UC Irvine
UC Irvine Electronic Theses and Dissertations

Title

Advances in discovery and characterization of small functional RNAs

Permalink

https://escholarship.org/uc/item/554570pg

Author

Moreira Passalacqua, Luiz Fernando

Publication Date

2019

Copyright Information

This work is made available under the terms of a Creative Commons Attribution License, availalbe at https://creativecommons.org/licenses/by/4.0/

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

UNIVERSITY OF CALIFORNIA,
IRVINE

Advances in discovery and characterization of small functional RNAs

DISSERTATION

submitted in partial satisfaction of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in Pharmacological Sciences

by

Luiz Fernando Moreira Passalacqua

Dissertation Committee:
Professor Andrej Lupták, Chair
Professor Klemens J. Hertel
Associate Professor Robert S. Spitale

2019

Chapter 1 ⓒ 2015 Elsevier Ltd.
Chapter 3 ⓒ 2017 American Chemical Society
All other materials ⓒ 2019 Luiz Fernando Moreira Passalacqua

DEDICATION

To

my wife and son, Aline and Nicolas,
my parents, Francisco Luciano and Marisa, and my brother, Luciano
for their endless support and love.

"God writes straight with crooked lines."

— a Portuguese proverb

TABLE OF CONTENTS

Page

LIST OF FIGURES v

LIST OF TABLES viii

ACKNOWLEDGMENTS ix

CURRICULUM VITAE x

ABSTRACT OF THE DISSERTATION xiv

CHAPTER 1: Introduction 1
- 1.1 Background 1
- 1.2 *In vitro* Selection 2
- 1.2 Self-cleaving ribozymes 6
- 1.3 Riboswitches 12
- 1.5 References 16

CHAPTER 2: Large phenotypic enhancement of structured random RNA pools 22
- 2.1 Abstract 23
- 2.2 Introduction 24
- 2.3 Results 25
- 2.4 Discussion 50
- 2.5 Conclusion 56
- 2.6 Acknowledgements 58
- 2.7 Materials and Methods 58
- 2.8 Supplementary Data 79
- 2.9 References 91

CHAPTER 3: Allosteric Modulation of Bacterial Ribozyme 95
- 3.1 Abstract 96
- 3.2 Introduction 96
- 3.3 Results 101
- 3.4 Discussion 113
- 3.5 Conclusion 118
- 3.6 Materials and Methods 119
- 3.7 Supplementary Data 122
- 3.7 References 127

CHAPTER 4: Co–transcriptional Analysis of Self–cleaving Ribozymes and Their Ligand Dependence 132
- 4.1 Abstract 132

4.2 Introduction 133
4.3 Materials 136
4.4 Methods 138
4.5 Notes 148
4.6 References 149

CHAPTER 5: Single Pass Transcription using T7 RNA polymerase 152
5.1 Introduction 152
5.2 Results 158
5.3 Discussion and Conclusion 167
5.4 Materials and Methods 171
5.5 References 173

CHAPTER 6: Regulation of mRNA translation by a photoriboswitch 177
6.1 Abstract 177
6.2 Introduction 178
6.3 Results and Discussion 180
6.4 Conclusion 194
6.5 Materials and Methods 195
6.8 Supplementary Data 209
6.9 References 228

LIST OF FIGURES

		Page
Figure 1.1	Schematic representation of *in vitro* selection	4
Figure 1.2	Mechanism of self–scission	7
Figure 1.3	Overview of the HDV ribozyme	11
Figure 1.4	Schematic representations of transcriptional and translation regulation of gene expression by riboswitches	14
Figure 2.1	Folding propensities of random-sequence RNA pools	28
Figure 2.2	Biochemical characterization of the monomeric RNA pools	36
Figure 2.3	*In vitro* selections from different compositions of starting pools	42
Figure 2.4	A highly abundant aptamer motif from the Ni-NTA selections	45
Figure 2.5	Characterization of dominant sequences form the *in vitro* selections for ATP-reacting RNAs	48
Figure S2.1	*In silico* and native-PAGE analysis of RNA pools	79
Figure S2.2	Solvent accessibility analysis of pools A, N5, and N6 assessed by 2′-acylation	81
Figure S2.3	Apta-Seq analysis of the top clusters from the Oregon Green selection	82
Figure 3.1	HDV–like ribozymes in *F. prausnitzii*	99
Figure 3.2	Influence of GlcN6P on the co–transcriptional self–scission of drz–Fpra–1 and drz–Fpra–2	103
Figure 3.3	Effect of GlcN6P and Mg^{2+} on the drz–Fpra ribozymes	105
Figure 3.4	Structural probing of GlcN6P binding of the drz–Fpra–1 ribozyme	108
Figure 3.5	Effect of GlcN6P on self–scission of drz–Fpra–1/2 hybrids and HDV ribozymes	112
Figure S3.1	Drz–Fpra–1 and drz–Fpra–2 self-cleavage activity	122

Figure S3.2 Drz–Fpra–1 and drz–Fpra–2 self-scission in the presence of UDP–GlcNAc 123

Figure S3.3 Effect of GlcN1P combined with GlcN6P on the activity of the drz–Fpra–1 ribozyme 123

Figure S3.4 Structural probing of GlcN1P binding of the drz–Fpra–1 ribozyme 124

Figure S3.5 Secondary structure alignment of drz–Fpra–1 and drz–Fpra–2 with HDV–like ribozymes mapping near glmM genes reported previouly 124

Figure 4.1 Resolved denaturing (urea) PAGE gel of co–transcriptional self–scission of the drz–Fpra–1 ribozyme 142

Figure 4.2 Data analysis of co–transcriptional kinetics using MS Excel 146

Figure 5.1 Illustration scheme of proposed single pass transcription 156

Figure 5.2 Cleavage of the promoter region by *HinfI* 160

Figure 5.3 Optimization of *HinfI* incubation time 162

Figure 5.4 Transcription analysis of drz–Fpra–2 165

Figure 6.1 An amino-*t*SS-responsive aptamer 182

Figure 6.2 Translation regulation by the Were-1 riboswitch 188

Figure 6.3 Regulation of luciferase expression by the Were-1 photoriboswitch *in vivo* 192

Figure S6.1 *In vitro* selection pool design for a photoriboswitch 209

Figure S6.2 E/Z isomerization of amino-*t*SS and amino-*c*SS 210

Figure S6.3 Screening of Were-1 based on optical activity 213

Figure S6.4 Structural probing of Werewolf-1 214

Figure S6.5 Binding PAGE-purified Were-1 RNA to a ribosome-mimic 220

Figure S6.6 Co-transcriptional binding of a ribosome-mimic *in vitro* under various Mg^{2+} conditions 221

Figure S6.7 Co-transcriptional Were-1 binding of a ribosome-mimic *in vitro* in the presence of *t*SS 222

Figure S6.8 Translation of a control plasmid, Luc2-pET, lacking the Were-1 riboswitch is not inhibited *in vitro* by amino-*t*SS 222

Figure S6.9 The effect of the canonical start codon on Were-1-Fluc expression *in vivo* 223

Figure S6.10 Co-transcriptional binding of a ribosome-mimic *in vitro* under amino-*t*SS photoisomerization conditions 223

Figure S6.11 Were-1 regulates translation *in vitro* 224

Figure S6.12 Were-1 regulates protein expression *in vivo* 225

LIST OF TABLES

Page

Table 2.1 Estimated complexity and composition of *in vitro* selected pools 32

Table S2.1 DNA sequences used for project, calculated minimum free energy (MFE) for the top 10 clones, and prediction of G-quartet 83

Table S3.1 DNA sequences of oligonucleotides used in the study 126

Table S6.1 DNA sequences used for *in vitro* and *in vivo* analysis of the Were-1 riboswitch 226

ACKNOWLEDGMENTS

I would first like to express the deepest appreciation to my advisor and committee chair, Professor Andrej Lupták, for supporting and advising me through all these years. You are a great scientist and an amazing person. You were instrumental in shaping and sharping my knowledge and critical skills needed to succeed in science and life. Thank you for believing in me. I could not think of a better person to mentor me.

I would like to thank my committee members, Professor Klemens J. Hertel and Professor Robert C. Spitale, for the valuable advices on both research and life. I am also grateful for my rotation at the Spitale lab.

I also thank my early mentors, Professor Leoberto C. Tavares and Dr. Sibila Grallert for the inspiration and guidance. Your enthusiasm and attitude were crucial to my initial life in science.

All present and past members of the Lupták lab. It has been a fun and engaging journey, both inside and outside of our “windowless” laboratory. I also thank present and past members from the Spitale and Hertel labs for the continuous help and fruitful discussions.

I thank the Department of Pharmaceutical Sciences and the University of California, Irvine for all the support and kindness through these years. I also thank the University of Sao Paulo for providing solid years of knowledge. I thank the financial support provided by the University of California, Irvine, and CAPES – Brazil.

I also would like to thank my wonderful family and friends, from here, Brazil, and around the globe. Friendship is essential to the soul.

My parents, Francisco Luciano and Marisa, my brother, Luciano, for being role models in so many ways. Your support, love, and care are of vital importance. I love you!

My wife, Aline, for being there for me in all moments. Since the beginning I knew it was you. It will always be you. There are no words to express everything I feel for you. Our synergistic relationship makes us better. The real need to be together is in our hearts. Thank you for everything! And my son, Nicolas, for showing me the most pure and intense form of love. You make me a better person. I love you both! Always!

CURRICULUM VITAE

Luiz Fernando Moreira Passalacqua

Education:

2014 – 2019	Ph.D. Pharmacological Sciences, University of California, Irvine
2003 – 2008	B.S. Pharmacy and Biochemistry, University of Sao Paulo – Brazil
2006	Bishop's University, Sherbrook, Canada (exchange program)

Research and professional experience:

2014 – 2019 Graduate student researcher, Advisor: Professor Andrej Lupták
University of California, Irvine, Dept. of Pharmaceutical Sciences

Research focused on small functional RNAs, particularly on three different areas: *in vitro* selections, self–cleaving ribozymes, and riboswitches.

2012 – 2014 Clinical Research Associate
Clinipace Worldwide, Inc. – Sao Paulo, Brazil

Conducted research site visits to ensure that clinical research sites were following protocol and applicable FDA regulations.

2010 – 2012 Clinical Researcher
Natura Inovacao e Tecnologia – Cajamar, Brazil

Managed clinical research projects related to the safety of products for human use. Designed a new methodology for safety assessment of products.

2007 – 2009 Clinical Trial Assistant
Quintiles Transnational Inc. (current IQVIA) – Sao Paulo, Brazil

Performed daily clinical trial tasks as an assistant and supported clinical research associates. Provided training to new employees.

2007 Intern in Pharmacovigilance and Customer Service
AstraZeneca Brazil – Cotia, Brazil

Reported all medical events and prepared attendance reports for physicians.

2003 – 2006 Undergraduate Research, Advisor: Professor Leoberto C. Tavares
University of Sao Paulo – Brazil

Researched the molecular reactivity of flavonoids as potential radioligands for the benzodiazepine receptor site on the GABAA receptor.

Publications

* Denotes equal contribution.

9. **Passalacqua, L. F. M.**, Dingilian, A. I., Lupták A. Single pass transcription using T7 RNA polymerase. *Manuscript in preparation*.

8. Chizzolini, F., **Passalacqua, L. F. M.**, Lupták, A. Synthesis of rNTPs by trimetaphosphate and nucleosides under prebiotic plausible conditions. *Manuscript in preparation*.

7. Chizzolini, F.*, **Passalacqua, L. F. M.***, Oumais, M. *, Dingilian, A. I., Szostak, J. W., Lupták, A. Large phenotypic enhancement of structured random RNA pools. *Submitted*.

6. Rotstan, K., Abdelsayed, M., **Passalacqua L. F. M.,** Chizzolini, F., Kasireddy, S., Chamberlain, R., Misek, J., Lupták, A. Regulation of mRNA translation by a photoriboswitch. *Submitted.*

5. **Passalacqua, L. F. M.**, Lupták, A. Co-transcriptional analysis of self-cleaving ribozymes and their ligand dependence. *Methods Mol Biol*. (2019) *In press*.

4. **Passalacqua, L. F. M.***, Jimenez, R. M.*, Fong, J. Y., Lupták, A. Allosteric modulation of the *Faecalibacterium prausnitzii* hepatitis delta virus-like ribozyme by glucosamine 6-phosphate: the substrate of the adjacent gene product. *Biochemistry* 56, 6006-6014 (2017).

3. Mosca, M. M.; Rosa, V. P.; Vendramini, J. M. M.; Guerra, L. O.; **Passalacqua, L. F. M.**; Martin, F. R.; Fragoso, M. F.; Santos, V.C. . Comparison between two methods of assessment of cutaneous irritation by rinse-off products: Flex Wash and Soap Chamber. In: XXI COLAMIQC - Congresso Latino Americano e Ibérico de Químicos Cosméticos, São Paulo, Brazil (2013).

2. Grallert, S. R. M. ; Paula, F. R. ; **Passalacqua, L. F. M.** ; Tavares, L. C.. Studies of molecular reactivity of flavonoids. In: 3rd Brazilian Symposium on Medicinal Chemistry, v. S3-186. BrazMedChem (2006).

1. Grallert, S. R. M. ; Paula, F. R. ; **Passalacqua, L. F. M.** ; Caldeira, J. ; Tavares, L. C. . Studies of molecular reactivity of flavonoids. In: XI Semana farmacêutica de ciência e tecnologia da FCF/USP, 2006, São Paulo, Brazil. Revista Brasileira de Ciências Farmacêuticas 42, 40 (2006).

Invited talks and presentations

6. **Luiz F. M. Passalacqua.** Enrichment of phenotypic capacity of random pools. Oral presentation at the UCI Synthetic and Chemical Biology Club. Irvine, CA, February 2019.

5. **Luiz F. M. Passalacqua**, Randi M. Jimenez, Jennifer Y. Fong, and Andrej Lupták. Allosteric modulation of bacterial HDV–like ribozymes by glucosamine 6–phosphate: The substrate of the adjacent gene product. Poster presented on the 23rd Annual Meeting of the RNA Society. Berkeley, CA, May 2018.

4. **Luiz F. M. Passalacqua**, Randi M. Jimenez, Jennifer Y. Fong, and Andrej Lupták. Allosteric modulation of bacterial HDV–like ribozymes by glucosamine 6–phosphate. Poster presented on the 4th Annual RNA Symposium at UCI, University of California, Irvine, CA, April 2018.

3. **Luiz F. M. Passalacqua.** Allosteric modulation of *Faecalibacterium prausnitzii* HDV–like ribozyme by glucosamine 6–phosphate: The substrate of the adjacent gene product. Oral presentation at the 2018 Vertex Day at UCI, Irvine, CA, March 2018.
2. **Luiz F. M. Passalacqua.** Allosteric modulation of bacterial HDV–like ribozymes by glucosamine 6–phosphate. Oral presentation at the UCI RNA Club. Irvine, CA, October 2017.

1. **Luiz F. M. Passalacqua**, Randi M. Jimenez, Jennifer Y. Fong, and Andrej Lupták. Allosteric modulation of bacterial HDV–like ribozymes by glucosamine 6–phosphate. Poster presented on the 11th Annual Symposium on Biological Complexity – RNA Biology, Salk institute, San Diego, CA, January 2017.

Honors and awards

2019	Graduate Dean's Dissertation Fellowship University of California, Irvine
2019	Miguel Velez Scholarship University of California, Irvine
2018	Outstanding Teaching Assistant Dept. of Pharmaceutical Sciences, University of California, Irvine
2014 – 2018	Science without Borders Ph.D. Scholarship LASPAU – USA / CAPES – Brazil
2005 – 2006	Undergraduate Research Fellowship National Council of Scientific and Technological Dev. – Brazil

Teaching experience

Summer 2019	Life 101 (Head Teaching Assistant), University of California, Irvine
Winter 2019	Life 101 (Teaching Assistant), University of California, Irvine
Winter 2018	Medicinal Chemistry (Head Teaching Assistant), University of California, Irvine
Winter 2017	Medicinal Chemistry (Teaching Assistant), University of California, Irvine
Winter 2016	Medicinal Chemistry (Teaching Assistant), University of California, Irvine
Spring 2015	Medicinal Chemistry (Teaching Assistant), University of California, Irvine

Mentoring experience

1. Chanika Assavarittirong (UROP Fellowship)	2018 – 2019
2. Phu Nguyen (UROP Fellowship)	2017 – 2019
3. Jiwoo Han (Youth Science Fellowship program)	Summer 2018
4. Jennifer Fong (UROP & SURP Fellowship)	2016 – 2018
5. Marwa Ahady (UROP Fellowship)	Summer 2015

Additional coursework

2011 – 2012	Certificate in Clinical Trial Administration University of California, San Diego, Extension
2009 – 2010	Diploma in Business Metropolitan College / University of Cambridge International Diploma – Toronto, ON – Canada

Skills

Laboratory:	PCR, qPCR, RT–qPCR, RT reaction, PCR mutagenesis, *in vitro* transcription, *in vitro* translation, native and denaturing PAGE gel electrophoresis, sequencing gel electrophoresis, RNA/DNA ^{32}P–labeling, RNA/DNA purification, RNA structural probing (in–line probing, SHAPE, LASER, S1, T1, and Tb^{3+} probing), phosphorothioate RNA interference, column biding assay, click chemistry (copper and copper–free catalyzed), next–generation sequencing, mammalian cell culture, *E. coli* cell culture, cloning, protein expression and purification, firefly and Renilla luciferases bioluminescence assay, cell imaging, and radiation safety officer.
Software:	MS Office Word and Excel, Image J, Image Quant, Inkscape, Fastaptamer, AptaSuite, Cutadapt, Jalview, Galaxy, and PyMol.
Languages:	Fluent in English, Portuguese, and Spanish. Intermediate level in French.

ABSTRACT OF THE DISSERTATION

Advances in discovery and characterization of small functional RNAs

By

Luiz Fernando Moreira Passalacqua

Doctor of Philosophy in Pharmacological Sciences

University of California, Irvine, 2019

Professor Andrej Lupták, Chair

Over the past five decades, new discoveries on RNA molecules have expanded the scope of their functionalities beyond information-carrying to include catalysis and regulation of gene expression. Yet much remains to be discovered and understood about functional RNAs. Due to the versatility of these molecules, the synthetic biology field has also shown a growing interest in functional RNAs. The studies presented herein are about small functional RNAs and focus on three different areas: *in vitro* selections, self–cleaving ribozymes, and riboswitches. First, I present studies that provide insights into improving *in vitro* selections by enhancing the phenotypic potential of RNA libraries used in the discovery of functional RNAs. The second focus of my studies is the characterization of self-cleaving ribozymes, and I report the first example of allosteric modulation of a natural self-cleaving ribozyme by a metabolite. I also present two different methodologies that can be used in the analysis of self-cleaving ribozymes. Lastly, I present work contributed towards the discovery and characterization of a novel synthetic photoriboswitch, an optogenetic tool that is able to control the downstream mRNA translation with temporal control.

Chapter 1: Introduction

1.1 Background

In the 1950's Francis Crick postulated that RNA acts as an intermediary between DNA and protein[1,2]. However, over the past five decades, studies have shown that the cellular functions attributed to RNA have greatly increased to include catalysis in the ribosome and spliceosome, and regulatory functions, such as gene silencing and gene regulation. These discoveries, among others, have led to a paradigm shift in molecular biology[3–5]. The findings from the Human Genome project are in agreement with this expanded role for RNA; it is now known that only about 3% of the DNA transcribed codes for protein genes and the other 97% codes for non–coding RNA[6]. These non–coding RNAs, which were once thought to be "junk", possess a vast range of modalities and functions, ranging from enzyme–like activity to control of gene expression[7,8]. Non–coding RNAs can be separated in two major groups: housekeeping non–coding RNAs, such as ribosomal RNA, tRNAs, small nucleolar RNAs, and small nuclear RNAs. The other group encompasses regulatory non–coding RNAs, which can be further sub–divided in two subgroups known as small non–coding RNA and long non–coding RNA[8]. Small non–coding RNAs, such as micro–RNAs, pi–RNAs, and small interfering RNAs, are limited to 200 nucleotides long, while long non–coding RNAs are longer than 200 nucleotides[7,8].

The versatility of RNA molecules lends support to the RNA World hypothesis, which suggests RNA preceded DNA and proteins as a precursor of the contemporary biology[9]. The RNA World hypothesis is especially attractive because RNA can carry genetic information like DNA, but it can also perform catalysis, a biological role generally attributed to proteins[9,10].

The synthetic biology field has also shown a growing interest in functional RNAs. The design of RNA molecules that exhibit functionalities such as regulatory, catalysis, sensing, information processing, and scaffold activities, is able to proportionate synthetic elements that can yield important tools[11]. These tools can be used in the development of biosensors, imaging probes, systems that program cellular functions, and even pharmaceutical agents[11,12].

In summary, there is now a wealth of the new information about RNA molecules and led the scientific community to pursue the research on functional RNAs. My thesis work focuses on three different areas of functional RNA biochemistry: *in vitro* selections, self–cleaving ribozymes, and riboswitches. All three are briefly described in this introductory chapter.

1.2 *In vitro* selection

RNA molecules with functional properties rely on their specific tertiary structure to promote activity, such as catalysis and ligand binding[13]. RNA aptamers are oligonucleotides that bind target ligands with high affinity and/or specificity[14,15]. These molecules have great potential for the development and incorporation into biosensors, image probes, biomolecular tools, and pharmaceutical agents. Additionally, aptamers can be used to further study and understand RNA functions in cells[12,13,16]. The origin of the aptamers may be from a synthetic or genomic library, but most RNA aptamers have been discovered using a technique based on molecular evolution called *in vitro* selection[14] or SELEX – systematic evolution of ligands by exponential enrichment[15].

Created almost 30 years ago, *in vitro* selection is based on selecting target–binding (or function specific) RNAs from highly diverse pools through serial rounds of enrichment and amplification[14,15]. A typical *in vitro* selection with RNA molecules starts with the generation of a large diverse oligonucleotide library. Each RNA molecule contains two constant primer–binding regions at each end for polymerase chain reaction (PCR) amplification and cDNA synthesis. The library (also called pool) is transcribed and introduced to a functional screen, such as binding or catalysis, and sequences demonstrating the desired function are separated from those lacking it. Once separated, the oligonucleotides of interest are collected and regenerated through reverse–transcription (RT) and PCR amplification for subsequent rounds of enrichment. These rounds of enrichment are repeated until the library converges to a collection of molecules that demonstrate the function of interest. Finally, the final library is cloned and sequenced[12,16]. A schematic representation of the *in vitro* selection approach is presented on **Figure 1.1**. The results are then bioinformatically treated to rank and cluster similar sequences in individual clones for subsequent confirmation of activity and biochemical characterization[16]. Although conceptually similar, the specific experimental designs and conditions of *in vitro* selection experiments differ considerably depending on the nature and characteristics of the target function[16].

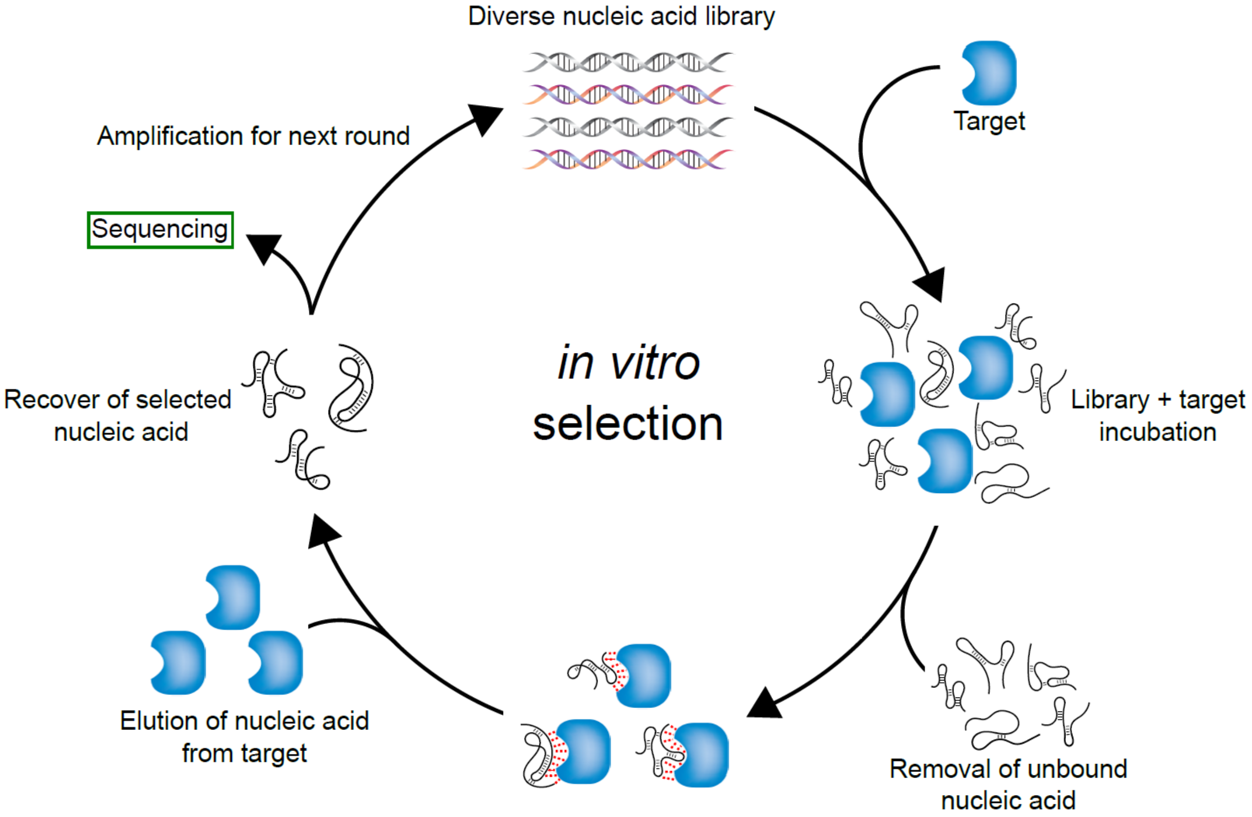

Diverse nucleic acid library
Target
Amplification for next round
Sequencing
in vitro selection
Recover of selected nucleic acid
Library + target incubation
Elution of nucleic acid from target
Removal of unbound nucleic acid

Figure 1.1. Schematic representation of *in vitro* selection. *In vitro* selection experiment begins with a diverse nucleic acid library of DNA or RNA molecules. The library is then incubated with the target of choice for functional screen. Sequences that demonstrate affinity towards the target molecule are isolated. The bound sequences are then eluted from the target molecules, recovered and amplified for subsequent rounds of enrichment. Multiple rounds are performed until the library converges on to a collection of selected sequences with affinity for the target molecule. The final library is sequenced.

While synthetic libraries are chemically synthesized, genomic libraries are generated from genomic DNA[16,17]. Genomic libraries consist of genomic sequences of selected size, flanked by constant sequences required for amplification and transcription[17]. To date, only three examples of RNA aptamers from the human genome have been discovered, including adenosine (triphosphate), guanosine triphosphate, and folic acid[13,18–20]. Many genomic RNA aptamers are part of larger constructs known as riboswitches, in which the aptamer domains typically cause ligand–induced conformational change, leading to modulation of a downstream expression[21–23]. On the other hand, aptamers derived from synthetic libraries are much more abundant and there are even databases of such aptamers, like *Apta-Index (https://www.aptagen.com/aptamer-index).*

When compared to other biomolecules that can exert similar functions, like antibodies, aptamers have great advantages. The production of aptamers is cheaper and they can be made *in vitro*, avoiding *in vivo* manipulation and its variables[24]. Aptamers are also easier to chemically manipulate than large biomolecules, making conjugation with molecules like fluorophores and quantum dots more feasible[25,26]. Additionally, aptamers tend to exhibit low to no toxicity or immunological effects, which makes them highly desirable for pharmaceutical usage[27]. The main limitation of aptamers is their susceptibility to nucleases; degradation by nucleases can occur both *in vivo* and *ex vivo*. The most common way to increase its resistance to nucleases is to add chemical modifications to the aptamer molecules[26,28].

One of the most important aspects of an *in vitro* selection experiment is the random library that is used to start the selection. In general, these libraries vary from 30 to 200 nucleotides and their sequence diversity easily reaches values that are impossible to produce in a laboratory. For example, a library with a random region of 100 nucleotides has the theoretical diversity of 4^{100} ($\sim 1.6 \times 10^{60}$). In a research setting, it is nearly impossible to get a diversity over $\sim 10^{16}$, which corresponds to a transcription reaction of ~10–100 mL. Thus, in most *in vitro* selections, the initial random pool is undersampled, decreasing the phenotypic potential of the pool. In chapter 2, I present a study that involves the manipulation of initial random pools to boost the potential of such libraries, therefore increasing the probability of getting molecules with the phenotype of interest.

1.3 Self–cleaving ribozymes

Discovered over 35 years ago, ribozymes are RNA molecules that can catalyze a chemical reaction in the absence of a protein cofactor[29–31]. Self–cleaving ribozymes are a specific group of RNA molecules that promote a site–specific self–scission reaction[32]. These molecules have been found throughout all branches of life, from bacteria to animals, but their distribution, biological functions, and regulation are yet to be fully elucidated[32,33,42,34–41]. The most common mechanism employed by self–cleaving ribozymes to promote the self–scission reaction is a general acid–base catalysis, where a transesterification involving a nucleophilic attack by a 2′–oxygen on the scissile phosphodiester bond, produces a 2′−3′ cyclic phosphate and a 5′–hydroxyl product, with metal ions and metabolites employed as potential cofactors **(Figure 1.2)**[32,33,43–46].

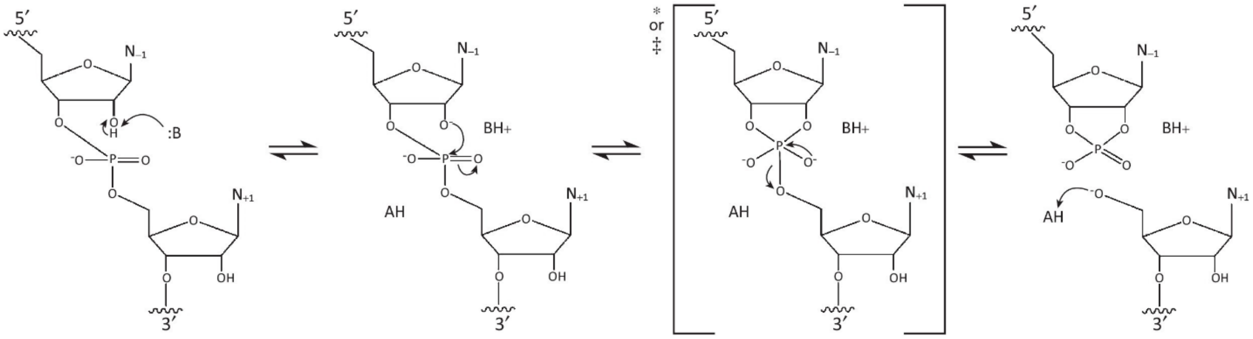


Figure 1.2. Mechanism of self–scission. A general acid–base catalysis involves a general base, which deprotonates the 2′–hydroxyl of the nucleophile, positioned in–line with the 5′–O leaving group. The transesterification proceeds via a phosphorane transition state or intermediate, depending on whether it is stabilized, yielding a 2′–3′ cyclic phosphate in the upstream nucleotide and an oxyanion on the downstream nucleotide. The leaving group is protonated by a general acid[32].

To date, nine self–cleaving ribozyme families have been discovered in nature, they include: the hammerhead[36,38], hairpin[37], hepatitis delta virus (HDV)[34,35], glucosamine–6–phosphate synthase (glmS)[46], Neurospora Varkud satellite (VS)[47], twister[42], twister sister (TS), pistol, and hatchet motifs[39].

The biological functions of self–cleaving ribozymes are still not fully understood and many details are yet to be elucidated, however self–cleaving ribozymes are likely involved in several biological roles. Some of the known functions consist of self–scission during rolling–circle replication of RNA genomes, co–transcriptional processing of retrotransposons, and metabolite–dependent gene expression regulation in bacteria[35,36,53–55,38,40,47–52]. Their genomic locations also suggest that they are involved in many other biological processes, some of which may not be directly associated with RNA scission. In other cases, like two mammalian ribozymes that are highly conserved[56,57], their biological roles have not been elucidated yet.

Part of my thesis work focused on the HDV–like ribozyme family. This class of ribozyme was first discovered in the hepatitis delta virus (HDV) in the late 1980s[34,35]. HDV is an RNA virusoid that is not harmful by itself, but it has been shown to make co–infection with hepatitis beta virus (HBV) the most severe form of viral hepatitis, causing liver cirrhosis and hepatocellular carcinoma[58]. This co–infection with HBV affects about 20 million people worldwide, corresponding to ~ 5% of all cases of HBV hepatitis[59]. It is known that the HDV life cycle depends on HBV envelope proteins for its assembly, justifying the need of co–infection with HBV[58,59]. Interestingly, a recent study suggests that the coinfection with enveloped viruses distinct from HBV can also disseminate HDV[60], a finding that may increase the pathological importance of the HDV. The virus relies on host cellular machinery for replication and the HDV self–cleaving ribozyme plays a crucial role in rolling–circle replication of the single–stranded RNA genome[32,58]. The crystal structure and self–scission mechanism of the HDV self–cleaving ribozyme have been solved **(Figure 1.3A)** making it one of the most extensively studied self–cleaving ribozymes[61–67]. It was only in 2006[56] when a second example of this ribozyme motif was found elsewhere, in humans, and in 2009 when its widespread occurrence was discovered[40].

The presence of the HDV–like ribozyme motif have been found in many eukaryotes, including humans, as well as in *Chilo iridescent* virus, several bacteria, and most recently in several microbial meta–genomic datasets[39,40,50,51,56,68,69]. These self–cleaving RNAs likely play a number of distinct roles in biology[50,51,56,68–70], but little is known about their regulation.

Despite a great degree of sequence diversity, the HDV–like self–cleaving ribozymes exhibit a conserved secondary structure that includes five helical regions (P1.1, P1, P2, P3, P4) forming two coaxial stacks (P1, P1.1 stack on P4 and P2 stacks on P3) that are joined by single stranded regions (J1/2, L3, J4/2) **(Figure 1.3B)**. Additionally, crystal structures revealed that a nested double–pseudoknot is also present, buried in the junction of P1, P1.1, and P3. In the active site, only six conserved nucleotides fulfill functional or structural roles. The ribozyme exists naturally in a minimal (lacking P4) or extended (extension of J1/2 and P4) form[40,61,62,65,69]. The proposed self–scission mechanism for the HDV ribozyme reaction relies on a protonated form of the nucleobase C75 that donates a proton to the 5′–oxyanion leaving group of G1[63]. The protonated cytosine is thought to be stabilized through interactions with the scissile phosphate[65] and is suggested to interact electrostatically with the divalent metal ion bound in the active site[71]. Additionally, the divalent cations Mg^{2+} have a profound impact on catalysis and recent evidence revealed a Mg^{2+} ion in a metal binding pocket in the ribozyme active site[65,72]. Thus, the HDV–like ribozymes are metalloenzymes under biological conditions. A Mg^{2+} ion coordinated by a hydroxide anion is suggested to act as the general base that deprotonates the 2′–OH that attacks the scissile phosphate[63,72]. **Figure 1.3C** depicts the mechanism of this self–cleaving reaction. Also, it is suggested that the coordinated divalent cation potentially stabilizes the pentavalent phosphorane transition state[73]. It has also been observed that there is little metal ion specificity, and catalysis can proceed in high concentrations of monovalent cations alone[74]. The HDV ribozyme motif is also remarkably stable, showing activity in up to 18 M formamide[75].

In my thesis work, I present in chapter 3 a study on an allosteric modulation of a bacterial HDV–like ribozyme by the substrate of the adjacent gene product. In chapter 4, I present a method to study the kinetics of such ribozymes. Additionally, in chapter 5, I present a new strategy to study such ribozymes (and other molecules) using an *in vitro* system that allows a single pulse of transcription.

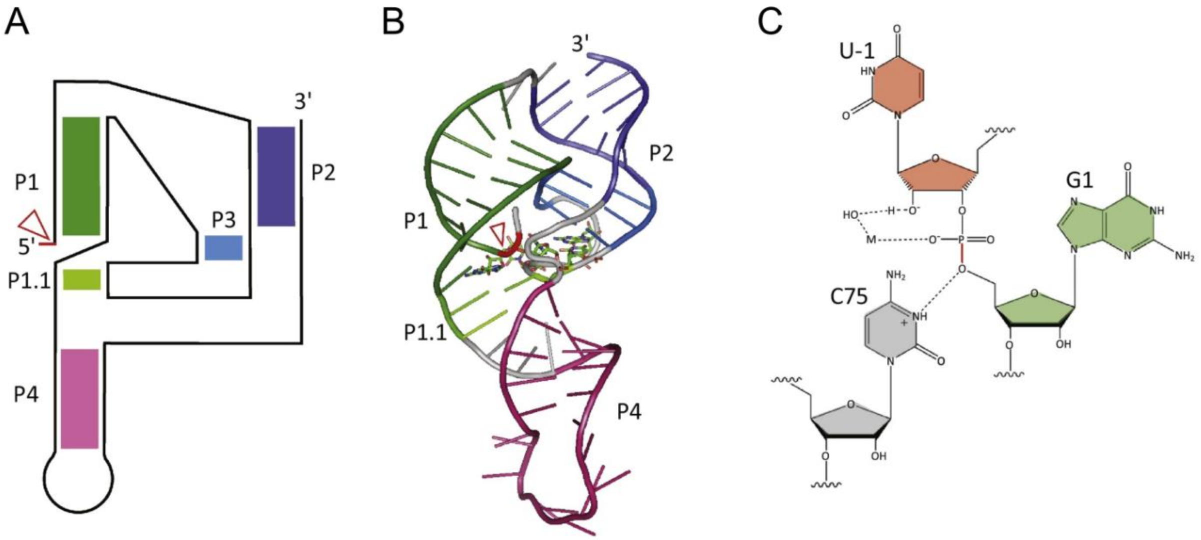


Figure 1.3. Overview of the HDV ribozyme. (A) Secondary structure of the HDV ribozyme[32]. The helical segments are numbered as ‘paired’ elements (P1, P2, etc.) and colored. (B) Crystal structure of the genomic HDV ribozyme[32,61] – Protein Data Bank ID is 1DRZ. The crystal structure follows the same color scheme as the secondary structure. The red arrowheads indicate the site of self–scission in both (A) and (B). (C) Proposed self–scission mechanism, which illustrates interactions that promote the cleavage event. The nucleobase flanking the scissile phosphate is splayed apart, which promotes the in–line orientation necessary to accomplish cleavage. The scissile bond is shown in red[32].

1.4 Riboswitches

Riboswitches are generally present in the 5′ untranslated regions (UTRs) of messenger RNAs (mRNAs), these structured RNA molecules are capable of regulating genes expression at both the transcriptional and translational level[22,23]. Riboswitches are able to sense target ligands in the cellular environment and up or down–regulate the gene expression of the downstream mRNA[21–23,76,77]. First discovered in 2002, in bacteria[78–80], 39 different classes of riboswitches have been discovered and validated , and it is proposed that several hundred new riboswitches are still to be discovered[22,81,82]. These RNA elements are mostly present in bacteria, however there is one known example of a riboswitch family in fungi and plants, demonstrating that this level of regulation is also present in eukaryotic cells[83,84].

The basic structure of a riboswitch is composed by two distinct regions: first an aptamer domain, which senses the target ligand in the environment and produces a conformational change upon binding. Second, an expression platform downstream of the aptamer domain, that may or may not allow transcription or translation to happen, as a result of the conformational change of the aptamer domain. The two regions present some overlapping nucleotides, known as the *switching sequence.* The pairing of this sequence upon binding or absence of the ligand may lead to two mutually exclusive structures, representing the on or off state of the mRNA regulation[23,77,85]. At the transcriptional level, the structural change promoted by the binding of the ligand to the aptamer domain may induce the formation of a transcription terminator stem, while at the translational level, the structural change may occlude or display the ribosome binding site (RBS)[23,77]. **Figure 1.4** illustrates the mechanism of action of riboswitches with each domain and possible downstream influence on gene regulation. Other types of riboswitch families do exist, like the ribozyme–riboswitch *glmS* that senses glucosamine 6–phosphate (GlcN6P) and self–cleaves, decreasing the half–life of the mRNA by recruitment of RNase J[46,86], or the eukaryotic thiamine pyrophosphate (TPP) riboswitch family, uncommonly situated at the 3′–UTR and present in fungi and plants, regulates the alternative splicing of the mRNA[84]. The variety of target ligands that participate in the riboswitches gene control is also extensive, ranging from metabolites, coenzymes, signaling molecules, and even metal ions[23,76,77,87].

In recent years, the field of synthetic biology has developed a great interest in riboswitches, which can be used as platforms for the development of new molecular biology tools, particularly in the control of gene–expression[87–90]. One example of such an application is the theophylline–dependent translation riboswitch, a genetic circuit that makes use of the theophylline aptamer to regulate translation[91]. Later on, a theophylline–dependent transcription riboswitch was also developed[92], showing that such applications are feasible for modulation of gene expression at the level of both transcription and translation. Despite the relative success of the scientific community in the development of such tools, a tool that offers spatiotemporal control of gene regulation using riboswitches is still lacking. Such a tool would be useful to the field of molecular biology. In chapter 6, I present work performed to develop a synthetic photoriboswitch, an optogenetic tool that is able to control the downstream mRNA translation with temporal control.

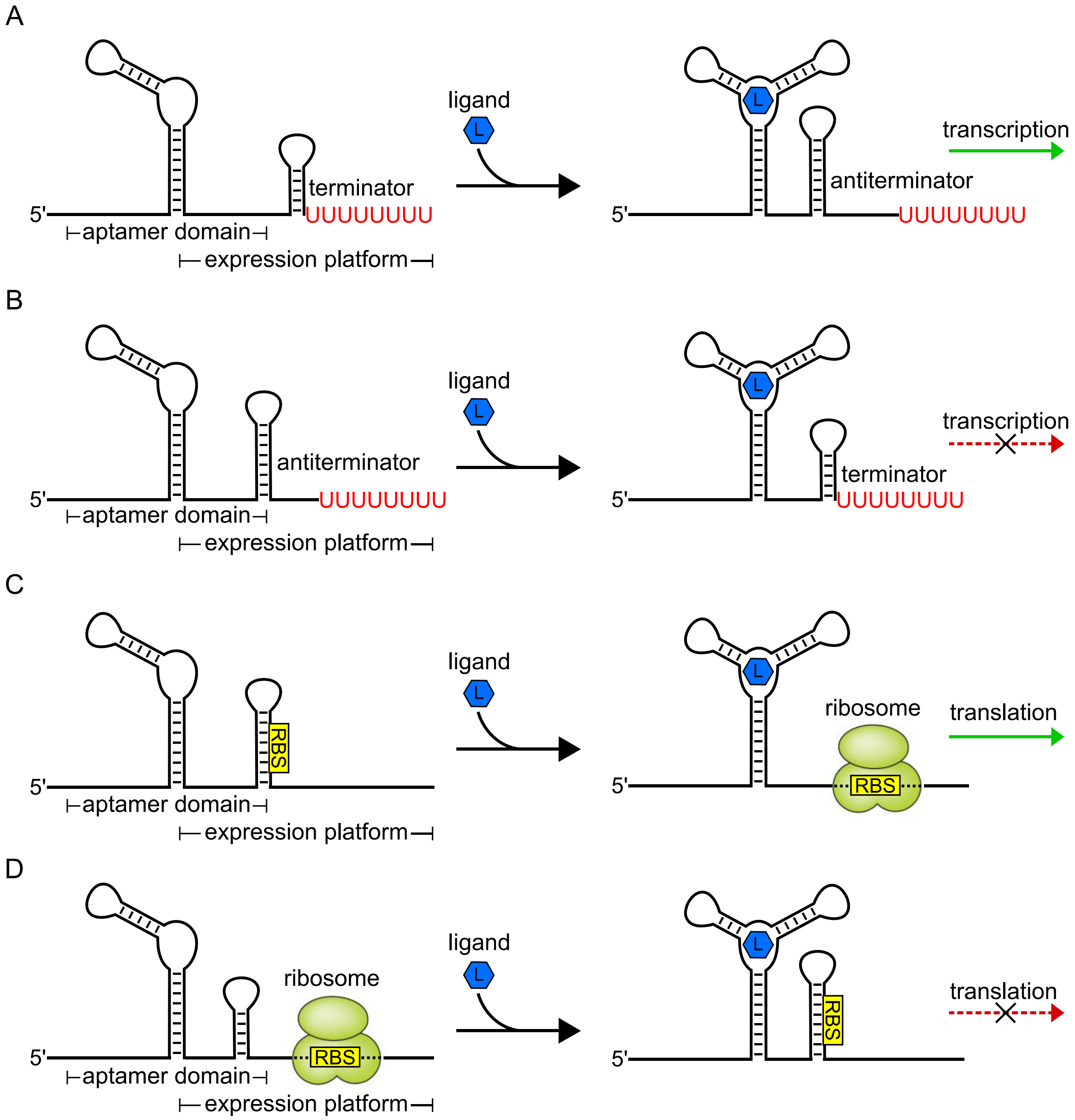

A
5'
terminator
UUUUUUUU
⊢aptamer domain⊣
⊢expression platform⊣
ligand
L
5'
transcription
antiterminator
UUUUUUUU
B
5'
antiterminator
UUUUUUUU
⊢aptamer domain⊣
⊢expression platform⊣
ligand
L
5'
transcription
terminator
UUUUUUUU
C
5'
RBS
⊢aptamer domain⊣
⊢expression platform⊣
ligand
L
5'
ribosome
RBS
translation
D
5'
ribosome
RBS
⊢aptamer domain⊣
⊢expression platform⊣
ligand
L
5'
RBS
translation

Figure 1.4. Schematic representations of transcriptional and translation regulation of gene expression by riboswitches. Upon ligand binding to the aptamer domain, a conformational change is promoted causing the expression platform to switch, resulting in gene activation or repression. **(A) Transcription activation** – after ligand binding and conformational change, the terminator motif is disrupted and the anti–terminator motif is

formed, allowing transcription to continue. **(B) Transcription repression** – after ligand binding and conformational change, the anti–terminator motif is disrupted and the terminator motif is formed, promoting the release of the transcriptional machinery and the termination of transcription. **(C) Translation activation** – after ligand binding and conformational change, the ribosome binding site (RBS) is displayed for ribosomal binding, promoting translation initiation. **(D) Translation repression** – after ligand binding and conformational change, the RBS is sequestered, inhibiting ribosomal binding and translation initiation.

1.5 References

1. Crick, F. Central Dogma of Molecular Biology. *Nature* **227**, 561–563 (1970).
2. Crick, F. H. On protein synthesis. *Symp. Soc. Exp. Biol.* **12**, 138–163 (1958).
3. Cech, T. R. & Steitz, J. A. Review The Noncoding RNA Revolution – Trashing Old Rules to Forge New Ones. *Cell* **157**, 77–94 (2014).
4. Sharp, P. A. The Centrality of RNA. *Cell* **136**, 577-580 (2009).
5. Morris, K. V, & Mattick, J. S. The rise of regulatory RNA. *Nat. Rev. Genet.* **15**, 423 (2014).
6. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).
7. Palazzo, A. F. & Lee, E. S. Non-coding RNA: what is functional and what is junk? *Front. Genet.* **6**, 2 (2015).
8. Fu, X.-D. Non-coding RNA: a new frontier in regulatory biology. *Natl. Sci. Rev.* **1**, 190–204 (2014).
9. Gilbert, W. Origin of life: The RNA world. *Nature* **319**, 618 (1986).
10. Cech, T. R. The RNA Worlds in Context. *Cold Spring Harb Perspect Biol,* **4**, 7 (2012).
11. Liang, J. C., Bloom, R. J. & Smolke, C. D. Engineering biological systems with synthetic RNA molecules. *Mol. Cell* **43**, 915–926 (2011).
12. Wu, Y. X. & Kwon, Y. J. Aptamers: The “evolution” of SELEX. *Methods* **106**, 21–28 (2016).
13. Vu, M. M., Jameson, N. E., Masuda, S.J., Lin, D., Larralde-Ridaura, R., Lupták, A. Convergent evolution of adenosine aptamers spanning bacterial, human, and random sequences revealed by structure-based bioinformatics and genomic SELEX. *Chem. Biol.* **19**, 1247–1254 (2012).
14. Ellington, A. D., Szostak, J. W. In vitro selection of RNA molecules that bind specific ligands. *Nature* **346**, 818–822 (1990).
15. Tuerk, C., Gold, L. Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science,* **249**, 505–510 (1990).
16. Jijakli, K., Khraiwesh, B., Fu, W., Luo, L., Alzahmi, A., Koussa, J., Chaiboonchoe, A., Kirmizialtin, S., Yen, L., Salehi-Ashtiani, K. The in vitro selection world. *Methods* **106**, 3–13 (2016).
17. Lorenz, C., von Pelchrzim, F. & Schroeder, R. Genomic systematic evolution of ligands by exponential enrichment (Genomic SELEX) for the identification of protein-binding RNAs independent of their expression levels. *Nat. Protoc.* **1**, 2204–2212 (2006).
18. Curtis, E. A., Liu, D. R. Discovery of Widespread GTP-Binding Motifs in Genomic DNA and RNA. *Chem. Biol.* **20**, 521–532 (2013).

19. Sassanfar, M., Szostak, J. W. An RNA motif that binds ATP. *Nature* **364**, 550–553 (1993).
20. Terasaka, N., Futai, K., Katoh, T. & Suga, H. A human microRNA precursor binding to folic acid discovered by small RNA transcriptomic SELEX. *RNA* **22**, 1918–1928 (2016).
21. Breaker, R. R. Prospects for Riboswitch Discovery and Analysis. *Mol. Cell* **43**, 867–879 (2011).
22. Breaker, R. R. Riboswitches and Translation Control. *Cold Spring Harb. Perspect. Biol.* **10**, 11 (2018).
23. Serganov, A. & Nudler, E. A Decade of Riboswitches. *Cell* **152**, 17–24 (2013).
24. Tombelli, S., Minunni, M. & Mascini, M. Aptamers-based assays for diagnostics, environmental and food analysis. *Biomol. Eng.* **24**, 191–200 (2007).
25. Liu, J., Cao, Z. & Lu, Y. Functional Nucleic Acid Sensors. *Chem. Rev.* **109**, 1948–1998 (2009).
26. Wan, W. B. & Seth, P. P. The Medicinal Chemistry of Therapeutic Oligonucleotides. *J. Med. Chem.* **59**, 9645–9667 (2016).
27. Wlotzka, B., Leva, S., Eschgfäller, B., Burmeister, J., Kleinjung, F., Kaduk, C., Muhn, P., Hess-Stumpp, H., Klussmann, S. *In vivo* properties of an anti-GnRH Spiegelmer: an example of an oligonucleotide-based therapeutic substance class. *Proc. Natl. Acad. Sci.* **99**, 8898–8902 (2002).
28. Pieken, W. A., Olsen, D. B., Benseler, F., Aurup, H. & Eckstein, F. Kinetic characterization of ribonuclease-resistant 2'-modified hammerhead ribozymes. *Science* **253**, 314–317 (1991).
29. Doudna, J. A. & Cech, T. R. The chemical repertoire of natural ribozymes. *Nature* **418**, 222–228 (2002).
30. Kruger, K. *et al.* Self-splicing RNA: Autoexcision and autocyclization of the ribosomal RNA intervening sequence of tetrahymena. *Cell* **31**, 147–157 (1982).
31. Guerrier-Takada, C., Gardiner, K., Marsh, T., Pace, N. & Altman, S. The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme. *Cell* **35**, 849–857 (1983).
32. Jimenez, R. M., Polanco, J. A. & Lupták, A. Chemistry and biology of self-cleaving ribozymes. *Trends Biochem. Sci.* **40**, 648–661 (2015).
33. Fedor, M. J. Comparative enzymology and structural biology of RNA self-cleavage. *Annu. Rev. Biophys.* **38**, 271–299 (2009).
34. Wu, H. N., Lin, Y. J., Lin, F. P., Makino, S., Chang, M. F., & Lai, M. M. Human hepatitis delta virus RNA subfragments contain an autocleavage activity. *Proc. Natl. Acad. Sci.* **86**, 1831–1835 (1989).
35. Sharmeen, L., Kuo, M. Y., Dinter-Gottlieb, G. & Taylor, J. Antigenomic RNA of human hepatitis delta virus can undergo self-cleavage. *J. Virol.* **62**, 2674–2679 (1988).

36. Prody, G. A., Bakos, J. T., Buzayan, J. M., Schneider, I. R. & Bruening, G. Autolytic processing of dimeric plant virus satellite RNA. *Science* **231**, 1577–1580 (1986).

37. Buzayan, J. M., Gerlach, W. L. & Bruening, G. Non-enzymatic cleavage and ligation of RNAs complementary to a plant virus satellite RNA. *Nature* **323**, 349–353 (1986).

38. Hutchins, C. J., Rathjen, P. D., Forster, A. C. & Symons, R. H. Self-cleavage of plus and minus RNA transcripts of avocado sunblotch viroid. *Nucleic Acids Res.* **14**, 3627–3640 (1986).

39. Weinberg, Z. Kim, P. B., Chen, T. H., Li, S., Harris, K. A., Lünse, C. E., & Breaker, R. R. New classes of self-cleaving ribozymes revealed by comparative genomics analysis. *Nat. Chem. Biol.* **11**, 606–610 (2015).

40. Webb, C. H. T., Riccitelli, N. J., Ruminski, D. J. & Lupták, A. Widespread occurrence of self-cleaving ribozymes. *Science* **326**, 953 (2009).

41. Hammann, C., Luptak, A., Perreault, J. & de la Peña, M. The ubiquitous hammerhead ribozyme. *RNA* **18**, 871–885 (2012).

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

43. Wilson, T. J., Liu, Y. & Lilley, D. M. J. Ribozymes and the mechanisms that underlie RNA catalysis. *Front. Chem. Sci. Eng.* **10**, 178–185 (2016).

44. Ren, A., Micura, R. & Patel, D. J. Structure-based mechanistic insights into catalysis by small self-cleaving ribozymes. *Curr. Opin. Chem. Biol.* **41**, 71–83 (2017).

45. Seith, D. D., Bingaman, J. L., Veenis, A. J., Button, A. C. & Bevilacqua, P. C. Elucidation of catalytic strategies of small nucleolytic ribozymes from comparative analysis of active sites. *ACS Catal.* **8**, 314–327 (2018).

46. Winkler, W. C., Nahvi, A., Roth, A., Collins, J. A. & Breaker, R. R. Control of gene expression by a natural metabolite-responsive ribozyme. *Nature* **428**, 281–286 (2004).

47. Saville, B. J. & Collins, R. A. A site-specific self-cleavage reaction performed by a novel RNA in neurospora mitochondria. *Cell* **61**, 685–696 (1990).

48. Rojas, A. A., Vazquez-Tello, A., Ferbeyre, G., Venanzetti, F., Bachmann, L., Paquin, B., … Cedergren, R. Hammerhead-mediated processing of satellite pDo500 family transcripts from Dolichopoda cave crickets. *Nucleic Acids Res.* **28**, 4037–4043 (2000).

49. Martick, M., Horan, L. H., Noller, H. F. & Scott, W. G. A discontinuous hammerhead ribozyme embedded in a mammalian messenger RNA. *Nature* **454**, 899 (2008).

50. Eickbush, D. G. & Eickbush, T. H. R2 retrotransposons encode a self-cleaving ribozyme for processing from an rRNA cotranscript. *Mol. Cell. Biol.* **30**, 3142–3150 (2010).

51. Sánchez-Luque, F. J., López, M. C., Macias, F., Alonso, C. & Thomas, M. C.

Identification of an hepatitis delta virus-like ribozyme at the mRNA 5'-end of the L1Tc retrotransposon from Trypanosoma cruzi. *Nucleic Acids Res.* **39**, 8065–8077 (2011).

52. Cervera, A. & De la Peña, M. Eukaryotic penelope-like retroelements encode hammerhead ribozyme motifs. *Mol. Biol. Evol.* **31**, 2941–2947 (2014).
53. Forster, A. C. & Symons, R. H. Self-cleavage of plus and minus RNAs of a virusoid and a structural model for the active sites. *Cell* **49**, 211–220 (1987).
54. Epstein, L. M. & Gall, J. G. Transcripts of Newt satellite DNA self-cleave in vitro. *Cold Spring Harb. Symp. Quant. Biol.* **52**, 261–265 (1987).
55. Ferbeyre, G., Smith, J. M. & Cedergren, R. Schistosome satellite DNA encodes active hammerhead ribozymes. *Mol. Cell. Biol.* **18**, 3880–3888 (1998).
56. Salehi-Ashtiani, K., Lupták, A., Litovchick, A. & Szostak, J. W. A genomewide search for ribozymes reveals an HDV-like sequence in the human CPEB3 gene. *Science* **313**, 1788–1792 (2006).
57. de la Peña, M. & García-Robles, I. Intronic hammerhead ribozymes are ultraconserved in the human genome. *EMBO Rep.* **11**, 711–716 (2010).
58. Taylor, J. M. Hepatitis delta virus. *Virology* **344**, 71–76 (2006).
59. Lempp, F. A., Ni, Y. & Urban, S. Hepatitis delta virus: insights into a peculiar pathogen and novel treatment options. *Nat. Rev. Gastroenterol. &Amp; Hepatol.* **13**, 580 (2016).
60. Perez-Vargas, J., Amirache, F., Boson, B., Mialon, C., Freitas, N., Sureau, C., … Cosset, F. L. Enveloped viruses distinct from HBV induce dissemination of hepatitis D virus in vivo. *Nat. Commun.* **10**, 2098 (2019).
61. Ferré-D'Amaré, A. R., Zhou, K. & Doudna, J. A. Crystal structure of a hepatitis delta virus ribozyme. *Nature* **395**, 567–574 (1998).
62. Ke, A., Zhou, K., Ding, F., Cate, J. H. D. & Doudna, J. A. A conformational switch controls hepatitis delta virus ribozyme catalysis. *Nature* **429**, 201 (2004).
63. Das, S. R. & Piccirilli, J. A. General Acid Catalysis by the Hepatitis Delta Virus Ribozyme. *Nat. Chem. Biol.* **1**, 45–52 (2005).
64. Ferré-D'Amaré, A. R., Doudna, J. A. Crystallization and structure determination of a hepatitis delta virus ribozyme: use of the RNA-binding protein U1A as a crystallization module. *J. Mol. Bio.* **295**, 3 (2000).
65. Chen, J. H., Yajima, R., Chadalavada, D. M., Chase, E., Bevilacqua, P. C., Golden, B. L. A 1.9 Å crystal structure of the HDV ribozyme precleavage suggests both lewis acid and general acid mechanisms contribute to phosphodiester cleavage. *Biochemistry* **49**, 6508–6518 (2010).
66. Koo, S. C., Lu, J., Li, N. S., Leung, E., Das, S. R., Harris, M. E., & Piccirilli, J. A. Transition State Features in the Hepatitis Delta Virus Ribozyme Reaction Revealed by Atomic Perturbations. *J. Am. Chem. Soc.* **137**, 8973–8982 (2015).

67. Golden, B. L., Hammes-Schiffer, S., Carey, P. R. & Bevilacqua, P. C. An Integrated Picture of HDV Ribozyme Catalysis. *Biophysics of RNA Folding* 135–167 (2013).

