insertion of an aptamer. Upon addition of theophylline, aptazyme's conformation is stabilized and results in ribozyme cleavage. This aptazyme design was originally developed in *E. coli* (Wieland & Hartig, 2008), where it functioned as an on-switch, since ligand-induced ribozyme cleavage results in the release of the sequestered ribosome binding site (RBS) and activation of translation. In contrast, in mammalian context, aptazyme cleavage within the UTR of mRNA results in degradation of the corresponding mRNA. Therefore, the aptazyme functions as an off-switch. The Yokobayashi group applied the same aptazyme in mammalian cells, but for control of RNAi. In their system, instead of inserting the aptazyme in the UTR of an mRNA, they fused the aptazyme with the structural analogue of pri-mRNA. Aptazyme cleavage resulted in the release of pri-mRNA and its processing by Drosha, Dicer and induction of RNAi. They optimized the communication module and developed two aptazymes to control RNAi, with theophylline and guanine aptamers (Kumar et al., 2009) (Nomura, Kumar, & Yokobayashi, 2012). Later Farzan's group used the same design to develop off-switches against three different compounds, tetracycline, theophylline and guanine, using systematic optimization of communication module stability based on their scoring function (Zhong, Wang, Bailey, Gao, & Farzan, 2016).

Another strategy to make off-switches with hammerhead ribozyme was introduced by the Smolke group in 2013 (Wei, Chen, & Smolke, 2013). They discovered that the aptazymes they had previously engineered in yeast worked in mammalian cells. In that design, they

fused theophylline aptamer with sTSRV hammerhead ribozyme at the stem II and controlled tertiary loop interactions by aptamer binding. By displacing the loop at the end of stem II they succeeded in making both on- and off- switches. The same group later used similar design to make MS2 protein-responsive off-switches (Kennedy, Vowles, d'Espaux, & Smolke, 2014). Another protein-responsive switch was developed by Auslander et al. They replaced stem II of an environmental HHR with N-peptide-binding aptamer and then randomized stem I loop to re-establish tertiary interactions between the two loops (Auslander et al., 2014).

While stem II was used to make both on- and off- switches, stem I of the hammerhead ribozyme was used for on-switches only. It was fused with MS2 aptamer (Kennedy et al., 2014), tetracycline aptamer (Beilstein et al., 2015) and guanine aptamer (Stifel, Sporing, & Hartig, 2019) to yield on-switches, of which Beilstein's work had the highest on-off ratio.

Until recently, the only other ribozyme used as an aptazyme scaffold for mammalian cells, was the hepatitis delta virus (HDV) ribozyme from the viral genome. Nomura et al. replaced stem-loop P4-L4 with the guanine aptamer and tested a few variants of the bridging sequence, eventually identifying one that inhibited gene expression by 29-fold (Nomura et al., 2013). This study showed that ribozymes other than hammerhead can be scaffolds for aptazyme design.

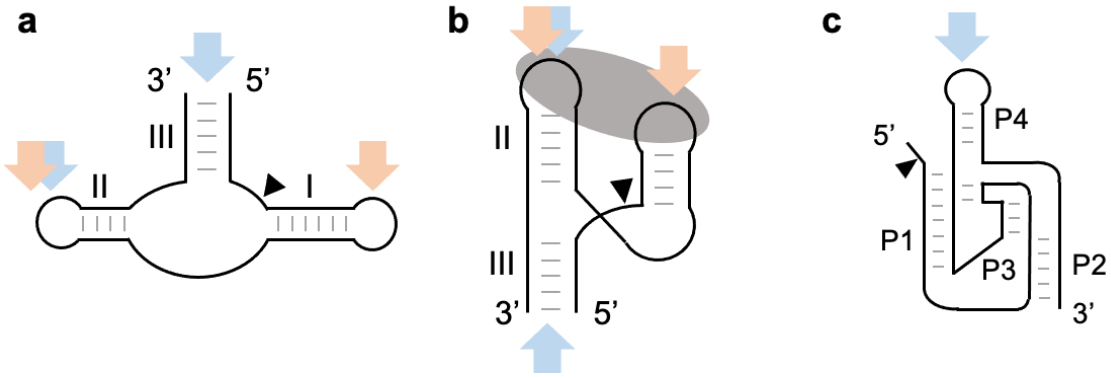


Figure 1-2. Schematic structures of hammerhead ((a) secondary structure, (b) tertiary structure) and HDV (c) ribozymes and reported aptamer insertion sites. Gray oval indicates loop-loop interaction. Black triangle indicates cleavage site. Arrows show sites of aptamer insertion for on- (orange) and off- (blue) switches.

Hence, hammerhead ribozyme has been the most popular choice for mammalian aptazyme engineering. All of its stem-loops have been used for aptamer fusion, while different strategies were employed to make on-switches versus off-switches. Stem III not involved in tertiary interactions was used for making off-switches. Later, targeting loop-loop interactions of stems I and II allowed developing on-switches with stem I aptamer fusion, and both on- and off-switches with stem II (Figure 1-2A and 1-B). In contrast, in HDV ribozyme only one stem-loop (P4-L4) has been used for making an off-switch (Figure 1-2C).

In addition to choosing aptazyme architecture and optimizing communication module sequence, there are other ways of improving performance of aptazyme switches. For example, the Suess group reported improved aptazyme performance upon simultaneous insertion of an aptazyme in the 5' and 3' UTR of mRNA (Beilstein et al., 2015). This finding was later corroborated by the Smolke and the Farzan group (Zhong et al., 2016) (Bell et al., 2015). However, 5' UTR aptazyme insertion can have undesired inhibitory effect by introducing alternative start codons from the aptazyme sequence and interfering with

ribosome scanning (Auslander et al., 2010) (Wittmann & Suess, 2011), so often 3' UTR is preferred. Since aptazyme activity is highly dependent on its 3D structure, some groups introduced spacer sequences flanking the aptazyme to minimize context dependence (Auslander et al., 2014; Beilstein et al., 2015; Bell et al., 2015).

1.5. High-throughput approaches for aptazyme screening

Since optimization of the communication module sequence involves screening thousands of variants even with a few randomized positions, it is laborious and time-consuming to test them all individually. Due to the ability to only express a single plasmid copy with a given origin of replication and high replication rate, bacteria are a good medium for screening of mutant libraries. Soon after discovering the ability to control gene expression with aptazymes in bacteria, several groups came up with strategies to scale up the throughput of aptazyme variants tested. Systems for screening aptamer switches in bacteria based on motility (Topp & Gallivan, 2008) and viability (Muranaka, Sharma, Nomura, & Yokobayashi, 2009; Nomura & Yokobayashi, 2007) had been developed previously, and it was soon applied for screening aptazyme switches as well. For example, Wieland and Hartig linked aptazyme activity to reporter EGFP gene instead of selection markers and screened a library of hammerhead-theophylline aptazymes with 3 randomized positions (Wieland & Hartig, 2008). They picked single colonies from a plate of *E. coli* transformed with the library-embedded plasmid and inoculated them into 96-well plates. Transformed bacteria, each carrying an individual plasmid copy were then grown in 96-well plates and measured for fluorescence levels in the same plates. Several groups further increased the throughput of fluorescence reporter-based screening in bacteria for aptamer switches by employing fluorescence-activated cell sorting (FACS) (Fowler, Brown, & Li, 2008; Lynch & Gallivan, 2009)).

Since budding yeast possesses the same ability to uptake external DNA and form single colonies as bacteria, similar screening approaches can be used to screen for aptazymes in yeast. The Smolke lab developed several FACS-based strategies to screen aptazymes in yeast (Liang, Chang, Kennedy, & Smolke, 2012; Townshend, Kennedy, Xiang, & Smolke, 2015). By sorting yeast cells into activity bins based on reporter fluorescence, the group identified aptazymes responsive to theophylline and tetracycline.

While bacteria and yeast are established systems for running high-throughput screenings due to ability to uptake external DNA and form single colonies, mammalian cells have not been used for screening as much. Attempts to increase the throughput of aptazyme variants testing in mammalian cells are limited to just a few examples. The first one was demonstrated as early as 2009, when Ellington's group ran a selection for highly active hammerhead ribozymes in HeLa cells (X. Chen, Denison, Levy, & Ellington, 2009). Although their construct lacked the aptamer domain, selection strategy is also applicable for aptazymes. They created two libraries by randomizing 8 positions in stem II and 6 positions in loop II of the hammerhead ribozyme, attached mammalian U6 promoter upstream and transfected these short DNA fragments directly into cells. They let the transcription and cleavage proceed in cells and then extracted total RNA, containing ribozyme fragments transcribed from the DNA library. Then they reverse transcribed and amplified the cleaved fragments and used them as a pool for the next round of selection. Upon completing five rounds of selection they achieved some enrichment of highly active hammerhead mutants (X. Chen et al., 2009).

Similar approach of transfecting a library of DNA fragments directly into mammalian cells for ribozyme activity screening was followed by Nomura et al. (Nomura, Chien, &

Yokobayashi, 2017). They generated a library of natural and artificial pistol ribozymes as double-stranded DNA fragments under U6 promoter and transfected them into HEK293 cells. Similarly to Chen at al, they extracted total RNA containing the transcribed and cleaved library; however, instead of running multiple selection rounds they sequences the extracted RNA pool in a single Illumina sequencing run. This method allowed identifying pistol ribozyme variants active in mammalian cells for the first time from ~3000 variants.

Although both of these studies were able to select ribozyme variants active in mammalian cells, it may be difficult to apply the methods to aptazymes. Aptazyme variants may differ in cleavage speed only slightly among each other and are quickly degraded upon cleavage, thus requiring sensitive measurement of cleavage rates. Furthermore, both double-stranded DNA and short RNA fragments have low stability in cytoplasm and are subject to intracellular nucleases' activity. Gel extraction step for total RNA imposes another bias on the screening. Therefore, although these methods provide a promising perspective on screening in mammalian cells, they should be optimized before being applied to aptazymes.

Since screening in mammalian cells directly is more challenging and lower throughput, there have been attempts to run screening in simpler systems. In the early days of aptazyme engineering for mammalian cells, Breaker's group ran rounds of *in vitro* selection to isolate theophylline-inducible hammerhead ribozymes for subsequent application in mammalian cells. However, in spite of running the screening in near-physiological conditions with very low Mg^{2+} concentration (0.55mM), the group discovered that *in vitro* selected aptazyme variants did not have the same activity in cells. They hypothesized that co-transcriptional RNA folding and the speed switching between the active and inactive aptazyme states may be the key factors differing between the two systems. In contrast, HDV-based aptazymes selected *in vitro* by Kobori et al. (2015) maintained their guanine-induced response when tested in mammalian cells, although *in vitro* Mg^{2+} concentration was higher than that of Breaker's group. Later Kobori et al. (2018) also screened a library of HDV ribozymes *in vitro* with subsequent successful transfer to mammalian cells. Hence, it is still not clear which physiological parameters have to be mimicked to ensure that aptazymes screened *in vitro* are functional in mammalian cells as well. Furthermore, since both of Kobori's successful examples of transfer from *in vitro* to *in vivo* were with HDV ribozyme, it is possible that the intracellular parameters affecting aptazyme performance vary among different ribozyme classes.

Not only transfer from *in vitro* to mammalian systems, but also from bacteria or yeast to mammalian cells seems challenging. For example, Hartig's group has engineered hammerhead-based aptazymes for mammalian cells from previously described bacterial switches; however, original bacterial sequences did not produce high on-off ratios in mammalian cells and had to be optimized for mammalian cells. For some other groups, like Suess (Wittman and Suess, 2007), rounds of selection *in vitro* were able to produce aptazymes functioning in yeast, but not in mammalian cells. They speculate that it is due to the differences in transcription-termination and cycles of RNA-binding proteins displacement during mRNA processing. Due to the growing confusion about transferability of ribozymes' activity across different systems, Wurmthaler and Hartig ran a comprehensive study of the dependence of ribozyme activity on organisms and motifs (Wurmthaler and Hartig 2018). After assessing activity of several hammerhead ribozyme variants from different organisms they discovered that indeed the activity varied greatly even between two eukaryotic systems, such as yeast and mammalian cells. However, they found that one of the hammerhead motifs maintained high activity across prokaryotic and eukaryotic systems, suggesting that the

factors affecting ribozyme activity are different for different ribozyme motifs. As one of the key factors they mention the differences in mRNA half-life between bacteria, yeast and mammalian cells.

Overall, there are several cases when *in vitro* screening in near-physiological conditions succeeded in selecting aptazymes functioning in mammalian cells. However, effect of such conditions may depend on a particular ribozyme motif, like type 3 hammerhead ribozyme from Hartig's group (2018) or HDV ribozyme from our group (2015, 2018). However, for many other aptazymes, transfer of aptazyme activity even across two eukaryotic systems (yeast and mammalian cells) remains challenging. Since the exact localization and conditions of each of the transcription-translation steps are not studied well, it is difficult to reliably predict behaviour of aptazymes in different contexts.

1.6. Update on the recently published studies

Close to finishing this thesis work, the Smolke and the Hartig groups published two papers with very similar strategies for aptazyme screening (Strobel, Sporing, et al., 2020; Xiang et al., 2019). They generated plasmid libraries containing a reporter fluorescence gene followed by an aptazyme variant in 3' UTR. Plasmid libraries were then transfected into HEK293 cells. Following transcription, active aptazyme variants would cleave and be degraded by the intracellular machinery, whereas inactive variants would stay in the cytoplasm longer. mRNAs containing inactive variants can then be extracted, reverse transcribed and sequenced. Abundance of a particular aptazyme variant in the extracted RNA pool is then normalized by its abundance in the starting DNA pool, and this value can be used to compare relative activities between the variants. The Hartig group first validated the strategy by rediscovering existing aptazymes, such as Guanine off-switch (GuaM8HDV) from Nomura et al. (Nomura et al., 2013) and Tet on-switch (K19) from Beilstein et al. (Beilstein et al., 2015). They then also identified Guanine on-switches based on hammerhead ribozyme and both on- and off- switches based on tetracycline aptamer and twister ribozyme(Strobel, Sporing, et al., 2020) . Xiang et al. used the method to screen for hammerhead-based aptazymes activated by four different ligands (theophylline, hypoxanthine, (6R,S)-folinic acid and cyclic di-GMP) (Xiang et al., 2019). They also compared read counts from RNA-seq to protein expression levels of the same aptazyme variants assessed by FACS of lentiviral vector-transduced HEK293T cells. These two papers resolved the issues with the previous screening methods in mammalian cells (X. Chen et al., 2009; Nomura et al., 2017). They addressed the problem with low nucleic acid stability in the cytoplasm by transfecting a library of plasmids, rather than linear double-stranded DNA fragments. Moreover, they accounted for the degradation of cleaved species in the cytoplasm and extracted DNA plasmid library for normalization. Hence, this RNA sequencing method is a promising tool to screen large pools of aptazyme variants.

1.7. Applications of aptazymes

Being short RNA sequences, aptazymes have the advantage over protein switches in terms of their small genetic size, flexible design, and low immunogenicity. This, along with recent advances in translation of aptazyme research to mammalian cell context (Sporing et al., 2020), made them an attractive tool for biomedical applications.

As one of the first applications of aptazymes, Smolke and coworkers optimized previously described theophylline responsive on-switch (Win & Smolke, 2007) for mammalian cells and

used it to control proliferation of T cells (Y. Y. Chen et al., 2010). By introducing a copy of IL-2 gene with multiple copies of the on-switch integrated in the 3'UTR, the group was able to control T cells' proliferation by addition of theophylline. The same group later employed aptazyme-directed gene expression to control mammalian cell cycle, which is required in some fields of biomedical research and biotechnological industry. Wei and Smolke screened proteins involved in cell cycle to find key regulators that, once inhibited, will exert strongest effect on normal cell cycle progression (Wei & Smolke, 2015). After identifying p27 as the target for G0/1-phase arrest and cyclin B1 for G2/M-phase arrest, they introduced exogenous copies of those genes with the 3'UTR insert of the theophylline-hammerhead off-switch from a previous study (Townshend et al., 2015). However, although these studies demonstrated the potential of aptazymes in controlling fundamental biological processes, they both suffered from low dynamic range of RNA devices.

One of the rapidly developing directions of aptazyme research is delivery of transgenes with viral vectors and control of viral replication. Existing studies using viral vectors are summarized in Figure 1-3.

For example, Ketzer et al. inserted theophylline-responsive hammerhead ribozyme, P1-F5, previously reported by the same group, around a transgene in viral vectors (Auslander et al., 2010; Ketzer, Haas, Engelhardt, Hartig, & Nettelbeck, 2012). Their study demonstrated that aptazymes can be delivered via viral vectors, such as AAVs, replication-deficient adenoviruses and oncolytic adenovirus, to control transgene expression in cell culture. Importantly, in addition to common reporter fluorescent genes they also tested CCL5, a therapeutically relevant chemokine. Later, the same group further explored aptazymes' potential for viral gene therapy (Ketzer et al., 2014). They incorporated the same aptazyme, P1-F5, in the UTR of genes important for viral replication of both RNA (measles) and DNA (adenovirus) viruses. Their study provided proof of concept that an aptazyme can be used for controlling viral replication. In 2015 Strobel et al. applied GuaM8HDV aptazyme previously developed by the Yokobayashi group to control transgene expression in AAV vectors (Strobel et al., 2015). Transgene expression in the producer cells is an unnecessary by-product of viral particles generation, as only essential viral proteins are required at that stage. Strobel et al. showed that addition to guanine was able to inhibit expression of transgenes in producer HEK 293 cells, which helped to increase AAV yield and decrease toxicity of some transgenes (Strobel et al., 2015). Recently, our group also demonstrated control of a viral transgene with the same GuaM8HDV off-switch (Takahashi & Yokobayashi, 2019). A transgene was delivered by the vesicular stomatitis virus (VSV) vector and was reversibly controlled by addition of guanine over the course of 12 days. Farzan's group also demonstrated control of a transgene in a viral vector (AAV) by their rationally designed aptazyme, but in mice (Zhong et al., 2016). Their study marked the first application of aptazymes *in vivo*.

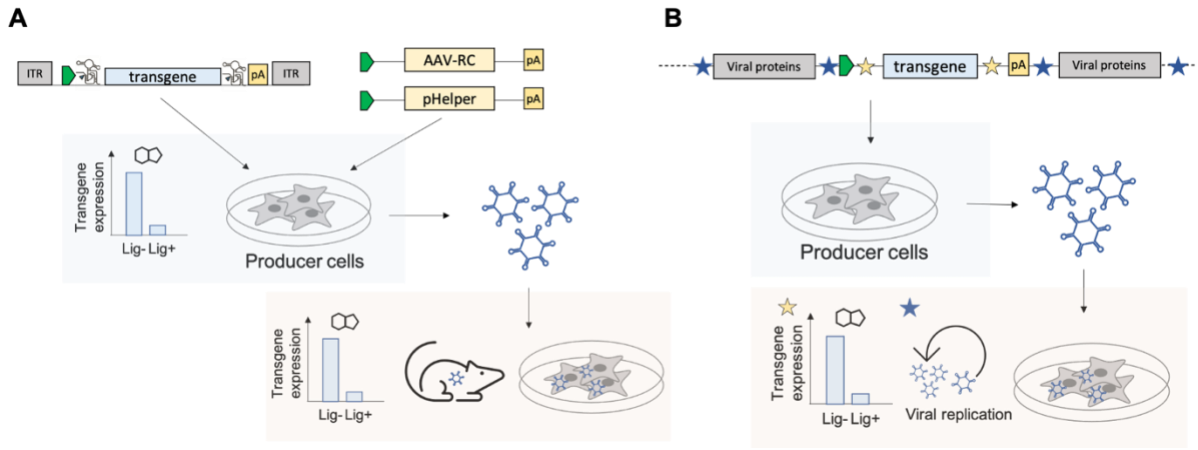


Figure 1-3. Applications of aptazymes using viral vectors. A) Control of transgene expression with an aptazyme in an AAV vector. Aptazymes can be inserted upstream and/or downstream of the transgene to control its expression either during viral particle production in the producer cell line or in the infected cell culture or mice (Ketzer et al., 2012) (Strobel et al., 2015) (Strobel, Duchs, et al., 2020) (Zhong et al., 2016). AAV vectors require helper proteins for generation of viral particles, so initial step requires co-transfection with other plasmids. (B) Control of a transgene or a viral gene expression with an aptazyme in other types of viruses, such as adenovirus, oncolytic adenovirus or vesicular stomatitis virus (Ketzer et al., 2012) (Ketzer et al., 2014) (Takahashi & Yokobayashi, 2019). Aptazyme can be inserted in any of the positions marked with stars. Yellow stars indicate positions for control of a transgene, and blue stars indicate positions for control of viral replication. In this schematic location of transgene relative to the viral proteins is arbitrary and can vary depending on the virus.

In their next *in vivo* application, aptazymes were used for therapeutic purposes. Reid et al. used aptazymes to control expression of Eylea, an anti-VEGF agent, delivered by AAV vectors in mice (Reid, Nettesheim, Connor, & Lipinski, 2018). VEGF is involved in the development of wet age-related macular degeneration and in standard therapy is inhibited by intra-vitreal injections of Aflibercept (Eylea), a recombinant fusion protein. Delivery of Eylea protein-coding sequence in an AAV vector controlled by an aptazyme allowed its long-term small molecule-inducible expression. Thus, instead of regular intra-vitreal injections of Eylea, mice were treated by a single injection of the AAV vector followed by oral administration of tetracycline (Reid et al., 2018). This study demonstrated potential of aptazymes as tools for tunable gene expression control for real-world clinical applications.

Aside from controlling mRNA stability, aptazymes have been recently used to control function of other RNAs in mammalian cells. Growing use of CRISPR/Cas9 technique fostered development of tools for its precise control. Since CRISPR/Cas9 is a hybrid system relying both on RNA and protein, aptazymes can be used to control its function through the RNA component. Several groups have employed aptazymes to control gRNA stability and processing. For example, Tang et al. incorporated theophylline-responsive aptazyme from Hartig's group (Wieland & Hartig, 2008) into gRNA (W. Tang, Hu, & Liu, 2017). The aptazyme was bridging the gRNA sequence and the antisense strand occluding it, and thus, aptazyme cleavage released the antisense strand and allowed gRNA activation. In contrast, Chen et al. coupled aptazyme cleavage to degradation, not activation of sgRNA (H. Chen et al., 2018). They incorporated the same theophylline-responsive aptazyme into the sgRNA backbone, thus linking stability of the sgRNA to aptazyme self-cleavage. By inducing degradation of sgRNA with theophylline they reduced off-target effect of CRISPR/Cas9 genome editing.

It is important to note that aptazymes often exhibit different activities in different contexts. This could be due to differences in cell types, protein complexes during RNA processing,

reporter protein or sequence context. For example, P1-F5 theo-HHR off-switch used in Ketzer's studies to control expression of viral transgene and viral replication, did not work when placed in AAV vector by Strobel et al (2015) (Ketzer et al., 2012; Ketzer et al., 2014; Strobel et al., 2015). Instead, Strobel et al. (2015) used GuaM8HDV off-switch (Nomura et al., 2013), which also had only 4.4-fold inhibition of the AAV transgene in response to guanine, in contrast to the originally reported 29-fold downregulation. An exception to this trend is the study by Strobel et al. (2020) with K19 ON-switch in mice, where anti-FITC scFv antibody delivered with AAV vector achieved 15-fold activation upon addition of tetracycline, which is higher than the value originally reported in cells (Beilstein et al., 2015; Strobel, Duchs, et al., 2020). However, this difference could be due to the secretion of their reporter protein anti-FITC scFv from the liver cells, as will be discussed in the later paragraphs.

In general, aptazyme on-off ratios tend to be lower *in vivo* compared to the values assessed in cell culture. For example, Zhong et al. achieved 7-fold repression of the transgene delivered to the mouse muscle by an AAV vector, while the same construct with double aptazyme insertion resulted in 33-fold reduction in gene expression in cell culture(Zhong et al., 2016). In their age-related macular degeneration study, Reid et al. (2018) assessed five previously reported aptazymes both in HEK293T cell culture and in mouse retina. While their on-off ratios in cells were lower than those originally reported by other groups (e.g. 1.5 vs 8.5 for K19 (Beilstein et al., 2015) and 2.5 vs 29.5 for GuaM8HDV (Nomura et al., 2013)), these values were in the same range in both contexts. This observation suggests that, in this case, sequence context or vector type might have had more effect on the aptazyme activity than the cell type.

Reporter protein could also affect measured on-off ratios of the aptazyme. For example, in their comprehensive study of aptazyme activity in different organs in mice, Strobel et al. (2019) found that on-off ratios of the K19 switch from the Suess group vary only slightly across lungs, liver, heart and muscles (Beilstein et al., 2015) (Strobel, Duchs, et al., 2020). However, on-off ratios measured in liver were different for the reporter proteins anti-FITC and NLuc. The authors explain the difference in the measured value by secretion of anti-FITC scFv versus accumulation of NLuc. Thus, difference in the processing of the reporter protein, such as secretion or degradation rate may also contribute to the dynamic range of a riboswitch.

Hence, it is currently difficult to predict aptazyme behaviour for a new application, as there are many variables in every new context. Aptazymes, therefore, may require further optimization for the desired vector, reporter protein and target tissue type. Development of a method or system that reduces context dependence would alleviate the need to optimize aptazymes for each context.

1.8. Conclusion and outlook

In around two decades of their development, aptazymes have been shown to function *in vitro*, in bacteria, yeast, human cell culture and *in vivo* in mice. Applications of aptazymes for various cellular processes and replication of viral vectors demonstrate their potential as compact and reliable tools for synthetic biology. Due to their small size, *in cis* mode of action and relative simplicity of engineering, various types of aptazymes functioning in mammalian cells have been developed over the years. Different aptazyme architectures have been demonstrated with hammerhead and HDV ribozymes to yield both on- and off-switches.

Furthermore, various screening techniques have been developed to optimize aptazyme performance in different environments. Those optimization studies showed that intracellular environment and transcription-translation processes affect aptazyme activity, since not all aptazymes can be transferred across different systems, such as prokaryotic and eukaryotic cells. However, key factors important for aptazyme activity in different environments still remain to be unraveled.

Since aptazymes consist of distinct parts, aptamer and ribozyme, they, in principle, can be engineered to respond to different molecules. This has been demonstrated for hammerhead ribozyme with various ligands, such as MS2 protein, tetracycline, theophylline, guanine, etc. While this ability of aptamers holds promise for their application, currently it is limited by the availability of ligand-aptamer pairs. One of the future challenges is developing aptamers against soluble, non-toxic and biorthogonal molecules that function *in vivo*. Furthermore, currently existing aptazyme scaffolds are restricted to only two ribozyme classes. Expanding the number of available ribozymes and aptamers may help find switches with higher dynamic range and promote their application. Combined with new design strategies and screening methods, it will allow streamlining development of aptazymes for biomedical applications.

2. Chapter 2. New ribozyme scaffolds for aptazyme design in mammalian cells

2.1 Introduction

2.1.1 Self-cleaving ribozymes

Since the discovery of the catalytic function of RNA in *Tetrahymena thermophila* by Thomas Cech in 1982 (Kruger et al., 1982), ribozymes have been found in all kingdoms of life and classified into distinct functional types. In addition to self-splicing introns, self-cleaving ribozymes and *trans*-acting RNase P found in nature, there are artificially engineered ribozymes with novel functions, such as class I ligase developed by artificial evolution (Bartel & Szostak, 1993), aminoacylating ribozyme (Li & Huang, 2005), and others (Park et al., 2019). Among the known ribozyme classes, self-cleaving ribozymes are the most common in nature and in synthetic biology applications. Out of nine currently known structural classes of self-cleaving ribozymes, the first four were discovered in the process of studying biological systems and organisms that they were part of (C. E. Weinberg, Weinberg, & Hammann, 2019). The other five classes of self-cleaving ribozymes were discovered by bioinformatic search (Barrick et al., 2004), (Roth et al., 2014) (Z. Weinberg et al., 2015). While some self-cleaving ribozymes' folds contains only helical junctions (e.g. hammerhead, hairpin, twister sister, Varkud satellite), some also require a pseudoknot formation for cleavage (HDV, *glmS*, pistol, twister) (Figure 2-1). These structural properties are important for their applications in synthetic biology. Ribozymes with helical junctions are relatively easy to engineer, as some helices tolerate mutations or extensions and can therefore be used for aptamer fusion. In contrast, more complex pseudoknots are more difficult to modify. Partly due to this reason, compact hammerhead ribozymes have been chosen almost exclusively for engineering aptazymes. As summarized in Chapter 1, with the only exception of HDV-based off-switch (Nomura et al., 2013), all existing aptazymes functioning in mammalian cells have been developed based on the hammerhead ribozyme class. Hammerhead was also one of the earliest ribozymes to be discovered and therefore has been studied extensively over the years. However, with new ribozyme classes being discovered and characterized in the last few years, there are more options for scaffolds for aptazyme engineering. Furthermore, aside from the traditional fusion at stems, new architectures can be explored to allow for modular design for new aptazymes.

Increasing the diversity of self-cleaving ribozyme scaffolds will allow addressing some issues with aptazyme development. Modular nature of aptazymes should allow replacement of the aptamer domain; however, in reality, smooth replacement is not always achieved (Stifel et al., 2019). Some aptamers may have unexpected interactions with parts of the ribozyme, which negatively affects their performance. In this regard, even surrounding sequence may affect ribozyme activity (Bell et al., 2015) (Auslander et al., 2014), and thus some ribozymes may not function well in certain sequence contexts. Furthermore, as highlighted in the study of Wurmthaler et al., ribozyme activity often varies across different biological systems (Wurmthaler, Klauser, & Hartig, 2018), so having more available ribozyme scaffolds will help to find the optimal solution for a given context.

2.1.2. Twister ribozymes

In 2014, Breaker and coworkers discovered a new, sixth, class of self-cleaving ribozymes by

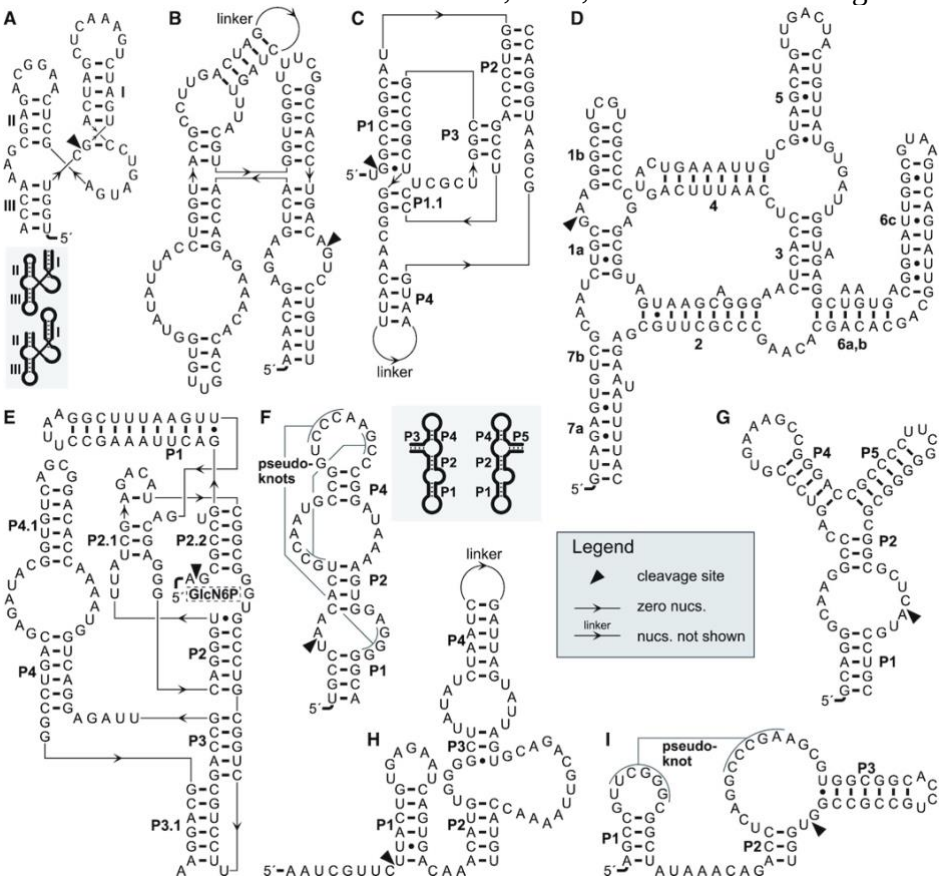


Figure 2-1.. Currently known natural self-cleaving ribozyme classes. Figure from Weinberg et al (C. E. Weinberg et al., 2019). (A) Hammerhead type III with inset showing circular permutations. (B) Hairpin ribozyme. (C) HDV ribozyme. (D) The Varkud satellite ribozyme. (E) The glmS ribozyme. (F) A type-P1 twister ribozyme with the inset showing type-P3 (left) and type-P5 (right) circularly permuted forms. (G) Twister sister ribozyme. (H) Hatchet ribozyme. (I) Pistol ribozyme.

bioinformatics analysis (Roth et al., 2014). The short RNA motif of these new twister ribozymes is present across many species of prokaryotes and eukaryotes, including insects, fish and plants. The group named this doubly pseudoknotted RNA motif "twister", because of its resemblance to the ancient Egyptian hieroglyph "twisted flax". They assessed activity of some twister ribozyme representatives *in vitro*, and although *in vitro* k_{obs} was measured to be 10 min^{-1} in standard conditions (30 mM HEPES, pH 7.5, 100 mM KCl, 20 mM $MgCl_2$ at 23°

C), it was speculated to reach 1000 min^{-1} under physiological conditions. This makes twister the fastest self-cleaving ribozyme characterized to date. Soon after its discovery, three groups published crystal structures of twister ribozymes and provided insights on its cleavage mechanism (Liu, Wilson, McPhee, & Lilley, 2014) (Eiler, Wang, & Steitz, 2014) (Ren et al., 2014). As predicted by bioinformatics earlier, the ribozyme forms two pseudoknots that help stabilizing the active site and positioning the nucleotides involved in catalysis. Two nucleotide bases, G62 and A63, were shown to be critical for cleavage. G62 serves as a general base, whereas A63 stabilizes transition state (Eiler et al., 2014).

Due to twisters' fast cleavage and presence in all domains of life, several groups started engineering the ribozyme for synthetic biology applications. For example, Hartig's group has designed aptazymes based on twister ribozyme env-9 (environmental sequence 9) that were functioning as riboswitches in bacteria and yeast (Felletti, Stifel, Wurmthaler, Geiger, & Hartig, 2016). They designed two riboswitches responsive to different ligands (theophylline and thiamine pyrophosphate) employing different architectures by adding aptamers to P1 and P5 stems of the ribozyme. The Jaffrey group made a circularized version of Twister from *Nematostella vectensis* to stabilize exogenous RNAs in a cell (Litke & Jaffrey, 2019). Aside from demonstrating another application for ribozymes, this paper was also an indication that twister ribozymes can function in mammalian cells and are, therefore, a promising scaffold for riboswitch engineering. Indeed, in 2020, close to finishing this thesis work and after publishing our twister-based switches (Mustafina et al., 2020), Strobel et al. published twister-based off-switches responsive to tetracycline developed by screening in mammalian cells (Strobel, Sporing, et al., 2020). Hence, as already highlighted by Gebetsberger and Micura, fast cleavage rate, presence in many domains of life and availability of strategic sites for aptamer insertion make twister ribozymes good candidates for developing aptazymes (Gebetsberger & Micura, 2017).

2.1.3. Pistol ribozymes

Pistol ribozyme class has also been discovered by Breaker's group by bioinformatic analysis soon after twister, along with two other self-cleaving ribozymes – twister sister and hatchet (Z. Weinberg et al., 2015). The group had noticed that hammerhead and twister ribozymes often appear next to each other in proximity to certain genes. They hypothesized that this phenomenon may be the case for other ribozymes as well and searched for structurally conserved sequences in the intergenic regions near those genes and hammerhead and twister ribozymes. Indeed, this strategy led to the discovery of conserved motifs, some of which later confirmed to be twister sister, hatchet and pistol ribozymes.

The group was searching for new ribozyme motifs in a collection of bacterial and archaeal genomes and environmental sequences, and subsequently identified 449 unique Pistol sequences. As later determined by two structural biology groups, cleavage mechanism has similar features to those of other nucleolytic ribozymes, such as pseudoknot structure and the use of guanosine as general base (Nguyen, Wang, & Steitz, 2017; Ren et al., 2016). The groups also discovered that highly conserved consecutive adenines form a A-minor motif that stabilizes pistol's overall fold (Nguyen et al., 2017).

In 2017, members of our group engineered an aptazyme based on a pistol ribozyme from a bacterium *Alistipes putredinis* using high-throughput screening *in vitro* (Kobori, Takahashi, & Yokobayashi, 2017). However, these aptazymes did not function in intracellular environment. To address this challenge, Nomura et al. generated a library of pistol ribozyme variants and tested them in mammalian cells (Nomura et al., 2017). The library contained 676

natural sequences from genomes of various organisms described by Weinberg et al. (2015) and 376 sequences generated by reshuffling of the ribozyme structural elements. Using their new screening method, Nomura et al. identified multiple pistol variants that were active in mammalian cells (Nomura, Chien, & Yokobayashi, 2017). This finding held promise for engineering pistol-based ribozymes, as even some of the synthetic reshuffled variants were active in mammalian cells (Supplementary Figure S1). However, our preliminary efforts to engineer aptazymes based on the pistol ribozymes active in mammalian cells yielded disappointing results (results not shown). Therefore, we were prompted to further modify the pistol ribozymes to create circularly permuted variants not found in nature.

Twister and pistol ribozymes activity in mammalian cells

To confirm their potential for engineering aptazymes for mammalian cells, I tested promising new ribozyme variants in mammalian cells. For the twister class, I selected ribozyme variants from the original Breaker's group's study (Roth et al., 2014). For the pistol ribozymes, I tested circularly permuted variants of synthetic ribozymes previously reported by Nomura et al (Nomura et al., 2017).

2.2. Results

All five of the selected twister variants were active when transfected to mammalian cells (Figure 2-2. Transfected plasmids contain EGFP reporter with the ribozyme embedded in 3'UTR, and self-cleavage of the ribozyme leads to mRNA degradation and reduction of EGFP expression. Activity of ribozymes can be assessed by comparing fluorescence of ribozyme-embedded plasmid to the control EGFP plasmid without the ribozyme insertion. Although all variants were active in cells and resulted in significant reduction of EGFP expression, the ribozymes from *Schistosoma mansoni P1* (sm P1), *Oryza sativa* (os) and *Nematostella vectensis* (nv) demonstrated highest activity, reducing EGFP expression by ≥95%.

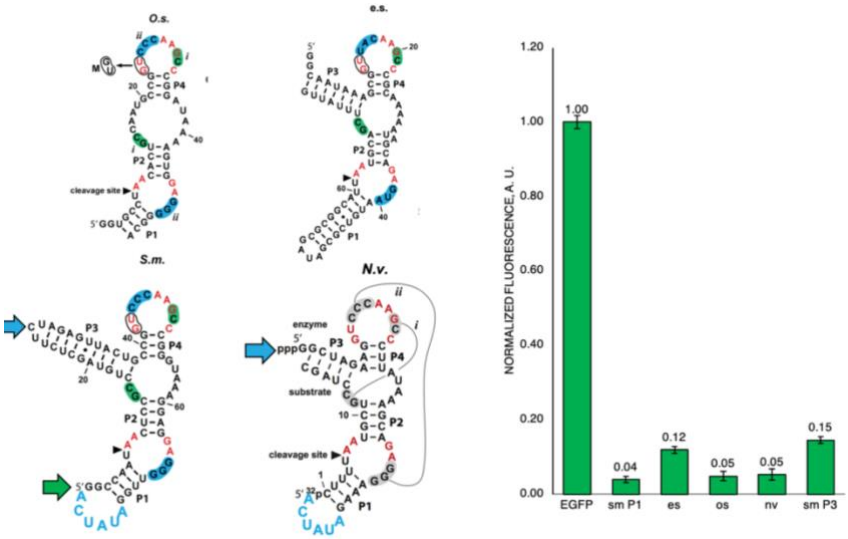


Figure 2-2. Activity of twister ribozymes in HEK293 cells. Left panel is adapted from Roth et al. (Roth et al., 2014). For sm twister two variants were created, starting at P3 (blue arrow and additional nucleotides in blue) and at P1 (green arrow). Bimolecular Nv twister from the original work has been converted into a single RNA molecule with 5' end at P3 stem (blue arrow) and with a loop (blue nucleotides) at the end of the P1 stem. Right panel shows reporter's fluorescence levels in HEK293 cells. EGFP is an empty vector, corresponding to maximum fluorescence. Os, sm P1 and nv twisters are highly active in mammalian cells. All values are normalized by mCherry fluorescence, a control for transfection efficiency.

Having confirmed high activity of twister ribozymes in mammalian cells, I decided to explore if this activity can be affected by the surrounding sequence. Ribozyme activity is dependent on its correct folding, and introduction of a competing structure may disrupt correct folding and affect ribozyme cleavage. Thus, for the three most active ribozyme variants I inserted a 10-nt complementary sequence upstream of the ribozyme to see if it can inhibit their activity (Figure 2-3).

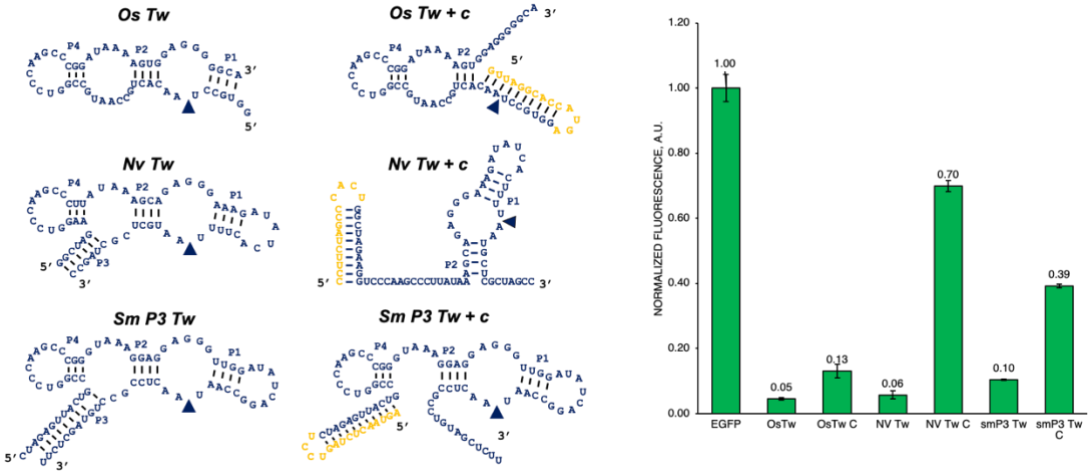


Figure 2-3. Twister ribozymes inhibition upon insertion of a complementary sequence upstream. Left: expected secondary structures of twister variants without and with the 10-nt complementary insert. The insert is expected to hybridize with parts of the ribozyme disrupting its self-cleaving activity. Right: activity of modified twisters in HEK293 cells. EGFP is an empty vector, corresponding to maximum fluorescence. Out of three variants Nv twister has the most significant inhibition upon insertion of the complementary sequence. All values are normalized by mCherry fluorescence, a control for transfection efficiency.