68. Ruminski, D. J., Webb, C.-H. T., Riccitelli, N. J. & Lupták, A. Processing and translation initiation of non-long terminal repeat retrotransposons by hepatitis delta virus (HDV)-like self-cleaving ribozymes. *J. Biol. Chem.* **286**, 41286–41295 (2011).

69. Riccitelli, N. J., Delwart, E. & Lupták, A. Identification of minimal HDV-like ribozymes with unique divalent metal ion dependence in the human microbiome. *Biochemistry* **53**, 1616–1626 (2014).

70. Webb, C., Riccitelli, N., Ruminski, D. & Luptak, A. Widespread Occurance of Self-Cleaving Ribozymes. *Science* **326**, 953 (2009).

71. Gong, B., Chen, J.-H., Bevilacqua, P. C., Golden, B. L. & Carey, P. R. Competition between $Co(NH_3)_6^{3+}$ and Inner Sphere Mg^{2+} Ions in the HDV Ribozyme. *Biochemistry* **48**, 11961–11970 (2009).

72. Golden, B. L. Two distinct catalytic strategies in the hepatitis δ virus ribozyme cleavage reaction. *Biochemistry*, **50,** 9424–9433 (2011).

73. Thaplyal, P., Ganguly, A., Hammes-Schiffer, S. & Bevilacqua, P. C. Inverse Thio Effects in the Hepatitis Delta Virus Ribozyme Reveal that the Reaction Pathway Is Controlled by Metal Ion Charge Density. *Biochemistry* **54**, 2160–2175 (2015).

74. Perrotta, A. T. & Been, M. D. HDV ribozyme activity in monovalent cations. *Biochemistry* **45**, 11357–11365 (2006).

75. Duhamel, J. *et al.* Secondary structure content of the HDV ribozyme in 95% formamide. *Nucleic Acids Res.* **24**, 3911–3917 (1996).

76. McCown, P. J., Corbino, K. A., Stav, S., Sherlock, M. E. & Breaker, R. R. Riboswitch diversity and distribution. *RNA* **23**, 995–1011 (2017).

77. Garst, A. D., Edwards, A. L. & Batey, R. T. Riboswitches: structures and mechanisms. *Cold Spring Harbor perspectives in biology*, **3,** 6 (2011).

78. Mironov, A. S., Gusarov, I., Rafikov, R., Lopez, L. E., Shatalin, K., Kreneva, R. A., Perumov, D. A., Nudler, E. Sensing Small Molecules by Nascent RNA: A Mechanism to Control Transcription in Bacteria. *Cell* **111**, 747–756 (2002).

79. Nahvi, A., Sudarsan, N., Ebert, M. S., Zou, X., Brown, K. L., Breaker, R. R. Genetic Control by a Metabolite Binding mRNA. *Chem. Biol.* **9**, 1043–1049 (2002).

80. Winkler, W., Nahvi, A., Breaker, R. R. Thiamine derivatives bind messenger RNAs directly to regulate bacterial gene expression. *Nature* **419**, 952– 956 (2002).

81. Mccown, P. J., Corbino, K. A., Stav, S., Sherlock, M. E. & Breaker, R. R. Riboswitch Diversity and Distribution Riboswitch Diversity and Distribution. *RNA* **23**, 995–1011 (2017).

82. Chen, X., Mirihana Arachchilage, G. & Breaker, R. R. Biochemical validation of a second class of tetrahydrofolate riboswitches in bacteria. RNA (2019).

83. Sudarsan, N., Barrick, J. E. & Breaker, R. R. Metabolite-binding RNA domains are present in the genes of eukaryotes. *RNA* **9**, 644–647 (2003).

84. Sudarsan, N., Cohen-Chalamish, S., Nakamura, S., Emilsson, G. M. & Breaker, R. R. Thiamine pyrophosphate riboswitches are targets for the antimicrobial compound pyrithiamine. *Chem. Biol.* **12**, 1325–1335 (2005).

85. Breaker, R. R. Riboswitches and the RNA world. *Cold Spring Harb. Perspect. Biol.* **4**, 2 (2012).

86. Collins, J. A., Irnov, I., Baker, S. & Winkler, W. C. Mechanism of mRNA destabilization by the glmS ribozyme. *Genes Dev.* **21**, 3356–3368 (2007).

87. Batey, R. T. Riboswitches: still a lot of undiscovered country. *RNA* **21**, 560–563 (2015).

88. Hallberg, Z. F., Su, Y., Kitto, R. Z. & Hammond, M. C. Engineering and In Vivo Applications of Riboswitches. *Annu. Rev. Biochem.* **86**, 515–539 (2017).

89. Villa, J. K., Su, Y., Contreras, L. M. & Hammond, M. C. Synthetic Biology of Small RNAs and Riboswitches. *Microbiol. Spectr.* **6,** 3 (2018).

90. Wittmann, A. & Suess, B. Engineered riboswitches: Expanding researchers' toolbox with synthetic RNA regulators. *FEBS Lett.* **586**, 2076–2083 (2012).

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

92. Wachsmuth, M., Findeiss, S., Weissheimer, N., Stadler, P. F. & Morl, M. De novo design of a synthetic riboswitch that regulates transcription termination. *Nucleic Acids Res.* **41**, 2541–2551 (2013).

Chapter 2: Large phenotypic enhancement of structured random RNA pools

Fabio Chizzolini*, **Luiz F. M. Passalacqua***, Mona Oumais*, Armine I. Dingilian, Jack W. Szostak, Andrej Lupták. Large phenotypic enhancement of structured random RNA pools. *Manuscript in preparation.*

* Authors contributed equally to this work.

Contribution Statement

Computational analysis of the pools and assessment of secondary structures in pools with ethidium bromide experiments were performed by F.C. Structural probing of the pools, low diversity *in vitro* selection for catalytic RNAs, characterization of clusters for both Oregon Green and low diversity catalytic RNAs were performed by F.C. and **L.F.M.P**. High diversity *in vitro* selection for catalytic RNAs and characterization of clusters were performed by **L.F.M.P**. *In vitro* selection for Fluc/Ni-NTA and NanoLuc/Ni-NTA, and characterization of clusters enriched from the NanoLuc/Ni-NTA pools were performed by M.O. Samples of the selected pools for sequencing were prepared by A.I.D. A.L. performed initial selection of the RNA pools and Oregon Green *in vitro* selection. A.L. and J.W.S. conceived the study.

2.1 Abstract

Laboratory evolution of functional RNAs is central to many areas of chemical and synthetic biology. *In vitro* selections critically depend on the abundance of functional molecules, such as aptamers and ribozymes, in the starting sequence pools. For selection of novel functions the pools are typically transcribed from random-sequence DNA templates, yielding a highly diverse set of RNAs that contain a multitude of folds and biochemical activities. The phenotypic potential—the frequency of functional RNAs—is very low, requiring large complexity of starting pools, often surpassing 10^{16} different sequences, to identify highly active isolates. Furthermore, the majority of random sequences are not structured and have a high propensity for aggregation; the *in vitro* selection process thus involves not just enrichment of functional RNAs, but also their purification from aggregation-prone “free-riders”. We reasoned that purification of the non-aggregating, monomeric subpopulation of a random-sequence RNA pool will yield pools of folded, functional RNAs. We performed six rounds of selection for monomeric sequences and show that the enriched population is compactly folded. *In vitro* selections originating from various mixtures of the compact pool and a fully-random pool showed that sequences from the compact pool always dominate the population once a biochemical activity is detectable. A head-to-head competition of the two pools starting from a low (5E12) sequence diversity revealed that the phenotypic potential of the compact pool is about 1000-times higher than the fully-random pool. A selection for folded and monomeric RNA pools thus greatly increases the frequency of functional RNAs from that seen in random-sequence pools, providing a facile experimental approach to isolation of highly active functional RNAs from low-diversity populations.

2.2 Introduction

Functional RNAs fold into specific structures that form binding sites for aptamer ligands and active sites of ribozymes[1–4]. These structures range from simple stem-loops and G-quadruplexes to multi-helical junctions, pseudoknots, and their combinations, giving rise to intricate folds that endow these RNAs with their biochemical activities. The RNAs typically fold hierarchically, using base-paired regions as structural modules, and extruding single-stranded elements to make tertiary interactions, functional sites, and peripheral loops[5,6].

In vitro selection experiments harness the enormous potential of molecular evolution to yield functional RNAs[7]. The key for the success of any selection is the ability to sample a sufficient number of sequences and structures that are diverse enough to allow isolation of macromolecules carrying the desired phenotype. High-complexity DNA pools, typically composed of random sequences flanked by sequence-invariant primer-binding regions, are used as templates for *in vitro* transcription and experimental isolation of desired functional RNAs[8–10]. Because at least a weak activity has to be present in the starting pool for a selection to yield a functional RNA, the composition of the starting population is critical for the success of the entire selection experiment[11]. The phenotypic potential of the starting pool—the frequency of functional sequences—is thus a key factor of the experiment, but this property has been difficult to modulate experimentally.

A dominant property of random-sequence pools is their concentration-dependent aggregation, which potentially obscures active sequences. While longer and more diverse random regions are desirable to increase the chance of isolating complex RNA folds[12], in practice, increasing the length or concentration of a random pool may in some cases lower the yield of functional RNAs[13]. To address this issue, two approaches have previously been used to increase the structured fraction of the starting pool and yield a higher frequency of functional sequences: (i) incorporation of a structured domain next to the random sequence[14,15], and (ii) biasing the nucleotide composition of the random region to increase the prevalence of base-paired elements within the pool[16,17]. However, both of these approaches are biased towards the built-in structural features, and cannot sample a full range of RNA topologies and folds, presumably limiting their phenotypic potential.

We developed an experimental approach that decreases the aggregation of a random-sequence pool and increases its phenotypic potential. We show that isolation of the monomeric fraction of a random-sequence pool yields a population enriched in compact, structured RNAs. *In vitro* selections comparing this compact pool with a fully-random pool of the same length show that the compact pool is at least as rich in functional RNAs as a fully random pool, but because the sequence diversity of the compact pool is significantly lower, the frequency of functional RNAs is greatly increased. Thus, a pre-selection for a folded, monomeric pool provides an experimental approach to increase the phenotypic potential of random-sequence libraries.

2.3 Results

Functional RNAs are more structured than random RNAs. We first generated *in silico* several pools of random RNAs of identical length, but different nucleotide composition; each with 1 million sequences. The first RNA pool was designed to mimic a pool of non-evolved RNAs, composed of a random stretch of 70 nucleotides (nts) flanked by 20-nt primer-binding sequences. Next, we designed ten different pools, each mimicking a population of potentially functional RNAs. The first 2 pools had functional motifs consisting of well-characterized RNA aptamers, flanked by two random regions that extended the aptamer sequences to 70 nts, flanked again by the same two constant regions as in the fully random library. The three pools contained an ATP-binding aptamer (40 nts)[18], a class I GTP-binding aptamer (41 nts)[14,19], and the Malachite Green aptamer (38 nts)[20]. We also generated seven RNA pools with functional regions consisting of well-characterized ribozymes: a class I ligase ribozyme (96 nts)[21], an aminoacylase ribozyme (90 nts)[22], a kinase ribozyme (89 nts)[23], the R3C ligase ribozyme (73 nts)[24], the L1X6c ligase ribozyme (71 nts)[25,26], a class III ligase ribozyme (56 nts)[21], and a class II ligase ribozyme (87 nts)[21,27]. Because some of these ribozymes are larger than the aptamers, the functional domains were only flanked by two random regions, without any additional constant regions. The total length of each RNA sequence contained in all the RNA pools was kept constant. Next, we calculated the predicted minimum free energy (MFE) of folding for each RNA sequence in the pools, and plotted the resulting distributions of MFEs to compare the three aptamer-containing pools and the seven ribozyme-containing pools with the fully random pool (**Figure 2.1a**; see **Figure S2.1** for the MFE distributions of individual aptamer and ribozyme pools). Compared to the fully random pool, the average predicted MFE of both the aptamer and ribozyme pools was significantly shifted towards more structured RNAs, (**Figures 2.1a** and **S2.1a**).

Furthermore, the least stable tails of the distributions, representing the sequences with the lowest predicted secondary structure content, differed even more between the random and the functional pools, with the ten least-structured sequences from the random pool mapping significantly below the bottom ten sequences from any of the functional pools (**Figure S2.1a**). On the other hand, the ten most stable sequences from the random pool had MFEs similar to the most stable sequences of the aptamer and ribozyme pools (**Figure S2.1**). These calculations showed that a random-sequence pool is on average less folded and has a broader distribution of secondary structures than pools containing functional domains. The results provided a computational foundation for our hypothesis that a random RNA pool preselected for non-aggregating, monomeric sequences contains a higher proportion of structured RNA molecules, and that these folded populations are enriched for biochemically active motifs, increasing the phenotypic potential of random libraries.

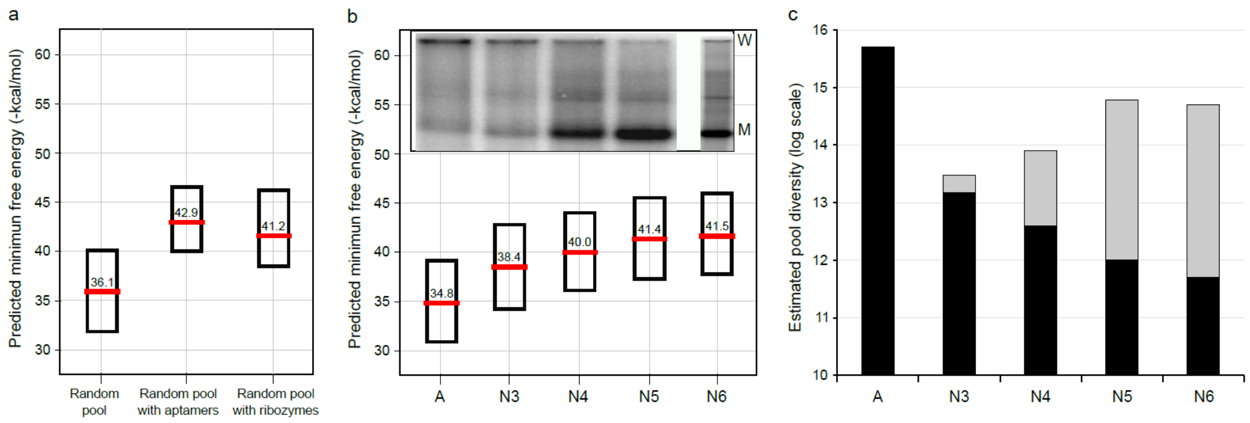


Figure 2.1. Folding propensities of random-sequence RNA pools. (a) A plot of predicted minimum free energy of folding (MFE) of *in-silico*–generated RNA pools consisting of either a fully random segment or functional motifs flanked by random regions. The functional motifs consisted of pooled aptamers or ribozymes. Red bar shows the average MFE for each pool and the boxes indicate one quartile deviation for each distribution. MFE distributions, including the sequences with the ten most extreme MFEs, of individual aptamer and ribozyme pools are shown in Figure S1a. RNA pools containing functional motifs show significantly higher MFEs, suggesting that the process of *in vitro* selection shifts the population towards more folded functional molecules. (b) *In vitro* evolution of a random RNA pool (A) from a population of aggregation-prone sequences of low average MFE into a population of monomeric RNAs results in gradual increase of average MFE, which plateaus after five rounds of enrichment (pools N3-N6). Native PAGE analysis (inset) shows that the pools become less aggregated, moving the majority of the RNAs from the proximity of the starting point of the PAGE (W, well) to a population migrating as monomeric RNAs (M). This large mobility increase is retained even at elevated pool concentrations (Fig. S2.1b). (c) Graph of estimated maximum complexities of individual pools. Black bars correspond to the sequences derived from pool A. Gray

bars show the additional sequence diversity introduced through mutagenic PCR. The complexity values assume perfect recovery of sequences from native PAGE gels and unbiased amplification. The native-PAGE selection decreased the fraction of the starting pool (A) by ~4 orders of magnitude and the mutagenic PCR introduced up to ~3 orders of magnitude sequence diversity of the monomeric RNA sequences.

Selection of monomeric subpopulation from a random RNA pool yields aggregation–resistant sequences. To test the hypothesis that the monomeric fraction of a random-sequence pool is more folded and has a higher phenotypic potential, we purified non-aggregating, monomeric sequences from a random RNA pool. First, a large-scale transcription reaction of a random-sequence DNA pool (pool A, containing 70 nts of random sequence flanked by two primer-binding regions to yield 110-nt RNA transcripts) was fractionated using a polyacrylamide gel electrophoresis (PAGE) gel composed of a denaturing top section containing urea, followed by a non-denaturing (native) section, lacking urea and containing Mg^{2+} (see Material and Methods for details). By isolating only the fastest migrating RNAs from the non-denaturing part of the gel, we reduced the overall diversity of the pool, presumably eliminating aggregation-prone molecules retained either in the well or in the denaturing portion of the gel, while increasing the prevalence of monomeric species. This purification technique was applied under progressively less denaturing conditions for a total of three times, before the first reverse-transcription and amplification were performed. Having undergone three native-PAGE purifications, this pool was termed N3. To distinguish the selected pool from the starting population, we introduced a point mutation in the 3′ constant region, changing a *StyI* restriction endonuclease binding site to that of the *MnlI* enzyme. The mutation also shifts the restriction-enzyme cut site such that a mixed pool, containing both pool-A (fully random) and pool-N (monomeric pools) sequences, could be digested by these two restriction enzymes to yield distinct bands on analytical agarose or PAGE gels, allowing a straightforward measurement of the relative proportion of the two pools.

The same native-PAGE purification was performed three more times, with reverse transcription and PCR amplification after each PAGE fractionation, yielding pools N4, N5, and N6. Furthermore, low-level (0.0076 mutations/nt/5 cycles of PCR) mutagenesis[28,29] was applied to all DNA pools (N3–N6) to increase the sequence diversity near (within a few mutations) the selected sequences. As expected, native-PAGE analysis of transcripts from the naïve (A) and selected pools (N3, N4, N5 and N6) showed that the population shifted from predominantly aggregated RNAs, immobilized near the well, to RNAs migrating predominantly as a monomeric band (**Figure 2.1b** inset). The native-PAGE analysis of the pools also showed that the monomeric pools migrated about as fast as the self-cleaved form of the genomic HDV ribozyme, a stably folded functional RNA[30]. This PAGE mobility was noticeably faster than that of a construct that does not fold into a stable structure and was used as an unfolded-control RNA of the same covalent length (**Figures S2.1b** and **2.2a**).

To estimate the sequence diversity of the selected pools, we measured the fraction of RNA recovered from each native-PAGE step and calculated the fraction of the pool A sequences remaining in each pool N3–N6 (**Figure 2.1c** and **Table 2.1**). Assuming faithful and unbiased amplification of each recovered sequence, and accounting for the additional diversity introduced by the mutagenic PCR, we also estimated the maximum complexity of each pool. Given that the starting diversity of the pool A (5×10^{15}) undersamples the theoretical complexity ($4^{70} \approx 10^{42}$) by ~26 orders of magnitude, the sequences sampled in this pool were highly sparse in the total sequence space. Any sequence diversity introduced by the mutagenic PCR in addition to the pool-A–derived sequences (~10^3-fold in pool N6) then represents local sequence variation around the monomeric, structured sequences enriched during the native-PAGE selection process, and not broad sampling of the theoretical sequence space. The mutagenic PCR is also a mechanism for potential increase of the number of functional RNAs beyond the sequences directly inherited from pool A. For the final pool, N6, the real complexity likely falls within a range defined by remaining pool A sequences ($\sim 5 \times 10^{11}$) and the additional diversity resulting from mutagenesis introduced during the amplification steps (total estimated diversity $\lesssim 5 \times 10^{14}$). Moreover, because some sequences are likely suppressed during transcription, reverse transcription, and PCR, the real molecular diversity of the N pools is likely somewhat lower than the estimated complexity range. We do not have an experimental approach to measure these biases; however, HTS analysis of the selected pools did not reveal any gross sequence bias that would result from the selection and amplification steps.

Table 2.1: Estimated complexity and composition of *in vitro* selected pools.

Pool	Maximum Complexity	Maximum Fraction of pool A
A	5×10^{15}	1
N3	$\lesssim 3 \times 10^{13}$	$\lesssim 0.003$
N4	$\lesssim 8 \times 10^{13}$	$\lesssim 0.0008$
N5	$\lesssim 6 \times 10^{14}$	$\lesssim 0.0002$
N6	$\lesssim 5 \times 10^{14}$	$\lesssim 0.0001$ $\lesssim 5 \times 10^{11}$ pool A molecules remaining

Monomeric subpopulation of random pools is enriched for more compact, structured sequences. Given the results obtained by native PAGE analysis (**Figures 2.1b, S2.1b**, and **2.2a**), we also expected that the selected pools contain more secondary structure elements than those of pool A. As a proxy measurement of secondary structure, we performed high-throughput sequencing of the naïve (A) and selected pools (N3, N4, N5, and N6), and calculated the predicted MFEs of 1 million sequences from each pool. The MFE distributions showed that the average MFE of the pool A was significantly lower than for the selected pools, while the predicted stability increased with selection round, reaching a plateau for pools N5 and N6 (**Figure 2.1b**). Notably, the average MFE of the *in-silico*–generated random pool and the sequenced random pool A had almost identical average MFEs, whereas the MFEs of the selected pools, particularly N5 and N6, were in strong accordance with the average MFEs of the *in-silico*–generated pools containing functional domains (**Figures 2.1a** and **2.1b**). These results suggested that not only was the selected subpopulation more compact and structured, but it was likely enriched for functional RNAs, as well.

We next asked whether the selection yielded monomeric RNAs because they are simply mutually non-aggregating or whether they are also resistant to aggregation-prone random sequences. As expected, analysis of the selected pools showed higher aggregation of pools A and N3 at higher concentrations (~20 *vs* ~10 µM), whereas pools N4 and N5 remained monomeric and migrated as folded RNAs at both concentrations (**Figure S2.1b**). We also transcribed pools A, N5, and N6 independently and analyzed them in mixtures with 20-fold excess of unlabeled pool A to assess the resistance to aggregation with fully random sequences. All three pools (A, N5, and N6) exhibited more aggregation in presence of excess of pool A; however, for pool A, this shift of the population was dramatic, whereas for the selected pools (N5 and 6), it was minimal (**Figure 2.2a**). These results demonstrate that the pre-selection for the monomeric sub-population of a random pool yields sequences that are not only non-aggregating among themselves, but also resistant to aggregation with fully random sequences, suggesting that the selection of the monomeric sub-population of the pool enriched for tightly and uniquely folding sequences.

To further characterize the monomeric pools, we performed several experiments designed to establish whether the monomer selection led to an increase in the fraction of structured regions. We first used S1 nuclease, which preferentially cleaves single-stranded regions of RNAs[31], to measure the susceptibility of the pools to RNase degradation. We incubated pools A, N5, and N6 with the nuclease and analyzed the populations at various time-points using high-resolution denaturing PAGE. We observed that for the pool A, the bands corresponding to the full-length RNAs disappeared faster and to a greater extent than for the selected pools N5 and N6 (**Figure 2.2b**). Furthermore, the intensity of the fragments corresponding to RNAs between 66 and 89 nts long, presumably representing longer stretches of nuclease-inaccessible secondary or tertiary structures, was also higher in the selected pools (**Figure 2.2b**). A similar effect was observed for RNA sizes of about 45 nts, whereas the opposite was observed for very short RNA sequences of 11-13 nts (**Figure 2.2b**). These results showed that the monomeric pools are enriched for more nuclease-resistant, structured RNAs.

Next, as an independent experimental approach for assessing the secondary structure content of the pools, we performed a titration of the intercalating dye ethidium bromide. Ethidium fluorescence greatly increases upon intercalation in double-stranded regions of nucleic acids[32]. If the process of monomer enrichment of the selected pools led to an increase in the prevalence of double-stranded regions, the fluorescence of the intercalating dye should increase as well. Titration of ethidium bromide into purified pools A, N5, and N6 showed a significant increase in fluorescence for the selected pools, compared to pool A (**Figure 2.2c**), again indicating a higher prevalence of base-paired regions in the selected pools.

Lastly, we performed 2′ acylation of the RNA backbone[33]. Since the acylation reaction preferentially targets solvent-accessible 2′ hydroxyls (often in dynamic, single-stranded segments), the structured positions tend to be less acylated. We hypothesized that an increase in the prevalence of double-stranded (or otherwise structured) regions in the selected pools would reduce the total number of backbone positions available for acylation. We performed a bulk acylation of purified transcripts from the A, N5, and N6 pools and analyzed the overall modification levels using high-resolution denaturing PAGE, revealing the degree of acylation as a mobility shift (**Figure S2.2a**). Pool A showed a larger shift due to a higher degree of modification, resulting from a higher solvent-accessible fraction of RNAs, than the N5 and N6 pools (**Figure S2.2b**), further supporting the observation that the monomeric pools are less accessible to backbone acylation, and therefore more compact.

Overall, the computational MFE analysis and four independent experimental methods—native PAGE, S1 nuclease digestion, ethidium intercalation, and bulk acylation—all supported the hypothesis that a selection of monomeric subpopulation of random sequences yields pools of compact, folded RNAs that exhibit no bias towards specific sequences. These results suggest that the selected subpopulation may be rich in functional RNAs and that the relative frequency of these functional molecules, that is, the phenotypic potential of the compact pools, may be significantly higher than the random-sequence pool. To test this second hypothesis, we turned to *in vitro* selections for functional RNAs from various pool mixtures.

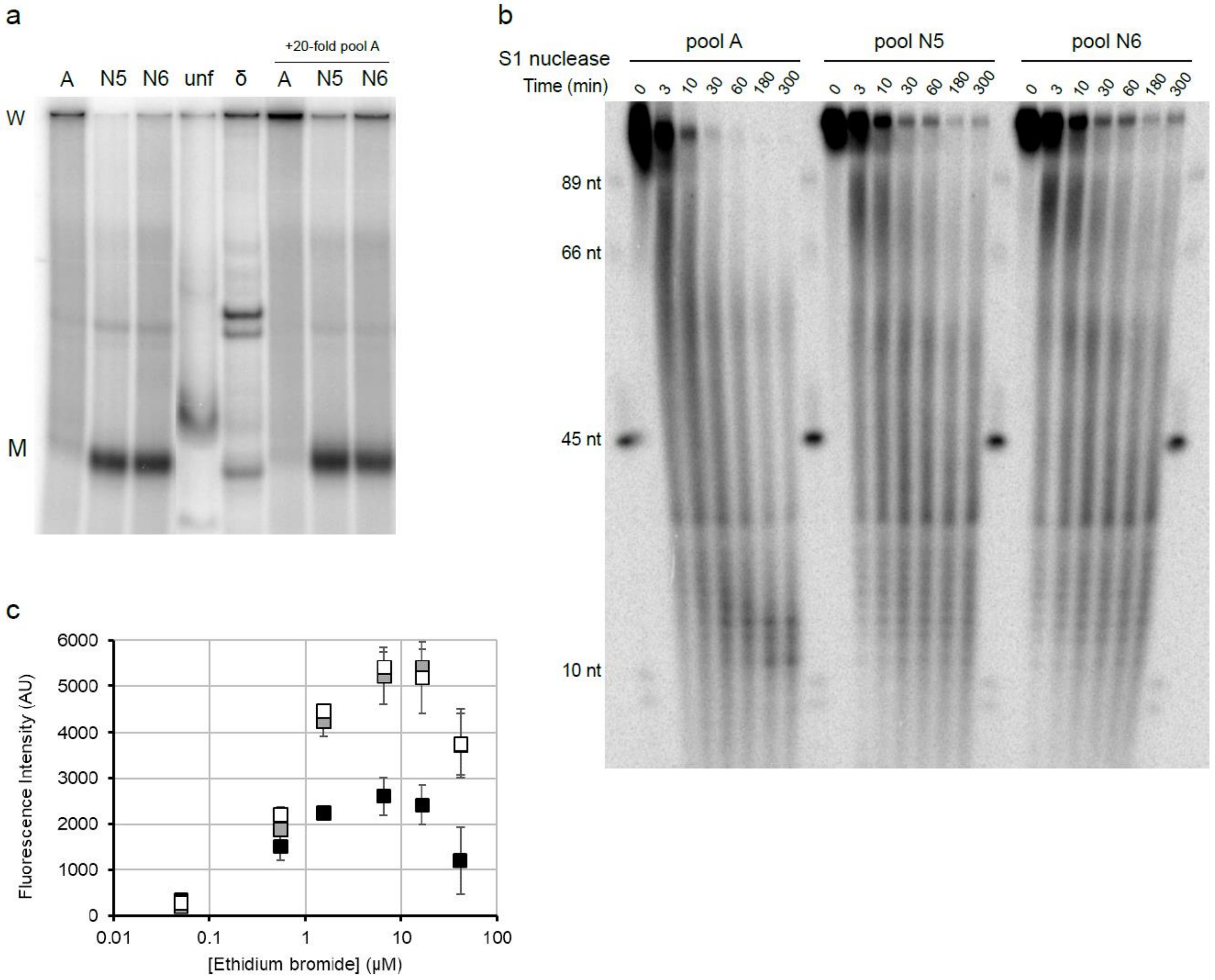


Figure 2.2. Biochemical characterization of the monomeric RNA pools. (a) Native PAGE analysis of the starting random pool (A) and the monomeric pools (N5, N6) shows that the monomeric pools migrate significantly faster than an unfolded RNA of the same length (unf) and near a genomic HDV ribozyme (δ), which represents a highly compact functional RNA. Incubation of the pools with a 20-fold excess of unlabeled pool A (right side of the gel) results in further aggregation, whereas the majority of the pools N5 and N6 is resistant to aggregation. (b) S1 nuclease probing of pools A, N5, and N6. High-resolution PAGE analysis of the S1 digestion kinetics shows faster digestion of the fully-random pool A than the monomeric pools N5 and N6, which appear to degrade to

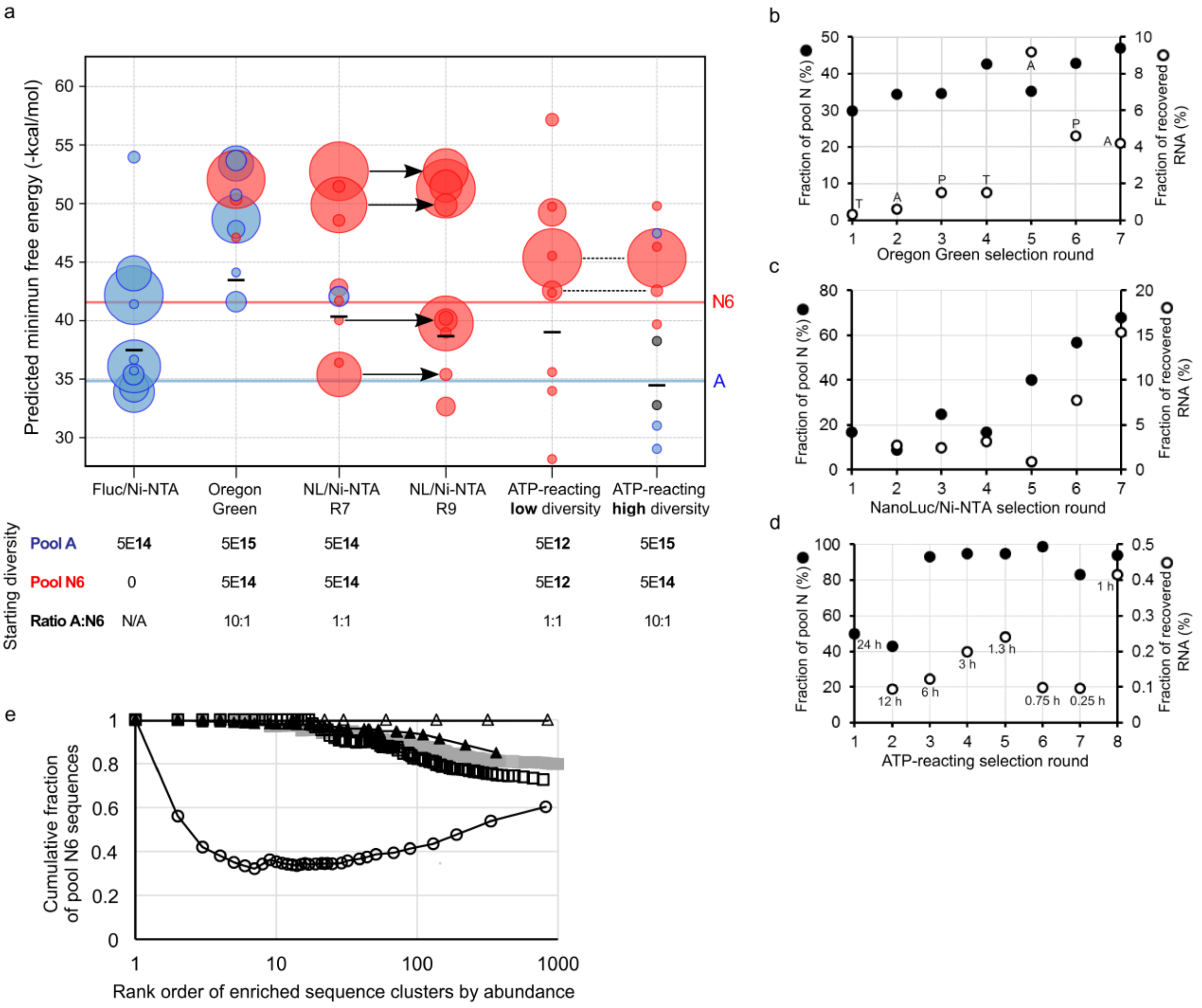
relatively stable domains above the 66-nt marker, as well as populations of shorter RNA domains, indicating that the monomeric pools are, on average, more tightly folded. (c) Fluorescence due to ethidium intercalation into the paired regions of the pools A (black squares), N5 (gray), and N6 (open). Duplicate experiments are shown for each ethidium titration, indicating that the N5 and N6 pools contain about twice as much secondary structure elements as the pool A.

Enrichment of monomeric sequences increases the phenotypic potential of the selected pools. We started by performing an *in vitro* selection of aptamers that bind the firefly luciferase (Fluc) protein immobilized on Ni^{2+}-nitrilotriacetic acid (NTA) beads, using only the naïve pool (A). The starting diversity of the DNA template was 5×10^{14}. After twelve rounds of selection, the resulting RNA pool showed detectable binding and elution from Ni-NTA beads. Analysis of the ten most abundant sequences (**Supplementary Table 2.1**) revealed a motif that bound Ni-NTA, eluted with imidazole, and was similar to a previously published Ni^{2+}-binding motif (N11 and N23)[34]. This motif—a stem-loop with a conserved CAAUUGNAAAAACG loop sequence—was present in the most enriched sequence, as well as seven others of the ten most abundant sequence clusters isolated in this *in* vitro selection experiment. Further description of this motif is presented in the NanoLuc aptamer selection below. *In silico* analysis of the sequenced pool revealed that the two most abundant sequences had MFEs well above the starting pool (A) and that the average MFE for the selected sequences was also above the average MFE for pool A, but lower than the average MFE of pool N6 (**Figure 2.3a**). Moreover, six of the ten most abundant clusters were predicted by GQRS Mapper[35] to potentially form G-quartet structures (**Supplementary Table 2.1**), which would provide further stability to the RNAs. The experiment supported the hypothesis guiding this project: that *in vitro* selection for functional RNAs shifts the population towards more structured sequences. We next turned to selections with mixed pools A and N to test whether enrichment for structured monomers also increased the phenotypic potential of the selected pools.

To explore the phenotypic capacity of the structured pools, we performed *in vitro* selections against different targets, with the pool N6 directly competing with the pool A for a selectable function. First, we performed an *in vitro* selection for aptamers binding the fluorescent dye Oregon Green (OG). Pools A and N6 were transcribed at 5 x 10^{15} and 5 x 10^{14} diversity, respectively. The pool A diversity was set to the same level as the starting diversity for the monomer selection described above; these diversities therefore represent the maximum practical diversity for pool A and maximum theoretical diversity for pool N6. The PAGE-purified transcripts from both pools were combined, and a selection for OG aptamers was performed for seven rounds, alternating the support media among agarose, polyacrylamide, and Tentagel to prevent isolation of bead-binding aptamers. After 7 rounds, the resulting pool was analyzed by HTS, revealing that the first and the third most frequent sequences originated from pool N6, with one other N6-derived sequence appearing in the ten most abundant clusters (**Supplementary Table 2.1**). The analysis further showed that these top clusters consisted of highly-structured RNAs, with MFEs well above the average MFE for the entire selected OG pool, as well as above the average MFEs of the two starting pools A and N6 (**Figures 2.1b** and **2.3a**). Apta-Seq analysis[36], a method based on 2′ hydroxyl acylation detected by reverse transcriptase termination mapping[33] of the selected pools in the presence of varying concentrations of the target ligand (**Figure S2.3**), revealed that the three most abundant sequences were likely highly structured, although the analysis yielded quantitative information about the strength of the aptamer-ligand interactions only for the most abundant sequence. Overall, HTS revealed that sequences from the structured pool (N6) represented approximately 60 % of the selected population of RNAs. In contrast, restriction enzyme analysis of the relative composition of the individual selection rounds showed a somewhat lower fraction of pool-N-derived sequences (**Figure 2.3b**), suggesting that even after seven rounds, the pool contained a large number of pool-A sequences that were not sampled by HTS. Considering that the starting population of the OG aptamer selection was 10-fold biased towards the pool A, the final distribution of the two pools was an indication that pool-N6–derived sequences were strongly enriched for OG aptamers. This result further indicated that in order for the top sequences to be selected, they had to be generally more structured than the sequences found in pool A, and implied a higher phenotypic potential of the structured pool N6.

Next, we performed a selection against Ni-NTA–immobilized Nano luciferase (NanoLuc). In this case we chose to combine pools A and N6 at the same diversity of 5 × 10^{14}, matching the maximum theoretical complexity of pool N6. As the selection progressed, the fraction of pool N6 in the total DNA population steadily increased, with N6 sequences dominating the whole population, not just the most abundant sequences (**Figure 2.3c**). Notably, even though the selection started with equal parts of the two pools, the population after just one round was heavily skewed towards pool A, perhaps because the RNAs carry forward many aggregation-prone pool-A-derived sequences, which are presumably non-functional. It then took another five rounds of selection to equalize the fractions of the two starting pools, at which point the binding became evident, and the population started to be dominated by a small number of N6-derived sequences. After 7 rounds, 9 out of the 10 most frequent sequences originated from pool N6, whereas only 1 originated from pool A (**Figure 2.3a**). The pool-A-derived sequence was the fourth most frequent, but represented only about 1/10 of the three most abundant, N6-derived sequences. Similarly to what was observed in the OG selection, the majority of the top-ten sequences had a predicted MFE either equal to, or greater than, the average MFE of the pool N6 (**Figure 2.3a**).

Sequence analysis of the NanoLuc/Ni-NTA pool revealed that the conserved RNA motif found in the Fluc/Ni-NTA selection was present in 6 of the most abundant 8 sequence clusters (**Figures 2.4a** and **2.4b**), notably, in the top 3 sequences that dominated the selection. This motif consists of a sequence-conserved (AAUUGNRAAAAC) loop flanked by a stem (P1) with a C-G base-pair closing the loop. The stem also often contains a single bulged A in the upstream strand, one or two base-pairs below the conserved loop (**Figure 2.4b**). As in the Fluc/Ni-NTA experiment, this motif maps next to different sequences in each of the six dominant clusters, implying that these six sequence clusters emerged independently. The P1 composition is diverse (except where it overlaps with the primer-binding sequences), exhibiting strong co-variation and minimal sequence conservation (**Figure 2.4a**); therefore, the helix likely serves only a structural role. To confirm the activity of the motif, we performed a column-binding assay with the Ni-NTA beads. The top three sequences (NL-C1 through C3), as well as a minimized construct corresponding to the identified motif (NL-C3-min), bound the Ni-NTA beads and eluted with imidazole (**Figure 2.4c**). Control experiments with the minimal motif (C3-min) showed similar binding and elution when presented with His_6-tagged proteins (NanoLuc, T7 RNA polymerase, Fluc, and Park7) and protein-free Ni-NTA beads. Furthermore, the motif did not exhibit binding to the proteins in gel-shift and filter-binding assays, confirming that the binding target of the motif is the chelated Ni^{2+} and not the proteins.



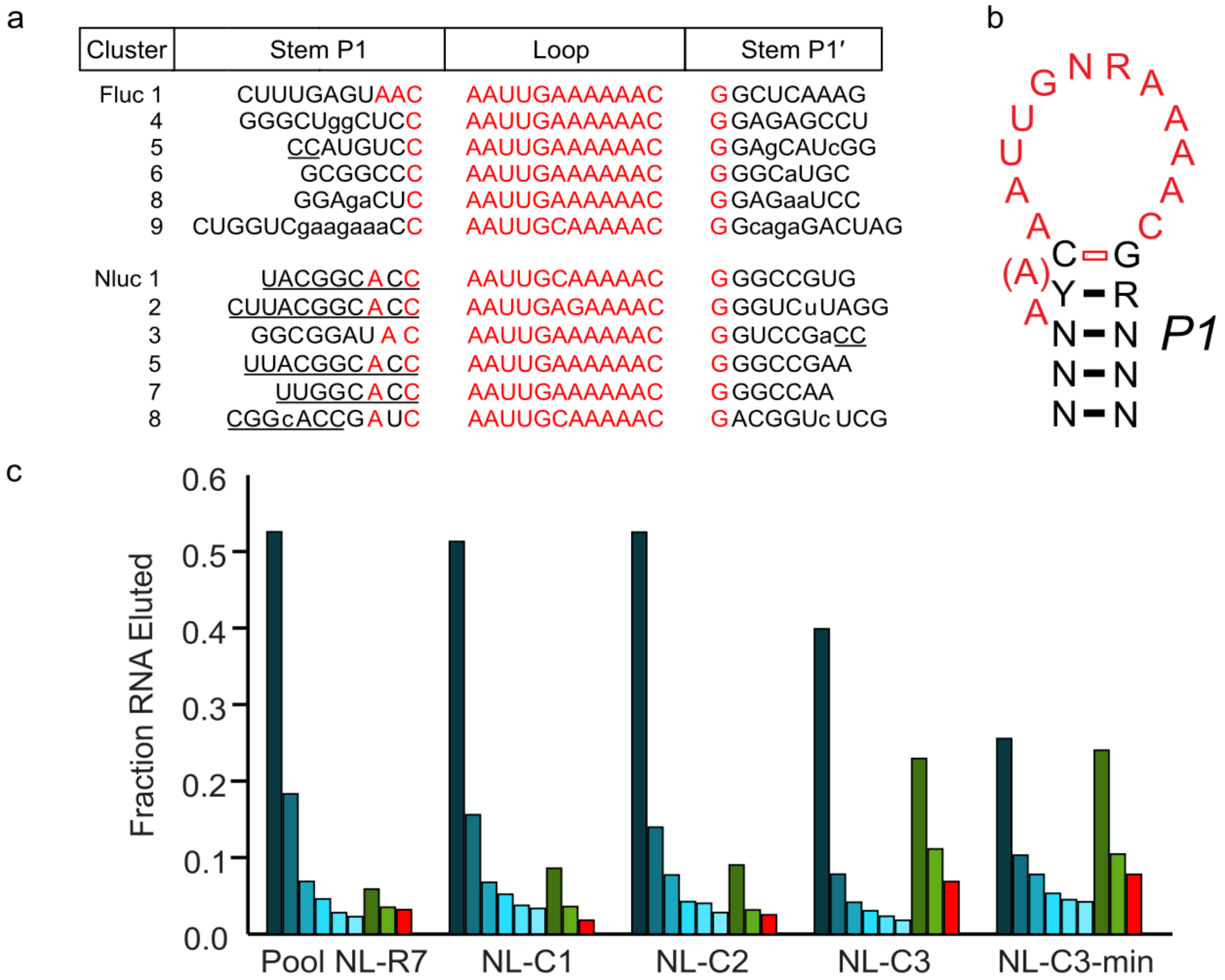
	Fluc/Ni-NTA	Oregon Green	NL/Ni-NTA R7	NL/Ni-NTA R9	ATP-reacting low diversity	ATP-reacting high diversity
Pool A	5E**14**	5E**15**	5E**14**		5E**12**	5E**15**
Pool N6	0	5E**14**	5E**14**		5E**12**	5E**14**
Ratio A:N6	N/A	10:1	1:1		1:1	10:1

Figure 2.3. *In vitro* selections from different compositions of starting pools. (a) Predicted MFEs of the ten most abundant sequences for each of the *in vitro* selection experiments. The relative abundance of each sequence indicated by the circle diameter; with blue and red circles representing sequences originating from the pools A and N6, respectively. Gray circles represent sequences which could not be assigned to neither of the two pools due to mutations. The blue and red horizontal lines represent the average MFE of pools A and N6, respectively (Figure 2.1b). The black lines represent the average MFE value of the top 10,000 most abundant sequences in each selection. The composition and diversity of the starting pool and the target of each selection is indicated

below each graph. The data show high folding propensity of the most abundant sequences in each selection experiment. Furthermore, selections from mixed pools are dominated by sequences originating from the N6 pool, even if the starting N6:A ratio is 1:10. (b-d) The relative population of the pool N6 sequences in the mixed pools (●) increased for all three selections, as the fraction of recovered RNA (○) increased with the selection rounds for Oregon Green aptamers (b), NanoLuc/Ni-NTA aptamers (c), and ATP-reacting RNAs (d). During the Oregon Green (OG) experiment (b), the beads were changed for each selection round to immobilize OG as a carboxy-OG coupled to TentaGel-NH_2 beads (T), or OG–biocytin bound to streptavidin on agarose (A) or polyacrylamide (P) beads. (d) The incubation times for the ATP-reacting RNAs are indicated above the fraction-recovered data points. (e) Analysis of the ~1000 most abundant sequence clusters for each *in vitro* selection experiment. Cumulative fraction of sequences derived from pool N6 in the last round of each selection experiment showing that for NanoLuc/Ni-NTA round 7 (gray squares) and round 9 (black squares), as well as the ATP-reacting selections starting from a high- (solid triangles) and low-diversity pools (open triangles), the population is largely dominated by the N6-derived sequences. In the OG selection (circles) the results were mixed, with the most abundant sequence originating from N6 but many other sequences among the top ~300 clusters derived from pool A.

The selection of the same Ni-NTA–binding motif in both the Fluc and NanoLuc selections allowed us to directly compare the theoretical stability of the individual motifs in the most abundant clusters. The motifs derived from the two selections had similar average MFEs, but the distribution of the predicted stabilities was wider for the sequences derived from pool A than pool N6, providing additional evidence that the selection for compact monomeric sequences yields more structured motifs.

Because the NanoLuc/Ni-NTA selection was dominated by the Ni^{2+}-binding motifs, we continued the experiment by immobilizing the NanoLuc protein on nitrocellulose membrane and isolating RNAs that remained attached to the membrane after extensive washes. After two rounds of selection the abundance of the Ni^{2+}-binding motif started to decrease, none of the sequences among the 10 most abundant clusters were derived from pool A, and the MFE analysis showed that the clusters formed well-folded RNAs (**Figure 2.3a**). The top pool-A–derived cluster was the 14th most abundant, representing under 3 % of the cumulative sequences up to that point in the ranking. Pool-N–derived sequences heavily dominated the top 1000 most abundant sequence clusters of the selection (**Figure 2.3e**). These results show that the pool-A sequences are present in the selected population, but not as strongly enriched as pool N6 sequences, supporting the conjecture that many pool A sequences are aggregating “free-riders” and that the N6 pool has a higher phenotypic potential.

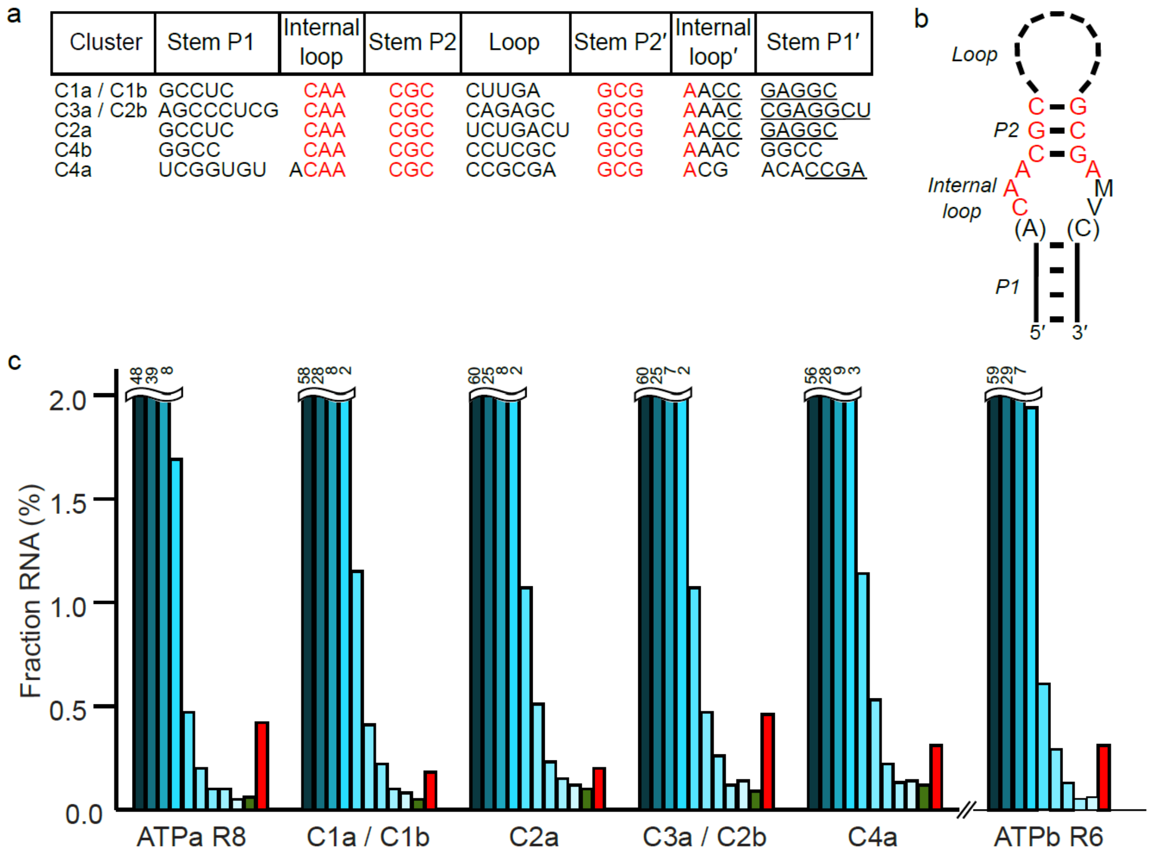


Cluster	Stem P1	Loop	Stem P1′
Fluc 1	CUUUGAGUAAC	AAUUGAAAAAAC	G GCUCAAAG
4	GGGCUggCUCC	AAUUGAAAAAAC	G GAGAGCCU
5	CCAUGUCC	AAUUGAAAAAAC	G GAgCAUcGG
6	GCGGCCC	AAUUGAAAAAAC	G GGCaUGC
8	GGAgaCUC	AAUUGAAAAAAC	G GAGaaUCC
9	CUGGUCgaagaaaCC	AAUUGCAAAAAC	G GcagaGACUAG
Nluc 1	UACGGC A CC	AAUUGCAAAAAC	G GGCCGUG
2	CUUACGGC A CC	AAUUGAGAAAAC	G GGUCuUAGG
3	GGCGGAU A C	AAUUGAAAAAAC	G GUCCGaCC
5	UUACGGC A CC	AAUUGAAAAAAC	G GGCCGAA
7	UUGGC A CC	AAUUGAAAAAAC	G GGCCAA
8	CGGcACCG A UC	AAUUGCAAAAAC	G ACGGUc UCG

Figure 2.4. A highly abundant aptamer motif from the Ni-NTA selections. (a) Sequence alignment of a motif found in both Fluc/Ni-NTA and NanoLuc/Ni-NTA aptamer selections. Residues from the constant regions are underlined, conserved residues are highlighted in red, and base-pair mismatches are shown in lowercase. The clusters depict a consensus sequence and conserved structural motif (b) consisting of a stem with potentially one bulging adenosine and a sequence-conserved loop, highlighted in red. The proposed closing C-G base-pair is conserved and therefore not supported by sequence co-variation. The bulging adenosine is observed after the Y-R base pair or immediately after the conserved C-G base pair. (c) Column binding profile of the round-7 NL/Ni-NTA pool, the top three NL clusters (C1-C3), and a minimal motif derived from

cluster 3. Fractions corresponding to flow-through and washes are indicated in blue, imidazole–eluted fractions in green, and RNA retained on resin in red. Cluster 3 shows the highest binding to the Ni-NTA beads and the NL-C3-min motif derived from this cluster exhibits similar binding and elution profile.

Catalytic potential of pools A and N6. To test the relative catalytic potential of the A and N6 pools, we performed an *in vitro* selection for ATP reactivity, following previously-described selections that yielded 5′ self-capping ribozymes[37,38], This experiment was designed to test the phenotypic capacity of the pools near the lower end of the theoretical diversity of the pool N6 (**Figure 2.1c** and **Table 2.1**). ATP-agarose beads were incubated with RNAs transcribed from DNA libraries of 5×10^{12} molecules from both pools A and N6. Similarly to the aptamer selections, purified RNAs from the two pools were mixed together at the beginning of the *in vitro* selection in order for the two pools to compete with each other. After incubation with ATP-agarose beads, 8 washes using two different denaturing buffers were performed to remove non-covalently bound RNAs. On-bead reverse transcription was followed by amplification, and after 8 rounds of selection, the pools exhibited detectable ATP conjugation (**Figure 2.3d**). Restriction enzyme analysis showed that in this selection, RNAs from the N6 pool started to dominate the selected population earlier than in the aptamer selections (**Figures 2.3b–2.3d**), essentially taking over the population by the third round. HTS of the last round of the selection revealed that the ten most abundant sequence clusters originated from pool N6 and the dominant sequences had MFEs well above the average MFE for the pool N6 (**Figure 2.3a**). Testing the top four sequences for reactivity towards immobilized ATP confirmed that these are functional RNAs (**Figure 2.5c**). The four most abundant sequences share a structural motif (a hairpin with a (A)CAA/AMV(C) internal loop, where M can be either A or C and V is any nucleotide but U; flanked by a CGC/GSG inside helix; **Figures 2.5a and 2.5b**), suggesting that this motif is important for the ATP reactivity.



Cluster	Stem P1	Internal loop	Stem P2	Loop	Stem P2′	Internal loop′	Stem P1′
C1a / C1b	GCCUC	CAA	CGC	CUUGA	GCG	AACC	GAGGC
C3a / C2b	AGCCCUCG	CAA	CGC	CAGAGC	GCG	AAAC	CGAGGCU
C2a	GCCUC	CAA	CGC	UCUGACU	GCG	AACC	GAGGC
C4b	GGCC	CAA	CGC	CCUCGC	GCG	AAAC	GGCC
C4a	UCGGUGU	ACAA	CGC	CCGCGA	GCG	ACG	ACACCGA

* a = low diversity selection / b = high diversity selection

Figure 2.5. Characterization of dominant sequences form the *in vitro* selections for ATP-reacting RNAs. (a) Sequence alignment of a motif found in both low- and high-diversity ATP-reacting selections. Residues from the 3' constant region of the pools are underlined. The clusters depict a conserved structural motif (b) consisting of a stem with an internal loop and a non-conserved loop. Conserved residues are highlighted in red. (c) ATP-column-binding profile of the selected pools (ATPaR8 – low-diversity selection, round 8; ATPbR6 – high-diversity selection, round 6) and four of the most abundant clusters that contain the conserved motif. Fractions corresponding to flow-through and washes are indicated in blue, ATP elutions are shown in green (when present), and RNA retained on ATP-agarose beads is red. The graphs were truncated at 2 %, with actual values indicated above each bar.

Given the limited diversity of the starting population and the denaturing washes during each round of enrichment, the results suggest that the selection quickly suppressed aggregation-prone “free riders” originating from pool A, and revealed the high relative abundance of functional RNAs in the pool N6. Indeed, only 0.12 % (4 out of 3340) of the sequences that showed enrichment by having at least two copies in the HTS output originated from pool A. These four sequences mapped to two distinct clusters of 2 sequences each, representing the smallest detectable enrichment. This observation indicated that the abundance of ATP-reacting sequences in the 5 × 10^{12} sequences of pool A is about 10^{3}-times lower than in pool N6, and suggested that pool N6 has much higher phenotypic potential.

To compare the catalytic potential of the two pools at the upper end of their complexities, we again combined pools A and N6 at 5 x 10^{15} and 5 x 10^{14} diversity, respectively, as we had done for the OG selection. We repeated the *in vitro* selection for reactivity with ATP and found that the dominant sequence clusters were again derived from pool N6 (**Figures 2.5a** and **2.5b**), with the two most abundant sequences matching the first and the third most abundant sequence from the previous, low-complexity selection. The appearance of the two sequences may have resulted from a contamination by the previous selection; however, none of the other high-abundance sequences from the low-complexity selection were detected in this selection, suggesting that the top two sequences (both originating from pool N6) were isolated independently. The fourth most abundant sequence (also from pool N6) contained the above-described internal loop motif, further suggesting that this motif is important for ATP reactivity. The selected pool was again heavily dominated by sequence clusters derived from pool N6, with pool-A–

derived sequences populating low-abundance sequence clusters (**Figure 2.3e**). Overall, the results from these two *in vitro* selections reinforced the conclusion reached in the aptamer selections: that enrichment for monomers leads to an increase in the phenotypic potential a random-sequence pool.

2.4 Discussion

Functional RNAs, such as aptamers and ribozymes, are found in many biological settings and have been evolved *in vitro* for a variety of roles, for example as regulatory RNAs, functional units of diagnostic and therapeutic tools, and paradigmatic catalysts in Origin of Life research. A hallmark of these RNAs is the formation of specific, compact structures that give rise to their biochemical activities[39,40]. The compaction is often hierarchical, with base-paired regions interspersed by single-stranded segments that form tertiary contacts, peripheral loops, and a variety of folds that form the binding and catalytic pockets[41]. The helical regions are the most common features of functional RNAs, with overall topologies ranging from simple stem-loops to multi-helix junctions and complex pseudoknots. Another example of a basic building block is a G-quadruplex, which can also stably extrude single-stranded segments for the formation of active motifs.

In all known cases, functional RNAs form partially solvent-inaccessible domains, which are often associated with the binding and catalytic sites, and solvent-exposed peripheral segments that often serve as topological connections or tertiary contacts[2,39]. The emergence of functional RNAs, whether *in vivo* or *in vitro*, coincides with the formation of these compact structures. Whereas the structures of cellular RNAs benefit from the long process of evolutionary fine-tuning, RNAs evolved in the laboratory typically undergo a very short selection and must therefore possess a near-perfect global fold and at least partial biochemical activity from the start of the experiment. Because at least some activity has to be present in the founding population, the phenotypic potential of the initial pool is critical for the success of the selection.

Generally, the phenotypic potential—the ability of a sequence pool to perform a selectable function—scales with the molecular diversity of the population. The theoretical sequence diversity of a pool increases as 4^N with the length (N) of a random sequence and quickly reaches levels above the practical limit, which typically depends on the synthesis scale of the templating DNA, but often remains below ~5×10^{16} (~85 nmol). A pool of 28 fully random positions has a theoretical diversity on this order (7×10^{16}) and can, in principle, sample the phenotypic potential for a biochemical activity defined by the selection process[11]. However, a 28-nt RNA tends to be too short for finding complex folds that are necessary to achieve higher biochemical activity[21,42]. For example, formation of just two 5-bp stems requires a total of 20 nt, and although some of those nucleotides can be “borrowed” from the primer-binding regions of the pool[43], at least one strand of each stem has to arise from the random region, leaving a limited number of nucleotides to define the sequence necessary for catalysis or target binding. Lengthening of the random region allows larger and more complex functional RNAs to be discovered, but the sequence space sampled by these long pools tends to be many orders of magnitude below the theoretical limit. In the case of our work, the theoretical diversity of the pool is $4^{70} \approx 10^{42}$; therefore, a starting population of ~10^{16} different molecules undersamples the theoretical diversity by 26 orders of magnitude. However, long random-sequence domains do not necessarily need to sample the entire sequence space, primarily because only a fraction of the full sequence space needs to be sampled to yield a given secondary structure[44]. For example, a simple 5-base-pair helix can be formed by 4^5 different sequences (not counting G•U wobble-pairs), providing for a large redundancy in sequences capable of forming the same secondary structure. Therefore, a random pool can undersample its theoretical sequence diversity by ~$1/4^B$, where B is the number of simple base-pairs (bps) occurring in helical segments (or $1/2^N$ for the number of nucleotides involved), while maintaining the same structural complexity.

Despite the potentially sparse sampling of the sequence space, random-sequence pools tend to aggregate, particularly in the presence of Mg^{2+} or other multivalent cations that support the formation of tertiary and quaternary structures. This propensity for aggregation may be the result of individual (unevolved) sequences to assume multiple competing conformations[45]. Aggregation leads to two effects that may adversely influence the selection outcome: one is the increased propensity of aggregation-prone sequences to act as “free-riders” binding the functional sequences that fit the selection criteria and requiring subsequent rounds of selection to purify the active species from the “free-riders”, and the other is a potential sequestration of active sequences within large aggregates that hinder their biochemical activity, such as target binding.

To increase the length of the starting pools, while presumably lowering aggregation and increasing the phenotypic potential of the starting RNA populations, several different approaches have been undertaken: 1) structured scaffolds, such as a simple hairpin[14], a binding domain[46], a three-way helical junction[47], and riboswitches[48,49], have been incorporated next to random-sequence segments; 2) a computationally-guided design of an RNA pool has been shown to yield higher proportion of multi-helical structures[17]; and 3) a highly degenerate structured DNA library was shown to increase the selection success when compared a with fully-random pool[16]. While successful at yielding functional nucleic acids, all of these approaches bias the population towards the starting scaffolds, potentially missing highly active sequences with folds that could not be sampled from these libraries. To address this problem, we explored an experimental approach aimed at increasing the prevalence of structures within a random-sequence pool by isolating the monomeric from aggregation-prone sequences and testing the phenotypic potential of these structured pools.

First, we computationally confirmed that RNA pools containing functional folds, either aptamers or ribozymes, have increased secondary-structure content when compared to random RNA pools (**Figures 2.1a** and **S2.1a**). We then confirmed that an aptamer selection from a fully-random pool leads to sequences with increased secondary-structure content and associated lower MFE (Fluc/Ni-NTA selection, **Figures 2.3a** and **2.4**). Building on these results, we experimentally tested whether purification of sequences that remain monomeric in the presence of aggregation-prone random sequences increases the prevalence of structure and whether this structured sub-population is enriched for functional RNAs. Isolation of the monomeric sub-population revealed not only that the RNAs are resistant to aggregation with random sequences, but that they are also compactly folded (**Figures 2.1b, S2.1b**, and **S2.2a**). HTS analysis of the monomeric pools showed that their average stability increases with the native-PAGE selection round, reaching a plateau after 5 rounds of selection, and matching the folding propensities of the *in silico* generated pools (**Figure 2.1b**). These pools were also notably devoid of sequences with low predicted secondary-structure content, providing a computational explanation for the low aggregation propensity of the selected pools. To experimentally verify the increased prevalence of structured RNAs in the pools, we used several biochemical approaches: analysis of the base-paired content using an intercalating dye, digestion of single-stranded segments by a nuclease, and evaluation of the solvent accessibility of the RNA backbone by an acylating agent (**Figures 2.2** and **S2.2**). All three experiments confirmed that the selection for monomeric RNAs increased the prevalence of structured regions in the random-sequence pool, and, together with the native-PAGE analyses, suggested that the selected pools were more compact and structured without regard for specific types of structure (e.g. simple secondary structure, pseudoknots, or G-quartets), although the ethidium intercalation experiment indicated a notable increase in base-paired regions in the population. The experimental data were thus in a strong accordance with the HTS analysis, showing that isolation of the non-aggregating, monomeric sub-population of a random-sequence pool yields structured, compactly folded sequences.

Previous experiments with equimolar mixtures of fully-random and stem-loop-containing RNA pools revealed that the hairpin-containing pool yielded aptamers with more structural modules and significantly higher activity[14,42]. To test the hypothesis that selection for structured RNAs leads to an increase of the phenotypic potential of the RNA pool, we also performed competitive selection experiments, but the pools were mixed in varying proportions at the beginning of each selection. The selected sequences were analyzed for their origin, MFE distribution, and biochemical function. Consistently, RNAs from the compact pool dominated each selection, even when the sequence diversity of the starting pool was biased against the compact pool. In most cases, sequences originating from the fully-random pool (A) dominated the early round of the selections, but after several rounds of enrichment, when the biochemical activity of the pools became detectable, the compact pool sequences took over the population, confirming the premise that random-sequence pools contain many aggregating “free-riders” with little or no phenotype sought by the selection experiment. Regardless of starting ratios, all mixed-pool selections showed that the most enriched sequence originated from the compact pool. Even though the restriction enzyme analysis and HTS showed that pool-A–derived sequences were abundant in most of the selected pools, in all cases, except for the OG selection, the N6-derived sequences dominated the most enriched sequences—well above the 100 most abundant sequence clusters (**Figure 2.3e**). Taken together our results consistently show that compact, monomeric sequences derived from random-sequence pools are biochemically rich, with a significantly higher phenotypic potential than the parent, fully-random pool.

Perhaps the most informative selection was performed with low starting pool diversity. The selection of ATP-reacting RNAs, which were previously shown to yield self-capping ribozymes[37,38], yielded functional sequences completely dominated by the compact-pool sequences (**Figures 2.3a, 2.3d, 2.3e** and **2.5; Supplementary Table S2.1**). The pool-A-derived sequences comprised just 0.1 % of the population and formed no significant sequence clusters, suggesting that they were not selected for but rather represented leftover “free-riders” from early rounds of the selection. These results imply that the compact pool has at least 1000-times higher frequency of functional RNAs than pool A, meaning that 5 x 10^{12} sequences of the compact pool N6 yield about as many functional RNAs as 5 x 10^{15} sequences of the fully-random pool A. Given the near-zero enrichment of pool-A sequences, the phenotypic potential is likely more than 1000-times higher in pool N6 than in pool A. This observation suggests that the original selection for monomeric sequences largely preserved the functional RNAs present in the original pool (A), because the selection decreased the number of pool-A sequences from 5 x 10^{15} to ~5 x 10^{11} (**Table 2.1**). Assuming unbiased propagation of sequences, pool N6 should be enriched for structured, functional RNAs by at least 10^4–times and the selection for ATP-reacting RNAs suggests that this enhancement is likely reflected in the phenotypic potential as well. The selection of compact, monomeric sequences thus, for the first time, represents an experimental approach to greatly increase the abundance of functional RNAs accessible by *in vitro* selections.

2.5 Conclusion

Our results show that a preparatory-scale purification can be used to enhance the phenotypic potential of a random RNA pool. By purifying RNAs transcribed from a random DNA pool using a native-PAGE fractionation, we isolated sequences that are not only more structured, but also enriched for functional RNAs. This method increases the phenotypic potential of random-sequence pools by several orders of magnitude and allows lengthening of the random-sequence regions of a pool to explore more complex structures, which can give rise to higher biochemical activity[21,42]. While the immediate application of the preparatory technique is to increase the phenotypic potential of random-sequence RNA pools in order to improve the outcomes of *in vitro* selection experiments, we note that the removal of aggregating RNAs could also be relevant for the Origin of Life and the evolution of early biosphere. Removal of aggregating RNAs molecules, whether through mineral-mediated molecular filtration, surface adsorption, or simple precipitation, increases the prevalence of stably folded sequences with higher phenotypic potential, reducing the sequence space that needs to be explored in order to find functional folds. This reduction could have implications for the emergence of the first catalytic RNAs. For example, it has been shown that a medium-sized ribozyme such as the class I ligase, with a size similar to that of the RNA pools discussed in this work, would appear only once in 3×10^{18} molecules of a random RNA sequence[21] and simpler functional motifs are likely far more common[11]. A physicochemical process that removes aggregating sequences may significantly lower the requirement of sequence diversity to yield functional molecules, thereby increasing the phenotypic potential of the emerging sequences of the RNA world. Furthermore, any process that introduces sequence diversity to such a population through recombination or single-nucleotide mutagenesis would also likely increase the phenotypic potential of the RNAs.

Finally, the enhancement of phenotypic potential of random-sequence pool brings closer the advent of *de novo* discovery of functional RNAs through screening of individual sequences. For example, given the speed of detection using ultrafast cell-sorters[50], at ~10^5 events/s, it is feasible to sort ~10^{10} droplets or beads in a day. While exploring the activity of 10^{15} sequences would be impractical, even if 100-1000 random RNAs are loaded into each droplet, it may be feasible to screen for a biochemical function by detecting the signal generated by individual sequences in an entire compact pool, such as pool N6.

2.6 Acknowledgements

This work was supported by grants from the John Templeton Foundation through the Foundation for Applied Molecular Evolution and NSF-CBET 1804220 (A. L.), NSF-IGERT Biophotonics predoctoral fellowship (M. O.), and Science Without Boarders Program – CAPES Foundation, Ministry of Education of Brazil – Process 99999.013571/2013–03 and Miguel Velez Scholarship-UCI (**L. F. M. P.**). The authors thank the following labs for their generous contributions: R.C. Spitale lab (UCI) for the acylating reagents, the Prescher lab (UCI) for purified firefly and NanoLuc luciferases, and Weiss lab (UCI) for purified PARK7 protein. The authors also thank K. A. Dunne-Dombrink for help with the Fluc/Ni-NTA selection. J. W. S. is an investigator of the Howard Hughes Medical Institute.

2.7 Materials and Methods

RNA library preparation

101-nt single-stranded DNA (Keck Biotechnology Resource Laboratory at Yale School of Medicine) containing a stretch of 70 random nucleotides flanked by constant regions (AL01, Supplementary Table 2.1) was converted into double-stranded DNA via primer extension reaction with a 39-nt DNA oligonucleotide containing the T7 promoter (AL013, Supplementary Table 2.1). The primer extension reaction was performed using homemade *Taq* DNA polymerase with a single annealing step, at 72 ºC. The resulting DNA library (pool A, ~8 nmol, corresponding to a diversity of ~5×10^{15} sequences) was used as a template for a 20 mL transcription reaction containing 50 mM Tris-HCl buffer, pH 7.5, 0.01 % Triton X-100, 2 mM spermidine, 5 mM of each rNTP (A-, C-, G-, and UTP), 10 mM GMP, 25 mM Mg(OAc)$_2$, 5 mM dithiotreitol (DTT), 1 unit of inorganic pyrophosphatase, ~40000 units of homemade T7 RNA polymerase, and 1 μL of [α-^{32}P]ATP (250 μCi/ml). The transcription reaction was incubated at 37 ºC for 5 hours. At the end of the transcription step, 10 units of RQ1 DNase I (Promega) were added to the reaction and incubated at 37 ºC for 1 hr. The reaction was then partially purified on eight Sephadex G-25 columns (NAP-25; GE Life Sciences) equilibrated with a buffer containing 100 mM NaCl, 10 mM Tris-HCl, pH 8, and 1 mM EDTA (STE buffer). This step served to partially desalt the transcription reaction and remove ~90 % of the small molecules. The flow-through solutions were combined to retain the diversity of the pool, and centrifuged using a 5-kDa cut-off filter to concentrate the macromolecules and estimate the remaining small-molecule fraction. All fractions from these purification steps were analyzed using an analytical PAGE to determine their efficiency and measure the RNA yield from the transcription reaction. Analysis of the relative intensities of the filtrates and the transcribed RNAs yielded >20 RNAs per DNA template.

Enrichment of monomeric RNAs from a random RNA library

Solid urea was added to the partially purified RNA pool to a final concentration of 4 M. The RNA library was heated briefly to about 90 ºC and loaded onto a polyacrylamide gel electrophoresis (PAGE) gel (60 x 35 x 0.3 cm^3), composed of a denaturing part (7 M urea, and 8 % acrylamide in 0.5 x TBE buffer: 44 mM Tris-borate, 1 mM EDTA, pH 7.5) with a height of 5 cm above a non-denaturing (native) part (0.5 x TBE, 10 mM Mg(OAc)$_2$, 10 % acrylamide) with a height of 30 cm. The PAGE was performed using a buffer containing 0.5 x TBE in the top reservoir and 0.5 x TBE with 10 mM MgCl$_2$ in the bottom reservoir at 11 W for 6 hr. The gel was then exposed to a phosphorimage screen, which was scanned on a Molecular Dynamics Storm scanner. About a 5-cm band of the fastest-migrating part of the RNA pool was excised from the non-denaturing section of the gel, cut into small (~1 x 5 cm^2) strips and electroeluted. Eluted RNA was ethanol precipitated, resuspended in 0.5 x TBE and 3 % glycerol buffer, and loaded onto a second non-denaturing PAGE gel (0.18 mm thick, 8 % acrylamide, 0.5 x TBE, 10 mM $MgCl_2$), and fractionated for 6 hrs at 8 W. The gel was exposed to a phosphorimage screen and analyzed as above. A thinner band, corresponding to about the fastest-migrating 1/3 of the RNA that entered the gel, was isolated, electroeluted, and precipitated. The isolated RNA corresponded to about 1.7 % of the starting RNA introduced to the first PAGE. The procedure was repeated one more time for a total of three fractionations using non-denaturing PAGE. For the third native PAGE, the RNA was loaded from a buffer that contained 5 % glycerol, 0.5 x TBE buffer, 10 mM NaCl and 10 mM $Mg(OAc)_2$. In this third fractionation, a clear band, corresponding to the monomeric pool (as checked by co-migration with the self-cleaved genomic HDV ribozyme of the same number of nucleotides as the pool), appeared and was isolated.

After the three native-gel purifications, about 0.7 nmol of the starting RNA remained. The RNA was reverse-transcribed into cDNA in a 1 mL reaction containing ~4000 units of Superscript II reverse transcriptase (RT, Invitrogen) for 1 hr at 42 ºC and a reverse primer (AL012, Supplementary Table 2.1) designed to introduce a point mutation in the StyI restriction site engineered into the starting pool (A), converting it into an MnlI restriction site. A test RT reaction with a ^{32}P-labeled reverse primer showed almost quantitative full extension reaction, suggesting that the vast majority of the RNA was reverse transcribed into cDNA. The cDNA was then PCR-amplified approximately

12-fold (using primers AL013 and AL012) in a 5 mL *Taq* DNA polymerase reaction with the following temperature cycling: 95 ºC for 45 s, 50 ºC for 90 s, 72 ºC for 180 s. The amplified DNA (pool N3) was extracted with phenol-chloroform, following by two extractions with chloroform alone, purified using Sephadex G-50 column, and ethanol-precipitated. The DNA pellet was resuspended in 0.24 ml of 0.1 x sodium chloride-tris-EDTA (STE) buffer and stored at -20 ºC. Pool N3 was *in vitro* transcribed in a volume of 1 mL, and the transcribed RNA was mixed with glycerol and urea (3 % and 1.2 M, respectively) and loaded onto another non-denaturing PAGE gel and fractionated. The band corresponding to the monomeric pool, as monitored by the location of a control sample in a parallel lane, was excised, eluted and amplified to yield the next DNA pool (N4). The procedure was repeated twice to yield pools N5 and N6, and 5 % glycerol (no urea) was used to load the RNA transcription reactions onto the native PAGE gels. To introduce local sequence diversity to the selected sequences at each round of amplification, a further hypermutagenic PCR amplification using disbalanced dNTPs and $MnCl_2$ was performed after each amplification step[28,29]. Assuming no amplification bias in transcription, reverse transcription and PCR, and unbiased mutagenesis during hypermutagenic PCR, the maximum complexity and the fraction of starting pool's sequences in the enriched DNA pools are described on Table 2.1.

High-throughput sequencing of DNA pools

HiSeq 4000 (Illumina) SR100 or MiSeq v2 Nano (Illumina) high-throughput sequencing (HTS) were used to sequence desired DNA pools. The DNA libraries of interest were prepared for Illumina sequencing by the addition of flanking TruSeq adapter sequences with two subsequent PCRs. The first PCR was performed using primers

AL2700 and AL2701 (Supplementary Table 2.1), which can prime onto the constant regions of the pool, while adding part of the TruSeq adapter sequences. Then, in the second PCR, a universal forward primer (AL2690, Supplementary Table 2.1) is used with a pool specific reverse primer inserts the unique indexing 6-nt barcode (AL2771-5, Supplementary Table 2.1), completing the flow cell binding hybridization sequence. Both PCRs were performed with DreamTaq Master Mix (Thermo Fisher Scientific), using approximately 10 nM of DNA template, and amplified for 7 cycles (denaturing 95 °C, 30 s, annealing 60 °C, 30 s, and elongation 72 °C, 30 s). The expected size of the resulting amplicons was confirmed by agarose gel electrophoresis. Barcoded DNA pools were then purified using QIAquick PCR Purification kit (Qiagen) following manufacturer instructions. Purified DNA pools were then quantified using a NanoDrop ND-1000 UV-Vis spectrophotometer (Thermo Fisher Scientific), and then mixed at an equimolar concentration (approx. 3 nM or 30 fmoles of each) for Illumina sequencing. The resulting library was then sequenced at the UCI Genomics High-Throughput Facility, where further quality control was performed on the library using a 2100 Bioanalyzer (Agilent Technologies), before sequencing, along with a PhiX DNA library spike-in of 20 %. Reads with more than a single mutation in the barcode were discarded, and the remaining reads were de-multiplexed based on their barcode.

Computational folding of RNA sequences from high-throughput sequencing data

To assess the actual enrichment of folded sequences achieved using the native-PAGE enrichment, reads from the sequenced library were filtered and de-multiplexed. Cutadapt[51] was then used to trim the 5′ and 3′ constant regions of the pools from the reads, while simultaneously discarding the reads that did not contain both constant regions, and saving the reads that did in a new file. The correctly trimmed reads, which were saved for further analysis, consisted at least 60% of the total reads generated for each barcode (ranging between 1 and 5 million reads per barcode). Next, RNAfold (ViennaRNA package)[52] was used in order to predict the minimum free folding energies of 1 million reads from each of the barcoded samples. Minimum free energies were calculated by including the 5′ and 3′ constant regions of the pool, because these regions were present during the selection process.

Computational folding of random unstructured and structured RNA pools generated *in silico*

Both “unstructured” and “structured” DNA libraries were randomly generated *in silico* using an in-house developed Python script. The “unstructured” random library was composed of 500,000 fully randomized 70-nt RNA sequences, whereas the “structured” library was composed of 500,000 70-nt RNA sequences containing two random regions flanking a defined structured RNA, chosen from a list of well-characterized aptamers and ribozymes (Supplementary Table 2.1). The lengths of the random regions were chosen to extend the length of the structured RNA to bring the total length of the generated sequence to 70-nt. A “structured” library was therefore generated for each of the chosen aptamers and ribozymes. For the aptamers, we chose the ATP-binding motif[18], a class 1 GTP-binding aptamer[14,19], and the Malachite Green aptamer[20]. For the ribozymes we chose a class I ligase ribozyme[21], an aminoacylase ribozyme[22], a kinase ribozyme[23], the R3C ligase ribozyme[24], the L1X6c ligase ribozyme[25,26], a class III ligase ribozyme[21], and a class II ligase ribozyme[21,27]. Finally, both the 5′ and 3′ constant regions of these theoretical pools were added to both the “structured” (only for the aptamers) and “unstructured” libraries. For the “structured” libraries containing the ribozymes, their bigger size would not allow for the insertion of the 5′ and 3′ constant regions, alongside the random regions, therefore the ribozymes were simply flanked by random regions up to the total length of the pool. The predicted minimum free energies of folding were then grouped and plotted into two categories: aptamers and ribozymes.

PAGE purification of RNA pools

The transcripts of interest were purified by 7.5 % polyacrylamide gel electrophoresis (PAGE) under denaturing conditions (7 M urea). RNA was eluted from the gel into 300 µL of 300 mM KCl, and precipitated by adding 1 µL glycoblue coprecipitant (Invitrogen) and 700 µL of 98 % ethanol at -80 °C.

PAGE purification of ^{32}P-labeled RNA pools

For ^{32}P labeling of the RNAs, transcription reactions were performed as described above, except only 250 µM ATP was used to drive the utilization of [α-^{32}P]ATP, and purified by 7.5 % polyacrylamide gel electrophoresis (PAGE) under denaturing conditions (7 M urea). The gels were exposed to a phosphorimage screen and visualized by a phosphorimager (Typhoon, GE Healthcare). The resulting bands of interest were excised and eluted as described above.

Secondary structure analysis of RNA pools through binding of intercalating dye

The intercalating fluorescent dye ethidium bromide (Sigma-Aldrich) was used to assess the relative fraction of base-paired regions in RNAs transcribed from different pools, as a proxy measure of the average secondary structure of each pool. An ethidium bromide titration was performed in the presence of 100 nM of purified RNAs from different pools, at the following concentrations: 0.05 µM, 0.55 µM, 1.6 µM, 6.6 µM, 16 µM and 42 μM. A negative control containing no RNA was included in the titration. The titration was carried out in 0.5 x TBE. Fluorescence emission spectra from 580 nm to 650 nm were collected on a microplate reader (Synergy H1 Microplate Reader, BioTek) with excitation 526 nm and emission 610 nm at 25 °C, using 384-well plates (clear Optilux flat bottom, BD Bioscience).

Solvent accessibility analysis of RNA pools by acylation of 2′-hydroxyls

For this experiment, 2-(azidomethyl)nicotinic acid acyl imidazole (NAI–N_3)[53] was chosen as the acylation agent (kindly donated by the Spitale lab, UC–Irvine). To reach the desired concentrations, the 2 M NAI–N_3 stock was diluted in DMSO. Acylation reactions were prepared in a total volume of 5 μL. PAGE-purified ^{32}P-labeled RNA pools were resuspended in 6 μL of water and heated to 70 °C for 3 min, and then snap-cooled on ice. For each reaction, 1 μL of the RNA pool (4000 cpm), 2 μL of water, and 1 μL of 50 mM Tris-Cl, pH 7.5, 0.5 mM EDTA were combined and allowed to equilibrate at 37 °C for 5 minutes. Each reaction was treated with 1 μL of the different NAI–N_3 solutions to reach the desired final concentration (ranging from 10 mM to 400 mM). The reactions were incubated for 60 min at 37 °C. Control reactions with no NAI–N_3 contained 1 μL DMSO. Reactions were precipitated by adding 10 μL 3 M KCl, 1 μL glycoblue coprecipitant (Invitrogen), 84 μL H_2O, and 300 μL of 98 % ethanol. The acylated RNAs pools were resolved using denaturing 7.5 % PAGE, exposed to phosphorimage screen (GE Healthcare), and scanned by the GE Typhoon phosphorimager. The band intensities were analyzed by creating lane profiles of each lane using ImageJ[54].

Nuclease probing of single-stranded regions of RNA pools

Single-stranded nuclease probing reactions were prepared in a total volume of 30 μL. Reactions were prepared by adding ^{32}P-labeled RNA pools (8000 cpm) into S1 Nuclease buffer containing 0.2 unit of S1 nuclease (Thermo Fisher Scientific). Reactions were performed at 25 °C. Individual time points were collected and quenched with loading buffer containing 7 M urea and 20 mM EDTA. The samples were resolved using denaturing 7.5 % PAGE, exposed to phosphorimage screen (GE Healthcare), and scanned by the GE Typhoon phosphorimager. The bands intensities were analyzed by creating lane profiles for each lane using ImageJ[54].

In vitro **selection of aptamers using the unstructured, random pool**

Selection of firefly luciferase (Fluc)/Ni-NTA aptamers from pool A:

For each round of *in vitro* selection, fresh His-Pur Ni-NTA resin (Thermo Fisher Scientific) was used for immobilization of hexahistidine-tagged Fluc (kindly provided by Prof. Jennifer Prescher, UC-Irvine). The resin was packed in Corning Spin-X columns and equilibrated twice with 2 resin volumes of luciferase reaction buffer (20 mM Tris-HCl, pH 7.5, 2 mM $MgSO_4$, 0.5 mg/mL BSA, 1 mM TCEP, 10 μM ATP, and 1 μM D-luciferin) and 10 mM imidazole, pH 7.5. The protein (26 nM) was immobilized on ~160 ng/ μL of resin for 10 minutes, followed by washing with the luciferase reaction buffer and 25 mM of imidazole (pH 7.5) to remove unbound proteins. An additional equilibration of the resin with 2 resin volumes of the binding buffer 50 mM Tris-HCl, pH 7.5, 140 mM KCl, 10 mM NaCl, 5 mM $MgCl_2$ was performed prior to incubation with the RNA pools.

Pool A with approximately 5×10^{14} unique sequences was used for the first round of selection. During round 1, an initial counter selection was performed for the transcribed

RNA pool on the Ni-NTA resin containing no immobilized proteins. Counter-selection, using immobilized Nanoluciferase (NanoLuc), was introduced during round 3 to remove nonspecific binders for approximately 1 hour at room temperature. The flow-through from the counter-selection was incubated with the Fluc resin with gentle shaking at room temperature. The incubation duration throughout the selection ranged from approximately 2 hours in the early rounds to 5 minutes in the later rounds. Additionally, the transcription volume was also decreased from 1 mL to 10 µL during the selection rounds. The resin was washed 4–5 times with 2 resin volumes of the binding buffer for removal of the weak binders. During rounds 5–12, increasing concentrations of guanidine HCl (50 - 200 mM) were added to the binding buffer for stringent washing of the resin. RNA-protein complexes were eluted with 4 resin volumes of 10 mM EDTA, which chelates Ni^{2+} in the resin. To increase the specificity of the selected RNAs, the elution buffer was changed to 250 mM imidazole, pH 7.5, after round 8. Also, during round 8, the transcription reaction mixture of the ^{32}P-body labeled RNA pool was passed through 3 columns of Sephadex-G10 (equilibrated with the binding buffer) for DTT removal to prevent the reduction of the Ni^{2+} in Ni-NTA resin. The eluted RNA was extracted with Trizol (Sigma), followed by precipitation with isopropanol. For selection rounds with ^{32}P-labeled RNAs, the fractions collected were analyzed on a denaturing PAGE. The elution fractions were excised and eluted in 200 mM KCl overnight, followed by precipitation with ethanol. The RNA retrieved from each round was reverse transcribed in a 20 µL reaction using M-MuLV (New England Biolabs) reaction buffer, 5 mM DTT, 2 µM reverse primer (AL045), 1 mM dNTPs, 100 unit of M-MuLV and 100 units Superscript II (Invirogen) reverse transcriptases. The reaction was performed using a temperature gradient of 37, 42, and 65 °C, for 20 minutes each, followed by inactivation of the enzyme at 95 °C for 10 minutes. From rounds 1 to 4, the cDNA was amplified using DreamTaq PCR Master Mix (Thermo Fisher Scientific) according to the manufacturer's procedures with 0.5 μM of primers AL013 and AL045 (Supplementary Table 2.1). Starting from round 5, the cDNA was amplified using both Q5 (2 units) and PFU (1 unit) DNA polymerases, ThermoPol buffer (New England Biolabs), 1 μM of each forward (AL013) and reverse primer (AL045), 200 μM each dNTP, and 1.5 mM $MgCl_2$.The DNA was initially denatured at 95 °C for 3 minutes, followed by 4-12 cycles of denaturation at 95 °C, annealing at 55 °C, and elongation at 72 °C. An additional elongation step at 72 °C was performed at the last cycle for 2 minutes. The optimal number of cycles for amplification for each round was determined using fractions collected every four PCR cycles using an agarose gel electrophoresis.

In vitro selection of aptamers using mixtures of pools A and N6

Selection of NanoLuc/Ni-NTA aptamers from pools A and N6:

Hexahistidine-tagged Nanoluciferase (NanoLuc, kindly provided by Prof. Jennifer Prescher, UC-Irvine) was immobilized on His-Pur Ni-NTA resin (Thermo Fisher Scientific). The resin was packed in Corning Spin-X columns and was equilibrated twice with 2 resin volumes of phosphate buffered saline (PBS) and 10 mM imidazole, pH 7.5. The protein (1 μM) was immobilized on the resin, followed by washing with PBS and 25 mM imidazole (pH 7.5) to remove unbound proteins. An additional equilibration of the resin with 2 resin volumes of the binding buffer 50 mM Tris-Cl, pH 7.5, 140 mM KCl, 10 mM NaCl, 5 mM $MgCl_2$ was performed prior to incubation with the RNA pools.

5×10^{14} sequences of pools A and N6 were combined for the first round of selection. The complexity of each pool corresponded to the maximum theoretical complexity of pool

N6. The transcription reaction mixture of the ^{32}P-labeled RNA pool was passed through 3 columns of Sephadex-G10 (equilibrated with the binding buffer) and incubated with the counter-selection target resin (Fluc/Ni-NTA), with gentle shaking for approximately 1 hour at room temperature to subtract nonspecific binders. This step was used until round 7, except the first, when the pools were incubated with streptavidin agarose beads bound by Oregon Green biocytin as an initial counter-selection step. During rounds 8 and 9, hexahistidine-tagged PARK7 (kindly provided by Prof. Gregory Weiss, UC-Irvine) was introduced in the counter-selection step in addition to Fluc. The flow-through from this step was then incubated with the NanoLuc/Ni-NTA beads with gentle shaking for 1 hour at room temperature. The transcription volumes ranged from 120 to 50 µL throughout the rounds of selection. The resin was washed 4–5 times with 2 resin volumes of the binding buffer for removal of weak binders. During rounds 1 and 2, the RNA was eluted twice with 2 resin volumes of 10 mM EDTA. The elution buffer was switched to 250 mM imidazole, pH 7.5 for the remaining rounds of selection. The fractions collected were analyzed on a denaturing PAGE. The elution fractions were excised and eluted in 200 mM KCl overnight, followed by precipitation with ethanol. In rounds 8 and 9, further enrichment of RNA-protein complexes was carried out using 0.45 µm pore size nitrocellulose membrane (Bio-Rad). The membrane was washed 6 times with binding buffer (15 µL) under vacuum, followed by eluting the RNA off the membrane-immobilized-NanoLuc in 8M urea and 10 mM EDTA at 65 °C. The counts for each fraction collected was were measured using the Beckman LS 6500 Multi-purpose Scintillation Counter. The elution fractions were precipitated in 200 mM KCl and ethanol, followed by reverse transcription as described above. The cDNA was amplified using both Q5 (2 units) and PFU (1 unit) DNA polymerases, ThermoPol buffer (New England Biolabs), 1 µM of forward (AL013) and reverse primer (AL045), 200 µM each dNTP, and 1.5 mM $MgCl_2$.The DNA was initially denatured at 95 °C for 3 minutes, followed by 4-12 cycles of denaturation at 95 °C, annealing at 55 °C, and elongation at 72 °C. An additional elongation step at 72 °C was performed at the last cycle for 2 minutes. The optimal number of cycles for amplification for each round was determined using fractions collected every four PCR cycles using an agarose gel electrophoresis.

In vitro selection of Oregon Green aptamers from pools A and N6:

In this experiment, pools A and N6 were employed at their maximum (practical) diversity. The unstructured pool A was transcribed from a DNA library of 5×10^{15} sequences and the structured pool N6 was transcribed from a library of $\lesssim5\times10^{14}$ DNA sequences that were derived from the random pool (A) of 5×10^{15} sequences, as described above. Thus, both pools originated from the same sequence diversity, but pool N6 was enriched for folded, monomeric sequences and its theoretical sequence complexity was no more than 5×10^{14}, and realistically, significantly lower. The two pools were transcribed, PAGE-purified, combined, and introduced to Oregon Green-linked beads, alternating Oregon Green immobilized on Tentagel-amine (Sigma-Aldrich), streptavidin agarose (Sigma-Aldrich), and streptavidin polyacrylamide beads (Pierce). The selection buffer used for washes and elution steps contained 10 mM Tris-HCl, pH 8.0, 100 mM KCl, and 10 mM $Mg(OAc)_2$. The counter-selection step contained streptavidin beads alone, whereas the selection step used the same beads bound by Oregon Green at a final concentration of ~20 µM. The RNA pools were incubated for 1 hr with the beads before washing with the binding buffer and eluting with 25 µM Oregon Green (OG) over 4 hrs. The RNA was reverse-transcribed and amplified as described above.

In vitro selection for catalytic potential of pools A and N6

In vitro selection of ATP-reacting RNAs using low diversity mixtures of pools A and N6:

5×10^{12} sequences sequence of pools A and N6 were combined for the first round of selection and were incubated with ATP-agarose beads (Sigma-Aldrich). The volume of ATP-agarose beads employed for the incubation was such that to achieve approximately 5 mM ATP. The incubation was done at room temperature, using a buffer composed of 50 mM Tris-HCl (pH 7.6), 25 mM $MgCl_2$ and 150 mM KCl, and the first round was incubated for 24 hrs, shaking. For the next rounds the incubation time was reduced: round 2 to 12 hrs, round 3 to 6 hrs, round 4 to 3 hrs, round 5 to 80 minutes, round 6 to 45 minutes, round 7 to 15 minutes, and round 8 to 60 minutes. After incubation, the beads were washed using two different denaturing buffers, employed one after the other in different washes. The first denaturing buffer contained 7 M urea and 50 mM EDTA, while the second denaturing buffer contained 95% formaldehyde and 22.5 mM EDTA. Both denaturing buffers were pre-warmed at 40 °C. The first four denaturing washes were performed by incubating with either the first or the second denaturing buffer followed by shaking for 5 minutes at room temperature. The fifth and the sixth washes were shook for 20 minutes at room temperature. The seventh and the eighth washes were performed with no additional incubation times. Next, an RT reaction on beads was incubated at 37 °C for 45 minutes. The RT reaction was composed of 0.5 µM of DNA primer (AL045, Supplementary Table 2.1), 1x First Strand Buffer (Invitrogen), 10 mM dithiothreitol (DTT), 0.5 mM dNTPs each, and 100 units of M-MuLV reverse transcriptase. Afterwards, the cDNA was amplified using the DreamTaq PCR Master Mix (Thermo Fisher Scientific) with 0.5 μM of primers AL013 and AL045 (Supplementary Table 2.1). To assess that the amplification was due to RNA bound to the beads, and not RNA that was being washed off, an additional non-denaturing wash (1x TE buffer) after the last denaturing wash was done, and that non-denaturing wash was also subjected to RT and PCR using the same conditions as the beads. While amplification from the beads was detectable after approximately 12 rounds of PCR, no amplicons were detected after up to 30 cycles of PCR for the negative control. The amplified DNA was then used as the template for next round of selection.

In vitro selection of ATP-reacting RNAs using high diversity mixtures of pools A and N6:

5×10^{15} sequences of pool A and ~5×10^{14} sequences pool N6 were combined for the first round of selection. The first round of section comprised on an initial counter-step incubation with ADP-agarose beads (Sigma-Aldrich) for 6 hours followed by the addition of ATP-agarose beads (Sigma-Aldrich) to the mixture for 22 hours, shaking. The volume of ADP- and ATP-agarose beads employed for R1 incubation was such that to achieve approximately 0.7 mM of each. The incubation was done at room temperature, using a buffer composed of 50 mM Tris-HCl (pH 7.6), 25 mM $MgCl_2$ and 150 mM KCl, and the first round was incubated for 24 hrs, shaking. For the next rounds, the volume of ATP-agarose beads employed was such that to achieve approximately 5 mM, and the incubation time was reduced: round 2 to 15 hrs, round 3 to 6 hrs, round 4 to 3.5 hrs, and round 5 to 2 hrs. The incubation time for round 6 was 4 hrs. After incubation, the exactly same procedure was performed as abovementioned for the *in vitro* selection of ATP-reacting RNAs using low diversity mixtures of pools A and N6.

Restriction digestion analysis of selected pools composition

To monitor the relative enrichment of sequences from Pool A or Pool N6 throughout the selections, DNA from each round was treated with the restriction enzymes *StyI* and *MnlI*, which cut the pool A and pool N in their reverse-primer binding regions, respectively. A 5 µL reaction comprising 2 µL DNA, CutSmart buffer 1× (New England Biolabs), and 0.5 µL of either restriction enzyme was incubated at 37 °C for 30 minutes. The samples were analyzed using an agarose gel electrophoresis, and the results were quantified using ImageJ[49].

Column binding assay of NanoLuc/Ni-NTA selection top clusters

Ligand binding for the top clusters and cluster 3-min from round 7 of the NanoLuc selection were tested using a column assay in which 10 µL of ^{32}P-labeled RNAs were incubated with 10 µM of the target protein immobilized on Ni-NTA resin as described above. The resin was washed 5 times with 2 resin volumes of the binding buffer, followed by 2 elutions with 250 mM imidazole, pH 7.5. The counts for each fraction collected were measured using the Beckman LS 6500 Multi-purpose Scintillation Counter.

Column binding assay of ATP-reacting selections and top clusters

^{32}P-labeled RNAs from rounds of interest were transcribed and PAGE purified as described above. PAGE purified, ^{32}P-labelled RNA was then incubated with ATP-beads (Sigma-Aldrich) in the same buffer used in the selections: 50 mM Tris-HCl (pH 7.6), 25 mM MgCl2, and 150 mM KCl. The reactions were incubated at room temperature with shaking for 6 hours (ATP low diversity R8 and top clusters) or 4 hours (ATP high diversity R6). Afterwards, the eight denaturing washes were performed on the beads as previously described in the selection procedure, with alternating the two different denaturing buffers.

The flow-through from the RNA incubation step, and each of the washes were collected for scintillation counting. The elution step was performed using the same buffer used in the selection, but with an additional of 5 mM ATP (no elution step was performed for the ATP high diversity R6). The elution was incubated for 5 minutes with shaking at room temperature. After the ATP elution was removed through centrifugation, the dry beads were also collected for counting. All the collected fractions were then counted using a Beckman - LS 6500 Multi-purpose Scintillation Counter.

Clustering and ranking of top RNA sequences from high-throughput sequencing of selected RNA pools

Selection rounds, sequenced using Illumina HTS as described above, were analyzed in order to rank and cluster the most abundant clones present in each selection. After quality-filtering and de-multiplexing reads, the 5′ and 3′ constant regions of the pools were trimmed using Cutadapt[51]. Reads lacking both the 5′ and 3′ constant regions were discarded, and the remaining reads comprised at least 60 % of the starting reads. Next, Fastaptamer[55] was used to group and count identical reads. Counted reads were ranked by abundance and saved in a different file, along with the associated RPM (reads per million) of each read. Finally, Fastaptamer[55] was used for counted reads with an RPM higher than 500 for the NanoLuciferase selection in order to cluster non-identical reads differing by at most 3 mutations. For the Oregon Green and the Firefly Luciferase selections, counted reads with an RPM higher than 10 were clustered using Fastaptamer[55], allowing a clustering of non-identical sequences differing by at most 15 mutations. For the ATP-reacting selections, counted reads with RPM higher than 2 were clustered using Fastaptamer[55], allowing a clustering of non-identical sequences differing by at most 7 mutations. Additionally, for sequences derived from selections from a mixture of pool A and N6, in-house developed Python script was employed to distinguish the origin (of those top sequences) either from the pool A or the pool N6.

Apta-Seq analysis of the selected RNA pools

Apta-seq[36] was performed to assess the response of selected RNA pools to the ligands they were selected for. RNA from the last round of the Oregon Green selection was transcribed and purified as previously described. Purified RNA was then subjected to selective 2′-hydroxyl acylation and primer extension (SHAPE). SHAPE reactions were prepared in a total volume of 10 µL and are described as follows: the purified RNA pool was resuspended in water and heated to 70 °C for 3 min. 70 nM of purified RNA was added to a buffer containing 10 mM Tris-HCl, pH 7.5, 100 mM KCl, and 10 mM $Mg(OAc)_2$. A dilution series of Oregon Green (OG) was prepared using a 1:1 stock of OG:DMSO. 10 nm to 100 µM OG were aliquoted to the RNA in buffer and incubated at room temperature for 30 minutes. 100 mM NAI-N_3 was added to the mixture and the reaction was incubated for 30 minutes at room temperature. Reactions containing no SHAPE reagent were substituted with 10 % DMSO. Reactions were precipitated with 10 µL 3 M KCl, 1 µL glycoblue, 89 µL H_2O, and 300 µL 98 % ethanol. After recovery of the acylated RNAs, all samples were reverse-transcribed to cDNA with M-MuLV reverse transcriptase (Invitrogen). The DNA primer used for reverse transcription was AL2701 (Supplementary Table 2.1), and the samples were subsequently ligated using T4 DNA Ligase (New England Biolabs). The RT reaction was composed of 0.5 µM of DNA primer, 1x First Strand Buffer (Invitrogen), 10 mM dithiothreitol (DTT), 0.5 mM dNTPs each, and 100 units of M-MuLV reverse transcriptase. Extensions were performed for 60 minutes at 37 °C.

The resulting DNA library was then ligated to a phosphorylated DNA oligonucleotide using T4 DNA Ligase.

T4 Ligase reaction and PCR amplification

T4 Ligase reaction was performed following Ritchey et al.[56], with some modifications. A phosphorylated DNA oligonucleotide (AL3340, Supplementary Table 2.1) was ligated to the DNA library generated with reverse transcription. The DNA oligonucleotide contained a random hexamer to serve as a splint for the 3′-end of the DNA library, and a hairpin to bring the 5′-phosphate of the DNA oligonucleotide into close proximity of the 3′-OH of the DNA library, thereby facilitating the ligation reaction. The ligation reaction contained the following components: 1 µM phosphorylated oligonucleotide AL3340, 500 mM betaine, 10 % PEG 8000, along with approximately 20 nM of DNA library. The reaction was heated at 70 °C for 3 minutes, and then cooled on ice for 1 minute. Afterwards, 1 X T4 DNA ligase buffer and 400 units of T4 DNA ligase (New England Biolabs) were added to initiate the ligation reaction. The ligation reaction was then incubated at 16 °C for 6 hours, followed by another 6 hours at 25 °C. Afterwards, the ligation reaction was diluted 1:200 and used as template for a PCR amplification with primers AL3220 and AL2701 (Supplementary Table 2.1) using the DreamTaq Master Mix (Thermo Fisher Scientific). Each primer was present at a final concentration of 200 nM, and 20 cycles of PCR were performed using the following conditions: denaturation at 95 °C, 30 s, annealing at 60 °C, 30 s, and elongation at 72 °C, 30 s. Finally, 5 units of the restriction enzyme *AfeI* (New England Biolabs) were added to the PCR, and the reaction was incubated at 37 °C for 1 hour. *AfeI* was used in order to reduce the prevalence of an unspecific amplicon composed of a ligation reaction between the RT primer and the phosphorylated primer. The phosphorylated DNA oligonucleotide (AL3340) was designed as to insert an *AfeI* restriction site upon ligation with the RT primer, AL2701. In our experience, such ligation was problematic for the second round of PCR amplification needed to barcode the samples for Illumina sequencing; therefore, we exploited the insertion of a new restriction site in order to reduce the prevalence of such an amplicon from the first PCR in the ligated DNA library. The *AfeI*-digested PCR was diluted 1:200 and used as a template for a final PCR amplification with the universal Illumina forward primer AL2690, using unique reverse barcoding primers (AL2771-2791, Supplementary Table 2.1), which complete the required Illumina adapter sequences and inserts unique TruSeq barcodes for the different samples of each Apta-Seq experiment. PCRs were amplified for 8 cycles with DreamTaq Master Mix (Thermo Fisher Scientific) using the following conditions: denaturation at 95 °C, 30 s, annealing at 60 °C, 30 s, and elongation at 72 °C, 30 s. Resulting amplicons were then analyzed using agarose gel electrophoresis.

PAGE purification of barcoded libraries

The final amplification of the barcoded DNA samples required further PAGE purification before sequencing. All SHAPE-analyzed samples were pooled together, without the no-SHAPE (un–acylated) control, and both DNAs were PAGE-purified to exclude a band at ~150 bp, which corresponded to the self-primed forward and reverse Illumina sequences. The purified fractions were then analyzed using either Illumina-specific forward and reverse primers, and pool-specific primers to assess the % of self-primed, versus cDNA-derived DNA in each fraction. The concentration of each fraction was measured by a NanoDrop 1000 Spectrophotometer (Thermo Fisher Scientific), agarose gel electrophoresis, and diagnostic PCR.

Code availability

All in-house developed Python scripts are available by direct requests to the corresponding author.

2.8 Supplementary Data

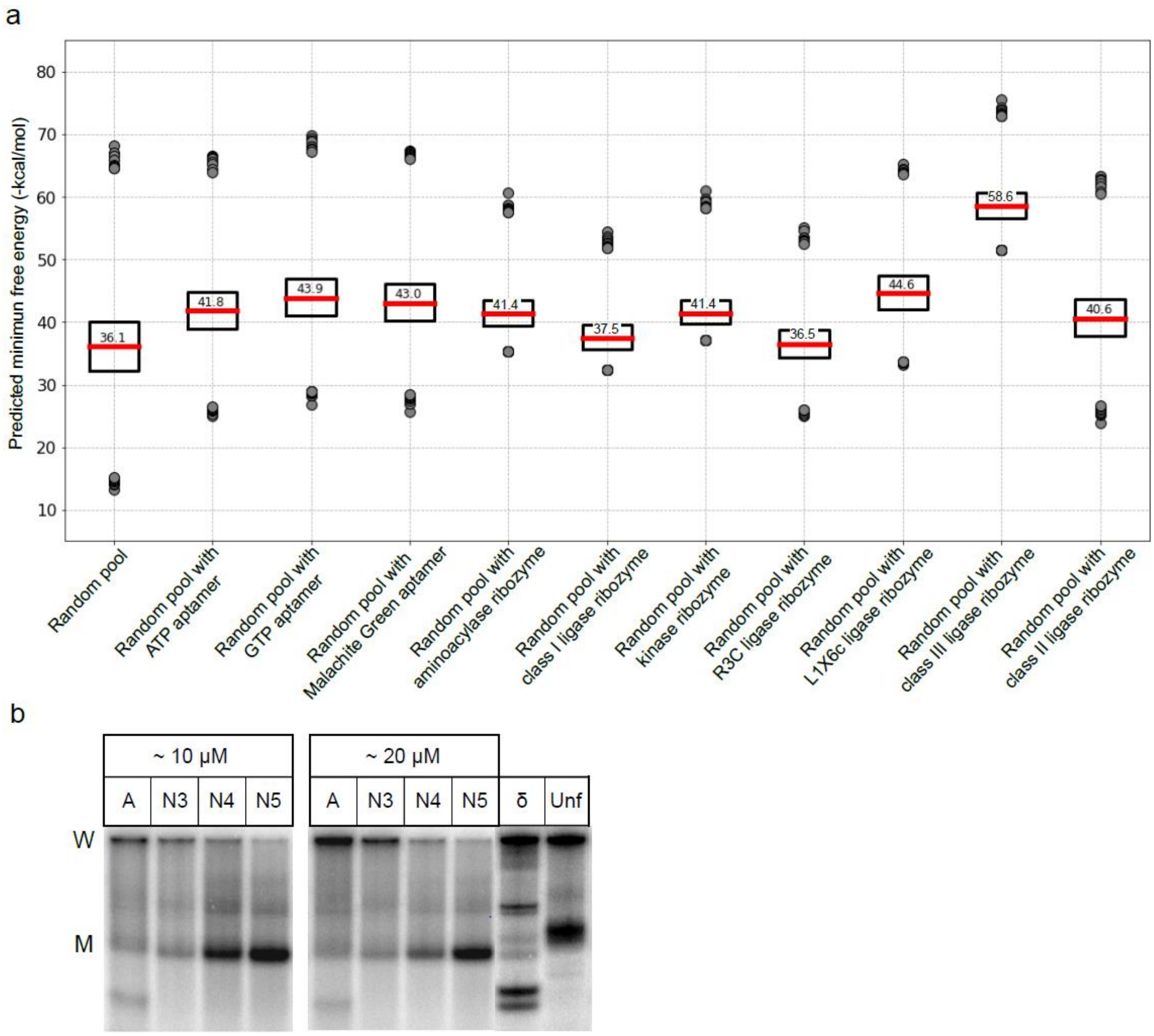


Figure S2.1. ***In silico* and native-PAGE analysis of RNA pools.** (a) A plot of predicted minimum free energy of folding (MFE) of *in-silico*–generated RNA pools consisting of either a fully-random segment (Random pool) or functional RNAs, with specific aptamers or ribozymes indicated below the graph, flanked by random regions to make the total

length of the RNAs identical to the pool A. Red bar shows the average MFE for each pool (the numbers give the calculated value of the average), the boxes indicate one quartile deviation for each distribution, and the gray circles show the MFEs of the ten most stable or unstable sequences. The random pool shows lower average MFE than any of the other *in-silico*–generated pool, but the ten most stably folded sequences are on par with the most stable sequences from the pools with functional domains. On the other hand, the ten least stable sequences of the random pool have significantly lower MFEs than the least stable sequences of any other pool, suggesting that unstructured RNAs are diminished during an *in vitro* selection. (b) Native PAGE analysis of the pool A and its derivative pools enriched for monomeric sequences (pools N3–N5) at ~10 µM and ~20 µM concentrations, showing that most of the pool A aggregates near the well (W), whereas the pools N3 through N5 exhibit progressively less aggregation, even at a higher concentration. Pools N4 and N5 predominantly migrate similarly to the genomic HDV ribozyme (δ), a folded RNA largely consisting of helical segments and a tightly folded tertiary structure. The PAGE mobility is significantly faster than an unfolded control sequence (unf) of the same length.

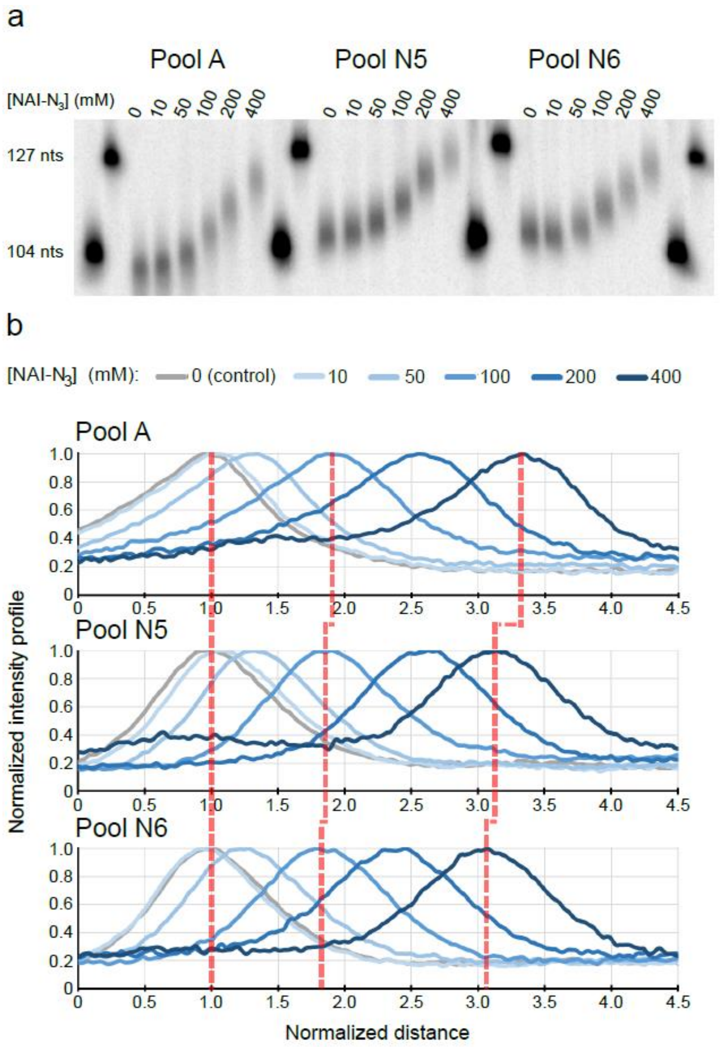


Figure S2.2. Solvent accessibility analysis of pools A, N5, and N6 assessed by 2′-acylation. (a) High-resolution PAGE gel of acylated pools under increasing concentrations of NAI-N_3, which acylates the RNAs at solvent-accessible 2′ positions. The gel shows increasing retardation due to increased bulk acylation of the RNAs at higher NAI-N_3 concentrations. (b) Normalized migration profiles of acylated RNAs for all three pools. Peak positions of control, 100 mM NAI-N_3, and 400 mM NAI-N_3 profiles are marked with the red trace line, highlighting the differences of RNA mobility due to acylation for each pool and showing lower bulk acylation of the pools N5 and N6 than pool A.

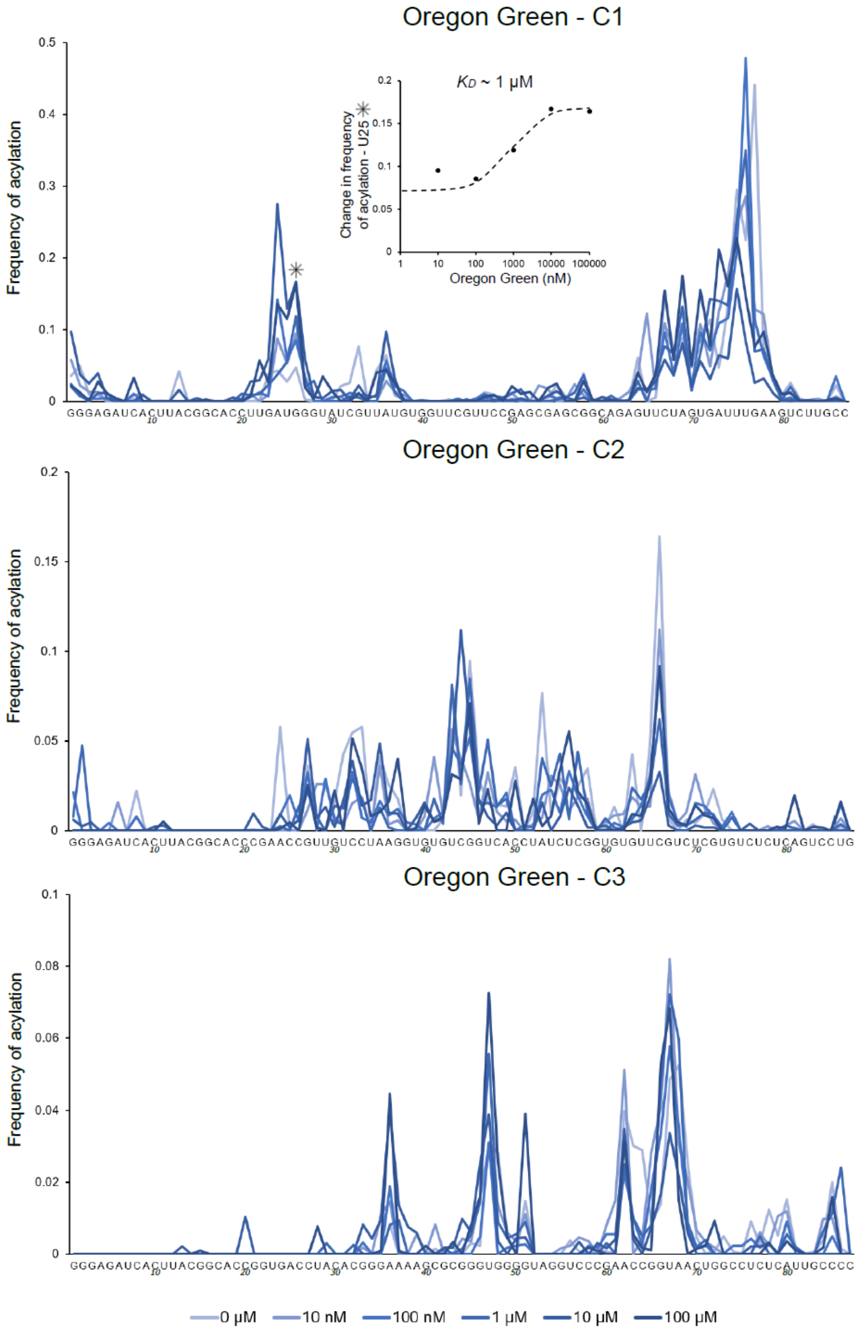


Figure S2.3. Apta-Seq analysis of the top clusters from the Oregon Green selection. Graphs depicting acylation positions for the top 3 clusters from the selected pool with increasing concentrations of the Oregon Green dye, ranging from 0 µM (light blue) to 100 µM (dark blue). Binding isotherm for the cluster 1 is modeled from frequency of acylation at the position U25, revealing a dissociation constant of ~ 1 µM.

Supplementary Table 2.1. DNA sequences used for project. Calculated minimum free energy (MFE) for the top 10 clones of all *in vitro* selections is shown. G-quartet prediction data is also presented when applicable.