Indeed, I observed some level of cleavage inhibition for all three variants, most dramatically for the Nv variant. This result shows that it is possible to inhibit this ribozyme's activity by insertion of a complementary sequence upstream. The inhibition was most significant for Nv twister, from 6% to 70% of the control EGFP expression level.

I then moved on to testing activity of circularly permuted synthetic pistol ribozymes in mammalian cells.

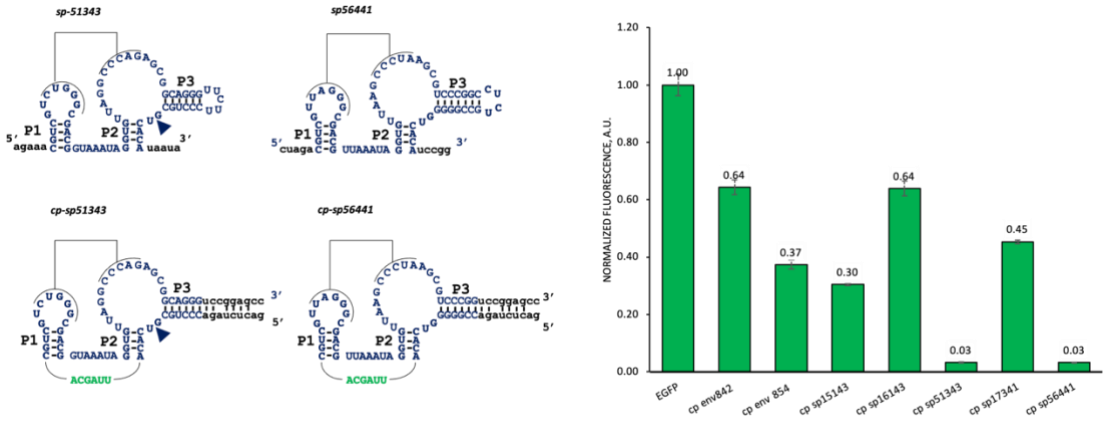


Figure 2-4. Circular permutation of synthetic pistol ribozymes. Right: activity of synthetic pistol ribozyme in HEK293 cells after the circular permutation. EGFP is an empty vector, corresponding to maximum fluorescence. Left: schematic representations of two most active circularly permuted pistol variants, cp-sp51343 and cp-sp56441. Artificial junction between stems P1 and P2 is marked in green (bottom panel). Inactivating mutations in sp-51343 are marked in red.

We picked some of the synthetic reshuffled variants that were active in mammalian cells in the previous work from our lab (Nomura et al., 2017), and then circularly permuted them by connecting P1 and P2 stems of these variants with a 6-nucleotide junction. As shown in Figure 2-4, after the circular permutation, only two of them, sp51343 and sp56441, maintained high activity. This finding makes these two variants novel synthetic pistol ribozymes with high activity in mammalian cells.

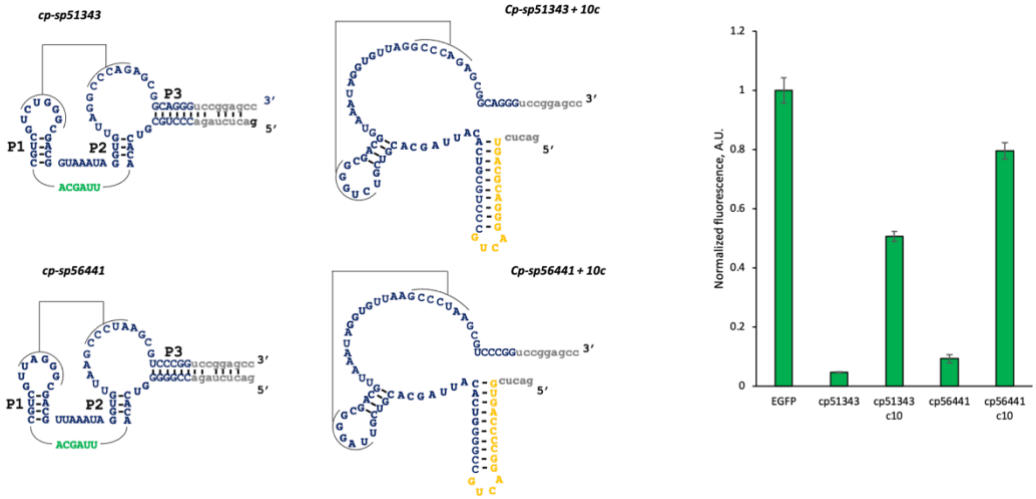


Figure 2-5. Circularly permuted pistol (CPP) ribozymes inhibition upon insertion of a complementary sequence upstream. Left: expected secondary structures of CPP variants without and with the 10-nt complementary insert. The insert is expected to hybridize with parts of the ribozyme disrupting its self-cleaving activity. Right: activity of modified CPPs in HEK293 cells. EGFP is an empty vector, corresponding to maximum fluorescence. Both cp51343 and cp56441are significantly inhibited by the complementary sequence. All values are normalized by mCherry fluorescence, a control for transfection efficiency.

Next, I tested the effect of the surrounding sequence on these two highly active circularly permuted pistols (CPP) variants, sp51343 and sp56441. Same as with twister ribozymes, I inserted a 10-nt complementary sequence upstream of the ribozyme to see if it will be able to inhibit their activity. Similar to twister ribozymes, the cleavage of the two CPP variants, cp51343 and cp56441 was dramatically inhibited upon insertion of the complementary sequence.

2.3. Discussion

As described in chapter 1, aside from one HDV-based example, all currently existing aptazymes for mammalian systems were based on inserting an aptamer in one of the stem-loops of the hammerhead ribozyme. While this approach yielded a range of both on- and off-switches (Auslander et al., 2010; Beilstein et al., 2015; Kennedy et al., 2014; Nomura et al., 2013; Stifel et al., 2019; Zhong et al., 2016), new ribozyme scaffolds and architectures for mammalian gene expression control are necessary to improve aptazyme modularity and foster new applications. Both twister and circularly permuted pistol ribozymes are yet unexplored scaffolds for aptazyme design. Although twister has been used for circular RNA generation by the Jaffrey group and for aptazymes in yeast by the Fussenegger group, there were no examples of aptazymes with it for mammalian cells by the start of this project (Litke & Jaffrey, 2019) (Felletti et al., 2016). (Strobel et al reported their twister-based switch in 2020 (Strobel, Sporing, et al., 2020))

Pistol ribozymes family is even less explored for synthetic biology applications and especially in mammalian cells. Aside from the high-throughput characterization of the Pistol-

based aptazyme *in vitro* and screening of Pistol variants *in vivo* in our lab (Nomura et al., 2017) (Kobori et al., 2017), there have not been any published attempts to engineer it as a synthetic biology tool. In this chapter I showed that this ribozyme can remain active even after a circular permutation of the reshuffled variants, which introduces a novel ribozyme scaffold not existing in nature. In contrast to other reported circular permutations of ribozymes that resulted in another naturally occurring topology (Roth et al., 2014) (C. E. Weinberg et al., 2019), circular permutation of the pistol ribozymes connects two distant termini by a flexible linker resulting in a new topology not found in the natural pistol ribozymes. The ability of pistol to maintain high activity in mammalian cells upon connection of two distant termini suggests that the connecting fragment (marked in green on Figures 2-4 and 2-5) is flexible and may tolerate further modifications, such as aptamer insertion. This makes circularly permuted pistol ribozymes a promising tool for mammalian cell applications.

2.4. Conclusion and outlook

In this chapter, I showed that both twister and pistol ribozyme families are promising and yet unexplored scaffolds for aptazyme engineering in mammalian cells. From both of these classes, I identified ribozyme variants that are highly active in mammalian cells. Of those, circularly permuted pistol ribozymes have a previously undescribed topology, which holds promise for aptazyme engineering. Furthermore, I established that the activity of some of the highly active ribozymes, such as Nv twister, cp56441 and cp51343 pistol, can be inhibited by insertion of an upstream sequence complementary to the ribozyme. This finding emphasizes the importance of the sequence context for the ribozyme activity and may be exploited to engineer new aptazymes. Addition of these two new ribozyme classes to the list of available scaffolds can help develop new designs and applications for aptazymes in mammalian cells.

2.5. Methods

Plasmid construction. All EGFP-aptazyme constructs were derived from pEGFP-N1 (Appendix). Ribozyme sequences were inserted between XbaI and BspEI sites as listed in Table 1. Ribozymes were inserted by whole plasmid PCR with primers containing the ribozyme sequence in the overhangs. Linear PCR products were then treated with DpnI, purified with DNA Clean & Concentrator kit (Zymo Research), phosphorylated and ligated. Ligation products were transformed into home-made competent Top10 *E. coli* cells. Several colonies were randomly picked to inoculate overnight liquid cultures. Plasmids used for transfection were purified using Zyppy Plasmid Miniprep kit (Zymo Research). After plasmid extraction, their sequences were verified by Sanger sequencing.

Table 1. Inserted twister ribozyme sequences

Twister variant	Sequence inserted between XbaI (red) and BspEI (green) sites
Schistosoma mansoni (P1)	tctagaggccaataactccgcctgtagctctactagagttactgccggtcccaagcccgggtaaaggag gagggttggcctccgga
Schistosoma mansoni (P3)	tctagactagagttactgccggtcccaagcccgggtaaaggaggagggttggatatcaggccaataact ccgcctgtagctctttccgga

Nematostella vectensis (P3)	tctagaggctagaaggtcccaagcccttataaagcagagggaaagatatcacttttaatgctgcctagct ccgga
Environmental sequence	tctagaggcaataaagcggttacaagcccgcaaaaatagcagagtaatgtcgcgatagcgcggcatta atgcagctttattgtccgga
Oryza sativa	tctagaggtgcctaacactgccaatgccggtcccaagcccggataaaagtggagggggcatccgga

Table 2. Inserted pistol ribozymes sequences

Twister variant	Sequence inserted between XbaI (red) and BspEI (green) sites
Cp env842	tctagacgggaggtagtcaattaaactcgatacggcgagtataaataggactttaagccgtaagcgttcccgt ccgga
Cp env854	tctagagggtaggtcacatagaacactcgactaggcgagtataaacaggtgtcaagcctagtgcgttaccct ccgga
Cp sp15143	tagaccctgcgtcacattaagaactcgactatgcgagtataaataggtgttaggcatagtgcggcagggtcgg
Cp sp16143	tctagaccctgcgtcacattaaaaactcgttagggcgagtataaataggtgttaagccctaagcggcagggtc cgga
Cp sp51343	tctagaccctgcgtcacattagcacgtcgtctgggcgacggtaaataggtgttaggcccagagcggcagggtc ggga
Cp sp17341	tctagaccgggggtcacattaaaaactcgtggcggcgagtgtaaataggtgttaagccgccagcgtcccgg tccgga
Cp sp56441	tctagaccgggggtcacattagcacgtcgttagggcgacgttaaataggtgttaagccctaagcgtcccggtcc gga

Transfection. HEK293 cells were cultured in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% heat-inactivated FBS (Gibco) containing 2 mM L-glutamine and 100 units/mL of penicillin-streptomycin (DMEM-FBS). Cells were kept in a 37°C incubator with 5% CO_2 and passaged regularly upon reaching 90% confluency.

Approximately 20 h prior to transfection, the cells were trypsinized, diluted to ~2.7×10^5 cells/mL and 100 µL/well were seeded onto a 96-well plate. Cells in each well were cotransfected with 100 ng of the EGFP- aptazyme plasmid and 20 ng of pCMV-mCherry (transfection control) using 0.3 µL of TransIT-293 Transfection Reagent (Mirus) according to the manufacturer's instructions. Forty-eight hours after transfection, the medium in each well was replaced with 100 µL of phosphate buffered saline (PBS), and fluorescence intensity was measured by Infinite M1000 PRO microplate reader (Tecan). Fluorescence intensity was measured at 484 nm excitation/ 510 nm emission/5 nm bandwidth for EGFP, and at 587 nm excitation/610 nm emission/10 nm bandwidth for mCherry. Background fluorescence

measured using untransfected cells was subtracted from the EGFP and mCherry fluorescence values. Then, EGFP fluorescence was normalized by mCherry fluorescence to account for variations in transfection efficiency and cell counts. All reported values are averages of three replicate wells.

2.6. Supplementary

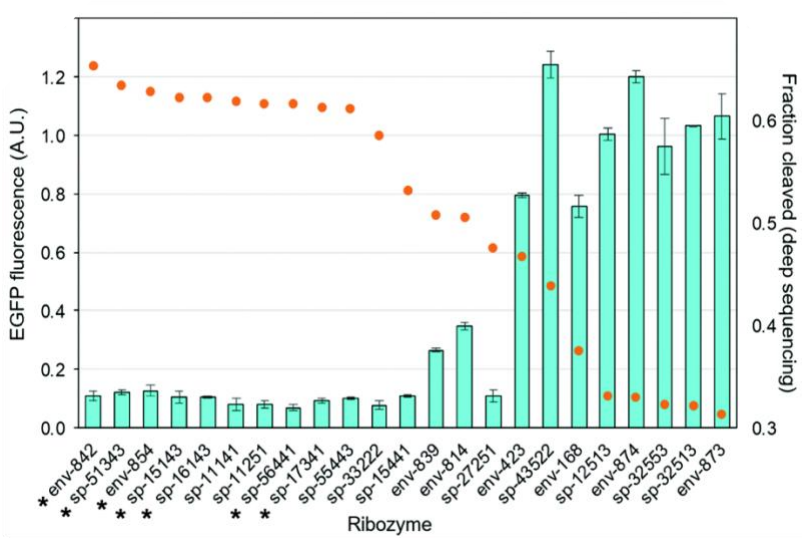


Figure S1. Activities of pistol ribozymes in mammalian cells from Nomura et al. (Nomura et al 2017).
Variants used in this chapter are marked with a star (*)
(A) Ribozyme activity based on EGFP assay in HEK293 cells. (orange circles are Fc values derived from deep sequencing). EGFP fluorescence of HEK293 cells transfected with an EGFP–ribozyme plasmid was normalized to the fluorescence of the cells transfected with an empty (no ribozyme) plasmid.

3. Chapter 3. *In vitro* screening of aptazyme libraries for mammalian riboswitches

3.1. Introduction

Most of the existing aptazymes were designed by inserting an aptamer into one of the stems of a ribozyme and optimizing so-called communication module, the sequence connecting these two elements. While there are a few different ribozyme scaffolds used for off-switches, on-switches have exclusively been developed with hammerhead ribozyme. For example, one of the earliest on-switches developed by the Smolke group, was made by fusing an MS2 aptamer in stem III of the hammerhead ribozyme (Kennedy, Liang, & Smolke, 2013). In the K19 on-switch developed by the Suess' group, tetracycline aptamer is fused at stem I of the hammerhead ribozyme (Beilstein et al., 2015). In their design, tetracycline binding prevents the loop-loop interactions critical for the ribozyme cleavage. The on-switch recently reported by the Hartig lab is based on the same design, but the tetracycline aptamer is replaced with a guanine aptamer (Stifel et al., 2019).

Instead of the classical fusion at the stem, in this work we suggest a new design, where an aptamer is inserted directly upstream of the ribozyme (Figure 3-1). This architecture has been previously demonstrated in our lab with guanine aptamer and a pistol ribozyme *in vitro* (Kobori et al., 2017*)*. When placed in tandem, the ribozyme (*Rz*) and aptamer (*Apt*) become mutually exclusive structures competing for the functional conformation. In such tandem architecture binding of the aptamer ligand stabilizes aptamer structure, thus preventing ribozyme folding and, therefore, inhibiting its cleaving activity. To control the stability of the two states, we insert a communication module upstream (anti-rz) that will hybridize with 5' of the ribozyme stem (P_{rz}). When the communication module is fully complementary to the 5' of P_{rz}, it will form a stem at the base of the aptamer (P_{ap}) and thus disrupt functional ribozyme conformation. Alternatively, when the communication module is not complementary to the ribozyme, P_{rz} is stable and the ribozyme can take its functional conformation. We hypothesize that with a certain composition of the communication module, binding of the ligand will shift the balance of the aptazyme structures towards *Apt* state, thus inactivating the ribozyme and allowing further mRNA processing (Figure 3-1).

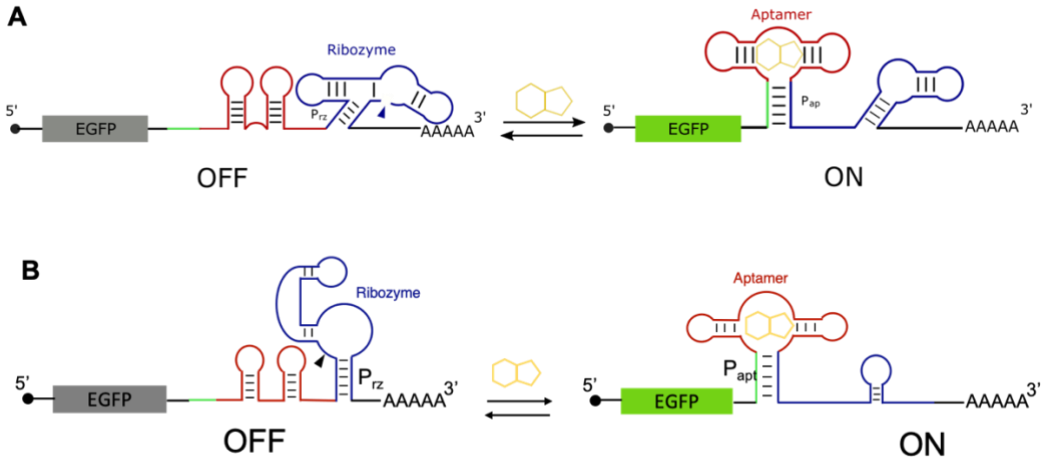


Figure 3-1. Tandem architecture for mammalian ON-switches with a twister ribozyme (A) and a synthetic circularly permuted pistol ribozyme (CPP) (B). In the absence of ligand, the aptazyme preferentially forms the Rz structure, which results in ribozyme self-cleavage and subsequent mRNA degradation. Aptamer-ligand binding changes the aptazyme conformation to Apt, disrupting ribozyme folding and cleavage and therefore allowing protein translation.

Finding the sequence composition allowing optimal balance between the *Apt* and *Rz* states is a challenge that requires testing a large number of variants. As discussed in Chapter 1, methods for optimizing the communication module in mammalian cells by the start of this project were restricted to rationally designing a handful of candidates and testing them individually (*in vivo* screening methods were published by Xiang et al and Strobel et al later (Xiang et al., 2019),(Strobel, Sporing, et al., 2020)). This traditional method is time-consuming and laborious, and the search for the optimal communication module would be greatly facilitated by high-throughput screening. Established high-throughput methods described up to the start of this project were ran *in vitro*, bacteria and yeast systems. As discussed in Chapter 1, due to differences in context, none of these screening systems can guarantee equal aptazyme performance in mammalian cells. Nevertheless, of the three systems, screening *in vitro* has the highest throughput and is the least time- and labor-consuming. Furthermore, screening run in the low Mg^{2+} concentration in our lab previously yielded aptazymes functional in mammalian cells (Kobori, Nomura, Miu, & Yokobayashi, 2015; Kobori & Yokobayashi, 2018). Aptazyme activity in the desired mammalian cell context can be confirmed by testing *Apt* and *Rz* state constructs in mammalian cells prior to screening. Then, we can generate a library of aptazyme variants by randomizing the communication module and screening them *in vitro* in low [Mg^{2+}] environment to mimic intracellular conditions. Finally, selected candidates can be tested in HEK293 cells and their on- and off-levels can be fine-tuned by introducing single-nucleotide mutations or insertions.

Synthetic pistol variant cp-sp41343 and twister variant from *Nematostella vectensis* introduced in Chapter 2 are promising candidates for the tandem architecture as their activity can be inhibited by insertion of a complementary sequence upstream. This property is critical for the tandem architecture, since it reveals the ability of the communication module (anti-Rz) to hybridize with 5' of P_{rz} and form P_{ap} instead, thus disrupting the ribozyme activity. If this property maintains upon insertion of the aptamer between the ribozyme and anti-Rz, the construct can be screened for potential aptazymes.

We chose guanine aptamer as the sensory module for the new aptazyme. Molecules for controlling gene expression must meet several criteria, such as cell membrane permeability, low toxicity, good solubility and little or no biological activity. Moreover, availability of corresponding aptamers restricts the list of commonly used ligands to a handful of molecules, such as tetracycline, theophylline and guanine. While none of these compounds are completely bio-orthogonal, there are currently no better ligands available. We decided to start with guanine, since it is a natural metabolite and has an aptamer from the naturally occurring riboswitch from 3'UTR of *Bacillus subtilis* (Mandal, Boese, Barrick, Winkler, & Breaker, 2003). This aptamer has been previously used in our group for mammalian riboswitch design (Nomura et al., 2012) (Nomura et al., 2013).

3.2. Results

3.2.1. Preliminary testing of tandem architecture

In Chapter 2, I demonstrated that ribozyme activity is sensitive to the sequence context, since its activity can be inhibited by insertion of a complementary sequence upstream. Therefore, first I confirmed that the activity of the twister variant from *Nematostella vectensis* stays high even when the guanine aptamer is inserted upstream of the ribozyme. Although the aptamer sequence does not contain regions complementary to the ribozyme, introduction of a structured RNA sequence upstream may inadvertently affect ribozyme folding. As shown in

Figure 3-2, insertion of the guanine aptamer had only a minor effect on the ribozyme activity (construct "0c"), and EGFP expression remained low. Next, I checked if the ribozyme can still be affected by the upstream complementary insertion, even when separated by the aptamer sequence (construct "10c", Figure 3-2). Indeed, "10c" construct's EGFP expression was very close to that of the empty control plasmid ("EGFP", Figure 3-1), suggesting that the ribozyme cleavage is mostly inhibited.

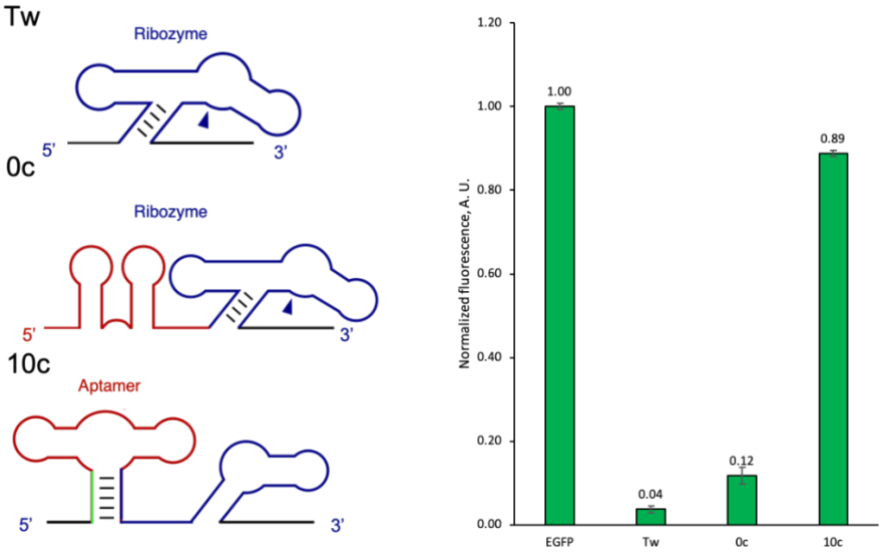


Figure 3-2 Effect of the upstream sequence on the twister-guanine aptazyme conformation. Left: schematic representations of the three tested constructs. 0c: Guanine aptamer (red) was inserted immediately upstream of the ribozyme (dark blue). 10c: 10 complementary nucleotides (green) inserted upstream of the aptamer disrupt ribozyme conformation and promote aptamer folding. Right: fluorescence levels of HEK293 cells transfected with a plasmid containing EGFP and one of the constructs in its 3'UTR. "Tw" and "0c" have decreased EGFP expression, reflecting high cleavage activity. In "10c" EGFP expression is restored due to ribozyme inhibition. "EGFP" control is transfected with a plasmid with EGFP only, without the ribozyme contstructs, for reference of basal expression

I repeated the same experiment for the circularly permuted pistol (CPP) variant cp-sp51343 to test if it preserves the same activity in the presence of guanine aptamer (Figure 3-3). Same as the Nv twister, CPP sp51343 maintained high activity with the aptamer inserted upstream and reduced activity after the addition of the complementary sequence.

This finding shows that both ribozymes' activity can be controlled by the composition of the sequence upstream of guanine aptamer, which sets the basis for the tandem aptazyme architecture. The levels of GFP expression for 0c and 10c constructs match the expected activities of the *Rz* and *Apt* states and show the wide range in which the ribozyme activity can be controlled by the composition of the upstream sequence. The next challenge is to explore other compositions of the 10nt upstream sequence (communication module) and find a sequence that allows transition between the mutually exclusive *Apt* and *Rz* states in response to ligand. To find such an optimal sequence we generated libraries of the communication module mutants for both ribozyme scaffolds and screened them with the high-throughput *in vitro* method previously developed in our lab (Figure 3-4) (Kobori et al., 2015).

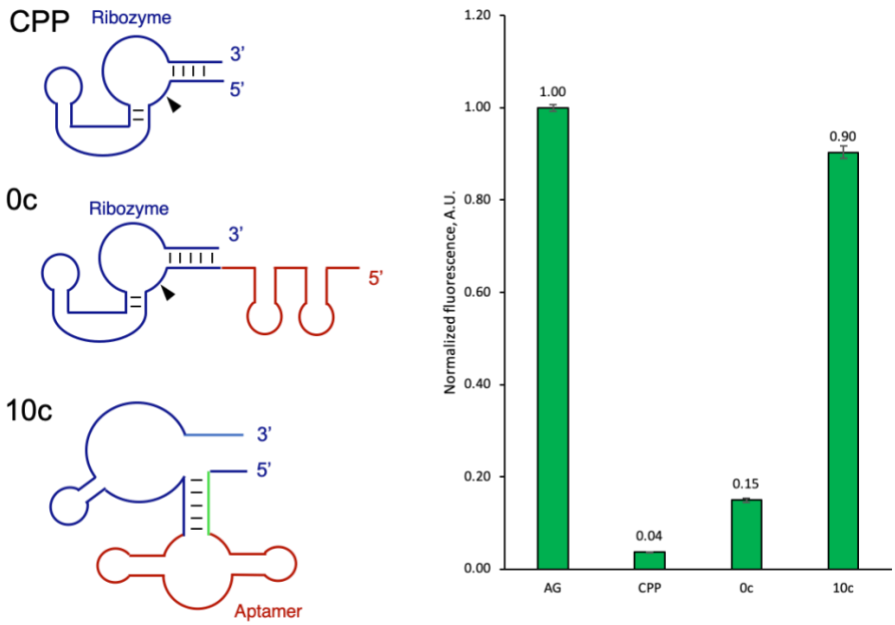


Figure 3-3. Effect of the upstream sequence on the pistol-guanine aptazyme conformation. Left: schematic representations of the three tested constructs. 0c: Guanine aptamer (red) was inserted immediately upstream of the ribozyme (dark blue). 10c: 10 complementary nucleotides (green) inserted upstream of the aptamer disrupt ribozyme conformation and promote aptamer folding. Right: fluorescence levels of HEK293 cells transfected with a plasmid containing EGFP analog Azami Green (AG) and one of the constructs in its 3'UTR. CPP and 0c have decreased AG expression, reflecting high cleavage activity. In "10c" AG expression is restored due to ribozyme inhibition. "AG" control is transfected with a plasmid with Azami Green only, without the ribozyme constructs, for reference of basal expression. We used Azami Green instead of EGFP due to unexpected interaction between EGFP and guanine discovered in later experiments

3.2.2. High-throughput *in vitro* screening.

The screening method allows running a self-cleavage reaction for thousands of the aptazyme variants in a single test tube (Figure 3-4). Since aptazyme cleavage is co-transcriptional, cleaved and uncleaved aptazyme variants can be size-separated on a polyacrylamide gel after transcription. Once extracted from the gel, the fragments are barcoded in the subsequent reverse transcription and PCR steps and sequenced.

First, I tested the method with the twister-guanine aptazyme library design summarized in Figure 3-5A. Three nucleotides closest to the aptamer were fixed to form a stem, since previous library screenings revealed conservation of the stem at the base of the aptamer in aptazymes with high dynamic range (Kobori et al., 2017). Furthermore, conserving just 3 nucleotides of the communication module, reduces the library size from 4^{10} (1048576) to 4^{7} (16384) variants which allows screening them all in a single MiSeq run.

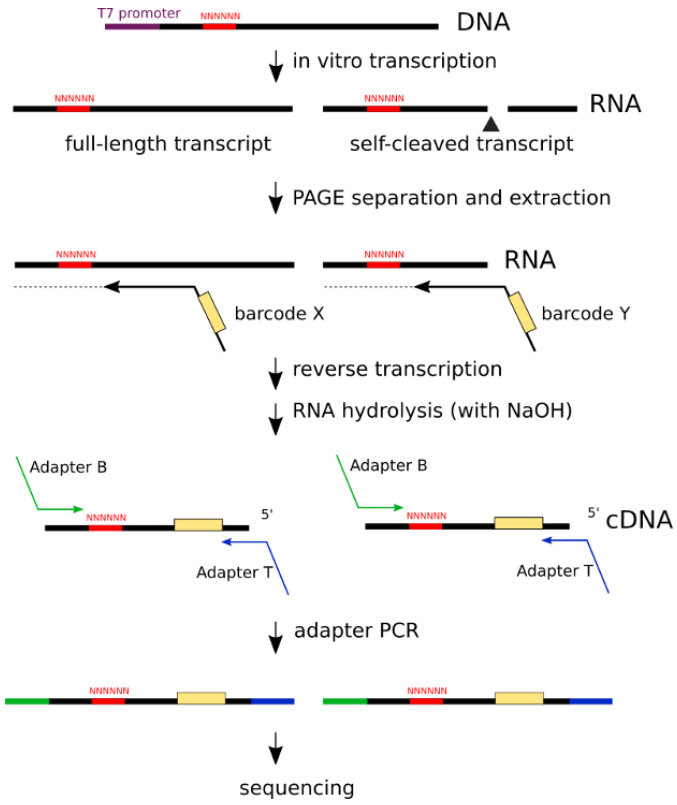


Figure 3-4. Schematic workflow of the high-throughput in vitro aptazyme screening. First, DNA library with a randomized region and a T7 promoter is transcribed *in vitro* in the presence and absence of ligand. Then cleaved and uncleaved fragments resulting from co-transcriptional cleavage are gel-separated and reverse transcribed with barcoded primers. Then RNA is hydrolysed and cDNAs are purified. Sequencing adapters are attached in the final PCR and the library is sequenced with Illumina MiSeq machine.

I processed the twister-guanine N7 library according to the workflow summarized in Figure 3-4. [Mg^{2+}] in a transcription reaction is a crucial factor for both transcription and aptazyme self-cleavage as it is known to stabilize RNA structure and serves as a co-factor in ribozyme catalysis (Gebetsberger & Micura, 2017) (Ren et al., 2016). Therefore, it was adjusted for every library individually. Free cytosolic [Mg^{2+}] in a mammalian cell is reported to range between 0.5 to 1 mM (Romani, 2011); however, this concentration is too low for *in vitro* transcription reaction to yield sufficient RNA products. Thus, for every library we used the lowest [Mg^{2+}] that could stably yield transcription product. For twister-guanine library it was determined to be 3 mM.

The library was sequenced on an Illumina MiSeq machine, and resuting reads were sorted based on barcodes, trimmed and counted (detailed information about the run is in the Supplementary Table S1). Barcoded reads distribution (Figure 3-5C) revealed that “full” reads without guanine constitute much smaller fraction of the total reads than those with guanine. This confirms our original expectation that the library is more active in the absence of ligand, and, therefore, only few of the variants remain intact, while most of the variants are self-cleaved. To assess the aptazyme activity we introduced the parameter “fraction cleaved”, which is calculated by dividing the number of cleaved reads over total reads for the given variant. By comparing this parameter between the “lig-“ and “lig+” conditions for a given

variant we can find aptazymes that change their activity in response to the ligand. When plotted based on their fractions cleaved, sequences nonresponsive to the ligand are expected to align along the y=x line, whereas potential switches will be located in the corners away from that line. On-switches for mammalian cells are expected to have lower cleavage in the presence of ligand, and thus, they should be located in the bottom right corner of the plot ("Fc-">"Fc+"). While twister-guanine N7 library has a large number of highly active variants regardless of presence of ligand, clustered in the right side of the plot, many of them have reduced activity in the presence of ligand (bottom right corner). This once again confirms that the library has a large number of potential switches.

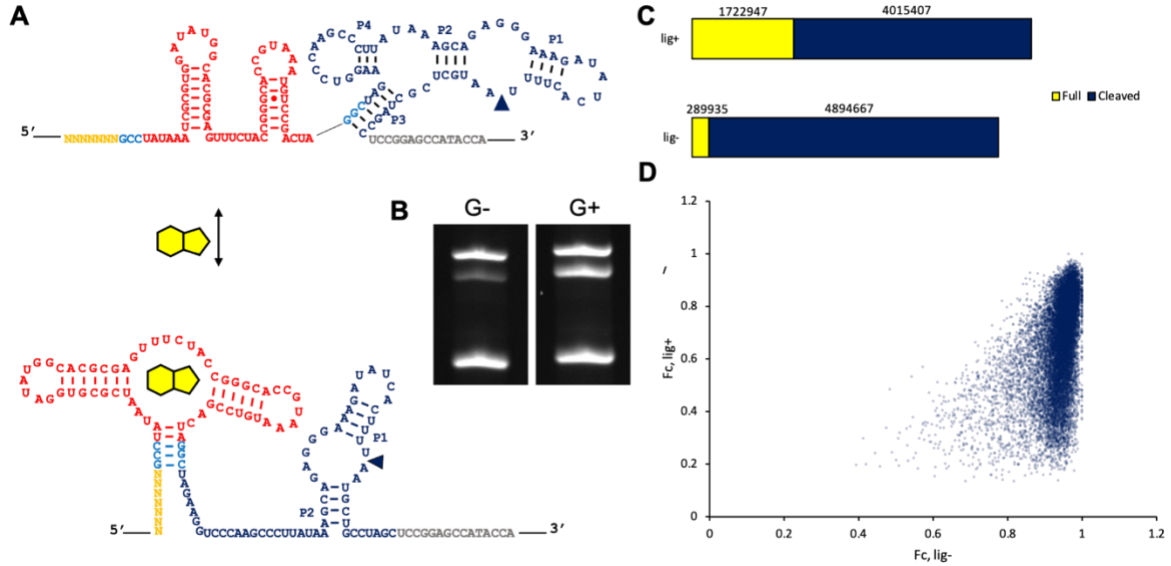


Figure 3-5. A summary of in vitro screening of a 16384 variants twister-guanine aptazyme library. (A) Proposed library conformation in the presence and absence of guanine. (B) PAGE analysis of the library transcribed *in vitro* at 3 mM $MgCl_2$. Top band (194nt) – DNA template, middle band (176nt) – full-length RNA fragment, bottom band (140nt) – cleaved RNA fragment. (C) Reads distribution of the four barcoded fractions. (D) Fraction cleaved for all variants in the absence and in the presence of ligand. The population with higher fraction cleaved in the absence than in the presence of ligand (Fc- high, Fc+ low) found in the bottom right part of the plot represents potential on-switches.

From this population of potential switches, I picked six variants based on their ratios of fractions cleaved and read counts (Figure 3-6B, Table S2), cloned them into the pEGFP plasmid and subsequently transfected to HEK293 cells. When tested in mammalian cells, the selected aptazymes were indeed able to increase EGFP levels upon addition of guanine (Figure 3-6A). It is difficult to establish a quantitative correlation between EGFP levels and fractions cleaved because the parameters measured are different (cleaved fraction of RNA *in vitro* and protein expression level in cells). However, qualitatively, the method is expected to narrow down a large sequence pool to a smaller pool of switching candidates. Indeed, variants that were responsive to the aptamer ligand *in vitro* demonstrated the same property in intracellular context, showing the potential of the method for identifying switches for mammalian cells.

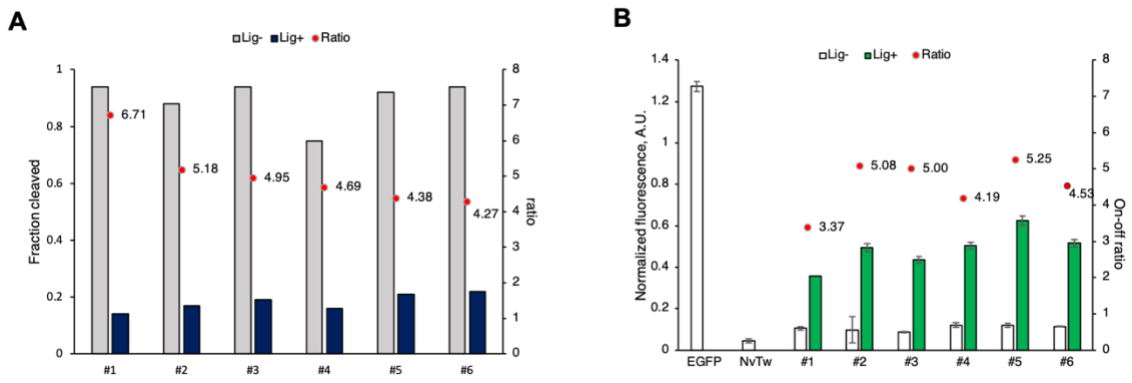


Figure 3-6. Evaluation of the candidate aptazyme switches in HEK293 cells. (A) fractions cleaved of the aptazyme candidates calculated from the sequencing reads. (B) activity of the aptazyme candidates in mammalian cells in the presence (lig+, 250μM) and absence (lig-, 0 μM) of guanine. Aptazyme sequence is inserted in the 3' UTR of a EGFP-embedded plasmid that is then transfected into HEK293 cells. High EGFP level corresponds to low self-cleavage, and low EGFP level corresponds to high aptazyme cleavage. Since the parameters measured are not the same for the two experiments, only qualitative correlation between the EGFP activity and sequencing reads is expected.

Aside from identifying promising switches, the first sequencing run showed that a library with 7 randomized position (16384 variants) is too large to have full coverage in one MiSeq run. Most of the variants had under 100 reads in all conditions, which could introduce noise when calculating the fractions cleaved. To improve coverage of the library, I designed a new, smaller library with 6 randomized positions in the communication module, yielding 4096 variants. In this new library the first nucleotide of the upstream sequence has been fixed as "C", since that position was conserved in most of the variants with high on-off ratios (Figure 3-7A).

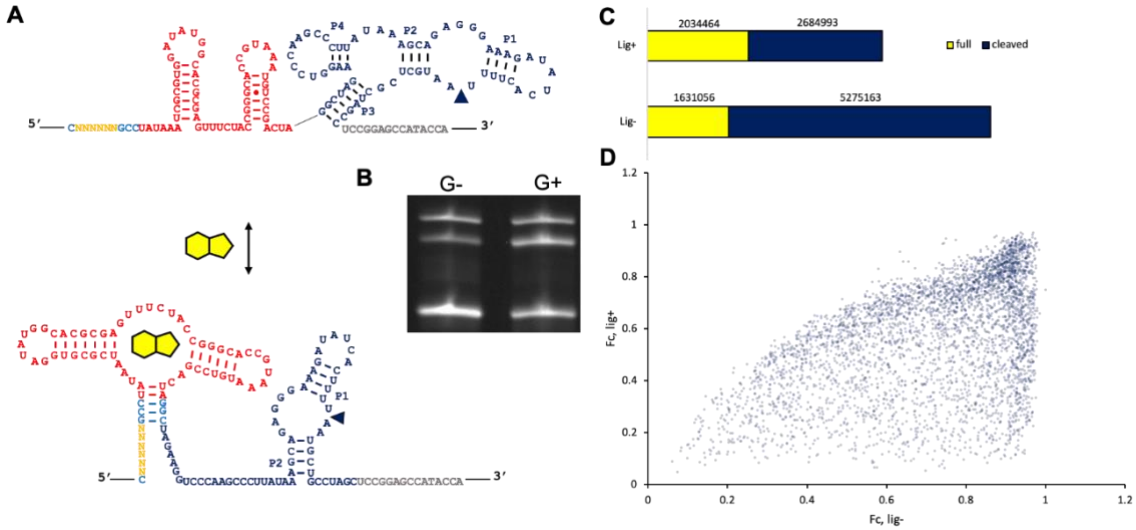


Figure 3-7. A summary of in vitro screening of a 4096 variants twister-guanine aptazyme library. Notation is the same for Figures 3-8 and 3-9: (A) Proposed conformation of the library in the presence and absence of guanine. (B) PAGE analysis of the library transcribed in *vitro* at 3 mM $MgCl_2$. Top band (194 nt) – DNA template, middle band (176 nt) – full-length RNA fragment, bottom band (140 nt) – cleaved RNA fragment. (C) Reads distribution of the four barcoded fractions. (D) Fraction cleaved for all variants in the absence and in the presence of ligand. The population with higher fraction cleaved in the absence than in the presence of ligand (Fc- high, Fc+ low) found in the bottom right part of the plot represents potential on-switches. The library seems to contain plenty of potential switches.