Oligo name	**Sequence**	
AL01 (pool A reverse complement)	5’-CGCTGTCCCGAGCCTTGG N70 GGTGCCGTAAGTGATCTCCC-3’	
AL013 (forward primer for pools A and N)	5’-GCTAATACGACTCACTATAGGGAGATCACTTACGGCACC-3’	
AL012 (reverse primer for pool N)	5’-CGCTGTCCCGAGCCTCGG-3’	
AL045 (reverse primer for pools A and N)	5’-CGACGCGCTGTCCCGAGC-3’	
AL2700	5’-CACGACGCTCTTCCGATCTGGGAGATCACTTACGGCA C-3’	
AL2701	5’-AGACGTGTGCTCTTCCGATCTCGCGCTGTCCCGAGC-3’	
AL2690	5’-AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT-3’	
AL3340	5’-pho-GCT AGATCGGAA GAGCGTCG TTCCCATCT AGC NNNNNN CTGCCCATAGAG-3’-spacer	
AL3220	5'-CTACACGACGCTCTTCCGATCT-3'	
Structured RNAs		
ATP-binding Sassanfar motif	5'-GGGUUGGGAAGAAACUGUGGCACUUCGGUGCCAGCAACCC-3'	
Class 1 GTP-binding aptamer	5'-GGGACGAAGUGGUUGGGCGCUUCGGCGUGUGAAAACGUCCC-3'	
Malachite green aptamer	5'-GGAUCCCGACUGGCGAGAGCCAGGUAACGAAUGGAUCC-3'	
Class I ligase ribozyme	5'-GGAAAAGACAAUCUGCCCUCAGAGCUUGAGAACAUCUUCGGAUGCAGGGGAGGCAGCCCCCGGUGGCUUUAACGCCAACGUUCUCAACAAUAGUGA-3'	
Aminoacylase ribozyme	5'-GGGAGAGGAUACUACACGUGUCCCAUCCAAGACCUCGAUAAAAUCGGUCGACUUGUCUGCGCUGCCCAUUGCAUGUAGCAGAAGCUUCCG-3'	
Kinase ribozyme	5'-GGGAGAGGAUACCACACGUAUCCAAUGCAAGACGUCGAUAAAAUCGGUCGACUUGUGUGAACUGCGAAGUGAUGUAGCAGAAGCUUCCG-3'	

R3C ligase ribozyme	5'-GAGACCGUAAUGAGUAGUACUUAUUAUGCUCGAUUGUUCGUAAGAACAGUUUGAAUGGGUUGAAUAUAGUGAG-3'	
L1X6c ligase ribozyme	5'-GACCUCGGCGAAAGCCGUUCGACCAGCGAAAGCUCUUAGACAGGAGGUUAGGUGCCUCCGAAAGGAGCACU-3'	
Class III ligase ribozyme	5'-GGAACACUAUCCGACUUCGGUCGGUGGAGAUGUAUAGUCUUAGGGUGAGGCUGGUA-3'	
Class II ligase ribozyme	5'-GAACCAGUCGGAACACUAUCCGACUGGCACCCGUUUUCGGGUGGGGAGUGCCUAGAAGUGGGUAGGUCUUUUUAGACCGCCUAGGCC-3'	
Top 10clusters from each selection for Figure 2.3a		**Calculated MFE (-kcal/mol)**
Firefly luciferase selection		
Cluster 1	5'-GGGAGAUCACUUACGGCACCACCGGGGAGAUCAAACCUUUGAGUAACAAUUGAAAAAACGGCUCAAAGCUAUUGCAGGUAUAACCUGCAUCCCAAGGCUCGGGACAGCG-3'	42.31
Cluster 2	5'-GGGAGAUCACUUACGGCACCUCUGCGAGCUUGGCAUAAAUCCGACAGGUAGCCAGACACAGUAGGCAGCACGUUUUUCCGCACGCAUGCACCAAGGCUCGGGACAGCG-3'	36.01
Cluster 3	5'-GGGAGAUCACUUACGGCACCACCGAGUAAAGGUGCGACUGUUUUCUGACCGUGACGUUUUGGGUUUGAAUAUGGAUGGAAAAACUGGCGGCCAAGGCUCGGGACAGCG-3'	33.99
Cluster 4	5'-GGGAGAUCACUUACGGCACCUGAUCGGGCUGGCUCCAAUUGAAAAAACGGAGAGCCUCUACAGCGCAAUUGUCGAGUGAGGUUGUGCUUGUCCCCAAGGCUCGGGACAGCG-3'	43.76
Cluster 5	5'-GGGAGAUCACUUACGGCACCAUGUCCAAUUGAAAAAACGGAGCAUCGGACGUCCUCUAAGUUGAUCAGCGCGCGGGGAGUCUAUGUCGCAAGCCAAGGCUCGGGACAGCG-3'	34.59
Cluster 6	5'-GGGAGAUCACUUACGGCACCUACCUAUGGGGCGGCCCAAUUGAAAAAACGGGCAUGCGGAGUCUCUUUAGGAUGCGUGAUGUGUAAUCGUGCCAAGGCUCGGGACAGCG-3'	35.88
Cluster 7	5'-GGGAGAUCACUUACGGCACCGGCAGGUCAAAGUCUGUGAGUGACUUUCGAACCGCGUUCGUAGUAGGGCACCGGGCGAUAGCUCCGUGUCAGCCAAGGCUCGGGACAGCG-3'	53.76

Cluster 8	5'-GGGAGAUCACUUACGGCACCGAUCAUGACUCUCCGAAUUCGUGGGGAGACUCAAUUGCAAAAACGAGAAUCCAGAGAUAGUUUUCCUGUCCCAAGGCUCGGGACAGCG-3'	35.71
Cluster 9	5'-GGGAGAUCACUUACGGCACCCAUGCUACUGCCCGAGUAUUUCGUCUGGUCGAAGAAACCAAUUGCAAAAACGGCAGAGACUAGUUCCAAGGCUCGGGACAGCG-3'	36.44
Cluster 10	5'-GGGAGAUCACUUACGGCACCUCGCCUGCGCGACGUACGCGCCGUACAACGUUUGUCGUACCCCCCCUUCGGUCCCAACUUGCUACCUCUUACCAAGGCUCGGGACAGCG-3'	37.73
Oregon Green selection		
Cluster 1	5'-GGGAGAUCACUUACGGCACCUUGAUGGGUAUCGUUAUGUGGUUCGUUCCGAGCGAGCGGCAGAGUUCUAGUGAUUUGAAGUCUUGCCUGCCCGAGGCUCGGGACAGCG-3'	52.11
Cluster 2	5'-GGGAGAUCACUUACGGCACCCGAACCGUUGUCCUAAGGUGUGUCGGUCACCUAUCUCGGUGUGUUCGUCUCGUGUCUCUCAGUCCUGUCGCCAAGGCUCGGGACAGCG-3'	48.66
Cluster 3	5'-GGGAGAUCACUUACGGCACCGGUGACCUACACGGGAAAAGCGCGGGUGGGGUAGGUCCCGAACCGGUAACUGGCCUCUCAUUGCCCCUGGCCGACCAAGGCUCGGGACAGCG-3'	53.45
Cluster 4	5'-GGGAGAUCACUUACGGCACCCUCGUUGUGGGCUUUAGGUGGCCCGGCCUAGCGUUCUGCGCCCUGAUUAAGAAAUAGCUGUCCACUUUUCCAAGGCUCGGGACAGCG-3'	41.87
Cluster 5	5'-GGGAGAUCACUUACGGCACCCUGUGGGGGUGGCUGUCCCCGUUUGCUGCAAUCCCCAGCAGGACCGUCUUGAACCUUGGAAACCAACGUCCAAGGCUCGGGACAGCG-3'	53.65
Cluster 6	5'-GGGAGAUCACUUACGGCACCUUGUUCGGAACUGCUGUCCAACCGAGAAUCGUGUCUCUGAUCUUCCCCGUUUUGUCCGAGUUUAGGGUGUCCAAGGCUCGGGACAGCG-3'	47.55
Cluster 7	5'-GGGAGAUCACUUACGGCACCACUGCCUUCGUGAUAAGUUCCGGGUCUCGUUUAGUGUAAACUGAGAAAUUGUAAAAAGCAGUUACAUUGCCCGAGGCUCGGGACAGCG-3'	50.35
Cluster 8	5'-GGGAGAUCACUUACGGCACCCGCAUAACUGCAUGAUGAGGCUGUCCACUCAGGUGCCUUGCGGUACCCAGUGAUCGACCGACACCAAGUCCCAAGGCUCGGGACAGCG-3'	50.52
Cluster 9	5'-GGGAGAUCACUUACGGCACCAGCGGUUGCGAGGGUCAUGGUGGGCGGGAGGCACCCGAUUGUGUCACUUAGUAAUCUCUAGUCGCGGCUGCCGAGGCUCGGGACAGCG-3'	47.18

Cluster 10	5'-GGGAGAUCACUUACGGCACCCUAGUUACGCCACACCGAGUCAGAUUUCGUUCAGUUACCUGCUUCGGCAGCAAUAUGAACGAACGGUACGCUCCAAGGCUCGGGACAGCG-3'	44.33
NanoLuciferase selection		
Round 7		
Cluster 1	5'-GGGAGAUCACUUACGGCACCAAUUGCAAAAACGGGCCGUGUUGAUGACUUCGACCUGUCUAGCACGGUCACCGGGGCCAAGCGCCAGGGGCCGAGGCUCGGGACAGCG-3'	52.64
Cluster 2	5'-GGGAGAUCACUUACGGCACCAAUUGAGAAAACGGGUCUUAGGAGUUCCCACGCGAAUAUGAGCUUUGAGGCGGCAGACAAGAUCUGUGCCCCGAGGCUCGGGACAGCG-3'	49.93
Cluster 3	5'-GGGAGAUCACUUACGGCACCGCGUGUGCGUAGAAUACCCUUCACGCUAGCUAACGAGAAAUAGGCGGAUACAAUUGAAAAAACGGUCCGACCGAGGCUCGGGACAGCG-3'	35.64
Cluster 4	5'-GGGAGAUCACUUACGGCACCAACGACUUUAUAGGACGCAUAUGAAGGUGUAAAGAACGUUGUAUAAGUCGCCUGGUUGCACAAGUGUUUUCCAAGGCUCGGGACAGCG-3'	42.18
Cluster 5	5'-GGGAGAUCACUUACGGCACCAAUUGAAAAAACGGGCCGAAACGAGUCCAAUCACCCGUCCGCGGAGCCAACCGAGAUGAAGAAUCGAGGACCGAGGCUCGGGACAGCG-3'	42.46
Cluster 6	5'-GGGAGAUCACUUACGGCACCCUCAAGAAAAAACCAUCCCAAAGGGUUCUUGAGACCAUGUGUGAUCCCACGCUCCAAGCCUUGUGUCACACCGAGGCUCGGGACAGCG-3'	48.45
Cluster 7	5'-GGGAGAUCACUUACGGCACCAAUUGAAAAAACGGGCCAAAGUAAUGUGCAUAGCCCGUGUUGCCCGGUCUUAGGCUUCCAAACUAGACGCCCGAGGCUCGGGACAGCG-3'	51.17
Cluster 8	5'-GGGAGAUCACUUACGGCACCGAUCAAUUGCAAAAACGACGGUCUCGACGUUUGGACUCUCUGAAGCUUACGACGAACAGCGGGCUAGCGCCCGAGCUCGGGACAGCG-3'	36.39
Cluster 9	5'-GGGAGAUCACUUACGGCACCCUGUGCGUUCCCUUAGGCUCCAGUCUUGAGCAUCUAAUGUAAAAGGAGACAAAAUGAACUCAGCUCGAGACCCGAGGCUCGGGACAGCG-3'	41.54
Cluster 10	5'-GGGAGAUCACUUACGGCACCUUGAAAAUCCAGAGAUGAUAAUGAGAAUUCAUCCUAAGGGUCGAGUUGAAACGGAGGUCCCCGACCUGACCCGAGGCUCGGGACAGCG-3'	40.02
Round 9		
Cluster 1	5'-GGGAGAUCACUUACGGCACCGAAACUGCGUAAGACGCUCCUGGCGUAACGAGUAUCUGGGGUUGCUGCGCAUUCCAAAAAAGGGCGGGCCCCGAGGCUCGGGACAGCG-3'	51.75

Cluster 2	5'-GGGAGAUCACUUACGGCACCAGAAACACGGCCCUAAUACGGCACGGAAAAGCCACUGAUGUCUAAAGUUGCCCCCCAGGCUUAAUCCGACCUCACCGAGGCUCGGGACAGCG-3'	39.9
Cluster 3	5'-GGGAGAUCACUUACGGCACCAAUUGCAAAAACGGGCCGUGUUGAUGACUUCGACCUGUCUAGCACGGUCACCGGGGCCAAGCGCCAGGGGCCGAGGCUCGGGACAGCG-3'	52.64
Cluster 4	5'-GGGAGAUCACUUACGGCACCUUGCACUGAUUUUGCAUUCGAAAUGAAAGGAAAGUGUGGGCAGUAGCUGGUCUGACCGUCCGAGAGUCUUGUACCCGAGGCUCGGGACAGCG-3'	51.78
Cluster 5	5'-GGGAGAUCACUUACGGCACCAAUUGAGAAAACGGGUCUUAGGAGUUCCCACGCGAAUAUGAGCUUUGAGGCGGCAGACAAGAUCUGUGCCCCGAGGCUCGGGACAGCG-3'	49.93
Cluster 6	5'-GGGAGAUCACUUACGGCACCUUGAAAAUCCAGAGAUGAUAAUGAGAAUUCAUCCUAAGGGUCGAGUUGAAACGGAGGUCCCCGACCUGACCCGAGGCUCGGGACAGCG-3'	40.02
Cluster 7	5'-GGGAGAUCACUUACGGCACCAAUUAAAAGACGGCAGACGCAAGAAACUGAUCGAACCAGGUUGUCGCCAACGCCGGAGUAAUUGCGAAAACACCGAGGCUCGGGACAGCG-3'	32.65
Cluster 8	5'-GGGAGAUCACUUACGGCACCGAUCAAUUGCAAAAACGACGGUCUCGACGUUUGGACUCUCUGAAGCUUACGACGAACAGCGGGCUAGCGCCCGAGGCUCGGGACAGCG-3'	40.23
Cluster 9	5'-GGGAGACACUUACGGCACCGCGUGUGCGUAGAAUACCCUUCACGCUAGCUAACGAGAAAUAGGCGGAUACAAUUGAAAAAACGGUCCGACCGAGGCUCGGGACAGCG-3'	35.43
Cluster 10	5'-GGGAGAUCACUUACGGCACCAAGGGUUACCGCUGUUUGUACGGCGAGUAAAAGACCAAGCCAAGCGUGAAACUACGAAAACGGUUCAGGGCCCCCGAGGCUCGGGACAGCG-3'	39.01
ATP-reacting selection (low complexity)		
Cluster 1	5'-GGGAGAUCACUUACGGCACCGGUUUCGAAGCACCGGUAACUGCAGAGAAGCUCCAACCUGAUCCACAGAUGCCUCCAACGCCUUGAGCGAACCGAGGCUCGGGACAGCG-3'	45.12
Cluster 2	5'-GGGAGAUCACUUACGGCACCUAAUACAGUCAUUAUCGGGACCGUACAUUUCUCCAAGCUUGCGUUCUGCGCCUCCAACGCUCUGACUGCGAACCGAGGCUCGGGACAGCG-3'	48.41
Cluster 3	5'-GGGAGAUCACUUACGGCACCACGACAACGCAGACGGCUCAAUCGUGGGACGACUUCCCCAGUCCAGCCCUCGCAACGCCAGAGCGCGAAACCGAGGCUCGGGACAGCG-3'	42.01

Cluster 4	5'-GGGAGAUCACUUACGGCACCAAACUGUCGUAAGCUCCCGUUGUUAUCUGAGUCGCAAUCACAUCGGUGUACAACGCCCGCGAGCGACGACACCGAGGCUCGGGACAGCG-3'	57.31
Cluster 5	5'-GGGAGAUCACUUACGGCACCCGUAAAUGGGCAACGCAAUAACAUUGCCGAACGAAAGUGACAAAAAAAUACAAGAAUAAAGAAAACACUGCCCUGAGGCUCGGGACAGCG-3'	28.24
Cluster 6	5'-GGGAGAUCACUUACGGCACCACGCCUGAGUGCUCAAAAAUGAAAAAAAGAAGAAAAAGACUGUUACUUUGAAGCCUAGAGUACCGUAUGGACCGAGGCUCGGGACAGCG-3'	35.46
Cluster 7	5'-GGGAGAUCACUUACGGCACCAAGCGAUGCAGAUGCAUUGCCUACUGACCCGUCAACGGCACCGGUGGUCUGCGUAAACGUUCCCGACCACCGAGGCUCGGGACAGCG-3'	49.84
Cluster 8	5'-GGGAGAUCACUUACGGCACCGCCAACUACGAGAAAAAUCGCGGGCCCAAAUAUGAUGUGAUCAACGUUGUCAGAAGAACAGCCUCUGAAAUCACGCCGAGGCUCGGGACAGCG-3'	42.28
Cluster 9	5'-GGGAGAUCACUUACGGCACCGCAGAAGAGUGCUAAAAUACUGUGAUAAUUAUUACUGUCCGUGCCAAUGUCUUAUAAAUGCAUCACAUGACCGAGGCUCGGGACAGCG-3'	34.17
Cluster 10	5'-GGGAGAUCACUUACGGCACCGCAAAUCUUCUGACGCUAAAUGAAGAUGCGUCCGUAAAGCGAACUAACUGUCCCGCAUAGGCUGAUGUAUAAACCGAGGCUCGGGACAGCG-3'	45.57
ATP-reacting selection (high complexity)		
Cluster 1	5'-GGGAGAUCACUUACGGCACCGGUUUCGAAGCACCGGUAACUGCAGAGAAGCUCCAACCUGAUCCACAGAUGCCUCCAACGCCUUGAGCGAACCGAGGCUCGGGACAGCG-3'	45.12
Cluster 2	5'-GGGAGAUCACUUACGGCACCACGACAACGCAGACGGCUCAAUCGUGGGACGACUUCCCCAGUCCAGCCCUCGCAACGCCAGAGCGCGAAACCGAGGCUCGGGACAGCG-3'	42.01
Cluster 3	5'-GGGAGAUCACUUACGGCACCUCUGAGGCCGAAUCUCUCGCACCCGUGGCCAGCGGAUUUGGUAAUACCAUGACCGCAUCGUGCGCAUGAUGCCAAGGCUCGGGACAGCGCG-3'	47.33
Cluster 4	5'-GGGAGAUCACUUACGGCACCUAACGGCCCAACGCCCUCGCGCGAAACGGCCCUUAGUGCGCAGUUAAACUUCACAUUCUGGGCCUCCAUACACCGAGGCUCGGGACAGCG-3'	46.37
Cluster 5	5'-GGGAGAUCACUUACGGCACCAACUAAAUAUCGUGUAACGGCCCUUCUUGCGACCAAUUCAGGCGACAUUAGGCCAAGGCGCAGAAGGAUACCACGGCUCGGGACAGCGCG-3'	32.85

Cluster 6	5'-GGGAGAUCACUUACGGCACCUUGAACCCUGACUGACACAGGAUCUCGGAAGUUUCGAAGUGAACAAGAUUUGAUAACAUAUACUGUUCCACCAAGGCUCGGGACAGCGCG-3'	29.01
Cluster 7	5'-GGGAGAUCACUUACGGCACCAAAAACAAGGUGAUUAUUGCGUGAAGUGACGAAUUAAAACUGAAACCCGGCACUAUGCGAGAUAUAAACCGUCCAAGGCUCGGGACAGCGCG-3'	30.9
Cluster 8	5'-GGGAGAUCACUUACGGCACCGAGAUUGGGUAUGGCAGUGUAUGAUGACUUGGUAUAAUCGAGCAAGUUUCUGCUAACUGCCUUUGACUGCCCCACGGCUCGGGACAGCGCG-3'	38.72
Cluster 9	5'-GGGAGAUCACUUACGGCACCACACGAACUGAUCCCUAGUUGUUUAGAGGAUCGAGUGUGCAUCAGCCGACAUCCAGUAUGUCCAUGAGUAACACCGAGGCUCGGGACAGCG-3'	49.87
Cluster 10	5'-GGGAGAUCACUUACGGCACCGUUGUGCUAGAUUCGUCACAAAGUCAUAUGUGUUACUCUAACUGUUCCAUGAGCAUAAGUCGAGAUGUAGACCGAGGCUCGGGACAGCG-3'	39.75

Top 10clusters from each selection for Figure 2.3a	**Predicted G-quartet**			
Firefly luciferase selection	**Position**	**Length**	**Region**	**G-Score**
Cluster 1	2	26	GGAGAUCACUUACGGCACCACCGGGG	10
Cluster 2	none			
Cluster 3	86	17	GGCGGCCAAGGCUCGGG	18
Cluster 4	none			
Cluster 5	74	30	GGGGAGUCUAUGUCGCAAGCCAAGGCUCGG	2
Cluster 6	2	30	GGAGAUCACUUACGGCACCUACCUAUGGGG	10
Cluster 7	2	25	GGAGAUCACUUACGGCACCGGCAGG	12
Cluster 8	none			
Cluster 9	none			
Cluster 10	none			
Oregon Green selection				
Cluster 1	none			
Cluster 2	none			
Cluster 3	45	11	GGUGGGGUAGG	19
Cluster 4	30	17	GGCUUUAGGUGGCCCGG	17
Cluster 5	15	18	GGCACCCUGUGGGGGUGG	14

Cluster 6	none			
Cluster 7	none			
Cluster 8	none			
Cluster 9	15	29	GGCACCAGCGGUUGCGAGGGUCAUGGUGG	21
Cluster 10	none			
NanoLuciferase selection				
Round 7				
Cluster 1	none			
Cluster 2	none			
Cluster 3	none			
Cluster 4	none			
Cluster 5	none			
Cluster 6	73	29	GGAGGUCCCCGACCUGACCCGAGGCUCGG	5
Cluster 7	none			
Cluster 8	none			
Cluster 9	none			
Cluster 10	73	29	GGAGGUCCCCGACCUGACCCGAGGCUCGG	5
ATP-reacting selection (low complexity)				
Cluster 1	96	17	GGCUCGGGACAGCGGGG	16
Cluster 2	none			
Cluster 3	none			
Cluster 4	none			
Cluster 5	none			
Cluster 6	none			
Cluster 7	none			
Cluster 8	none			
Cluster 9	none			
Cluster 10	none			

2.9 References

1. Hermann, T. & Patel, D. J. Stitching together RNA tertiary architectures. *J Mol Biol* **294**, 829–849 (1999).
2. Edwards, T. E., Klein, D. J. & Ferre-D'Amare, A. R. Riboswitches: small-molecule recognition by gene regulatory RNAs. *Curr Opin Struct Biol* **17**, 273–279 (2007).
3. Serganov, A. & Patel, D. J. Ribozymes, riboswitches and beyond: regulation of gene expression without proteins. *Nat Rev Genet* **8**, 776–790 (2007).
4. Jimenez, R. M., Polanco, J. A. & Lupták, A. Chemistry and Biology of Self-Cleaving Ribozymes. *Trends Biochem. Sci.* **40**, (2015).
5. Russell, R. *et al.* Rapid Compaction During RNA Folding. *Proc Natl Acad Sci U S A* **99**, 4266–4271 (2002).
6. Woodson, S. A. Compact Intermediates in RNA Folding. *Annu. Rev. Biophys.* **39**, 61–77 (2010).
7. Wilson, D. S. & Szostak, J. W. In vitro selection of functional nucleic acids. *Annu Rev Biochem* **68**, 611–647 (1999).
8. Ellington, A. D. & Szostak, J. W. In vitro selection of RNA molecules that bind specific ligands. *Nature* **346**, 818–822 (1990).
9. Tuerk, C. & Gold, L. Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science* **249**, 505 LP – 510 (1990).
10. Robertson, D. L. & Joyce, G. F. Selection in vitro of an RNA enzyme that specifically cleaves single-stranded DNA. *Nature* **344**, 467–468 (1990).
11. Pobanz, K. & Lupták, A. Improving the odds: Influence of starting pools on in vitro selection outcomes. *Methods* **106**, 14–20 (2016).
12. Sabeti, P. C., Unrau, P. J. & Bartel, D. P. Accessing rare activities from random RNA sequences: The importance of the length of molecules in the starting pool. *Chem. Biol.* **4**, 767–774 (1997).
13. Legiewicz, M., Lozupone, C., Knight, R. & Yarus, M. Size, constant sequences, and optimal selection. *Rna* **11**, 1701–1709 (2005).
14. Davis, J. H. & Szostak, J. W. Isolation of high-affinity GTP aptamers from partially structured RNA libraries. *Proc. Natl. Acad. Sci.* **99**, 11616–11621 (2002).
15. Porter, E. B., Polaski, J. T., Morck, M. M. & Batey, R. T. Recurrent RNA motifs as scaffolds for genetically encodable small-molecule biosensors. *Nat. Chem. Biol.* **13**, 295–301 (2017).
16. Ruff, K. M., Snyder, T. M. & Liu, D. R. Enhanced functional potential of nucleic acid aptamer libraries patterned to increase secondary structure. *J Am Chem Soc* **132**, 9453–9464 (2010).

17. Luo, X. *et al.* Computational approaches toward the design of pools for the in vitro selection of complex aptamers. *RNA* **16**, 2252–2262 (2010).
18. Sassanfar, M. & Szostak, J. W. An RNA motif that binds ATP. *Nature* **364**, 550–553 (1993).
19. Carothers, J. M., Davis, J. H., Chou, J. J. & Szostak, J. W. Solution structure of an informationally complex high-affinity RNA aptamer to GTP. *RNA* **12**, 567–579 (2006).
20. Grate, D. & Wilson, C. Laser-mediated, site-specific inactivation of RNA transcripts. *Proc. Natl. Acad. Sci.* **96**, 6131–6 (1999).
21. Ekland, E. H., Szostak, J. W. & Bartel, D. P. Structurally complex and highly active RNA ligases derived from random RNA sequences. *Science* **269**, 364–370 (1995).
22. Illangasekare, M., Sanchez, G., Nickles, T. & Yarus, M. Aminoacyl-RNA synthesis catalyzed by an RNA. *Science* **267**, 643–647 (1995).
23. Curtis, E. A. & Bartel, D. P. New catalytic structures from an existing ribozyme. *Nat. Struct. Mol. Biol.* **12**, 994–1000 (2005).
24. Rogers, J. & Joyce, G. F. The effect of cytidine on the structure and function of an RNA ligase ribozyme. *RNA* **7**, 395–404 (2001).
25. Robertson, M. P. & Ellington, A. D. In vitro selection of an allosteric ribozyme that transduces analytes to amplicons. *Nat Biotechnol* **17**, 62–66 (1999).
26. Robertson, M. P. & Scott, W. G. The structural basis of ribozyme-catalyzed RNA assembly. *Science* **315**, 1549–1553 (2007).
27. Pitt, J. N. & Ferre-D'Amare, A. R. Structure-guided engineering of the regioselectivity of RNA ligase ribozymes. *J Am Chem Soc* **131**, 3532–3540 (2009).
28. Cadwell, R. C. & Joyce, G. F. Randomization of genes by PCR mutagenesis. *Genome Res.* **2**, 28–33 (1992).
29. Wilson, D. S. & Keefe, A. D. Random Mutagenesis by PCR. *Curr. Protoc. Mol. Biol.* 8.3.1-8.3.9 (2001).
30. Duhamel, J. *et al.* Secondary structure content of the HDV ribozyme in 95% formamide. *Nucleic Acids Res.* **24**, 3911–3917 (1996).
31. Vogt, V. M. Purification and Further Properties of Single-Strand-Specific Nuclease from Aspergillus oryzae. *Eur. J. Biochem.* **33**, 192–200 (1973).
32. Lepecq, J. B. & Paoletti, C. A fluorescent complex between ethidium bromide and nucleic acids. Physical-Chemical characterization. *J. Mol. Biol.* **27**, 87–106 (1967).
33. Merino, E. J., Wilkinson, K. A., Coughlan, J. L. & Weeks, K. M. RNA structure analysis at single nucleotide resolution by selective 2'-hydroxyl acylation and primer extension (SHAPE). *J Am Chem Soc* **127**, 4223–4231 (2005).
34. Hofmann, H. P., Limmer, S., Hornung, V. & Sprinzl, M. Ni2+-binding RNA motifs with an asymmetric purine-rich internal loop and a G-A base pair. *RNA* **3**, 1289–300 (1997).

35. Kikin, O., D’Antonio, L. & Bagga, P. S. QGRS Mapper: a web-based server for predicting G-quadruplexes in nucleotide sequences. *Nucleic Acids Res.* **34**, W676–W682 (2006).

36. Abdelsayed, M. M. *et al.* Multiplex Aptamer Discovery through Apta-Seq and Its Application to ATP Aptamers Derived from Human-Genomic SELEX. *ACS Chem. Biol.* **12**, 2149–2156 (2017).

37. Zaher, H. S., Watkins, R. A. & Unrau, P. J. Two independently selected capping ribozymes share similar substrate requirements. *RNA* **12**, 1949–1958 (2006).

38. Huang, F. & Yarus, M. 5‘-RNA Self-Capping from Guanosine Diphosphate. *Biochemistry* **36**, 6557–6563 (2002).

39. Ferré-D’Amaré, A. R. & Doudna, J. A. RNA folds: insights from recent crystal structures. *Annu. Rev. Biophys. Biomol. Struct.* **28**, 57–73 (1999).

40. Alemán, E. A., Lamichhane, R. & Rueda, D. Exploring RNA folding one molecule at a time. *Curr Opin Chem Biol* **12**, 6:647-54 (2008).

41. Russell, R. *Biophysics of RNA folding*. *Biophysics of RNA Folding* (2013). doi:10.1007/978-1-4614-4954-6

42. Carothers, J. M., Oestreich, S. C., Davis, J. H. & Szostak, J. W. Informational complexity and functional activity of RNA structures. *J Am Chem Soc* **126**, 5130–5137 (2004).

43. Pressman, A. D. *et al.* Mapping a Systematic Ribozyme Fitness Landscape Reveals a Frustrated Evolutionary Network for Self-Aminoacylating RNA. *J. Am. Chem. Soc.* **141**, 6213–6223 (2019).

44. Vorobyeva, M. A., Davydova, A. S., Vorobjev, P. E., Pyshnyi, D. V. & Venyaminova, A. G. Key aspects of nucleic acid library design for in vitro selection. *Int. J. Mol. Sci.* **19**, 470 (2018).

45. Schultes, E. A., Spasic, A., Mohanty, U. & Bartel, D. P. Compact and ordered collapse of randomly generated RNA sequences. *Nat. Struct. Mol. Biol.* **12**, 1130–1136 (2005).

46. Lorsch, J. R. & Szostak, J. W. In vitro evolution of new ribozymes with polynucleotide kinase activity. *Nature* **371**, 31–36 (1994).

47. Braselmann, E. *et al.* A multicolor riboswitch-based platform for imaging of RNA in live mammalian cells. *Nat. Chem. Biol.* **14**, 964–971 (2018).

48. Martini, L. *et al.* In Vitro Selection for Small-Molecule-Triggered Strand Displacement and Riboswitch Activity. *ACS Synth. Biol.* **4**, 1144–1150 (2015).

49. Lau, M. W. & Ferre-D’Amare, A. R. In vitro evolution of coenzyme-independent variants from the glmS ribozyme structural scaffold. *Methods* **106**, 86–81 (2016).

50. Lau, A. K. S., Wong, T. T. W., Shum, H. C., Wong, K. K. Y. & Tsia, K. K. Ultrafast Microfluidic Cellular Imaging by Optical Time-Stretch. *Methods Mol Biol.* **1389,** 23-45 (2016).

51. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing

reads. *EMBnet.journal* **17**, 10–12 (2011).

52. ViennaRNA Package 2.0. *Algorithms Mol. Biol.* **6**, 26 (2011).
53. Spitale, R. C. *et al.* Structural imprints in vivo decode RNA regulatory mechanisms. *Nature* **519**, 486–490 (2015).
54. Eliceiri, K. W., Schneider, C. A. & Rasband, W. S. NIH Image to ImageJ : 25 years of image analysis. *Nat. Methods* **9**, 671–675 (2012).
55. Alam, K. K., Chang, J. L. & Burke, D. H. FASTAptamer: A Bioinformatic Toolkit for High-throughput Sequence Analysis of Combinatorial Selections. *Mol Ther Nucleic Acids* **4**, e230 (2015).
56. Ritchey, L. E. *et al.* Structure-seq2 : sensitive and accurate genome-wide profiling of RNA structure in vivo. **45**, 1–9 (2017).

Chapter 3: Allosteric Modulation of Bacterial Ribozyme

Publication note:

This paper was originally published in the journal *Biochemistry*.

Passalacqua L.F.M.*, Jimenez R.M.*, Fong J.Y., Lupták A. Allosteric Modulation of the *Faecalibacterium prausnitzii* Hepatitis Delta Virus-like Ribozyme by Glucosamine 6-Phosphate: The Substrate of the Adjacent Gene Product. *Biochemistry* **56,** 6006–6014 (2017).

* Authors contributed equally to this work.

Copyright ⓒ 2017 American Chemical Society

Contribution Statement

L F. M. P., R. M. J. and A. L. designed the study; **L F. M. P.**, R. M. J. and J. Y. F. performed the experiments; all authors analyzed the data; and **L F. M. P.**, R. M. J. and A. L. wrote the manuscript. **L F. M. P.** and R. M. J. authors contributed equally to this work.

3.1 Abstract

Self–cleaving ribozymes were discovered 30 years ago and have been found throughout nature, from bacteria to animals, but little is known about their biological functions and regulation, particularly how cofactors and metabolites alter their activity. An HDV–like self–cleaving ribozyme maps upstream of a phosphoglucosamine mutase *(glmM)* open reading frame (ORF) in the genome of the human gut bacterium *Faecalibacterium prausnitzii*. The presence of a ribozyme in the untranslated region of *glmM* suggests a regulation mechanism of gene expression. In the bacterial hexosamine biosynthesis pathway, the enzyme glmM catalyzes the isomerization of glucosamine 6–phosphate into glucosamine 1–phosphate. In this study, we investigated the effect of these metabolites on the co–transcriptional self–cleavage rate of the ribozyme. Our results suggest that glucosamine 6–phosphate, but not glucosamine 1–phosphate, is an allosteric ligand that increases the self–cleavage rate of drz–Fpra–1, providing the first known example of an allosteric modulation of a self–cleaving ribozyme by the substrate of the adjacent gene product. Given that the ribozyme is activated by the glmM substrate, but not the product, this allosteric modulation may represent a potential feed–forward mechanism of gene expression regulation in bacteria.

3.2 Introduction

Ribozymes are RNA molecules that can catalyze a chemical transformation in the absence of a protein cofactor[1-3]. Self–cleaving ribozymes comprise a group of RNA molecules that promote a site–specific self–scission reaction. In all known self–cleaving ribozymes, the cleavage reaction is a transesterification that involves a nucleophilic attack by a 2′ oxygen on the adjacent phosphodiester bond, producing a 2′–3′ cyclic phosphate and a 5′–hydroxyl products[4-14]. To date, nine self–cleaving ribozyme families have been discovered, comprising the hairpin[10], hammerhead[8, 9], hepatitis delta virus (HDV)[6, 7], glucosamine–6–phosphate synthase (*glmS*)[11], *Neurospora* Varkud satellite (VS)[12], twister[13], twister sister (TS), pistol, and hatchet motifs[14]. First characterized 30 years ago[6, 7, 15], the HDV self–cleaving ribozyme has been extensively studied, with elucidated crystal structures and mechanism of self–scission[16-21]. HDV–like self–cleaving ribozymes exhibit great sequence diversity, but fold into a conserved secondary structure that includes a nested double–pseudoknot[22]. HDV–like ribozyme have been found in many eukaryotes, including humans, as well as in *Chilo* iridescent virus, several bacteria, and most recently in several microbial metagenomic datasets[14, 22-27]. These self–cleaving RNAs likely play a number of distinct roles in biology[22-27], but little is known about their regulation, particularly with regard to the role ligands or metabolites may have in modulating HDV–like ribozyme self–cleavage as either allosteric effectors or cofactors.

One particular case of interest is the drz–Fpra–1 HDV–like ribozyme, found in the human gut bacterium *Faecalibacterium prausnitzii*[22]*. F. prausnitzii* is a Gram positive Firmicute that represents more than 5 % of the total bacterial population in the fecal microbiota of a healthy human and is suggested to be negatively correlated with certain pathologies, such as Crohn’s disease and ulcerative colitis[28-30]. The ribozyme cleavage site maps 106 nucleotides upstream of phosphoglucosamine mutase (*glmM*) open reading frame (ORF)[22] (**Figure 3.1A**).

The enzyme glmM is a component of the hexosamine biosynthesis pathway, catalyzing the transformation of glucosamine 6–phosphate (GlcN6P) into glucosamine 1–phosphate (GlcN1P)[31] (**Figure 3.1B**). The final product of the hexosamine biosynthesis pathway is uridine diphosphate N–acetyl–glucosamine (UDP–GlcNAc), a key substrate used for cell wall biosynthesis[32]. The secondary structure of the ribozyme is shown in **Figure 3.1C**.

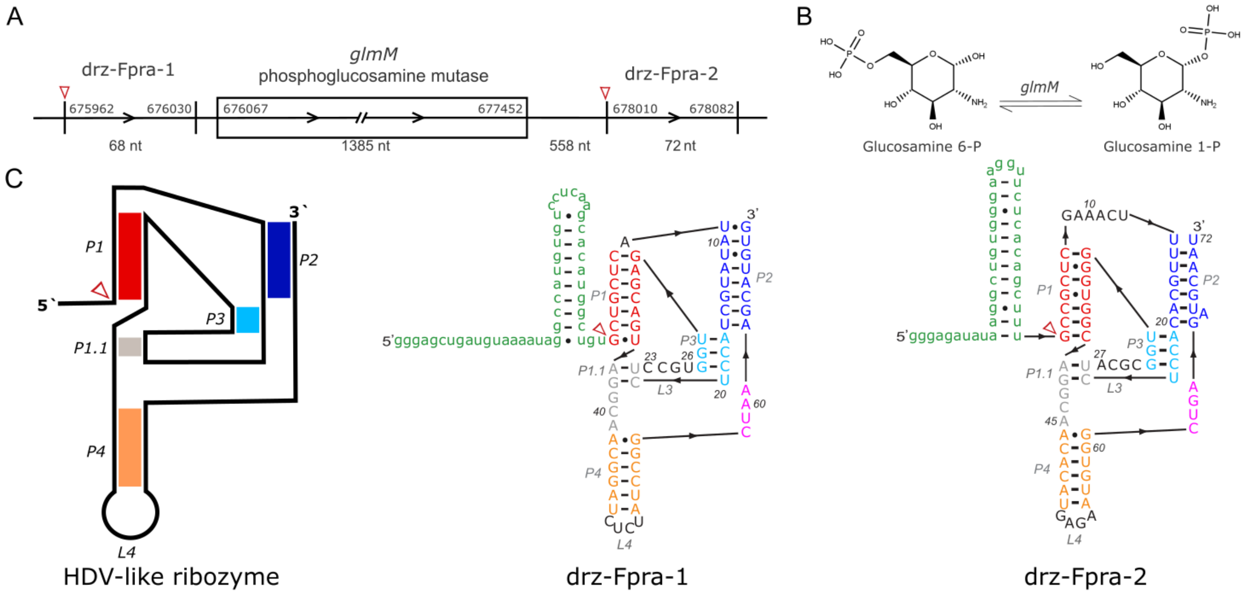

A
drz-Fpra-1
glmM
phosphoglucosamine mutase
drz-Fpra-2
675962
676030
676067
677452
678010
678082
68 nt
1385 nt
558 nt
72 nt
B
glmM
Glucosamine 6-P
Glucosamine 1-P
C
P1
P2
P3
P1.1
P4
L4
L3
5`
3`
HDV-like ribozyme
5'gggagcugauguaaaauag
5'gggagauaua
drz-Fpra-1
drz-Fpra-2

Figure 3.1. HDV–like ribozymes *in F. prausnitzii*. (A) Genome locus of *F. prausnitzii* showing the glmM ORF, the upstream drz–Fpra–1 ribozyme and the newly discovered drz–Fpra–2 ribozyme. (B) The isomeration reaction promoted by the glmM enzyme. (C) A schematic representation of the secondary structure of HDV–like ribozymes (left), and predicted secondary structures of drz–Fpra–1 and drz–Fpra–2 ribozymes, including hairpins in the 5′ leader sequence. Red arrowheads mark the cleavage sites.

In bacteria, a common mode of regulation of gene expression involves sensing a metabolite related to the adjacent gene product through an RNA regulatory element called a riboswitch[33]. In the case of the *glmS,* a GlcN6P–sensing riboswitch[34], the RNA is also a self–cleaving ribozyme that resides at the 5′ untranslated region (UTR) of the transcript encoding glutamine fructose–6–phosphate amidotransferase, which catalyzes the transformation of fructose–6–phosphate into GlcN6P[11]. The metabolite GlcN6P acts as a co–factor of the *glmS* ribozyme, accelerating its self–scission nearly one–million–fold[11, 35-39], and promoting the degradation of the adjacent ORF through an RNase J–dependent mechanism[40]. Thus, this feedback system senses the amount of GlcN6P in the cell and represses gene expression through the ribozyme–dependent activity. This metabolite–responsive regulation system has been well–characterized, by structural, biochemical and mechanistic studies[11, 35-39, 41-44]. Other metabolites, such as glucose 6–phosphate (Glc6P), may compete with GlcN6P and thus upregulate the gene expression by inhibiting the self–cleavage of the RNA[39, 41]. Surprisingly, only 3 mutations in the active site are necessary to convert the ribozyme into a coenzyme–independent self–cleaving ribozyme, revealed by an *in vitro* selection of a GlcN6P–insensitive *glmS* ribozyme that used a divalent cation for catalysis without changing the overall fold of the ribozyme[42].

Due to the proximity of the HDV–like ribozyme and the *glmM* gene, we hypothesized that the ribozyme drz–Fpra–1 and the metabolites involved in the hexosamine biosynthesis pathway may contribute to gene expression regulation through modulation of the ribozyme activity. During the course of this work, we found another HDV–like ribozyme, named drz–Fpra–2 (**Figure 1C**), downstream of the *glmM* gene and we chose to study the two ribozymes in parallel. To explore the mechanism of this putative regulation, we studied the *in vitro* ribozyme self–cleavage kinetics in the presence of several metabolites. The kinetic and structural probing of the drz–Fpra–1 in the presence of the metabolites shows an allosteric modulation of the ribozyme by the substrate, but not the product, of the adjacent gene product. The effect on drz–Fpra–2 was largely the opposite.

3.3 Results

Discovery of a second HDV–like ribozyme in the same locus of *F. prausnitzii* genome. We decided to analyze the *F. prausnitzii* genome in order to verify the mapping of the drz–Fpra–1 ribozyme near the *glmM* gene. To our surprise, we found another HDV–like ribozyme, drz–Fpra–2, with high sequence similarity to drz–Fpra–1 in the same locus. This second ribozyme was found 558 nucleotides downstream of the *glmM* gene (**Figure 3.1A**). Although similar in structure and sequence, some differences near the active site were observed in the proposed secondary structures (**Figure 3.1C**), as discussed below. One unique feature found in both drz–Fpra ribozymes is an A–U base–pair in the top of the P1.1 region, where a G–C base–pair is usually found in other HDV–like ribozymes[49].

Self–scission kinetics of drz–Fpra ribozymes in the presence of metabolites. To study the effect of the metabolite on the cleavage rate of the ribozymes, we started with an *in vitro* co–transcriptional cleavage kinetics performed in the presence or absence of 20 mM GlcN6P at 5 mM Mg^{2+}. Interestingly, the results showed that the effect of the metabolite on each ribozyme was different. GlcN6P accelerates the self–scission of the drz–Fpra–1 ribozyme, whereas, it decreases the self–cleavage rate of the drz–Fpra–2 ribozyme (**Figure 3.2**). The low amplitude of the effect of the metabolite on the two ribozymes indicated that GlcN6P is an allosteric effector (modulator) and not a co–factor or co–enzyme that directly participates in catalysis. To further investigate the ligand effect on the activity of drz–Fpra–1, we performed a titration of GlcN6P, keeping the ionic strength of the solution constant and [Mg^{2+}] fixed at 5 mM. Self–cleavage kinetics under conditions approximating co–transcriptional self–scission showed a dose–response between the cleavage rate constant and the metabolite, increasing the k_{GlcN6P} more than 2–fold at higher concentrations of GlcN6P, when compared to no–metabolite $k_{control}$ (**Figure 3.3A**). At low concentrations of GlcN6P, the cleavage rate was similar to the no–metabolite control, indicating that the metabolite may not be affecting the catalytic mechanism of the ribozyme.

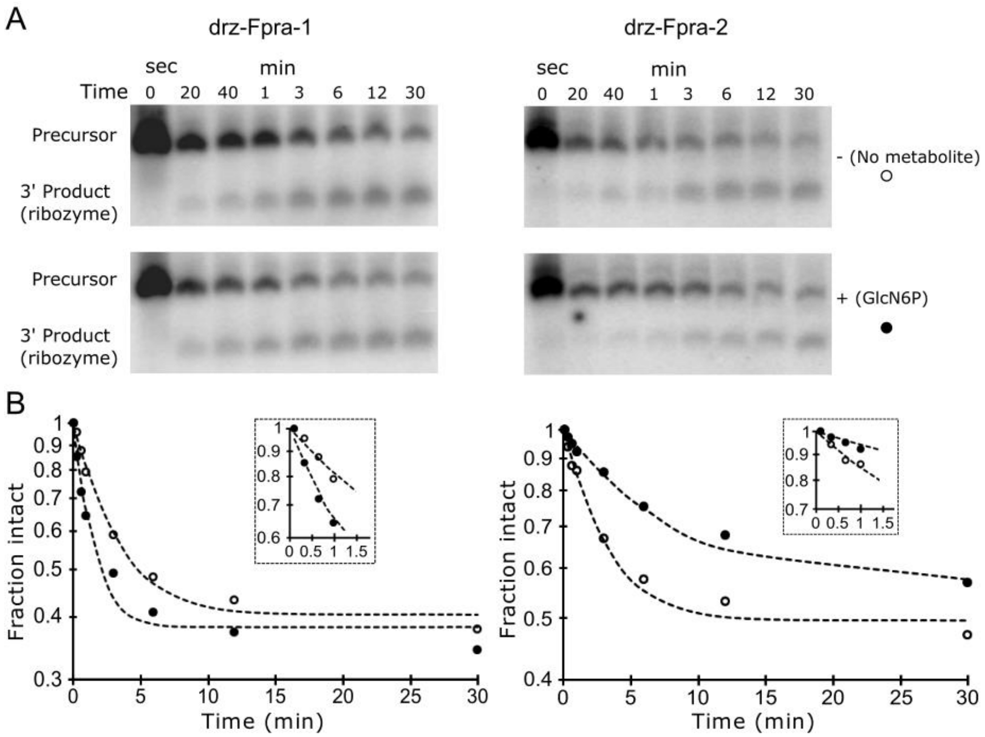


Figure 3.2. Influence of GlcN6P on the co–transcriptional self–scission of drz–Fpra–1 and drz–Fpra–2. To study the effect of the metabolite on the cleavage rate of the ribozymes, *in vitro* co–transcriptional cleavage kinetics were performed in presence and absence of GlcN6P. (A) PAGE analysis of the co–transcriptional self–scission of the two ribozymes in the absence (top) and presence (bottom) of 20 mM GlcN6P. (B) Log–linear graphs of the ribozyme self–scission. GlcN6P accelerates the self–cleavage rate of drz–Fpra–1 (left; no GlcN6P, open circles; 20 mM GlcN6P, solid circles) but inhibits drz–Fpra–2 (right). Early time points are shown in insets.

To investigate the origin of this dose–response, we varied the concentration of Mg^{2+}, the divalent metal ion used by the ribozyme for catalysis[17, 18, 50-52], in the presence of 20 mM GlcN6P. At low concentrations of Mg^{2+}, both k_{GlcN6P} and $k_{control}$ behave similarly, with a strong overlap between both data sets (**Figure 3.3B**). This trend changes at about physiological Mg^{2+}, where the cleavage rate increases in the presence of GlcN6P. At higher concentrations of the divalent metal ion, $k_{control}$ decreases, possibly due to misfolding. Surprisingly, this effect is not seen for k_{GlcN6P}, which continues to gradually increase. This result is consistent with a model, in which the metabolite interacts with the ribozyme, but does not act directly in the catalytic step of the reaction (**Figure 3.3B**). We also performed the same titration of Mg^{2+} in the presence of 20 mM of GlcN6P for drz–Fpra–2. As expected, an opposite effect was observed, where a decrease in the cleavage rate was found in the same conditions when compared to a no–metabolite control (**Figure 3.3C**).

To investigate the specificity of the ribozyme–metabolite interaction, we decided to test compounds related to GlcN6P. The ribozyme self–cleavage kinetics were tested in the presence of the following molecules: GlcN6P, the precursor of the reaction promoted by the glmM enzyme; GlcN1P, the isomer of GlcN6P and the product of the reaction; glucose (Glc); Glc6P and glucosamine (GlcN) – the last two were chosen due to their chemical similarity to the metabolites, because each carries a different chemical group present in both metabolites of the enzymatic reaction (the amino and the phosphate groups). The kinetic rate constants were normalized to no–metabolite cleavage kinetics at the same conditions, and showed that GlcN1P, the product of the

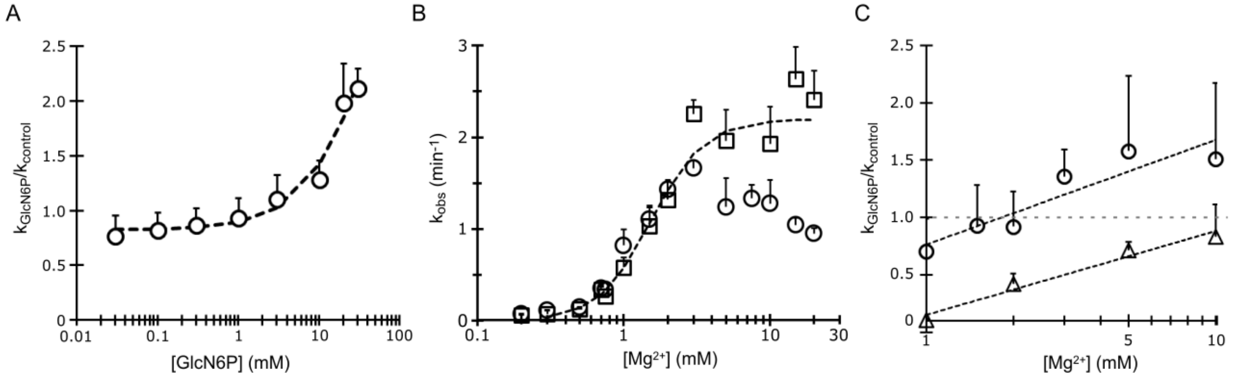


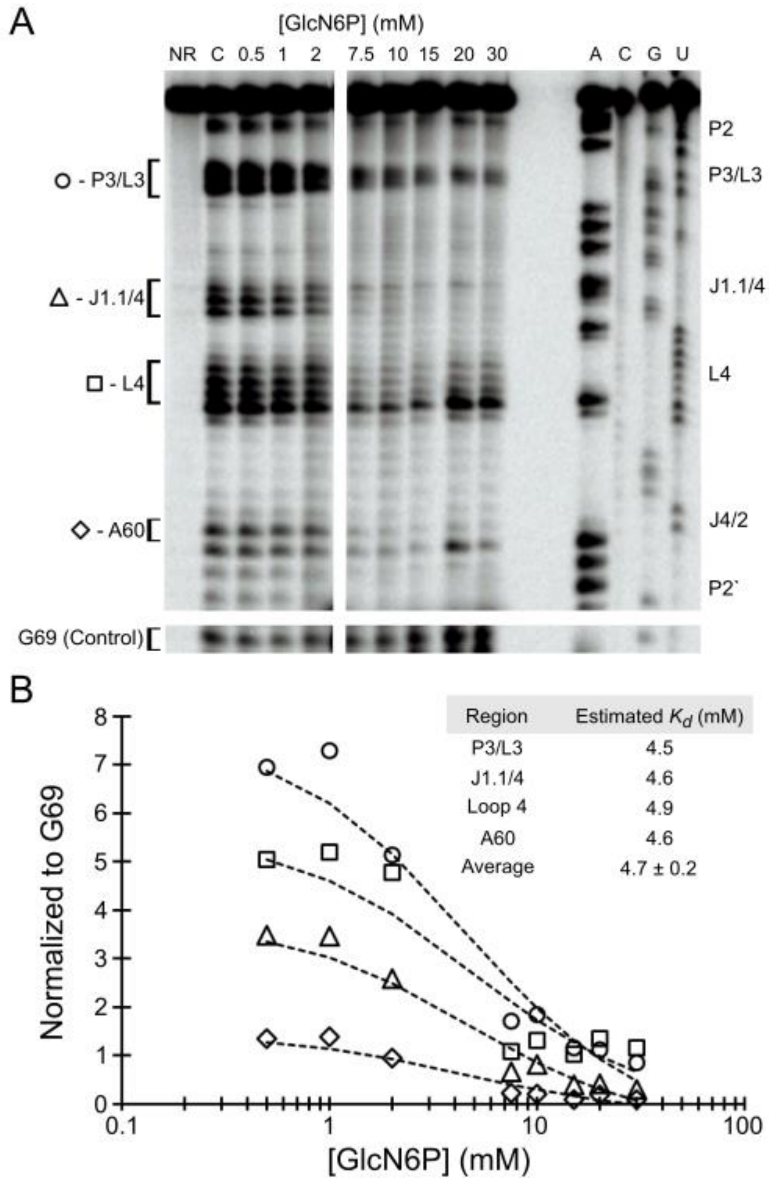
Figure 3.3. Effect of GlcN6P and Mg^{2+} on the drz–Fpra ribozymes. (A) Drz–Fpra–1 self–scission dose–response to GlcN6P at constant (5 mM) Mg^{2+}, normalized to the no–metabolite control. (B) Mg^{2+} dependence of drz–Fpra–1 self–scission in presence of 20 mM GlcN6P (open squares) or no metabolite (open circles). (C) Mg^{2+} dependence of drz–Fpra–1 (open circles) and drz–Fpra–2 (open triangules) self–scission in 20 mM GlcN6P normalized to a no–metabolite control.

reaction catalyzed by glmM, activated only the Fpra–2 ribozyme – albeit with low statistical significance, whereas Glc slightly increased the activity of both ribozymes (**Figure S3.1**). GlcN and Glc6P increased the cleavage rate of both drz–Fpra–1 and drz–Fpra–2 ribozymes to a similar extent (**Figure S3.1**). This is an unexpected result for drz–Fpra–2, because it was not inhibited by GlcN and Glc6P, like in the presence of GlcN6P. These results suggest that both the amino and the phosphate groups are important in modulating the self–scission rate of the ribozymes, and their relative positioning around the sugar ring leads to the differential effect on the two ribozymes. Finally, we decided to probe the ribozymes in the presence of UDP–GlcNAc, the final product of the hexosamine biosynthesis pathway, but no significant effect was observed (**Figure S3.2**).

Another way to compare the effects of the GlcN6P and GlcN1P is to calculate the ratio of the self–scission rate constants at the same concentrations of the metabolite, while varying the Mg^{2+} concentration. At 20 mM metabolite concentration, the k_{GlcN6P}/k_{GlcN1P} ratios for drz–Fpra–1 were 1.4 ± 0.1, 2.2 ± 0.4, and 2.5 ±0.5 at 1, 5, and 10 mM Mg^{2+}. Interestingly, when both GlcN1P and GlcN6P were included in the drz–Fpra–1 reaction, the effect of GlcN6P was enhanced (**Figure S3.3**).

The structure of the ribozyme drz–Fpra–1 is stabilized by GlcN6P. In order to investigate the effect of the metabolite on the ribozyme structure, we performed an in–line probing experiment in the presence of different concentrations of GlcN6P. This technique profiles the natural degradation of an RNA molecule at different conditions, providing information on the relative stability of individual phosphodiesters[46]. More flexible, solvent–exposed regions (such as single–strand regions) and specific conformations are more likely to promote a 2′–OH attack on the scissile phosphate, cleaving the molecule. The pattern of the cleavage at different conditions can indicate structural changes, suggesting differences in the flexibility and protection promoted by a ligand. Our results showed that GlcN6P promotes an apparent stabilization of the ribozyme at specific sites, suggesting that the metabolite directly interacts with the RNA (**Figure 3.4A**). The regions showing the most prominent change in degradation at higher GlcN6P concentration are P3/L3, J1.1/4, L4 and A60 (J4/2). Three of those regions (P3/L3, J1.1/4 and A60) surround the active site of the ribozyme, with A60 only 2 nucleotides away from the catalytic cytosine residue (C58). No region of increased degradation was observed in this experiment.

The data obtained for each concentration of metabolite, including the no–metabolite control, were normalized to G69 near the 3′ terminus of the ribozyme, used as a control position with ligand–independent degradation. The data were modeled by a dissociation constant equation (Equation 2) for a single ligand and a K_d was calculated independently for each affected region (**Figure 3.4B**). The K_d estimations for all 4 regions were consistent, resulting in an average K_d of 4.7 ± 0.2 mM. Performing the same experiment with the metabolite GlcN1P, the isomer of GlcN6P and product of the enzymatic reaction that does not affect the cleavage rate of the ribozyme, we did not observe any region with significant degradation pattern changes (**Figure S3.4A and B**). We measured the degradation for the regions P3/L3, L4 and A60 (J4/2), and used 3 different control bands for normalization. No apparent K_d was revealed for any of the regions, demonstrating that GlcN1P does not affect the degradation pattern of the ribozyme.



Region	Estimated K_d (mM)
P3/L3	4.5
J1.1/4	4.6
Loop 4	4.9
A60	4.6
Average	4.7 ± 0.2

Figure 3.4. Structural probing of GlcN6P binding of the drz–Fpra–1 ribozyme. (A) In–line probing of the 3′–labeled drz–Fpra–1 ribozyme in the presence of GlcN6P. The band intensities that change with GlcN6P concentration are indicated in the figure by the nucleotide identity. The sequence of the RNA was determined using iodoethanol cleavage of ribozymes with phosphorothioate–modified backbone at positions indicated above each lane. The intensities of the control band (G69) from the same experiment are shown

below the rest of the gel. (B) Graph of band intensities for the regions P3/L3 (open circles), J1.1/4 (open triangles), L4 (open squares), and A60 (open diamonds), normalized to a control band at G69. The data were fit to a model based on Equation 2. The average dissociation constant is estimated in 4.7 ± 0.2 mM. The positions where the band intensities respond to metabolite concentration are also indicated left of the gel in (A).

Metabolite effect is ribozyme– and position–specific. To verify that the effects promoted by GlcN6P in both drz–Fpra–1 and drz–Fpra–2 were specific, we decided to test the ligand influence on different ribozymes of the same family, starting with two inactive drz–Fpra–1 mutants, C58A and C58U. Based on previous mutagenesis of HDV ribozymes[16, 17, 19, 22, 23, 51, 53], we expected a very slow self–cleavage of C58A and an abolishment of catalysis for C58U. In two–hour–long experiments, neither mutant of drz–Fpra–1 showed self–scission in the presence of GlcN6P (data not shown). These results demonstrate that GlcN6P does not rescue the activity of the ribozyme, further supporting a role as an allosteric modulator and not a catalytic co–factor for the ligand. To further investigate the specificity of GlcN6P, we probed two control ribozymes with the metabolite, the genomic HDV (gHDV) and the antigenomic HDV (aHDV), the two distinct HDV ribozymes from the hepatitis delta virus[6, 7]. No significant change in the cleavage rate was observed for these ribozymes (**Figure 3.5**), suggesting that the kinetic modulation by the metabolite is specific to the drz–Fpra ribozymes.

To narrow down the opposite effect promoted by GlcN6P on drz–Fpra–1 and drz–Fpra–2, we decided to study the discrepant nucleotides near the active sites of the two otherwise similar ribozymes. The catalytic cores of the two ribozymes differ by only three nucleotides. We tested three hybrids C23A and U26C in the L3, and A60G in the J4/2 regions of the drz–Fpra–1 sequence. We found that U26C retains, and even slightly increases the effect of GlcN6P on the cleavage rate constant; C23A seems to abrogate the metabolite effect; and A60G reverts the metabolite effect, decreasing the cleavage rate of the ribozyme when compared to the k_{WT} (**Figure 3.5**), to a level similar to the drz–

Fpra–2. These results indicate that the nucleotides C23 and A60 are involved in the allosteric modulation of drz–Fpra–1.

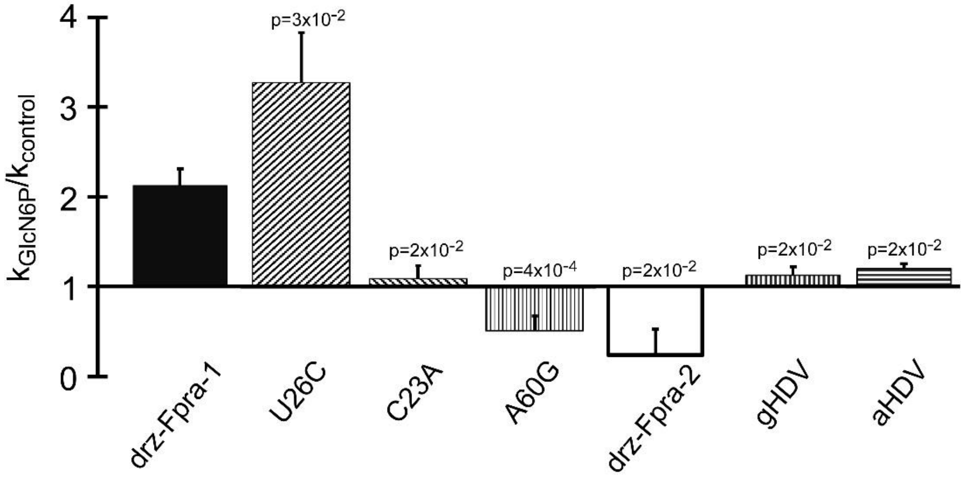


Figure 3.5. Effect of GlcN6P on self–scission of drz–Fpra–1/2 hybrids and HDV ribozymes. Comparison of modulation of the ribozyme activity by 20 mM GlcN6P normalized to the no–metabolite control at 5mM Mg^{2+}. U26C, C23A and A60G are drz–Fpra–1/drz–Fpra–2 hybrids referenced to the drz–Fpra–1 positions. gHDV and aHDV are the genomic and antigenomic HDV ribozymes, respectively. All p-values were calculated with respect to the wild–type Fpra–1 ribozyme and show statistical significance of the results.

3.4 Discussion

One of the nine natural families of self–cleaving ribozymes, the HDV–like ribozymes are found in several bacteria, many eukaryotes (including humans), *Chilo* iredescent virus, and in microbial metagenomic datasets[14, 22-27]. Little is known about their biological functions and regulation, particularly any roles that cofactors and metabolites have in altering their activity. Drz–Fpra–1[22] and drz–Fpra–2 HDV–like ribozymes were found in the human gut bacterium *F. prausnitzii* genome*,* surrounding the phosphoglucosamine mutase (*glmM*) open reading frame (ORF)[49]. The enzyme glmM catalyzes the transformation of glucosamine 6–phosphate (GlcN6P) into glucosamine 1–phosphate (GlcN1P) in the hexosamine biosynthesis pathway[31]. We have studied the effect of the metabolites from this pathway on the self–cleavage rate of the drz–Fpra ribozymes. Whereas GlcN6P increases the cleavage rate of the drz–Fpra–1 ribozyme and appears to stabilize its structure, GlcN1P, the product of the glmM enzymatic reaction, does not significantly affect the ribozyme.

The GlcN6P effect on drz–Fpra–1 showed that the ligand modestly increases the self–cleavage rate of the ribozyme when compared to the no–metabolite control. Titration of Mg^{2+} at constant GlcN6P or titration of ligand at constant Mg^{2+} showed similar results, increasing the self–cleavage rate of the ribozyme. When we examined the influence of GlcN6P on the Mg^{2+} dependence of the ribozymes, we observed that the ligand effect is not seen at low concentrations of the divalent cation, behaving similarly to the no–metabolite control. Mg^{2+} coordinated with a hydroxide anion likely acts as the general base in the self–cleavage reaction of HDV ribozymes, deprotonating the 2′–OH of the base that promotes the attack in the scissile phosphate[17], while the essential cytosine in

J4/2 (C58 in drz–Fpra–1) acts as a general acid, mediating a proton transfer to the 5′–oxygen of G1, the leaving group of the self–scission reaction[17]. If the effect promoted by the metabolite directly influenced the role of Mg^{2+} in the mechanism of the reaction, the k_{GlcN6P} would be expected to increase the cleavage rate in all instances throughout the ligand titration, including at low concentrations of Mg^{2+}. This result suggests that the effect caused by the ligand does not directly influence the catalytic role of the metal ion in the self–cleavage reaction.

In the case of titration of GlcN6P, the ribozyme responds to the ligand, cleaving faster than the no–metabolite control. Again, this effect is not seen at low concentrations of the metabolite, indicating that the effect is not directly related to the catalysis in the active site, otherwise we would expect to see an appreciably higher k_{GlcN6P} even at lower concentrations of GlcN6P. Moreover, testing the effect of the metabolite in the two inactive mutants of drz–Fpra–1, C58A and C58U, we found that the ligand does not rescue their activity, indicating that GlcN6P does not act as the general acid. This finding contrasts with the role of GlcN6P in the *glmS* ribozyme, where the cofactor actively participates in the active site, mediating the protonation of the 5′–oxyanion leaving group[36-39, 42-44, 54]. The *glmS* cofactor has been implicated in other roles in catalysis, including helping to align the active site, stabilizing the developing charge during the self–cleavage reaction, and participating in a set of competing hydrogen bonds to ensure potent activation and regulation of the catalysis[54, 55]. In contrast, our results suggest that the effect produced by GlcN6P on drz–Fpra–1 is not directly in the catalytic step of the self–cleavage reaction, suggesting an allosteric modulation of the ribozyme activity.

The structural probing of the drz–Fpra–1 in the presence of GlcN6P and GlcN1P showed decreased conformational flexibility of the RNA solely in the presence of GlcN6P (**Figures 3.4 and S3.4**). The active site of HDV ribozymes is formed by the P3/L3, P1.1 and J4/2 elements[16, 18, 19], and our probing data show that the ligand decreases the degradation of P3/L3, J1.1/4, L4 and A60 in J4/2. Even though GlcN6P clearly affects the active site stability and promotes faster self–scission, the in–line probing experiment does not pinpoint the site of interaction between the metabolite and the ribozyme, because the changes in degradation pattern may result from a conformational change promoted by the intermolecular interaction, and additional studies would be needed to map the exact position of interaction.

For all regions of intensity change of the in–line probing bands, the obtained estimated K_d was similar, with an average of 4.7 ± 0.2 mM. The intracellular concentration of GlcN6P in *F. prausnitzii* is not known, but the reported steady–state concentration in *E. coli* cells is 1.2 mM[56]. Although it is around 4 times lower than the estimated K_d for drz–Fpra–1 ribozyme, subcellular localization and spikes in metabolic activity may result in a higher concentration of the metabolite than the cell culture average. Thus, the metabolite concentration may reach intracellular levels that significantly alter the activity of the ribozyme.

The drz–Fpra–2 ribozyme, located 558 nucleotides downstream of the glmM gene, was found by sequence similarity when searching for drz–Fpra–1. Interestingly, GlcN6P promoted an opposite effect on drz–Fpra–2 than on drz–Fpra–1, decreasing the self–cleavage rate when compared to the $k_{control}$. This effect was unexpected and intriguing, bringing to our attention those nucleotides of the active site, where the ribozymes differ. We tested the effect of GlcN6P on the activity of the hybrids of drz–Fpra–1 and drz–Fpra–

2, with the following mutations made to drz–Fpra–1: C23A and U26C in the L3, and A60G in the J4/2 regions. The results demonstrate that U26C retains the effect of GlcN6P on the cleavage rate, maintaining the elevated k_{GlcN6P} and suggesting that the metabolite does not interact with the ribozyme at this nucleotide. On the other hand, the C23A mutation abrogates the metabolite effect, bringing the self–cleavage rate to the same level as the no–metabolite control. The third mutation, A60G, reverts the metabolite effect, decreasing the cleavage rate of the ribozyme below the $k_{control}$, similarly to what is observed in drz–Fpra–2. Thus, our data suggest that C23A and A60G are the discrepant nucleotides from the catalytic core that play a role in the opposite effect caused by GlcN6P in the two ribozymes. C23 is part of L3 and is important to maintain the active site structure, stacking on C24 and possibly hydrogen bonding with neighboring residues[16, 18, 19]. C23 may also be important for the stacking with U–1, similarly to the role of U23 in the HDV ribozyme[16, 18, 19]. A mutation at C23 may promote a conformational change that prevents the interaction of GlcN6P with the ribozyme. A60 is part of the trefoil turn of the ribozyme structure and makes an A–minor interaction with P3, participating in a network of hydrogen bonding between P3/L3 and J4/2 regions[16, 18, 19]. Thus, the A60G mutation would sterically hinder the interaction of the nucleobase, causing the J4/2 to move away from P3 and potentially disrupting the interactions of the A60 ribose as well. This change may be sufficient to abrogate the already weak interaction between the ribozyme and the metabolite.

Testing ribozyme kinetics in the presence of other metabolites elucidated the functional moieties critical in GlcN6P sensitivity in drz–Fpra–1. The aforementioned GlcN1P, the product of the enzymatic reaction of glmM, showed no significant effect on drz–Fpra–1 and a small effect on drz–Fpra–2. Similarly, Glc and UDP–GlcNAc, had little to no effect on the ribozymes. On the other hand, GlcN and Glc6P increased the cleavage rate for both ribozymes, leading us to speculate that both the amino and phosphate groups are relevant for the effect of the ligand in the ribozyme, as long as they are not at adjacent carbons of the sugar ring, as seen in GlcN1P. The concentration of Glc6P in the cytosol is likely very low and a previous report on the concentration of metabolites in *E. coli* with a detection limit of 130 nM did detect Glc6P[56]*.* Thus, despite an *in vitro* effect of Glc6P on the ribozymes, it is unlikely that it affects their activity *in vivo*.

The cleavage rate constants obtained for gHDV and aHDV ribozymes in the presence of GlcN6P were similar to those of the no–metabolite control. These results suggest that the interaction between the metabolite and drz–Fpra–1 is specific. Concerning the structure of the ribozymes, to our knowledge, the first P1.1 base pair is found to be A–U only in bacteria[22, 24, 25, 27, 49], and the drz–Fpra–1 was the only case of a bacterial HDV–like ribozyme upstream of a neighboring gene. However, a recent discovery of several ribozymes using comparative genomic analysis[14] also revealed HDV–like ribozymes in environmental samples, of which 4 showed the *glmM* gene 34–to–42 nts downstream of the ribozymes, similarly to drz–Fpra–1 in *F. prausnitzii* (37 nucleotides). Seven other HDV–like ribozymes showed the *glmM* gene upstream of the ribozymes. All of these ribozymes have an A–U base pair in P1.1, suggesting that they originate from closely related bacteria. The alignment of the drz–Fpra–1, drz–Fpra–2 and other microbial HDV–like ribozymes to the secondary structure is shown in **Figure S3.5**. The alignment, combined with the results presented here, suggests that env–26 HDV–like ribozyme would respond to GlcN6P, because it is the only sequence with a cytosine in the same position of L3 as in drz–Fpra–1 (C23), whereas all other ribozymes bear an adenosine at this position, like in drz–Fpra–2, where we found no effect of the metabolite. Regarding the J4/2 strand, only drz–Fpra–2 contains a guanosine residue at the position equivalent to A60 in drz–Fpra–1, here reported to be responsible for the inhibitory effect of GlcN6P on the activity of the ribozyme.

3.5 Conclusion

We show that the metabolite GlcN6P interacts with the drz–Fpra–1 ribozyme and increases its self–cleavage activity; however, we do not know the biological significance of these findings, and future studies will include biological assays addressing this question. In contrast to the *glmS* riboswitch–ribozyme, which is part of a feedback regulatory loop utilizing the product of the downstream metabolic step to induce the ribozyme self–scission and mRNA degradation, the activity of drz–Fpra–1 increases with the concentration of the substrate, providing an example of a putative feed–forward mechanism. The downstream RNA product, which starts with the cleaved drz–Fpra–1 ribozyme and is followed by the GlmM open reading frame, is terminated with a 5′–OH that is sequestered by the structure of the cleaved ribozyme[16, 18, 19]. This structure may protect the 5′ terminus against endonucleases, thus increasing the mRNA stability. Moreover, RNAs with 5′–OH termini have an extended half–life when compared with 5′–phosphorylated mRNAs, because 5′-hydroxyls are inferior substrates for the endonucleolytic degradation by RNase E[57, 58]. The extended half–life of mRNAs with 5′–OH termini over phosphorylated 5′ termini has been used in bacterial metabolic engineering to design a number of aptazymes that increased stability of the downstream transcripts upon ligand–dependent self–scission[59], and similar mechanism may be acting in the case of the *F. prausnitzii* ribozyme–terminated *glmM* mRNA. To the best of our knowledge, the drz–Fpra–1 activation by GlcN6P is the first example of an allosteric modulation of a natural self–cleaving ribozyme by a metabolite. We believe that this is not a unique case and that other examples of natural allostericly–regulated self–cleaving ribozymes exist, providing another example of gene expression regulation at the RNA level.

3.6 Materials and Methods

***In vitro* RNA transcription.** RNA was transcribed at 37 ˚C for one hour in a 20 μL volume containing 10 mM DTT, 2 mM spermidine, 1.25 mM each rNTP, 7.5 mM $MgCl_2$, 1 unit of T7 RNA polymerase, and 0.5 pmole of DNA template. The transcripts were purified by 10 % polyacrylamide gel electrophoresis (PAGE) under denaturing conditions (7M urea). RNA was eluted from the gel into 300 μL of 300 mM KCl and precipitated by adding 700 μL of 100% ethanol at –20 °C.

***In vitro* co–transcriptional cleavage kinetics.** *In vitro* transcription was performed similarly to the above–described RNA transcription assay with the following modifications: 4.5 mM $MgCl_2$; 1.25 mM of each GTP, UTP, CTP; 250 μM ATP; 4.5 μCi [α–^{32}P]–ATP (Perkin Elmer); and 40 mM HEPES pH 7.4. A 10 μL transcription reaction was initiated by the addition of DNA and incubated at 24 ˚C for 10 min. A 1.0 μL aliquot of the reaction was withdrawn and its transcription and self–scission terminated by the addition of urea loading buffer. The remaining 4.0 μL volume was diluted 25–fold (100 μL final volume) into a physiological–like buffer: 50 mM HEPES pH 7.4, 10 mM NaCl, 140 mM KCl, and the desired concentration of $MgCl_2$ and metabolite. Control experiments showed that this dilution efficiently prevented any new RNA synthesis; therefore, the kinetics of transcription did not need to be accounted for in our kinetic analysis, contrasting with the previously described analysis of co–transcriptional cleavage by Long and Uhlenbeck[45]. For conditions requiring consistent ionic strength, the buffer and metabolite stocks were pH–adjusted by the addition of KOH and the contribution of K^+ and Na^+ from the metabolite stocks was tracked. The concentration of K^+ was adjusted by the addition of KCl for a final reaction concentration of 140 mM. 5 μL aliquots were collected at the indicated times following the dilution of the transcription reaction into the physiological–like buffer at 37 ˚C and self–scission was terminated by adding 5 μL volume of stop buffer containing 20 mM EDTA, 5 mM Tris pH 7.4, 8 M urea, with xylene cyanol and bromophenol blue loading dyes. The denaturing PAGE gel of cleavage products was exposed to phosphorimage screens and analyzed using Typhoon phosphorimager and ImageQuant software (GE Healthcare). The band intensities were analyzed by creating line profiles of each lane using ImageQuant, exporting the data to Microsoft Excel. Self–cleavage data were fit to a mono–exponential decay function (Equation 1)

$$k_{obs} = A \cdot e^{(-kt)} + C \qquad \text{(Equation 1)}$$

where A and C represent the relative fraction of the ribozyme population cleaving with a rate constant *k* or remain uncleaved, respectively. The model was fit to the data using a linear least–squares analysis and the Solver module of Microsoft Excel.

RNA 3′–terminus labeling. RNA was *in vitro* transcribed and PAGE–purified. The appropriate RNA species was excised, precipitated, and re–suspended in water. RNA was then ligated at 37 °C for 3 hours in a volume of 10 µL, in RNA ligase buffer containing 50 mM Tris–HCl, 10 mM $MgCl_2$, 1 mM DTT and pH 7.5 (NEB), 2 µCi [5′–^{32}P] cytidine 3′, 5′– bisphosphate (Perkin Elmer) and one unit of T4 RNA ligase (NEB) and PAGE purified again.

In–line probing. The 3′–end labeled RNA was incubated with varying amounts of ligand for up to 2 days at 37 °C in a buffer containing 140 mM KCl, 10 mM NaCl, 20 mM Tris chloride, pH 8.5, 1 mM $MgCl_2$, and 1 mM spermidine, based on the in–line probing technique of Soukup and Breaker[46]. The partially hydrolyzed RNAs were resolved using denaturing PAGE, exposed to phosphorimage screens (Molecular Dynamics/GE Healthcare), and scanned by GE Typhoon phosphorimager. The sequences in the degradation pattern were assigned by running α–phosphorothioate nucleotide modified RNA cleaved by treatment with iodoethanol in parallel lanes[47, 48]. The band intensities were analyzed by creating line profiles of each lane using ImageQuant and exporting the data to Microsoft Excel. The areas of the fitted curves were used to measure intensity changes related to the binding, divided by intensities of a control band. The resulting ratios were plotted in Excel as a function of ligand concentration and modeled with a dissociation constant equation for a single ligand:

$$Fraction\ Bound = \frac{\left(\frac{[ligand]}{([ligand]+K_{Dd}} - Baseline\right)}{Range} \qquad \text{(Equation 2)}$$

The model was fit to the data using a linear least–squares analysis and the Solver module of Microsoft Excel.

Metabolites. All the metabolites used in this study were purchased from Sigma Aldrich. To measure the concentration of free phosphate, which may affect the ribozyme kinetics by forming an insoluble complex with the Mg^{2+} ions, we used a fluorescent phosphate sensor based on the bacterial phosphate–sensing protein (Thermo Fisher). The molar ratio of free phosphate in the GlcN1P, GlcN6P, and GlcP was found to be 0.0001, 0.002, and 0.0004.

3.7 Supplementary Data

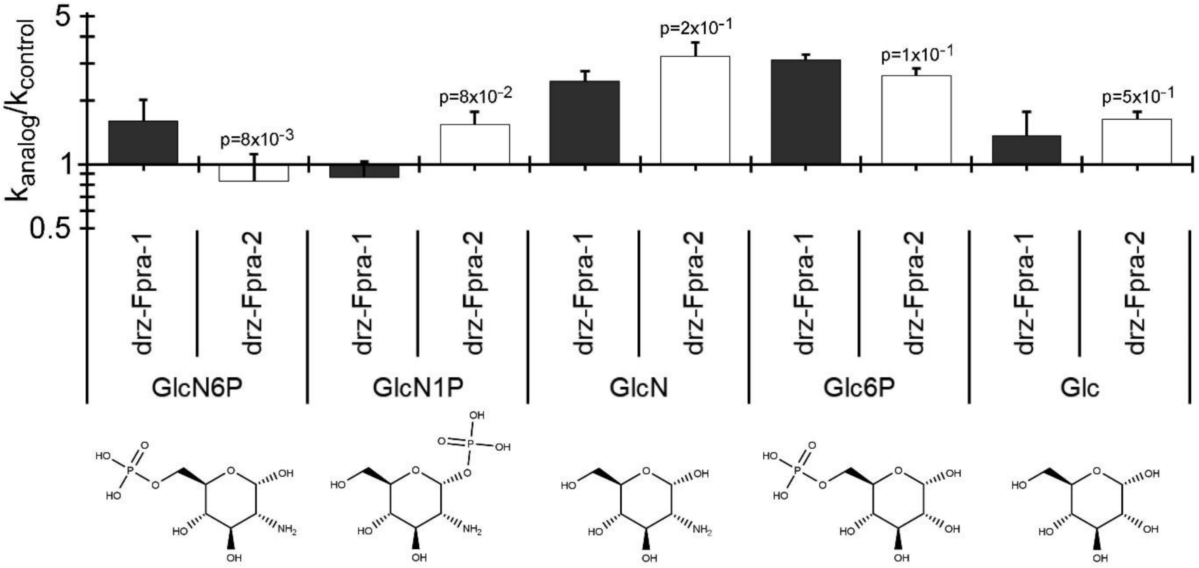


Figure S3.1. Drz–Fpra–1 and drz–Fpra–2 self-cleavage activity at 5 mM Mg^{2+} in the presence of different analogs at 20 mM normalized to the no-metabolite control.

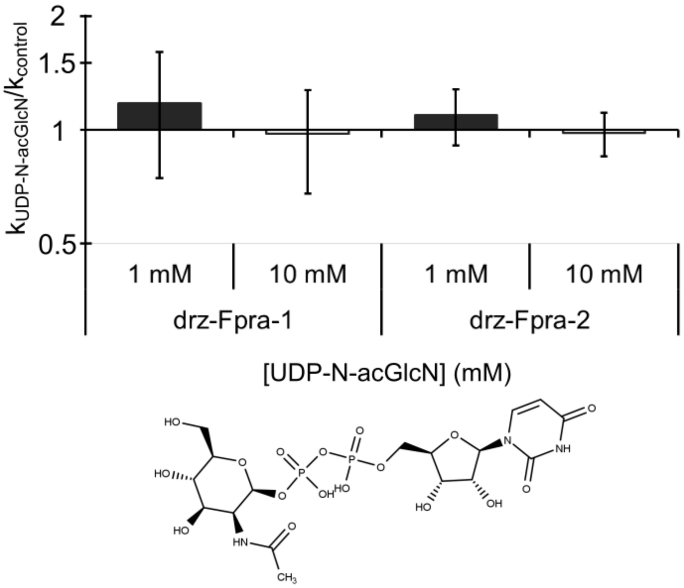


Figure S3.2. Drz–Fpra–1 and drz–Fpra–2 self-scission in the presence of UDP–GlcNAc normalized to the no-metabolite control.

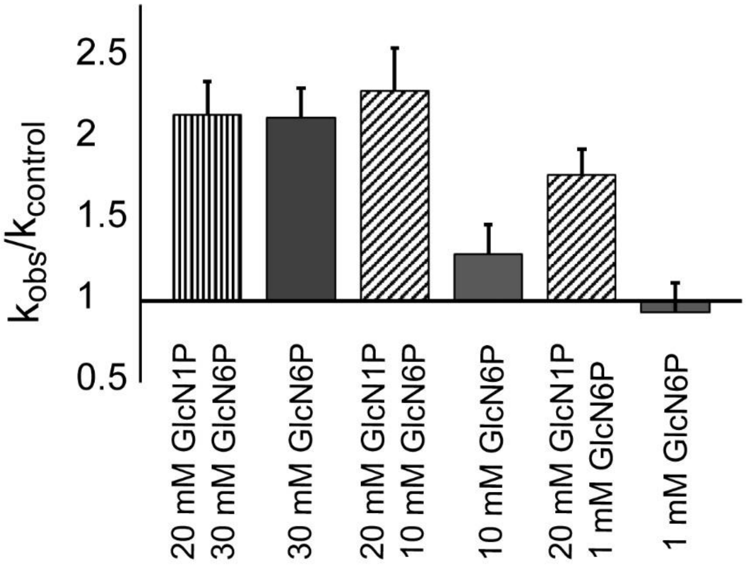


Figure S3.3. Effect of GlcN1P combined with GlcN6P on the activity of the drz–Fpra–1 ribozyme.

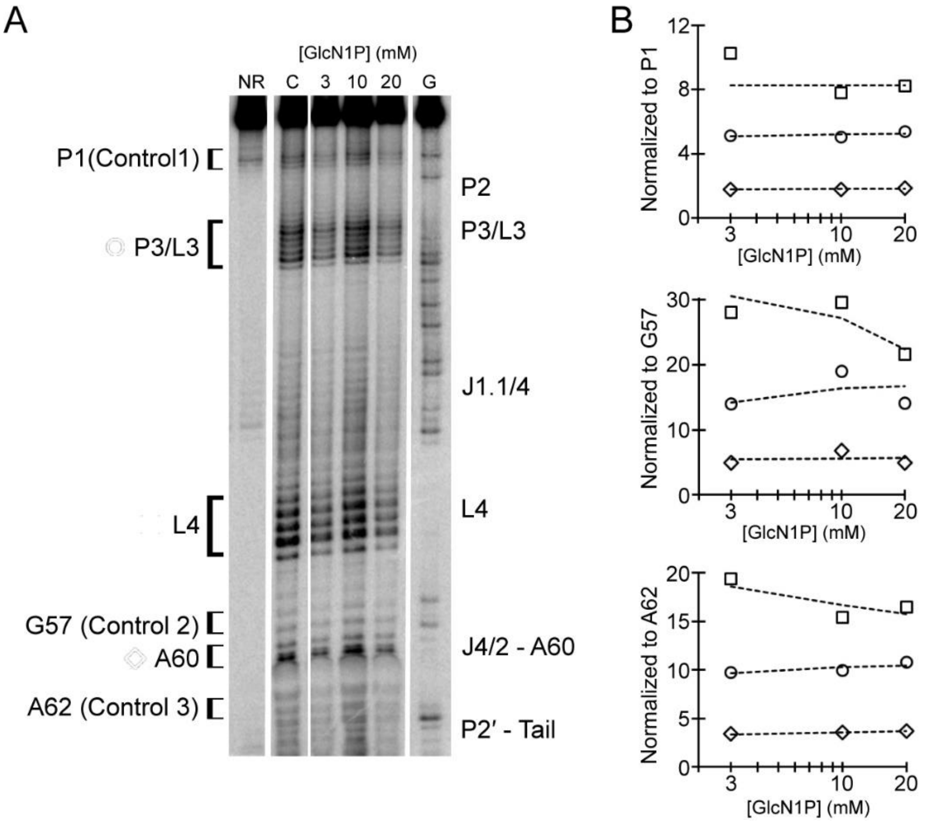
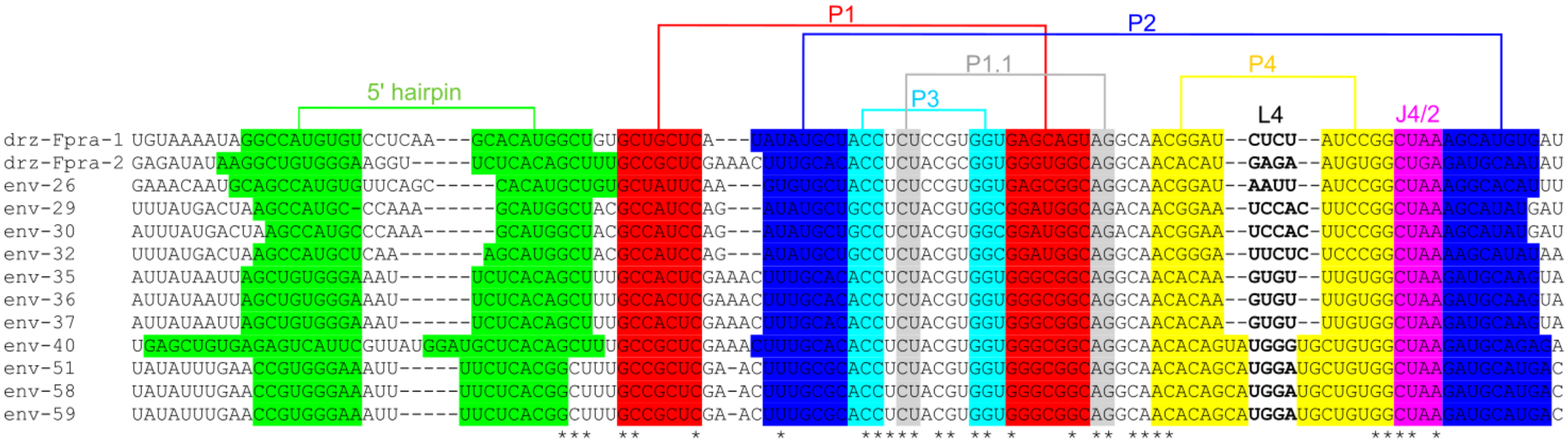


Figure S3.4. Structural probing of GlcN1P binding of the drz–Fpra–1 ribozyme (A) In-line probing of the 3′–labeled drz–Fpra–1 ribozyme in the presence of GlcN1P. No change in band intensity was observed. The sequence of the RNA was determined using iodoethanol cleavage of ribozymes with phosphorothioate-modified backbone at guanosines. (B) Three equivalent regions that showed shifted degradation pattern in the presence of GlcN6P (Figure 4) were chosen for comparison: P3/L3 (open circles), L4 (open squares), and A60 (open diamonds). Each region was normalized to three different controls (P1, G57 and A62). The data were fit to a model based on Equation 2. No dissociation constant was discernable.



```
drz-Fpra-1 UGUAAAAUAGGCCAUGUGUCCUCAA---GCACAUGGCUGUGCUGCUCA---UAUAUGCUACCUCUCCGUGGUGAGCAGUAGGCAACGGAU--CUCU--AUCCGGCUAAAGCAUGUGAU
drz-Fpra-2 GAGAUAUAAGGCUGUGGGAAGGU-----UCUCACAGCUUUGCCGCUCGAAACUUUGCACACCUCUACGCGGUGGGUGGCAGGCAACACAU--GAGA--AUGUGGCUGAGAUGCAAUAU
env-26     GAAACAAUGCAGCCAUGUGUUCAGC-----CACAUGCUGUGCUAUUCAA---GUGUGCUACCUCUCCGUGGUGAGCGGCAGGCAACGGAU--AAUU--AUCCGGCUAAAGGCACAUUU
env-29     UUUAUGACUAAGCCAUGC-CCAAA------GCAUGGCUACGCCAUCCAG---AUAUGCUGCCUCUACGUGGCGGAUGGCAGACAACGGAA--UCCAC-UUCCGGCUAAAGCAUAUGAU
env-30     AUUUAUGACUAAGCCAUGCCCAAA------GCAUGGCUACGCCAUCCAG---AUAUGCUGCCUCUACGUGGCGGAUGGCAGACAACGGAA--UCCAC-UUCCGGCUAAAGCAUAUGAU
env-32     UUUAUGACUAAGCCAUGCUCAA-------AGCAUGGCUACGCCAUCCAG---AUAUGCUGCCUCUACGUGGCGGAUGGCAGGCAACGGGA--UUCUC-UCCCGGCUAAAAGCAUAUAA
env-35     AUUAUAAUUAGCUGUGGGAAAU------UCUCACAGCUUUGCCACUCGAAACUUUGCACACCUCUACGUGGUGGGCGGCAGGCAACACAA--GUGU--UUGUGGCUAAGAUGCAAGUA
env-36     AUUAUAAUUAGCUGUGGGAAAU------UCUCACAGCUUUGCCACUCGAAACUUUGCACACCUCUACGUGGUGGGCGGCAGGCAACACAA--GUGU--UUGUGGCUAAGAUGCAAGUA
env-37     AUUAUAAUUAGCUGUGGGAAAU------UCUCACAGCUUUGCCACUCGAAACUUUGCACACCUCUACGUGGUGGGCGGCAGGCAACACAA--GUGU--UUGUGGCUAAGAUGCAAGUA
env-40     UGAGCUGUGAGAGUCAUUCGUUAUGGAUGCUCACAGCUUUGCCGCUCGAAACUUUGCACACCUCUACGUGGUGGGCGGCAGGCAACACAGUAUGGGUGCUGUGGCUAAGAUGCAGAGA
env-51     UAUAUUUGAACCGUGGGAAAUU-----UUCUCACGGCUUUGCCGCUCGA-ACUUUGCGCACCUCUACGUGGUGGGCGGCAGGCAACACAGCAUGGAUGCUGUGGCUAAGAUGCAUGAC
env-58     UAUAUUUGAACCGUGGGAAAUU-----UUCUCACGGCUUUGCCGCUCGA-ACUUUGCGCACCUCUACGUGGUGGGCGGCAGGCAACACAGCAUGGAUGCUGUGGCUAAGAUGCAUGAC
env-59     UAUAUUUGAACCGUGGGAAAUU-----UUCUCACGGCUUUGCCGCUCGA-ACUUUGCGCACCUCUACGUGGUGGGCGGCAGGCAACACAGCAUGGAUGCUGUGGCUAAGAUGCAUGAC
                                              ***  **    *      *      ***** ** ** *    * ** ****                ****  *
```

Figure S3.5. Secondary structure alignment of drz–Fpra–1 and drz–Fpra–2 with HDV–like ribozymes mapping near *glmM* genes reported previouly (Weinberg *et al.* (2015) *Nat. Chem. Biol*. 11(8), 606–610).

Table S3.1. DNA sequences of oligonucleotides used in the study. T7 promoters are shown italicized and self-cleaved ribozymes are underlined. Point mutations are boxed in gray.

Name	DNA Sequence (5′ – 3′)
Drz –Fpra –1	TTCCCGCGAAAT*TAATACGACTCACTATA*GGGAGCTGATGTAAAATAGGCCATGTGTCCTCAAGCACATGGCTGTGCTGCTCATATATGCTACCTCTCCGTGGTGAGCAGTAGGCAACGGATCTCTATCCGGCTAAAGCATGTGATTGTC
Drz –Fpra –2	TTCCCGCGAAAT*TAATACGACTCACTATA*GGGGAGATATAAGGCTGTGGGAAGGTTCTCACAGCTTTGCCGCTCGAAACTTTGCACACCTCTACGCGGTGGGTGGCAGGCAACACATGAGAATGTGGCTGAGATGCAATATTGTC
Drz –Fpra –1 C58U	TTCCCGCGAAAT*TAATACGACTCACTATA*GGGAGCTGATGTAAAATAGGCCATGTGTCCTCAAGCACATGGCTGTGCTGCTCATATATGCTACCTCTCCGTGGTGAGCAGTAGGCAACGGATCTCTATCCGGTTAAAGCATGTGATTGTC
Drz –Fpra –1 C58A	TTCCCGCGAAAT*TAATACGACTCACTATA*GGGAGCTGATGTAAAATAGGCCATGTGTCCTCAAGCACATGGCTGTGCTGCTCATATATGCTACCTCTCCGTGGTGAGCAGTAGGCAACGGATCTCTATCCGGATAAAGCATGTGATTGTC
gHDV	TTCCCGCGAAAT*TAATACGACTCACTATA*GGGACCCCACTCTGCAGGGTCCGCGTTCCATCCTTTCTTACCTGATGGCCGGCATGGTCCCAGCCTCCTCGCTGGCGCCGGCTGGGCAACATTCCGAGGGGACCGTCCCCTCGGTAATGGCGAATGGGACCC
aHDV	TTCCCGCGAAAT*TAATACGACTCACTATA*GGGAGGTTTGCGTCTCGCGTCCTTCTTTCCTCTTCGGGTCGGCATGGCATCTCCACCTCCTCGCGGTCCGACCTGGGCATCCGAAGGAGGACGCACGTCCACTCGGATGGCTAAGGGAGAGCCA
Hybrid: Drz –Fpra –1/2 C23A	TTCCCGCGAAAT*TAATACGACTCACTATA*GGGAGCTGATGTAAAATAGGCCATGTGTCCTCAAGCACATGGCTGTGCTGCTCATATATGCTACCTCTACGTGGTGAGCAGTAGGCAACGGATCTCTATCCGGCTAAAGCATGTGATTGTC
Hybrid: Drz –Fpra –1/2 U26C	TTCCCGCGAAAT*TAATACGACTCACTATA*GGGAGCTGATGTAAAATAGGCCATGTGTCCTCAAGCACATGGCTGTGCTGCTCATATATGCTACCTCTCCGCGGTGAGCAGTAGGCAACGGATCTCTATCCGGCTAAAGCATGTGATTGTC
Hybrid: Drz –Fpra –1/2 A60G	TTCCCGCGAAAT*TAATACGACTCACTATA*GGGAGCTGATGTAAAATAGGCCATGTGTCCTCAAGCACATGGCTGTGCTGCTCATATATGCTACCTCTCCGTGGTGAGCAGTAGGCAACGGATCTCTATCCGGCTGAAGCATGTGATTGTC

3.7 References

1. Doudna, J. A., and Cech, T. R. The chemical repertoire of natural ribozymes. *Nature* **418**, 222-228 (2002).
2. Kruger, K., Grabowski, P. J., Zaug, A. J., Sands, J., Gottschling, D. E., and Cech, T. R. Self-splicing RNA: autoexcision and autocyclization of the ribosomal RNA intervening sequence of Tetrahymena. *Cell* **31**, 147-157 (1982).
3. Guerrier-Takada, C., Gardiner, K., Marsh, T., Pace, N., and Altman, S. The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme. *Cell* **35**, 849-857 (1983).
4. Jimenez, R. M., Polanco, J. A., and Luptak, A. Chemistry and biology of self-cleaving ribozymes. *Trends Biochem Sci* **40**, 648-661 (2015).
5. Fedor, M. J. Comparative enzymology and structural biology of RNA self-cleavage. *Annu Rev Biophys* **38**, 271-299 (2009).
6. Wu, H. N., Lin, Y. J., Lin, F. P., Makino, S., Chang, M. F., and Lai, M. M. Human hepatitis delta virus RNA subfragments contain an autocleavage activity. *Proc Natl Acad Sci U S A* **86**, 1831-1835 (1989).
7. Sharmeen, L., Kuo, M. Y., Dinter-Gottlieb, G., and Taylor, J. Antigenomic RNA of human hepatitis delta virus can undergo self-cleavage. *J. Virol.* **62**, 2674-2679 (1988).
8. Prody, G. A., Bakos, J. T., Buzayan, J. M., Schneider, I. R., and Bruening, G. Autolytic processing of dimeric plant virus satellite RNA. *Science* **231**, 1577-1580 (1986).
9. Hutchins, C. J., Rathjen, P. D., Forster, A. C., and Symons, R. H. Self-cleavage of plus and minus RNA transcripts of avocado sunblotch viroid. *Nucleic Acids Res* **14**, 3627-3640 (1986).
10. Buzayan, J. M., Gerlach, W. L., and Bruening, G. Non-enzymatic cleavage and ligation of RNAs complementary to a plant virus satellite RNA. *Nature* **232**, 349-353 (1986).
11. Winkler, W. C., Nahvi, A., Roth, A., Collins, J. A., and Breaker, R. R. Control of gene expression by a natural metabolite-responsive ribozyme. *Nature* **428**, 281-286 (2004).
12. Saville, B. J., and Collins, R. A. A site-specific self-cleavage reaction performed by a novel RNA in Neurospora mitochondria. *Cell* **61**, 685-696 (1990).
13. Roth, A., Weinberg, Z., Chen, A. G. Y., Kim, P. B., Ames, T. D., and Breaker, R. R. A widespread self-cleaving ribozyme class is revealed by bioinformatics. *Nat Chem Biol* **10**, 56-U92 (2014).

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

15. Kuo, M. Y., Sharmeen, L., Dinter-Gottlieb, G., and Taylor, J. Characterization of self-cleaving RNA sequences on the genome and antigenome of human hepatitis delta virus. *J Virol* **62**, 4439-4444 (1988).

16. Ferre-D'Amare, A. R., Zhou, K., and Doudna, J. A. Crystal structure of a hepatitis delta virus ribozyme. *Nature* **395**, 567-574 (1998).

17. Das, S. R., and Piccirilli, J. A. General acid catalysis by the hepatitis delta virus ribozyme. *Nat Chem Biol* **1**, 45-52 (2005).

18. Chen, J. H., Yajima, R., Chadalavada, D. M., Chase, E., Bevilacqua, P. C., and Golden, B. L. A 1.9 A crystal structure of the HDV ribozyme precleavage suggests both Lewis acid and general acid mechanisms contribute to phosphodiester cleavage. *Biochemistry* **49**, 6508-6518 (2010).

19. Ke, A., Zhou, K., Ding, F., Cate, J. H., and Doudna, J. A. A conformational switch controls hepatitis delta virus ribozyme catalysis. *Nature* **429,** 201-205 (2004).

20. Koo, S. C., Lu, J., Li, N. S., Leung, E., Das, S. R., Harris, M. E., and Piccirilli, J. A. Transition state features in the hepatitis delta virus ribozyme reaction revealed by atomic perturbations. *J Am Chem Soc* **137**, 8973-8982 (2015).

21. Golden, B. L., Hammes-Schiffer, S., Carey, P. R., and Bevilacqua, P. C. An integrated picture of HDV ribozyme catalysis. *In Biophysics of RNA Folding (Russell, R., Ed.)*, pp 135-167, Springer New York, New York, NY (2013).

22. Webb, C. H., Riccitelli, N. J., Ruminski, D. J., and Lupták, A. Widespread occurrence of self-cleaving ribozymes. *Science* **326**, 953 (2009).

23. Salehi-Ashtiani, K., Lupták, A., Litovchick, A., and Szostak, J. W. A genomewide search for ribozymes reveals an HDV-like sequence in the human CPEB3 gene. *Science* **313**, 1788-1792 (2006).

24. Ruminski, D. J., Webb, C. H., Riccitelli, N. J., and Lupták, A. Processing and translation initiation of non-long terminal repeat retrotransposons by hepatitis delta virus (HDV)-like self-cleaving ribozymes. *J Biol Chem* **286**, 41286-41295 (2011).

25. Eickbush, D. G., and Eickbush, T. H. R2 retrotransposons encode a self-cleaving ribozyme for processing from an rRNA cotranscript. *Mol Cell Biol* **30**, 3142-3150 (2010).

26. Riccitelli, N. J., Delwart, E., and Luptak, A. Identification of minimal HDV-like ribozymes with unique divalent metal ion dependence in the human microbiome. *Biochemistry* **53**, 1616-1626 (2014).

27. Sanchez-Luque, F. J., Lopez, M. C., Macias, F., Alonso, C., and Thomas, M. C. Identification of an hepatitis delta virus-like ribozyme at the mRNA 5'-end of the L1Tc retrotransposon from Trypanosoma cruzi. *Nucleic Acids Res* **39**, 8065-8077.

28. Miquel, S., Martin, R., Rossi, O., Bermudez-Humaran, L. G., Chatel, J. M., Sokol, H., Thomas, M., Wells, J. M., and Langella, P. (2013) Faecalibacterium prausnitzii and human intestinal health. *Curr Opin Microbiol* **16**, 255-261 (2011).

29. Heinken, A., Khan, M. T., Paglia, G., Rodionov, D. A., Harmsen, H. J., and Thiele, I. Functional metabolic map of Faecalibacterium prausnitzii, a beneficial human gut microbe. *J Bacteriol* **196**, 3289-3302 (2014).

30. Sokol, H., Pigneur, B., Watterlot, L., Lakhdari, O., Bermudez-Humaran, L. G., Gratadoux, J. J., Blugeon, S., Bridonneau, C., Furet, J. P., Corthier, G., Grangette, C., Vasquez, N., Pochart, P., Trugnan, G., Thomas, G., Blottiere, H. M., Dore, J., Marteau, P., Seksik, P., and Langella, P. Faecalibacterium prausnitzii is an anti-inflammatory commensal bacterium identified by gut microbiota analysis of Crohn disease patients. *Proc Natl Acad Sci U S A* **105**, 16731-16736 (2008).

31. Mengin-Lecreulx, D., and van Heijenoort, J. Characterization of the essential gene glmM encoding phosphoglucosamine mutase in Escherichia coli. *J Biol Chem* **271**, 32-39 (1996).

32. Raetz, C. R. Molecular genetics of membrane phospholipid synthesis. *Annu Rev Genet* **20**, 253-295 (1986).

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

34. McCown, P. J., Corbino, K. A., Stav, S., Sherlock, M. E., and Breaker, R. R. Riboswitch diversity and distribution. *RNA* **23**, 995-1011 (2017).

35. Klein, D. J., and Ferre-D'Amare, A. R. Structural basis of glmS ribozyme activation by glucosamine-6-phosphate. *Science* **313**, 1752-1756 (2006).

36. Cochrane, J. C., Lipchock, S. V., and Strobel, S. A. Structural investigation of the GlmS ribozyme bound to Its catalytic cofactor. *Chem Biol* **14**, 97-105 (2007).

37. Viladoms, J., and Fedor, M. J. The glmS ribozyme cofactor is a general acid-base catalyst. *J Am Chem Soc* **134**, 19043-19049 (2012).

38. Cochrane, J. C., Lipchock, S. V., Smith, K. D., and Strobel, S. A. Structural and chemical basis for glucosamine 6-phosphate binding and activation of the glmS ribozyme. *Biochemistry* **48**, 3239-3246 (2009).

39. McCarthy, T. J., Plog, M. A., Floy, S. A., Jansen, J. A., Soukup, J. K., and Soukup, G. A. Ligand requirements for glmS ribozyme self-cleavage. *Chem Biol* **12**, 1221-1226 (2005).

40. Collins, J. A., Irnov, I., Baker, S., and Winkler, W. C. Mechanism of mRNA destabilization by the glmS ribozyme. *Genes Dev* **21**, 3356-3368 (2007).

41. Watson, P. Y., and Fedor, M. J. The glmS riboswitch integrates signals from activating and inhibitory metabolites in vivo. *Nat Struct Mol Biol* **18**, 359-363 (2011).

42. Lau, M. W., and Ferre-D'Amare, A. R. An in vitro evolved glmS ribozyme has the wild-type fold but loses coenzyme dependence. *Nat Chem Biol* **9**, 805-810 (2013).

43. Soukup, G. A. Core requirements for glmS ribozyme self-cleavage reveal a putative pseudoknot structure. *Nucleic Acids Res* **34**, 968-975 (2006).

44. Jansen, J. A., McCarthy, T. J., Soukup, G. A., and Soukup, J. K. Backbone and nucleobase contacts to glucosamine-6-phosphate in the glmS ribozyme. *Nat Struct Mol Biol* **13**, 517-523 (2006).

45. Long, D. M., and Uhlenbeck, O. C. Kinetic characterization of intramolecular and intermolecular hammerhead RNAs with stem II deletions. *Proc Natl Acad Sci U S A* **91**, 6977-6981 (1994).

46. Soukup, G. A., and Breaker, R. R. Relationship between internucleotide linkage geometry and the stability of RNA. *RNA* **5**, 1308-1325 (1999).

47. Ryder, S. P., and Strobel, S. A. Nucleotide analog interference mapping. *Methods* **18**, 38-50 (1999).

48. Vu, M. M., Jameson, N. E., Masuda, S. J., Lin, D., Larralde-Ridaura, R., and Luptak, A. Convergent Evolution of Adenosine Aptamers Spanning Bacterial, Human, and Random Sequences Revealed by Structure-Based Bioinformatics and Genomic SELEX. *Chem Biol* **19**, 1247-1254 (2012).

49. Webb, C. H., and Lupták, A. HDV-like self-cleaving ribozymes. *RNA Biol* **8**, 719 – 727 (2011).

50. Cerrone-Szakal, A. L., Siegfried, N. A., and Bevilacqua, P. C. Mechanistic characterization of the HDV genomic ribozyme: solvent isotope effects and proton inventories in the absence of divalent metal ions support C75 as the general acid. *J Am Chem Soc* **130**, 14504-14520 (2008).

51. Nakano, S., Chadalavada, D. M., and Bevilacqua, P. C. General acid-base catalysis in the mechanism of a hepatitis delta virus ribozyme. *Science* **287**, 1493-1497 (2000).

52. Chen, J., Ganguly, A., Miswan, Z., Hammes-Schiffer, S., Bevilacqua, P. C., and Golden, B. L. Identification of the catalytic Mg(2)(+) ion in the hepatitis delta virus ribozyme. *Biochemistry* **52**, 557-567 (2013).

53. Perrotta, A. T., Shih, I., and Been, M. D. Imidazole rescue of a cytosine mutation in a self-cleaving ribozyme. *Science* **286**, 123-126 (1999).

54. Bingaman, J. L., Zhang, S., Stevens, D. R., Yennawar, N. H., Hammes-Schiffer, S., and Bevilacqua, P. C. The GlcN6P cofactor plays multiple catalytic roles in the glmS ribozyme. *Nat Chem Biol* **13**, 439-445 (2017).

55. Bingaman, J. L., Gonzalez, I. Y., Wang, B., and Bevilacqua, P. C. Activation of the glmS ribozyme Nnucleophile via overdetermined hydrogen bonding. *Biochemistry* **56**, 4313-4317 (2017).

56. Bennett, B. D., Kimball, E. H., Gao, M., Osterhout, R., Van Dien, S. J., and Rabinowitz, J. D. Absolute metabolite concentrations and implied enzyme active site occupancy in Escherichia coli. *Nat Chem Biol* ***5***, 593-599 (2009).

57. Celesnik, H., Deana, A., and Belasco, J. G. Initiation of RNA decay in Escherichia coli by 5' pyrophosphate removal. *Mol Cell* ***27***, 79-90 (2007).

58. Deana, A., Celesnik, H., and Belasco, J. G. The bacterial enzyme RppH triggers messenger RNA degradation by 5' pyrophosphate removal. *Nature* ***451***, 355-358 (2008).

59. Carothers, J. M., Goler, J. A., Juminaga, D., and Keasling, J. D. Model-driven engineering of RNA devices to quantitatively program gene expression. *Science* **334**, 1716-1719 (2011).

Chapter 4: Co–transcriptional Analysis of Self–cleaving Ribozymes and Their Ligand Dependence

Publication note:

The content of this chapter is currently accepted for publication on *Methods in Molecular Biology* book on ribozymes.

Passalacqua, L. F. M., Lupták, A. Co–transcriptional Analysis of Self–cleaving Ribozymes and Their Ligand Dependence. *In press - Methods in Molecular Biology*

4.1 Abstract

Self–cleaving ribozymes are RNA molecules that catalyze a site–specific self–scission reaction. Analysis of self–cleavage is a crucial aspect of the biochemical study and understanding of these molecules. Here we describe a co–transcriptional assay that allows the analysis of self–cleaving ribozymes in different reaction conditions and in the presence of desired ligands and/or cofactors. Utilizing a standard T7 RNA polymerase *in vitro* transcription system under limiting Mg^{2+} concentration, followed by a 25–fold dilution of the reaction in desired conditions of self–cleavage (buffer, ions, ligands, pH, temperature, etc.) to halt the synthesis of new RNA molecules, allows the study of self–scission of these molecules without the need for purification or additional preparation steps, such as refolding procedures. Furthermore, because the transcripts

are not denatured, this assay likely yields RNAs in conformations relevant to co–transcriptionally folded species *in vivo*.

4.2 Introduction

Discovered over 30 years ago[1–5], self–cleaving ribozymes are catalytic RNA molecules that promote a site–specific self–scission reaction. The most common mechanism of the self–scission reaction is a general acid–base catalysis, where a transesterification involves a nucleophilic attack by a 2′–oxygen on the adjacent phosphodiester bond, producing a 2′−3′ cyclic phosphate and a 5′–hydroxyl product[6–9], with metal ions and metabolites employed as potential cofactors[6, 7, 10, 11]. To date, self–cleaving ribozyme families have been discovered in nature, comprising the hammerhead[1, 2], hairpin[3], hepatitis delta virus (HDV)[4, 12], glucosamine–6–phosphate synthase (*glmS*)[11], *Neurospora* Varkud satellite (VS)[5], twister[13], twister sister (TS), pistol, and hatchet motifs[14].

Self–cleaving ribozymes are broadly distributed throughout all branches of life[13–16]. Likely involved in several roles in biology, some of the known functions include self–scission during rolling–circle replication of RNA genomes, co–transcriptional processing of retrotransposons, and metabolite–dependent gene expression regulation in bacteria[1, 2, 4, 5, 11, 17–25]. Recently, it has also been shown that metabolites may modulate the activity of self–scission of some ribozymes[26]. Genomic locations of these ribozymes suggest that they affect many other biological processes, some of which may not be directly associated with RNA scission. Other examples, including highly conserved mammalian ribozymes[27, 28], suggest that many new biological roles are yet to be discovered.

The discovery and understanding of self–cleaving ribozymes and their roles depend on the biochemical characterization of these molecules. An important aspect of this characterization is the kinetics investigation of self–scission under different conditions and in the presence of metabolites, cofactors, and other potential ligands, such as protein chaperones. The most common options in the investigation of the kinetics of ribozymes are the study of pre–purified ribozymes and the co–transcriptional self–cleavage analysis. The first method typically uses denaturing PAGE to fractionate transcripts of isolate uncleaved ribozymes, precipitation, and a refolding step, although less harsh conditions, such as non–denaturing chromatography and precipitation–less concentration of the purified samples have been utilized[29]. A denaturing step is best avoided if refolding of the ribozymes can lead to conformations different from the co–transcriptional folding. This may happen when the RNA loses the directional order of folding (5′ to 3′) and the co–existence of different folding states is likely to increase. For example, a comparison of the co–transcriptional folding and Mg^{2+}–initiated refolding after precipitation of the RNase P ribozyme, revealed that even though all folding processes have kinetic traps and misfold, the Mg^{2+}–initiated refolding involves residues in different regions of the molecule, while the co–transcriptional folding allows the 5′ region to fold before the 3′ region, eliminating major misfold traps[30]. Furthermore, it has been shown that the co–transcriptional folding notably enhances the self–scission of the human HDV–like ribozyme CPEB3 when compared to a pre–purified sample[31], and that self–cleavage transcripts of the HDV ribozyme with an attenuator in the 3′ end could not be restored efficiently by renaturation[32]. Hence, the use of the co–transcriptional self–cleavage analysis is preferred when possible.

Standard co–transcriptional analysis relies on the study of the self–cleavage reaction in transcriptional buffer, while transcription occurs. This methodology is limited to the experimental conditions compatible with RNA polymerase activity. Additionally, the concurrent synthesis of new molecules of RNA has to be accounted for, adding a second kinetic element to the analysis. Herein, we provide an alternative co–transcriptional assay that allows the analysis of self–cleaving ribozymes in different reaction conditions and in the presence of desired ligands. The *in vitro* transcription is performed under limiting amount of Mg^{2+}, and the reaction is followed by a 25–fold dilution in desired condition of self–cleavage (buffer, ions, ligand, pH, temperature, etc.). The limiting amount of Mg^{2+} reduces the self–scission reaction during the initial transcription. Control experiments showed that a 25–fold dilution efficiently prevents any new RNA synthesis; therefore, our co–transcriptional kinetic analysis does not need to account for the kinetics of transcription, contrasting with the previously described analysis of co–transcriptional cleavage by Long and Uhlenbeck[33]. This method allows study of self–scission of these RNAs without the need of purification and a second kinetic parameter. This approach is also useful to synthetic biology, as self–cleaving ribozyme can be used as platforms to the development of new molecular biology tools, particularly gene expression–regulating aptazymes[34]. The method is also applicable to the study of other types of ribozymes, such as self–splicing introns[35, 36].

For the illustration of the methodology, we present an example of the HDV–like ribozyme drz–Fprau–1. Found in the human gut bacterium *Faecalibacterium prausnitzii*, the ribozyme cleavage site maps 106 nucleotides upstream of the phosphoglucosamine mutase (*glmM*) open reading frame[15]. The enzyme glmM catalyzes the transformation of glucosamine 6–phosphate (GlcN6P) into glucosamine 1–phosphate (GlcN1P)[37]. It was shown that GlcN6P, a natural metabolite, increases the self–scission rate of the ribozyme when compared to a no metabolite control[26].

4.3 Materials

Working with RNA requires care to avoid contamination by RNases. All solutions should be prepared using double–distilled RNase–free water (ddH_2O) and analytical grade reagents. All solutions should be tested for the presence of RNases before use. Chemicals and reagents are purchased from commercial suppliers. Radioactive [α–^{32}P] ATP should be handled and disposed with safety and according to current regulations of purchase, use, and disposal.

4.3.1 *In vitro t*ranscription

1. 10× transcription buffer: 400 mM Tris–HCl or HEPES pH 7.5, 100 mM DTT (dithiothreitol), 20 mM spermidine, 1 % Triton X–100 (see Notes 1 and 2).
2. 100 mM $MgCl_2$ stock.
3. 25 mM stocks of each rGTP, rUTP, and rCTP.
4. 2.5 mM stock of rATP.
5. [α–^{32}P] rATP.
6. T7 RNA polymerase.
7. Purified stock of DNA template of the ribozyme to be studied with T7 promoter (see Note 3).

4.3.2 Self–cleavage assay

The cleavage buffer depends on the system to be studied and also on the experiment proposed. For the example illustrated here, a physiological–like buffer is used.

1. 2× self–cleavage buffer: 100 mM Tris–HCl or HEPES (pH 7.4), 20 mM NaCl, 280 mM KCl (see Note 4).
2. 100 mM $MgCl_2$ stock.
3. Stocks of the metabolite(s) of interest (if applicable).

4.3.3 Denaturing loading buffer

1. 2× denaturing loading buffer: 8 M urea, 20 mM EDTA (see Note 5) (ethylenediaminetetraacetic acid), 0.05 % (w/v) bromophenol blue, and 0.05 % (w/v) xylene cyanol.

4.3.4 Denaturing Polyacrylamide Gel Electrophoresis (PAGE)

1. 40 % bis–acrylamide 19:1.
2. Urea.
3. 10× TBE:890mM Tris–borate, 890mM boric acid, 20mM EDTA, pH 8.3.
4. TEMED (N,N,N′,N′–tetramethylethylenediamine).
5. 10 % (w/v) ammonium persulfate (APS).
6. Polyacrylamide gel solution: 8 M urea, 0.5× TBE, 15 % bis–acrylamide 19:1 (store away from light).
7. Diluent gel solution: 8 M urea, 0.5× TBE.

8. Set of small plates (16.5 × 22 cm) or medium plates (16.5 × 28 cm) glass plates for PAGE, 1.5– and 0.8–mm Teflon spacers, wide–toothed and narrow toothed combs to cast wells, cellulose chromatography paper, and plastic wrap.
9. Electrophoresis power supply.
10. Storage phosphor image screen.

4.4 Methods

4.4.1 *In vitro* RNA transcription under minimal Mg^{2+} conditions

1. Prepare 100 μL the *in vitro* transcription mix by adding: 10 μL of the 10× transcription buffer, 3.5 μL of the 100 mM $MgCl_2$ stock, 4 μL of the 25 mM stock of each rGTP, rUTP, and rCTP, 10 μL of 2.5 mM stock of rATP, 1 μL of [α–^{32}P] rATP, 1 unit of T7 RNA polymerase, complete volume to 90 μL with ddH_2O. Note that 10 % of the solution is accounted for the DNA template to be added in the next step. This mix can be saved at –20 °C prior to T7 RNA polymerase addition or at 4 °C after T7 RNA polymerase addition for later use. The goal of this step is to prepare an efficient transcription reaction under conditions with most of Mg^{2+} chelated by the rNTPs, leaving minimal Mg^{2+} to promote ribozyme catalysis.
2. Initiate transcription by adding 0.5 μL of the DNA template with a T7 promoter of the ribozyme construct to be studied (~ 0.5 pmol) into 4.5 μL of the *in vitro* transcription mix.
3. Incubate the reaction at 24 °C (room temperature) for 10 min, to initiate transcription (see Note 6).