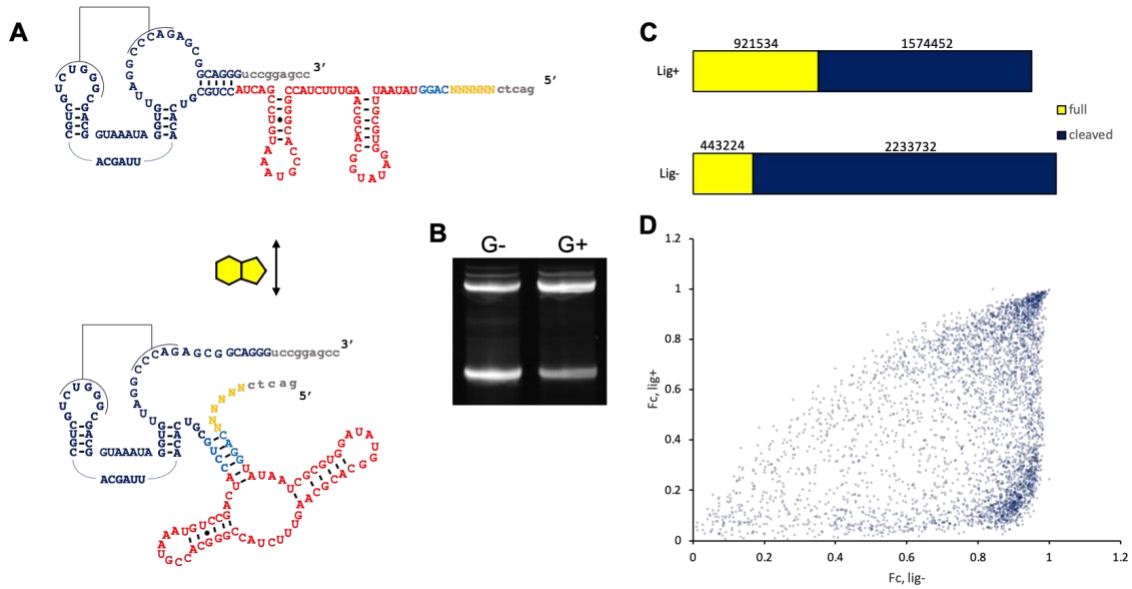


Figure 3-8. A summary of in vitro screening of 4096 variants of the pistol-guanine library. Notation is the same as in Figure 3-7. (B) Library was transcribed at 4 mM $MgCl_2$. Top band (235 nt) – DNA template, middle band (218nt) – full-length RNA fragment, bottom band (151 nt) – cleaved RNA fragment. (D) The population with higher fraction cleaved in the absence than in the presence of ligand (Fc- high, Fc+ low) found in the bottom right part of the plot represents potential on-switches. The library seems to be enriched in on-switches.

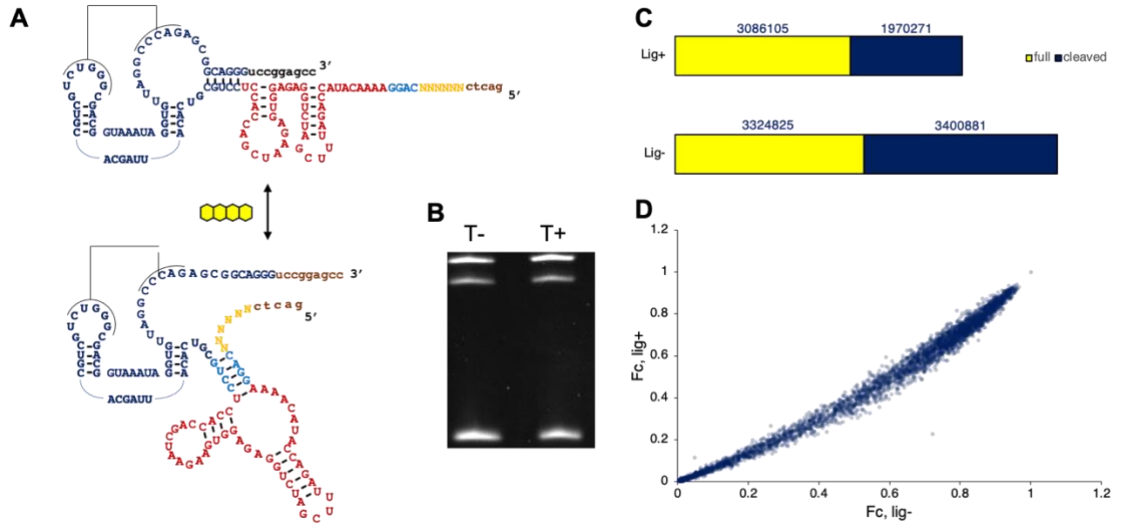


Figure 3-9. A summary of high-throughput in vitro screening of 4096 variants of the pistol-tetracycline library. Notation is the same as in Figure 3-7. (B) The library was transcribed at 3 mM $MgCl_2$. Top band (245nt) – DNA template, middle band (228nt) – full-length RNA fragment, bottom band (150 nt) – cleaved RNA fragment. (D) The population with higher fraction cleaved in the absence than in the presence of ligand (Fc- high, Fc+ low) found in the bottom right part of the plot represents potential on-switches. The library does not contain variants responsive to tetracycline.

The new, smaller library was processed following the same protocol as the original N7 library and sequenced with the help of OIST SQC section. The run yielded 11625676 reads after sorting and quality filtering, so most of the variants passed the coverage filter of over 100 reads per condition. Despite only one nucleotide difference in the library design, reads distribution was different from N7 library, most noticeably in the absence of ligand. While in the original N7 library the number of "full" reads was very low in the absence of ligand, in the new N6 library this number has grown, implying that there were fewer highly active variants (Figure 3-7C). The reads distribution represented by fractions cleaved was also different, with aptazyme activities ranging wider along the x-axis. There was still a cluster of highly active aptazymes in the top right corner, but candidates inhibited by the presence of

ligand (potential switches) had varying levels of fractions cleaved instead of being clustered in the right corner.

Screening with twister revealed that libraries with six randomized positions are optimal for sufficient reads coverage for a single MiSeq run. Hence, for the circular pistol libraries, I fixed four nucleotides closest to the aptamer to form a stem and randomized the remaining six positions (Figure 3-8A). In addition to guanine aptamer, I also ran screening for a library with the tetracycline aptamer (Figure 3-9A). For transcription of tetracycline library [Mg^{2+}] was kept at 3 mM, whereas for guanine library it had to be increased up to 4 mM, since full-length transcript was not visible on a gel with lower [Mg^{2+}] concentrations. While for the guanine library it was clear from the gel (Figure 3-8B) and the reads distribution (Figure 3-8C) that the library contains plenty of switching variants, for the tetracycline library both gel bands brightness (Figure 3-9B) and reads distributions (Figure 3-9C) were similar in the presence and absence of ligand. This trend was even more apparent in the scatter plot of fractions cleaved (Figure 3-8D), where the guanine library has a large cluster in the bottom right corner of the plot. In contrast, in the same plot for the tetracycline library all variants are aligned along the y=x line, indicating that although the library covers a wide range of cleaving activities, it does not respond to the ligand (Figure 3-9D). This result suggests that not every aptamer is able to induce the desired ligand inhibition of ribozyme cleavage. Binding affinity of the aptamer is unlikely to be the reason, as K_d of tetracycline aptamer is estimated to be ~0.8 nM (Berens, Thain, & Schroeder, 2001), which even lower than K_d (5 nM) of guanine aptamer (Mandal et al., 2003). On the other hand, guanine aptamer comes from a natural bacterial riboswitch, which might give it some feature that *in vitro* selected tetracycline aptamer does not have. It is also possible that tetracycline aptamer sequence happens to have some unexpected interaction with the ribozyme. Hence, ability of tandem architecture to generate a switch *in vitro* has some dependency on the aptamer.

3.2.3. Validating riboswitch candidates in mammalian cells.

Having run *in vitro* screening for the three libraries, I selected variants with highest on-off ratios to further test them in mammalian cells (candidates' profiles are available in Table S3). Since tetracycline library screening did not identify ligand-responsive variants, I only tested candidates from the two guanine libraries.

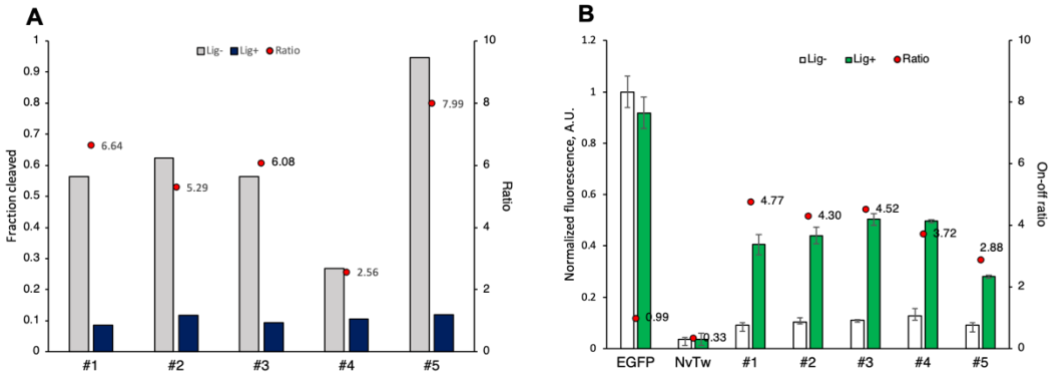


Figure 3-10. Evaluation of the candidates switches from the N6 twister-guanine library in HEK293 cells. Part 1. (A) Fractions cleaved of the best aptazyme variants selected by NGS screening tested in HEK293 cells. (B) The aptazyme sequence of variants "#1"-"#5" is inserted in the 3' UTR of an EGFP-embedded plasmid that is then transfected into HEK293 cells. All selected candidates activate gene expression in response to the ligand (250 μM guanine). "EGFP" is a control plasmid without aptazyme. "NvTw" is a control plasmid with Nv twister ribozyme only (without aptamer).

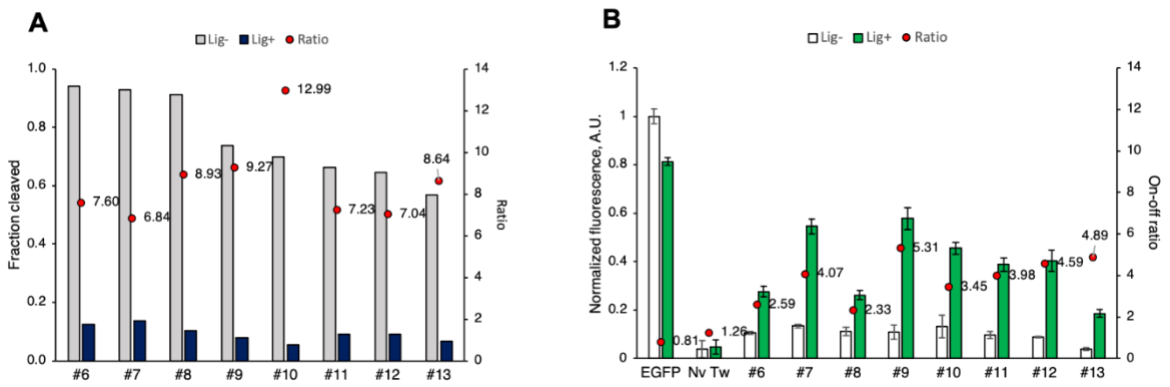


Figure 3-11. Evaluation of the candidates switches from the N6 twister-guanine library in HEK293 cells. Part 2. Notation is the same as Figure 3-10. All tested candidates function as on-switches in mammalian cells.

All of the 13 candidates selected from the twister-guanine library (Figures 3-10 and 3-11) were indeed functioning as on-switches in HEK293 cells. While some candidates had modest on-off ratios (<3 for variants #5, #6, #8), some activated gene expression by 5-fold (#9, #1). On-off ratios of EGFP fluorescence in mammalian cells were generally not as high as ratios of fractions cleaved *in vitro*. This is not surprising, since protein expression level measured in cells involves many more processes than simple transcription *in vitro*.

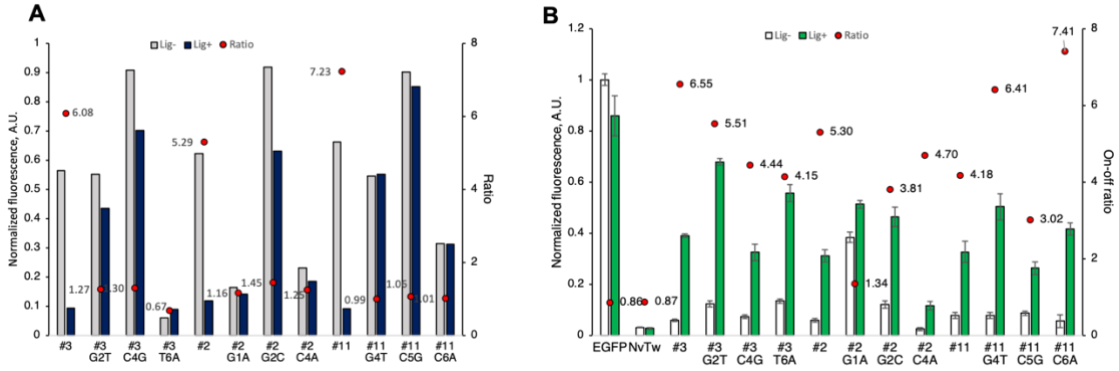


Figure 3-12. Evaluation of the candidate non-switches from the N6 twister-guanine library. Part 1. (A) Fractions cleaved of non-switching single mutants of the previously shown switches. Letters after the number of the switch indicate the mutation in the variable region. For example, variant "#3 G2T" is the mutant of variant #3, where G in the second position has been changed to T. (B) activity of the mutants in HEK293 cells. Although mutated variants were not expected to switch, many of them (#3 G2T, #3 T6A, #11 G4T, #11 C6A, etc) were also on-switches in HEK293 cells, some even demonstrating better performance than the original top candidates

To verify the ability of the method to qualitatively select switches from a mixed population I compared activities of "good" and "poor" switches from the screening. I picked three previously tested switches (#2, #3 and #11) that had single-nucleotide mutants with low on-off ratios *in vitro* and tested them in cells (Figure 3-12. In spite of expected poor response to the ligand predicted by sequencing-based screening, most of the single-nucleotides mutants turned out to function as on-switches in cells. Importantly, some of the supposedly "poor" switches exhibited even higher on-off ratios than the original sequences. For example, #11_C6A significantly outperformed the original switch #11, having almost double of its on-off ratio and a higher on-level. #2_G1A was the only mutant with expected low cleavage activity in both presence and absence of ligand.

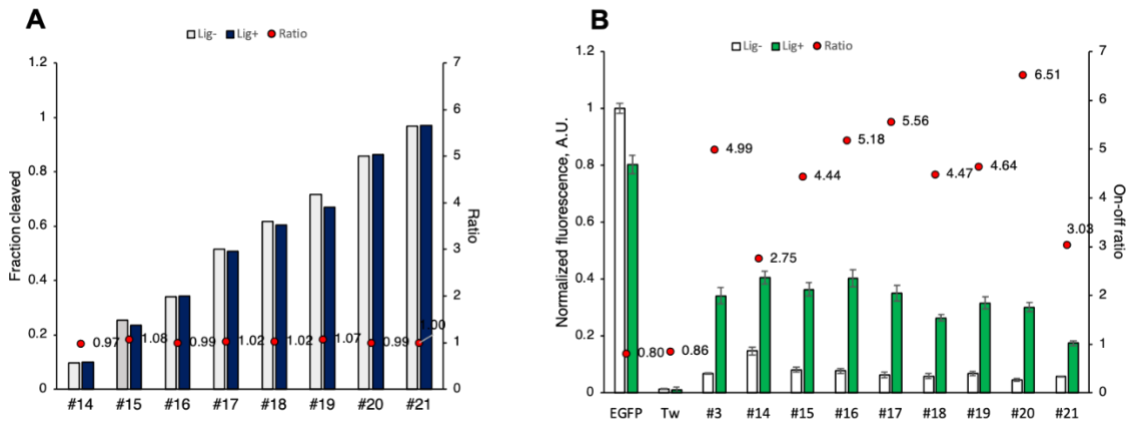


Figure 3-13. Evaluation of the candidate non-switches from the N6 twister-guanine library. Part 2. (A) non-switching variants from different ranges of fractions cleaved. (B) activity of the candidates in HEK293 cells. All newly selected non-switching candidates demonstrated switching activity in cells, some with on-off ratios exceeding 5.

To further explore the relationship between the activity of aptazymes *in vitro* and in cells, I picked eight more non-switching variants. This time I sorted them by fractions cleaved and picked a variant from each activity level, ranging from 0.1 to 0.9 in fractions cleaved (Figure 3-13), to see if some *in vitro* activity levels have better correlation with the intracellular activity. However, again, all eight new variants turned out to be on-switches in cells, regardless of their activity level assessed by sequencing.

I repeated the validation experiment in smaller scale with the candidates from the pistol-guanine library (Figure 3-14 and Table S4). Same as with the twister library, regardless of their activity profiles from screening, all of the candidates turned out to be modest on-switches when tested in cells. Thus, for both pistol- and twister-based libraries validation experiments showed that *in vitro* screening approach is not able to select switches functioning in mammalian cells.

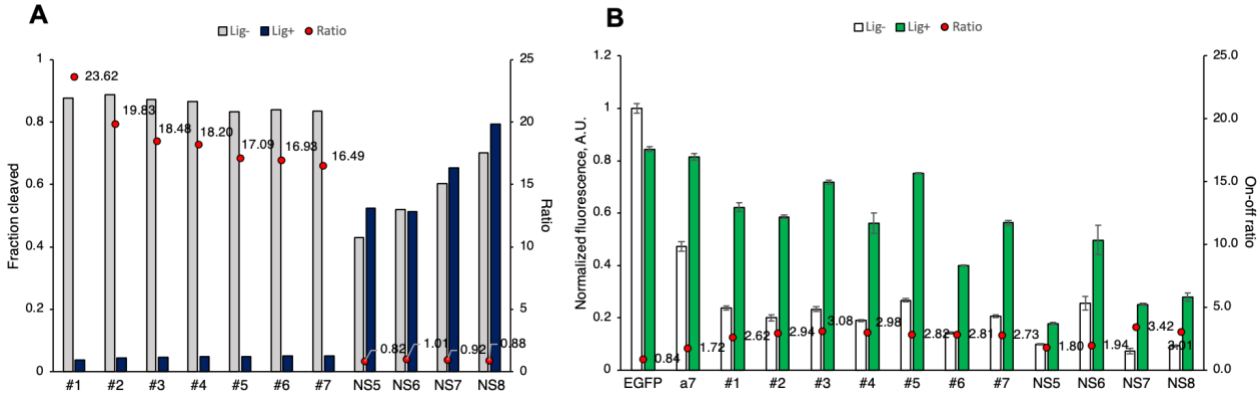


Figure 3-14. Evaluation of the candidate non-switches from the N6 pistol-guanine library. (A) fractions cleaved of the seven best (#1-#7) and the four worst (NS1-NS4) on-switches based on the sequencing result. (B) Activity of the aptazyme candidates in HEK293 cells. Same as with twister-guanine, activity of the candidates assessed by screening is not consistent with their activity in cells.

The first successful validation experiments for the candidates switches from N7 and N6 twister-guanine libraries probably were artifacts of the library design, as it was highly enriched in on-switches. Therefore, any random variant picked from the library was more likely to be a switch than a non-switch. Interestingly, we saw the same pattern with the pistol-guanine library, which may also be related to the high fraction of switches in the library. Nevertheless, the high-throughput screening method failed to select on-switches from the

aptazyme libraries based on two different ribozyme scaffolds, as some of the variants nonresponsive to the ligand *in vitro* turned out to be best switches in cells.

Although this method worked for HDV-based aptazymes in papers previously published by our lab, it did not work for the twister- and circular pistol- based libraries in this project. Many factors could contribute to this observation. First, *in vitro* screening only contains a transcription step and does not take into account post-transcriptional events in the cell, such as post-transcriptional RNA modifications and subsequent translation. Second, even the transcription itself is fundamentally different between the mammalian cell and the test tube. Mammalian cell's transcription machinery involves coordinated action of protein complexes and translocation from the nucleus to the cytoplasm with local concentration gradients, whereas *in vitro* transcription buffer only contains essential ions, T7 RNA polymerase and NTPs. In addition to that, RNA half-life and stability in the cytoplasm depends on protein degradation complexes, which are difficult to mimic in an *in vitro* experiment.

3.2.4. Optimization of reaction conditions in vitro

Since reduction of [Mg^{2+}] to 4 mM in previous work was sufficient to select aptazymes for mammalian cells, I did not consider other parameters that may affect aptazyme performance in a mammalian cell. However, different factors may be important for the activity of different aptamer and ribozyme classes. Although we cannot consider all possible conditions and mimic all complex processes happening in mammalian cells, by testing different parameters of *in vitro* transcription we may come across some that are crucial for aptazyme activity. Hence, I decided to optimize *in vitro* reaction conditions by varying length of transcription, molecular crowding conditions and concentrations of reaction components. I ran optimization experiments with three aptazyme variants with different profiles (Figure 3-15). #3 was a switch with on-off ratio of ~6 both *in vitro* and in cells, whereas #3_G2T and #11_C6A were not responsive to the ligand based on the *in vitro* screening result and gel quantification but were activating gene expression in cells. I varied parameters of *in vitro* transcription for these variants to find such a condition where activities of #3_G2T and #11 C6A *in vitro* will be closer to those in cells.

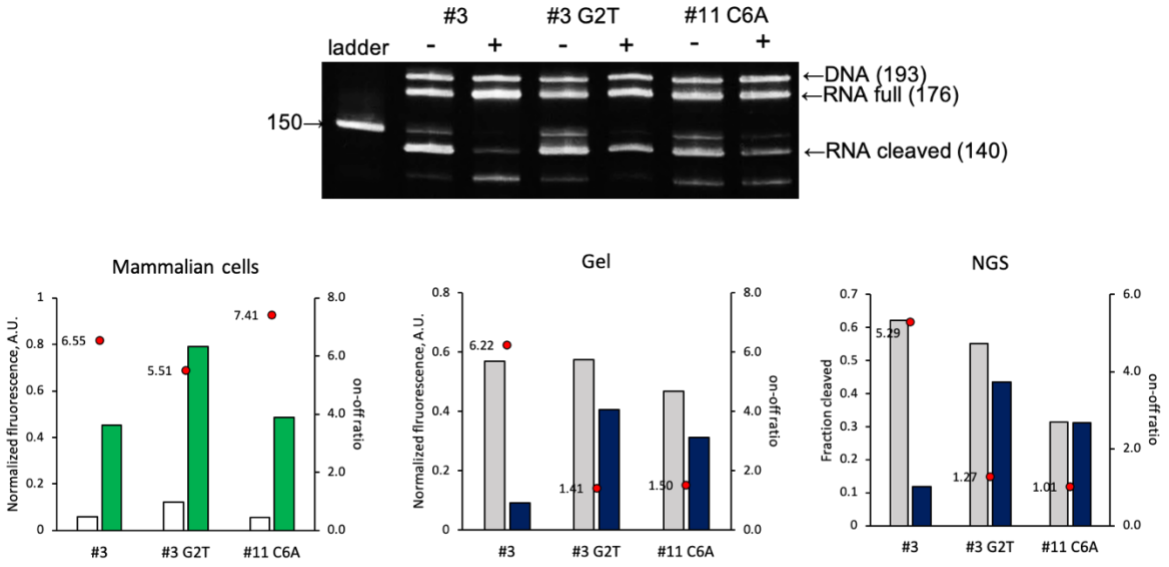


Figure 3-15. Activity of three aptazyme candidates assessed by in vitro screening (bottom right), PAGE (bottom middle) and in mammalian cells (bottom left). Gel image (top) of the *in vitro* transcribed candidates was used to quantify aptazyme's cleaving activity for the bottom chart. While the activity of "#3" is consistent between three assays, it varies drastically between *in vitro* and intracellular contexts for "#3 G2T" and "#11 C6A".

Although [Mg^{2+}] had been optimized for each library prior to screening, I only considered its effect on the transcription efficiency and not on the ligand response. Hence, I ran *in vitro* transcription for #3_G2T variant in lower [Mg^{2+}] (Figure 3-16) and quantified the products on a polyacrylamide gel. Same as before, decreasing concentrations of [Mg^{2+}] negatively affected the yield of reaction, with 2 mM condition only having very faint bands and 1mM not having RNA product at all. However, interestingly, reduction of [Mg^{2+}] helped to improve ligand response, and #3_G2T's on-off ratio approaching 2 in 2 mM [Mg^{2+}]. Unfortunately, due to dependency of transcription yield on [Mg^{2+}], there is no information on aptazyme activity in even lower concentrations. Moreover, even in 2 mM [Mg^{2+}] the reaction is not stable and due to batch-to-batch variations the product was sometimes difficult to detect on a gel.

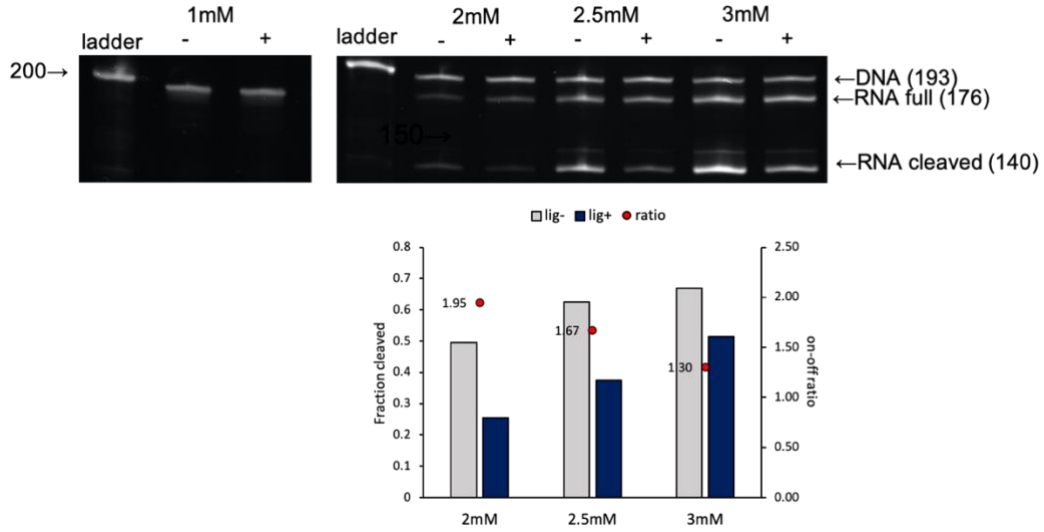


Figure 3-16. Effect of the Mg^{2+} concentration on the "#3 G2T" candidate's activity assessed by PAGE. Top: PAGE result of the "#3 G2T" variant in different [Mg^{2+}] in the absence (-) and presence (+) of 250µM guanine. Bottom: quantification of the "full" and "cleaved" RNA bands based on the gel image. While transcription efficiency drops with lower [Mg^{2+}], on-off ratio of the aptazyme increases slightly.

Another potentially important but overlooked factor in this screening was molecular crowding. Molecular crowding refers to the excluded volume effect that results from high total concentration of macromolecules inside cells (Ellis, 2001). Molecular crowding is known to affect free ion concentrations (Leamy, Assmann, Mathews, & Bevilacqua, 2016), equilibrium constants of biochemical reactions (Leamy et al., 2016; Nakano & Sugimoto, 2016), ribozyme docking (Paudel & Rueda, 2014) catalytic activity of nucleic acids (Dupuis, Holmstrom, & Nesbitt, 2014; Nakano & Sugimoto, 2016), and has been studied extensively in the recent years. Macromolecular crowding is commonly mimicked by 20 w/v% of PEG8000 (later referred to as "PEG") in catalytic RNA studies (Nakano & Sugimoto, 2016), and therefore, I tested aptazyme activity (#3 variant) in this condition first. However, when product of *in vitro* transcription at 20% PEG was run on a gel, no bands were detected (Figure 3-17). Importantly, even DNA band was not visible on the gel, probably because the solution was too viscous to diffuse through the gel. Thus, it is not clear if RNA has been transcribed and if cleavage activity was affected. However, since at 5% PEG DNA bands were visible, but RNA product was not detected still, it seems that transcription is inhibited by high PEG concentrations. At lower 2% PEG RNA bands were observed though, suggesting that transcription is inhibited at PEG concentrations between 2 and 5%.

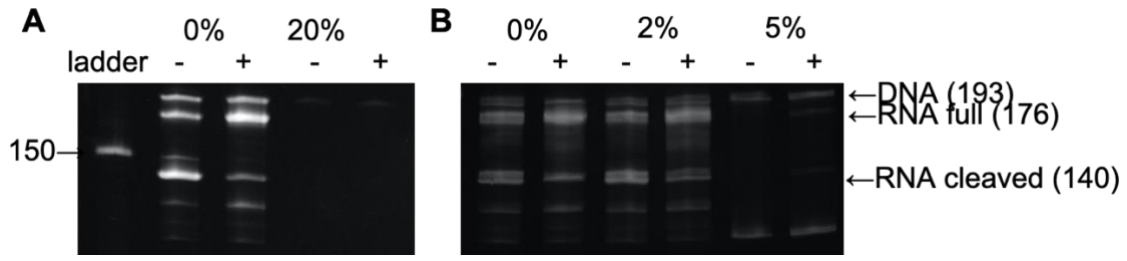


Figure 3-17. Effect of PEG8000 on transcription assessed with "#3 G2T" candidate by PAGE. (A) cleavage activity of "#3 G2T" at 0% and 20% of PEG8000. (B) Cleavage activity of "#3 G2T" at lower concentrations of PEG8000. First line of labels: PEG8000 w/v percentage. Second line of labels: presence or absence of guanine (250 μM). Transcription proceeded at 2w/v% PEG8000, but no products were observed at 5 w/v% and 20 w/v% PEG8000.

Since *in vitro* transcription did not produce RNA at higher than 5% PEG, I tested the effect of molecular crowding with lower PEG concentrations, ranging from 0 to 4% for the #11_C6A variant (Figure 3-18). Addition of PEG at such concentrations did not affect neither fractions cleaved, nor response to the ligand. Thus, molecular crowding effect is difficult to simulate with PEG8000 for T7 polymerase-based *in vitro* transcription. At the 20w/v % normally used in molecular crowding studies, it inhibits transcription, whereas at lower concentrations it has no effect. This finding is corroborated by the study by Tan et al (2013) in artificial cells, where a sharp drop of transcription efficiency was observed at a point between 1 and 10% of molecular crowding agent (Dextran 6000) due to reduced diffusion rate of the T7 RNA polymerase (Tan, Saurabh, Bruchez, Schwartz, & LeDuc, 2014). In the same study bigger molecular crowding reagents, such as Dextran 200,000, do not have inhibitory effect on transcription, and therefore could potentially produce more conclusive results for co-transcriptional aptazyme cleavage. Unfortunately, currently there are not many studies on the effect of molecular crowding on T7 polymerase *in vitro* transcription. More commonly used model is cell-free protein expression in *E. coli* extract. Interestingly, in such systems PEG8000 in low concentrations (1-2%) is used to increase the protein yield; however, since these systems are based on *E. coli* polymerase in cell extract and measure protein (not RNA) output, they are difficult to compare with our *in vitro* transcription system (Shin & Noireaux, 2010) (Deng et al., 2018).

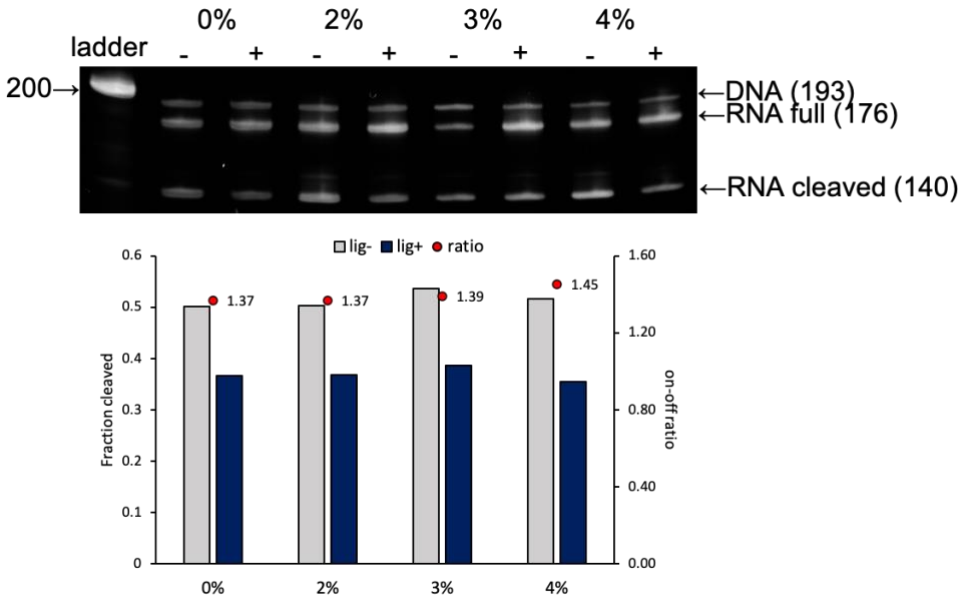


Figure 3-18. Effect of different concentrations of PEG8000 on the "#11 C6A" candidate's activity assessed by PAGE at 3 mM [Mg^{2+}]. Top: PAGE result. Top row of labels indicates the concentrations of PEG. The second row of labels indicates addition of 250 µM guanine. Bottom: quantification of the gel bands' intensity. None of the tested concentrations of PEG8000 affected aptazyme's on-off ratio.

Another factor that could affect aptazyme activity is the duration of reaction. It is difficult to predict the speed of aptazyme folding and cleavage in intracellular conditions. Furthermore, self-cleavage will only affect subsequent mRNA processing if it happens before mRNA is degraded by intracellular machinery. mRNA half-life in the cytoplasm of a mammalian cell ranges from a few minutes to several hours (C. Y. Chen, Ezzeddine, & Shyu, 2008), and it is not known to which side of the spectrum transfected EGFP transcript belongs. Therefore, I tested the time-dependence of the aptazyme cleavage reaction (#11_C6A variant) by quenching the transcription reaction with an equal volume of 20 mM EDTA at different time points. Since the aptazyme cleavage is co-transcriptional, duration of cleaving reaction is also affected by the duration of transcription. Transcription time was varied from the standard 180 minutes to 20 minutes at 37°C (Figure 3-19). Furthermore, aptazyme cleavage was tested at reduced temperature (30°C) for 30 mins to slow down the transcription process. Interestingly, even after 20 mins of transcription at 37°C RNA product was clearly visible on the gel, although with reduced yield. All shorter reactions contained an unidentified shorter band, which was probably an incomplete transcript. Overall, fractions cleaved did not change significantly among the four conditions, indicating that aptazyme cleavage is not affected on this time scale.

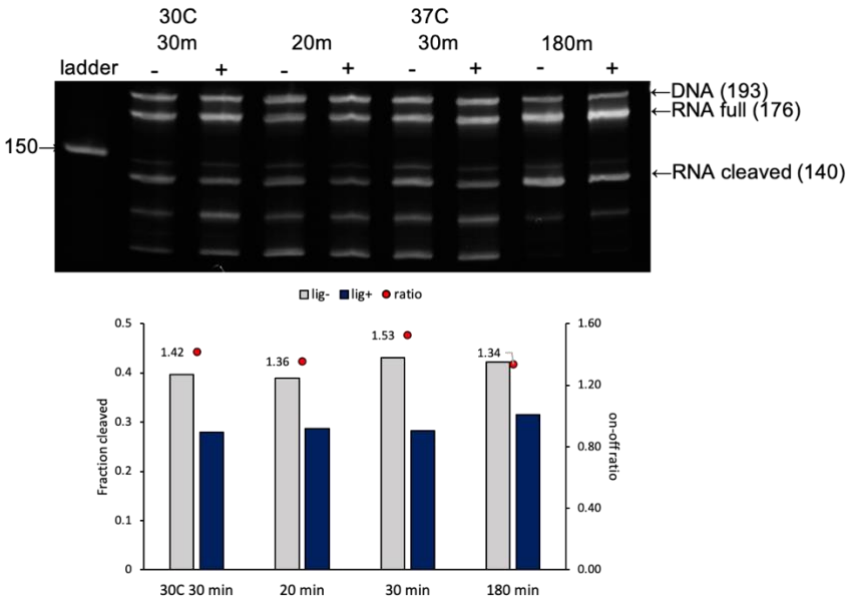


Figure 3-19. Effect of shorter transcription time on the "#11 C6A" candidate's activity assessed by PAGE at 3 mM [Mg^{2+}]. Top: PAGE result. Top row of labels indicates the temperature of the reaction. The second row of labels indicates duration of the reaction in minutes. The third row indicates addition of 250 µM guanine. Bottom: quantification of the gel bands' intensity. Transcription yields additional short products (see extra bands) when run for shorter time, compared to the original reaction duration (180 minutes). However, aptazyme's ligand response is not affected in any of the conditions.

Lastly, I tested aptazyme activity in lower concentrations of nucleotides (NTP) (Figure 3-20). According to Traut (1994), intracellular concentrations of free NTPs are in submillimolar range (Traut, 1994), whereas in T7 polymerase-based transcription they are present at 2 mM each to maximize the product yield. I, therefore, assessed aptazyme cleavage of #11_C6A variant at 8-, 64- and 128-fold reduced nucleotides concentration. Reduced NTPs could affect not only the speed of transcription, but also effective [Mg^{2+}] concentration as they can sequester free Mg^{2+} ions. Indeed, at decreased nucleotides concentrations, fraction cleaved was much higher than in the standard reaction (Figure 3-20). Nucleotides must have been depleted over the course of transcription, releasing Mg^{2+} ions and thus aiding aptazyme cleavage. Interestingly, transcription still proceeded even after the nucleotides were reduced over 100-fold, with almost all transcripts undergoing self-cleavage. Nevertheless, although decrease in nucleotides' concentration affected aptazyme cleavage, it did not improve response to the ligand.

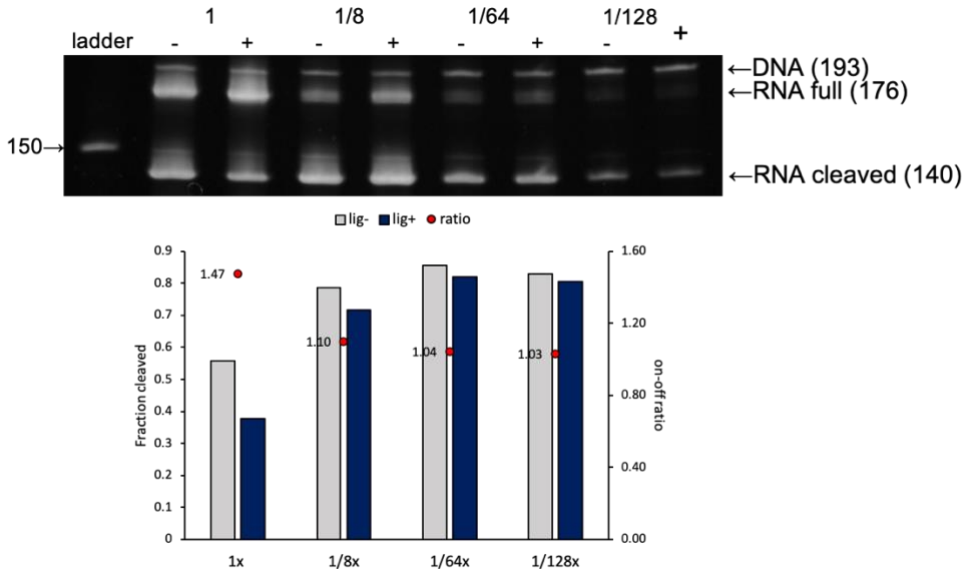


Figure 3-20. Effect of lower NTPs concentration on the "#11 C6A" candidate's activity assessed by PAGE at 3 mM [Mg^{2+}]. First row of labels indicates fold change in NTPs concentration relative to the original 2 mM each. Second row indicates addition of 250 µM guanine. Although with reduced yield, transcription is still able to proceed even at less than 1% (1/128) of the original nucleotides' concentration. Aptazyme cleavage increases with decreasing nucleotides' concentration, potentially due to the higher available [Mg^{2+}]. However, reduced NTPs do not increase on-off ratio of the candidate.

3.3. Discussion

3.3.1. Comparison of *in vitro* and *in vivo* results.

Interestingly, both *in vitro* screening and individual variants' testing in cells revealed that the library is enriched in on-switches. Although that finding coincided for both contexts, those on-switches were apparently not always the same across the two systems. Validation experiments in cells showed that not only some promising *in vitro* switches do not perform as well in mammalian cells (e.g. #8), but also that they can be significantly outperformed by some of the supposedly poor variants (#11_C6A). That being the case, the architecture seems to have a potential for both *in vitro* and mammalian systems, but variations in only six nucleotides of the communication module have surprisingly different effects across the two systems.

Although *in vitro* screening results could not be transferred to mammalian cells, validation experiments revealed plenty of good on-switches, confirming potential of the aptazyme libraries. Optimal communication module compositions ensuring high on-off ratios in mammalian cells remain to be found by other methods. Given the sensitivity of aptazymes to the environment, it is probably most reliable to run this optimization in mammalian cells directly.

3.3.2. *In vitro* screening optimization

Prior to starting this project there was contradicting evidence of transferability of aptazymes across different systems. On one hand, there were successful transitions like *in vitro* screened HDV-based aptazymes applied in mammalian cells in our group (Kobori et al 2015, Kobori et al 2018) and CChMVd(-) and sLTSV(-) hammerhead motifs functioning across yeast,

bacterial and mammalian systems in the Hartig's group's study (Wurmthaler et al., 2018). On the other hand, pistol-guanine switches screened *in vitro* in our lab by Kobori et al. did not work in mammalian cells (Kobori et al., 2017) and there were multiple examples of bacterial switches failing to work in eukaryotic systems (Wurmthaler et al., 2018) (Auslander et al., 2010). Transition has been reported to be challenging even across yeast and mammalian cells, which are both eukaryotic systems (Wittmann & Suess, 2011).

From that evidence it looked like although transferability of aptazymes and ribozymes across different systems is challenging, it is possible for some ribozyme motifs. In this study we were hoping that preliminary testing of new ribozyme variants and aptazyme constructs in mammalian cells will be sufficient to assume that communication module mutants will have similar activity in low [Mg^{2+}]. However, screening and validation experiments showed that even single nucleotide mutations may have different effect *in vitro* and in mammalian cells (#3, #11 variants, Figure 3-12). Moreover, keeping [Mg^{2+}] at 3 mM was not sufficient to simulate intracellular environment for aptazyme library.

Although optimization of *in vitro* transcription led to some interesting findings, such as effect of reduced NTPs concentration and shorter reaction time, it was not able to simulate the aptazyme-based riboswitch function in mammalian cells. For example, T7 polymerase-based transcription could not proceed in near-physiological 1 mM [Mg^{2+}] and high molecular crowding condition simulated by 20% w/v PEG8000. Overall, attempts to optimize *in vitro* conditions only covered a small space of possible reaction conditions, so there is a possibility that a certain combination of parameters, for example low [Mg^{2+}], high weight molecular crowding agent, low NTPs concentration and very short reaction time, may be able to mimic intracellular environment. Finding this optimal combination without knowing key details about mRNA processing in a mammalian cell will require blindly testing thousands of conditions.