4.4.2 Kinetics assay

1. Prepare the self–cleavage mix by adding: 50 μL of the 2× self–cleavage buffer, 5 μL of the 100 mM $MgCl_2$ stock, and desired concentration of any metabolite(s) if applicable. Complete volume to 96 μL with ddH_2O. Incubate at desired temperature of self–scission assay (see Note 7).
2. Set up a cone–bottom well–plate to aliquot time–point fractions and terminate reactions. Each well should contain 5 μL of 2× denaturing loading buffer.
3. After 10 minutes of the *in vitro* transcription, withdraw a 1 μL aliquot of the reaction mixture, and terminate its transcription and self–scission by the addition of the 2× denaturing loading buffer. This is the zero time–point, used as reference for the kinetic analysis, showing the extent of the RNA production and self–scission that occurred during the transcription period. Control experiments showed that 10 minutes *in vitro* transcription is enough time to make sufficient ^{32}P–labeled RNA without significant self–scission catalysis.
4. Transfer the remaining 4.0 μL volume of the transcription reaction into the pre–incubated 96 μL of self–cleavage mix (25–fold dilution) and start timing the self–cleavage reaction under the new conditions. Collect aliquots of 5 μL at the desired time–points, and terminate the self–scission by depositing the aliquots into the pre–prepared cone–bottom well–plate with 5 μL of the 2× denaturing loading buffer.

4.4.3 Resolving the results

Results are resolved using denaturing PAGE.

1. Prepare 50 mL of bis–acrylamide gel solution of the appropriate percentage for the RNA sample. For a 10 % polyacrylamide gel solution, dilute the 15 % bis–acrylamide gel solution stock solution with the diluent gel solution (see Note 8).
2. Wash the glass plates, 0.8 mm spacers and small–tooth combs thoroughly with distilled water, then 70 % ethanol solution. With bottom and side spacers properly placed, clip the plates together with clamps, making sure that there are no gaps between spacers in the bottom corners of the plates.
3. In a 15 mL tube, add 3 mL of the 15 % bis–acrylamide gel solution. Add 20 μL of TEMED and mix. To this, add 30 μL of the 10 % APS solution to initialize polymerization. Rapidly mix and pour into the gel plate assembly to create a plug at the bottom of the gel. Allow ~2 min for polymerization.
4. Once the plug is polymerized, to the 50 mL of the bis–acrylamide gel solution, add 50 μL of TEMED and mix. Next, add 500 μL of the 10 % APS solution and mix to initialize polymerization. Pour into the gel plate assembly. Insert combs at desired depth and allow the gel to polymerize completely.
5. Take off clamps and carefully remove the combs and the bottom spacer. Move the assembly to an electrophoresis gel box. Add 0.5× TBE buffer to cover the top and bottom of the gel. Rinse the wells and the room left by the bottom spacer to remove air bubbles. Pre–run the gel at least for 30 minutes at 20 W (small plates) or 40 W (medium plates) before loading the samples (ideally, the gel has to be hot when touching for loading – this assures that the samples keep denatured during loading).

6. Turn off the power supply. Rinse the wells and load samples. Run the gel at 20 W (small plates) or 40 W (medium plates). The time of run may vary according to the length of the transcription product and fragments, and both dyes of the loading buffer can be used to estimate where the RNA products are in the gel. In general, a 40 minutes run is enough for constructs less than 110 nucleotides (see Note 9).
7. After the electrophoresis separation is done, turn off the power supply, remove the plate assembly and uncast the set of plates carefully. Cover one side of the gel with cellulose chromatography paper (this removes excess liquid in the gel, preventing excessive sample diffusion during storage). Cover the entire gel in plastic wrap and expose it to a phosphor screen. Place the phosphor image screen cassette in a refrigerator during exposure (up to an overnight exposure). If a longer exposure is needed, the gel should be dried and then exposed to the phosphor image screen (see Note 10).
8. Use a biomolecular imager system (like Typhoon series from GE Healthcare) to retrieve the gel image from the exposed phosphor image screen. Analyze the gel image by creating lane profiles of each lane and measuring band intensities using an appropriate software, such as ImageQuant (GE Healthcare) or Image J (Open source – NIH). For self–cleaving ribozymes, the single precursor RNA band (full length product) cleaves into two visible bands (5′ and 3′ products), which increases in intensity over time as the self–scission reaction is allowed to proceed.

Figure 4.1 provides an illustration of a resolved gel image. *In vitro* co–transcriptional cleavage kinetics of drz–Fprau–1 were performed in presence and absence of GlcN6P.

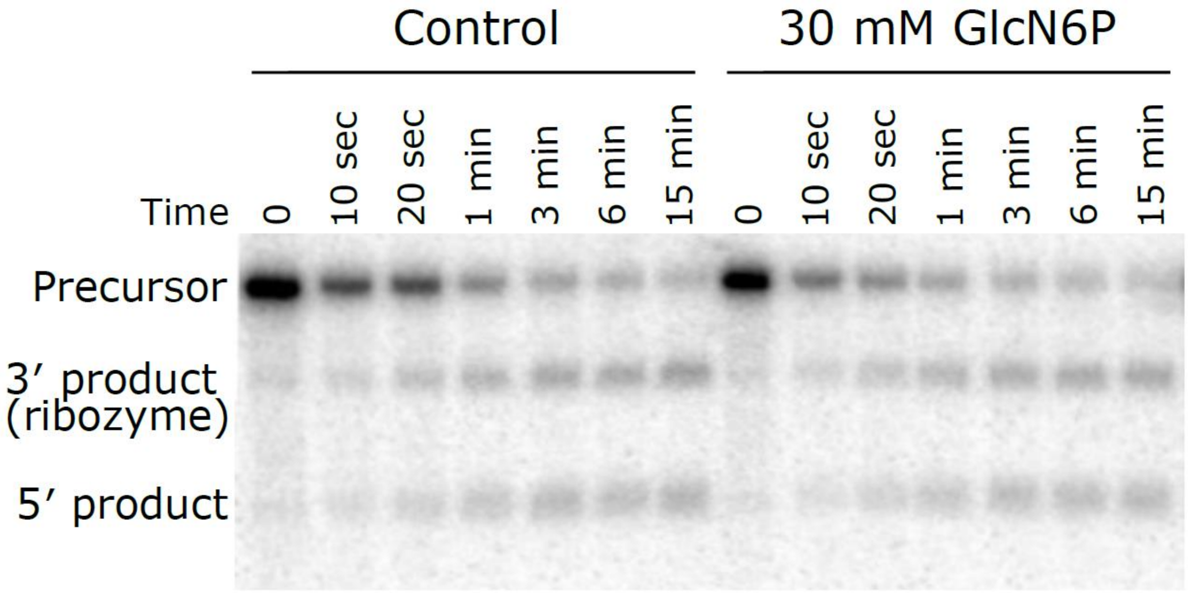

Control
30 mM GlcN6P
Time
0
10 sec
20 sec
1 min
3 min
6 min
15 min
0
10 sec
20 sec
1 min
3 min
6 min
15 min
Precursor
3′ product (ribozyme)
5′ product

Figure 4.1. Resolved denaturing (urea) PAGE gel of co–transcriptional self–scission of the drz–Fpra–1 ribozyme. *In vitro* co–transcriptional cleavage kinetics were performed in absence (control) and presence of 30 mM GlcN6P at a constant Mg^{2+} concentration (10 mM).

4.4.4 Data analysis

Several data–fitting software packages are currently available to perform data analysis. Herein, we explain how to analyze the data utilizing Microsoft Office Excel (MS Excel). We utilize linear least–squares optimization and the Solver module of MS Excel to fit the data. The retrieved band intensities of the self–scission experiment are used to solve for the observed rates of self–cleavage (k_{obs}) using linear regression analysis of a mono–exponential decay function, as shown below:

$$\text{Fraction intact} = A \times e^{-kt} + C$$

where A and C represent the relative fractions of the ribozyme population cleaving with a rate constant k and remaining uncleaved population, respectively.

1. In a MS Excel spreadsheet, horizontally, insert the band intensities values for each time–point, starting at column B. Leave column A for labeling the rows utilized: one for the single precursor RNA band, and two for the cleaved bands. Use row 1 to label and reference each lane from the gel.
2. In a convenient location on the spreadsheet, print arbitrary values corresponding to "A," "*k*," and "C" for the mono–exponential decay and residuals model. For this example, the function and the cells used are: =K3*EXP(−L3*B8) + M3. Where K3 represents the value for "A," L3 represents the value for "k", B8 represents the time, and M3 represents the value for "C". The dollar sign ensures that the value in that cell is used regardless of where the formula is pasted in the spreadsheet. Therefore, the time value will change as the formula is pasted across the time–point columns. This programming is important for utilizing the "Solver" tool.

3. On the 5th row (=sum of bands), calculate the sum of the bands (precursor + cleaved ones).
4. On the 6th row (=time), insert the time–point for each aliquot withdrawn (use the same units for all time–points).
5. On the 7th row (=fraction), calculate the fraction of precursor RNA band (full length) for each time–point by dividing the value of the precursor band by the sum of the precursor and cleaved bands at a single time–point (= precursor band/sum of bands). The formula can be pasted into subsequent cells for each time–point without retyping.
6. On the 8th row (=model), as previously introduced, calculate the model for each time–point where "t" is the time and "A," "*k*", and "C" are arbitrary values for a mono–exponential decay model: `=$K$3*EXP(−$L$3*B8) + $M$3`. The formula can be pasted into subsequent cells for each time–point without retyping, because time is the variable that will change.
7. On the 9th row (=square difference), calculate the square of the difference at each time–point between the model value and the fraction cleaved [= (fraction – model)^2].
8. Calculate the sum of all square differences by programming using the SUM function in a new cell `$J$11` (=sum of values of row 9).
9. Solver tool is an add–in which may need to be loaded into MS Excel via the Excel options and Add–ins tab menu. Solver is a tool for optimization and equation solving that finds the optimum value in one cell by adjusting the values in the cells

the user specifies. Therefore, this tool can be used to solve the regression of the data points to the model, resulting in the k_{obs} value for the particular ribozyme.

a. In MS Excel under the Data menu, select Solver in the Analyze subset.

b. Set the target cell to the sum of the square differences by selecting the cell containing that value ($J$11).

c. The goal is to minimize the value of that selected cell (sum of square differences), therefore, set “To:” to “min”.

d. “By changing variable cells:” should be set to the arbitrary model values (A, *k*, C; $K$3–$M$3 in our example).

e. Click “Solve.” Allow the process to complete. The model value cell $L$3 should now contain the k_{obs} value for the particular ribozyme (the value for *k*).

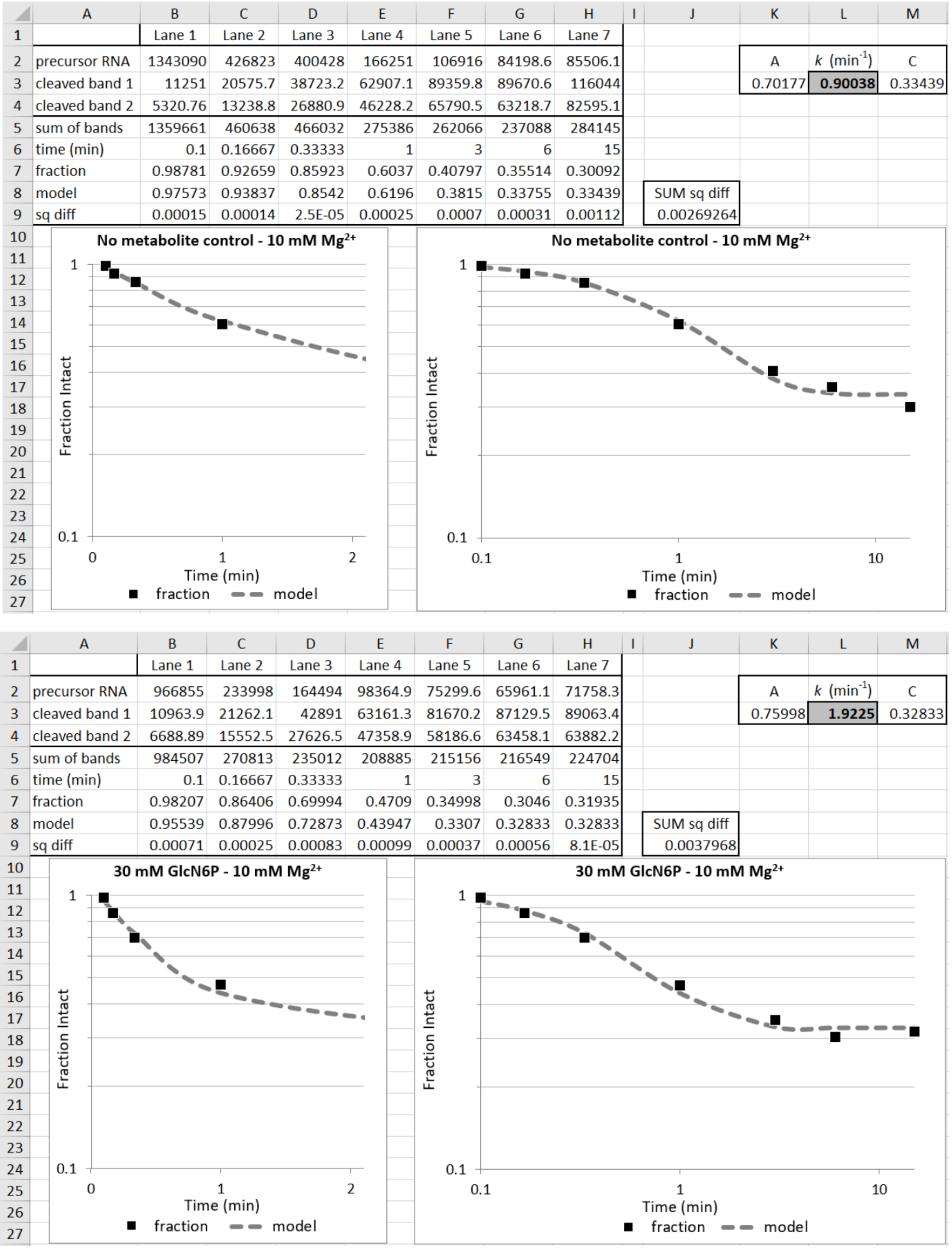
f. The “Solve” processing can be visualized if both the calculated model values as well as the fraction cleaved values are plotted vs. time.

10. Alternatively, Solver tool is also available as an add–in for Google docs spreadsheet.

Figure 4.2 shows the MS Excel spreadsheet used for the data analysis of both kinetics experiments presented on Figure 4.1.

A

B



A

	A	B	C	D	E	F	G	H	I	J	K	L	M
1		Lane 1	Lane 2	Lane 3	Lane 4	Lane 5	Lane 6	Lane 7					
2	precursor RNA	1343090	426823	400428	166251	106916	84198.6	85506.1			A	k (min^{-1})	C
3	cleaved band 1	11251	20575.7	38723.2	62907.1	89359.8	89670.6	116044			0.70177	**0.90038**	0.33439
4	cleaved band 2	5320.76	13238.8	26880.9	46228.2	65790.5	63218.7	82595.1					
5	sum of bands	1359661	460638	466032	275386	262066	237088	284145					
6	time (min)	0.1	0.16667	0.33333	1	3	6	15					
7	fraction	0.98781	0.92659	0.85923	0.6037	0.40797	0.35514	0.30092					
8	model	0.97573	0.93837	0.8542	0.6196	0.3815	0.33755	0.33439		SUM sq diff			
9	sq diff	0.00015	0.00014	2.5E-05	0.00025	0.0007	0.00031	0.00112		0.00269264			

10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
No metabolite control - 10 mM Mg2+
1
0.1
Fraction Intact
0
1
2
Time (min)
■ fraction
▬ ▬ model
No metabolite control - 10 mM Mg2+
1
0.1
Fraction Intact
0.1
1
10
Time (min)
■ fraction
▬ ▬ model

B

	A	B	C	D	E	F	G	H	I	J	K	L	M
1		Lane 1	Lane 2	Lane 3	Lane 4	Lane 5	Lane 6	Lane 7					
2	precursor RNA	966855	233998	164494	98364.9	75299.6	65961.1	71758.3			A	k (min^{-1})	C
3	cleaved band 1	10963.9	21262.1	42891	63161.3	81670.2	87129.5	89063.4			0.75998	**1.9225**	0.32833
4	cleaved band 2	6688.89	15552.5	27626.5	47358.9	58186.6	63458.1	63882.2					
5	sum of bands	984507	270813	235012	208885	215156	216549	224704					
6	time (min)	0.1	0.16667	0.33333	1	3	6	15					
7	fraction	0.98207	0.86406	0.69994	0.4709	0.34998	0.3046	0.31935					
8	model	0.95539	0.87996	0.72873	0.43947	0.3307	0.32833	0.32833		SUM sq diff			
9	sq diff	0.00071	0.00025	0.00083	0.00099	0.00037	0.00056	8.1E-05		0.0037968			

10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
30 mM GlcN6P - 10 mM Mg2+
1
0.1
Fraction Intact
0
1
2
Time (min)
■ fraction
▬ ▬ model
30 mM GlcN6P - 10 mM Mg2+
1
0.1
Fraction Intact
0.1
1
10
Time (min)
■ fraction
▬ ▬ model

Figure 4.2. Data analysis of co–transcriptional kinetics using MS Excel. Calculated k_{obs} is highlighted at cell L3 for both the no metabolite control (A) and the 30 mM GlcN6P

experiments (B). Solver tool was used to find the best parameters to fit a simple monoexponential decay (A – fraction reacted, k_{obs} – self–scission rate constant) with an unreacted fraction (C) equation. The two graphs in each panel show identical data presented on log–linear (early time–points for visual comparison of initial self–scission rate) and log–log scales. The data derived from the PAGE images are shown with squares. The best–fit models are shown as dashed lines.

4.5 Notes

1. Make all buffers in a 5x or 10x concentration to facilitate the preparation of the reactions. Store premade buffers at appropriate temperature.
2. The *in vitro* transcription buffer used usually cannot be the commercially available buffer supplied with the enzyme because of the high amounts of Mg^{2+}. However, the rNTPs can be stretched to concentrations matching the Mg^{2+} concentration in the buffer to prevent significant ribozyme activity during the initial transcription reaction.
3. Do not forget to insert the T7 promoter sequence upstream of the DNA templates. To increase transcription yield, consider the sequence immediately downstream of the T7 RNA promoter. The +1 to +3 promoter sequence with nucleotides GGG or GGC affords the highest yield ***(38)***.
4. For conditions requiring a consistent ionic strength, the buffer and metabolite(s) stocks may have to be pH–adjusted by the addition of KOH or HCl. Additionally, the contribution of ions from the metabolite(s) stocks have to be tracked and considered in the final reaction composition. For example, glcN6P is typically available as sodium salt. Thus, titration of this metabolite has to be accounted for when determining the ionic strength.
5. Increase the EDTA concentration accordingly if reaction to be quenched has more than 10 mM Mg^{2+} and/or another divalent ion.
6. You can reduce the *in vitro* transcription temperature (down to 16 °C) to decrease the self–cleavage reaction during transcription. Note that the transcription yield will be reduced as well.

7. Use a thermocycler with a heated lid to avoid condensation of water on the lid of the tube. Solvent evaporation can drastically change the concentration of solutes.
8. Polyacrylamide gel percentage may differ according to the length of products and fragments generated by the self–scission reaction. In general, lower percentages of gels are used to have greater separation of fragments with similar lengths.
9. It is important to design constructs in a way that will allow the two product bands to be distinguishable by size separation.
10. Exposure time of the gel to the phosphor screen depends on the amount of ^{32}P–labeled material on the gel. To increase the labeling yield of [α–^{32}P] rATP, you may reduce the concentration of the non–radioactive rATP to 0.1 mM or even 0.05 mM in the *in vitro* transcription mix. Adjust the Mg^{2+} concentration accordingly.
11. ^{32}P is a high energy β–emitter. Avoid exposure to the radiation and radioactive contamination. Wear proper PPE to minimize exposure to radiation. Dispose of radioactive waste in accordance with the rules and regulations.

4.6 References

1. Prody GA, Bakos JT, Buzayan JM, et al. Autolytic processing of dimeric plant virus satellite RNA. *Science* **231**, 1577–1580 (1986).
2. Hutchins CJ, Rathjen PD, Forster AC, et al. Self–cleavage of plus and minus RNA transcripts of avocado sunblotch viroid. *Nucleic Acids Res* **14**, 3627–3640 (1986).
3. Buzayan JM, Gerlach WL, and Bruening G. Non–enzymatic cleavage and ligation of RNAs complementary to a plant virus satellite RNA. *Nature* **323**, 349–353 (1986).
4. Sharmeen L, Kuo MY, Dinter–Gottlieb G, et al. Antigenomic RNA of human hepatitis delta virus can undergo self–cleavage. *J Virol* **62**, 2674–2679 (1988).
5. Saville BJ and Collins RA. A site–specific self–cleavage reaction performed by a novel RNA in neurospora mitochondria. *Cell* **61**, 685–696 (1990).
6. Jimenez RM, Polanco JA, and Lupták A. Chemistry and biology of self–cleaving ribozymes. *Trends Biochem Sci* **40**, 648–661 (2015).

7. Wilson TJ, Liu Y, and Lilley DMJ. Ribozymes and the mechanisms that underlie RNA catalysis. *Front Chem Sci Eng* **10**, 178–185 (2016).

8. Ren A, Micura R, and Patel DJ. Structure–based mechanistic insights into catalysis by small self–cleaving ribozymes. *Curr Opin Chem Biol* **41**, 71–83 (2017).

9. Seith DD, Bingaman JL, Veenis AJ, et al. Elucidation of catalytic strategies of small nucleolytic ribozymes from comparative analysis of active sites. *ACS Catal* **8**, 314–327 (2018).

10. Fedor MJ. Comparative enzymology and structural biology of RNA self–cleavage. *Annu Rev Biophys* **38**, 271–299 (2009).

11. Winkler WC, Nahvi A, Roth A, et al. Control of gene expression by a natural metabolite–responsive ribozyme. *Nature* **428,** 281–286 (2004).

12. Wu HN, Lin YJ, Lin FP, et al. Human hepatitis delta virus RNA subfragments contain an autocleavage activity. *Proc Natl Acad Sci* **86,** 1831–1835 (1989).

13. Roth A, Weinberg Z, Chen AGY, et al. A widespread self–cleaving ribozyme class is revealed by bioinformatics. *Nat Chem Biol* **10**, 56–60 (2014).

14. Weinberg Z, Kim PB, Chen TH, et al. New classes of self–cleaving ribozymes revealed by comparative genomics analysis. *Nat Chem Biol* **11**, 606–610 (2015).

15. Webb CHT, Riccitelli NJ, Ruminski DJ, et al. Widespread occurrence of self–cleaving ribozymes. *Science* **326**, 953 (2009).

16. Hammann C, Luptak A, Perreault J, et al. The ubiquitous hammerhead ribozyme. *RNA* **18**, 871–885 (2012).

17. Martick M, Horan LH, Noller HF, et al. A discontinuous hammerhead ribozyme embedded in a mammalian messenger RNA. *Nature* **454**, 899 (2008).

18. Vazquez–Tello A, Rojas AA, Paquin B, et al. Hammerhead–mediated processing of satellite pDo500 family transcripts from *Dolichopoda* cave crickets. *Nucleic Acids Res* **28**, 4037–4043 (2000).

19. Ruminski DJ, Webb C–HT, Riccitelli NJ, et al. Processing and translation initiation of non–long terminal repeat retrotransposons by hepatitis delta virus (HDV)–like self–cleaving ribozymes. *J Biol Chem* **286**, 41286–41295 (2011).

20. Eickbush DG and Eickbush TH. R2 retrotransposons encode a self–cleaving ribozyme for processing from an rRNA cotranscript. *Mol Cell Biol* **30**, 3142–3150 (2010).

21. Sánchez–Luque FJ, López MC, Macias F, et al. Identification of an hepatitis delta virus–like ribozyme at the mRNA 5′–end of the L1Tc retrotransposon from *Trypanosoma cruzi*. *Nucleic Acids Res* **39**, 8065–8077 (2011).

22. Cervera A and la Peña M De. Eukaryotic penelope–like retroelements encode hammerhead ribozyme motifs. *Mol Biol Evol* **31**, 2941–2947 (2014).

23. Forster AC and Symons RH. Self–cleavage of plus and minus RNAs of a virusoid and a structural model for the active sites. *Cell* **49**, 211–220 (1987).

24. Epstein LM and Gall JG. Transcripts of Newt satellite DNA self–cleave *in vitro*. *Cold Spring Harb Symp Quant Biol* **52**, 261–265 (1987).

25. Ferbeyre G, Smith JM, and Cedergren R. *Schistosome satellite* DNA encodes active hammerhead ribozymes. *Mol Cell Biol* **18**, 3880–3888 (1998).

26. Passalacqua LFM, Jimenez RM, Fong JY, et al. Allosteric modulation of the Faecalibacterium prausnitzii hepatitis delta virus–like ribozyme by glucosamine 6–phosphate: the substrate of the adjacent gene product. *Biochemistry* **56**, 6006–6014 (2017).

27. Salehi–Ashtiani K, Lupták A, Litovchick A, et al. A genomewide search for ribozymes reveals an HDV–like sequence in the human CPEB3 gene. *Science* **313**, 1788–1792 (2006).

28. la Peña M de and García-Robles I. Intronic hammerhead ribozymes are ultraconserved in the human genome. *EMBO Rep* **11**, 711–716 (2010).

29. Lupták A, Ferré–D'Amaré AR, Zhou K, et al. Direct pKa measurement of the active–site cytosine in a genomic hepatitis delta virus ribozyme. *J Am Chem Soc* **123**, 8447–8452 (2001).

30. Pan T, Artsimovitch I, Fang X, et al. Folding of a large ribozyme during transcription and the effect of the elongation factor NusA. *Proc Natl Acad Sci* **96**, 9545–9550 (1999).

31. Chadalavada DM, Gratton EA, and Bevilacqua PC. The human HDV–like CPEB3 ribozyme is intrinsically fast–reacting. *Biochemistry* **49**, 5321–5330 (2010).

32. Diegelman–Parente A and Bevilacqua PC. A mechanistic framework for co–transcriptional folding of the HDV genomic ribozyme in the presence of downstream sequence. *J Mol Biol* **324**, 1–16 (2002)

33. Long DM and Uhlenbeck OC. Kinetic characterization of intramolecular and intermolecular hammerhead RNAs with stem II deletions. *Proc Natl Acad Sci* **91**, 6977–81 (1994).

34. Carothers JM, Goler JA, Juminaga D, et al. Model–driven engineering of RNA devices to quantitatively program gene expression. *Science* **334**, 1716–1719 (2011).

35. Cech TR. Self–splicing of group I introns. *Annu Rev Biochem* **59**, 543–568 (1990).

36. Zhao C and Pyle AM. Structural insights into the mechanism of group II intron splicing. *Trends Biochem Sci* **42**,470–482 (2017).

37. Mengin–lecreulx D and Heijenoort J Van. Characterization of the essential gene *glmM* encoding phosphoglucosamine mutase in *Escherichia coli*. *Mol Biol* **271**,32–39 (1996).

38. Imburgio D, Rong M, Ma K, et al. Studies of promoter recognition and start site selection by T7 RNA polymerase using a comprehensive collection of promoter variants. *Biochemistry* **39**, 10419–10430 (2000).

Chapter 5: Single Pass Transcription using T7 RNA polymerase

Luiz F. M. Passalacqua, Armine I. Dingilian, Andrej Lupták. Single Pass Transcription using T7 RNA polymerase. *Manuscript in preparation.*

Contribution Statement

L F. M. P. and A. L. designed the study; **L F. M. P.** and A.I.D. performed the experiments; all authors analyzed the data; and **L F. M. P.** and A. L. wrote the manuscript.

5.1 Introduction

RNA molecules can be conveniently synthesized by *in vitro* transcription. This procedure is simple and routinely used to generate RNA molecules that range from few to several thousands of nucleotides[1]. *In vitro* transcripts can be used in studies that include the RNA molecule itself (e.g. biochemical studies of functional RNAs) and downstream events that involve the transcript (e.g. translational studies)[1,2]. T7 RNA polymerase (T7 RNAP) is one of the most common and robust enzymes used for *in vitro* transcription and it has been extensively studied over the last 50 years[2,3]. T7 RNAP was originally derived from the T7 bacteriophage and is a 98 kDa single–subunit enzyme (in contrast to multi–subunit enzymes in bacteria and eukaryotes) that does not need any additional factors to perform transcription[2,3]. Additionally, it presents a high specificity towards the T7 promoter sequence[1–6].

The promoter sequence of the T7 RNAP consists of a 23–base pair (bp) region that extends from the position –17 to the position +6, where +1 is the first nucleotide to be transcribed[4,7]. It can be separated in two domains, an upstream binding region that extends from the position –17 to –5, and a downstream initiation region, from –4 to +6[4,5,7–10]. The consensus sequence found in T7 bacteriophage promoters in nature is 5′–TAATACGACTCACTATAGGGAGA, and the underlined nucleotides represent the first nucleotides to be transcribed[4]. While some variations are not tolerated at all (–7, –8, and –9), some others are accepted at an efficiency penalty, reducing the yield of transcription[4]. At the initiation region, substitutions are in general well tolerated in the first nucleotides to be transcribed, with the exception of substitutions at the position +1, in which no variation is tolerable, and the position +2, that reduces the transcription yield by half[4]. Thus, for a message of interest to be efficiently transcribed, the first nucleotide to be transcribed in the initiation region should remain unchanged; however, optimal results are obtained if the entire binding region and the first six nucleotides (–4 to +2) of the initiation region are maintained[4].

The T7 RNAP system can be conveniently employed to generate up to mg amounts of specific transcripts and is widely used in molecular biology[1,6,11]. While the T7 RNAP system has been invaluable for *in vitro* studies of RNA, there are many situations in which large quantities and continuous synthesis of RNA are not desired. For example, an easy and simple methodology using the T7 RNAP system is yet to be described for co–transcriptional studies with constant number of RNA molecules throughout the experiment. In order to use a constant amount of RNA, current studies require T7 transcribed RNAs to be purified under denaturing conditions, a step that can lead to misfolded RNAs[12]. This is due to the fact that *in vitro* refolding causes RNA molecules to lose their directional order of folding (5′ to 3′) which is likely to result in the co–existence of different folding states [13,14]. Less harsh conditions of purification, such as non–denaturing chromatography have been utilized[15], but with an increased cost and labor involved. Alternatively, if T7 transcribed RNAs are not purified, the kinetics of the continuous transcription reaction should to be accounted for, which leads to increasingly complex calculations that require the addition of a second kinetic element[16,17]. One additional alternative would be to adapt the methodology presented in chapter 4, where the transcription reaction is terminated by a 25–fold dilution, in desired conditions, preventing any new RNA synthesis and void the need to account for the kinetics of transcription. This method however, may not be suitable for experiments that are time–sensitive or volume–limited.

Thus, considering the need for transcription assays that allow for the generation of quantitative amounts of RNA and enable temporal control, we developed an assay that allows for a single pass transcription which yields single transcript per template molecule of DNA. We hypothesized that if we stall the T7 RNAP downstream of the promoter region, we can use a specific restriction enzyme to cleave the promoter region and inhibit the binding of another T7 RNAP to the DNA. RNA polymerases can be stalled by depleting a specific nucleotide from the reaction mixture[18–20]. In this case, guanosine can't be the nucleotide to be depleted because the T7 RNAP requires Gs at the initiation region to start transcription[4]. We chose to have the T7 RNAP stalled at the position +27 by having the first cytosine to be incorporated at this position, and starting transcription using only GTP, ATP, and UTP. The position of the stalling was chosen based on biochemical and structural studies of the T7 RNAP transcription bubble, reported to be of a stretch of 17 nucleotides, with an overall interaction with about 20 nucleotides of the DNA template[21–26]. Once the T7 RNAP is stalled, the restriction enzyme is added, cleaving the promoter region. Next, we add CTP and allow the full transcript to be synthesized. **Figure 5.1** illustrates the proposed scheme. We wanted to stall the T7 RNAP away from the promoter region, to give enough room to the restriction enzyme to cleave the promoter region without affecting the T7 RNAP complex. Additionally, T7 RNAP complexes stalled within the first 8–10 bases are much less stable (during the initiation complex), and complexes stalled at a further position, during the elongation complex, are more stable and tend to have less abortive events even if another polymerase enzyme bumps into the first stalled enzyme (the second enzyme would be in initiation complex and not stable enough to bump and remove the T7 RNAP downstream)[23,27,28].

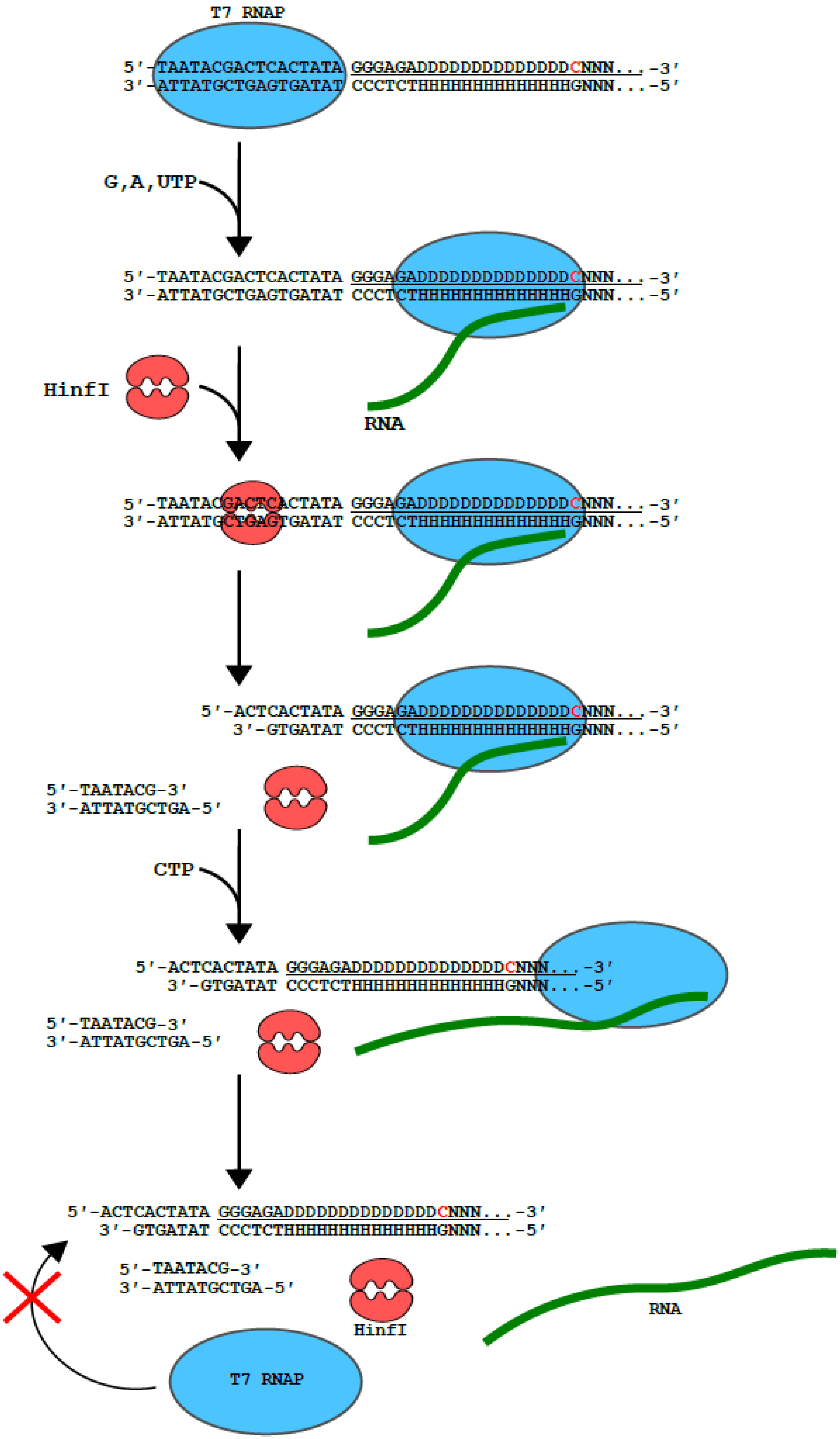

T7 RNAP
5'-TAATACGACTCACTATA GGGAGADDDDDDDDDDDDDDDCNNN...-3'
3'-ATTATGCTGAGTGATAT CCCTCTHHHHHHHHHHHHHHHGNNN...-5'
G,A,UTP
HinfI
RNA
CTP
5'-TAATACG-3'
3'-ATTATGCTGA-5'
5'-ACTCACTATA
3'-GTGATAT
HinfI
RNA
T7 RNAP

Figure 5.1. Illustration scheme of proposed single pass transcription. Transcribed region is underlined, and first cytosine is in red. D = A or G or T; H = A or C or T (not G).

The T7 RNAP promoter can be target by four different restriction enzymes that recognize at least 4 nucleotides stretches and are commercially available (*MlyI, PleI, SfcI,* and *HinfI*)[29]. The restriction enzyme of choice is *HinfI*. It is a type II restriction endonuclease that recognizes 5′– GANTC, where N is any base, and cleaves between guanine and adenine residues on both strands of the DNA duplex (positions –10 and –11 of the promoter sequence)[29–31]. The type II restriction endonucleases family is found in bacteria and archaea, and serves to protect the organism against invading foreign DNA[32]. The enzyme has a monomeric molecular weight of 31 kDa and is active in a dimer form[31]. Additionally, *HinfI* depends on Mg^{2+} ions to be functional and promote catalysis[32]. The structure of *HinfI* is yet to be solved. Despite of low sequence homology among type II restriction endonucleases, the enzymes that have had their structures investigated present structural similarities[32,33]. Studies of DNase footprinting of distinct type II restriction enzymes have shown that they protect 13 to 21 bases of the targeted DNA[34], a crucial information that allows us to better estimate a functional stalling position for the T7 RNAP, minimizing the risk of steric hindrance between the two enzymes.

In this study, we used the HDV–like self–cleaving ribozyme drz–Fpra–2[35] (previously described on chapter 3) as model molecule to be transcribed. Self–cleaving ribozymes are catalytic RNA molecules that promote a site–specific self–scission reaction[36]. The use of a self–cleaving ribozyme is a good choice for model investigations because we can track both the synthesis of the full–length construct and the formation of the cleaved product. In this study, we were able to investigate the cleavage of the promoter region by *HinfI* and the single pass transcription of the self–cleaving ribozyme drz–Fpra–2.

5.2 Results

T7 RNAP promoter cleavage by restriction enzyme. To study the cleavage of the T7 promoter region in the DNA to be transcribed, we performed kinetics studies of the cleavage reaction by the restriction enzyme *HinfI* under different conditions. The concentration of DNA template used for these experiments was 100 nM. We first investigated the cleavage reaction in just the reaction mixture (reaction buffer + NTPs) without the T7 RNAP (**Figure 5.2a**). After 1 minute of reaction, ~ 40% of the DNA was cleaved, increasing to ~ 85% after 3 minutes of reaction. The cleavage reaction was demonstrated to be fast with no full–length construct detected at 6 minutes of incubation. Next, we performed the same reaction with the addition of T7 RNAP (to a final concentration of 100 nM) in a binding competition for the promoter region with *HinfI*. To support the binding of the T7 RNAP to the promoter region prior to the restriction enzyme, we waited 3 minutes to add *HinfI* to the reaction. We also avoided the addition of NTPs to ensure that the T7 RNAP would not start transcription and leave the promoter region. To our surprise, the cleavage reaction occurred faster in the presence of T7 RNAP. After 1 minute of incubation about ~ 60% of the DNA was cleaved, and a similar value at 3 minutes, where ~85% of the full–length was cleaved (**Figure 5.2b**). As previously observed, no full–length DNA was detected after 6 minutes reaction. This result also indicates that *HinfI* doesn't need to be in excess for the cleavage reaction to be completed.

Lastly, we performed the same experiment in presence of GTP, ATP, and UTP, which led the T7 RNAP to stall downstream of the promoter region and possibly facilitating DNA cleavage by *HinfI* when compared to the previous competing experiment. In the stalling experiment, the results showed that the cleavage pattern is faster than the two previous experiments, with no detectable full–length DNA after 3 minutes of reaction. This observation suggests that the T7 RNAP stalled downstream of the promoter region facilitates the action of *HinfI* in the promoter region. Additionally, the results indicate that the T7 RNAP was not bound to the promoter region anymore when *HinfI* was added to the reaction. For all these experiments, *HinfI* was used at 1% (v/v), corresponding to a concentration of about 65 nM, lower than the DNA and T7 RNAP concentration, indicating that the enzyme is robust enough to be used in low concentration. Therefore, we decided to use this concentration for the entire study. The results of the experiments of DNA cleavage by *HinfI* performed here indicate that the cleavage of the T7 promoter is quickly achievable and is not delayed by binding competition with T7 RNAP.

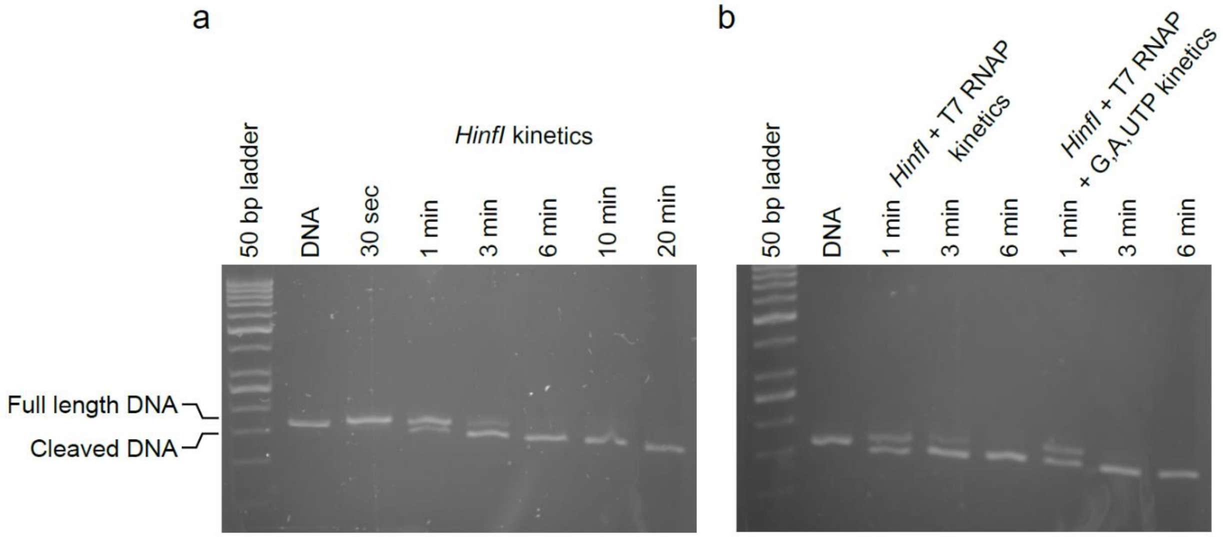

a
b
Hinfl kinetics
50 bp ladder
DNA
30 sec
1 min
3 min
6 min
10 min
20 min
Full length DNA
Cleaved DNA
50 bp ladder
DNA
Hinfl + T7 RNAP kinetics
1 min
3 min
6 min
Hinfl + T7 RNAP + G,A,UTP kinetics
1 min
3 min
6 min

Figure 5.2. **Cleavage of the promoter region by *HinfI.*** (a) Kinetics of the promoter cleavage by *HinfI* in reaction mixture + NTPs. (b) Kinetics of the promoter cleavage by *HinfI* + T7 RNAP in the absence of NTPs (left) and in the presence of G,A,UTP (right). The DNA template of the self–cleaving ribozyme drz–Fpra–2 was used for this experiment.

Optimization of *HinfI* incubation time. The previous experiment performed showed that 6 minutes is sufficient to cleave the promoter region to an undetectable level using agarose gel electrophoresis and SYBR staining. To further investigate the influence of promoter cleavage on a single transcription event, we performed several experiments that varied the incubation time with *HinfI* from 10 minutes to 30 minutes, prior to the addition of CTP which allows the stalled transcription to finish. Our results show that there is no distinguishable difference in transcriptional readout between varying incubation times with *HinfI*. This indicates that a 10 minutes incubation is enough time to ensure that the promoter region is fully cleaved and transcription occurs with no apparent disturbance (**Figure 5.3**). It has been suggested that the T7 RNAP can form a stable stalled complex in elongation mode[27], while others suggest that the T7 RNAP can abort the transcript[37]. Thus, we opted for the shortest time of incubation (10 minutes) to continue with our single pass transcription studies.

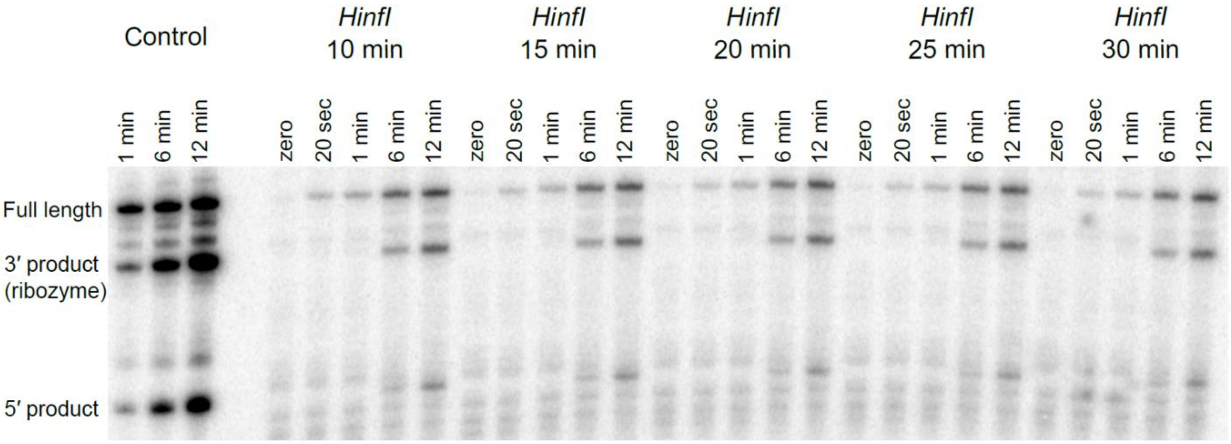

Control
HinfI 10 min
HinfI 15 min
HinfI 20 min
HinfI 25 min
HinfI 30 min
1 min
6 min
12 min
zero
20 sec
1 min
6 min
12 min
Full length
3′ product (ribozyme)
5′ product

Figure 5.3. Optimization of HinfI incubation time. Resolved PAGE of drz–Fpra–2 transcription under different incubation times of *HinfI*. The control experiment is performed in the absence of *HinfI*.

Stoichiometry of DNA template, T7 RNAP, and NTPs. To decrease the chances of multiple T7 RNAPs on the same DNA template, we also investigated the ratio of DNA template molecules to T7 RNAP molecules. We noticed that when the concentration of the T7 RNAP was higher than the DNA template, the reaction generated more aborted and background transcripts (data not shown). Thus, we decreased the concentration of T7 RNAP to increase the probability of having a single T7 RNAP per template. We optimized the ratio of DNA template to T7 RNAP to be around 1:0.75. We also decreased the concentration of NTPs, which resulted in less background transcription, possibly due to reduction of misincorporation during the stalled phase of the process. After several rounds of optimization, we settled the NTP concentrations at 200 µM of GTP, 100 µM of ATP and UTP, and 50 µM of CTP.

Kinetics of self–cleaving ribozyme using single pulse transcription. In order to confirm that our single pass transcription system works as proposed, we analyzed the co–transcriptional self–cleavage reaction of the ribozyme drz–Fpra–2 and compared it to a control experiment without the addition of *HinfI* to promote the cleavage of the promoter region. As expected, the control experiment showed that without *HinfI*, drz–Fpra–2 transcripts accumulate overtime and plateau after 6 minutes of reaction, which could be due to the lack of NTPs. The experiment using the proposed single pass transcription, shows that the synthesis of RNA plateaus after 1 minute of reaction (considering the addition of CTP as the t=0 for the reaction) and proceeds like a single pass transcription. Because our model is a self–cleaving ribozyme, we were able to follow the cleavage of the full–length transcript and the accumulation of the cleaved product, making it easier to verify if new transcripts were still being synthesized. **Figure 5.4** shows the comparison between the two experiments. The experimental results are in agreement with our expected results and validate the herein proposed method.

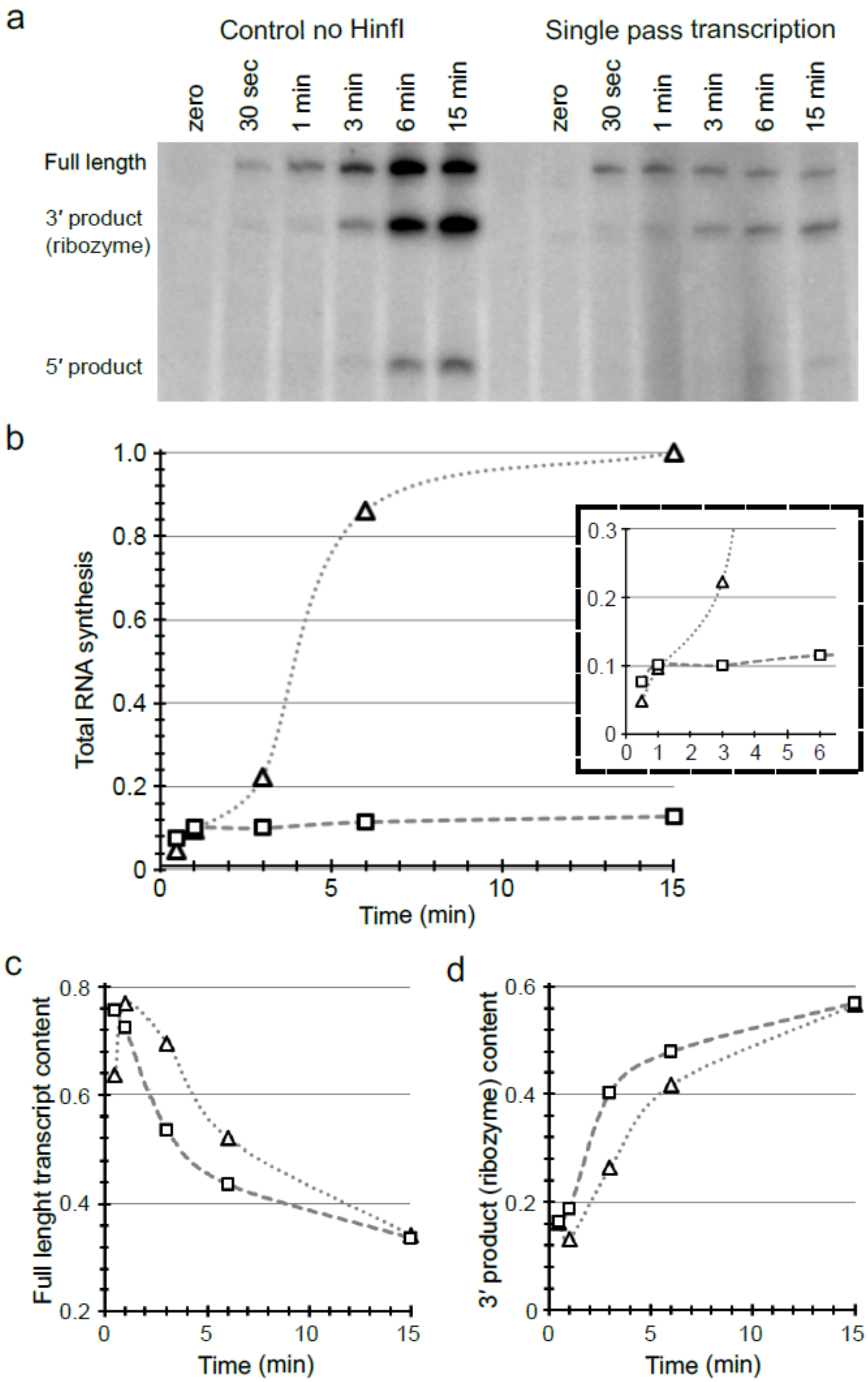


Figure 5.4. Transcription analysis of drz–Fpra–2. (a) PAGE gel of the transcription reaction of drz–Fpra–2 in the absence (control) and in the presence of HinfI (single pass transcription). (For b, c, and d: triangle = control no *HinfI*, square = single pass transcription). (b) Comparison of total RNA synthesis over time. Early time points are

shown in inset. The data is normalized to the highest intensity of bands (full length + products). (c) Full length transcript content over time and (d) 3′ product (cleaved ribozyme) content over time, showing the difference of self–cleaving kinetics between the two experiments.

5.3 Discussion and Conclusion

In this study, we hypothesized that it would be possible to create a system using the T7 RNAP that allows a single transcript to be synthesized per molecule of DNA template. Using both the T7 RNAP and a specific restriction enzyme that targets the T7 promoter region, *HinfI*, our results suggest that we were able to achieve this goal. We started by studying the cleavage of the promoter region by the restriction enzyme *HinfI* in different conditions – in the absence of the T7 RNAP, in the presence of the T7 RNAP competing for the binding at the promoter region, and in the presence of the T7 RNAP stalled downstream of the promoter region in the position +27. For all conditions, the restriction enzyme was able to fully cleave the promoter region, indicating that the site–specific cleavage reaction is quickly and reliably completed in normal transcription conditions.

Next, we carried out an investigation to verify the optimal incubation time of the restriction enzyme prior to the transcription reaction. Considering the previous cleavage data using agarose gels and SYBR staining, we decided to analyze transcription results after incubating the reaction from 10 to 30 minutes before the addition of the fourth nucleotide, CTP. The lowest incubation time we tested showed no detectable difference to any other incubation time, suggesting that 10 minutes is enough time for the *HinfI* enzyme to cleave the promoter region of the DNA template. The combination of these results and the results of the DNA cleavage are in agreement with *HinfI* successfully cleaving the promoter region in the presence of transcription conditions.

We also studied the stoichiometry between the DNA template, T7 RNAP, and *HinfI.* The data collected showed that *HinfI* is robust enough to fully cleave 0.1 µM DNA at 65 nM, in the presence or absence of T7 RNAP. Furthermore, cleavage by *HinfI was* independent of *T7 RNAP* competing for binding at the promoter site or stalled downstream of the promoter. Our tests also indicate that the DNA template should be in higher concentrations than the T7 RNAP, to avoid accumulation of more than one enzyme per DNA molecule. This effect may cause transcription of more than one transcript per DNA molecule and also may induce transcription termination of the downstream stalled transcript[27]. The DNA template: T7 RNAP concentration ratio that we suggest is 1: 0.75.

Lastly, we verified if the single pass transcription experiment was successful. We analyzed the self–cleaving reaction of the ribozyme drz–Fpra–2 under two different conditions. First, we performed a control experiment without the addition of *HinfI*, that allowed transcription to occur continuously. Second, we utilized the herein proposed system, which makes use of *HinfI* to cleave the promoter region and allows for transcription of a single RNA molecule per DNA template. For the control reaction, without addition of the restriction enzyme, we verified that RNA synthesis occurs continuously during the reaction up to 6 minutes and subsequently plateaus, most likely due to residual amounts of NTPs still available at that point. In this case, the analysis of such self–cleavage reaction to obtain a self–scission rate constant should include not only the cleavage kinetics, but also the kinetics of the RNA synthesis, bringing a second kinetic element to the calculations.

Contrary to the control experiment, when we used *HinfI* to cleave the promoter region before the addition of CTP, we observed a plateau in RNA synthesis 1 minute after addition of the fourth nucleotide, followed by a decrease of the full–length band intensity over time. This suggests that no more transcript is made or that the background transcription is slower than the cleavage reaction. Interestingly, after 30 seconds of transcription reaction, the intensity observed for the full–length band was higher for the single pass experiment when compared to the control experiment. This finding may be explained by previous findings that showed that the initiation of the transcription has a lag phase that can take between 1 and 5 seconds[37–40], and about 22% of the transcripts are aborted[25,37,41], two events that supposedly happen before the addition of CTP for the single pass experiment, where it's suggested that the T7 RNAP was stalled at an elongation complex and ready to finish the synthesis of the transcript. When comparing the levels of the full length transcript and the formation of the 3′ ribozyme product, we can verify a different kinetic pattern between the two experiments, suggesting how different the observations of such experiments may be depending on the choice of methodology. Additionally, the accumulation of the ribozyme cleaved product over time indicates that the self–scission reaction was occurring during the experiment without any apparent influence of *HinfI*. Altogether, the results obtained in this study suggests that it is possible to achieve a single pulse transcription after stalling the T7 RNAP downstream of the promoter region and subsequent cleavage of the promoter by the restriction enzyme *HinfI*.

It is important to note that the system proposed here also has disadvantages, particularly if the DNA to be transcribed has another *HinfI* restriction site. If this is the case, duplex DNA should be avoided and a single strand antisense DNA template with sense complementary region that comprises solely the promoter region should be used, as it has been shown that once in elongation phase, a duplex DNA is not required by the T7 RNAP to perform transcription[6,42]. Additionally, experiments carried after the single pass transcription should be able to tolerate the transcription buffer and present enzymes (T7 RNAP and *HinfI*).

To the best of our knowledge, this is the first time that such a method is proposed using the T7 RNAP system. The simplicity of this method makes it a valuable alternative to purification or dilution of the reaction to halt transcription. Additionally, this approach allows the addition of more components (such as ligands for example) to the reaction after the single pulse transcription is performed without the concern of it affecting an eventual continuous transcription reaction. The method is not only suitable for studying functional RNAs, such as ribozymes and riboswitches, but can also be adapted to study phenomena such as RNA degradation and translation. Such studies would benefit from the ability to synthesize known quantities of RNA co–transcriptionally. Furthermore, the ability to directly limit the RNA synthesized in an experiment without denaturing the transcripts, likely yields conditions that are more relevant in a native environment, such as inside cells. Future experiments could test this system with different models and contexts, such as riboswitches and translational studies. Moreover, the quantitative comparison between transcripts concentration and the initial DNA and T7 RNAP concentration could provide a better understanding of the single pass transcription, revealing to what extent the synthesized transcripts amounts correlate to the initial quantities of both the DNA and T7 RNAP molecules. In summary, the herein developed method has the potential to be applicable to the study of several classes of RNAs and relevant downstream phenomena, such as mRNA translation.

5.4 Materials and Methods

T7 RNAP promoter cleavage by *HinfI*. 40 µL reactions were preparing by adding 8 µL 5× reaction buffer (1× = 40 mM Tris–HCl, 15 mM Mg(OAc)2, 50 mM KOAc, 2 mM Spermidine, pH 7.5), 4 µL of 100 mM DTT (10 mM final), and drz–Fpra–2 DNA template (final concentration = 100 nM). NTPs were added accordingly to the experiment to a final concentration of 250 µM each (all four NTPs for the cleavage analysis without T7 RNAP; no NTPs for the competing for binding at the promoter region experiment; and G,A,UTP for the cleavage analysis for downstream stalled T7 RNAP). *HinfI* was added at final concentration of 1% (v/v) (~ 65 nM). T7 RNAP, when added, at a final concentration of 100 nM. Water was used to complete reaction volume when needed. Time points were collected and reaction terminated with equal volume of stop buffer containing 25 mM EDTA, pH 7.4, with xylene cyanol and bromophenol blue loading dyes. Results were resolving with 2.5% agarose gel and SYBR Gold staining according to manufacture protocol (Invitrogen). The bands intensities were analyzed by creating lane profiles for each lane using ImageJ[43].