First, intracellular environment probably has local gradients of concentration, which may also differ based on the cell type and stage. Second, during mRNA maturation process, when aptazyme binding and cleavage are happening, mRNA molecule is translocated from the nucleus to the cytoplasm, which may have different environments. Furthermore, we do not know when exactly aptamer binding and ribozyme cleavage are happening. Does the intracellular environment affect aptamer binding? Is the ligand concentration different in the nucleus and in the cytoplasm? How fast does the aptazyme fold? Does self-cleavage happen before or after mRNA is exported from the nucleus? Answers to these questions are necessary to establish an *in vitro* screening system that would allow modelling aptazyme behaviour in mammalian cells. While it is possible that not all of these factors are important for aptazyme activity and even a simple model may be able to simulate mammalian cell environment, it is still important to find that parameter, or otherwise screen many more conditions to find it serendipitously. Meanwhile, as those questions remain open, the only reliable way to engineer aptazymes is to carry out all reactions in the same environment as where they are expected to be applied eventually.

3.4. Conclusion

Controlling aptazyme conformation and activity by inserting a complementary sequence upstream proved to be a promising strategy for development of riboswitches. Using EGFP as a reporter gene, we were able to decrease its expression levels to ~10% and then restore again to ~90% by changing the composition of the upstream sequence for twister and circularly

permuted pistol ribozymes. With a certain composition of this inserted sequence, the aptazyme may achieve a desired balance state between the aptazyme and ribozyme conformation, which can be shifted by the addition of the ligand. In line with our expectations, we found such sequences in the process of testing aptazyme candidates selected by high-throughput *in vitro* screening. However, the method was not providing consistent results for mammalian cells, most probably due to the difference in the test tube reaction and mRNA processing in the cell. High-throughput *in vitro* screening is an efficient way to characterize activities of thousands of aptazyme variants in a single experiment; however, in its current form it cannot be used to make predictions about aptazyme activities in contexts of living systems. Elucidation of factors crucial for aptazyme activity in the mammalian cell and systematic optimization of the screening conditions to account for those factors are necessary to ensure smooth transition between different contexts. However, such optimization would lie outside of the scope of this project and, therefore, to meet the goals of this project it is more reasonable to explore the promising tandem architecture with new ribozyme scaffolds in mammalian cells directly.

3.5. Materials and methods

Plasmid construction: same as chapter 2

HEK293 cell culture and transfection: same as chapter 2

Library information:

"TwGua" plasmid containing the tandemly arranged aptamer and ribozyme was used as a PCR template for the *in vitro* transcription library construction. T7 promoter and randomized region were included in the overhang of the forward primer.

Library Name	DNA Sequence	Length
TwGua N7	TAATACGACTCACTATAGGGaaagcggccgcgactctagaNNN NNNNGCCTATAATCGCGTGGATATGGCACGCAAGTTT CTACCGGGCACCGTAAATGTCCGACTAGGCTAGAAG GTCCCAAGCCCTTATAAAGCAGAGGGAAAGATATCA CTTTT^AATGCTGCCTAGCtccggagccataccacatttgta	193
TwGua N6a	TAATACGACTCACTATAGGGaaagcggccgcgactctagaCNNN NNNGCCTATAATCGCGTGGATATGGCACGCAAGTTTC TACCGGGCACCGTAAATGTCCGACTAGGCTAGAAGG TCCCAAGCCCTTATAAAGCAGAGGGAAAGATATCAC TTTT^AATGCTGCCTAGCtccggagccataccacatttgta	193

PistGua N6	TAATACGACTCACTATAGGGctgctggagttcgtgaccgccgccgg gatcactctcggcatggacgagctgtacaagtaaagcggccgcgactcNNNNNN CAGGTATAATCGCGTGGATATGGCACGCAAGTTTCTA CCGGGCACCGTAAATGTCCGACTACCTGCG^TCACAtta gcaCGTCGTCTGGGCGACGGTAAATAGGTGTTAGGCCC AGAGCGGCAGGGtccggagccataccacattt	245
PistTet N6	TAATACGACTCACTATAGGGctgctggagttcgtgaccgccgccgg gatcactctcggcatggacgagctgtacaagtaaagcggccgcgactcNNNNNN CAGGAAAACATACCAGATTTCGATCTGGAGAGGTGA AGAATACGACCACCTCCTGCG^TCACAttagcaCGTCGTC TGGGCGACGGTAAATAGGTGTTAGGCCCAGAGCGGC AGGGtccggagccataccacattt	235

Colour coding for the table: orange – T7 promoter, purple – communication module to control the newly introduced stem stability, blue – aptamer, green – ribozyme.

Library construction and *in vitro* transcription.

Individual ribozyme variants tested were first cloned in a plasmid and sequence verified. The plasmids were used as templates to prepare *in vitro* transcription templates by PCR. Randomized sequences and T7 promoter were added as overhangs in the primer sequence. The libraries were *in vitro* transcribed in the 50 µl reaction in transcription buffer (40 mM Tris-HCl pH 8.0, 2 mM spermidine, 10 mM DTT) containing T7 RNA polymerase (2.5 U/µl, New England Biolabs), RNase inhibitor (2 U/µl, Thermo Scientific), 2 mM NTPs each, 3 mM $MgCl_2$ and 5 pmol DNA template. The reaction was also supplemented by the ligand or ligand's solvent in the following concentrations:

Ligand	Lig-, final c	Lig+, final c
Tetracycline	0	100 µM
Guanine	3 mM NaOH	500 µM and 3 mM NaOH

Transcription reaction was run at 37°C for 3 hours. Upon completion of transcription reaction, 6 U of Dnase I (New England Biolabs) and 6 µl DNase I transcription buffer (final 10 mM Tris-HCl, 2.5 mM $MgCl_2$, 0.5 mM $CaCl_2$) were added and the reaction mixture was incubated on ice for 1h. After the reactions RNAs were purified with RNA Clean & Concentrator kit (Zymo Research) and eluted with 15 µl of 10 mM EDTA.

Isolation of RNA fragments

Purified and concentrated RNAs were loaded onto a denaturing polyacrylamide gel. Fifteen microliters of eluted RNA was mixed with an equal volume of 2x RNA loading dye. To avoid overloading, 30 µl of RNA-dye mix was split among 5 lanes, so eventually 6 µl of the mix was loaded per well. Denaturing PAGE (8% polyacrylamide, 8M urea, 19:1

acrylamide:bis-acrylamide) was run for 1 h at 200V unless otherwise noted. Before imaging, the gel was stained in 1xTAE with 1x SYBR Gold (ThermoScientific) for 15 min. Cleaved and uncleaved RNA fragments were separated based on size by PAGE and excised from the gel. Gel slices were then crushed and mixed with 300 µL of TENa buffer (200 mM NaCl, 10 mM Tri-HCl pH 7.5, 5 mM EDTA). The mixture was placed on a shaker overnight at 4°C to recover RNAs from the gel by passive elution. RNA fragments were subsequently precipitated with Quick-Precip Plus Solution (Edge Biosystems) according to the manufacturer's instructions. Precipitated RNAs were dissolved in 16 µL of 10 mM EDTA.

Reverse transcription and barcoding

Whole volume of eluted RNA fragments was used as a template for reverse transcription reaction. 12.5 µl of RNA fragments were mixed with 10 nmol NTPs (each), 100pmol corresponding reverse primer containing barcodes and RNase-free water to make up to total volume of 15.5 µl. This mixture was heated to 72°C for 3 min and then placed on ice. Then reverse transcription reaction was initiated by the addition of 250 U of Maxima H Minus Reverse Transcriptase (Thermo Fisher Scientific) and 10 U of murine RNase inhibitor (New England Biolabs), and reverse transcription buffer (Thermo Fisher Scientific) to make up to final 1x in 20ul (50 mM Tris-HCl (pH 8.3 at 25°C), 75 mM KCl, 3 mM $MgCl_2$, 10 mM DTT). Reaction was run at 50°C for 1 h and was followed by heat-inactivation of the enzyme at 65°C for 20 minutes. Upon the reaction completion 1 µL of 5 M NaOH solution was added to each tube, and the tubes were incubated at 95°C for 5 min. The resulting cDNAs were first cleaned by Oligo Clean & Concentrator kit (Zymo Research) and then ran on a denaturing PAGE (8% polyacrylamide, 8 M urea, 19:1 acrylamide:bis-acrylamide) to separate desired cDNAs from excess reverse primers. The band of desired size was extracted in the same manner as RNA after *in vitro* transcription, except that gel slices of cleaved and uncleaved bands of the same condition (e.g. lig-) were crushed in the same tube to minimize bias during extraction. After extraction, cDNAs were resuspended in 10 µL RNase-free water.

Barcodes encoding the presence of the ligand were attached as primer overhangs in the final PCR with Q5 HF MasterMix (New England Biolabs). Length barcodes (full vs cleaved) differed in length (8 and 10 nt, or 9 and 11 nt) to shift identical nucleotides reads and avoid overclustering during MiSeq sequencing. The number of PCR cycles was kept to a minimum, and varied between 5 and 9, depending on the library. Amplified library was run on a 2% agarose gel, and the band of the corresponding size was extracted using Zymoclean DNA Gel Recovery kit (Zymo Research) to remove unreacted primers with adapters. After gel extraction, the library was dissolved in 10 µL RNase-free water. Final library composition is summarized in the table below.

Library Name	Sequence for MiSeq read	Length
TwGuaN7	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT[length barcode]GTAGAAACTTGCGTGCCATATCCACGCGATTATAGGCNNNNNNNtctagagtcgcggccgcttt[ligand barcode]ATCTCGTATGCCGTCTTCTGCTTG	161/163

TwGua N6a	AATGATACGGCGACCACCGAGATCTACACTCTTTCCC TACACGACGCTCTTCCGATCT[length barcode]GTAGAAACTTGCGTGCCATATCCACGCGATTA TAGGCNNNNNNGtctagagtcgcggccgcttt[ligand barcode]ATCTCGTATGCCGTCTTCTGCTTG	161/ 163
PistGua N6	AATGATACGGCGACCACCGAGATCTACACTCTTTCCC TACACGACGCTCTTCCGATCT[length barcode]GTAGAAACTTGCGTGCCATATCCACGCGATTA TACCTGNNNNNNgagtcgcggccgctttactt[ligand barcode]ATCTCGTATGCCGTCTTCTGCTTG	165/ 167
PistTet N6	AATGATACGGCGACCACCGAGATCTACACTCTTTCCC TACACGACGCTCTTCCGATCT[length barcode]CTCTCCAGATCGAAATCTGGTATGTTTTCCTG NNNNNNgagtcgcggccgctttactt[ligand barcode]ATCTCGTATGCCGTCTTCTGCTTG	159/ 161

Color coding: yellow – P5 adapter, gray – R1seq, blue – aptamer, purple – communication module, dark blue – P7 adapter

Sequencing and data analysis

The libraries were sequenced on an Illumina MiSeq sequencer using MiSeq Reagent Kit v3 (150 cycles, single-end) at 12 pM loading concentration with 15% PhiX control by OIST DNA sequencing section (twister libraries) and by myself (CPP libraries). The reads were first sorted into cleaved and uncleaved pools by the 8-nt and 10-nt barcodes. Each of these pools was subsequently sorted into ligand+ and ligand- populations by the 6-nt barcode at the 3' end. After verifying and trimming the constant sequences, the remaining randomized fragments of reads were quality filtered to contain at least 80% of the base calls with Phred score over 30. Filtered reads were converted into fasta format and counted. Sequences with total read counts below 100 in any of the two conditions (lig- and lig+) were not included in the further analysis (except for N7 library where no filtering was run due to low reads). Read counts in each condition were used to calculate fraction cleaved and uncleaved of each variant according to the following formula:

$$Fc = \frac{reads\ cleaved}{reads\ cleaved + reads\ uncleaved}$$

Then fraction cleaved in presence of ligand (Fc+) was plotted versus fraction cleaved in absence of ligand (Fc-) in a scatter plot.

3.6. Supplementary information

Supplementary Table S1. MiSeq run statistics

Library	Library size	Total reads	Filtered and sorted	Quality filter
N7 twister-guanine	16384	16283727	12127029	Q30 90% (full read)
N6 twister-guanine	4096	16521075	11625676	Q30 80% (N6)
N6 cpp-guanine	4096	14728767	5172942	Q30 80% (N6)
N6 cpp-tetracycline	4096	18583555	11782082	Q30 80% (N6)

Supplementary Table S2. Tested candidates from N7 twister-guanine library

	Guanine -				Guanine +				
N7 Sequence	Full	Cleaved	Total	Fc-	Full	Cleaved	Total	Fc+	Fc-/Fc+
CGGCACA	6	88	94	0.94	100	16	116	0.14	6.79
TTGCAGC	30	216	246	0.88	187	38	225	0.17	5.20
CGGCCTT	21	353	374	0.94	262	63	325	0.19	4.87
CGGCTTA	70	215	285	0.75	304	58	362	0.16	4.71
TGGCTGA	29	329	358	0.92	285	78	363	0.21	4.28
TGGCCTT	38	584	622	0.94	467	133	600	0.22	4.24

Supplementary Table S3. Tested candidates from N6 twister-guanine library

		Guanine-				Guanine +				
	N6	Full	Cleaved	Total	Fc-	Full	Cleaved	Total	Fc+	Fc-/Fc+
#1	AGTCGA	337	436	773	0.564	377	35	412	0.085	6.64

#2	GGCCAG	250	413	663	0.623	405	54	459	0.118	5.29
#3	CGGCTT	935	1212	2147	0.565	1006	103	1109	0.093	6.08
#4	AGTCCA	581	213	794	0.268	308	36	344	0.105	2.56
#5	AGTTGC	89	1590	1679	0.947	1027	138	1165	0.118	7.99
#6	GCAGCT	76	1245	1321	0.942	784	111	895	0.124	7.60
#7	CGCACC	23	306	329	0.930	178	28	206	0.136	6.84
#8	ACGGCG	24	249	273	0.912	334	38	372	0.102	8.93
#9	CGGCCA	126	356	482	0.739	312	27	339	0.080	9.27
#10	TGGGCC	88	204	292	0.699	475	27	502	0.054	12.99
#11	TCTGCC	488	957	1445	0.662	744	75	819	0.092	7.23
#12	CGGCTC	351	642	993	0.647	514	52	566	0.092	7.04
#13	AGGCCA	312	412	724	0.569	312	22	334	0.066	8.64
#14	CGCCTA	957	104	1061	0.098	321	36	357	0.101	0.97
#15	GTTCTG	279 2	954	3746	0.255	1125	349	1474	0.237	1.08
#16	CTGCTC	111 6	573	1689	0.339	514	269	783	0.344	0.99
#17	CGACAG	188	201	389	0.517	104	107	211	0.507	1.02
#18	TGTATG	105 3	1696	2749	0.617	581	885	1466	0.604	1.02
#19	TTAAAG	360	916	1276	0.718	277	559	836	0.669	1.07
#20	GAATAA	93	561	654	0.858	62	389	451	0.863	0.99
#21	TCCCAC	19	592	611	0.969	12	403	415	0.971	1.00
#3 G2T	CTGCTT	137 2	1694	3066	0.553	896	688	1584	0.434	1.27
#3 C4G	CGGGTT	228	2253	2481	0.908	484	1136	1620	0.701	1.30
#3 T6A	CGGCTA	150 7	97	1604	0.060	406	40	446	0.090	0.67
#2 G1A	AGCCAG	579	114	693	0.165	243	40	283	0.141	1.16
#2 G2C	GCCCAG	39	432	471	0.917	123	211	334	0.632	1.45
#2 C4A	GGCAAG	987	297	1284	0.231	398	90	488	0.184	1.25
#11 G4T	TCTTCC	934	1122	2056	0.546	516	635	1151	0.552	0.99

#11 C5G	TCTGGC	174	1607	1781	0.902	175	1004	1179	0.852	1.06
#11 C6A	TCTGCA	135 3	620	1973	0.314	651	294	945	0.311	1.01

Supplementary Table S4. Tested candidates from N6 pistol-guanine library

		Guanine-				Guanine +				
Name	N6	Full	Cleaved	Total	Fc-	Full	Cleaved	Total	Fc+	Fc-/Fc+
#1	TTATGA	65	463	528	0.877	389	15	404	0.037	23.62
#2	ACTACT	45	357	402	0.888	320	15	335	0.045	19.83
#3	ACGAAA	37	251	288	0.872	202	10	212	0.047	18.48
#4	ACGTCA	53	344	397	0.866	240	12	252	0.048	18.20
#5	TCTGAA	50	248	298	0.832	254	13	267	0.049	17.09
#6	GCAACA	65	338	403	0.839	211	11	222	0.050	16.93
#7	TCGTCA	70	353	423	0.835	319	17	336	0.051	16.49
NS5	GTGAGG	257	193	450	0.429	230	253	483	0.524	0.82
NS6	TTTGGG	232	252	484	0.521	319	337	656	0.514	1.01
NS7	TGTAGC	187	283	470	0.602	216	408	624	0.654	0.92
NS8	CTTAGC	232	543	775	0.701	231	887	1118	0.793	0.88

4. Chapter 4. Rational design of aptazyme-based riboswitches in mammalian cells

4.1.Introduction

Chapter 3 introduced a new architecture for mammalian aptazymes, where aptamer and ribozyme are arranged in tandem. This architecture allows controlling conformation of the aptazyme between aptamer (*Ap*) and ribozyme (*Rz*) states by adjusting stability of the competing stems P_{ap} and P_{rz} (Figure 4-1). In our preliminary testing, 10 complementary nucleotides upstream were sufficient to form P_{ap}, disrupt *Rz* state and inhibit activity of both circularly permuted pistol and twister ribozymes.

We envisaged that among thousands of sequence variants of the competing stem P_{ap}, there would be some variants with optimal balance between the *Ap* and *Rz* states to allow switching in response to the aptamer ligand. Indeed, as shown in Chapter 3, libraries with this tandem design were enriched in on-switches for both twister and circularly permuted pistol ribozymes. However, since the high-throughput *in vitro* screening method did not correlate with the aptazyme activity in cells, we needed another systematic approach to identify optimal communication module sequence.

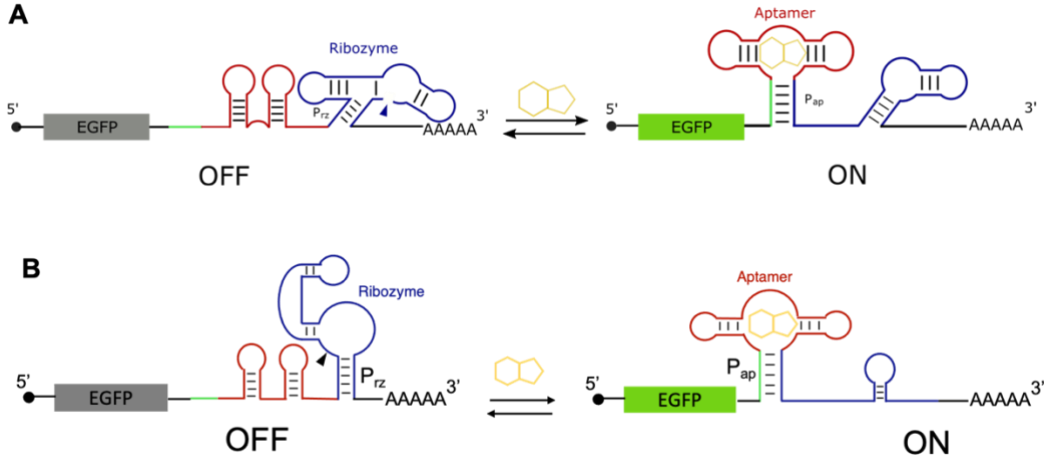


Figure 4-1. Tandem architecture for mammalian on-switches with a twister ribozyme (A) and a synthetic circularly permuted pistol ribozyme (CPP) (B). In the absence of ligand, the aptazyme preferentially forms the Rz structure, which results in ribozyme self-cleavage and subsequent mRNA degradation. Aptamer-ligand binding changes the aptazyme conformation to Apt, disrupting ribozyme folding and cleavage and therefore allowing protein translation.

As summarized in Chapter 1, all existing strategies for mammalian aptazyme on-switches are based on aptamer fusion in one of the ribozyme's stem-loops (Stifel et al., 2019) (Kennedy et al., 2014) (Beilstein et al., 2015) (Wei et al., 2013). In those designs, communication module is involved in an essential tertiary loop-loop interaction, and ligand binding is interfering with it, thereby rendering the ribozyme inactive. In contrast, in our current design ribozyme activity is expected to be controlled by hybridization with communication module, and hence, needs a different strategy for sequence optimization.

In this chapter, we suggest a rational design strategy to optimize the communication module (P_{ap}) to control relative stabilities of Apt and Rz states in the tandem aptamer-ribozyme architecture. We apply this approach to both twister and circular pistol ribozymes to yield on-switches in mammalian cells. Furthermore, we demonstrate the generality of the design

strategy for circularly permuted pistol (CPP) ribozyme-based switch by adjusting the response to two different ligands.

4.2.Results

4.2.1. Length of P_{ap} affects ribozyme activity

Chapter 3 established that 10 nt complementary nucleotides upstream are sufficient to inhibit the ribozyme activity. For our systematic optimization of the P_{ap} vs P_{rz} stability, we decided to explore inserts of different length to see their effect on the ribozyme activity. I made constructs with the insert length ranging from 0 to 14 complementary nucleotides and tested them in mammalian cells (Figure 4-2). I expected that with the shorter inserts the P_{ap} stem will not be stable enough to disrupt P_{rz}, and the ribozyme will be able to take its functional conformation. In contrast, with longer inserts, P_{ap} was expected to outcompete P_{rz}, thus disrupting ribozyme activity and restoring EGFP expression.

Experimental findings were in line with these expectations. From 0 to 4 nt insertions the ribozyme activity was close to control (Tw) both in the presence and absence of ligand, whereas with insertions over 10nt, ribozyme activity was inhibited resulting in high EGFP expression levels (Figure 4-2B). Interestingly, for the inserts between 5 and 9 nt the constructs were responsive to the ligand.

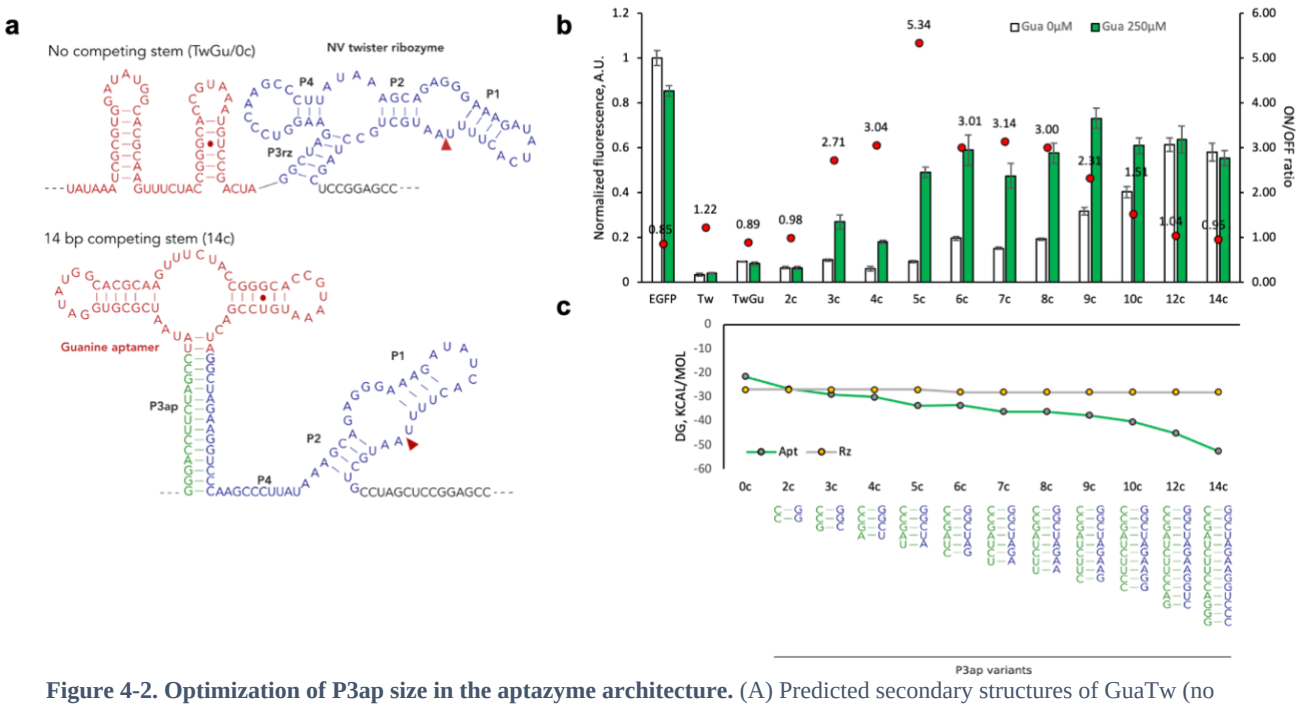


Figure 4-2. Optimization of P3ap size in the aptazyme architecture. (A) Predicted secondary structures of GuaTw (no upstream complementary bases) and 14c (14-base upstream complementary bases) aptazyme candidates. Red arrowhead indicates the cleavage site. (B) Normalized EGFP expression levels of the riboswitch variants in the absence and presence (250 µM) of guanine. EGFP: empty vector with no aptazyme. (C) Gibbs free energies (ΔG) of *Apt* and *Rz* structures calculated by Mfold. The structures were constrained to either Ap- or Rz-like folds by entering the constraint information provided in Table S1

I calculated folding energies of the two competing *Apt* and *Rz* structures with mfold software to estimate their relative stabilities (Figure 4-2C) (Zuker, 2003). The structures were forced into either Rz or Apt state by entering constraint information summarized in Table S1. For the construct with no complementary insert (0c) Gibbs free energy (ΔG) of the *Rz* state was higher than that of the *Apt* state. For 2c and 3c, it was equal or very similar, which was

reflected in the transfection result as well, where the ribozymes were mostly active. In the intermediate range, from 5 to 9 nt (5c-9c), where ΔG was between -30 and -40 kcal/mol, aptazymes were able to activate gene expression in cells in response to guanine. For the variants with low ribozyme activity both in the presence and absence of ligand (insert over 10nt, 10c-14c), we saw larger difference between the folding energies. These observations suggest that there is an optimal energy difference between the two states, where the construct can balance between the *Apt* and *Rz* states, triggered by the addition of ligand. Above that value the *Apt* state would become dominant, and below that value the construct would take *Rz* conformation.

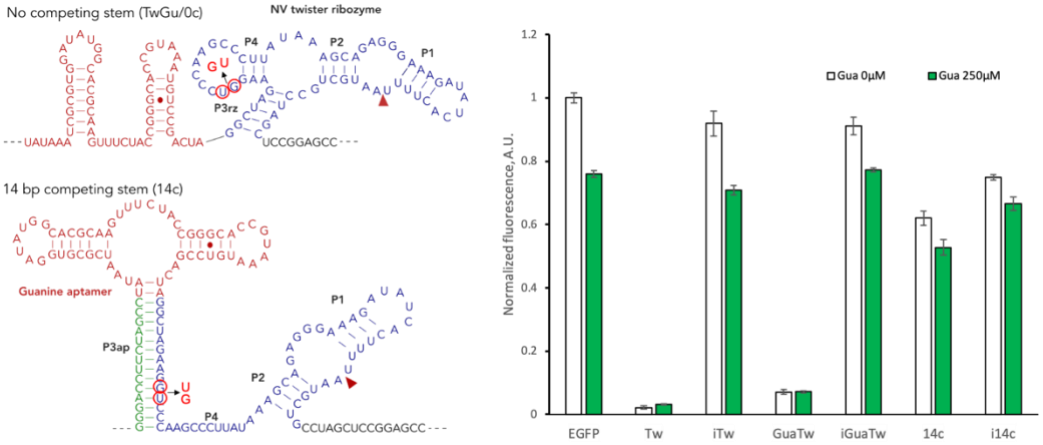


Figure 4-3. EGFP expression levels of inactivated twister ribozyme constructs. Normalized EGFP expression levels of inactivated ribozyme and aptazyme variants in the absence and presence (250 μM) of guanine. Tw – Nv twister ribozyme only, GuaTw – Nv twister ribozyme with guanine aptamer upstream (top structure), 14c – Nv twister ribozyme with guanine aptamer and 14bp complementary insert upstream (bottom structure), inactivated variants are marked with "i", EGFP - empty vector with no aptazyme. The ribozyme was inactivated by mutating two bases (Roth et al., 2014) circled in red in the schematic on the left.

Interestingly, although the level of gene expression seemed to plateau after stem length over 10, it never reached the level of the empty vector. I tested gene expression level for the same construct, but with a two-nucleotide inactivating mutation of the ribozyme (Figure 4-3). Transfection showed that there is a decrease in the gene expression level for the inactivated version of 14c (Figure 4-3, i14c) as well, only slightly different from the active 14c. This confirms that the decrease in the maximum level of expression is due to insertion of a structured RNA rather than residual aptazyme activity. Interestingly, the decrease was not as obvious for the inactivated versions of twister (iTw) and twister-guanine (iTwGua) constructs, potentially due to a lack of the stable 14 bp stem.

4.2.2. Fine-tuning the stability of P_{ap}

We then sought to improve the on-off ratio of the promising variants (5c, 6c and 7c) by further fine-tuning stability of the competing P_{ap} stem. I introduced single mismatch mutations in different positions of the stem to facilitate ribozyme folding in the absence of ligand and thereby decrease off-levels. As shown in Figure 4-4, this strategy helped to reduce off-levels and yielded some switches with on-off ratios over 6 (7c4x and 6c4x). Mutants of 5c variant had their on- and off-levels drop too much, especially for the mutations closer to the aptamer. This construct was probably more sensitive to a single nucleotide mismatch because of its reduced length.

The best switch that the strategy produced was variant 7c4x with an on-off ratio of 6.7. We further characterized it by recording its dose-dependent activation of gene expression (Figure 4-5).

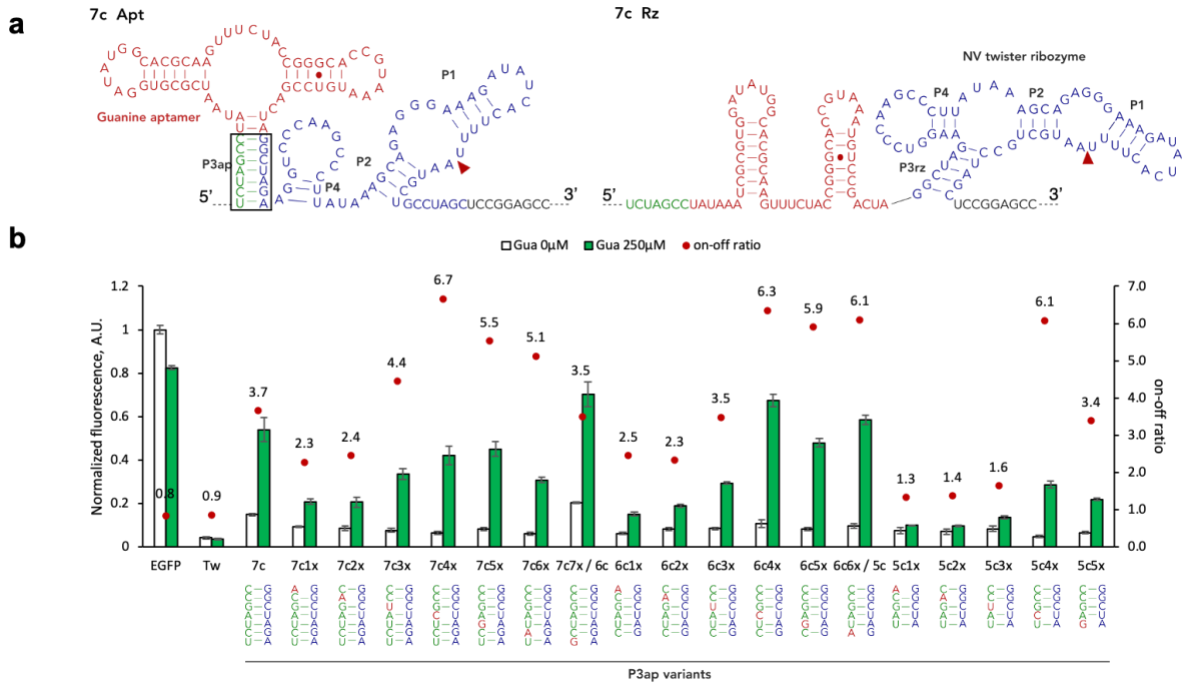


Figure 4-4. Optimization of P3ap stability by introduction of mismatches. (A) Representative Apt and Rz structures illustrated for the 7c variant. (B) Normalized EGFP expression levels of the riboswitch variants in the absence and presence (250 μM) of guanine. Aptazyme variants differ in $P3_{ap}$ stem composition. $P3_{ap}$ variants are shown in the bottom panel. Variants' labels indicate length of the stem and position of the mismatch. EGFP: empty vector with no aptazyme. Tw: ribozyme-only control.

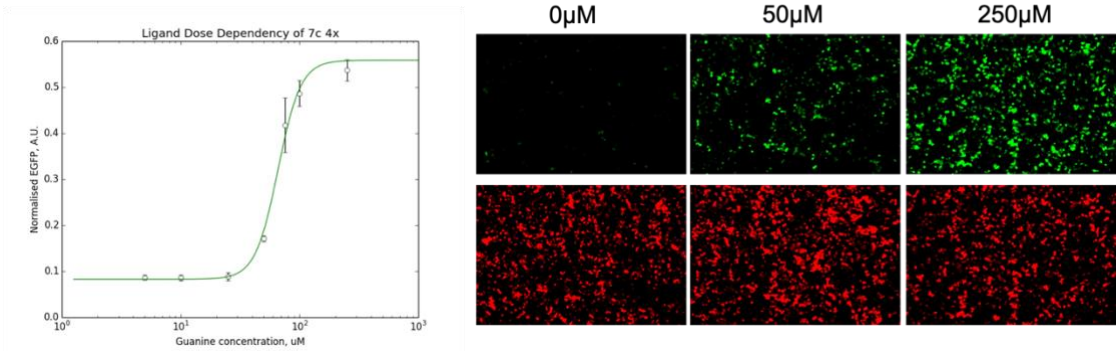


Figure 4-5. Dose-dependent EGFP expression of 7c4x. Left: The data of three replicate experiments with different guanine concentrations (0, 5, 10, 25, 50, 75, 100, 250 μM) are fitted to the Hill equation which is shown as the solid curve. Right: Fluorescence micrographs of the cells from the left figure.

It is noteworthy that while the position of mutation affected on-levels for all mutants, off-levels remained mostly the same. We interpreted it as an indication that mutation position has an effect on aptamer binding. To test this hypothesis a postdoc in our lab, Keisuke Fukunaga, checked binding affinity of the aptamers with 6 bp stem with mismatches in different positions, like in the previously tested aptazymes (Figure 4-6).

From his ITC measurements we found that while for the perfectly matched 6 bp-stem, the dissociation constant (K_d) was 19 nM, it increased to 39 nM when a mutation was added in

the distal position (6c5x). The dissociation constant increased even more with the mutation moving closer to the aptamer, eventually reaching as much as 377nM for the 6c1x variant. This confirmed that mutation closer to the aptamer is detrimental to aptamer binding, which leads to lower on-levels for such variants.

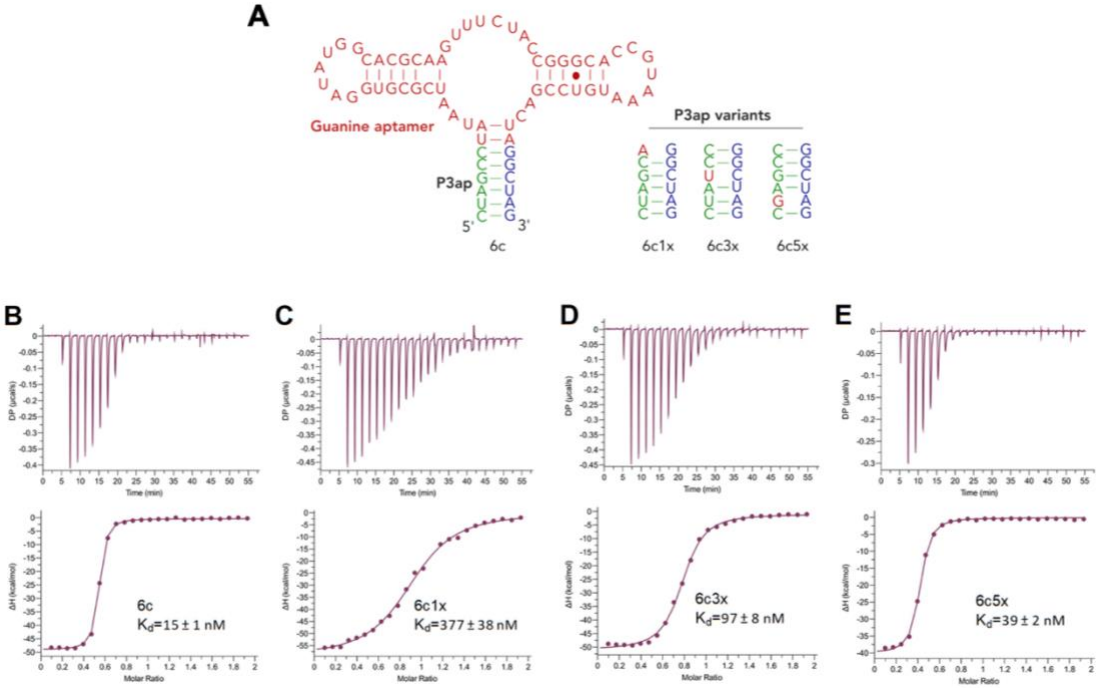


Figure 4-6. Effect of the mismatch position in the base stem of the guanine aptamer on Kd. (A) Stem variants of the aptamer studied by ITC. ITC results of the variants (B) 6c, (C) 6c1x, (D) 6c3x, and (E) 6c5x. Additional thermodynamic parameters obtained are provided in Table S2. The observations of binding stoichiometry (N) values lower than 1.0 have been reported for the guanine and other aptamers, probably due to the formation of nonfunctional complexes or misfolding (Chandra, Hannan, Xu, & Mandal, 2017) (Dwidar et al., 2019).

Thus, ITC measurements revealed the importance of the local stem stability for the aptamer binding, which affects the on-levels of the aptazyme. This finding can be used to fine-tune other aptazymes in future.

4.2.3. P_{ap} optimization for CPP ribozyme with guanine and tetracycline aptamers

We next tested if the same stem balance optimization is applicable to the circularly permuted pistol ribozyme as well. We repeated the same experiment with varying length of P_{ap} for the circular pistol ribozyme, but with Azami Green as a reporter protein (Figure 4-7). In the previous experiments with the twister ribozyme, we noticed that the addition of guanine led to ~20% decrease of EGFP fluorescence even in the empty (no-aptamer, no-ribozyme) vector. We speculated that guanine non-specifically interferes with EGFP expression, folding, or fluorescence. We confirmed that Azami Green, a green fluorescent protein derived from Galaxeidae coral (Karasawa, Araki, Yamamoto-Hino, & Miyawaki, 2003) does not exhibit such interference from guanine, therefore, decided to use it for the further experiments.
Similar to the twister ribozyme, 10 nt complementary insertion inhibited circularly permuted pistol. However, for this ribozyme even with the 2 bp insert, the newly introduced P_{ap} stem was allowing balance between the *Apt* and *Rz* states, allowing ligand-responsive inhibition of the ribozyme.

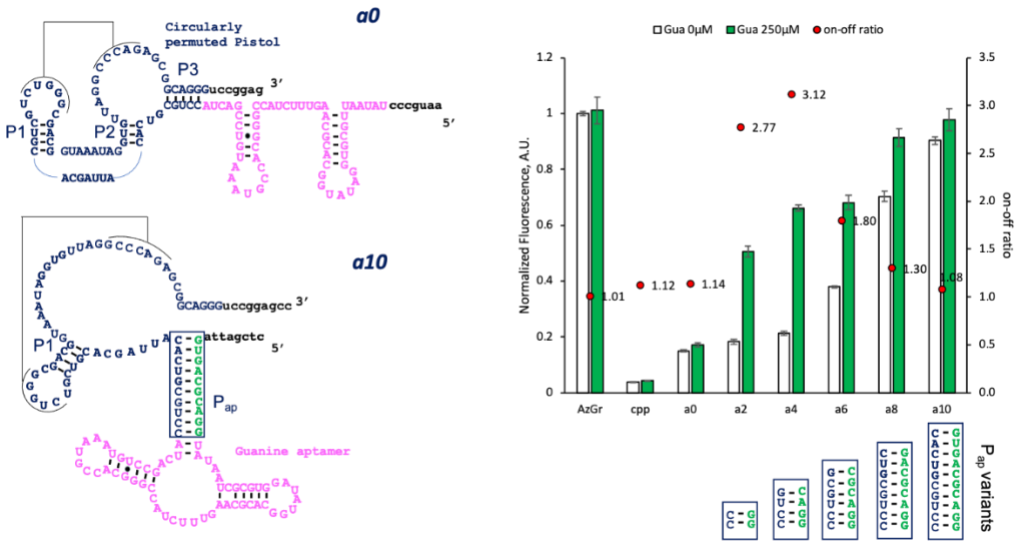


Figure 4-7. Effect of the Pap size on CPP-guanine aptazyme activity. Left: predicted secondary structures of a0 and a10 variants. Right: Normalized EGFP expression levels of the riboswitch variants with increasing P_{ap} stem length in the absence and presence (250 μM) of guanine. AzGr: empty vector with no aptazyme. The data are averages of 3 replicate samples and the error bars represent SD.

Although a2 and a4 had highest on-off ratios among the different length variants, I chose the a6 variant for further optimization. As was observed with twister's "5c", shorter stems are more sensitive to mutations and tend to retreat to *Rz* state easily, resulting in low gene expression levels. Thus, I introduced single-nucleotide mutations in the stem of the a6 variant (Figure 4-8). This resulted in improved on-off ratios, the highest being a6x3 with the dynamic range of 4.8. This finding was similar to the trend observed with twister and proved that tandem architecture can be used with different ribozymes as well.

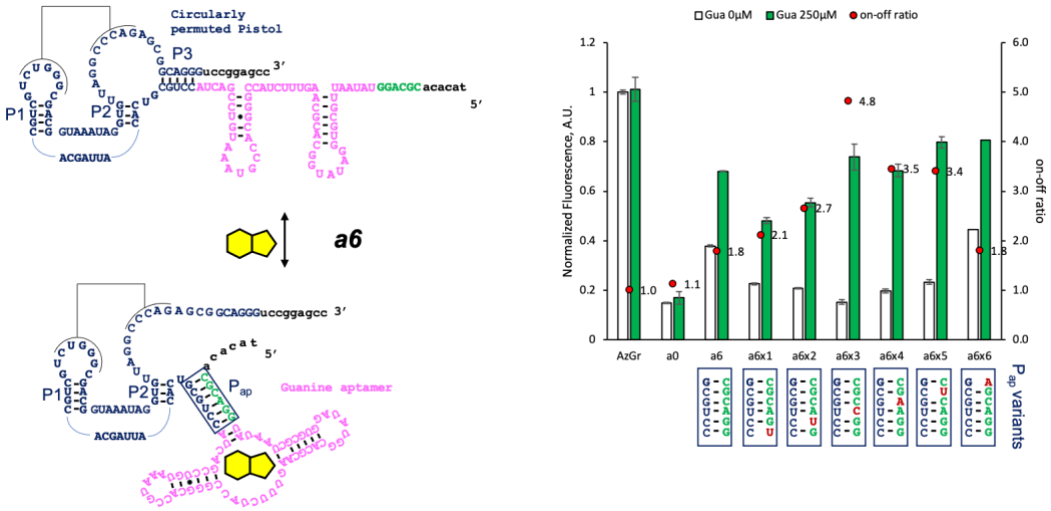


Figure 4-8. Optimization of P3ap stability in the CPP-guanine aptazyme architecture by introducing mismatches. Left: predicted secondary structures of a6 variant with (bottom) and without (top) the ligand. Right: normalized EGFP expression levels of the riboswitch variants in the absence and presence (250 μM) of guanine. Numbers after "x" in the names of variants indicate position of the mismatch, counting from the aptamer. EGFP: empty vector with no aptazyme. The data are averages of 3 replicate samples, and the error bars represent SD.

Having validated the strategy for different ribozyme scaffolds, I decided to test the strategy with different aptamers. I tested different length of the P_{ap} for the tetracycline aptamer (Figure 4-9). Although the trend with ribozyme inhibition was maintained, there was no ligand-dependent activation for any of the variants. This finding is similar to what we observed the library of tetracycline-CPP aptazymes Chapter 3 *in vitro*, where in spite of

covering a wide range of activities no variants showed tetracycline response. Again, as discussed in Chapter 3, the lack of tetracycline response could be due to some aptamer or ligand property or unexpected sequence interactions. Importantly, this phenomenon is consistent across *in vitro* and mammalian cell contexts.

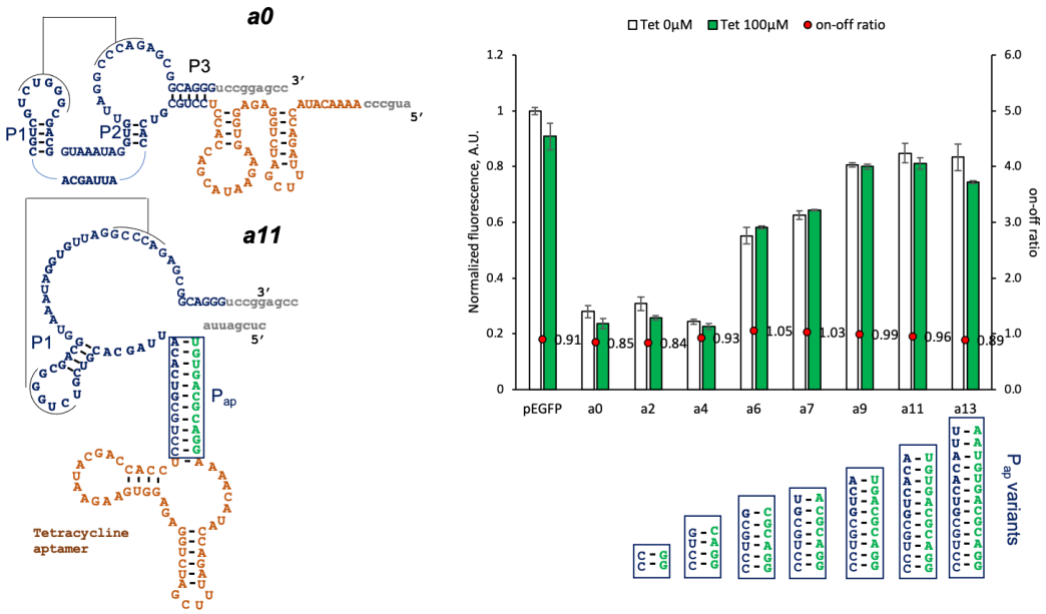


Figure 4-9. Effect of the Pap size on the CPP-tetracycline aptazyme activity. Left: predicted aptazyme structures of a0 and a11 variants. Right: Normalized EGFP expression levels of the riboswitch variants in the absence and presence (250 μM) of guanine. Numbers in variants' names indicate the length of the complementary insert. AzGr: empty vector with no aptazyme.

4.2.4. Competing stem strategy for other positions in CPP ribozyme

Since adjusting relative stabilities of P_{ap} and P_{rz} generated on-switches for tandemly arranged aptamer and ribozyme, we hypothesized that the same strategy can be applied for other positions in ribozyme. Specifically, we decided to insert the aptamer and the competing complementary sequence in the newly introduced junction of the CPP ribozyme. By introducing a circular permutation to the pistol ribozyme, we demonstrated that pistol ribozyme is tolerant to insertions in that junction. Hence, we decided to insert a guanine aptamer in the junction and introduce a sequence that hybridizes with the stems P2 and P3. In such design, the newly introduced P_{ap} is expected to compete with the stems P2 and P3 of the ribozyme (Figure 4-10). Indeed, the ribozyme tolerated insertion in the junction and remained active. Although we did not see complete inhibition of the ribozyme like we did with the tandem arrangement, we found several sequences with switching profile. Most notably, variant g6 was activating gene expression 8-fold in response to the ligand. Interestingly, for all the constructs on-levels were much lower than those of the tandemly arranged variants, suggesting that ribozyme inhibition was not as successful as with the upstream insert.

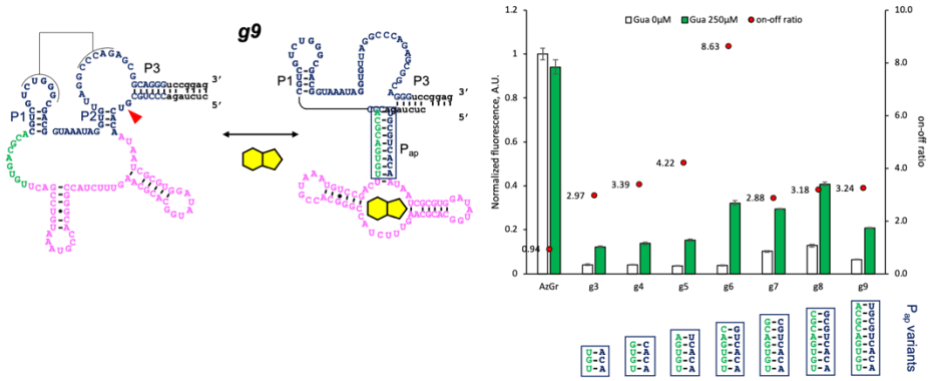


Figure 4-10. Effect of Pap competing with P2 stem on CPP-guanine aptazyme activity. Left: predicted secondary structure of the aptazyme with and without ligand. Right: Normalized EGFP expression levels of the riboswitch variants in the absence and presence (250 μM) of guanine. AzGr: empty vector with no aptazyme.

We then decided to target the other stem of the ribozyme, P1, from the junction insertion. For this design complementary sequence was inserted upstream of the aptamer and was expected to hybridize with P1 stem instead of P2. Previously, targeting P1 stem for the pistol ribozyme from *Alistipes putredinis* yielded guanine-responsive switches *in vitro*, but unfortunately that construct did not work in mammalian cells (Kobori et al., 2017).

The trend for P1 stem targeting was similar to that with tandem arrangement. While short complementary inserts were well tolerated by the ribozyme, longer stems (6-10) were inhibiting the ribozyme activity in a ligand-responsive manner. Although for the a10 variant gene expression leveled off between the "lig+" and "lig-" conditions, the ribozyme was not inhibited to the same extent as with the aptamer upstream.

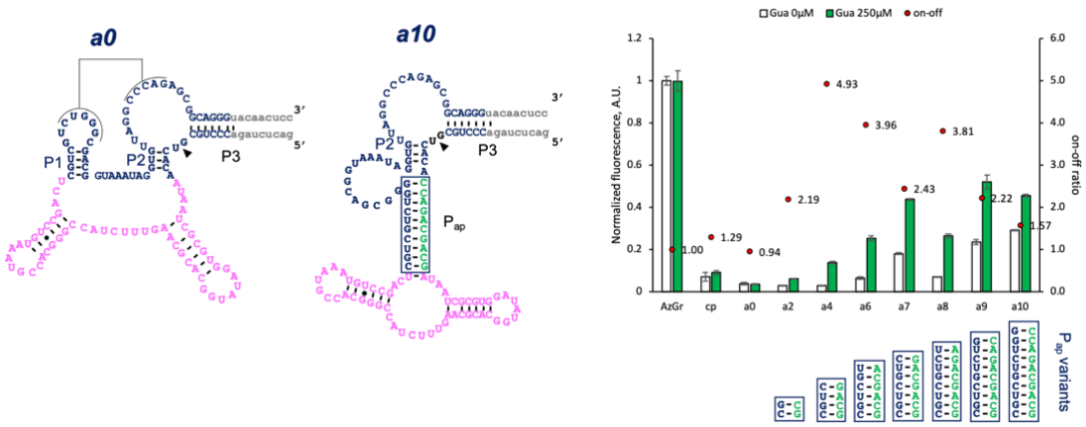


Figure 4-11. Effect of Pap competing with P1 stem on CPP-guanine aptazyme activity. Left: predicted secondary structure of the aptazyme variants a0 and a10 with and without ligand. Right: Normalized EGFP expression levels of the riboswitch variants in the absence and presence (250 μM) of guanine. AzGr: empty vector with no aptazyme.

Since some of the longer insert variants (a7, a9 and a10) had some extent of ligand activation, we decided to further engineer them to improve on-off ratios. Since formation of the P_{ap} was presumably leaving a large bulge from the leftover parts of P1 stem, we thought that the resulting steric hindrance can be reduced by a spacer upstream of the competing stem (Figure 4-12, left). We expected that this would give the ribozyme more flexibility to fold into the active conformation and thus decrease off-levels. Although this strategy did result in the decrease of the off-levels, it was accompanied by a similar decrease in the on-levels, not producing desired improvement of on-off ratios.

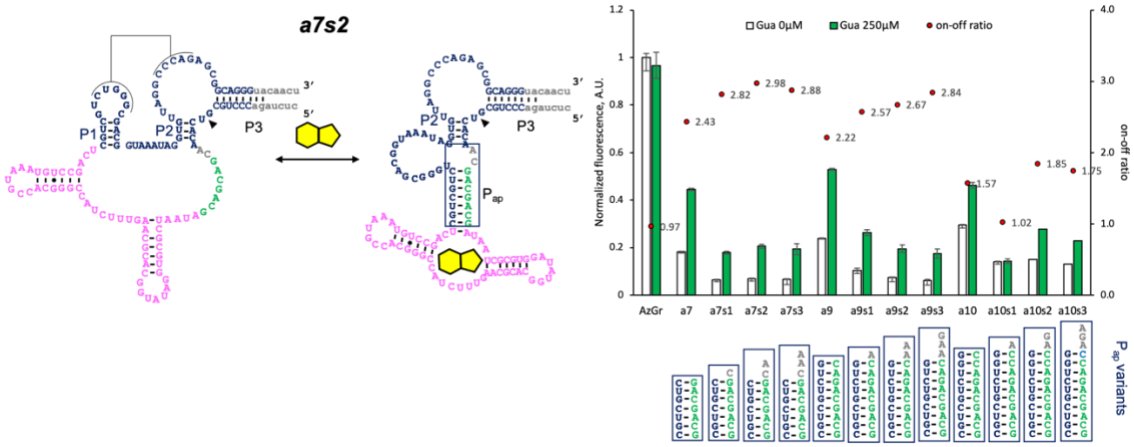


Figure 4-12. Optimization of Pap competing with P1 stem by introduction of spacers. Left: predicted secondary structure of a7s2 aptazyme variant with and without ligand. Right: Normalized EGFP expression levels of the riboswitch variants in the absence and presence (250 μM) of guanine. AzGr: empty vector with no aptazyme.

We then tested the same strategy with the tetracycline aptamer. In contrast to the tandem arrangement, tetracycline aptamer inserted in the junction resulted in good on-switches (Figure 4-13). The pattern for the longer stems was similar to that of the guanine aptamer, but interestingly, for shorter inserts the ribozyme remained highly active up until 6nt insert (a6). Then when the insert length reaches 7nt, gene expression suddenly increased to levels comparable to those of the guanine aptamer construct.

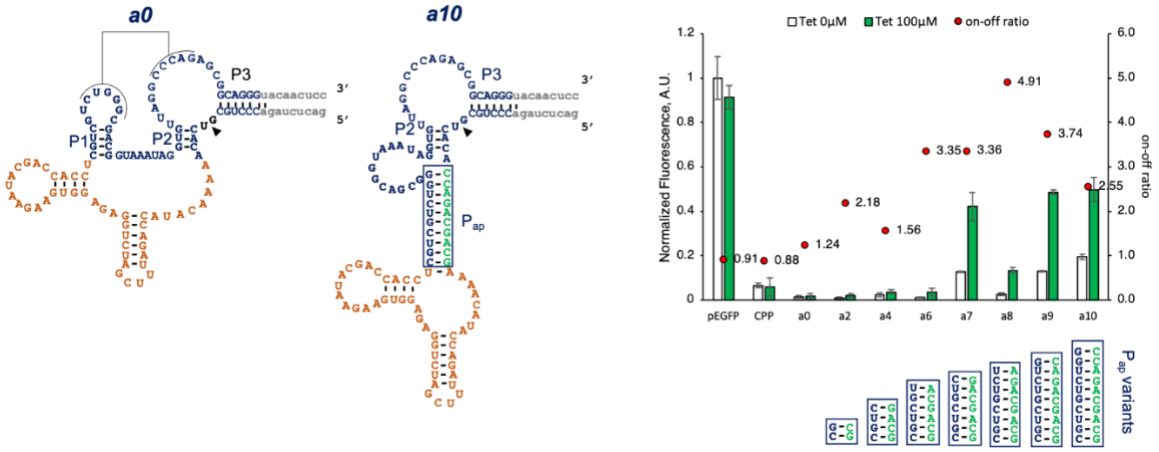


Figure 4-13. Effect of Pap competing with P1 stem on CPP-tetracycline aptazyme activity. Left: predicted secondary structure of the aptazyme variants a0 and a10 with and without ligand. Right: Normalized EGFP expression levels of the riboswitch variants in the absence and presence (100 μM) of tetracycline. EGFP: empty vector with no aptazyme.

Since variants from a7 to a10 followed the same pattern as the guanine constructs, we decided to try the same strategy with spacers to reduce off-levels (Figure 4-14). Same as for the previous construct, this reduced basal expression, but this time it also improved on-off ratios. For the a7, both on- and off- levels dropped significantly, probably due to the shorter stem, for which addition of the spacer made a more significant impact on ribozyme folding. For the a9 and a10, addition of spacers helped to reduce off-level, while maintaining decent on-levels, thus resulting in good on-off ratios, such ~5 for a9s1 and ~4 for a10s2.

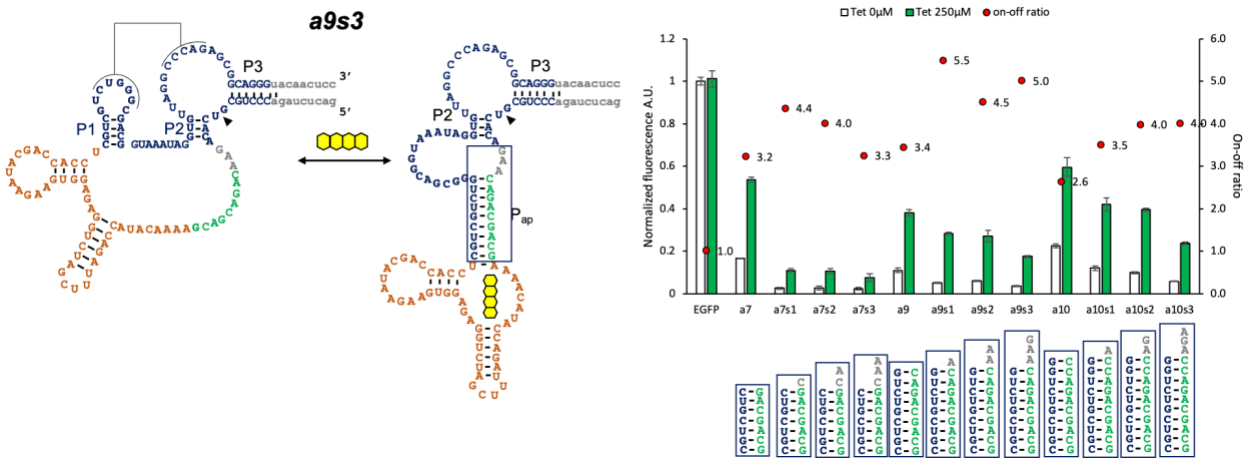


Figure 4-14. Optimization of Pap competing with P1 stem by introduction of spacers. Left: predicted secondary structure of a9s3 aptazyme variant with and without ligand. Right: Normalized EGFP expression levels of the riboswitch variants in the absence and presence (100 μM) of tetracycline. EGFP: empty vector with no aptazyme. The data are averages of 3 replicate samples and the error bars represent SD

4.3.Discussion

The aptazymes developed in this chapter are based on the two self-cleaving ribozyme classes that have not been used for mammalian cells before. Most of the existing aptazymes and all existing on-switches are based on the hammerhead ribozyme and follow similar design strategy. For example, on-switch developed by the Suess group with the highest on-off ratio up to date, was made by fusing tetracycline aptamer in the P1 stem of hammerhead (Beilstein et al., 2015). Hartig's group later engineered this switch more and changed the aptamer to guanine aptamer (Stifel et al., 2019). In 2015 Smolke's group published a theophylline-responsive hammerhead ribozyme, where the aptamer was fused in the P2 stem of the aptamer (Bell et al., 2015). In all these examples, an aptamer was fused into one of the stems of the hammerhead ribozyme. Close to finishing this thesis and after our twister paper (Mustafina, Fukunaga, & Yokobayashi, 2020) was published, Hartig's group made twister-based switches activated by tetracycline by high-throughput screening in mammalian cells (Strobel, Sporing, et al., 2020). Aside from that, the only other aptazyme scaffold used in mammalian cells has been the HDV ribozyme, developed by our group (Nomura et al., 2013). Hence, addition of twister and circularly permuted pistol ribozymes doubles the choice of available scaffolds for aptazyme engineering. Since currently aptazyme activity remains dependent on the environment and sequence context, having more ribozyme scaffolds increases the likelihood of finding the optimal one for the desired application.
In addition to new ribozyme scaffolds, this chapter also presents a new approach for aptazyme engineering. Tandem arrangement allows controlling the balance between the mutually exclusive aptamer and ribozyme states by adjusting the stability of the competing stem. In this work I adjusted stability of the stem varying its length and inserting mismatches at different positions; however, in future, stability can be controlled in other ways as well, such U-G wobble pairs and bulges.

Systematic introduction of mismatches in the stem at the base of the guanine aptamer (P_{ap}) and assessment of the aptamer-ligand binding revealed that mismatch mutation affects K_d of the aptamer. This is reflected in the on-levels of the aptazyme-controlled gene expression, that depend on P_{ap} stability and tend to increase with more distal mismatch positions. Similar observation is mentioned in Farzan's group's work, where they derived a scoring function to

optimize stability of the aptamer-ribozyme fusion stem (Zhong et al., 2016). Their score calculations included different weighing of base pair interactions depending on their positions relative to the aptamer. Despite the differences in the aptazyme design, both works emphasize the importance of local stability of the aptamer-ribozyme communication module. Since this approach does not rely on specific stem-loops in ribozymes, I was able to apply it for both twister and CPP ribozymes. Moreover, in CPP ribozyme the approach worked for three different positions. In addition to tandem arrangement, flexible circular junction of CPP allowed insertion of the aptamer with competing stem targeting two different sites of the ribozyme, P1 and P2 stems.

Unfortunately, I was not able to demonstrate the same modularity for the aptamer domain. Unlike guanine aptamer, tetracycline aptamer inserted upstream of the CPP ribozyme was not able to produce ligand-responsive activation. However, with optimized P_{ap} targeting P1 stem of the CPP ribozyme, tetracycline aptamer constructs had better on-off ratios than those with the guanine aptamer. Thus, although the strategy was successful for both aptamers in some positions, aptazyme performance varied depending on the aptamer. Since I only tested two aptamers in this study, there is not enough evidence to tell what parameter is important to ensure modularity of aptamer. Testing this strategy with more aptamers would provide more evidence to understand this phenomenon. One advantage of having tetracycline-responsive switches is that at the moment tetracycline is the only ligand that has been demonstrated to control aptazyme-dependent gene expression upon oral intake *in vivo* in mice (Reid et al., 2018). Furthermore, pharmacokinetic-pharmacodynamic relationships of the reporter protein and tetracycline were characterized in the study by the Hartig and the Kreuz groups (Strobel, Duchs, et al., 2020). Therefore, although similar validation experiments can be run with other ligands, for now tetracycline-responsive switches are the most established ones for *in vivo* applications.

We hope that this strategy will become a more systematic way for aptazyme engineering that would work for other scaffolds in future. Since there is a large space of unexplored sequences, such as combinations of spacers, mismatches, bulges and wobble basepairs, increasing throughput of tested variants will facilitate finding better optimal communications modules.

4.4.Conclusion and outlook

In this chapter I developed on-switches with two new ribozyme scaffolds, twister and circularly permuted pistol, for the first time in mammalian cells. For both of these scaffolds. I used the same strategy of introducing a sequence complementary to the ribozyme and controlling the stability of the resulting stem. Thus, the aptazyme is able to fold into one of the two mutually exclusive states, ribozyme (*Rz*) or aptamer (*Apt*), and the transition between these states is controlled by the addition of the aptamer ligand. I first applied this approach to the tandemly arranged aptamer and ribozyme, yielding guanine-responsive switches with on-off ratio of 6.7 for twister and 4.8 for circularly permuted pistol. This demonstrated that not only tandem architecture works for on-switches, but that it can also be applied to different ribozymes. However, I could not achieve the same modularity for the aptamers. When combined with the tetracycline aptamer in a tandem arrangement, CPP did not have the desired response.

I also expanded this approach to other positions of the CPP ribozyme, taking advantage of its flexible circular junction. By inserting an aptamer in the junction, I targeted two different

stems of the ribozyme and made on-switches for two different ligands, tetracycline and guanine. This shows that the new method is applicable to different positions in the ribozyme and different aptamers as well, although the aptamer performance seems to depend on the sequence context. Similar experiments with other aptamers and ribozymes may help to shed light on importance of the sequence context.

I showed that stability of the newly introduced stem can be controlled by systematic introduction of mismatches, and that position of the mismatch affects the on-level by affecting the binding affinity of the aptamer. While on-levels can be increased by the position of a mismatch, for designs with aptamer insertion within the ribozyme, off-levels can be decreased by introduction of spacers, as was demonstrated with CPP ribozyme.
Taken altogether, these findings will aid development and application of RNA devices in future. Combined with a higher-throughput method of assessing variants' activity in mammalian cells, these general principles can help to streamline aptazyme engineering.

4.5.Methods (Mustafina et al., 2020):

Riboswitch plasmid preparation. All plasmids were constructed by standard recombinant DNA techniques and sequence verified by Sanger sequencing. The EGFP-aptazyme constructs were cloned in pEGFP-N1-BspEI (Appendix, Figure S1). The inserted riboswitch sequences are listed in Table S2. The AG-aptazyme constructs were cloned in phm-AG1- (Appendix, Figure S2). Plasmids used for transfection were purified using Zyppy Plasmid Miniprep kit (Zymo Research).

HEK293 cell culture and transfection. HEK293 cells were cultured in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% heat-inactivated FBS (Gibco) containing 2 mM *L*-glutamine and 100 units/mL of penicillin-streptomycin (DMEM-FBS). Cells were kept in a 37°C incubator with 5% CO_2 and passaged regularly upon reaching 90% confluency. Approximately 20 h prior to transfection, the cells were trypsinized, diluted to ~2.7×10^5 cells/mL and 100 μL/well were seeded onto a 96-well plate. Cells in each well were cotransfected with 100 ng of the EGFP-aptazyme plasmid and 20 ng of pCMV-mCherry (transfection control) using 0.3 μL of *Trans*IT-293 Transfection Reagent (Mirus) according to the manufacturer's instructions. Five hours after transfection, the medium in each well was replaced with fresh medium with (250 μM) or without guanine. Guanine was dissolved at 25 mM in 0.2 M NaOH which was diluted by 100-fold in DMEM-FBS. The same volume of 0.2 M NaOH solution was diluted in DMEM-FBS for the medium without guanine. Forty-eight hours after transfection, the medium in each well was replaced with 100 μL of phosphate buffered saline (PBS), and fluorescence intensity was measured by Infinite M1000 PRO microplate reader (Tecan). Fluorescence intensity was measured at 484 nm excitation/510 nm emission/5 nm bandwidth for EGFP, and at 587 nm excitation/610 nm emission/10 nm bandwidth for mCherry. Background fluorescence measured using untransfected cells was subtracted from the EGFP and mCherry fluorescence values. Then, EGFP fluorescence was normalized by mCherry fluorescence to account for variations in transfection efficiency and cell counts. All reported values are averages of three replicate wells. Error bars on plots represent standard deviation among the three replicates.

RNA structure prediction and free energy calculation
For the free energy calculations (Figure 4-2C), aptazyme sequences were submitted to the Mfold Web Server's "RNA Folding Form"(Zuker, 2003) The structure was constrained to

either the *Apt* or *Rz* structure by forcing appropriate base pairs. The "constraint information" parameters used are listed in Table S1**.**

ITC measurements (Done by Keisuke Fukunaga). Guanine (Sigma-Aldrich) was dissolved in 1 N HCl/ethyl acetate solution and incubated at room temperature for few minutes. The solvent was removed by evaporation and the pellet was suspended in *n*-hexane. The solvent was removed again by evaporation to obtain guanine hydrochloride. Saturated solution of guanine hydrochloride was prepared in water and filtered through a 0.2 μm syringe filter (Merck-Millipore). The concentration was determined by absorbance using the reported molar extinction coefficient (ε_{248} = 11,400 cm^{-1} M^{-1}) (Windholz, 1976). For titration experiments, 70.4 μM guanine solution was prepared in ITC buffer (10 mM HEPES-KOH, pH 7.5, 140 mM KCl, 10 mM NaCl, 1 mM $MgCl_2$).

Template DNAs for *in vitro* transcription were prepared by primer extension using Q5 High-Fidelity DNA Polymerase (NEB). The resulting products were size-purified from agarose gel using a DNA purification kit (Zymoclean Gel DNA Recovery kit, Zymo Research). RNA aptamers were prepared using HiScribe T7 High Yield RNA Synthesis Kit (New England Biolabs) according to the manufacturer's instructions. The transcription products (100 μL) were treated with 2 U TURBO DNase (Thermo Fisher Scientific) at 37 °C for 1 h. The RNAs were precipitated with ammonium acetate and ethanol, and then dissolved in water. The solutions were further extracted with phenol-chloroform and ethanol precipitated. The RNAs were separated by denaturing polyacrylamide gel and extracted from the gel with TE buffer (10 mM Tris-HCl, pH 7.0, 0.1 mM EDTA-Na). The eluents were filtered through a 0.45 μm syringe filter. The gel-purified RNA solutions were concentrated, and the buffer was exchanged to water using an ultrafiltration device (Amicon Ultra 0.5 mL filter, 3 kDa cutoff, Merck-Millipore). RNA concentrations were determined by absorbance (A_{260}) according to OligoCalc.(Kibbe, 2007) For the titration experiments, 7.5 μM RNA solutions were prepared in the ITC buffer. RNAs were denatured at 65 °C for 3 min and then incubated at room temperature before measurement.

ITC experiments were performed as described previously with some modifications.(Dwidar et al., 2019) The titrations were performed at 37 °C using MicroCal PEAQ-ITC (Malvern). Injection parameters were as follows: initial 300 sec. delay, single 0.4 μL injection, and 24 serial injections of 1.5 μL at intervals of 120 sec. Stirring speed and reference power were set to 700 rpm and 4.5 μcal s^{-1}, respectively. Raw data were analyzed with MicroCal PEAQ-ITC analysis software ver. 1.0.0.1259 (Malvern) depending on one-site binding model. Background signals (guanine hydrochloride was titrated against the buffer) were subtracted from the data. Measurements were repeated twice to ensure reproducibility.

4.6.Supplementary information:

Table S1. Parameters used for Gibbs free energy calculations of *Rz* and *Apt* structures using Mfold. The parameters were entered in the "constraint information" box on the web form. Default values were used for the other parameters.

Variant / sequence	Rz	Apt
GuaTw/0c: UAUAAUCGCGUGGAUAUGGCACGCAAGUUUCUACCGGGCACCGUAA AUGUCCGACUAGGCUAGAAGGUCCCAAGCCCUUAUAAAGCAGAGGG AAAGAUAUCACUUUUAAUGCUGCCUAGC	F 59 120 5 F 64 79 3 F 84 113 4 F 93 106 4	F 1 57 2 F 8 24 5 F 35 53 6

2c: CCUAUAAUCGCGUGGAUAUGGCACGCAAGUUUCUACCGGGCACCGU AAAUGUCCGACUAGGCUAGAAGGUCCCAAGCCCUUAUAAAGCAGAG GGAAAGAUAUCACUUUUAAUGCUGCCUAGC	F 61 122 5 F 66 81 3 F 86 115 4 F 95 108 4	F 3 59 2 F 10 26 5 F 37 55 6
3c: GCCUAUAAUCGCGUGGAUAUGGCACGCAAGUUUCUACCGGGCACCG UAAAUGUCCGACUAGGCUAGAAGGUCCCAAGCCCUUAUAAAGCAGA GGGAAAGAUAUCACUUUUAAUGCUGCCUAGC	F 62 123 5 F 67 82 3 F 87 116 4 F 96 109 4	F 4 60 2 F 11 27 5 F 38 56 6
4c: AGCCUAUAAUCGCGUGGAUAUGGCACGCAAGUUUCUACCGGGCACC GUAAAUGUCCGACUAGGCUAGAAGGUCCCAAGCCCUUAUAAAGCAG AGGGAAAGAUAUCACUUUUAAUGCUGCCUAGC	F 63 124 5 F 68 83 3 F 88 117 4 F 97 110 4	F 5 61 2 F 12 28 5 F 39 57 6
5c: UAGCCUAUAAUCGCGUGGAUAUGGCACGCAAGUUUCUACCGGGCAC CGUAAAUGUCCGACUAGGCUAGAAGGUCCCAAGCCCUUAUAAAGCA GAGGGAAAGAUAUCACUUUUAAUGCUGCCUAGC	F 64 125 5 F 69 84 3 F 89 118 4 F 98 111 4	F 6 62 2 F 13 29 5 F 40 58 6
6c: CUAGCCUAUAAUCGCGUGGAUAUGGCACGCAAGUUUCUACCGGGCA CCGUAAAUGUCCGACUAGGCUAGAAGGUCCCAAGCCCUUAUAAAGC AGAGGGAAAGAUAUCACUUUUAAUGCUGCCUAGC	F 65 126 5 F 70 85 3 F 90 119 4 F 99 112 4	F 7 63 2 F 14 30 5 F 41 59 6
7c: UCUAGCCUAUAAUCGCGUGGAUAUGGCACGCAAGUUUCUACCGGGC ACCGUAAAUGUCCGACUAGGCUAGAAGGUCCCAAGCCCUUAUAAAG CAGAGGGAAAGAUAUCACUUUUAAUGCUGCCUAGC	F 66 127 5 F 71 86 3 F 91 120 4 F 100 113 4	F 8 64 2 F 15 31 5 F 42 60 6
8c: UUCUAGCCUAUAAUCGCGUGGAUAUGGCACGCAAGUUUCUACCGGG CACCGUAAAUGUCCGACUAGGCUAGAAGGUCCCAAGCCCUUAUAAA GCAGAGGGAAAGAUAUCACUUUUAAUGCUGCCUAGC	F 67 128 5 F 72 87 3 F 92 121 4 F 101 114 4	F 9 65 2 F 16 32 5 F 43 61 6
9c: CUUCUAGCCUAUAAUCGCGUGGAUAUGGCACGCAAGUUUCUACCGG GCACCGUAAAUGUCCGACUAGGCUAGAAGGUCCCAAGCCCUUAUAA AGCAGAGGGAAAGAUAUCACUUUUAAUGCUGCCUAGC	F 68 129 5 F 73 88 3 F 93 122 4 F 102 115 4	F 10 66 2 F 17 33 5 F 44 62 6
10c: CCUUCUAGCCUAUAAUCGCGUGGAUAUGGCACGCAAGUUUCUACCG GGCACCGUAAAUGUCCGACUAGGCUAGAAGGUCCCAAGCCCUUAUA AAGCAGAGGGAAAGAUAUCACUUUUAAUGCUGCCUAGC	F 69 130 5 F 74 89 3 F 94 123 4 F 103 116 4	F 11 67 2 F 18 34 5 F 45 63 6
12c: GACCUUCUAGCCUAUAAUCGCGUGGAUAUGGCACGCAAGUUUCUAC CGGGCACCGUAAAUGUCCGACUAGGCUAGAAGGUCCCAAGCCCUUA UAAAGCAGAGGGAAAGAUAUCACUUUUAAUGCUGCCUAGC	F 71 132 5 F 76 91 3 F 96 125 4 F 105 118 4	F 13 69 2 F 20 36 5 F 47 65 6
14c: GGGACCUUCUAGCCUAUAAUCGCGUGGAUAUGGCACGCAAGUUUCU ACCGGGCACCGUAAAUGUCCGACUAGGCUAGAAGGUCCCAAGCCCU UAUAAAGCAGAGGGAAAGAUAUCACUUUUAAUGCUGCCUAGC	F 73 134 5 F 78 93 3 F 98 127 4 F 107 120 4	F 15 71 2 F 22 38 5 F 49 67 6

Table S2. Summary of ITC measurements

RNA	*N*	K_d (nM)	*ΔH* (kcal/mol)	*ΔG* (kcal/mol)	*-TΔS* (kcal/mol)
6c	0.48	14 ± 1	-51.9 ± 0.2	-11.1	40.7
6c1x	0.86	354 ± 36	-63.7 ± 1.3	-9.2	54.6
6c3x	0.71	91 ± 9	-53.4 ± 0.6	-10.0	43.4

6c5x	0.36	37 ± 2	-41.8 ± 0.3	-10.6	31.3

Table S3. Sequences of the aptazyme variants for twister ribozyme. The aptazyme variants were inserted in the 3' UTR of the EGFP transcript between the XbaI (green) and BspEI (red) sites in pEGFP-N1-BspEI (**Figure S1** and **Appendix**). The guanine aptamer is marked in **bold**. The nucleotides involved in the $P3_{ap}$ formation are underlined.

Insert	Sequence
Tw	tctagaGGCTAGAAGGTCCCAAGCCCTTATAAAGCAGAGGGAAAGATATCACTTTTA ATGCTGCCTAGCtccgga
GuaTw/ 0c	tctaga**TATAATCGCGTGGATATGGCACGCAAGTTTCTACCGGGCACCGTAAATGTC CGACTA**GGCTAGAAGGTCCCAAGCCCTTATAAAGCAGAGGGAAAGATATCACTTTTA ATGCTGCCTAGCtccgga
2c	tctagaAACCGGCTCC**TATAATCGCGTGGATATGGCACGCAAGTTTCTACCGGGCAC CGTAAATGTCCGACTA**GGCTAGAAGGTCCCAAGCCCTTATAAAGCAGAGGGAAAGAT ATCACTTTTAATGCTGCCTAGCtccgga
3c	tctagaAACCGGCGCC**TATAATCGCGTGGATATGGCACGCAAGTTTCTACCGGGCAC CGTAAATGTCCGACTA**GGCTAGAAGGTCCCAAGCCCTTATAAAGCAGAGGGAAAGAT ATCACTTTTAATGCTGCCTAGCtccgga
4c	tctagaAACCGGAGCC**TATAATCGCGTGGATATGGCACGCAAGTTTCTACCGGGCAC CGTAAATGTCCGACTA**GGCTAGAAGGTCCCAAGCCCTTATAAAGCAGAGGGAAAGAT ATCACTTTTAATGCTGCCTAGCtccgga
5c	tctagaAACCATAGCC**TATAATCGCGTGGATATGGCACGCAAGTTTCTACCGGGCAC CGTAAATGTCCGACTA**GGCTAGAAGGTCCCAAGCCCTTATAAAGCAGAGGGAAAGAT ATCACTTTTAATGCTGCCTAGCtccgga
6c	tctagaAACCCTAGCC**TATAATCGCGTGGATATGGCACGCAAGTTTCTACCGGGCAC CGTAAATGTCCGACTA**GGCTAGAAGGTCCCAAGCCCTTATAAAGCAGAGGGAAAGAT ATCACTTTTAATGCTGCCTAGCtccgga
7c	tctagaAACTCTAGCC**TATAATCGCGTGGATATGGCACGCAAGTTTCTACCGGGCAC CGTAAATGTCCGACTA**GGCTAGAAGGTCCCAAGCCCTTATAAAGCAGAGGGAAAGAT ATCACTTTTAATGCTGCCTAGCtccgga
8c	tctagaAATTCTAGCC**TATAATCGCGTGGATATGGCACGCAAGTTTCTACCGGGCAC CGTAAATGTCCGACTA**GGCTAGAAGGTCCCAAGCCCTTATAAAGCAGAGGGAAAGAT ATCACTTTTAATGCTGCCTAGCtccgga
9c	tctagaACTTCTAGCC**TATAATCGCGTGGATATGGCACGCAAGTTTCTACCGGGCAC CGTAAATGTCCGACTA**GGCTAGAAGGTCCCAAGCCCTTATAAAGCAGAGGGAAAGAT ATCACTTTTAATGCTGCCTAGCtccgga
10c	tctagaTCCCTTCTAGCC**TATAATCGCGTGGATATGGCACGCAAGTTTCTACCGGGC ACCGTAAATGTCCGACTA**GGCTAGAAGGTCCCAAGCCCTTATAAAGCAGAGGGAAAG ATATCACTTTTAATGCTGCCTAGCtccgga
12c	tctagaGACCTTCTAGCC**TATAATCGCGTGGATATGGCACGCAAGTTTCTACCGGGC ACCGTAAATGTCCGACTA**GGCTAGAAGGTCCCAAGCCCTTATAAAGCAGAGGGAAAG ATATCACTTTTAATGCTGCCTAGCtccgga
14c	tctagaGGGACCTTCTAGCC**TATAATCGCGTGGATATGGCACGCAAGTTTCTACCGG GCACCGTAAATGTCCGACTA**GGCTAGAAGGTCCCAAGCCCTTATAAAGCAGAGGGAA AGATATCACTTTTAATGCTGCCTAGCtccgga
5c1x	tctagaAACCATAGCa**TATAATCGCGTGGATATGGCACGCAAGTTTCTACCGGGCAC CGTAAATGTCCGACTA**GGCTAGAAGGTCCCAAGCCCTTATAAAGCAGAGGGAAAGAT ATCACTTTTAATGCTGCCTAGCtccgga

5c2x	tctagaAACCATAGaC**TATAATCGCGTGGATATGGCACGCAAGTTTCTACCGGGCAC** **CGTAAATGTCCGACTA**GGCTAGAAGGTCCCAAGCCCTTATAAAGCAGAGGGAAAGAT ATCACTTTTAATGCTGCCTAGCtccgga
5c3x	tctagaAACCATAtCC**TATAATCGCGTGGATATGGCACGCAAGTTTCTACCGGGCAC** **CGTAAATGTCCGACTA**GGCTAGAAGGTCCCAAGCCCTTATAAAGCAGAGGGAAAGAT ATCACTTTTAATGCTGCCTAGCtccgga
5c4x	tctagaAACCATcGCC**TATAATCGCGTGGATATGGCACGCAAGTTTCTACCGGGCAC** **CGTAAATGTCCGACTA**GGCTAGAAGGTCCCAAGCCCTTATAAAGCAGAGGGAAAGAT ATCACTTTTAATGCTGCCTAGCtccgga
5c5x	tctagaAACCAgAGCC**TATAATCGCGTGGATATGGCACGCAAGTTTCTACCGGGCAC** **CGTAAATGTCCGACTA**GGCTAGAAGGTCCCAAGCCCTTATAAAGCAGAGGGAAAGAT ATCACTTTTAATGCTGCCTAGCtccgga
6c1x	tctagaAACCCTAGCa**TATAATCGCGTGGATATGGCACGCAAGTTTCTACCGGGCAC** **CGTAAATGTCCGACTA**GGCTAGAAGGTCCCAAGCCCTTATAAAGCAGAGGGAAAGAT ATCACTTTTAATGCTGCCTAGCtccgga
6c2x	tctagaAACCCTAGaC**TATAATCGCGTGGATATGGCACGCAAGTTTCTACCGGGCAC** **CGTAAATGTCCGACTA**GGCTAGAAGGTCCCAAGCCCTTATAAAGCAGAGGGAAAGAT ATCACTTTTAATGCTGCCTAGCtccgga
6c3x	tctagaAACCCTAtCC**TATAATCGCGTGGATATGGCACGCAAGTTTCTACCGGGCAC** **CGTAAATGTCCGACTA**GGCTAGAAGGTCCCAAGCCCTTATAAAGCAGAGGGAAAGAT ATCACTTTTAATGCTGCCTAGCtccgga
6c4x	tctagaAACCCTcGCC**TATAATCGCGTGGATATGGCACGCAAGTTTCTACCGGGCAC** **CGTAAATGTCCGACTA**GGCTAGAAGGTCCCAAGCCCTTATAAAGCAGAGGGAAAGAT ATCACTTTTAATGCTGCCTAGCtccgga
6c5x	tctagaAACCCgAGCC**TATAATCGCGTGGATATGGCACGCAAGTTTCTACCGGGCAC** **CGTAAATGTCCGACTA**GGCTAGAAGGTCCCAAGCCCTTATAAAGCAGAGGGAAAGAT ATCACTTTTAATGCTGCCTAGCtccgga
6c6x	tctagaAACCaTAGCC**TATAATCGCGTGGATATGGCACGCAAGTTTCTACCGGGCAC** **CGTAAATGTCCGACTA**GGCTAGAAGGTCCCAAGCCCTTATAAAGCAGAGGGAAAGAT ATCACTTTTAATGCTGCCTAGCtccgga
7c1x	tctagaAACTCTAGCa**TATAATCGCGTGGATATGGCACGCAAGTTTCTACCGGGCAC** **CGTAAATGTCCGACTA**GGCTAGAAGGTCCCAAGCCCTTATAAAGCAGAGGGAAAGAT ATCACTTTTAATGCTGCCTAGCtccgga
7c2x	tctagaAACTCTAGaC**TATAATCGCGTGGATATGGCACGCAAGTTTCTACCGGGCAC** **CGTAAATGTCCGACTA**GGCTAGAAGGTCCCAAGCCCTTATAAAGCAGAGGGAAAGAT ATCACTTTTAATGCTGCCTAGCtccgga
7c3x	tctagaAACTCTAtCC**TATAATCGCGTGGATATGGCACGCAAGTTTCTACCGGGCAC** **CGTAAATGTCCGACTA**GGCTAGAAGGTCCCAAGCCCTTATAAAGCAGAGGGAAAGAT ATCACTTTTAATGCTGCCTAGCtccgga
7c4x	tctagaAACTCTcGCC**TATAATCGCGTGGATATGGCACGCAAGTTTCTACCGGGCAC** **CGTAAATGTCCGACTA**GGCTAGAAGGTCCCAAGCCCTTATAAAGCAGAGGGAAAGAT ATCACTTTTAATGCTGCCTAGCtccgga
7c5x	tctagaAACTCgAGCC**TATAATCGCGTGGATATGGCACGCAAGTTTCTACCGGGCAC** **CGTAAATGTCCGACTA**GGCTAGAAGGTCCCAAGCCCTTATAAAGCAGAGGGAAAGAT ATCACTTTTAATGCTGCCTAGCtccgga
7c6x	tctagaAACTaTAGCC**TATAATCGCGTGGATATGGCACGCAAGTTTCTACCGGGCAC** **CGTAAATGTCCGACTA**GGCTAGAAGGTCCCAAGCCCTTATAAAGCAGAGGGAAAGAT ATCACTTTTAATGCTGCCTAGCtccgga
7c7x	tctagaAACcCTAGCC**TATAATCGCGTGGATATGGCACGCAAGTTTCTACCGGGCAC** **CGTAAATGTCCGACTA**GGCTAGAAGGTCCCAAGCCCTTATAAAGCAGAGGGAAAGAT ATCACTTTTAATGCTGCCTAGCtccgga