Optimization of HinfI incubation time*.* Similar to above mentioned, reactions were incubated with *HinfI* for different periods of time prior to the addition of CTP to allow transcription to finish. T7 RNAP were incubated for 5 minutes before the addition of *HinfI*. For this experiment, final concentrations were: DNA template = 100 nM; T7 RNAP = 200 nM; *HinfI* = 150 nM; GTP = 200 µM; A,U,CTP = 100 µM. 1 µL of [α–^{32}P]ATP (250 µCi/ml) (Perkin Elmer) was used to label transcripts and follow the reactions. Aliquots were collected at the indicated times and terminated by adding equal volume of stop buffer containing 25 mM EDTA, 5 mM Tris pH 7.4, 7 M urea, with xylene cyanol and bromophenol blue loading dyes. Results were resolved on a 10% denaturing PAGE gel. The gel was exposed to phosphorimage screen and analyzed using Typhoon phosphorimager and ImageQuant software (GE Healthcare). The band intensities were analyzed by creating line profiles of each lane using ImageQuant.

Stoichiometry optimization of DNA, T7 RNAP, and *HinfI*. Several reactions were performed as described above. Ranging concentrations of DNA template, T7 RNAP, and *HinfI* were tested to optimize the reaction conditions. Additionally, various NTPs concentrations were also tested. Results suggest that a ratio of DNA: T7 RNAP of 1: 0.75 should be maintained. *HinfI* showed to be a robust enzyme with 1% (v/v) being enough to cleave up to 0.5 µM of DNA in 10 minutes (data now shown). Optimal NTPs concentration for reaction is: GTP = 200 µM; A,UTP = 100 µM; CTP = 50 µM. Either $AT^{32}P$ or $CT^{32}P$ can be used to track the reaction, but $CT^{32}P$ provides a cleaner gel as aborted transcripts are not radiolabeled.

***In vitro* co–transcriptional cleavage analysis.** 40 µL reactions were preparing by adding 8 µL 5× reaction buffer (1× = 40 mM Tris–HCl, 15 mM Mg(OAc)2, 50 mM KOAc, 2 mM Spermidine, pH 7.5), 4 µL of 100 mM DTT stock (10 mM final), 2 µL of 100 mM $MgCl_2$ stock (5 mM final), and drz–Fpra–2 DNA template (final concentration = 200 nM). Because the self–cleaving reaction requires Mg^{2+} to catalysis[44–47], additional ions were added to allow self–cleavage reaction to occur without problems of eventual Mg^{2+} scarcity. NTPs were added accordingly to the optimized condition (GTP = 200 µM; A,UTP = 100 µM; CTP = 50 µM). 2 µL of [α–^{32}P]CTP (250 µCi/ml) (Perkin Elmer) was used to label and track the reactions. T7 RNAP was added to a final concentration of 150 nM. *HinfI* was maintained at the optimal final concentration of 1% (v/v) (~ 65 nM) and added to the reaction 3 minutes after the addition of the T7 RNAP. Water was used to complete reaction volume when needed. CTP addition was used as t=0 and occurred after a 10 minutes incubation time of *HinfI.* Time points were collected and reaction terminated with equal volume of stop buffer containing 25 mM EDTA, 5 mM Tris pH 7.4, 7 M urea, with xylene cyanol and bromophenol blue loading dyes. Results were resolved in a 12% denaturing PAGE gel. The gel was exposed to phosphorimage screens and analyzed using Typhoon phosphorimager and ImageQuant software (GE Healthcare). The band intensities were analyzed by creating line profiles of each lane using ImageQuant, exporting the data to Microsoft Excel.

DNA sequences used in this project. Full length drz-Fpra-2 DNA sequence with T7 promoter (underlined), and leader sequence is: 5'–TAATACGACTCACTATAGGGAGATATAAGGTTGTGGGAAGGTTCTCACAGCTTTGCCGCTCGAAACTTTGCACACCTCTACGCGGTGGGTGGCAGGCAACACATGAGAATGTGGCTGAGATGCAATATTGTC–3'. The first cytosine is boxed. The forward primer sequence used in this study is 5'–TAATACGACTCACTATAGG–3', and the reverse sequence used in this study 5'–GACAATATTGCATCTCAGCCACAT–3'.

5.5 References

1. Beckert, B. & Masquida, B. Synthesis of RNA by in vitro transcription. *Methods Mol. Biol.* **703**, 29–41 (2011).
2. Wang, W. *et al.* Bacteriophage T7 transcription system: an enabling tool in synthetic biology. *Biotechnol. Adv.* **36**, 2129–2137 (2018).
3. Borkotoky, S. & Murali, A. The highly efficient T7 RNA polymerase: A wonder macromolecule in biological realm. *Int. J. Biol. Macromol.* **118**, 49–56 (2018).

4. Imburgio, D., Rong, M., Ma, K. & McAllister, W. T. Studies of promoter recognition and start site selection by T7 RNA polymerase using a comprehensive collection of promoter variants. *Biochemistry* **39**, 10419–10430 (2000).

5. Rong, M., He, B., McAllister, W. T. & Durbin, R. K. Promoter specificity determinants of T7 RNA polymerase. *Proc. Natl. Acad. Sci.* **95**, 515–519 (1998).

6. Milligan, J. F., Groebe, D. R., Witherell, G. W. & Uhlenbeck, O. C. Oligoribonucleotide synthesis using T7 RNA polymerase and synthetic DNA templates. *Nucleic Acids Res.* **15**, 8783–8798 (1987).

7. Rosa, M. D. Four T7 RNA polymerase promoters contain an identical 23 bp sequence. *Cell* **16**, 815–825 (1979).

8. Chapman, K. A., Gunderson, S. I., Anello, M., Wells, R. D. & Burgess, R. R. Bacteriophage T7 late promoters with point mutations: quantitative footprinting and in vivo expression. *Nucleic Acids Res.* **16**, 4511–4524 (1988).

9. Chapman, K. A. & Burgess, R. R. Construction of bacteriophage T7 late promoters with point mutations and characterization by in vitro transcription properties. *Nucleic Acids Res.* **15**, 5413–5432 (1987).

10. Li, T., Ho, H. H., Maslak, M., Schick, C. & Martin, C. T. Major Groove Recognition Elements in the Middle of the T7 RNA Polymerase Promoter. *Biochemistry* **35**, 3722–3727 (1996).

11. Milligan, J. F. & Uhlenbeck, O. C. B. T.–M. in E. [5] Synthesis of small RNAs using T7 RNA polymerase. in *RNA Processing Part A: General Methods* **180**, 51–62 (1989).

12. Uhlenbeck, O. C. Keeping RNA happy. *RNA* **1**, 4–6 (1995).

13. Pan, T., Artsimovitch, I., Fang, X., Landick, R. & Sosnick, T. R. Folding of a large ribozyme during transcription and the effect of the elongation factor NusA. *Proc. Natl. Acad. Sci.* **96**, 9545–9550 (1999).

14. Diegelman–Parente, A. & Bevilacqua, P. C. A mechanistic framework for co–transcriptional folding of the HDV genomic ribozyme in the presence of downstream sequence. *J. Mol. Biol.* **324**, 1–16 (2002).

15. Lupták, A., Ferré–D'Amaré, A. R., Zhou, K., Zilm, K. W. & Doudna, J. A. Direct pKa Measurement of the Active–Site Cytosine in a Genomic Hepatitis Delta Virus Ribozyme. *J. Am. Chem. Soc.* **123**, 8447–8452 (2001).

16. Bevilacqua, P. C., Brown, T. S., Chadalavada, D., Parente, A. D. & Yajima, R. Kinetic analysis of ribozyme cleavage. In Kinetic Analysis of Macromolecules: A Practical Approach. *Oxford University Press.* **Chpt 3**, 49-74 (2003).

17. Long, D. M. & Uhlenbeck, O. C. Kinetic characterization of intramolecular and intermolecular hammerhead RNAs with stem II deletions. *Proc. Natl. Acad. Sci.* **91**, 6977–81 (1994).

18. Levin, J. R., Krummel, B. & Chamberlin, M. J. Isolation and properties of transcribing ternary complexes of Escherichia coli RNA polymerase positioned at

a single template base. *J. Mol. Biol.* **196**, 85–100 (1987).

19. Erie, D. A., Yager, T. D. & Hippel, P. H. Von. The single-nucleotide addition cycle in transcription: a biophysical and biochemical perspective. *Annu Rev Biophys Biomol Struct.* **21**, 379–415 (1992).
20. Sohn, Y., Shen, H. & Kang, C. Stepwise walking and cross–linking of RNA with elongating T7 RNA polymerase. *Methods Enzymol.* **371**, 170–179 (2003).
21. Yin, Y. W. & Steitz, T. A. Structural basis for the transition from initiation to elongation transcription in T7 RNA polymerase. *Science* **298**, 1387–1395 (2002).
22. Sousa, R. Structural and mechanistic relationships between nucleic acid polymerases. *Trends Biochem. Sci.* **21**, 186–190 (1996).
23. Mentesana, P. E., Chin–Bow, S. T., Sousa, R. & McAllister, W. T. Characterization of halted T7 RNA polymerase elongation complexes reveals multiple factors that contribute to stability. *J. Mol. Biol.* **302**, 1049–1062 (2000).
24. Ikeda, R. A. & Richardson, C. C. Interactions of the RNA polymerase of bacteriophage T7 with its promoter during binding and initiation of transcription. *Proc. Natl. Acad. Sci. U. S. A.* **83**, 3614–3618 (1986).
25. Martin, C. T., Muller, D. K. & Coleman, J. E. Processivity in early stages of transcription by T7 RNA polymerase. *Biochemistry* **27**, 3966–3974 (1988).
26. Durniak, K. J., Bailey, S. & Steitz, T. A. The structure of a transcribing T7 RNA polymerase in transition from initiation to elongation. *Science* **322**, 553–557 (2008).
27. Zhou, Y. & Martin, C. T. Observed instability of T7 RNA polymerase elongation complexes can be dominated by collision–induced 'bumping'. *J. Biol. Chem.* **281**, 24441–24448 (2006).
28. Ling, M. L., Risman, S. S., Klement, J. F., McGraw, N. & McAllister, W. T. Abortive initiation by bacteriophage T3 and T7 RNA polymerases under conditions of limiting substrate. *Nucleic Acids Res.* **17**, 1605–1618 (1989).
29. Roberts, R. J., Vincze, T., Posfai, J. & Macelis, D. REBASE—a database for DNA restriction and modification: enzymes, genes and genomes. *Nucleic Acids Res.* **43**, 298–299 (2014).
30. Roberts, R. J. Restriction and modification enzymes and their recognition sequences. *Nucleic Acids Res.* **9**, 75–96 (1981).
31. Frankel, A. D., Ackers, G. K. & Smith, H. O. Measurement of DNA–protein equilibria using gel chromatography: application to the HinfI restriction endonuclease. *Biochemistry* **24**, 3049–3054 (1985).
32. Pingoud, A., Fuxreiter, M., Pingoud, V. & Wende, W. Type II restriction endonucleases: structure and mechanism. *Cell. Mol. Life Sci.* **62**, 685–707 (2005).
33. Pingoud, A. & Jeltsch, A. Structure and function of type II restriction endonucleases. *Nucleic Acids Res.* **29**, 3705–3727 (2001).

34. Fox, K. R. DNAase I footprinting of restriction enzymes. *Biochem. Biophys. Res. Commun.* **155**, 779–785 (1988).

35. Passalacqua, L. F. M., Jimenez, R. M., Fong, J. Y. & Lupták, A. Allosteric modulation of the Faecalibacterium prausnitzii hepatitis delta virus–like ribozyme by glucosamine 6–phosphate: the substrate of the adjacent gene product. *Biochemistry* **56**, 6006–6014 (2017).

36. Jimenez, R. M., Polanco, J. A. & Lupták, A. Chemistry and biology of self–cleaving ribozymes. *Trends Biochem. Sci.* **40**, 648–661 (2015).

37. Koh, H. R. *et al.* Correlating Transcription Initiation and Conformational Changes by a Single–Subunit RNA Polymerase with Near Base–Pair Resolution. *Mol. Cell* **70**, 695–706.e5 (2018).

38. Jia, Y. & Patel, S. S. Kinetic mechanism of transcription initiation by bacteriophage T7 RNA polymerase. *Biochemistry* **36**, 4223–4232 (1997).

39. Tang, G.–Q., Roy, R., Bandwar, R. P., Ha, T. & Patel, S. S. Real–time observation of the transition from transcription initiation to elongation of the RNA polymerase. *Proc. Natl. Acad. Sci.* **106**, 22175–22180 (2009).

40. Skinner, G. M., Baumann, C. G., Quinn, D. M., Molloy, J. E. & Hoggett, J. G. Promoter binding, initiation, and elongation by bacteriophage T7 RNA polymerase: A single–molecule view of the transcription cycle. *J. Biol. Chem.* **279**, 3239–3244 (2004).

41. Munson, L. M. & Reznikoff, W. S. Abortive initiation and long RNA synthesis. *Biochemistry* **20**, 2081–2085 (1981).

42. Maslak, M. & Martin, C. T. Kinetic analysis of T7 RNA polymerase transcription initiation from promoters containing single–stranded regions. *Biochemistry* **32**, 4281–4285 (1993).

43. Eliceiri, K. W., Schneider, C. A. & Rasband, W. S. NIH Image to ImageJ : 25 years of image analysis. *Nat. Methods* **9**, 671–675 (2012).

44. Das, S. R. & Piccirilli, J. A. General Acid Catalysis by the Hepatitis Delta Virus Ribozyme. *Nat. Chem. Biol.* **1**, 45–52 (2005).

45. Chen, J. H. *et al.* A 1.9 Å crystal structure of the HDV ribozyme precleavage suggests both lewis acid and general acid mechanisms contribute to phosphodiester cleavage. *Biochemistry* **49**, 6508–6518 (2010).

46. Thaplyal, P., Ganguly, A., Hammes–Schiffer, S. & Bevilacqua, P. C. Inverse Thio Effects in the Hepatitis Delta Virus Ribozyme Reveal that the Reaction Pathway Is Controlled by Metal Ion Charge Density. *Biochemistry* **54**, 2160–2175 (2015).

47. Golden, B. L. Two distinct catalytic strategies in the hepatitis δ virus ribozyme cleavage reaction. *Biochemistry*, **50,** 9424–9433 (2011).

Chapter 6: Regulation of mRNA Translation by a Photoriboswitch

Rotstan, K. A., Abdelsayed, M. M., **Passalacqua, L. F. M.**, Chizzolini, F., Sudarshan, K., Chamberlin, A. R., Míšek, J. & Lupták, A. Regulation of mRNA translation by a photoriboswitch. *Submitted.*

Contribution Statement

A. L. and J. M. designed the study. J. M. and M. M. A. designed the *in vitro* selection experiment. J. M. and K. S. synthesized the ligands with the help of A. R. C. J. M. characterized the ligand. M. M. A. carried out the *in vitro* selection. K. A. R. performed the *in vitro* and *in vivo* riboswitch characterization experiments. F. C. and K. A. R. designed the luciferase reporter system. **L. F. M. P.**, K. A. R., and M. M. A. performed the structure probing of Were-1. A. L. and K. A. R. prepared figures and wrote the manuscript with input from all other authors. K. A. R. and M. M. A. authors contributed equally to this work.

6.1 Abstract

Optogenetic tools have revolutionized the study of receptor-mediated biological processes, but such tools are lacking for the study of RNA-controlled systems. To fill this gap, we used *in vitro* selection to isolate a novel RNA that selectively binds the *trans* isoform of a stiff-stilbene (amino-*t*SS), a rapidly and reversibly photoisomerizing small molecule. Structural probing revealed that Were-1 binds amino-*t*SS about 100-times

stronger than amino-*c*SS, giving the system robust selectivity for the *trans* isomer. *In vitro* and *in vivo* functional analysis showed that the riboswitch, termed Werewolf-1 (Were-1), inhibits translation of a downstream open reading frame when bound to amino-*t*SS and photoisomerization of the ligand with a sub-millisecond pulse of light induced the protein expression. Similarly, bacterial culture containing the *cis* isoform (amino-*c*SS) supported protein expression, which was inhibited upon photoisomerization to amino-*t*SS. Reversible regulation of gene expression using a genetically encoded light-responsive RNA will broaden the analysis of complex RNA processes in living cells.

6.2 Introduction

Optogenetic techniques have transformed the biomedical sciences by controlling biological events with high spatiotemporal resolution through triggering signal transduction pathways *via* light-sensing proteins[1–4]; however, there are currently no photoactive molecules that can reversibly regulate cellular events at the RNA level. Photo-caged ligands have previously been used to regulate RNA, but their photo-uncaging procedures require relatively long UV light exposure and are irreversible, allowing only a single molecular event to be initiated[5–10]. Furthermore, a light-responsive ribozyme has shown reversible activity,[11] and aptamers that bind photo-reversible ligands have been identified[12–14], but these RNAs have only been used *in vitro*. Here we used *in vitro* selection[15–17] to isolate a novel RNA that selectively binds only one photoisoform of a ligand, amino *trans* stiff-stilbene (amino-*t*SS)[18–20]. Chemical probing identified amino-*t*SS–induced RNA structural changes in both the aptamer domain and a downstream expression platform derived from a bacterial riboswitch. *In vitro* and *in vivo* functional analysis showed that the riboswitch, termed Were-1, can induce or inhibit translation of a downstream open reading frame upon exposure to a sub-millisecond pulse of light, through reversible photoisomerization of the ligand. Our results demonstrate how a genetically encoded light-responsive RNA can reversibly regulate gene expression using light, providing a new optogenetic tool to broaden the analysis of complex RNA processes in living cells[20–23].

To isolate a new aptamer fused to a functional expression platform, we constructed an RNA pool derived from a bacterial SAM-I riboswitch[24] by replacing its ligand-binding domain with a 45-nucleotide random sequence, partially randomizing its anti-terminator and terminator hairpins, and retaining its translation initiation sequences (**Supplementary Fig. 6.1**). We synthesized a photoactive ligand – a *trans* stiff stilbene with an amino-terminated linker (amino-*t*SS) designed to maintain good cell permeability (**Fig. 6.1a**). The ligand was also designed to have a narrow window for photoregulation of both isomerizations in order to keep the rest of the visible spectrum available for the potential readouts of luminescent assays. Amino-*t*SS was characterized using UV-Vis and NMR spectroscopy to confirm photoisomerization to the *cis* conformation at 342 nm and back to the *trans* conformation at 372 nm, and to ensure that both isoforms are stable on timescales relevant to pulsed gene expression (**Supplementary Fig. 6.2**). The RNA pool was selected *in vitro* to bind amino-*t*SS coupled to carboxylate agarose beads and eluted under denaturing conditions[16]. We hypothesized that a pool of amino-*t*SS–binding aptamers would include motifs that do not bind the *cis* photoisoform of the ligand and that the *t*SS–binding conformation stabilizes the expression platform in a single state that affects either transcription or translation of a downstream open reading frame (ORF).

After six rounds of *in vitro* selection, we cloned the pool into bacterial plasmids and tested individual sequences for amino-*t*SS binding by monitoring RNA–dependent changes in the fluorescence of the amino-*t*SS. One sequence showed markedly increased fluorescence of amino-*t*SS (**Supplementary Fig. 6.3, Fig. 6.1c**). This sequence, termed Werewolf-1 (Were-1) for its potential light-dependent conformational changes, was chosen for further analysis.

6.3 Results and Discussion

To assess the ligand-dependent structural modulation of Were-1, we performed multiple RNA structure-probing experiments, including digestions with T1 and S1 nucleases[25,26], terbium (III) footprinting[27], in-line probing[28], and selective 2' hydroxyl acylation by primer extension (SHAPE)[29] using a range of amino-*t*SS concentrations (**Fig. 6.1b, Supplementary Fig. 6.4**). The changes in the pattern of RNA probing suggested that Were-1 undergoes conformational modulation upon introduction of amino-*t*SS in both the sequence derived from the randomized region and the expression platform (**Fig. 6.1b, Supplementary Figs. 6.4a–6.4c, and 6.4g**). Control experiments with amino-*t*SS analogs, such as *trans*-stilbene (*t*S), 4,4-*trans*-dihydroxystilbene (*t*DHS), and S-adenosyl methionine (SAM), showed no change in the probing patterns (**Supplementary Figs. 6.4c–6.4f**), suggesting that Were-1 is specific for amino-*t*SS. Analysis of the T1 probing experiments revealed a K_D of 1.1 μM, based on the change in band intensity with increasing amino-*t*SS for nucleotide G42, normalized to a control band G72 (**Figure 6.1b**). Additionally, a 108-µM K_D was calculated based on the same band intensity change with increasing amino-*c*SS (**Supplementary Fig. 6.4g**), revealing a 100-fold specificity for the target ligand, amino-*t*SS. The average amino-*t*SS K_D derived from the T1 nuclease probing (at positions 42, 46, 77, and 80; normalized to the band intensity at position 72) was of 1.5 μM, whereas the nuclease S1 probing revealed a K_D of 0.4 μM (based on band intensity change for positions A44-G46), and the terbium (III) footprinting yielded a somewhat higher apparent K_D of 4.8 μM (calculated based on the change in intensity at positions A113 and U107, normalized to G134 control band; **Supplementary Fig. 6.4c**). In order to establish the location of the amino-*t*SS aptamer domain, we modeled the secondary structure of Were-1 based on the probing data and created mutants hypothesized to affect ligand binding affinity or RNA structural stability (**Fig. 6.1c**) and tested them *in vitro* and *in vivo*. The secondary structure modeling did not support a conformation containing a Rho-independent transcriptional terminator, in part because the selected sequence contained two mutations (C90A and U92A), which are predicted to disrupt the stability of a full-length transcription-terminating helix (**Fig. 6.1c**).

To test whether Were-1 can directly couple light-induced states of the ligand to the activity of the expression platform *in vitro*, the aptamer was tested for amino-*t*SS–dependent conformational changes using a strand-displacement assay that mimics mRNA binding by the bacterial ribosome[17,30,31]. We designed a DNA duplex in which the longer strand has a toehold sequence corresponding to the reverse complement of the Shine-Dalgarno sequence of Were-1 (**Supplementary Table 6.1**) and a fluorophore to assess whether the shorter DNA strand, containing a quencher chromophore, is displaced through RNA:DNA hybridization with Were-1 (**Supplementary Fig. 6.5a**). Polyacrylamide gel electrophoresis (PAGE)-purified Were-1 bound the toehold readily, but the strand displacement was diminished in the presence of the ligand amino-*t*SS

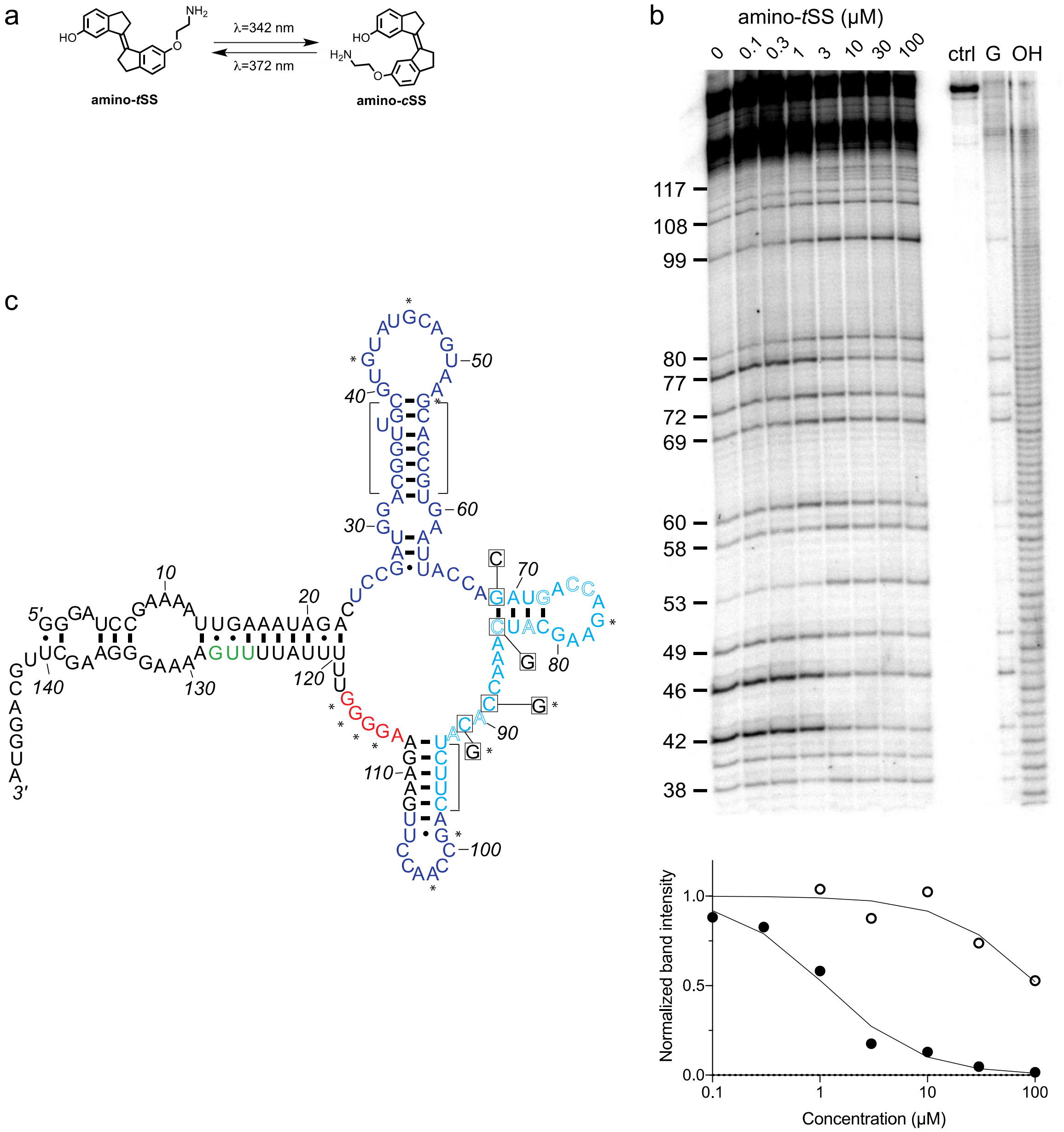


Figure 6.1. An amino-*t*SS-responsive aptamer. a, Amino-*t*SS isomerizes from *trans* to *cis* conformation when exposed to 342 nm light, and back to the *trans* isoform at 372 nm. **b**, RNase T1 probing of Were-1 structure. Right lanes contain a control with undigested RNA (ctrl), a T1-digested sequencing control (G), and a hydroxide-mediated partial digestion ladder (OH) of the RNA. The left lanes show partial T1 digestion in the presence of increasing amino-*t*SS at concentrations indicated above the gel image. The probing

shows clear ligand-dependent changes—both increases (e.g. G53, G99, G114-117) and decreases (e.g. G 42, G46, G77)—interspersed throughout the sequence. Below, an apparent K_D of 1.1 μM was calculated based on the change in band intensity with increasing amino-*t*SS (dark, filled circles) for nucleotide G46, normalized to a control band (G72). Additionally, a K_D of 108 μM was calculated based on the change in band intensity with increasing amino-*c*SS (open circles) for the same nucleotide and control (Supplementary Fig. 6.4g), suggesting high specificity for amino-*t*SS. An average K_D value of 1.5 µM amino-*t*SS was calculated for changes in nucleotides G42, G46, G77, and G80. **c**, Secondary structure prediction of Were-1 derived from all structural probing data in absence of the ligand (see also Supplementary Fig. 4). Partially randomized regions (light blue), the Shine-Dalgarno sequence (red), the start codon (green), and the 3' terminus sequence are derived from the *B. subtilis mswA* SAM-I riboswitch. The 5' part of the aptamer (dark blue) was selected from the random region of the starting pool (Fig. S1). Outlined letters are positions where the selected sequence differs from the *B. subtilis* riboswitch expression platform. Boxed positions were mutated to the indicated nucleotides to identify regions of structural and functional importance. Bracketed regions indicate areas that do not change in the presence of amino-*t*SS, and asterisks (*) indicate nucleotide positions that do change in the presence of amino-*t*SS.

(**Supplementary Fig. 6.5b**). Testing the toehold binding at various concentrations of amino-*t*SS revealed a dose dependence with a half–maximal inhibition of toehold binding at ~6 μM (**Supplementary Fig. 6.5c**).

We next asked whether the ribosome mimic binds this RNA during *in vitro* transcription (**Fig. 6.2a, Supplementary Fig. 6.6**). In the absence of the ligand, the RNA bound the toehold efficiently, showing a robust increase of fluorescence immediately after transcription initiation. In contrast, the addition of high concentration (14.8 μM) of amino-*t*SS strongly abrogated new binding of the toehold, as revealed by almost a full reduction in the slope of the fluorescence expansion curve (**Fig. 6.2b**). Intermediate concentrations of amino-*t*SS were tested to assess RNA binding and specificity, yielding a ligand-dependent response with a half-maximum of ~4 μM amino-*t*SS (**Fig. 6.2c**). Probing-derived secondary structure modeling of Were-1 suggested substantial ligand-dependent conformational changes in multiple parts of the sequence, except for one predicted hairpin. To confirm the presence of this structural element, we created a variant containing a single mutation (G69C), that was predicted to disrupt this helix, and a presumed compensatory mutant (G69C/C84G) (**Fig. 6.1c**). The G69C variant showed diminished response to amino-*t*SS, whereas the G69C/C84G double-mutant exhibited partially restored activity, suggesting that these two positions are indeed part of a helix. Other variants, C89G and C91G, designed based on parts of the sequence that showed amino-*t*SS–dependent changes in the structure-probing experiments, both showed decreased sensitivity to the ligand, suggesting that they are essential for ligand binding. Furthermore, when testing toehold binding using the purified *cis* isoform of the stiff stilbene (amino-*c*SS), as well as other stilbenes, such as *t*S, *t*DHS, and *trans* stiff stilbene (*t*SS; Were-1 ligand lacking the aminolated linker), no significant changes in fluorescence were observed (**Fig. 6.2c, Supplementary Fig. 6.7**). These results demonstrate that amino-*t*SS stabilizes the RNA in an “OFF” (ribosome-inaccessible) conformation in a dose-dependent manner and with high ligand specificity.

In order to test whether the Were-1 RNA interaction is selective for amino-*t*SS, and potentially acts as an amino-*t*SS riboswitch, we created a construct consisting of the putative riboswitch, including its minor start codon from *Bacillus subtilis* that was present in the starting pool, followed by a firefly luciferase (Fluc) ORF lacking its endogenous start codon. Based on the toehold assays, we hypothesized that in absence of amino-*t*SS, Were-1 would be transcribed in a conformation promoting the translation of the luciferase enzyme, whereas the presence of amino-*t*SS would stabilize a conformation preventing efficient translation initiation, downregulating the luciferase expression (**Fig. 6.2d**). Using a purified bacterial *in vitro* transcription/translation system, luciferase production was measured in the presence and absence of amino-*t*SS. In absence of the ligand, the construct exhibited robust luciferase production, demonstrating that the unbound aptamer promotes protein production from a downstream ORF (**Fig. 6.2e**). In contrast, when amino-*t*SS was added, protein production decreased in a dose-dependent manner (**Fig. 6.2f**). To confirm that the ligand itself did not impact the *in vitro* translation system, luminescence was tested with a control plasmid lacking the riboswitch, and no amino-*t*SS sensitivity was observed (**Supplementary Fig. 6.8**).

To further test the riboswitch, we incorporated the construct into a bacterial plasmid and induced its expression in *E. coli* cells (**Fig. 6.2g**). Bioluminescence, due to luciferase expression and activity, was robust in the absence of ligand, and when the cells were incubated in the presence of amino-*t*SS, bioluminescence was again diminished in a dose-dependent manner (**Fig. 6.2h, 6.2i**). To confirm the specificity of Were-1 for amino-*t*SS, we incubated cells in the presence of amino-*c*SS, *t*S and *t*DHS, and observed no significant change in bioluminescence (**Fig. 6.2i**). To determine the effect of an alternate start codon on Fluc expression, we changed the UUG start codon in the Were-1-Fluc plasmid to AUG. Bioluminescence was higher in the AUG samples compared to the wild-type UUG construct and showed a similar amino-*t*SS–dependent response (**Supplementary Fig. 6.9**). Testing the above-mentioned mutants confirmed the G69/C84 interaction, and the importance of the C89 and C91 positions for ligand binding. Taken together, our results demonstrate that Were-1 controls amino-*t*SS–dependent protein expression *in vitro* and *in vivo,* acting as a translational riboswitch.

We next asked whether Were-1 could regulate gene expression in a light-dependent manner, acting as a photoriboswitch. For this to occur, the *trans* isoform of the stiff stilbene must photoisomerize to its *cis* state, preventing binding of the Were-1 aptamer domain, and promoting expression of a downstream ORF.

Riboswitches are sensitive to co-transcriptional events because they are capable of adopting different RNA folds as they are transcribed by RNA polymerase [32]. To first monitor changes in RNA folding over time, we used the toehold-fluorophore system to determine whether the DNA duplex was in a bound state (no strand displacement) versus an unbound state (strand displacement releasing the quencher DNA, yielding fluorescence) during transcription. Our results show that when the amino-*t*SS-bound Were-1 structure was irradiated at 342 nm of light, the toehold fluorescence increased, suggesting that the RNA increased the binding to the ribosome mimic present on the DNA toehold. Furthermore, when exciting amino-*c*SS at 372 nm of light to switch the ligand to its *trans* isoform, we observed a decrease in toehold fluorescence growth, indicating that the photo-generated amino-*t*SS was able to re-bind the Were-1 RNA. The irradiation was repeated until all toehold was bound, with each switch showing consistent results (**Supplementary Fig. 6.10**). This experiment shows reversible, wavelength-dependent binding of the ribosome mimic, emulating light-dependent protein expression from a downstream open reading frame.

To study the system further, we used the Were-1-luciferase construct and the purified bacterial *in vitro* transcription/translation system to test whether luciferase expression could be regulated by our putative photoriboswitch. We found that when the reaction was irradiated at 342 nm of light, luminescence increased, suggesting that Were-1's conformation changed to expose the RBS, enabling luciferase expression. When the *cis* stilbene was photoisomerized to the *trans* state (amino-*t*SS) at 372 nm of light, luciferase protein production slightly decreased (**Supplementary Fig. 6.11**). These data are consistent with an *in vitro* activity of a photoriboswitch.

Finally, we used *E. coli* containing the Were-1-Fluc construct to determine whether gene expression can be regulated with a pulse of light. As shown above, bioluminescence was greatly diminished in the presence of amino-*t*SS (**Fig. 6.2i**). Upon exposing the bacteria to 342 nm light, we saw a robust increase in bioluminescence (**Fig. 6.3a, 6.3b and Supplementary Fig. 6.12**). This result strongly suggests that upon photoisomerization of amino-*t*SS to the *cis* isoform with a pulse of light, Were-1 was able to change conformation and expose its RBS to allow luciferase production.

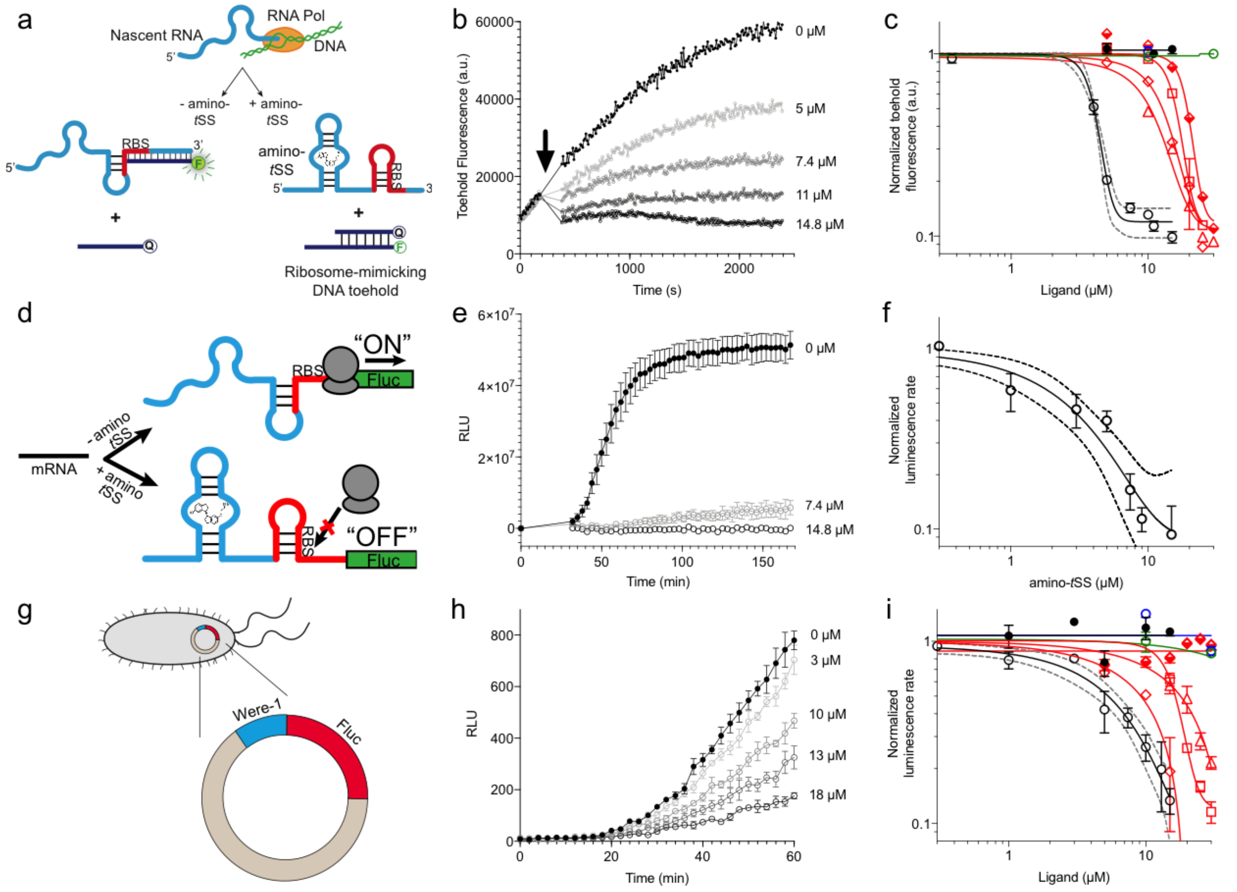


Figure 6.2. Translation regulation by the Were-1 riboswitch. a, Schematic of co-transcriptional binding of Were-1 RNA to amino-*t*SS in the presence of a toehold-reporter complex. In absence of amino-*t*SS, the transcribed RNA exposes the ribosomal binding site (RBS), enabling binding of the complementary region of the toehold reporter, displacing the quencher strand, and producing a fluorescence signal. In presence of amino-*t*SS, the RNA binds the ligand, sequestering the RBS and preventing displacement of the quencher strand. **b**, Co-transcriptional response of Were-1 to different concentrations of amino-*t*SS using the toehold reporter. Initial transcriptions of Were-1

without ligand show identical increase in toehold fluorescence for all samples. When amino-*t*SS is added (arrow), a dose-dependent decrease in fluorescence is observed. **c**, Response (± SEM; n=81) of Were-1 (black, open circles), and its variants (red) C89G (triangles), C91G (squares), G69C (half-shaded diamonds), and G69C/C84G (open diamonds), in the presence of amino-*t*SS shows a shift in dose-dependence for single mutations, particularly G69C, and partial recovery of activity for the G69C/C84G double mutant. Were-1 shows no response in the presence of amino-*c*SS (black circles), *trans*-stilbene (green, open circles) and *trans*-4,4-dihydroxystilbene (blue, open circles). **d**, Schematic of amino-*t*SS-dependent inhibition of protein expression *in vitro* using a Were-1-firefly luciferase (Were-1-Fluc) construct. In absence of the ligand, the RBS is exposed and luciferase is translated, whereas in presence of amino-*t*SS, the RBS is sequestered, abrogating Fluc expression. **e**, *In vitro* translation of the Were-1-Fluc construct. Robust luminescence is observed when no ligand is present, but the signal is significantly lower in presence of amino-*t*SS. **f**, Response (± SEM; n = 58) of the Were-1–regulated protein expression to amino-*t*SS. **g**, Schematic of the Were-1-Fluc construct incorporated into a bacterial plasmid. **h**, Were-1–controlled Fluc gene expression in *E. coli*. Bioluminescence is observed in absence of amino-*t*SS, and progressively diminished with increasing amino-*t*SS. **i**, Expression of Were-1-Fluc (± SEM; n=257) *in vivo* (black, open circles), and its variants (red) C89G (triangles), C91G (squares), and G69C/C84G (open diamonds), in the presence of amino-*t*SS, show a dose-dependent response. Were-1 mutant G69C (half-shaded diamond) and Were-1 in the presence of amino-*c*SS (black circles), *trans*-stilbene (green, open circles) and *trans*-4,4-dihydroxystilbene (blue, open circles) showed no change in bioluminescence, whereas the G69C/C84G double mutant shows restoration of activity similar to wild-type levels. Note, dose-response graphs (**c**, **f**, **i**) are on a log-log scale. The apparent amino-*t*SS IC_{50}s are 3.9 ± 0.2, 2.5 ± 1.0, and 5.3 ± 1.1 μM for the toehold (**c**), *in vitro* translation (**f**), and *in vivo* expression (**i**), respectively. Dashed lines correspond to the 95% confidence interval of the binding model.

To confirm these results, we performed control experiments, in which cells were covered during the excitation to distinguish between regular *E. coli* growth behavior and the increase in bioluminescence from the riboswitch (**Fig. 6.3a, 6.3b**). Additionally, in another experiment, we excited cells with a different wavelength of light (500 nm) that should not impact the isomerization of the ligand (**Supplementary Fig. 6.12**). Both controls showed lower bioluminescence compared to the 342 nm excited samples. To further analyze the system, we tested the temporal dependence of Were-1–regulated Fluc production by exposing an amino-*t*SS–containing bacterial culture to 342 nm light for various lengths of time. Relative to controls that were unexposed, the highest luciferase expression was seen at an exposure time of 500 μs (**Fig. 6.3b**). Furthermore, when testing the same system using amino-*c*SS, bioluminescence decreased in a dose-dependent manner with increasing amino-*c*SS concentration after exposure to 390 nm light (**Fig. 6.3c**). This result strongly suggests that by isomerizing amino-*c*SS to its *trans* isoform with a pulse of light, Were-1 was able to sequester its RBS to inhibit luciferase production. Testing the temporal response of Were-1–regulated Fluc expression in the presence of amino-*c*SS revealed that Were-1 regulates expression optimally at short exposures, showing the highest inhibition after a millisecond of light exposure (**Fig. 6.3d**). Based on isomerization data (**Supplementary Fig. 6.2b**), this effect is likely due to the ligand reaching a semi-photostationary state after longer light exposure. No difference in cell density was observed among the experiments, implying that neither the ligand, nor the light pulses affect the bacterial growth, and suggesting negligible photo-damage to the cells.

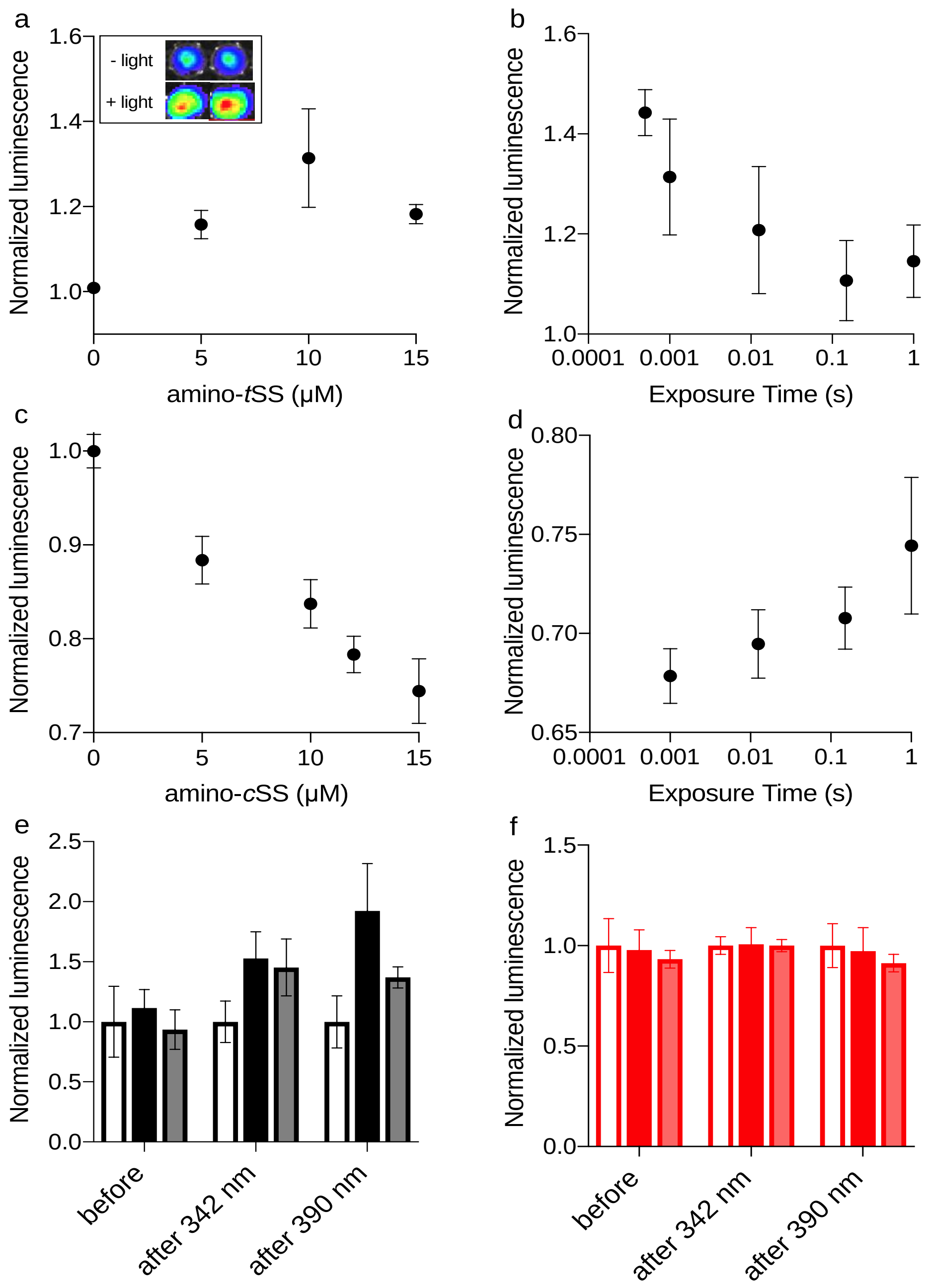


Figure 6.3. Regulation of luciferase expression by the Were-1 photoriboswitch *in vivo*. a, Normalized amino-*t*SS–dependent bioluminescence (± SEM) of the Were-1-Fluc construct after 1 ms exposure of 342 ± 5 nm light ($\Phi_q = 1.4*10^{-2}$ W/cm^2). The largest change in expression was observed in the presence of 10 µM amino-*t*SS. Inset shows the light–dependent bioluminescence of the bacterial cultures at 10 µM ligand. **b**, Were-

1 regulation of luciferase expression (± SEM) *in vivo* at various exposure times in presence of 10 µM amino-*t*SS. **c**, Normalized bioluminescence of the Were-1-Fluc *E. coli* incubated with amino-*c*SS after 1 s exposure of 390 ± 9 nm light ($\Phi_q = 5.5*10^{-2}$ W/cm^2) showing progressively higher protein expression inhibition at higher amino-*c*SS concentrations, presumably due to higher concentration of amino-*t*SS after photoisomerization. **d**, Change of Fluc expression after photoisomerization of 15 µM amino-*c*SS at 390 ± 9 nm light for various exposure times, showing largest photoswitching at 1 ms exposure. **e,** Regulation of luciferase expression (± SEM) by the Were-1-Fluc construct before exposure, forty-five minutes after a 1-ms pulse of 342 ± 5 nm light (dark and gray), which resulted in increased bioluminescence compared to control (clear), and forty-five minutes after 0.5 ms of 390 ± 9 nm exposure that resulted in decreased bioluminescence (gray) compared to samples that were only exposed to 342 nm. **f,** Luciferase expression (± SEM) by the Were-1-Fluc G69C mutant using the same conditions as above (e), showing no significant change in expression after exposure to 342 nm (red, light red) or 342 nm and 390 nm (light red) compared to unexposed controls (clear).

We next asked whether Were-1 could reversibly regulate gene expression *in vivo*, providing the first optogenetic tool to reversibly regulate cellular events at the RNA level. Using the same *E. coli* construct, we measured bioluminescence over two hours in samples that were covered during excitations, and therefore unexposed to light, samples that were exposed to a millisecond of 342 nm light, and samples that were exposed to 342 nm and then later excited at 390 nm for a sub-millisecond. Initial values prior to excitation showed no significant difference in bioluminescence twenty-five minutes post induction (**Fig. 6.3e**). Samples that were then exposed to 342-nm light showed a significant increase in luciferase expression forty-five minutes after exposure, compared to the unexposed control. Lastly, the samples that were subsequently exposed to 390-nm light decreased in bioluminescence forty-five minutes post exposure in comparison to those that were unexposed and those that were exposed only to 342 nm. As an additional control, the G69C mutant was exposed alongside Were-1 and showed no significant difference when G69C excited with 342 nm light, or 342 nm and 390 nm (**Fig. 6.3f**). These results strongly suggest that Were-1 is a photoriboswitch that can reversibly regulate protein expression *in vivo*.

6.4 Conclusion

Taken together, we show that novel riboswitches, regulated by synthetic ligands, can be evolved from random libraries fused to expression platforms. In the case of Were-1, binding of the target ligand stabilizes the RNA in a conformation that impedes the translation of a downstream ORF. This approach will likely also yield transcription-regulating riboswitches, and further molecular engineering will allow regulation of a wide range of cellular events in both *cis* and *trans*. We expect that Were-1 and similar photoriboswitches will allow reversible photoregulation of a variety of RNA-centered cellular events with a very high spatiotemporal resolution in bacteria and multicellular organisms alike.

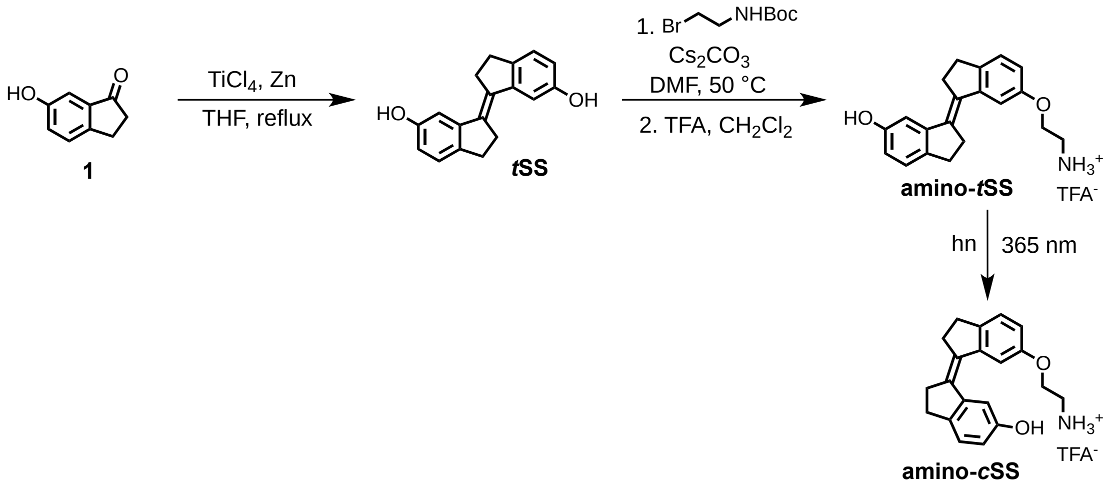
6.5 Materials and Methods

Reagents and equipment. Unless otherwise stated, all reagents were purchased from Sigma-Aldrich. (E)-6'-(2-aminoethoxy)-2,2',3,3' tetrahydro-[1,1'-biindenylidene]-6-ol (amino-*t*SS) was synthesized and prepared as described below. Commercially available reagents were used without further purification. Absorbance spectra were recorded with a Thermo Scientific NanoDrop 1000 spectrophotometer. Fluorescence excitation and emission spectra were measured with a Varian Cary Eclipse fluorescence spectrometer, unless otherwise specified. Bioluminescence was measured using an Andor 866 EMCDD camera, BioTek Synergy H1 plate reader, or IVIS Lumina II.

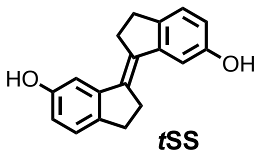
Synthesis of (E)-6'-(2-aminoethoxy)-2,2',3,3' tetrahydro-[1,1'-biindenylidene]-6-ol. All starting reagents were commercially available, and of analytical purity, and were used without further treatment. Solvents were dried according to standard methods. ^{1}H and spectra were recorded on Varian UNITY INOVA-300 and Bruker Avance-600 instruments. Chemical shifts (δ) are reported in ppm relative to residual solvent peak (DMSO: δ_H = 2.50 ppm) as internal standard. Accurate mass measurements (HRMS) were obtained by ESI on an Agilent 6530 Q-TOF MS spectrometer. Analytical TLC was performed using a precoated silica gel 60 Å F_{254} plates (0.2 mm thickness) visualized with UV at 254 nm. Preparative column chromatography was carried out using silica gel 60 Å (particle size

0.063–0.200 mm). Purifications by HPLC were performed under the following conditions: Agilent ZORBAX SB-C18 column (5 µL, 9.4x150 mm); UV/Vis detection at λ_{obs} = 254 nm; flow rate 4 mL/min; gradient elution method H_2O (0.1 % TFA) – CH_3CN (0.1 % TFA) from 95:5 to 0:100 in 20 min. Purity of compounds was confirmed using Agilent eclipse plus C18 column (3.5 µL, 4.6x100 mm); UV/Vis detection at λ_{obs} = 254 nm; flow rate 0.5 mL/min; gradient elution method H_2O (0.1 % TFA) – CH_3CN (0.1 % TFA) from 95:5 to 0:100 in 20 min.

Synthesis details

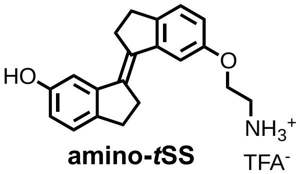

HO
O
1
TiCl4, Zn
THF, reflux
HO
OH
tSS
1. Br NHBoc
Cs2CO3
DMF, 50 °C
2. TFA, CH2Cl2
HO
O
NH3+
TFA-
amino-tSS
hn
365 nm
O
OH
NH3+
TFA-
amino-cSS

Scheme S1. Synthesis of stiff stilbene derivatives ***t*SS, amino-*t*SS,** and **amino-*c*SS**.



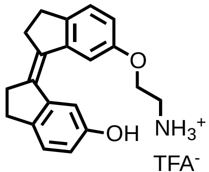
***t*SS**

(*E*)-2,2',3,3'-tetrahydro-[1,1'-biindenylidene]-6,6'-diol (*t*SS). To a stirred suspension of zinc powder (3.20 g, 48.92 mmol) in dry THF (50 ml), $TiCl_4$ (4.67 g, 24.58 mmol) was added over 10 minutes at 0 °C. The resulting slurry was heated at reflux for 1.5 h. Then a THF solution (50 ml) of indanone **1** (600 mg, 4.05 mmol) was added over 3 h period by syringe pump to the refluxing mixture. The reflux was continued for 30 minutes after the addition was complete. After cooling to room temperature, the reaction mixture was poured into a saturated solution of NH_4Cl and extracted with CH_2Cl_2. The organic solutions were dried over $MgSO_4$ and concentrated by rotary evaporation under reduced pressure. The crude product was purified by column chromatography using silica gel (hexane/iPrOH = 10:0.5) to afford ***t*SS** in 51 % yield (273 mg) as a white solid. ^{1}H NMR (400 MHz, DMSO-*d*$_6$): δ 2.95 (m, 4H), 3.02 (m, 4H), 6.63 (dd, *J* = 8.1, 2.2 Hz, 2H), 7.02 (d, *J* = 2.3 Hz, 2H), 7.11 (d, *J* = 8.0 Hz, 2H), 9.17 (s, 2H). ^{13}C NMR (100 MHz, DMSO-*d*$_6$): δ 29.6, 31.9, 111.0, 114.5, 125.2, 135.0, 137.0, 143.7, 156.1. HRMS (ESI): *m/z* $[M]^+$ calculated for $C_{18}H_{16}O_2$ 264.1145; found: 264.1141.



amino-*t*SS

***E*)-6'-(2-aminoethoxy)-2,2',3,3'-tetrahydro-[1,1'-biindenylidene]-6-ol (amino-*t*SS).** A mixture of ***t*SS** (125 mg, 0.47 mmol), 2-(Boc-amino)ethyl bromide (106 mg, 0.47 mmol), Cs_2CO_3 (772 mg , 2.37 mmol) and *n*Bu_4NBr (6 mg, 0.02 mmol) in DMF (15 ml) was heated at 50 °C for 6 h. After cooling to room temperature, CH_2Cl_2 was added and the mixture was washed with a saturated solution of NH_4Cl, water, brine and dried over $MgSO_4$. The organic solutions were concentrated by rotary evaporation under reduced pressure. The crude product was dissolved in a mixture of CH_2Cl_2/TFA (16 ml; 3:1) and stirred at room temperature for 30 minutes. The reaction mixture was concentrated by rotary evaporation under reduced pressure and purified by HPLC (gradient elution method H_2O (0.1 % TFA) – CH_3CN (0.1 % TFA) from 95:5 to 0:100) to afford **amino-*t*SS** (as a TFA salt) in 42 % yield (84 mg) as an off-white solid. ^{1}H NMR (400 MHz, DMSO-*d*$_6$): δ 2.96 (m, 4H), 3.01 (m, 4H), 3.22 (t, *J* = 5.1 Hz, 2H), 4.17 (t, *J* = 5.1 Hz, 2H), 6.66 (dd, *J* = 8.1, 2.2 Hz, 1H), 6.87 (dd, *J* = 8.2, 2.3 Hz, 1H), 7.03 (d, *J* = 2.2 Hz, 1H), 7.13 (d, *J* = 8.1 Hz, 1H), 7.16 (d, *J* = 2.4 Hz, 1H), 7.27 (d, *J* = 8.3 Hz, 1H), 8.10 (s, 3H), 9.28 (s, 1H). ^{13}C NMR (100 MHz, DMSO-d_6): δ 29.58, 29.64, 31.8, 31.9, 38.5, 54.7, 110.8, 111.1, 113.6, 114.8, 125.3, 125.4, 134.5, 135.9, 137.1, 139.6, 143.6, 144.0, 156.2, 156.9. HRMS (ESI): *m/z* $[M+H]^+$ calculated for $C_{20}H_{21}NO_2$ 308.1645; found: 308.1651.



amino-*c*SS

(*Z*)-6'-(2-aminoethoxy)-2,2',3,3'-tetrahydro-[1,1'-biindenylidene]-6-ol ((Z)-1; amino-*c*SS). A solution of **amino-*t*SS** (15 mg, 35.6 μmol) in DMSO (1 ml) in NMR cuvette was irradiated with handheld UV lamp (8 W) for 15 min. The resulting mixture of **amino-*t*SS** and **amino-*c*SS** was purified by HPLC (gradient elution method H_2O (0.1 % TFA) – CH_3CN (0.1 % TFA) from 95:5 to 0:100) to afford **amino-*c*SS** (as a TFA salt) in 20 % yield (2 mg) as an off-white solid. ^{1}H NMR (400 MHz, DMSO-d_6): δ 2.75 (m, 4H), 2.85 (m, 4H), 3.22 (t, *J* = 5.1 Hz, 2H), 4.13 (t, *J* = 5.1 Hz, 2H), 6.63 (dd, *J* = 8.2, 2.2 Hz, 1H), 6.84 (dd, *J* = 8.2, 2.2 Hz, 1H), 7.12 (d, *J* = 8.2 Hz, 1H), 7.25 (d, *J* = 8.3 Hz, 1H), 7.40 (d, *J* = 2.0 Hz, 1H), 7.53 (d, *J* = 2.0 Hz, 1H), 8.02 (s, 3H), 9.25 (s, 1H). ^{13}C NMR (100 MHz, DMSO-d_6): δ 29.2, 29.3, 34.9, 35.0, 38.5, 64.4, 108.9, 109.4, 114.5, 115.2, 125.8, 125.9, 134.4, 135.7, 138.6, 140.79, 140.84, 141.1, 155.46, 156.1. HRMS (ESI): *m/z* $[M+H]^+$ calcd for $C_{20}H_{21}NO_2$ 308.1645; found: 308.1644.