Table S4. Sequences of the aptazyme variants for pistol ribozyme. The aptazyme variants were inserted in the 3' UTR of the EGFP transcript between the XbaI (green) and MfeI (red) sites in pEGFP-N1 (**Figure S2** and **Appendix**). The tetracycline aptamer is marked in **bold**. The nucleotides involved in the $P3_{ap}$ formation are underlined. Spacers are marked in lowercase letters

Tet-P1-a0	tctagaCCCTGCGTCACA**AAAACATACCAGATTTCGATCTGGAGAGGTGAAGAATACGACCACCT**CGTCGTCTGGGCGACGGTAAATAGGTGTTAGGCCCAGAGCGGCAGGGTAcAACTccataccacatttgtagaggttttacttgctttaaaaaacctcccacacctccccctgaacctgaaacataaaatgaatgcaattg
Tet-P1-a2	tctagaCCCTGCGTCACACG**AAAACATACCAGATTTCGATCTGGAGAGGTGAAGAATACGACCACCT**CGTCGTCTGGGCGACGGTAAATAGGTGTTAGGCCCAGAGCGGCAGGGTAcAACTccataccacatttgtagaggttttacttgcttaaaaaacctcccacacctccccctgaacctgaaacataaaatgaatgcaattg
Tet-P1-a4	tctagaCCCTGCGTCACAGACG**AAAACATACCAGATTTCGATCTGGAGAGGTGAAGAATACGACCACCT**CGTCGTCTGGGCGACGGTAAATAGGTGTTAGGCCCAGAGCGGCAGGGTAcAACTccataccacatttgtagaggttttacttgctttaaaaaacctcccacacctccccctgaacctgaaacataaaatgaatgcaattg
Tet-P1-a6	tctagaCCCTGCGTCACAACGACG**AAAACATACCAGATTTCGATCTGGAGAGGTGAAGAATACGACCACCT**CGTCGTCTGGGCGACGGTAAATAGGTGTTAGGCCCAGAGCGGCAGGGTAcAACTccataccacatttgtagaggttttacttgctttaaaaaacctcccacacctccccctgaacctgaaacataaaatgaatgcaattg
Tet-P1-a7	tctagaCCCTGCGTCACAGACGACG**AAAACATACCAGATTTCGATCTGGAGAGGTGAAGAATACGACCACCT**CGTCGTCTGGGCGACGGTAAATAGGTGTTAGGCCCAGAGCGGCAGGGTAcAACTccataccacatttgtagaggttttacttgctttaaaaaacctcccacacctccccctgaacctgaaacataaaatgaatgcaattg
Tet-P1-a8	tctagaCCCTGCGTCACAAGACGACG**AAAACATACCAGATTTCGATCTGGAGAGGTGAAGAATACGACCACCT**CGTCGTCTGGGCGACGGTAAATAGGTGTTAGGCCCAGAGCGGCAGGGTAcAACTccataccacatttgtagaggttttacttgctttaaaaaacctcccacacctccccctgaacctgaaacataaaatgaatgcaattg
Tet-P1-a9	tctagaCCCTGCGTCACACAGACGACG**AAAACATACCAGATTTCGATCTGGAGAGGTGAAGAATACGACCACCT**CGTCGTCTGGGCGACGGTAAATAGGTGTTAGGCCCAGAGCGGCAGGGTAcAACTccataccacatttgtagaggttttacttgctttaaaaaacctcccacacctccccctgaacctgaaacataaaatgaatgcaattg
Tet-P1-a10	tctagaCCCTGCGTCACACCAGACGACG**AAAACATACCAGATTTCGATCTGGAGAGGTGAAGAATACGACCACCT**CGTCGTCTGGGCGACGGTAAATAGGTGTTAGGCCCAGAGCGGCAGGGTAcAACTccataccacatttgtagaggttttacttgctttaaaaaacctcccacacctccccctgaacctgaaacataaaatgaatgcaattg

Tet-P1-a7s1	tctagaCCCTGCGTCACAcGACGACG**AAAACATACCAGATTTCGATCTGGA** **GAGGTGAAGAATACGACCACCT**CGTCGTCTGGGCGACGGTAAATAGGTGTT AGGCCCAGAGCGGCAGGGTAcAACTccataccacatttgtagaggttttac ttgctttaaaaaacctcccacacctccccctgaacctgaaacataaaatga atgcaattg
Tet-P1-a7s2	tctagaCCCTGCGTCACAacGACGACG**AAAACATACCAGATTTCGATCTGG** **AGAGGTGAAGAATACGACCACCT**CGTCGTCTGGGCGACGGTAAATAGGTGT TAGGCCCAGAGCGGCAGGGTAcAACTccataccacatttgtagaggtttta cttgctttaaaaaacctcccaca cctccccctgaacctgaaacataaaatgaatgcaattg
Tet-P1-a7s3	tctagaCCCTGCGTCACAaacGACGACG**AAAACATACCAGATTTCGATCTG** **GAGAGGTGAAGAATACGACCACCT**CGTCGTCTGGGCGACGGTAAATAGGTG TTAGGCCCAGAGCGGCAGGGTAcAACTccataccacatttgtagaggtttt acttgctttaaaaaacctcccacacctccccctgaacctgaaacataaaat gaatgcaattg
Tet-P1-a9s1	tctagaCCCTGCGTCACAaCAGACGACG**AAAACATACCAGATTTCGATCTG** **GAGAGGTGAAGAATACGACCACCT**CGTCGTCTGGGCGACGGTAAATAGGTG TTAGGCCCAGAGCGGCAGGGTAcAACTccataccacatttgtagaggtttt acttgctttaaaaaacctcccacacctccccctgaacctgaaacataaaat gaatgcaattg
Tet-P1-a9s2	tctagaCCCTGCGTCACAaaCAGACGACG**AAAACATACCAGATTTCGATCT** **GGAGAGGTGAAGAATACGACCACCT**CGTCGTCTGGGCGACGGTAAATAGGT GTTAGGCCCAGAGCGGCAGGGTAcAACTccataccacatttgtagaggttt tacttgctttaaaaaacctcccacacctccccctgaacctgaaacataaaa tgaatgcaattg
Tet-P1-a9s3	tctagaCCCTGCGTCACAtaaCAGACGACG**AAAACATACCAGATTTCGATC** **TGGAGAGGTGAAGAATACGACCACCT**CGTCGTCTGGGCGACGGTAAATAGG TGTTAGGCCCAGAGCGGCAGGGTAcAACTccataccacatttgtagaggtt ttacttgctttaaaaaacctcccacacctccccctgaacctgaaacataaa atgaatgcaattg
Tet-P1-a10s1	tctagaCCCTGCGTCACAaCCAGACGACG**AAAACATACCAGATTTCGATCT** **GGAGAGGTGAAGAATACGACCACCT**CGTCGTCTGGGCGACGGTAAATAGGT GTTAGGCCCAGAGCGGCAGGGTAcAACTccataccacatttgtagaggttt tacttgctttaaaaaacctcccacacctccccctgaacctgaaacataaaa tgaatgcaattg
Tet-P1-a10s2	tctagaCCCTGCGTCACAaaCCAGACGACG**AAAACATACCAGATTTCGATC** **TGGAGAGGTGAAGAATACGACCACCT**CGTCGTCTGGGCGACGGTAAATAGG TGTTAGGCCCAGAGCGGCAGGGTAcAACTccataccacatttgtagaggtt ttacttgctttaaaaaacctcccacacctccccctgaacctgaaacataaa atgaatgcaattg
Tet-P1-a10s3	tctagaCCCTGCGTCACAaaaCCAGACGACG**AAAACATACCAGATTTCGAT** **CTGGAGAGGTGAAGAATACGACCACCT**CGTCGTCTGGGCGACGGTAAATAG GTGTTAGGCCCAGAGCGGCAGGGTAcAACTccataccacatttgtagaggt tttacttgctttaaaaaacctcccacacctccccctgaacctgaaacataa aatgaatgcaattg

Table S5. Sequences of the aptazyme variants for pistol ribozyme. The aptazyme variants were inserted in the 3' UTR of the AzamiGreen transcript between the NotI (green) and MfeI (red) restriction sites in phmAG1-MC1 (**Figure S2** and **Appendix**). The guanine aptamer is

marked in **bold**. The nucleotides involved in the P_{ap} formation are underlined. Spacers and mismatches are marked in lowercase letters

Gua-a0	gcggccgcgactcgattacacaatgccc**TATAATCGCGTGGATATGGCACGCAA** **GTTTCTACCGGGCACCGTAAATGTCCGACTA**CCTGCGTCACAttagcaCGTCGT CTGGGCGACGGTAAATAGGTGTTAGGCCCAGAGCGGCAGGGtccggagccatac cacatttgtagaggttttacttgctttaaaaaacctcccacacctcccccctgaa cctgaaacataaaatgaatgcaattg
Gua-a2	gcggccgcgactcgattacacaatgcGG**TATAATCGCGTGGATATGGCACGCAA** **GTTTCTACCGGGCACCGTAAATGTCCGACTA**CCTGCGTCACAttagcaCGTCGT CTGGGCGACGGTAAATAGGTGTTAGGCCCAGAGCGGCAGGGtccggagccatac cacatttgtagaggttttacttgctttaaaaaacctcccacacctcccccctgaa cctgaaacataaaatgaatgcaattg
Gua-a4	gcggccgcgactcgattacacaatCAGG**TATAATCGCGTGGATATGGCACGCAA** **GTTTCTACCGGGCACCGTAAATGTCCGACTA**CCTGCGTCACAttagcaCGTCGT CTGGGCGACGGTAAATAGGTGTTAGGCCCAGAGCGGCAGGGtccggagccatac cacatttgtagaggttttacttgctttaaaaaacctcccacacctcccccctgaa cctgaaacataaaatgaatgcaattg
Gua-a6	gcggccgcgactcgattacacaCGCAGG**TATAATCGCGTGGATATGGCACGCAA** **GTTTCTACCGGGCACCGTAAATGTCCGACTA**CCTGCGTCACAttagcaCGTCGT CTGGGCGACGGTAAATAGGTGTTAGGCCCAGAGCGGCAGGGtccggagccatac cacatttgtagaggttttacttgctttaaaaaacctcccacacctcccccctgaa cctgaaacataaaatgaatgcaattg
Gua-a8	gcggccgcgactcgattacacacGCAGG**TATAATCGCGTGGATATGGCACGCAA** **GTTTCTACCGGGCACCGTAAATGTCCGACTA**CCTGCGTCACAttagcaCGTCGT CTGGGCGACGGTAAATAGGTGTTAGGCCCAGAGCGGCAGGGtccggagccatac cacatttgtagaggttttacttgctttaaaaaacctcccacacctcccccctgaa cctgaaacataaaatgaatgcaattg
Gua-a10	gcggccgcgactcgattaGTGACGCAGG**TATAATCGCGTGGATATGGCACGCAA** **GTTTCTACCGGGCACCGTAAATGTCCGACTA**CCTGCGTCACAttagcaCGTCGT CTGGGCGACGGTAAATAGGTGTTAGGCCCAGAGCGGCAGGGtccggagccatac cacatttgtagaggttttacttgctttaaaaaacctcccacacctcccccctgaa cctgaaacataaaatgaatgcaattg
Gua-a6x1	gcggccgcgactcgattacacaCGCAGt**TATAATCGCGTGGATATGGCACGCAA** **GTTTCTACCGGGCACCGTAAATGTCCGACTA**CCTGCGTCACAttagcaCGTCGT CTGGGCGACGGTAAATAGGTGTTAGGCCCAGAGCGGCAGGGtccggagccatac cacatttgtagaggttttacttgctttaaaaaacctcccacacctcccccctgaa cctgaaacataaaatgaatgcaattg
Gua-a6x2	gcggccgcgactcgattacacaCGCAtG**TATAATCGCGTGGATATGGCACGCAA** **GTTTCTACCGGGCACCGTAAATGTCCGACTA**CCTGCGTCACAttagcaCGTCGT CTGGGCGACGGTAAATAGGTGTTAGGCCCAGAGCGGCAGGGtccggagccatac

	cacatttgtagaggttttacttgctttaaaaaacctcccacacctccccctgaa cctgaaacataaaatgaatgcaattg
Gua- a6x3	gcggccgcgactcgattacacaCGCcGG**TATAATCGCGTGGATATGGCACGCAA** **GTTTCTACCGGGCACCGTAAATGTCCGACTA**CCTGCGTCACAttagcaCGTCGT CTGGGCGACGGTAAATAGGTGTTAGGCCCAGAGCGGCAGGGtccggagccatac cacatttgtagaggttttacttgctttaaaaaacctcccacacctccccctgaa cctgaaacataaaatgaatgcaattg
Gua- a6x4	gcggccgcgactcgattacacaCGaAGGT**ATAATCGCGTGGATATGGCACGCAA** **GTTTCTACCGGGCACCGTAAATGTCCGACTA**CCTGCGTCACAttagcaCGTCGT CTGGGCGACGGTAAATAGGTGTTAGGCCCAGAGCGGCAGGGtccggagccatac cacatttgtagaggttttacttgctttaaaaaacctcccacacctccccctgaa cctgaaacataaaatgaatgcaattg
Gua- a6x5	gcggccgcgactcgattacacaCtCAGG**TATAATCGCGTGGATATGGCACGCAA** **GTTTCTACCGGGCACCGTAAATGTCCGACTA**CCTGCGTCACAttagcaCGTCGT CTGGGCGACGGTAAATAGGTGTTAGGCCCAGAGCGGCAGGGtccggagccatac cacatttgtagaggttttacttgctttaaaaaacctcccacacctccccctgaa cctgaaacataaaatgaatgcaattg
Gua- a6x6	gcggccgcgactcgattacacacGCAGG**TATAATCGCGTGGATATGGCACGCAA** **GTTTCTACCGGGCACCGTAAATGTCCGACTA**CCTGCGTCACAttagcaCGTCGT CTGGGCGACGGTAAATAGGTGTTAGGCCCAGAGCGGCAGGGtccggagccatac cacatttgtagaggttttacttgctttaaaaaacctcccacacctccccctgaa cctgaaacataaaatgaatgcaattg
G3	gcggccgcgactcgattacacaatgccctctagaCCCTGCGTCACA**TATAATCG** **CGTGGATATGGCACGCAAGTTTCTACCGGGCACCGTAAATGTCCGACTA**TGTCG TCGTCTGGGCGACGGTAAATAGGTGTTAGGCCCAGAGCGGCAGGGtccggagcc ataccacatttgtagaggttttacttgctttaaaaaacctcccacacctccccc tgaacctgaaacataaaatgaatgcaattg
G4	gcggccgcgactcgattacacaatgccctctagaCCCTGCGTCACA**TATAATCG** **CGTGGATATGGCACGCAAGTTTCTACCGGGCACCGTAAATGTCCGACTA**TGTGC GTCGTCTGGGCGACGGTAAATAGGTGTTAGGCCCAGAGCGGCAGGGtccggagc cataccacatttgtagaggttttacttgctttaaaaaacctcccacacctcccc ctgaacctgaaacataaaatgaatgcaattg
G5	gcggccgcgactcgattacacaatgccctctagaCCCTGCGTCACA**TATAATCG** **CGTGGATATGGCACGCAAGTTTCTACCGGGCACCGTAAATGTCCGACTA**TGTGA CGTCGTCTGGGCGACGGTAAATAGGTGTTAGGCCCAGAGCGGCAGGGtccggag ccataccacatttgtagaggttttacttgctttaaaaaacctcccacacctccc cctgaacctgaaacataaaatgaatgcaattg

G6	gcggccgcgactcgattacacaatgccctctagaCCCTGCGTCACA**TATAATCG CGTGGATATGGCACGCAAGTTTCTACCGGGCACCGTAAATGTCCGACTA**TGTGA CCGTCGTCTGGGCGACGGTAAATAGGTGTTAGGCCCAGAGCGGCAGGGtccgga gccataccacatttgtagaggttttacttgctttaaaaaacctcccacacctcc ccctgaacctgaaacataaaatgaatgcaattg
G7	gcggccgcgactcgattacacaatgccctctagaCCCTGCGTCACA**TATAATCG CGTGGATATGGCACGCAAGTTTCTACCGGGCACCGTAAATGTCCGACTA**TGTGA CGCGTCGTCTGGGCGACGGTAAATAGGTGTTAGGCCCAGAGCGGCAGGGtccgg agccataccacatttgtagaggttttacttgctttaaaaaacctcccacacctc ccccctgaacctgaaacataaaatgaatgcaattg
G8	gcggccgcgactcgattacacaatgccctctagaCCCTGCGTCACA**TATAATCG CGTGGATATGGCACGCAAGTTTCTACCGGGCACCGTAAATGTCCGACTA**TGTGA CGCCGTCGTCTGGGCGACGGTAAATAGGTGTTAGGCCCAGAGCGGCAGGGtccg gagccataccacatttgtagaggttttacttgctttaaaaaacctcccacacct cccccctgaacctgaaacataaaatgaatgcaattg
G9	gcggccgcgactcgattacacaatgccctctagaCCCTGCGTCACA**TATAATCG CGTGGATATGGCACGCAAGTTTCTACCGGGCACCGTAAATGTCCGACTA**TGTGA CGCACGTCGTCTGGGCGACGGTAAATAGGTGTTAGGCCCAGAGCGGCAGGGtcc ggagccataccacatttgtagaggttttacttgctttaaaaaacctcccacacc tcccccctgaacctgaaacataaaatgaatgcaattg
Gua-P1-a7	gcggccgcgactcgattacacaatgccctctagaCCCTGCGTCACAGACGACG**T ATAATCGCGTGGATATGGCACGCAAGTTTCTACCGGGCACCGTAAATGTCCGAC TA**CGTCGTCTGGGCGACGGTAAATAGGTGTTAGGCCCAGAGCGGCAGGGTAcAA CTccataccacatttgtagaggttttacttgctttaaaaaacctcccacacctc ccccctgaacctgaaacataaaatgaatgcaattg
Gua-P1-a8	gcggccgcgactcgattacacaatgccctctagaCCCTGCGTCACAAGACGACG **TATAATCGCGTGGATATGGCACGCAAGTTTCTACCGGGCACCGTAAATGTCCGA CTA**CGTCGTCTGGGCGACGGTAAATAGGTGTTAGGCCCAGAGCGGCAGGGTAcA ACTccataccacatttgtagaggttttacttgctttaaaaaacctcccacacct ccccctgaacctgaaacataaaatgaatgcaattg
Gua-P1-a9	gcggccgcgactcgattacacaatgccctctagaCCCTGCGTCACACAGACGAC G**TATAATCGCGTGGATATGGCACGCAAGTTTCTACCGGGCACCGTAAATGTCCG ACTA**CGTCGTCTGGGCGACGGTAAATAGGTGTTAGGCCCAGAGCGGCAGGGTAc AACTccataccacatttgtagaggttttacttgctttaaaaaacctcccacacc tcccccctgaacctgaaacataaaatgaatgcaattg
Gua-P1-a10	gcggccgcgactcgattacacaatgccctctagaCCCTGCGTCACACCAGACGA CG**TATAATCGCGTGGATATGGCACGCAAGTTTCTACCGGGCACCGTAAATGTCC GACTA**CGTCGTCTGGGCGACGGTAAATAGGTGTTAGGCCCAGAGCGGCAGGGTA cAACTccataccacatttgtagaggttttacttgctttaaaaaacctcccacac ctcccccctgaacctgaaacataaaatgaatgcaattg

Gua-P1-a7s1	gcggccgcgactcgattacacaatgccctctagaCCCTGCGTCACACGACGACG **TATAATCGCGTGGATATGGCACGCAAGTTTCTACCGGGCACCGTAAATGTCCGA CTA**CGTCGTCTGGGCGACGGTAAATAGGTGTTAGGCCCAGAGCGGCAGGGTAcA ACTccataccacatttgtagaggttttacttgctttaaaaaacctcccacacct ccccctgaacctgaaacataaaatgaatgcaattg
Gua-P1-a7s2	gcggccgcgactcgattacacaatgccctctagaCCCTGCGTCACAACGACGACC G**TATAATCGCGTGGATATGGCACGCAAGTTTCTACCGGGCACCGTAAATGTCCG ACTA**CGTCGTCTGGGCGACGGTAAATAGGTGTTAGGCCCAGAGCGGCAGGGTAc AACTccataccacatttgtagaggttttacttgctttaaaaaacctcccacacc tcccccctgaacctgaaacataaaatgaatgcaattg
Gua-P1-a7s3	gcggccgcgactcgattacacaatgccctctagaCCCTGCGTCACAAACGACGA CG**TATAATCGCGTGGATATGGCACGCAAGTTTCTACCGGGCACCGTAAATGTCC GACTA**CGTCGTCTGGGCGACGGTAAATAGGTGTTAGGCCCAGAGCGGCAGGGTA cAACTccataccacatttgtagaggttttacttgctttaaaaaacctcccacac ctcccccctgaacctgaaacataaaatgaatgcaattg
Gua-P1-a9s1	gcggccgcgactcgattacacaatgccctctagaCCCTGCGTCACAACAGACGA CG**TATAATCGCGTGGATATGGCACGCAAGTTTCTACCGGGCACCGTAAATGTCC GACTA**CGTCGTCTGGGCGACGGTAAATAGGTGTTAGGCCCAGAGCGGCAGGGTA cAACTccataccacatttgtagaggttttacttgctttaaaaaacctcccacac ctcccccctgaacctgaaacataaaatgaatgcaattg
Gua-P1-a9s2	gcggccgcgactcgattacacaatgccctctagaCCCTGCGTCACAAACAGACG ACG**TATAATCGCGTGGATATGGCACGCAAGTTTCTACCGGGCACCGTAAATGTC CGACTA**CGTCGTCTGGGCGACGGTAAATAGGTGTTAGGCCCAGAGCGGCAGGGT AcAACTccataccacatttgtagaggttttacttgctttaaaaaacctcccaca cctcccccctgaacctgaaacataaaatgaatgcaattg
Gua-P1-a9s3	gcggccgcgactcgattacacaatgccctctagaCCCTGCGTCACAGAACAGAC GACG**TATAATCGCGTGGATATGGCACGCAAGTTTCTACCGGGCACCGTAAATGT CCGACTA**CGTCGTCTGGGCGACGGTAAATAGGTGTTAGGCCCAGAGCGGCAGGG TAcAACTccataccacatttgtagaggttttacttgctttaaaaaacctcccac acctcccccctgaacctgaaacataaaatgaatgcaattg
Gua-P1-a10s1	gcggccgcgactcgattacacaatgccctctagaCCCTGCGTCACAaCCAGACG ACG**TATAATCGCGTGGATATGGCACGCAAGTTTCTACCGGGCACCGTAAATGTC CGACTA**CGTCGTCTGGGCGACGGTAAATAGGTGTTAGGCCCAGAGCGGCAGGGT AcAACTccataccacatttgtagaggttttacttgctttaaaaaacctcccaca cctcccccctgaacctgaaacataaaatgaatgcaattg
Gua-P1-a10s2	gcggccgcgactcgattacacaatgccctctagaCCCTGCGTCACAaaCCAGAC GACG**TATAATCGCGTGGATATGGCACGCAAGTTTCTACCGGGCACCGTAAATGT CCGACTA**CGTCGTCTGGGCGACGGTAAATAGGTGTTAGGCCCAGAGCGGCAGGG TAcAACTccataccacatttgtagaggttttacttgctttaaaaaacctcccac acctcccccctgaacctgaaacataaaatgaatgcaattg

Gua-P1-a10s3	gcggccgcgactcgattacacaatgccctctagaCCCTGCGTCACAagaCCAGACGACG**TATAATCGCGTGGATATGGCACGCAAGTTTCTACCGGGCACCGTAAATGTCCGACTA**CGTCGTCTGGGCGACGGTAAATAGGTGTTAGGCCCAGAGCGGCAGGGTACAACTccataccacatttgtagaggttttacttgctttaaaaaacctcccacacctccccctgaacctgaaacataaaatgaatgcaattg

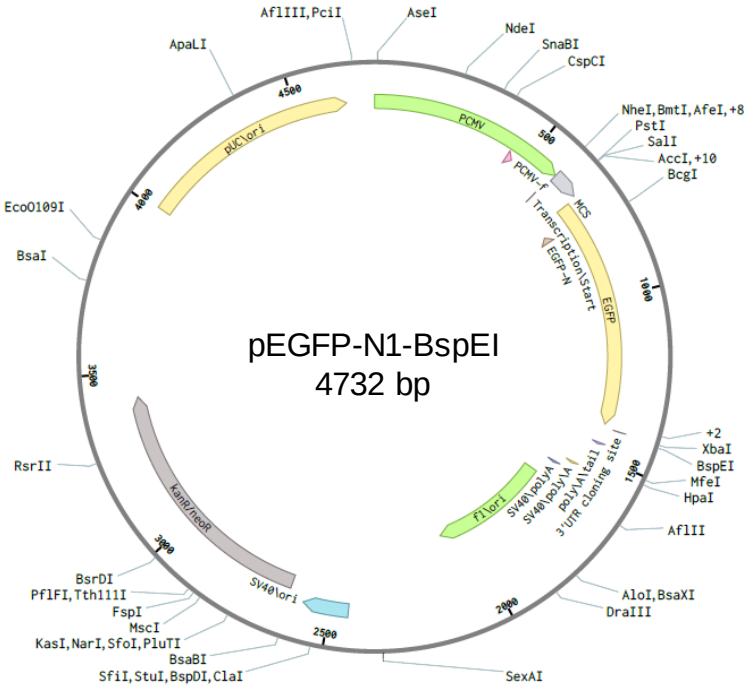


Figure S1. Plasmid map of pEGFP-N1-BspEI, the parental (empty) vector used to construct the riboswitch plasmids. The full DNA sequence is given in the Appendix below.

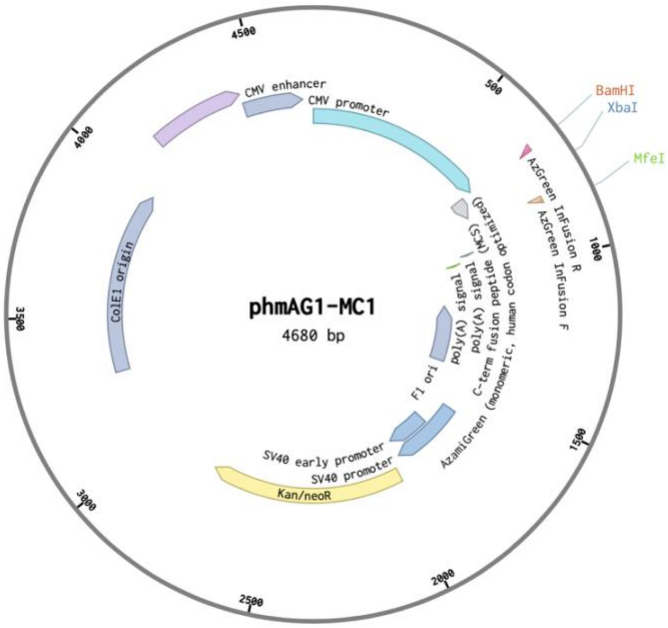


Figure S2. Plasmid map of phmAG1-MC1, the parental (empty) vector used to construct the riboswitch plasmids for all CPP-guanine constructs. The full DNA sequence is given in the Appendix at the end of the thesis.

5. Chapter 5. High-throughput screening of aptazymes in mammalian cells

5.1.Introduction

As new aptazyme architectures and scaffolds are being designed, aptazyme become more versatile tools for controlling gene expression. However, development of aptazymes for biomedical applications is hindered by low throughput of their testing in mammalian cells. Systematic rational design approaches, like the one presented in Chapter 4 or by the Farzan group (Zhong et al., 2016), provide valuable insights on the rules of RNA device engineering, but they cannot account for more complex RNA interactions and larger sequence spaces. Although various screening strategies have been employed in bacteria and yeast (Fowler et al., 2008) (Kobori et al., 2015) (Liang et al., 2012) (Lynch & Gallivan, 2009) (Muranaka et al., 2009), mammalian cells remain a challenging medium for high-throughput screening. As demonstrated in Chapter 3, transcription-translation system in mammalian cells is difficult to simulate *in vitro*, as many biochemical parameters of this process are not fully understood yet. Hence, until the critical factors for aptazyme activity are unraveled, screening in mammalian cells remains the most reliable approach for aptazyme engineering.

As introduced in Chapter 1, until recently, there had been two attempts to screen active ribozymes in mammalian cells (Nomura et al., 2017) (X. Chen et al., 2009). They were based on transfecting a dsDNA library of ribozyme variants into mammalian cells, letting them transcribe and self-cleave in the cell, extracting total RNA and separating them into desired fractions by gel. These methods allowed assessing library activity in a single experiment, without the need to operate large volumes of cell culture and screen for reporter protein expression, like in a typical mammalian cell screening (Priola et al., 2016). However, both methods suffered from low stability of nucleic acids in the cytoplasm and noise introduced

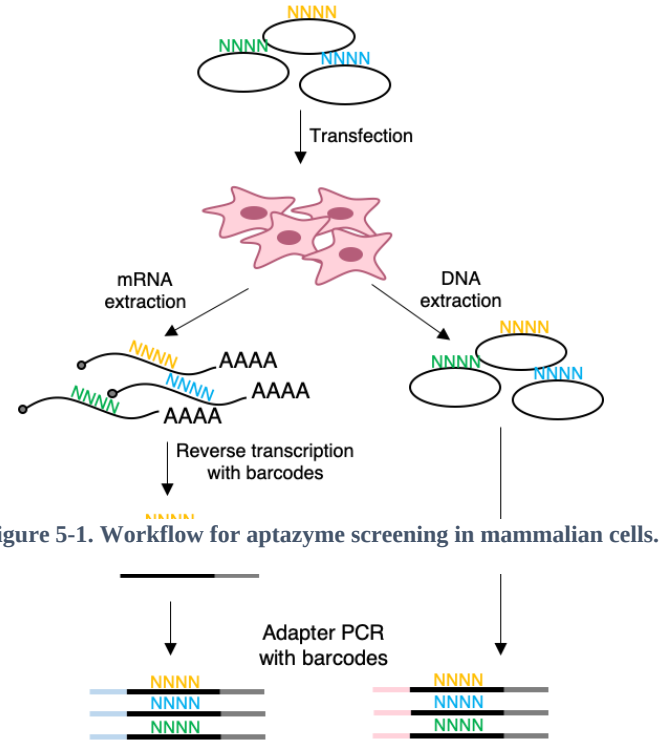


Figure 5-1. Workflow for aptazyme screening in mammalian cells.

during total RNA extraction and separation, which hindered their further applications.

To address these issues, I introduced several changes to the method to apply it for aptazyme screening in mammalian cells. The new method assesses aptazyme activity by sequencing only uncleaved variants of an aptazyme library (Figure 5-1). To address low nucleic acid stability in the cytoplasm, instead of linear dsDNA I will transfect a plasmid library containing aptazyme variants under the control of the CMV promoter. The construct will be transcribed and processed as mRNA, stability of which will be dependent on aptazyme cleavage. Therefore, only mRNAs containing uncleaved aptazyme variants will be harvested, since transcripts with active variants would be rapidly degraded in the cytoplasm. Extraction of only uncleaved species is expected to reduce sensitivity of the method to the RNA extraction procedure and reaction time. To account for the variability in the variants' distribution in the starting library, transfected DNA pool will be used for normalization.

5.1.3. Publication of the method by other groups

Soon after I started working on this project, the Smolke and the Hartig groups published articles describing very similar RNA sequencing-based screening approaches in mammalian cells (Strobel, Sporing, et al., 2020; Xiang et al., 2019). The Smolke group had the same workflow, except for extracting total RNA instead of mRNA as in my plan. They validated the method on hammerhead-theophylline library by screening inactivated ribozyme library and replicates of active ribozyme library (Xiang et al., 2019). Then they compared the results of the screening with a lentiviral-based FACS assay to confirm correlation of the library activity on RNA and protein levels. They applied the method to screen libraries with the same design, but different aptamers against four ligands and were able to select on-switches for all of them.

The Hartig group also tested the method on several libraries, but with a more sophisticated approach for assessing aptazyme's ligand response (Strobel, Sporing, et al., 2020). Instead of running the screening in the presence and absence of ligand, they ran it in increasing concentrations of the ligand and compared each variant's abundance across all of them. This allowed them to perform more complex statistical analyses to reduce the noise. Using this method, they were able to rediscover previously reported switches from randomized libraries, like K19 from the Suess group (Beilstein et al., 2015) and GuaM8HDV from the Yokobayashi group (Nomura et al., 2013), as well as find new aptazymes from guanine-hammerhead and twister-tetracycline libraries.

These two works demonstrated the potential and practicality of the method for RNA device screening in mammalian cells. Furthermore, they validated the correlation of RNA and protein levels for the aptazyme constructs and demonstrated robustness of the method with different aptamers and ribozymes. Hence, in my work, rather than optimizing the method, I focused on applying it to my aptazyme libraries referring to procedures described in these two papers.

5.2.Results

5.2.1. Library validation with N4 library

I first tested the method with the twister-guanine library with four randomized positions (Figure 5-2A). This library is expected to contain a large fraction of switching variants, since variants 5c, 6c, 7c and 9c from preliminary testing in chapter 4 are all included in the library

(Figure 5-2B). To ensure good sequencing coverage of this preliminary library, only four positions were randomized yielding 256 variants.

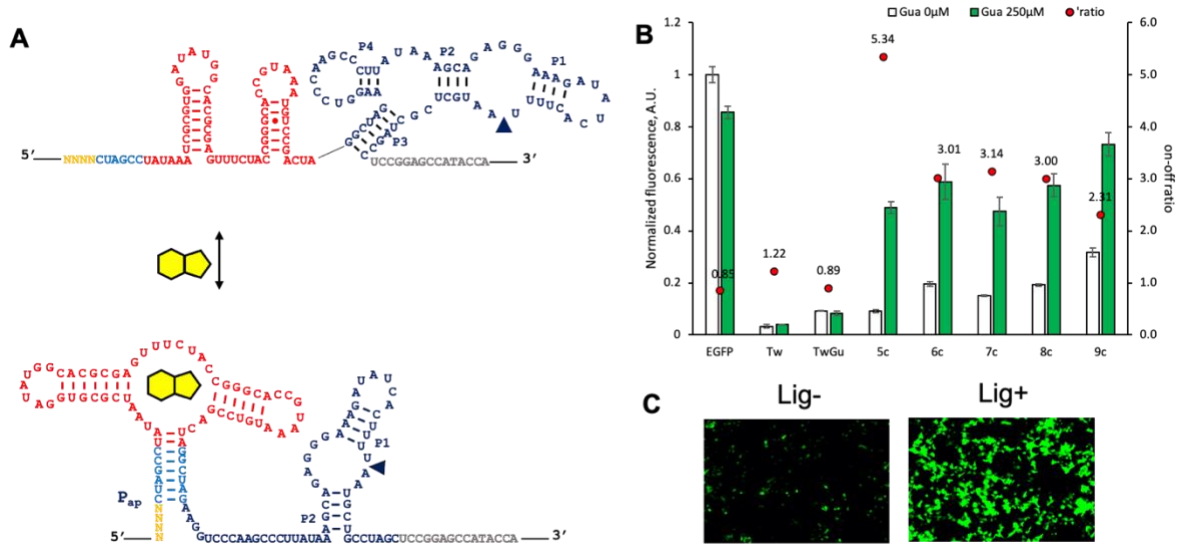


Figure 5-2. Twister-guanine N4 library. (A) Proposed library conformation in the absence and presence of guanine. Guanine aptamer Is marked in red and twister ribozyme is marked in dark blue. Nucleotides involved in formation of competing aptamer stem (P_{ap}) are marked in blue. (B) Activity of some aptazyme variants included in the library in HEK293 cells. EGFP – control with no aptazyme insertion, Tw – ribozyme only control, TwGua – aptazyme with no competing stem. 5c-9c are aptazymes with the corresponding number of basepairs in $P_{ap.}$ (C) Fluorescent micrographs of HEK293T cells transfected with the library.

The library was prepared according to the workflow summarized in Figure 5-1. Before transfection, a plasmid encoding an inactivated aptazyme variant with 2-nt mutation was spiked into the plasmid library at 1:128 ratio. This inactive variant was used to normalize the read counts of the active variants for each condition after sequencing. The micrograph of the transfected HEK293T cells 48 hours after transfection confirmed high number of switches in the library (Figure 5-2C), as fluorescence level was noticeably higher in the presence of the ligand. I extracted the RNA and DNA pools from the transfected cells and prepared the libraries for Illumina sequencing. I repeated this experiment twice on different days to test reproducibility of the method.

Thanks to the small library size, after sorting and quality filtering (QS > 30 for all randomized positions) all 256 variants passed the read count filter (>200) for each condition. First, every variant's read counts were normalized by the read count of the inactivated control variant in the same condition. Then, this normalized read count (DNA norm or RNA norm) was used to calculate their activity profile in the presence and absence of ligand. Unlike the *in vitro* method from Chapter 3, this method cannot provide the information on the cleaved variants because of the degradation of the corresponding mRNA in the cell. Instead, it sequences the uncleaved species of the RNA pool. In combination with the read distribution in the starting DNA pool, every variant can be characterized by its fraction uncleaved, F_{un}, which is the ratio between the number of RNA reads (representing uncleaved reads) and DNA reads (representing total reads) (Figure 5-5D).

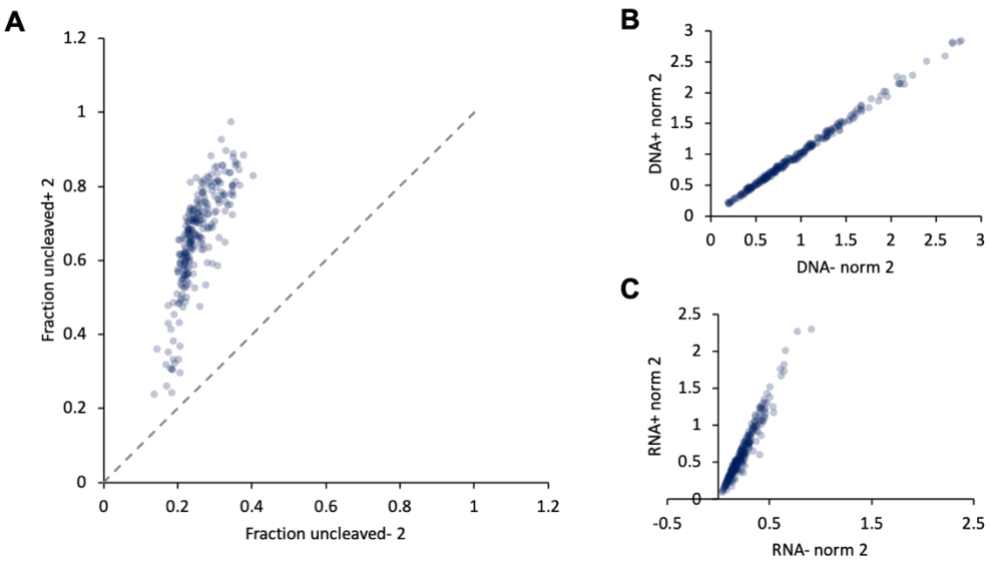


Figure 5-3. Summary of twister-guanine library screening, replicate 2. (A) Fractions uncleaved in the presence vs in the absence of 250 µM guanine. Dotted line indicates y=x line. Note abundance of variants above the line (B) Normalized reads from DNA library in the presence vs in the absence of 250 µM guanine. (C) Same as B, but for RNA.

The scatter plot of fractions uncleaved in the presence versus absence of the ligand reveals that the library almost entirely consists of on-switches (Figure 5-3A). All of the candidates are located above the y=x line, having more uncleaved transcripts in the presence than in the absence of ligand. DNA reads in the presence and absence of the ligand follow almost perfect one-to-one correlation, while the RNA reads are more abundant in the presence of ligand (Figure 5-3B and C). This confirms that the difference in fractions uncleaved is due to the RNA, not DNA variation.

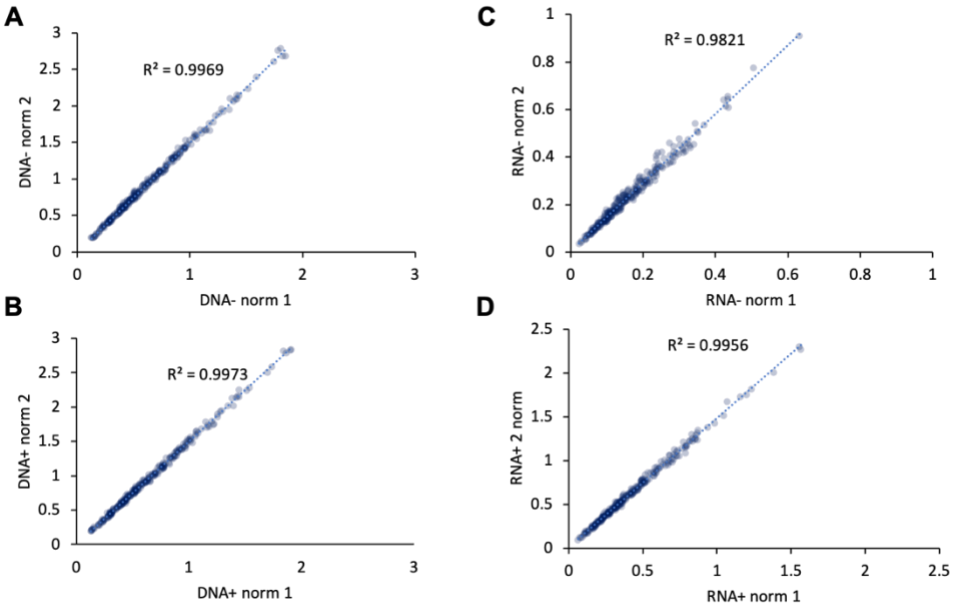


Figure 5-4. Reproducibility of normalized reads for twister-guanine library. (A) DNA without guanine. (B) DNA with 250 µM guanine. (C) RNA without guanine. (D) RNA with 250 µM guanine.

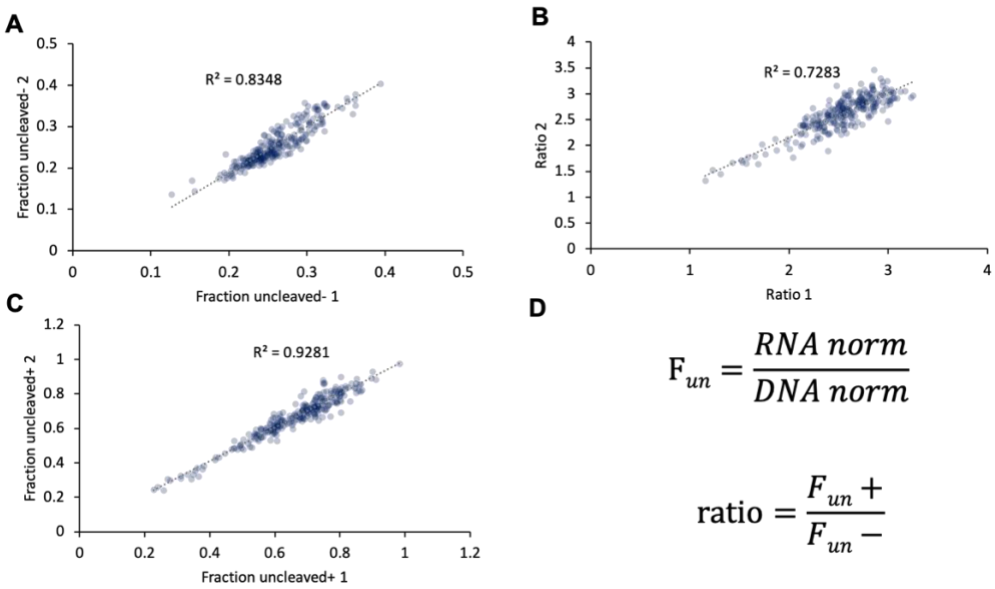


Figure 5-5. Reproducibility of fractions uncleaved in the absence (A) and presence (B) of ligand and their ratios (C) for twister-guanine library. (D) Formula for calculating fraction uncleaved (F_{un}) and on-off ratio

I then checked reproducibility of the screening by comparing numbers of reads and fractions cleaved from two experimental replicates (Figure 5-4 and 5-5). Comparison of RNA and DNA normalized reads in both conditions showed high correlation (Figure 5-4 A-D) with R^2>0.98. Next, I did the same comparison for the fractions uncleaved and on-off ratios (Figure 5-5). Since these values incorporate noise from all their upstream components, their alignment was not as clean as for individual parameters.

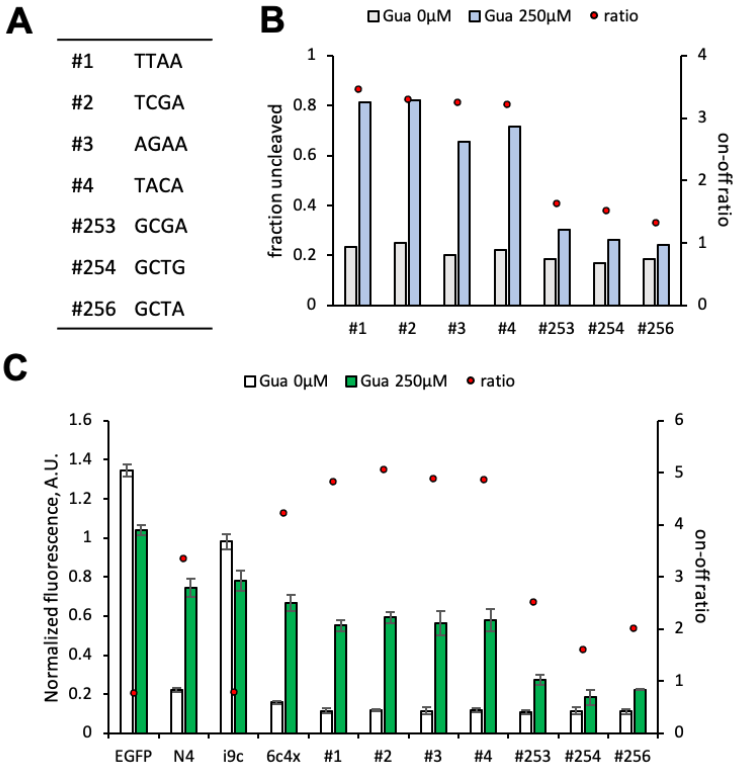


Figure 5-6. Validation of selected variants. (A) Randomized fragment composition of selected candidates. (B) Candidates' profiles based on sequencing reads. (C) Aptazymes' profiles assessed by fluorescence of EGFP in HEK293 cells. Aptazymes were inserted in the 3'UTR of EGFP. "EGFP" is a control plasmid without aptazyme insertion. N4 is the starting library. I9c is the inactivated version of 9c variant from Figure 5-2B. 6c4x is the control on-switch from Chapter 4.

To validate the screening results, I tested several candidates in HEK293 cells individually. I selected four candidates from the top and the bottom of the list sorted by their fractions uncleaved ratios (Figure 5-6A and B). Since the library is highly enriched in switches, even the lowest scoring candidates had on-off ratios are over 1. (Candidate #255 is missing because I failed to isolate a clone with correct sequence).

When tested in HEK293 cells, all candidates behaved as predicted by their RNA-sequencing profiles. The best four switches had activity levels and on-off ratios similar to 6c4x, one of the best switches found by rational design in chapter 4. Three "poor" candidates had lower expression levels and low on-off ratios as reflected in the sequencing result as well. Interestingly, the best switches only slightly outperformed the starting N4 library, once more highlighting high abundance of on-switches in this design. This validation experiment showed that RNA-sequencing method is able to reproducibly screen for aptazymes in mammalian cells.

5.2.2. Screening for switches with CPP-tetracycline N5 library

Next, I applied the screening method to the P1 CPP-tetracycline library, introduced in Chapter 4. In that design, the aptamer was inserted in the flexible junction of the CPP ribozyme with the communication module interacting with P1 stem (Figure 5-7A). In Chapter 4, I tried to optimize stability of P_{ap} by varying its length and adding spacers; however, there were many other potential modifications that were difficult to explore with one-by-one testing. The new screening method would allow assaying a large number of such modifications at once.

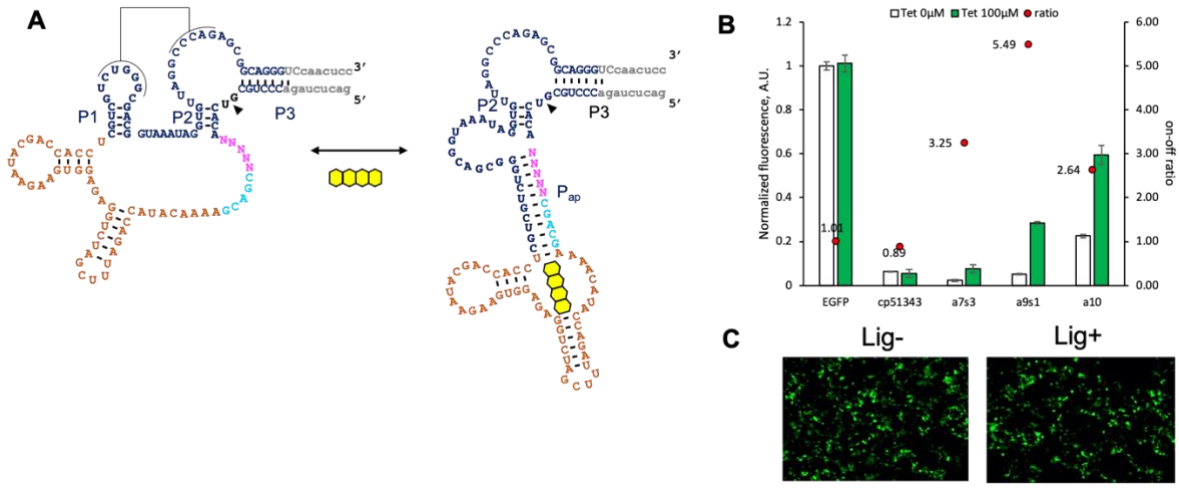


Figure 5-7. Circularly permuted pistol (CPP)-tetracycline N5 library. (A) Proposed library conformation in the absence and presence of tetracycline. Tetracycline aptamer is marked in brown and CPP ribozyme is marked in dark blue. Nucleotides involved in formation of competing aptamer stem (P_{ap}) are marked in cyan. (B) Activity of some aptazyme variants included in the library in HEK293 cells. EGFP – control with no aptazyme insertion, cp41343 – ribozyme only control, a7s3, a9s1, a10 are aptazymes with the corresponding number of basepairs (a) and spacers (s) in P_{ap}. (C) Fluorescent micrographs of HEK293T cells transfected with the library.

For the CPP-Tet library, I randomized five positions distal from the aptamer, yielding 1024 variants (Figure 5-7A). With this design, the library includes variants a7s3, a9s1 and a10 with different activity levels that were tested in chapter 4 as a part of the rational design strategy (Figure 5-7B). Since a10 variant contains a fully complementary sequence in P_{ap}, it likely represents the highest on-levels present in the library.

The micrographs of transfected HEK293T cells 48 hours after transfection showed similar levels of GFP expression in the presence and absence of ligand (Figure 5-7C), suggesting that the library has fewer switches than the twister-guanine one. The library was processed according to the workflow in Figure 5-1 twice on different days, and the two replicates were sequenced together.

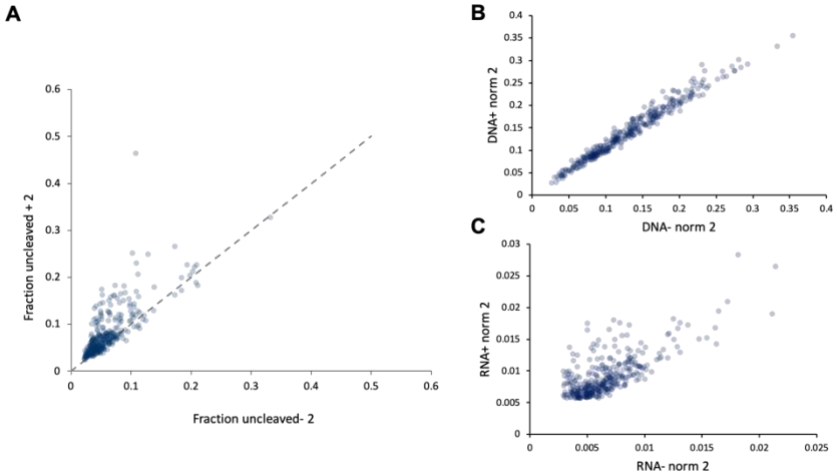


Figure 5-8. Summary of CPP-tetracycline library screening, replicate 2. (A) Fractions uncleaved in the presence vs in the absence of 100µM tetracycline. Dotted line indicates x=y line. Note the cluster of variants below 0.1 on both axes (B) Normalized reads from DNA library in the presence vs in the absence of 100 µM tetracycline. (C) Same as B, but for RNA.

Due to the poor sequencing run, I was not able to achieve a good coverage of the library, and only 351 out of 1024 variants passed the filter of >200 reads per each condition. Thus, all subsequent plots and calculations only reflect activities of these 351 variants. Sequencing revealed that the range of fractions uncleaved in the library was quite narrow, with only a few variants exceeding 0.2. Most of the variants clustered at the bottom left corner, meaning that most of the library was active in both lig- and lig+ conditions (Figure 5-8A). There were, however, some potential switches with higher fraction uncleaved in the presence of ligand. DNA alignment in the presence and absence of ligand was not as clean as with twister-guanine (Figure 5-8B). Since DNA distribution is not expected to vary between the two conditions, noise reflected in this plot was probably introduced during library preparation and will also be present in other samples.

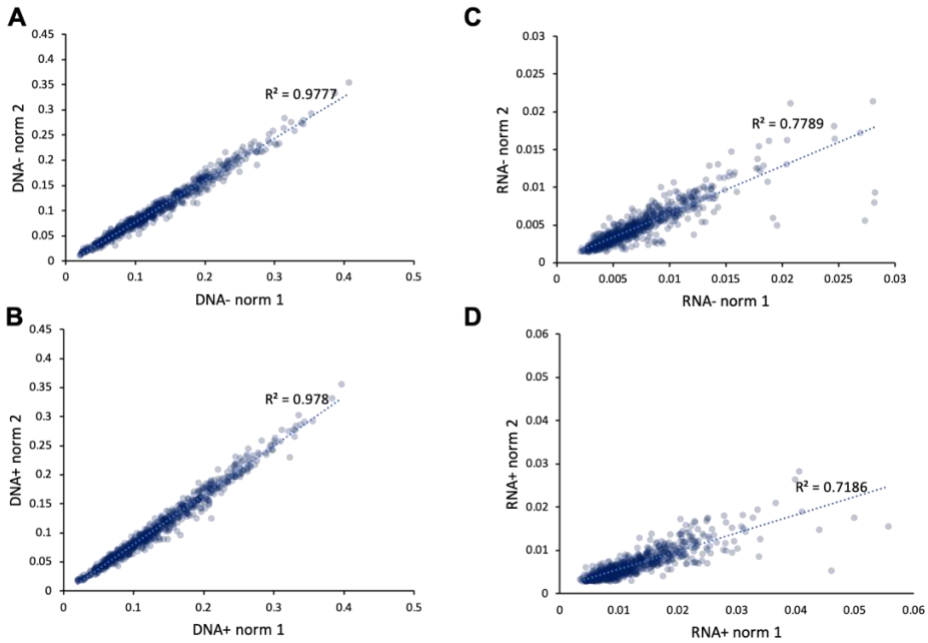


Figure 5-9. Reproducibility of normalized reads for CPP-guanine library. (A) DNA without guanine. (B) DNA with 100 µM tetracycline. (C) RNA without tetracycline. (D) RNA with 100 µM tetracycline

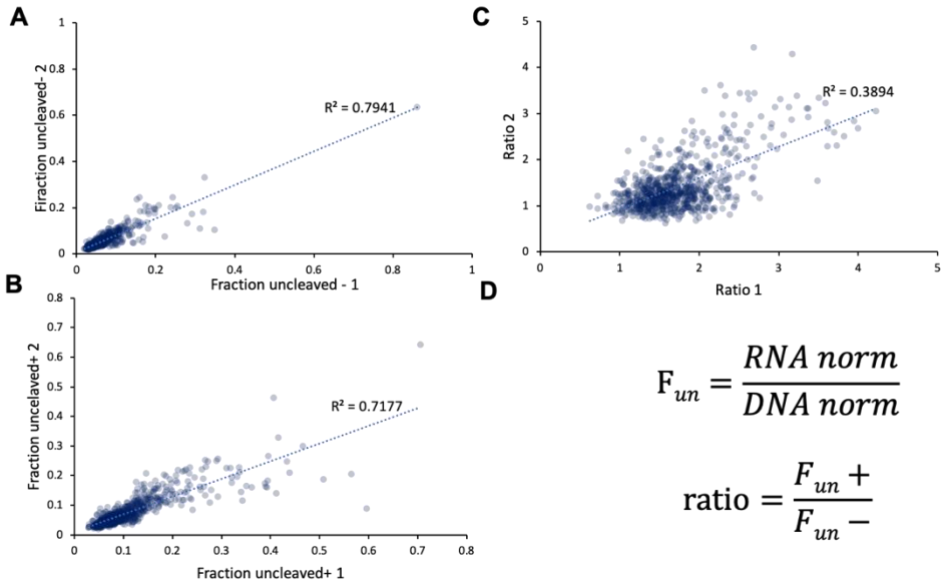


Figure 5-10. Reproducibility of fractions uncleaved in the absence (A) and presence (B) of ligand and their ratios (C) for CPP-tetracycline library. (D) Formula for calculating fraction uncleaved (F_{un}) and on-off ratio

CPP-tetracycline library demonstrated slightly lower reproducibility across the replicates than the twister-guanine library (Figure 5-9). Although DNA reads across the replicates still aligned well ($R^2 > 0.97$), RNA alignment did not have the same excellent correlation (Figure 5-9B, D). This could potentially be to the high activity of the circularly permuted pistol ribozyme, with some residual cleavage happening during library preparation before it is converted into cDNA. Consequently, alignment of fractions uncleaved and their ratios had lower correlations (Figure 5-10A-C).

Although the reproducibility was not as good as for twister-guanine library, the screening still identified promising variants that were consistent between the two replicates. I selected four

candidates with high and low on-off ratios and tested them individually in HEK293 cells (Figure 5-11). Overall, all candidates followed the trend predicted by sequencing, with the "best" four switches having higher on-off ratios than the "poorest" four. However, all tested candidates had low EGFP expression overall, even when activated by tetracycline. Although higher on-levels are often desirable, this finding was in line with the original expectation of low on-levels limited by a10 variant (Figure 5-7B) and low fractions uncleaved revealed by sequencing (Figure 5-8A). The best of selected switches, #5 (Figure 5-11), was comparable to a9s1 (Figure 5-8A) previously found by rational design, although both suffered from low on-levels. Hence, despite some variability across the replicates and insufficient library coverage, the method was still able to select on-switches from a heterogeneous library. Nevertheless, another trial of this experiment, with higher library coverage is necessary for a more comprehensive assessment.

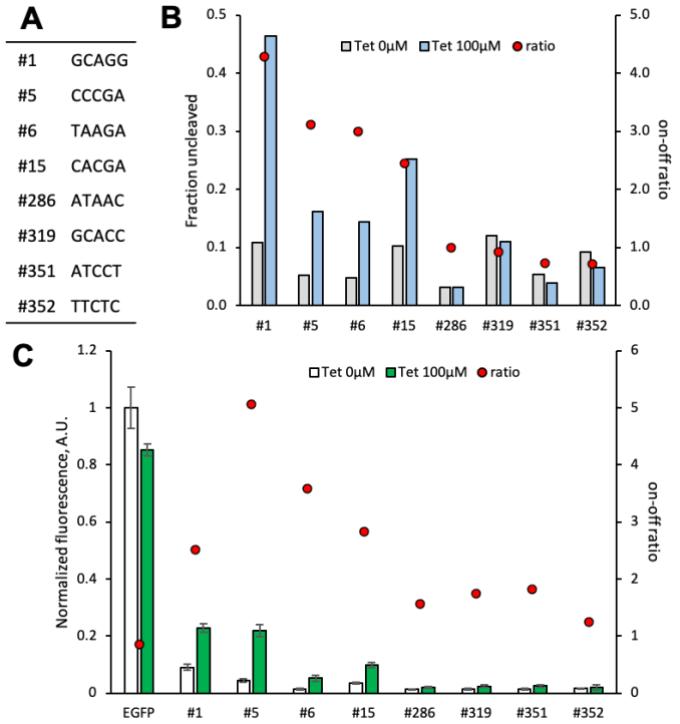


Figure 5-11. Validation of selected variants. (A) Randomized fragment composition of selected candidates. (B) Candidates' profiles based on sequencing reads. (C) Aptazymes' profiles assessed by fluorescence of EGFP in HEK293 cells. Aptazymes were inserted in the 3'UTR of EGFP. "EGFP" is a control plasmid without aptazyme insertion

5.3.Discussion

Preliminary screening of twister-guanine library demonstrated ability of the method to reproducibly sort switches based on their dynamic range. Strong correlation (R^2>0.98) between the two replicates of the screening was achieved for both DNA and RNA reads, which indicates that library preparation and extraction procedures are able to preserve variants' abundance. Twister-guanine N4 library contained several variants previously tested in Chapter 4, which could be used as controls for the screening. Their ranking based on on-off ratio met our expectations and helped to estimate ranges of on-off ratios for other variants in the library.

In terms of dynamic range, profiles of the best four switches selected by this method were slightly better than that of 6c4x, one of the best switches developed by rational design in the previous chapter. This indicates that although the library mostly consists of on-switches, it is

unlikely to have variants with on-off ratios significantly higher than those already found. However, testing more of the top candidates predicted by screening may reveal switches with slightly improved performance.

The lower reproducibility for the CPP-tetracycline library P1 library could be due to several reasons. First, due to low yield of the sequencing run, reads number was not sufficient for full library coverage. Many variants did not pass read count threshold and were excluded from the ranking. Second, CPP-tetracycline and twister-guanine libraries had different design, which could affect library preparation. Since randomized region in the CPP library was only five positions away from the cleavage site, reverse transcription primer annealing site was located in the 3' cleaved fragment. Thus, if some variants self-cleaved during RNA extraction or DNase treatment, they would not be reverse transcribed, and thus, their activity profiles based on sequencing would be skewed. In contrast, in the twister-guanine library, reverse transcription primer annealing site is located on the 5' side of the cleavage site, ensuring that all fragments will be reverse transcribed regardless of their cleavage during library preparation. Third, the libraries have different aptamer and ribozyme components and may have different kinetic profiles. For example, one of the ribozymes may have faster cleavage, resulting in a very low number of uncleaved transcripts, which would be more prone to random variations upon amplifications in library preparation.

To address these issues, the screening has to be repeated with different libraries and good reads coverage. Furthermore, more candidates from the screening have to be tested in cells individually to confirm reliability of the method. Lastly, statistical analyses and motif search similar to the ones run by the Hartig and Smolke groups may help to draw more conclusions from the screening.

5.4.Conclusion and outlook

This study showed that high-throughput screening of aptazyme in mammalian cells based on RNA sequencing is able to select ON-switches from a pool of randomized variants. This method allows characterizing activity of hundreds of ribozyme variants in mammalian cells directly, overcoming the problems with low throughput or reaction context discussed in previous chapters. Validation experiment on a 256 variants twister-guanine library selected on-switches comparable to 6c4x, one of the best variants found by rational design in chapter 4. Although there seems to be some dependence on the library design, overall, the method is reproducible and reliable at selecting switches with high on-off ratios. Applying this method to more libraries described in this thesis may reveal more aptazymes with high on-off ratios and will be done at the earliest opportunity.

This method will allow making aptazyme engineering more systematic and higher throughput, overcoming the major bottleneck in mammalian aptazyme engineering. Moreover, once sensitivity of the method is improved, aside from optimizing communication modules of new aptazymes, it may be used for screening of functional elements of an aptazyme, such as high-affinity aptamers or ribozyme variants, or of other RNA devices.

5.5.Methods

<u>Plasmid library preparation</u>

Randomized plasmid library was made by running whole-plasmid PCR with the forward primer containing randomized overhang. The resulting library of linear randomized plasmids was then treated with DpnI restriction enzyme to digest the template and purified with a Zymo microspin column. After phosphorylation with T4 polynucleotide kinase and subsequent heat-inactivation, the library was circularized with T4 DNA ligase overnight at 16°C. Slow ligation step is crucial for maximizing the number of colonies in the transformation. Two microliters of ligation mix was then used to transform 30 μL of DH5α competent *E. coli* cells (Toyobo). After quick thawing on ice (5 min), incubation on ice with the plasmid mixture (5 min), heat-shock at 42°C (40 sec) and recovery incubation on ice (5 min), 150 μl of LB was added to the cells, and the mix was incubated at 37°C at 250 rpm for the expression of KanR gene. After 90 minutes, 16 μL of the transformation mix was diluted with 50 μl LB and spread on a kanamycin (50 μg/mL) LB agar plate with subsequent overnight incubation (16 h) at 37°C. 160 μL of the transformation mix (10x of the amount plated) were used to inoculate a 5 mL of liquid culture in kanamycin (50 μg/mL) LB medium and incubated overnight (16 h) at 37°C at 250 rpm. Since the ratio of the cells plated vs cells used for seeding is 1:10, approximate number of colonies in the liquid culture can be estimated as 10x of the colonies on the plate. Based on the colony number, 10-fold library coverage was achieved for both libraries. Plasmid library was extracted with a miniprep kit (Qiagen) according to the manufacturer's instructions.

HEK293T cell culture and library transfection.

HEK293T cells were cultured in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% heat-inactivated FBS (Gibco) containing 2 mM l-glutamine and 100 units/mL of penicillin-streptomycin (DMEM-FBS). Cells were kept in a 37°C incubator with 5% CO_2 and passaged regularly upon reaching 90% confluency. Approximately 20 h prior to transfection, the cells were trypsinized, diluted to ~1×10^5 cells/mL and 2.5 mL/well were seeded into two wells of a 6-well plate. Each of the two wells was then transfected with identical mixture of 250 μl of Opti-MEM, 2.5 μg of the plasmid library and 7.5 μl of TransIT-293 Transfection Reagent (Mirus) according to the manufacturer's instructions. Five hours after transfection, the medium in each well was replaced with fresh medium with (250 μM) or without guanine. Guanine was dissolved at 25 mM in 0.2 M NaOH, which was diluted by 100-fold in DMEM-FBS prior to addition to wells. The same volume of 0.2 M NaOH solution was diluted in DMEM-FBS for the medium without guanine. Alternatively, for the tetracycline aptamer library, stock 10 mM tetracycline dissolved in H_2O was used instead of 25 mM Guanine in 0.2 M NaOH, with final concentration in medium of 100 μM. Equivalent amount of H_2O was used for the "no ligand" well.

RNA and DNA extraction and library preparation

Cells were trypsinized 48 h after transfection and washed with 3 mL of medium, counted and resuspended in 2 mL of PBS solution supplemented with 2% FBS. The resuspended cells (1 mL) were used for mRNA extraction with Dynabeads mRNA Direct extraction kit (Invitrogen). mRNA was extracted following the manufacturer's instructions and kept annealed to the beads for DNase treatment step. After the final wash with washing buffer B, the beads were resuspended in total volume of 50 μl containing 1X Turbo DNase buffer and 2 units of Turbo DNase (Thermo Scientific). Then the tube was incubated for 1 h on ice to minimize aptazyme cleavage in DNase buffer with high [Mg^{2+}] (for TwGua library DNase treatment was run for 30 min at 37°C because the library design was safe).

After DNase treatment the beads were washed with 50 µl of ice-cold water and the RNAs were eluted with 22 µL of elution buffer at 80°C for 3 mins.

For each reverse transcription reaction, 10 µl of eluted RNA fragments were mixed with 10 nmol NTPs (each), 10 pmol corresponding barcoded reverse primer as summarized in Table S2 and RNase-free water to make up to total volume of 15.5 µl. This mixture was heated to 72 °C for 3 min and then placed on ice. Then reverse transcription reaction was initiated by the addition of 250U of Maxima H Minus Reverse Transcriptase (Thermo Fisher Scientific), 10U of murine RNase inhibitor (New England Biolabs), and reverse transcription buffer (Thermo Fisher Scientific) to make up to final 1x in 20 µL (50 mM Tris-HCl (pH 8.3 at 25°C), 75 mM KCl, 3 mM $MgCl_2$, 10 mM DTT). Reaction was run at 50°C for 1 hour and was followed by heat-inactivation of the enzyme at 65°C for 20 minutes. Upon the reaction completion 1 µL of 5M NaOH solution was added to each tube, and the tubes were incubated at 95 °C for 5 min. The resulting cDNAs were cleaned by Oligo Clean & Concentrator kit (Zymo Research) and eluted with 10 µL H_2O. Remaining 1 mL of cells was used for DNA extraction with Quick-DNA Miniprep kit (Zymo Research) according to the manufacturer's instructions. Extracted total DNA containing transfected plasmid DNA was eluted with 50 µL H_2O.

Extracted total DNA and reverse transcribed cDNAs were used as templates for two subsequent PCR with Q5 HF MasterMix (New England Biolabs) to attach barcodes and sequencing adapters as primer overhangs (summarized in Table S2). Length barcodes (full vs cleaved) differed in length (8 and 10 nt, or 9 and 11 nt) to shift identical nucleotides reads and avoid overclustering during MiSeq sequencing. The number of PCR cycles varied between the total of 20 and 35, depending on the library. Final libraries were run on a 2% agarose gel, and the band of the corresponding size was extracted using Zymoclean DNA Gel recovery kit (Zymo Research) to remove unreacted primers with adapters. After gel extraction, the libraries were dissolved in 10 µL RNase-free water. Final library compositions are summarized in the Table S1.

Table S1. Read maps of the libraries used in this chapter.

Library Name	Sequence for MiSeq read	Length
TwGua N4	AATGATACGGCGACCACCGAGATCTACACACA**CTCT** **TTCCCT**ACACGACGCTCTTCCGATCT[ligand barcode]GTAGAAACTTGCGTGCCATATCCACGCGATTA TAGGCTANNNNtctagagtcgcggccgcttt[DNA/RNA barcode]ATCTCGTATGCCGTCTTCTGCTTG	165/ 167
CPP-Tet P1	AATGATACGGCGACCACCGAGATCTACACACA**CTCT** **TTCCCT**ACACGACGCTCTTCCGATCT[ligand barcode]TTCACCTCTCCAGATCGAAATCTGGTATGTTT	163/ 165

	TCGTCGNNNNNTGTGACGCAGGGtctagag[DNA/RNA barcodeATCTCGTATGCCGTCTTCTGCTTG	

Color coding: yellow – P5 adapter, gray – R1seq, blue – aptamer, purple – communication module, green – ribozyme, dark blue – P7 adapter, underlined – Illumina i5 and i7 indexes.

Table S2. Barcodes used in this chapter

Library	Condition	Barcode sequence in the final construct (5'-3')
All libraries:	Ligand -	TATCCTCT
	Ligand +	AGAGTAGACA
Twister-Guanine N4	DNA 1	CAGTTCA
	RNA 1	CTATACG
	DNA 2	GATGACT
	RNA 2	TTACGTC
CPP-Tetracycline N5	DNA 1	CAGTTCA
	RNA 1	CTATACG
	DNA 2	ATCAATG
	RNA 2	TGATTGC

Illumina sequencing and data processing.

Twister-guanine library was sequenced on an Illumina MiSeq v3 machine at the loading concentration of 12 pM with 15% spiked-in PhiX genome to increase the library diversity. Raw fastq reads were first sorted into lig- and lig+ pools by the 8-nt and 10-nt barcodes. Each of these pools was subsequently sorted into DNA and RNA populations by the 6-nt barcode at the 3' end. After verifying and trimming the constant sequences, the remaining randomized fragments were quality filtered to only contain base calls with Phred score over 30. Filtered reads were converted into fasta format and counted. Sequences with read counts below 200 in any of the four conditions (lig- and lig+, DNA and RNA) were not included in the further analysis. Read counts were first normalized by the number of reads of the inactive control in each corresponding condition:

$$reads\ norm = \frac{reads\ of\ a\ variant}{reads\ of\ the\ inactive\ control}$$

Normalized RNA reads represent the fraction of uncleaved reads for the given variant, whereas normalized DNA reads represent the fraction of total reads in the starting pool.

Then activity profile for a variant was calculated as fraction uncleaved (F_{un}) according to the following formula:

$$F_{un} = \frac{RNA\ norm}{DNA\ norm}$$

Then fraction intact in presence of ligand ($F_{un}+$) was plotted versus fraction full-length in absence of ligand ($F_{un}-$) in a scatter plot.

HEK293 cell culture and candidates' validation.

HEK293 cells were cultured in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% heat-inactivated FBS (Gibco) containing 2 mM l-glutamine and 100 units/mL of penicillin-streptomycin (DMEM-FBS). Cells were kept in a 37°C incubator with 5% CO_2 and passaged regularly upon reaching 90% confluency. Approximately 20 h prior to transfection, the cells were trypsinized, diluted to ~2.7×10^5 cells/mL and 100 µL/well were seeded onto a 96-well plate. Cells in each well were cotransfected with 100 ng of the EGFP-aptazyme plasmid and 20 ng of pCMV-mCherry (transfection control) using 0.3 µL of TransIT-293 Transfection Reagent (Mirus) according to the manufacturer's instructions. Five hours after transfection, the medium in each well was replaced with fresh medium with (250 µM) or without guanine. Guanine was dissolved at 25 mM in 0.2 M NaOH which was diluted by 100-fold in DMEM-FBS. The same volume of 0.2 M NaOH solution was diluted in DMEM-FBS for the medium without guanine. 48 hours after transfection, the medium in each well was replaced with 100 µL of phosphate buffered saline (PBS), and fluorescence intensity was measured by Infinite M1000 PRO microplate reader (Tecan). Fluorescence intensity was measured at 484 nm excitation/510 nm emission/5 nm bandwidth for EGFP, and at 587 nm excitation/610 nm emission/10 nm bandwidth for mCherry. Background fluorescence measured using untransfected cells was subtracted from the EGFP and mCherry fluorescence values. Then, EGFP fluorescence was normalized by mCherry fluorescence to account for variations in transfection efficiency and cell counts. All reported values are averages of three replicate wells.

5.6.Supplementary materials

Table S3. Candidates used in validation experiment for twister-guanine library

Name	Sequence	F_{un}- 1	F_{un}+ 1	Ratio 1	F_{un}- 2	F_{un}+ 2	Ratio 2
#1	TTAA	0.276	0.788	2.856	0.234	0.811	3.463
#2	TCGA	0.269	0.786	2.923	0.249	0.823	3.298
#3	AGAA	0.205	0.612	2.990	0.202	0.655	3.247
#4	TACA	0.242	0.711	2.943	0.222	0.715	3.224
#253	GCGA	0.203	0.343	1.687	0.187	0.304	1.631
#254	GCTG	0.195	0.239	1.231	0.171	0.261	1.525
#255	GCTT	0.213	0.279	1.308	0.206	0.297	1.440

#256	GCTA	0.197	0.227	1.156	0.184	0.243	1.318

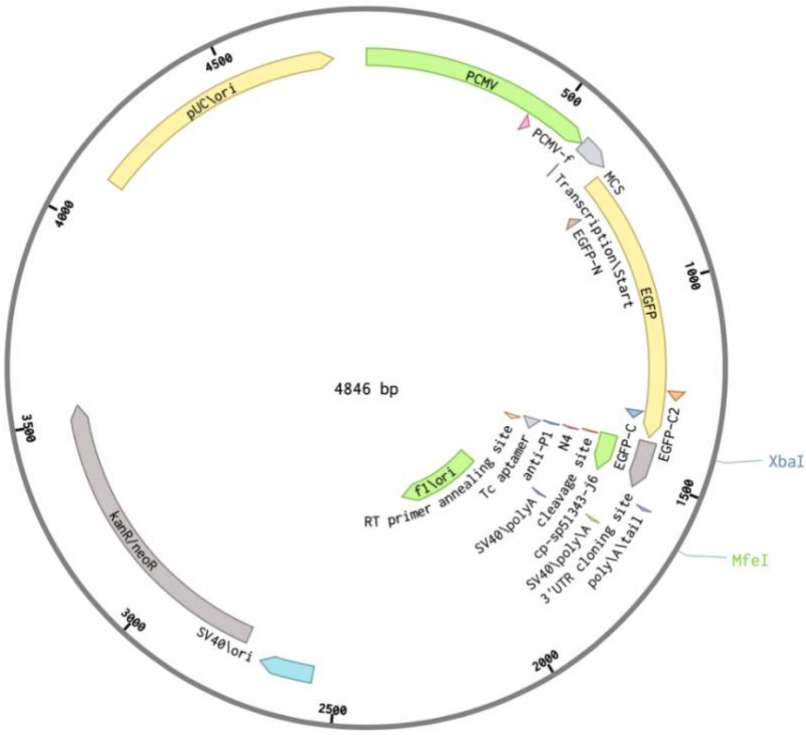
Table S3. Candidates used in validation experiment for cpp-tetracycline library

Name	Sequence	F_{un}- 1	F_{un}+ 1	Ratio1	F_{un}- 2	F_{un}+ 2	Ratio2
#1	GCAGG	0.128	0.406	3.165	0.108	0.267	4.286
#5	CCCGA	0.116	0.390	3.373	0.052	0.134	3.114
#6	TAAGA	0.100	0.229	2.291	0.048	0.211	2.990
#15	CACGA	0.126	0.266	2.109	0.103	0.385	2.451
#286	ATAAC	0.033	0.048	1.470	0.031	0.646	1.001
#319	GCACC	0.122	0.135	1.101	0.120	0.892	0.916
#351	ATCCT	0.049	0.067	1.369	0.054	0.809	0.723
#352	TTCTC	0.069	0.120	1.745	0.092	0.767	0.712

Table S4. Variable region of the pEGFP plasmid libraries.
Aptazyme plasmid libraries were designed by inserting the following sequences flanked by XbaI (green) and MfeI (red) restriction sites into the pEGFP-N1 plasmid shown below. Aptamers are marked in **bold**. The nucleotides involved in the P_{ap} formation are underlined. Spacers and mismatches are marked in lowercase letters

Library	Sequence
Twister-guanine N4	tctagaCCCTGCGTCACANNNNNCGACG**AAAACATACCAGATTTCG ATCTGGAGAGGTGAAGAATACGACCACCT**CGTCGTCTGGGCGA CGGTAAATAGGTGTTAGGCCCAGAGCGGCAGGGTAcAACTccatacc acatttgtagaggttttacttgctttaaaaaacctcccacacctccccctgaacctgaaacataaaatgaatgca attg
CPP-tetracycline N5	tctagaNNNNTAGCC**TATAATCGCGTGGATATGGCACGCAAGTTT CTACCGGGCACCGTAAATGTCCGACTA**GGCTAGAAGGTCCCAA GCCCTTATAAAGCAGAGGGAAAGATATCACTTTTAATGCTGCCTA GCtccggagccataccacatttgtagaggttttacttgctttaaaaaacctcccacacctccccctgaacctga aacataaaatgaatgcaattg

Figure S1. Plasmid map of the pEGFP-N1 plasmid used as a backbone for this study. Full sequence information is available in Appendix


4500
500
1000
1500
2000
2500
3000
3500
4000
pUC\ori
PCMV
PCMV-r
MCS
Transcription\Start
EGFP-N
EGFP
4846 bp
f1\ori
RT primer annealing site
Tc aptamer
anti-P1
N4
cleavage site
EGFP-C
EGFP-C2
SV40\polyA
cp-sp51343-j6
SV40\polyA
3'UTR cloning site
polyA\tail
XbaI
MfeI
KanR/neoR
SV40\ori

References

An, C. I., Trinh, V. B., & Yokobayashi, Y. (2006). Artificial control of gene expression in mammalian cells by modulating RNA interference through aptamer-small molecule interaction. *RNA, 12*(5), 710-716. doi:10.1261/rna.2299306

Auslander, S., & Fussenegger, M. (2016). Engineering Gene Circuits for Mammalian Cell-Based Applications. *Cold Spring Harb Perspect Biol, 8*(7). doi:10.1101/cshperspect.a023895

Auslander, S., & Fussenegger, M. (2017). Synthetic RNA-based switches for mammalian gene expression control. *Curr Opin Biotechnol, 48*, 54-60. doi:10.1016/j.copbio.2017.03.011

Auslander, S., Ketzer, P., & Hartig, J. S. (2010). A ligand-dependent hammerhead ribozyme switch for controlling mammalian gene expression. *Mol Biosyst, 6*(5), 807-814. doi:10.1039/b923076a

Auslander, S., Stucheli, P., Rehm, C., Auslander, D., Hartig, J. S., & Fussenegger, M. (2014). A general design strategy for protein-responsive riboswitches in mammalian cells. *Nat Methods, 11*(11), 1154-1160. doi:10.1038/nmeth.3136

Barrick, J. E., Corbino, K. A., Winkler, W. C., Nahvi, A., Mandal, M., Collins, J., . . . Breaker, R. R. (2004). New RNA motifs suggest an expanded scope for riboswitches in bacterial genetic control. *Proc Natl Acad Sci U S A, 101*(17), 6421-6426. doi:10.1073/pnas.0308014101

Bartel, D. P., & Szostak, J. W. (1993). Isolation of new ribozymes from a large pool of random sequences [see comment]. *Science, 261*(5127), 1411-1418. doi:10.1126/science.7690155

Beilstein, K., Wittmann, A., Grez, M., & Suess, B. (2015). Conditional control of mammalian gene expression by tetracycline-dependent hammerhead ribozymes. *ACS Synth Biol, 4*(5), 526-534. doi:10.1021/sb500270h

Beisel, C. L., Chen, Y. Y., Culler, S. J., Hoff, K. G., & Smolke, C. D. (2011). Design of small molecule-responsive microRNAs based on structural requirements for Drosha processing. *Nucleic Acids Res, 39*(7), 2981-2994. doi:10.1093/nar/gkq954

Bell, C. L., Yu, D., Smolke, C. D., Geall, A. J., Beard, C. W., & Mason, P. W. (2015). Control of alphavirus-based gene expression using engineered riboswitches. *Virology, 483*, 302-311. doi:10.1016/j.virol.2015.04.023

Belmont, B. J., & Niles, J. C. (2010). Engineering a direct and inducible protein-RNA interaction to regulate RNA biology. *ACS Chem Biol, 5*(9), 851-861. doi:10.1021/cb100070j

Berens, C., Groher, F., & Suess, B. (2015). RNA aptamers as genetic control devices: the potential of riboswitches as synthetic elements for regulating gene expression. *Biotechnol J, 10*(2), 246-257. doi:10.1002/biot.201300498

Berens, C., Thain, A., & Schroeder, R. (2001). A tetracycline-binding RNA aptamer. *Bioorg Med Chem, 9*(10), 2549-2556. doi:10.1016/s0968-0896(01)00063-3

Bloom, R. J., Winkler, S. M., & Smolke, C. D. (2014). A quantitative framework for the forward design of synthetic miRNA circuits. *Nat Methods, 11*(11), 1147-1153. doi:10.1038/nmeth.3100

Chandra, V., Hannan, Z., Xu, H., & Mandal, M. (2017). Single-molecule analysis reveals multi-state folding of a guanine riboswitch. *Nat Chem Biol, 13*(2), 194-201. doi:10.1038/nchembio.2252

Chen, C. Y., Ezzeddine, N., & Shyu, A. B. (2008). Messenger RNA half-life measurements in mammalian cells. *Methods Enzymol, 448*, 335-357. doi:10.1016/S0076-6879(08)02617-7

Chen, H., Li, Y., Du, C., Li, Y., Zhao, J., Zheng, X., . . . Xia, H. (2018). Aptazyme-mediated direct modulation of post-transcriptional sgRNA level for conditional genome editing and gene expression. *J Biotechnol, 288*, 23-29. doi:10.1016/j.jbiotec.2018.10.011

Chen, X., Denison, L., Levy, M., & Ellington, A. D. (2009). Direct selection for ribozyme cleavage activity in cells. *RNA, 15*(11), 2035-2045. doi:10.1261/rna.1635209

Chen, Y. Y., Jensen, M. C., & Smolke, C. D. (2010). Genetic control of mammalian T-cell proliferation with synthetic RNA regulatory systems. *Proc Natl Acad Sci U S A, 107*(19), 8531-8536. doi:10.1073/pnas.1001721107

Culler, S. J., Hoff, K. G., & Smolke, C. D. (2010). Reprogramming cellular behavior with RNA controllers responsive to endogenous proteins. *Science, 330*(6008), 1251-1255. doi:10.1126/science.1192128

Deng, N. N., Vibhute, M. A., Zheng, L., Zhao, H., Yelleswarapu, M., & Huck, W. T. S. (2018). Macromolecularly Crowded Protocells from Reversibly Shrinking Monodisperse Liposomes. *J Am Chem Soc, 140*(24), 7399-7402. doi:10.1021/jacs.8b03123

Dohno, C., Kimura, M., & Nakatani, K. (2018). Restoration of Ribozyme Tertiary Contact and Function by Using a Molecular Glue for RNA. *Angew Chem Int Ed Engl, 57*(2), 506-510. doi:10.1002/anie.201709041

Dupuis, N. F., Holmstrom, E. D., & Nesbitt, D. J. (2014). Molecular-crowding effects on single-molecule RNA folding/unfolding thermodynamics and kinetics. *Proc Natl Acad Sci U S A, 111*(23), 8464-8469. doi:10.1073/pnas.1316039111

Dwidar, M., Seike, Y., Kobori, S., Whitaker, C., Matsuura, T., & Yokobayashi, Y. (2019). Programmable Artificial Cells Using Histamine-Responsive Synthetic Riboswitch. *J Am Chem Soc, 141*(28), 11103-11114. doi:10.1021/jacs.9b03300

Eiler, D., Wang, J., & Steitz, T. A. (2014). Structural basis for the fast self-cleavage reaction catalyzed by the twister ribozyme. *Proc Natl Acad Sci U S A, 111*(36), 13028-13033. doi:10.1073/pnas.1414571111

Ellington, A. D., & Szostak, J. W. (1990). In vitro selection of RNA molecules that bind specific ligands. *Nature, 346*(6287), 818-822. doi:10.1038/346818a0

Ellis, R. J. (2001). Macromolecular crowding: an important but neglected aspect of the intracellular environment. *Curr Opin Struct Biol, 11*(1), 114-119. doi:10.1016/s0959-440x(00)00172-x

Endo, K., Hayashi, K., Inoue, T., & Saito, H. (2013). A versatile cis-acting inverter module for synthetic translational switches. *Nat Commun, 4*, 2393. doi:10.1038/ncomms3393

Endo, K., Hayashi, K., & Saito, H. (2016). High-resolution Identification and Separation of Living Cell Types by Multiple microRNA-responsive Synthetic mRNAs. *Sci Rep, 6*, 21991. doi:10.1038/srep21991

Felletti, M., & Hartig, J. S. (2017). Ligand-dependent ribozymes. *Wiley Interdiscip Rev RNA, 8*(2). doi:10.1002/wrna.1395

Felletti, M., Stifel, J., Wurmthaler, L. A., Geiger, S., & Hartig, J. S. (2016). Twister ribozymes as highly versatile expression platforms for artificial riboswitches. *Nat Commun, 7*, 12834. doi:10.1038/ncomms12834

Fowler, C. C., Brown, E. D., & Li, Y. (2008). A FACS-based approach to engineering artificial riboswitches. *Chembiochem, 9*(12), 1906-1911. doi:10.1002/cbic.200700713

Gebetsberger, J., & Micura, R. (2017). Unwinding the twister ribozyme: from structure to mechanism. *Wiley Interdiscip Rev RNA, 8*(3). doi:10.1002/wrna.1402

Goldfless, S. J., Belmont, B. J., de Paz, A. M., Liu, J. F., & Niles, J. C. (2012). Direct and specific chemical control of eukaryotic translation with a synthetic RNA-protein interaction. *Nucleic Acids Res, 40*(9), e64. doi:10.1093/nar/gks028

Green, A. A., Silver, P. A., Collins, J. J., & Yin, P. (2014). Toehold switches: de-novo-designed regulators of gene expression. *Cell, 159*(4), 925-939. doi:10.1016/j.cell.2014.10.002

Groher, F., Bofill-Bosch, C., Schneider, C., Braun, J., Jager, S., Geissler, K., . . . Suess, B. (2018). Riboswitching with ciprofloxacin-development and characterization of a novel RNA regulator. *Nucleic Acids Res, 46*(4), 2121-2132. doi:10.1093/nar/gkx1319

Hsu, H. T., Lin, Y. H., & Chang, K. Y. (2014). Synergetic regulation of translational reading-frame switch by ligand-responsive RNAs in mammalian cells. *Nucleic Acids Res, 42*(22), 14070-14082. doi:10.1093/nar/gku1233

Kallunki, T., Barisic, M., Jaattela, M., & Liu, B. (2019). How to Choose the Right Inducible Gene Expression System for Mammalian Studies? *Cells, 8*(8). doi:10.3390/cells8080796

Karasawa, S., Araki, T., Yamamoto-Hino, M., & Miyawaki, A. (2003). A green-emitting fluorescent protein from Galaxeidae coral and its monomeric version for use in fluorescent labeling. *J Biol Chem, 278*(36), 34167-34171. doi:10.1074/jbc.M304063200

Kashida, S., Inoue, T., & Saito, H. (2012). Three-dimensionally designed protein-responsive RNA devices for cell signaling regulation. *Nucleic Acids Res, 40*(18), 9369-9378. doi:10.1093/nar/gks668

Kawasaki, S., Ono, H., Hirosawa, M., & Saito, H. (2020). RNA and protein-based nanodevices for mammalian post-transcriptional circuits. *Curr Opin Biotechnol, 63*, 99-110. doi:10.1016/j.copbio.2019.11.019

Kennedy, A. B., Liang, J. C., & Smolke, C. D. (2013). A versatile cis-blocking and trans-activation strategy for ribozyme characterization. *Nucleic Acids Res, 41*(2), e41. doi:10.1093/nar/gks1036

Kennedy, A. B., Vowles, J. V., d'Espaux, L., & Smolke, C. D. (2014). Protein-responsive ribozyme switches in eukaryotic cells. *Nucleic Acids Res, 42*(19), 12306-12321. doi:10.1093/nar/gku875

Ketzer, P., Haas, S. F., Engelhardt, S., Hartig, J. S., & Nettelbeck, D. M. (2012). Synthetic riboswitches for external regulation of genes transferred by replication-deficient and oncolytic adenoviruses. *Nucleic Acids Res, 40*(21), e167. doi:10.1093/nar/gks734

Ketzer, P., Kaufmann, J. K., Engelhardt, S., Bossow, S., von Kalle, C., Hartig, J. S., . . . Nettelbeck, D. M. (2014). Artificial riboswitches for gene expression and replication control of DNA and RNA viruses. *Proc Natl Acad Sci U S A, 111*(5), E554-562. doi:10.1073/pnas.1318563111

Kibbe, W. A. (2007). OligoCalc: an online oligonucleotide properties calculator. *Nucleic Acids Res, 35*(Web Server issue), W43-46. doi:10.1093/nar/gkm234

Kobori, S., Nomura, Y., Miu, A., & Yokobayashi, Y. (2015). High-throughput assay and engineering of self-cleaving ribozymes by sequencing. *Nucleic Acids Res, 43*(13), e85. doi:10.1093/nar/gkv265

Kobori, S., Takahashi, K., & Yokobayashi, Y. (2017). Deep Sequencing Analysis of Aptazyme Variants Based on a Pistol Ribozyme. *ACS Synth Biol, 6*(7), 1283-1288. doi:10.1021/acssynbio.7b00057

Kobori, S., & Yokobayashi, Y. (2018). Analyzing and Tuning Ribozyme Activity by Deep Sequencing To Modulate Gene Expression Level in Mammalian Cells. *ACS Synth Biol, 7*(2), 371-376. doi:10.1021/acssynbio.7b00367

Kruger, K., Grabowski, P. J., Zaug, A. J., Sands, J., Gottschling, D. E., & Cech, T. R. (1982). Self-splicing RNA: autoexcision and autocyclization of the ribosomal RNA intervening sequence of Tetrahymena. *Cell, 31*(1), 147-157. doi:10.1016/0092-8674(82)90414-7

Kumar, D., An, C. I., & Yokobayashi, Y. (2009). Conditional RNA interference mediated by allosteric ribozyme. *J Am Chem Soc, 131*(39), 13906-13907. doi:10.1021/ja905596t

Leamy, K. A., Assmann, S. M., Mathews, D. H., & Bevilacqua, P. C. (2016). Bridging the gap between in vitro and in vivo RNA folding. *Q Rev Biophys, 49*, e10. doi:10.1017/S003358351600007X

Li, N., & Huang, F. (2005). Ribozyme-catalyzed aminoacylation from CoA thioesters. *Biochemistry, 44*(11), 4582-4590. doi:10.1021/bi047576b

Liang, J. C., Chang, A. L., Kennedy, A. B., & Smolke, C. D. (2012). A high-throughput, quantitative cell-based screen for efficient tailoring of RNA device activity. *Nucleic Acids Res, 40*(20), e154. doi:10.1093/nar/gks636

Lienert, F., Lohmueller, J. J., Garg, A., & Silver, P. A. (2014). Synthetic biology in mammalian cells: next generation research tools and therapeutics. *Nat Rev Mol Cell Biol, 15*(2), 95-107. doi:10.1038/nrm3738

Lin, Y. H., & Chang, K. Y. (2016). Rational design of a synthetic mammalian riboswitch as a ligand-responsive -1 ribosomal frame-shifting stimulator. *Nucleic Acids Res, 44*(18), 9005-9015. doi:10.1093/nar/gkw718

Litke, J. L., & Jaffrey, S. R. (2019). Highly efficient expression of circular RNA aptamers in cells using autocatalytic transcripts. *Nat Biotechnol, 37*(6), 667-675. doi:10.1038/s41587-019-0090-6

Liu, Y., Wilson, T. J., McPhee, S. A., & Lilley, D. M. (2014). Crystal structure and mechanistic investigation of the twister ribozyme. *Nat Chem Biol, 10*(9), 739-744. doi:10.1038/nchembio.1587

Lynch, S. A., & Gallivan, J. P. (2009). A flow cytometry-based screen for synthetic riboswitches. *Nucleic Acids Res, 37*(1), 184-192. doi:10.1093/nar/gkn924

Mandal, M., Boese, B., Barrick, J. E., Winkler, W. C., & Breaker, R. R. (2003). Riboswitches control fundamental biochemical pathways in Bacillus subtilis and other bacteria. *Cell, 113*(5), 577-586. doi:10.1016/s0092-8674(03)00391-x

Matsumoto, S., Caliskan, N., Rodnina, M. V., Murata, A., & Nakatani, K. (2018). Small synthetic molecule-stabilized RNA pseudoknot as an activator for -1 ribosomal frameshifting. *Nucleic Acids Res, 46*(16), 8079-8089. doi:10.1093/nar/gky689

Miki, K., Endo, K., Takahashi, S., Funakoshi, S., Takei, I., Katayama, S., . . . Yoshida, Y. (2015). Efficient Detection and Purification of Cell Populations Using Synthetic MicroRNA Switches. *Cell Stem Cell, 16*(6), 699-711. doi:10.1016/j.stem.2015.04.005

Mou, H., Zhong, G., Gardner, M. R., Wang, H., Wang, Y. W., Cheng, D., & Farzan, M. (2018). Conditional Regulation of Gene Expression by Ligand-Induced Occlusion of a MicroRNA Target Sequence. *Mol Ther, 26*(5), 1277-1286. doi:10.1016/j.ymthe.2018.02.021

Muranaka, N., Sharma, V., Nomura, Y., & Yokobayashi, Y. (2009). An efficient platform for genetic selection and screening of gene switches in Escherichia coli. *Nucleic Acids Res, 37*(5), e39. doi:10.1093/nar/gkp039

Mustafina, K., Fukunaga, K., & Yokobayashi, Y. (2020). Design of Mammalian ON-Riboswitches Based on Tandemly Fused Aptamer and Ribozyme. *ACS Synth Biol, 9*(1), 19-25. doi:10.1021/acssynbio.9b00371

Nakano, S. I., & Sugimoto, N. (2016). Model studies of the effects of intracellular crowding on nucleic acid interactions. *Mol Biosyst, 13*(1), 32-41. doi:10.1039/c6mb00654j

Nguyen, L. A., Wang, J., & Steitz, T. A. (2017). Crystal structure of Pistol, a class of self-cleaving ribozyme. *Proc Natl Acad Sci U S A, 114*(5), 1021-1026. doi:10.1073/pnas.1611191114

Nomura, Y., Chien, H. C., & Yokobayashi, Y. (2017). Direct screening for ribozyme activity in mammalian cells. *Chem Commun (Camb), 53*(93), 12540-12543. doi:10.1039/c7cc07815c

Nomura, Y., Kumar, D., & Yokobayashi, Y. (2012). Synthetic mammalian riboswitches based on guanine aptazyme. *Chem Commun (Camb), 48*(57), 7215-7217. doi:10.1039/c2cc33140c

Nomura, Y., & Yokobayashi, Y. (2007). Dual selection of a genetic switch by a single selection marker. *Biosystems, 90*(1), 115-120. doi:10.1016/j.biosystems.2006.07.006

Nomura, Y., Zhou, L., Miu, A., & Yokobayashi, Y. (2013). Controlling mammalian gene expression by allosteric hepatitis delta virus ribozymes. *ACS Synth Biol, 2*(12), 684-689. doi:10.1021/sb400037a

Pardee, K., Green, A. A., Ferrante, T., Cameron, D. E., DaleyKeyser, A., Yin, P., & Collins, J. J. (2014). Paper-based synthetic gene networks. *Cell, 159*(4), 940-954. doi:10.1016/j.cell.2014.10.004

Pardee, K., Slomovic, S., Nguyen, P. Q., Lee, J. W., Donghia, N., Burrill, D., . . . Collins, J. J. (2016). Portable, On-Demand Biomolecular Manufacturing. *Cell, 167*(1), 248-259 e212. doi:10.1016/j.cell.2016.09.013

Park, S. V., Yang, J. S., Jo, H., Kang, B., Oh, S. S., & Jung, G. Y. (2019). Catalytic RNA, ribozyme, and its applications in synthetic biology. *Biotechnol Adv, 37*(8), 107452. doi:10.1016/j.biotechadv.2019.107452

Paudel, B. P., & Rueda, D. (2014). Molecular crowding accelerates ribozyme docking and catalysis. *J Am Chem Soc, 136*(48), 16700-16703. doi:10.1021/ja5073146

Priola, J. J., Calzadilla, N., Baumann, M., Borth, N., Tate, C. G., & Betenbaugh, M. J. (2016). High-throughput screening and selection of mammalian cells for enhanced protein production. *Biotechnol J, 11*(7), 853-865. doi:10.1002/biot.201500579

Reid, C. A., Nettesheim, E. R., Connor, T. B., & Lipinski, D. M. (2018). Development of an inducible anti-VEGF rAAV gene therapy strategy for the treatment of wet AMD. *Sci Rep, 8*(1), 11763. doi:10.1038/s41598-018-29726-7

Ren, A., Kosutic, M., Rajashankar, K. R., Frener, M., Santner, T., Westhof, E., . . . Patel, D. J. (2014). In-line alignment and Mg(2)(+) coordination at the cleavage site of the env22 twister ribozyme. *Nat Commun, 5*, 5534. doi:10.1038/ncomms6534

Ren, A., Vusurovic, N., Gebetsberger, J., Gao, P., Juen, M., Kreutz, C., . . . Patel, D. J. (2016). Pistol ribozyme adopts a pseudoknot fold facilitating site-specific in-line cleavage. *Nat Chem Biol, 12*(9), 702-708. doi:10.1038/nchembio.2125

Romani, A. M. P. (2011). Intracellular magnesium homeostasis. In R. Vink & M. Nechifor (Eds.), *Magnesium in the Central Nervous System*. Adelaide (AU).

Roth, A., Weinberg, Z., Chen, A. G., Kim, P. B., Ames, T. D., & Breaker, R. R. (2014). A widespread self-cleaving ribozyme class is revealed by bioinformatics. *Nat Chem Biol, 10*(1), 56-60. doi:10.1038/nchembio.1386

Saito, H., Kobayashi, T., Hara, T., Fujita, Y., Hayashi, K., Furushima, R., & Inoue, T. (2010). Synthetic translational regulation by an L7Ae-kink-turn RNP switch. *Nat Chem Biol, 6*(1), 71-78. doi:10.1038/nchembio.273

Shanidze, N., Lenkeit, F., Hartig, J. S., & Funck, D. (2020). A Theophylline-Responsive Riboswitch Regulates Expression of Nuclear-Encoded Genes. *Plant Physiol, 182*(1), 123-135. doi:10.1104/pp.19.00625

Shin, J., & Noireaux, V. (2010). Efficient cell-free expression with the endogenous E. Coli RNA polymerase and sigma factor 70. *J Biol Eng, 4*, 8. doi:10.1186/1754-1611-4-8

Sporing, M., Finke, M., & Hartig, J. S. (2020). Aptamers in RNA-based switches of gene expression. *Curr Opin Biotechnol, 63*, 34-40. doi:10.1016/j.copbio.2019.11.008
Stifel, J., Sporing, M., & Hartig, J. S. (2019). Expanding the toolbox of synthetic riboswitches with guanine-dependent aptazymes. *Synth Biol (Oxf), 4*(1), ysy022. doi:10.1093/synbio/ysy022
Stripecke, R., Oliveira, C. C., McCarthy, J. E., & Hentze, M. W. (1994). Proteins binding to 5' untranslated region sites: a general mechanism for translational regulation of mRNAs in human and yeast cells. *Mol Cell Biol, 14*(9), 5898-5909. doi:10.1128/mcb.14.9.5898
Strobel, B., Duchs, M. J., Blazevic, D., Rechtsteiner, P., Braun, C., Baum-Kroker, K. S., . . . Kreuz, S. (2020). A Small-Molecule-Responsive Riboswitch Enables Conditional Induction of Viral Vector-Mediated Gene Expression in Mice. *ACS Synth Biol, 9*(6), 1292-1305. doi:10.1021/acssynbio.9b00410
Strobel, B., Klauser, B., Hartig, J. S., Lamla, T., Gantner, F., & Kreuz, S. (2015). Riboswitch-mediated Attenuation of Transgene Cytotoxicity Increases Adeno-associated Virus Vector Yields in HEK-293 Cells. *Mol Ther, 23*(10), 1582-1591. doi:10.1038/mt.2015.123
Strobel, B., Sporing, M., Klein, H., Blazevic, D., Rust, W., Sayols, S., . . . Kreuz, S. (2020). High-throughput identification of synthetic riboswitches by barcode-free amplicon-sequencing in human cells. *Nat Commun, 11*(1), 714. doi:10.1038/s41467-020-14491-x
Takahashi, K., & Yokobayashi, Y. (2019). Reversible Gene Regulation in Mammalian Cells Using Riboswitch-Engineered Vesicular Stomatitis Virus Vector. *ACS Synth Biol, 8*(9), 1976-1982. doi:10.1021/acssynbio.9b00177
Tan, C., Saurabh, S., Bruchez, M. P., Schwartz, R., & LeDuc, P. (2014). Reply to 'Complexity of molecular crowding in cell-free enzymatic reaction networks'. *Nat Nanotechnol, 9*(6), 407-408. doi:10.1038/nnano.2014.111
Tang, J., & Breaker, R. R. (1997). Examination of the catalytic fitness of the hammerhead ribozyme by in vitro selection. *RNA, 3*(8), 914-925. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/9257650
Tang, W., Hu, J. H., & Liu, D. R. (2017). Aptazyme-embedded guide RNAs enable ligand-responsive genome editing and transcriptional activation. *Nat Commun, 8*, 15939. doi:10.1038/ncomms15939
Topp, S., & Gallivan, J. P. (2008). Random walks to synthetic riboswitches--a high-throughput selection based on cell motility. *Chembiochem, 9*(2), 210-213. doi:10.1002/cbic.200700546
Townshend, B., Kennedy, A. B., Xiang, J. S., & Smolke, C. D. (2015). High-throughput cellular RNA device engineering. *Nat Methods, 12*(10), 989-994. doi:10.1038/nmeth.3486
Traut, T. W. (1994). Physiological concentrations of purines and pyrimidines. *Mol Cell Biochem, 140*(1), 1-22. doi:10.1007/BF00928361
Tuerk, C., & Gold, L. (1990). Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science, 249*(4968), 505-510. doi:10.1126/science.2200121
Valencia-Sanchez, M. A., Liu, J., Hannon, G. J., & Parker, R. (2006). Control of translation and mRNA degradation by miRNAs and siRNAs. *Genes Dev, 20*(5), 515-524. doi:10.1101/gad.1399806
Venkata Subbaiah, K. C., Hedaya, O., Wu, J., Jiang, F., & Yao, P. (2019). Mammalian RNA switches: Molecular rheostats in gene regulation, disease, and medicine. *Comput Struct Biotechnol J, 17*, 1326-1338. doi:10.1016/j.csbj.2019.10.001

Vogel, M., Weigand, J. E., Kluge, B., Grez, M., & Suess, B. (2018). A small, portable RNA device for the control of exon skipping in mammalian cells. *Nucleic Acids Res, 46*(8), e48. doi:10.1093/nar/gky062

Wang, S., Emery, N. J., & Liu, A. P. (2019). A Novel Synthetic Toehold Switch for MicroRNA Detection in Mammalian Cells. *ACS Synth Biol, 8*(5), 1079-1088. doi:10.1021/acssynbio.8b00530

Wei, K. Y., Chen, Y. Y., & Smolke, C. D. (2013). A yeast-based rapid prototype platform for gene control elements in mammalian cells. *Biotechnol Bioeng, 110*(4), 1201-1210. doi:10.1002/bit.24792

Wei, K. Y., & Smolke, C. D. (2015). Engineering dynamic cell cycle control with synthetic small molecule-responsive RNA devices. *J Biol Eng, 9*, 21. doi:10.1186/s13036-015-0019-7

Weinberg, C. E., Weinberg, Z., & Hammann, C. (2019). Novel ribozymes: discovery, catalytic mechanisms, and the quest to understand biological function. *Nucleic Acids Res, 47*(18), 9480-9494. doi:10.1093/nar/gkz737

Weinberg, Z., Kim, P. B., Chen, T. H., Li, S., Harris, K. A., Lunse, C. E., & Breaker, R. R. (2015). New classes of self-cleaving ribozymes revealed by comparative genomics analysis. *Nat Chem Biol, 11*(8), 606-610. doi:10.1038/nchembio.1846

Werstuck, G., & Green, M. R. (1998). Controlling gene expression in living cells through small molecule-RNA interactions. *Science, 282*(5387), 296-298. doi:10.1126/science.282.5387.296

Wieland, M., & Hartig, J. S. (2008). Improved aptazyme design and in vivo screening enable riboswitching in bacteria. *Angew Chem Int Ed Engl, 47*(14), 2604-2607. doi:10.1002/anie.200703700

Win, M. N., & Smolke, C. D. (2007). A modular and extensible RNA-based gene-regulatory platform for engineering cellular function. *Proc Natl Acad Sci U S A, 104*(36), 14283-14288. doi:10.1073/pnas.0703961104

Windholz, M. (1976). *The Merck index : an encyclopedia of chemicals and drugs* (9th ed.). Rahway, N.J.: Merck.

Wittmann, A., & Suess, B. (2011). Selection of tetracycline inducible self-cleaving ribozymes as synthetic devices for gene regulation in yeast. *Mol Biosyst, 7*(8), 2419-2427. doi:10.1039/c1mb05070b

Wurmthaler, L. A., Klauser, B., & Hartig, J. S. (2018). Highly motif- and organism-dependent effects of naturally occurring hammerhead ribozyme sequences on gene expression. *RNA Biol, 15*(2), 231-241. doi:10.1080/15476286.2017.1397870

Wurmthaler, L. A., Sack, M., Gense, K., Hartig, J. S., & Gamerdinger, M. (2019). A tetracycline-dependent ribozyme switch allows conditional induction of gene expression in Caenorhabditis elegans. *Nat Commun, 10*(1), 491. doi:10.1038/s41467-019-08412-w

Xiang, J. S., Kaplan, M., Dykstra, P., Hinks, M., McKeague, M., & Smolke, C. D. (2019). Massively parallel RNA device engineering in mammalian cells with RNA-Seq. *Nat Commun, 10*(1), 4327. doi:10.1038/s41467-019-12334-y

Xie, Z., Wroblewska, L., Prochazka, L., Weiss, R., & Benenson, Y. (2011). Multi-input RNAi-based logic circuit for identification of specific cancer cells. *Science, 333*(6047), 1307-1311. doi:10.1126/science.1205527

Yen, L., Svendsen, J., Lee, J. S., Gray, J. T., Magnier, M., Baba, T., . . . Mulligan, R. C. (2004). Exogenous control of mammalian gene expression through modulation of RNA self-cleavage. *Nature, 431*(7007), 471-476. doi:10.1038/nature02844

Yokobayashi, Y. (2019). Aptamer-based and aptazyme-based riboswitches in mammalian cells. *Curr Opin Chem Biol, 52*, 72-78. doi:10.1016/j.cbpa.2019.05.018

Zhong, G., Wang, H., Bailey, C. C., Gao, G., & Farzan, M. (2016). Rational design of aptazyme riboswitches for efficient control of gene expression in mammalian cells. *Elife, 5*. doi:10.7554/eLife.18858

Zuker, M. (2003). Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res, 31*(13), 3406-3415. doi:10.1093/nar/gkg595

Appendix

DNA sequence of pEGFP-N1

EGFP-coding sequence is shown in green. XbaI and BspeI sites between which most of the aptazymes were inserted are indicated in red. MfeI site is shown in blue

>pEGFP-N1
TAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCCATATATGGAGTTCCGCGTTACATAACTTAC
GGTAAATGGCCCGCCTGGCTGACCGCCCAACGACCCCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAGT
AACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCA
AGTGTATCATATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTA
CATGACCTTATGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTT
TTGGCAGTACATCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCCCATTGACGTCAA
TGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAACAACTCCGCCCCATTGACGCAAAT
GGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTGGTTTAGTGAACCGTCAGATCCGCTAGCGCTA
CCGGACTCAGATCTCGAGCTCAAGCTTCGAATTCTGCAGTCGACGGTACCGCGGGCCCGGGATCCACCGGTCGCC
ACC**ATGGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAAC**
GGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTCATCTGC
ACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCACCCTGACCTACGGCGTGCAGTGCTTCAGCCGC
TACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACCATC
TTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATC
GAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACTACAACAGCCAC
AACGTCTATATCATGGCCGACAAGCAGAAGAACGGCATCAAGGTGAACTTCAAGATCCGCCACAACATCGAGGAC
GGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCCGACAAC
CACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTGGAGTTC
GTGACCGCCGCCGGGATCACTCTCGGCATGGACGAGCTGTACAAGTAAAGCGGCCGCGAC**TCTAGA**CC**TCCGGA**G
CCATACCACATTTGTAGAGGTTTTACTTGCTTTAAAAAACCTCCCACACCTCCCCCTGAACCTGAAACATAAAAT
GAATG**CAATTG**TTGTTGTTAACTTGTTTATTGCAGCTTATAATGGTTACAAATAAAGCAATAGCATCACAAATTT
CACAAATAAAGCATTTTTTTCACTGCATTCTAGTTGTGGTTTGTCCAAACTCATCAATGTATCTTAAGGCGTAAA
TTGTAAGCGTTAATATTTTGTTAAAATTCGCGTTAAATTTTTGTTAAATCAGCTCATTTTTTAACCAATAGGCCG
AAATCGGCAAAATCCCTTATAAATCAAAAGAATAGACCGAGATAGGGTTGAGTGTTGTTCCAGTTTGGAACAAGA
GTCCACTATTAAAGAACGTGGACTCCAACGTCAAAGGGCGAAAAACCGTCTATCAGGGCGATGGCCCACTACGTG
AACCATCACCCTAATCAAGTTTTTTGGGGTCGAGGTGCCGTAAAGCACTAAATCGGAACCCTAAAGGGAGCCCCC
GATTTAGAGCTTGACGGGGAAAGCCGGCGAACGTGGCGAGAAAGGAAGGGAAGAAAGCGAAAGGAGCGGGCGCTA
GGGCGCTGGCAAGTGTAGCGGTCACGCTGCGCGTAACCACCACACCCGCCGCGCTTAATGCGCCGCTACAGGGCG
CGTCAGGTGGCACTTTTCGGGGAAATGTGCGCGGAACCCCTATTTGTTTATTTTTCTAAATACATTCAAATATGT
ATCCGCTCATGAGACAATAACCCTGATAAATGCTTCAATAATATTGAAAAAGGAAGAGTCCTGAGGCGGAAAGAA
CCAGCTGTGGAATGTGTGTCAGTTAGGGTGTGGAAAGTCCCCAGGCTCCCCAGCAGGCAGAAGTATGCAAAGCAT
GCATCTCAATTAGTCAGCAACCAGGTGTGGAAAGTCCCCAGGCTCCCCAGCAGGCAGAAGTATGCAAAGCATGCA
TCTCAATTAGTCAGCAACCATAGTCCCGCCCCTAACTCCGCCCATCCCGCCCCTAACTCCGCCCAGTTCCGCCCA
TTCTCCGCCCCATGGCTGACTAATTTTTTTTATTTATGCAGAGGCCGAGGCCGCCTCGGCCTCTGAGCTATTCCA
GAAGTAGTGAGGAGGCTTTTTTGGAGGCCTAGGCTTTTGCAAAGATCGATCAAGAGACAGGATGAGGATCGTTTC
GCATGATTGAACAAGATGGATTGCACGCAGGTTCTCCGGCCGCTTGGGTGGAGAGGCTATTCGGCTATGACTGGG
CACAACAGACAATCGGCTGCTCTGATGCCGCCGTGTTCCGGCTGTCAGCGCAGGGGCGCCCGGTTCTTTTTGTCA
AGACCGACCTGTCCGGTGCCCTGAATGAACTGCAAGACGAGGCAGCGCGGCTATCGTGGCTGGCCACGACGGGCG
TTCCTTGCGCAGCTGTGCTCGACGTTGTCACTGAAGCGGGAAGGGACTGGCTGCTATTGGGCGAAGTGCCGGGGC
AGGATCTCCTGTCATCTCACCTTGCTCCTGCCGAGAAAGTATCCATCATGGCTGATGCAATGCGGCGGCTGCATA
CGCTTGATCCGGCTACCTGCCCATTCGACCACCAAGCGAAACATCGCATCGAGCGAGCACGTACTCGGATGGAAG
CCGGTCTTGTCGATCAGGATGATCTGGACGAAGAGCATCAGGGGCTCGCGCCAGCCGAACTGTTCGCCAGGCTCA
AGGCGAGCATGCCCGACGGCGAGGATCTCGTCGTGACCCATGGCGATGCCTGCTTGCCGAATATCATGGTGGAAA
ATGGCCGCTTTTCTGGATTCATCGACTGTGGCCGGCTGGGTGTGGCGGACCGCTATCAGGACATAGCGTTGGCTA
CCCGTGATATTGCTGAAGAGCTTGGCGGCGAATGGGCTGACCGCTTCCTCGTGCTTTACGGTATCGCCGCTCCCG
ATTCGCAGCGCATCGCCTTCTATCGCCTTCTTGACGAGTTCTTCTGAGCGGGACTCTGGGGTTCGAAATGACCGA
CCAAGCGACGCCCAACCTGCCATCACGAGATTTCGATTCCACCGCCGCCTTCTATGAAAGGTTGGGCTTCGGAAT
CGTTTTCCGGGACGCCGGCTGGATGATCCTCCAGCGCGGGGATCTCATGCTGGAGTTCTTCGCCCACCCTAGGGG
GAGGCTAACTGAAACACGGAAGGAGACAATACCGGAAGGAACCCGCGCTATGACGGCAATAAAAAGACAGAATAA
AACGCACGGTGTTGGGTCGTTTGTTCATAAACGCGGGGTTCGGTCCCAGGGCTGGCACTCTGTCGATACCCCACC

GAGACCCCATTGGGGCCAATACGCCCGCGTTTCTTCCTTTTCCCCACCCCACCCCCCAAGTTCGGGTGAAGGCCC
AGGGCTCGCAGCCAACGTCGGGGCGGCAGGCCCTGCCATAGCCTCAGGTTACTCATATATACTTTAGATTGATTT
AAAACTTCATTTTTAATTTAAAAGGATCTAGGTGAAGATCCTTTTTGATAATCTCATGACCAAAATCCCTTAACG
TGAGTTTTCGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCG
CGTAATCTGCTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGCTACCAAC
TCTTTTTCCGAAGGTAACTGGCTTCAGCAGAGCGCAGATACCAAATACTGTCCTTCTAGTGTAGCCGTAGTTAGG
CCACCACTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAG
TGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCGGGCTGAAC
GGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAACTGAGATACCTACAGCGTGAGCTATG
AGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCG
CACGAGGGAGCTTCCAGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCG
TCGATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAAACGCCAGCAACGCGGCCTTTTTACGGTTCCT
GGCCTTTTGCTGGCCTTTTGCTCACATGTTCTTTCCTGCGTTATCCCCTGATTCTGTGGATAACCGTATTACCGC
CATGCAT

DNA sequence of phmAG1-MC1

AG-coding sequence is shown in green. NotI and MfeI sites between which the aptazymes were inserted are indicated in red.

> phmAG1-MC1

ATGGTGAGCGTGATCAAGCCCGAGATGAAGATCAAGCTGTGCATGAGGGGCACCGTGAACGGCCACAACTTCGTG
ATCGAGGGCGAGGGCAAGGGCAACCCCTACGAGGGCACCCAGATCCTGGACCTGAACGTGACCGAGGGCGCCCCC
CTGCCCTTCGCCTACGACATCCTGACCACCGTGTTCCAGTACGGCAACAGGGCCTTCACCAAGTACCCCGCCGAC
ATCCAGGACTACTTCAAGCAGACCTTCCCCGAGGGCTACCACTGGGAGAGGAGCATGACCTACGAGGACCAGGGC
ATCTGCACCGCCACCAGCAACATCAGCATGAGGGGCGACTGCTTCTTCTACGACATCAGGTTCGACGGCACCAAC
TTCCCCCCCAACGGCCCCGTGATGCAGAAGAAGACCCTGAAGTGGGAGCCCAGCACCGAGAAGATGTACGTGGAG
GACGGCGTGCTGAAGGGCGACGTGAACATGAGGCTGCTGCTGGAGGGCGGCGGCCACTACAGGTGCGACTTCAAG
ACCACCTACAAGGCCAAGAAGGAGGTGAGGCTGCCCGACGCCCACAAGATCGACCACAGGATCGAGATCCTGAAG
CACGACAAGGACTACAACAAGGTGAAGCTGTACGAGAACGCCGTGGCCAGGTACTCCATGCTGCCCAGCCAGGCC
AAGGGATCCTCAGGTACCGGAACTGCAGCAGAGAATTCGGGAAACTCGAGAACAAAGCTTGAATAA**GCGGCCGC**G
ACTCTAGATCATAATCAGCCATACCACATTTGTAGAGGTTTTACTTGCTTTAAAAAACCTCCCACACCTCCCCCT
GAACCTGAAACATAAAATGAATG**CAATTG**TTGTTGTTAACTTGTTTATTGCAGCTTATAATGGTTACAAATAAAG
CAATAGCATCACAAATTTCACAAATAAAGCATTTTTTTCACTGCATTCTAGTTGTGGTTTGTCCAAACTCATCAA
TGTATCTTAAGGCGTAAATTGTAAGCGTTAATATTTTGTTAAAATTCGCGTTAAATTTTTGTTAAATCAGCTCAT
TTTTTAACCAATAGGCCGAAATCGGCAAAATCCCTTATAAATCAAAAGAATAGACCGAGATAGGGTTGAGTGTTG
TTCCAGTTTGGAACAAGAGTCCACTATTAAAGAACGTGGACTCCAACGTCAAAGGGCGAAAAACCGTCTATCAGG
GCGATGGCCCACTACGTGAACCATCACCCTAATCAAGTTTTTTGGGGTCGAGGTGCCGTAAAGCACTAAATCGGA
ACCCTAAAGGGAGCCCCCGATTTAGAGCTTGACGGGGAAAGCCGGCGAACGTGGCGAGAAAGGAAGGGAAGAAAG
CGAAAGGAGCGGGCGCTAGGGCGCTGGCAAGTGTAGCGGTCACGCTGCGCGTAACCACCACACCCGCCGCGCTTA
ATGCGCCGCTACAGGGCGCGTCAGGTGGCACTTTTCGGGGAAATGTGCGCGGAACCCCTATTTGTTTATTTTTCT
AAATACATTCAAATATGTATCCGCTCATGAGACAATAACCCTGATAAATGCTTCAATAATATTGAAAAAGGAAGA
GTCCTGAGGCGGAAAGAACCAGCTGTGGAATGTGTGTCAGTTAGGGTGTGGAAAGTCCCCAGGCTCCCCAGCAGG
CAGAAGTATGCAAAGCATGCATCTCAATTAGTCAGCAACCAGGTGTGGAAAGTCCCCAGGCTCCCCAGCAGGCAG
AAGTATGCAAAGCATGCATCTCAATTAGTCAGCAACCATAGTCCCGCCCCTAACTCCGCCCATCCCGCCCCTAAC
TCCGCCCAGTTCCGCCCATTCTCCGCCCCATGGCTGACTAATTTTTTTTATTTATGCAGAGGCCGAGGCCGCCTC
GGCCTCTGAGCTATTCCAGAAGTAGTGAGGAGGCTTTTTTGGAGGCCTAGGCTTTTGCAAAGATCGATCAAGAGA
CAGGATGAGGATCGTTTCGCATGATTGAACAAGATGGATTGCACGCAGGTTCTCCGGCCGCTTGGGTGGAGAGGC
TATTCGGCTATGACTGGGCACAACAGACAATCGGCTGCTCTGATGCCGCCGTGTTCCGGCTGTCAGCGCAGGGGC
GCCCGGTTCTTTTTGTCAAGACCGACCTGTCCGGTGCCCTGAATGAACTGCAAGACGAGGCAGCGCGGCTATCGT
GGCTGGCCACGACGGGCGTTCCTTGCGCAGCTGTGCTCGACGTTGTCACTGAAGCGGGAAGGGACTGGCTGCTAT
TGGGCGAAGTGCCGGGGCAGGATCTCCTGTCATCTCACCTTGCTCCTGCCGAGAAAGTATCCATCATGGCTGATG
CAATGCGGCGGCTGCATACGCTTGATCCGGCTACCTGCCCATTCGACCACCAAGCGAAACATCGCATCGAGCGAG
CACGTACTCGGATGGAAGCCGGTCTTGTCGATCAGGATGATCTGGACGAAGAGCATCAGGGGCTCGCGCCAGCCG
AACTGTTCGCCAGGCTCAAGGCGAGCATGCCCGACGGCGAGGATCTCGTCGTGACCCATGGCGATGCCTGCTTGC
CGAATATCATGGTGGAAAATGGCCGCTTTTCTGGATTCATCGACTGTGGCCGGCTGGGTGTGGCGGACCGCTATC
AGGACATAGCGTTGGCTACCCGTGATATTGCTGAAGAGCTTGGCGGCGAATGGGCTGACCGCTTCCTCGTGCTTT
ACGGTATCGCCGCTCCCGATTCGCAGCGCATCGCCTTCTATCGCCTTCTTGACGAGTTCTTCTGAGCGGGACTCT
GGGGTTCGAAATGACCGACCAAGCGACGCCCAACCTGCCATCACGAGATTTCGATTCCACCGCCGCCTTCTATGA
AAGGTTGGGCTTCGGAATCGTTTTCCGGGACGCCGGCTGGATGATCCTCCAGCGCGGGGATCTCATGCTGGAGTT

CTTCGCCCACCCTAGGGGGAGGCTAACTGAAACACGGAAGGAGACAATACCGGAAGGAACCCGCGCTATGACGGC
AATAAAAAGACAGAATAAAACGCACGGTGTTGGGTCGTTTGTTCATAAACGCGGGGTTCGGTCCCAGGGCTGGCA
CTCTGTCGATACCCCACCGAGACCCCATTGGGGCCAATACGCCCGCGTTTCTTCCTTTTCCCCACCCCACCCCCC
AAGTTCGGGTGAAGGCCCAGGGCTCGCAGCCAACGTCGGGGCGGCAGGCCCTGCCATAGCCTCAGGTTACTCATA
TATACTTTAGATTGATTTAAAACTTCATTTTTAATTTAAAAGGATCTAGGTGAAGATCCTTTTTGATAATCTCAT
GACCAAAATCCCTTAACGTGAGTTTTCGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTG
AGATCCTTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTTGCC
GGATCAAGAGCTACCAACTCTTTTTCCGAAGGTAACTGGCTTCAGCAGAGCGCAGATACCAAATACTGTCCTTCT
AGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTT
ACCAGTGGCTGCTGCCAGTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGC
GCAGCGGTCGGGCTGAACGGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAACTGAGATA
CCTACAGCGTGAGCTATGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCGGCAG
GGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTTTCG
CCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAAACGCCAGCAACGC
GGCCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCACATGTTCTTTCCTGCGTTATCCCCTGATTCTGT
GGATAACCGTATTACCGCCATGCATTAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCCATATA
TGGAGTTCCGCGTTACATAACTTACGGTAAATGGCCCGCCTGGCTGACCGCCCAACGACCCCCGCCCATTGACGT
CAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGT
AAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTAAAT
GGCCCGCCTGGCATTATGCCCAGTACATGACCTTATGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCA
TCGCTATTACCATGGTGATGCGGTTTTGGCAGTACATCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTC
CAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTA
ACAACTCCGCCCCATTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTGGTTTAG
TGAACCGTCAGATCCGCTAGCATTGCCACC