***In vitro* RNA transcription**. RNA was transcribed at 37 ˚C for one hour in a 50 μL volume containing 40 mM tris-HCl, 6 mM dithiothreitol (DTT), 2 mM spermidine, 1.25 mM each rNTP, 8 mM $MgCl_2$, 1 unit of T7 RNA polymerase, and 5 pmol of DNA template. The transcripts were purified by 10 % PAGE under denaturing conditions (7 M urea). RNA was eluted from the gel into 300 µL of 300 mM KCl and precipitated by adding 700 µL of 95 % ethanol at –20 °C.

***In vitro* selection of amino-*t*SS aptamers.** An RNA pool derived from a *B. subtilis mswA* SAM-1 riboswitch, located in the 5′ untranslated region of the *metI* (cystathionine gamma-synthase, also denoted as *yjcI*) gene [24,33] was designed by replacing the riboswitch ligand-binding domain with a random region of 45 nucleotides. The anti-terminator stem and upstream half of the transcriptional terminator sequence were partially randomized at a 15 % level, and the loop of the terminator stem was fully randomized. The remaining part of the riboswitch, including the downstream half of the transcriptional terminator stem, containing a ribosome binding site (RBS) that binds the 3' end of *B. subtilis* 16S rRNA (3'-UUUCCUCCACUAG-5') [34] and an alternative UUG start codon, was retained (Supplementary Fig. 1). The pool was synthesized by Yale School of Medicine's Keck Oligonucleotide Synthesis facility as a single template strand that was then purified by 10 % PAGE and converted into dsDNA by a primer-extension reaction using a primer corresponding to the T7 RNA polymerase promoter. The pool was transcribed at an estimated sequence diversity of 10^{15}.

From that pool, RNAs were selected to bind amino-*t*SS, as follows. PAGE-purified ^{32}P-labeled RNA transcripts of the pool were precipitated, dried, and resuspended in a solution containing 140 mM KCl, 10 mM NaCl, 10 mM tris-chloride, pH 7.5, and 5 mM $MgCl_2$ (binding buffer). The RNA mixture was heated to 70 °C for three minutes and loaded onto agarose beads for a counter-selection step. Binders were discarded and the flow-through was incubated on agarose beads linked to amino-*t*SS. The beads were shaken for five minutes at room temperature, and the unbound RNA was collected. Amino-*t*SS beads were then washed with binding buffer for five minutes at room temperature. This washing step was repeated six times. Potential aptamers were then eluted twice with denaturing buffer, consisting of 7 M urea and 5 mM ethylenediaminetetraacetic acid (EDTA) in 45 mM tris, 45 mM borate buffer, pH 8, and heated at 95 °C for five minutes. Each fraction was analyzed for radioactivity using a liquid scintillation counter. Elutions were pooled, precipitated, dried, and resuspended in water for reverse transcription. The pool was reverse transcribed, and the cDNA was amplified by PCR and used for the next round of selection.

Screening of potential amino-*t*SS binders. After six rounds of *in vitro* selection, the selected pool was cloned into a TOPO TA plasmid (Invitrogen) and transformed into DH5α *E. coli* cells. Cells were plated on agar containing kanamycin and incubated overnight at 37 ˚C. Individual colonies were picked from the master plate and inoculated overnight in Luria Broth containing kanamycin. Plasmids were extracted and purified using a Miniprep Kit (QIAGEN), and sequenced (GENEWIZ). Individual clones were PCR-amplified using the library-specific primers and transcribed to test their optical activity in the presence and absence of amino-*t*SS. *t*SS emission spectra were collected using an excitation at 355 nm. Were-1 showed the highest increase in amino-*t*SS fluorescence at 430 nm and was chosen for further analysis.

Structure probing of Were-1:

T1 nuclease probing. Were-1 RNA was dephosphorylated in a solution of the reaction buffer (50 mM potassium acetate, 20 mM Tris-acetate, 10 mM magnesium acetate, 100 µg/ml BSA, pH 7.9), 1 µg of purified RNA, and 1.5 unit of Shrimp Alkaline Phosphatase

(NEB). The reaction was incubated at 37 °C for 30 minutes, and heat-inactivated at 65 °C for 5 minutes. 5′–labeled RNA (8000 cpm) was prepared in reaction buffer (70 mM Tris-HCl, pH 7.6, 10 mM $MgCl_2$, 5 mM DTT) using 1 μg of 5′–dephosphorylated RNA, 2 μCi [γ-^{32}P] ATP (Perkin Elmer), and 15 units of T4 PNK enzyme (NEB). The reaction was incubated at 37 °C for two hours and PAGE–purified. The 5′–labeled Were-1 RNA was added into binding buffer and the indicated concentrations of amino-*t*SS or controls (no ligand, amino-*c*SS, tS, tDHS, and SAM), and were incubated at 55 °C for 5 minutes and subsequently cooled at room temperature for 5 minutes. Next, T1 nuclease was added (0.05 units; Thermo Fisher Scientific) and samples were incubated at 37 °C for 15 minutes. All conditions were then quenched with a mixture of 7 M urea and 10 mM EDTA. Afterwards, the RNA was added to an equal volume of phenol:chloroform:isoamyl alcohol (25:24:1) and vortexed. Samples were centrifuged for 3 minutes at 8,000 RPM, and the aqueous phase was collected and transferred to a new tube. Samples were fractionated on a 10% PAGE gel and exposed to a phosphor image screen (GE Healthcare) for a minimum of 24 hours. The screen was scanned on a GE Typhoon phosphor imager. A guanosine-specific sequencing lane was resolved in parallel to all samples using 5′–labeled or 3′–labeled RNA (8000 cpm*)*, as specified, in T1 digestion buffer (250 mM sodium citrate, pH 7) and 0.5 units T1. Reactions were incubated at 55 °C for 5 minutes and quenched with a solution containing 7 M urea and 10 mM EDTA. RNA was extracted using phenol-chloroform, as noted above. Partial alkaline hydrolysis was also resolved in parallel by adding 5′–labeled or 3′–labeled RNA (8000 cpm), as specified, into a hydrolysis buffer (50 mM $NaHCO_3$, 1 mM EDTA, pH 10). Reactions were incubated at 95 °C for 10 minutes and quenched in a solution containing 7 M Urea and 10 mM EDTA. RNA was extracted using phenol-chloroform.

SHAPE. A selective 2'-hydroxyl acylation and primer extension (SHAPE) reaction, as described [35], was carried out on Were-1 in the presence of increasing amino-*t*SS concentrations and 30 µM controls (amino-*c*SS, tDHS, tS, and SAM).

S1 nuclease probing. Reactions were prepared by adding 3′–labeled Were-1 RNA (8000 cpm) into S1 nuclease buffer (40 mM sodium acetate, pH 4.5, 300 mM NaCl, and 2 mM $ZnSO_4$), and the indicated concentrations of amino-*t*SS, and were incubated at 55 °C for 5 minutes and subsequently cooled at room temperature for 5 minutes. Next, S1 nuclease was added (0.2 units; Thermo Fisher Scientific) and samples were incubated at 37 °C for 10 minutes and quenched in a solution of 7 M Urea and 10 mM EDTA. Samples were extracted using phenol-chloroform and resolved on a denaturing 10 % PAGE gel. The gel was then exposed to a phosphor image screen and scanned on a GE Typhoon phosphor imager. The sequences in the degradation pattern were assigned by running T1 digestion and partial alkaline hydrolysis in parallel lanes, as noted above.

Terbium (III) footprinting. Reactions were prepared by adding 5′–labeled Were-1 RNA (8000 cpm) into the binding buffer, and the indicated concentrations of amino-*t*SS or controls (no ligand, amino-*c*SS, tS, tDHS, and SAM), and were incubated at 55 °C for 5 minutes and subsequently cooled at room temperature for 5 minutes. Terbium (III) chloride was added to a final concentration of 10 mM and samples were incubated at 37 °C for 30 minutes and then quenched with a solution of 7 M urea and 10 mM diethylenetriaminepentaacetic acid (DTPA). Were-1 RNA was extracted using phenol-chloroform, as noted above, and samples were fractionated on a denaturing 10 % PAGE gel. The gel was exposed to a phosphor image screen (GE Healthcare) and scanned on a GE Typhoon phosphor imager. The sequences in the degradation pattern were assigned by running TI digestion and partial alkaline hydrolysis in parallel lanes, as noted above.

In–line probing. Reactions were prepared by adding 3′–labeled RNA (8000 cpm) into the binding buffer, pH 8.5, and the indicated concentrations of controls (no ligand, cSS, tS, tDHS, and SAM). Samples were initially incubated at 55 °C for 5 minutes and then cooled at room temperature for 5 minutes, and then incubated at 37 °C for 20 hours. All conditions were quenched in a solution of 7 M Urea and 10 mM EDTA. Were-1 RNA was extracted using phenol-chloroform, as noted above, and run on a denaturing 10 % PAGE gel. The gel was exposed to a phosphor image screen (GE Healthcare), and scanned on a GE Typhoon phosphor imager. The sequences in the degradation pattern were assigned by running T1 digestion and alkaline hydrolysis in parallel lanes, as noted above. All gels were analyzed in ImageJ. Structure predictions of Were-1 in the absence of amino-*t*SS were performed using RNAfold of the Vienna RNA package [36] (Fig. 6.1c).

***In vitro* strand displacement reaction.** A dsDNA reporter was designed to contain a toehold that complements the Shine-Dalgarno sequence of the riboswitch, in which the longer (toehold) strand (Rep F) contained the 3' toehold sequence, a reverse complement of the Shine-Dalgarno sequence, as well as a 5' fluorescein. The shorter strand (Rep Q) contained a 3' Iowa black quencher (Supplementary Table 1). A solution of 2:1 Rep Q:Rep F oligos in binding buffer was incubated at 95 °C for 1 minute, followed by 25 °C for 5 minutes, to anneal the strands and form the dsDNA reporter construct. In a Falcon 384-well Optilux Flat Bottom plate, strand displacement was initiated by adding 100 nM of purified Were-1 RNA to 50 nM of toehold-fluorophore reporter. Amino-*t*SS was quickly added to some samples to test for ligand-dependent displacement. Fluorescence emission was recorded in a BioTek Synergy plate reader over a 45-minute period under continuous illumination using the following parameters: excitation wavelength, 485 nm; emission wavelength, 520 nm.

***In vitro* co–transcriptional toehold-binding kinetics of Were-1**. *In vitro* transcription was performed similarly to the above–described RNA transcription assay with the following modifications: 3 pmol template DNA and 50 nM toehold-fluorophore reporter were used. A 30 μL transcription reaction was initiated by the addition of 4 mM rNTP mix (containing 1mM of each rNTP) and fluorescence emission of the toehold-fluorophore reporter was recorded in a Varian Cary Eclipse fluorimeter under continuous illumination at 37 °C using the following parameters: excitation wavelength, 485 nm; emission wavelength, 520 nm; increment of data point collection, 0.01 s; slit widths, 10 nm. These conditions were used for the entire experiment unless stated otherwise. After an initial fluorescence increase, corresponding to the initial burst of transcription, amino-*t*SS was rapidly added to the solution and fluorescence emission was recorded for 200 s. To switch amino-*t*SS to the *cis* isoform (amino-*c*SS), the solution was excited at 342 nm (slit width, 2.5 nm; $\Phi_q = 6.8 * 10^{-5}$ W/cm^2) for 60 s. Fluorescence emission of the toehold-fluorophore reporter was again recorded for 200 s. To switch the *cis* isoform back to the *trans* state, the solution was excited at 372 nm (slit width, 2.5 nm; $\Phi_q = 10 * 10^{-5}$ W/cm^2) for 60 s. Again, fluorescence emission of the toehold-fluorophore reporter was recorded for 200 s. This process was repeated two to three more times until fluorescence plateaued.

IC_{50} measurements. A dose-response of the Were-1 riboswitch to the target ligand (amino-*t*SS) was assessed by measuring fluorescence as a function of ligand concentration in the presence of a toehold-fluorophore reporter construct (50 nM). Fluorescence emission was recorded under continuous illumination at 37 °C using the following parameters: excitation wavelength, 485 nm; emission wavelength, 520 nm; increment of data point collection, 0.01 s; slit widths, 10 nm. The apparent rate constants were measured and plotted against the amino-*t*SS concentrations (or other ligands, as specified). The data were normalized to the no-amino-*t*SS control. The IC_{50} was extracted from fitting a curve to the graph using the equation:

$$\text{Normalized fluorescence} = 1 - \frac{[\text{ligand}]}{[\text{ligand}] + IC_{50}}$$

***In vitro* co–transcriptional magnesium dependence of Were-1 toehold-binding.** Using the same conditions as above, the fluorescence response of toehold-binding to the Were-1 riboswitch in the presence of 8.4 µM amino-*t*SS under various Mg^{2+} concentrations was measured. Fluorescence emission was recorded under continuous illumination at 37 °C on a BioTek Synergy H1 plate reader.

Cloning the Were-1 riboswitch for expression in *E. coli* cells. Were-1 DNA was cloned into the pBV-Luc (Addgene) vector in order to obtain a fused riboswitch-firefly luciferase (Fluc) reporter construct. The PCR primers were designed to add a 5' *EcoRI* site to the template Were-1 DNA upstream of the T7 promoter and a 3' overhang containing 35 nucleotides of the Fluc gene directly downstream of its start codon to replace the Fluc start codon sequence. Both the PCR product and plasmid were digested by *EcoRI HF* and *KasI* (New England BioLabs) and purified. The purified construct was then inserted at the 5' end of the Fluc coding sequence with T4 DNA ligase (New England

BioLabs). The resulting vector was termed Were-1-Fluc (Table 6.1). Were-1-Fluc was transformed into DH5α *E. coli* cells and grown overnight on agar plates containing ampicillin at 37 °C. Ten colonies were picked from a master plate and individual clones were inoculated overnight in Luria Broth containing ampicillin. Plasmids were purified using a Miniprep Kit (QIAGEN) and individually sequenced (GENEWIZ). Correct constructs were transcribed *in vitro* and fractionated on an agarose gel to confirm sequencing results by measuring the size of the fused construct. Using the same procedure as above, one clone was analyzed in an *in vitro* co-transcriptional toehold-binding experiment to test whether the new fused construct was able to function similarly to the stand-alone riboswitch.

***In vitro* transcription and translation kinetics.** The PURExpress *in vitro* protein synthesis kit (New England BioLabs) was used to transcribe and translate Were-1-Fluc. Experiments were performed similarly to the kit assay conditions with the following modifications: 200 ng/µL DNA, 100 µM D-luciferin, and 2 mM $MgCl_2$. Amino-*t*SS (or other ligands, as specified) was added in conditions when specified. A control plasmid, pET-Luc2, was also tested in the presence and absence of 11 µM amino-*t*SS. All luminescence data were acquired using an ANDOR camera (EMCCD) at 25 °C and analyzed using Solis software, and images were further processed and analyzed using ImageJ. To test whether Were-1 could regulate luciferase protein expression, samples were prepared under identical conditions and luminescence was measured for approximately 40 mins. Samples were then excited at 342 nm ($\Phi_q = 1.4*10^{-2}$ W/cm^2) for 1 s, and luminescence was recorded for approximately 30 mins. Samples were excited at 390 nm ($\Phi_q = 5.5*10^{-2}$

W/cm^2) for 1 s to switch Were-1 back to the bound ‘off’ state, and again, luminescence was measured for approximately 30 mins.

IC$_{50}$ measurements. A dose-response of the Were-1 riboswitch to amino-*t*SS was assessed by measuring luminescence as a function of increasing target concentration. All data were acquired using an ANDOR camera and analyzed with Solis software, and images were further processed and analyzed using ImageJ, as described above.

***In vivo* translation kinetics.** Were-1-Fluc was transformed into BL21(DE3) *E. coli* cells and grown overnight in Luria Broth containing ampicillin (OD$_{600}$ = 0.26). 1 mM IPTG was added to each well (containing 45 µL culture) to induce T7 RNA polymerase-driven expression and 100 µM D-luciferin to provide a substrate for Fluc. Amino-*t*SS or amino-*c*SS was also added where specified. Bioluminescence was recorded every five mins for one hour at 37 °C using a BioTek Synergy H1 plate reader. To test whether the Were-1 riboswitch regulates the production of Fluc, samples were prepared under the same conditions. In the presence of amino-*t*SS, bioluminescence was measured on a BioTek Synergy H1 plate reader for approximately 15 mins before samples were excited at 342 nm for 1s in order to isomerize amino-*t*SS to amino-*c*SS. Luminescence was recorded again for approximately 20 mins. Similar experiments were used regarding amino-*c*SS, with the exception of using 390 nm exposure for 0.5 ms in order to isomerize amino-*c*SS to amino-*t*SS, unless further specified.

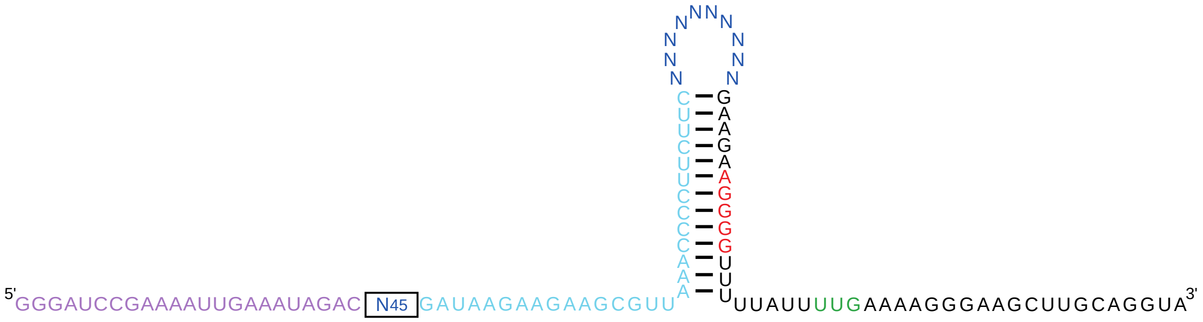
***In vivo* light exposure analysis.** To determine the dependence of the amino-*t*SS exposure on the Were-1-Fluc expression, samples were prepared as described above and loaded into two black-bottom 96-well plates. One plate was used as a control and the other was exposed to 342 nm light ($\Phi_q = 1.4*10^{-2}$ W/cm^2) for their specified time using a

Nikon FM-10 camera shutter in order to isomerize amino-*t*SS to amino-*c*SS. The same procedure was performed in the presence of amino-*c*SS, except 390 nm light was used (Φ_q = 5.5*10^{-2} W/cm^2) for their specified time to isomerize amino-*c*SS to amino-*t*SS. Bioluminescence was measured on an IVIS Lumina II imaging system 1 hr after light exposure. To test whether the Were-1 riboswitch regulates the production of Fluc multiple times, samples were prepared as described above and loaded in a black-bottom 96-well plate. The top half of the plate was used as a control, containing the unresponsive G69C mutant, and the bottom half contained Were-1. All samples contained 10 μM amino-*t*SS and had their bioluminescence was measured 25 minutes after induction on a BioTek Synergy H1 plate reader. Next, the first group of wells remained unexposed to light and the middle and far right samples were exposed to 342 nm (Φ_q = 1.4*10^{-2} W/cm^2) for 1 ms. Measurements for all samples were taken 40 minutes after 342 nm exposure. Finally, the last group of wells (far right) were excited at 390 nm for 0.5 ms. Bioluminescence was measured again, for all samples, 40 minutes after exposure. The same experiment was repeated with the inactive mutant, G69C, as an additional control. Data were normalized to the unexposed samples, and OD_{600} values were obtained to confirm that there was no cell death from UV damage.

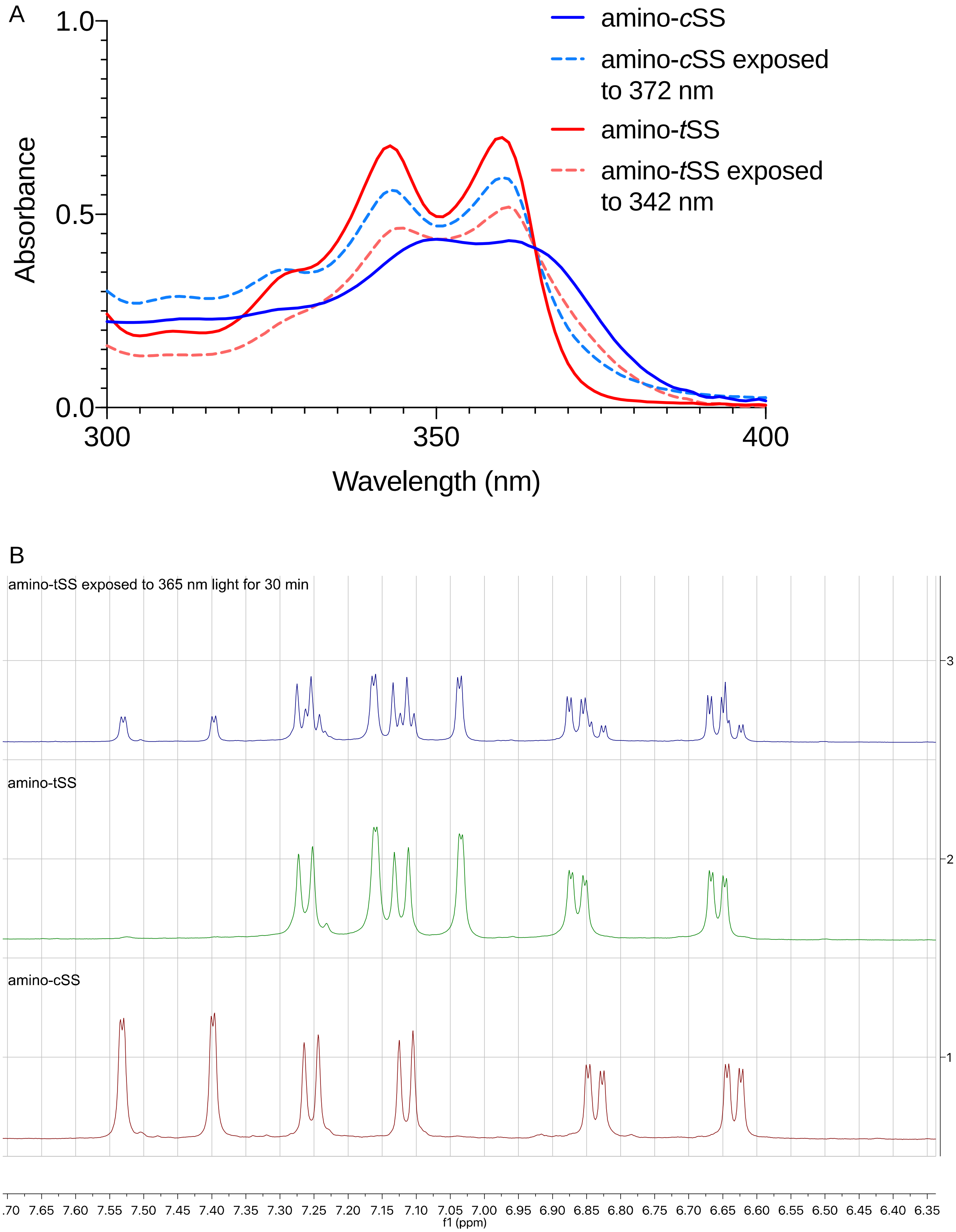
IC_{50} measurements. A dose-response of the Were-1 riboswitch to the target metabolite (amino-*t*SS) was assessed by measuring bioluminescence inhibition as a function of increasing target concentration in BL21(DE3) *E. coli* cells. Bioluminescence was recorded under continuous conditions at 37 °C.

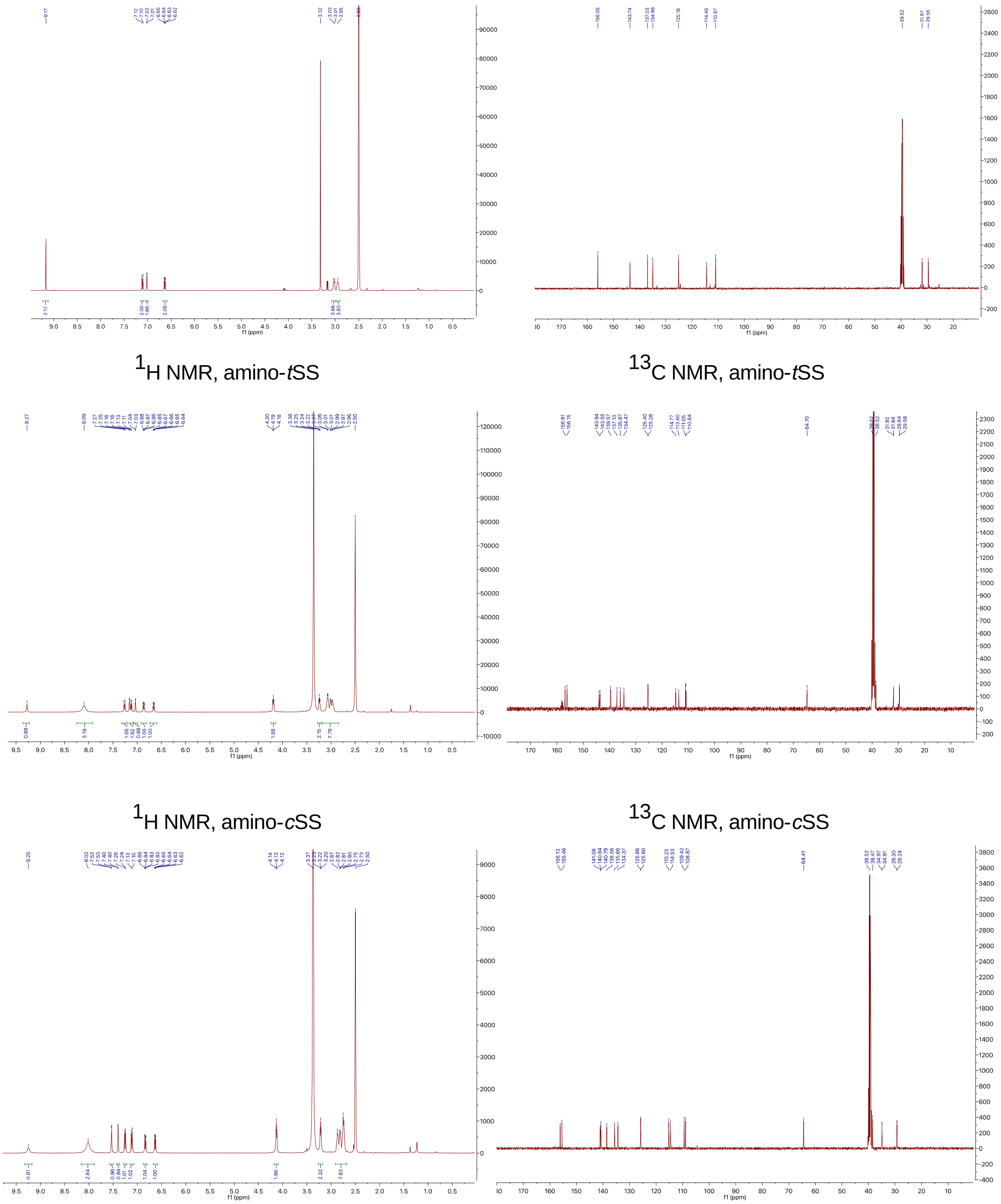
Data Availability: The authors declare that all data are available in the manuscript or the supplementary materials.

6.6 Supplementary Data


5'
GGGAUCCGAAAAUUGAAAUAGAC
N45
GAUAAGAAGAAGCGUU
AAACCCCUUCUUC
NNNNNNNNNN
GAAGAGGGGUUU
UUAUUUUGAAAAGGGAAGCUUGCAGGUA
3'

Supplementary Figure 6.1. *In vitro* selection pool design for a photoriboswitch. An RNA pool was derived from *B. subtilis mswA* SAM-I riboswitch by replacing its ligand-binding domain with a 45-nucleotide random sequence (dark blue, boxed), partially randomizing (at a 15% level) its anti-terminator hairpin and the 5' half of the terminator hairpin (light blue), replacing the terminator-helix tetraloop sequence (UUAU) with a random decamer (dark blue), and retaining the 3' half of the terminator hairpin and its translation initiation sequences (red – Shine-Dalgarno; green – start codon). The pool’s forward primer sequence (in RNA form) is shown in purple.


A
1.0
0.5
0.0
Absorbance
300
350
400
Wavelength (nm)
amino-*c*SS
amino-*c*SS exposed to 372 nm
amino-*t*SS
amino-*t*SS exposed to 342 nm
B
amino-tSS exposed to 365 nm light for 30 min
amino-tSS
amino-cSS
3
2
1
.70 7.65 7.60 7.55 7.50 7.45 7.40 7.35 7.30 7.25 7.20 7.15 7.10 7.05 7.00 6.95 6.90 6.85 6.80 6.75 6.70 6.65 6.60 6.55 6.50 6.45 6.40 6.35
f1 (ppm)



C

^{1}H NMR, *t*SS

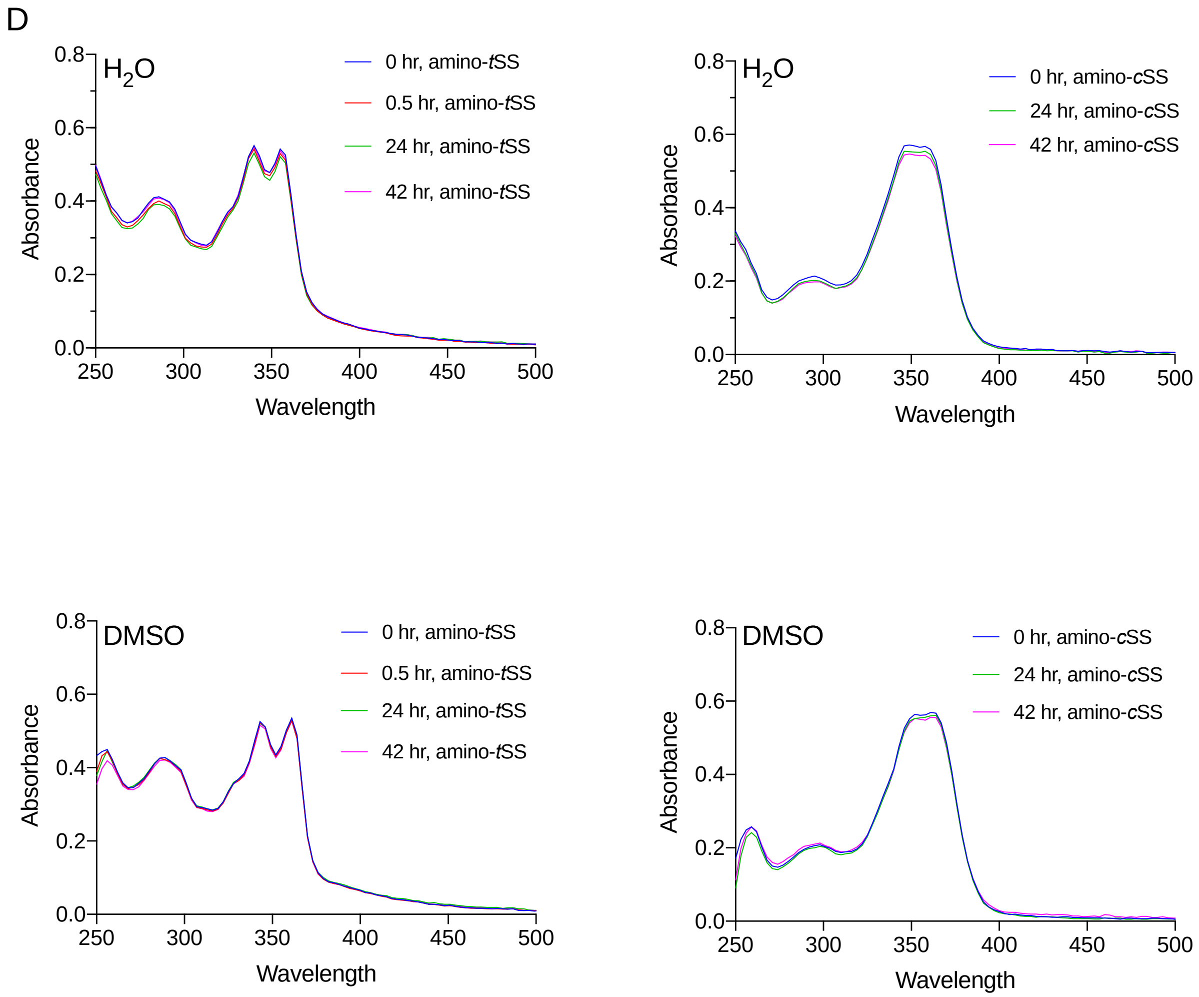
^{13}C NMR, *t*SS

^{1}H NMR, amino-*t*SS

^{13}C NMR, amino-*t*SS

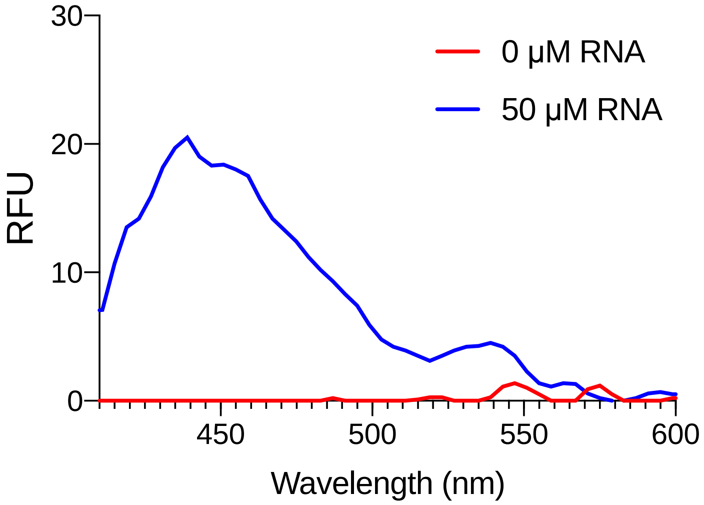
^{1}H NMR, amino-*c*SS

^{13}C NMR, amino-*c*SS

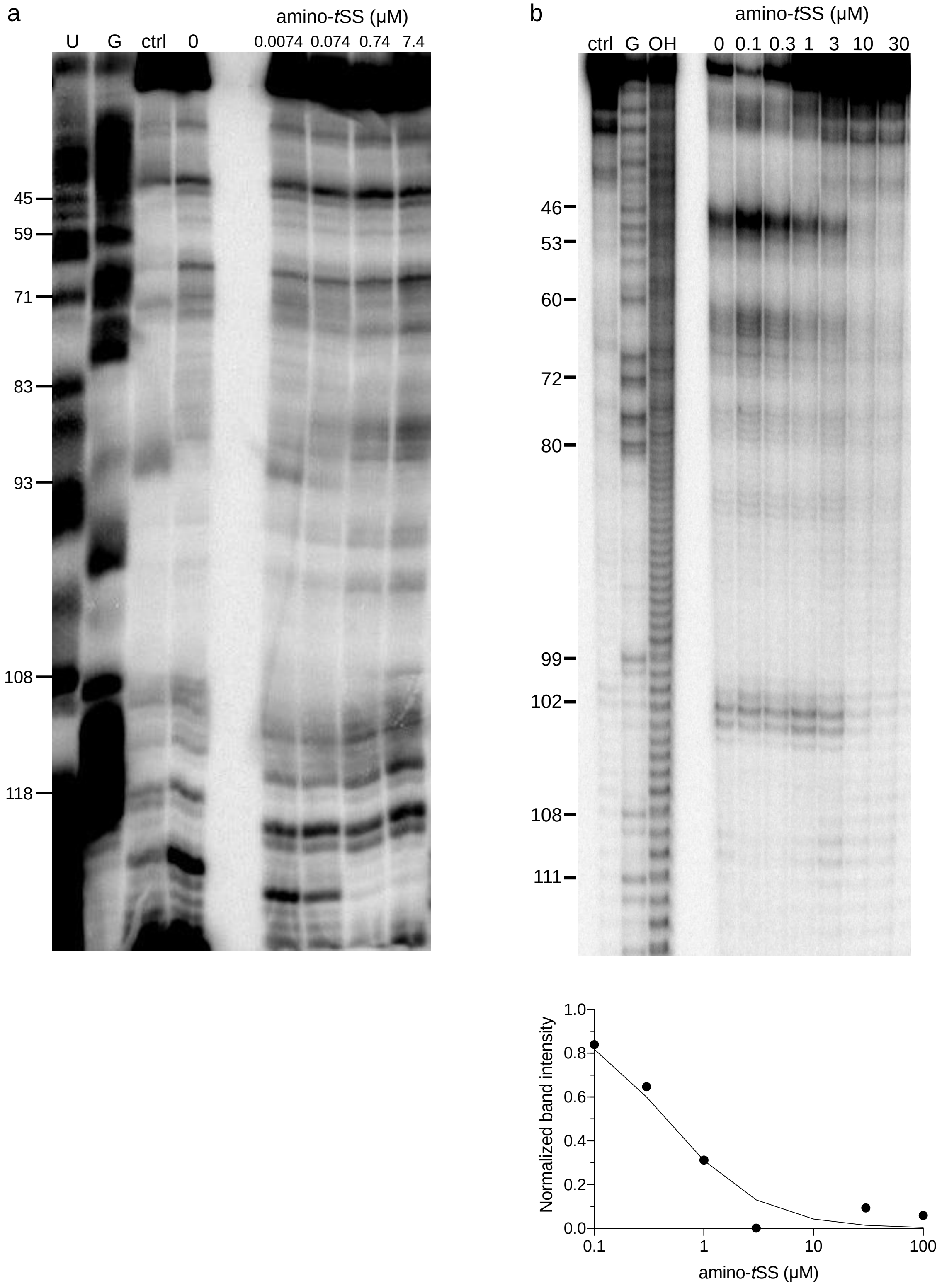


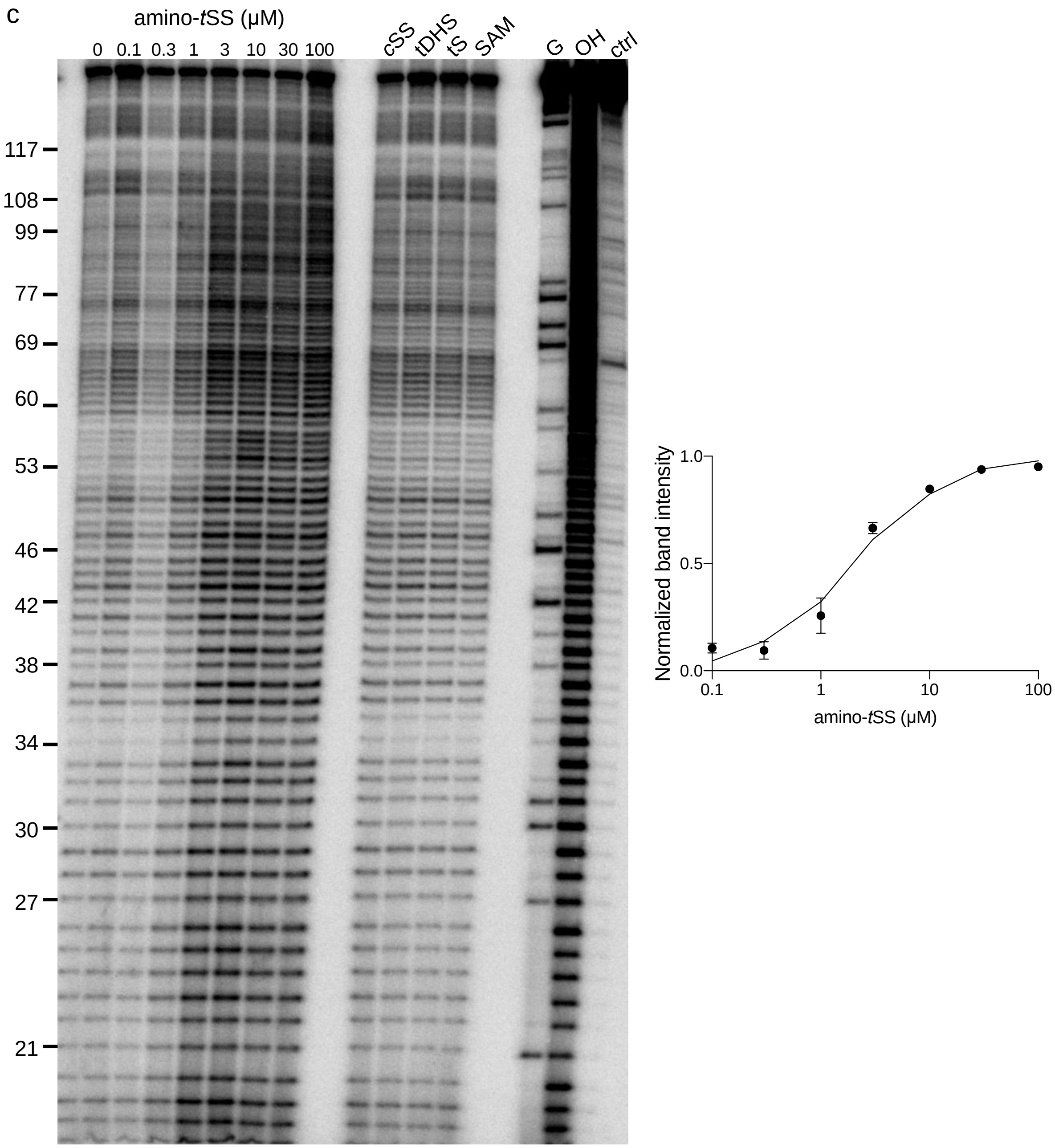
Supplementary Figure 6.2. E/Z isomerization of amino-*t*SS and amino-*c*SS. a, UV-Vis spectra of amino-*t*SS (red) and amino-*t*SS after 3 h exposure to 342 nm (red, dashed), and amino-*c*SS (blue) and amino-*c*SS after 1.5 h exposure to 372 nm (blue, dashed). Photo isomerization was performed in 30 mM DMSO solutions and subsequently diluted 1000 times for UV-Vis spectra measurement. **b**, ^{1}H NMR of amino-*t*SS, amino-*c*SS, and E/Z isomerization. At the photostationary state (amino-*t*SS exposed to 365 nm) the ratio of amino-*t*SS to amino-*c*SS is approximately 2:1 as determined by NMR. **c**, ^{1}H NMR and ^{13}C NMR spectra of *t*SS (top), amino-*t*SS (middle), and amino-*c*SS (bottom). **d,** UV-Vis spectra of amino-*t*SS and amino-*c*SS in H_2O (top) and DMSO (bottom). Both isoforms of the ligand are relatively stable in either solvent for the indicated times.

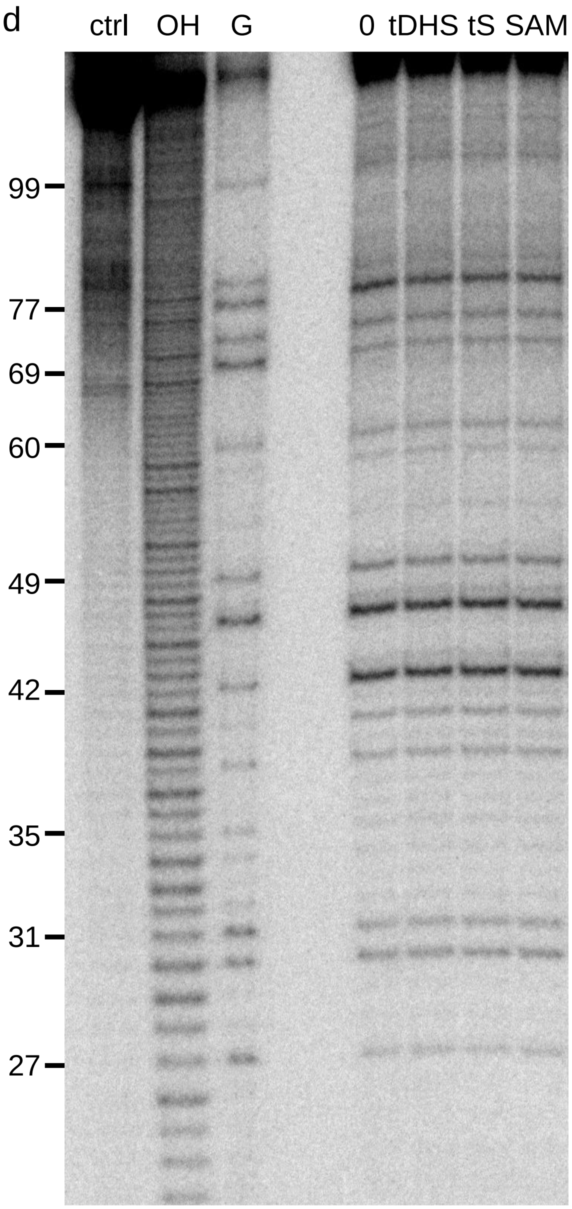
Supplementary Figure 6.3. Screening of Were-1 based on optical activity.

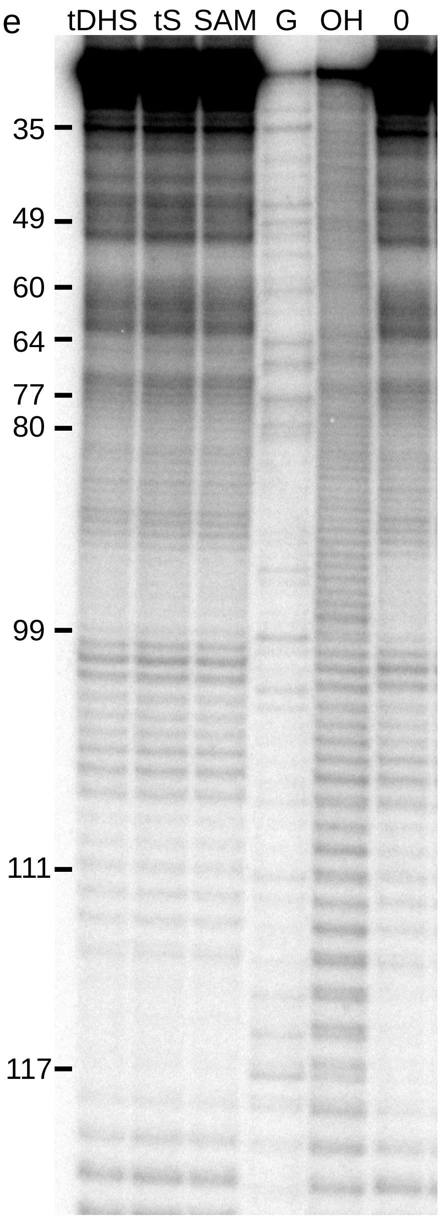


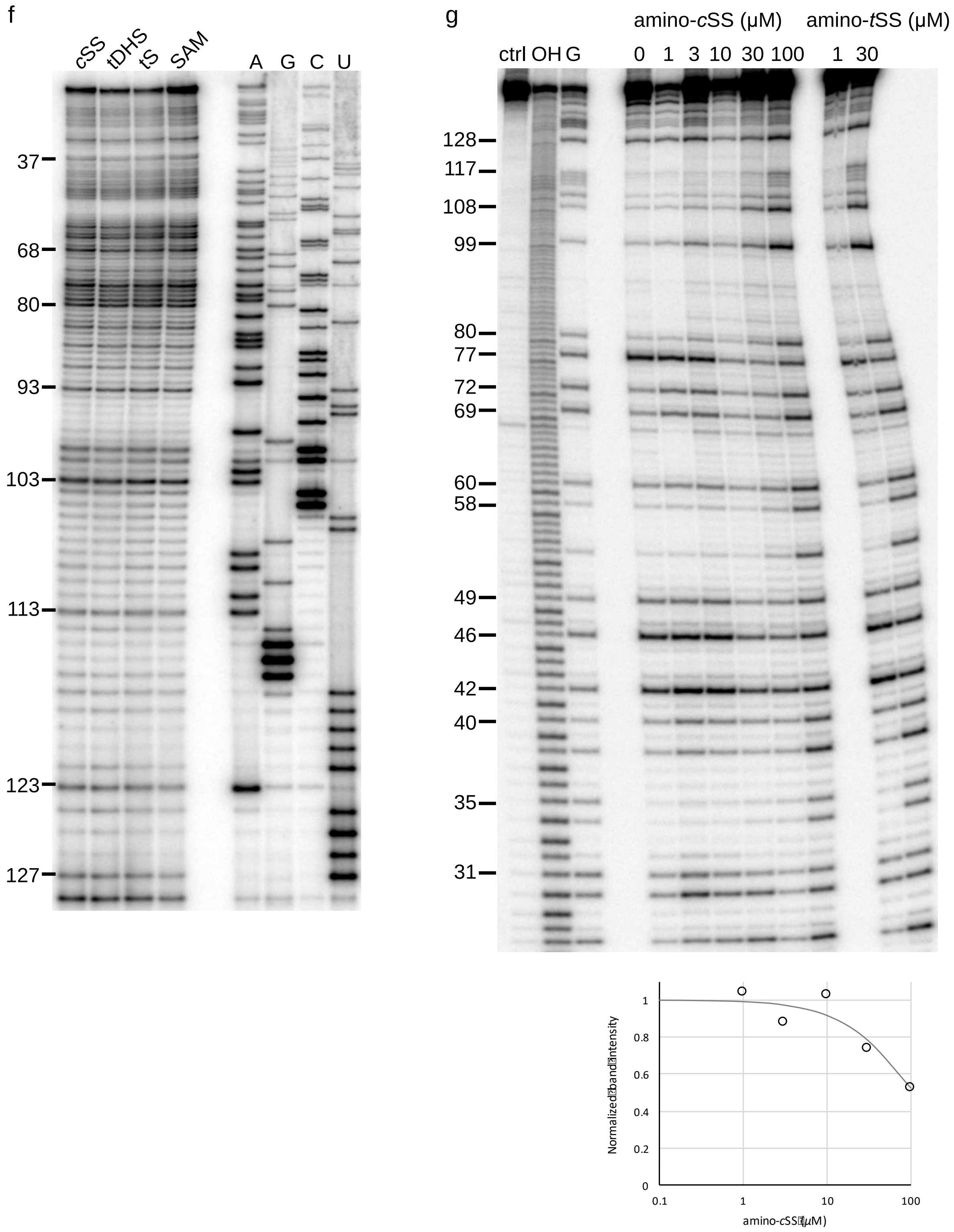
Fluorescence emission of 100 nM amino-*t*SS incubated with purified Were-1 RNA showing an increase in presence of Were-1 and suggesting RNA affinity for the target ligand. Emission spectra were collected using an excitation of 365/10 nm.


a
amino-tSS (μM)
U G ctrl 0 0.0074 0.074 0.74 7.4
45
59
71
83
93
108
118
b
amino-tSS (μM)
ctrl G OH 0 0.1 0.3 1 3 10 30
46
53
60
72
80
99
102
108
111
Normalized band intensity
1.0
0.8
0.6
0.4
0.2
0.0
0.1
1
10
100
amino-tSS (μM)

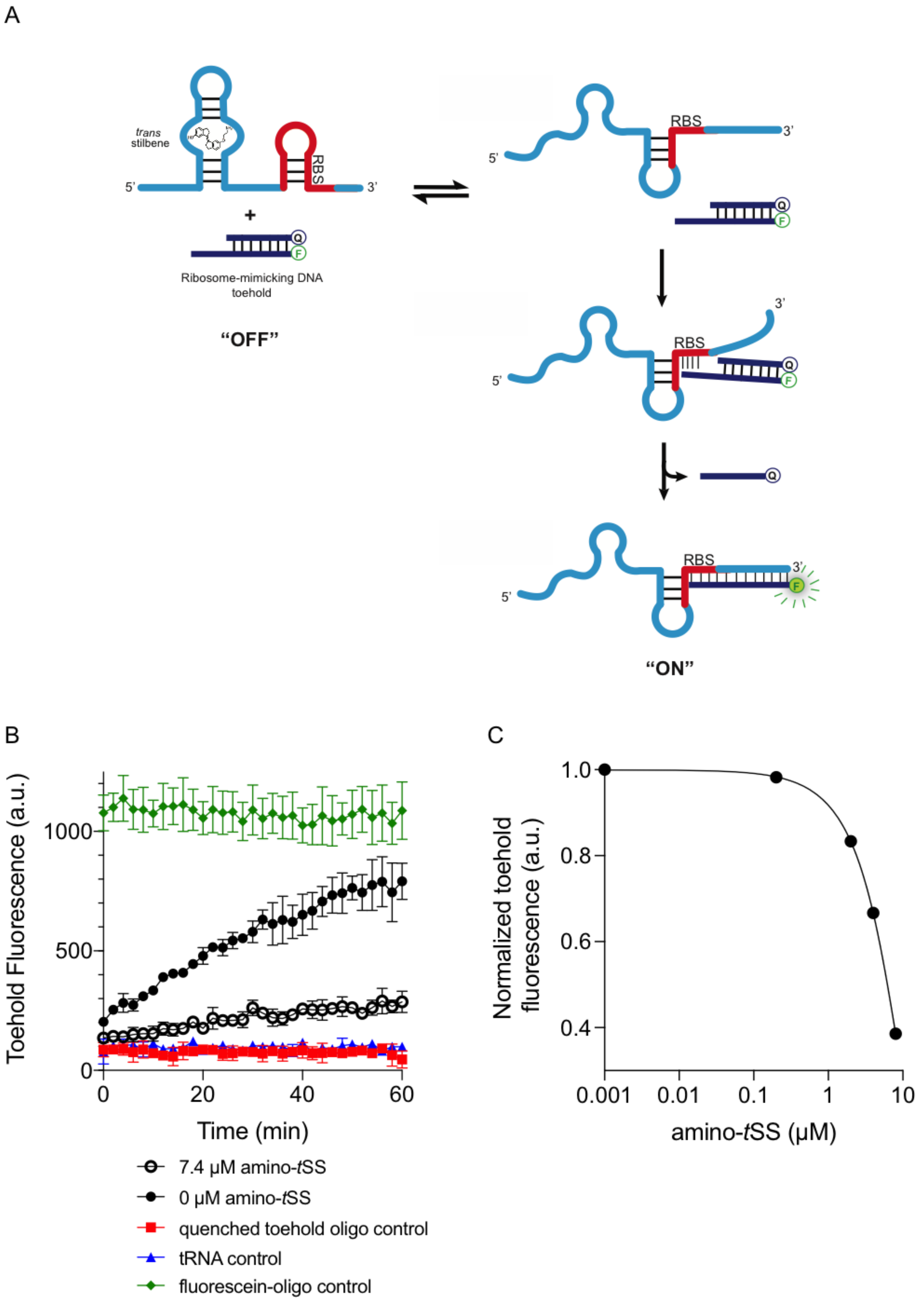

c
amino-*t*SS (µM)
0
0.1
0.3
1
3
10
30
100
*c*SS
*t*DHS
*t*S
SAM
G
OH
ctrl
117
108
99
77
69
60
53
46
42
38
34
30
27
21
Normalized band intensity
1.0
0.5
0.0
0.1
1
10
100
amino-*t*SS (µM)


d
ctrl
OH
G
0
tDHS
tS
SAM
99
77
69
60
49
42
35
31
27

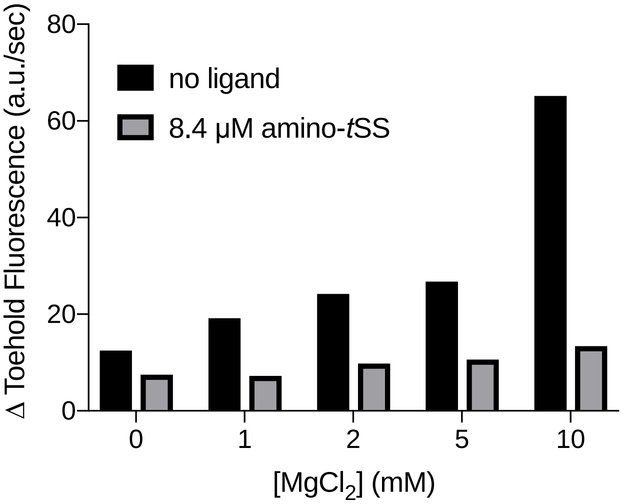

e
tDHS
tS
SAM
G
OH
0
35
49
60
64
77
80
99
111
117


f
cSS
tDHS
tS
SAM
A
G
C
U
37
68
80
93
103
113
123
127
g
amino-cSS (μM)
amino-tSS (μM)
ctrl
OH
G
0
1
3
10
30
100
1
30
128
117
108
99
80
77
72
69
60
58
49
46
42
40
35
31
Normalized band intensity
1
0.8
0.6
0.4
0.2
0
0.1
1
10
100
amino-cSS (μM)

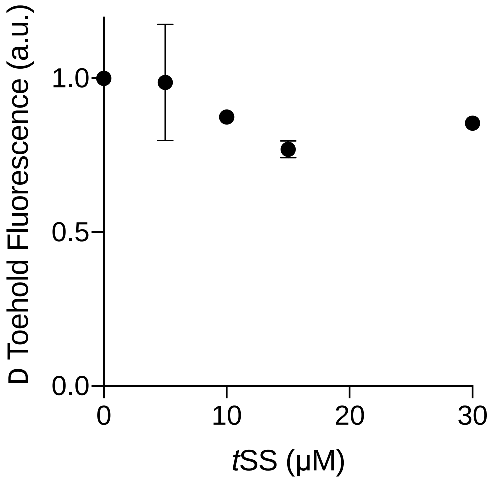
Supplementary Figure 6.4. Structural probing of Werewolf-1. a, SHAPE analysis of Were-1. Left lanes show reverse transcriptase (RT) stops due to ddA (“U”) and ddC (“G”) incorporation for sequence reference, followed by a no-acylation RT control (ctrl) lane, and SHAPE reactions with amino-*t*SS at concentrations indicated above the gel image. **b,** S1 digestion of Were-1. Left lanes contain a control with no RNA (ctrl), RNase T1 digestion for sequence reference, and a partial hydrolysis lane (OH). The right lanes show S1 digestion in the presence of increasing amino-*t*SS at concentrations indicated above the gel image. Analysis of the band intensity change at positions A44-G46 revealed an apparent K_D value of 0.4 μM (below). **c,** Terbium (III) footprinting of Were-1 in the presence of increasing amino-*t*SS. Left lanes show RNase T1 digestion for sequence reference, and partial hydrolysis (OH). Middle lanes show Tb^{3+} footprinting in the presence of increasing amino-*t*SS at concentrations indicated above the gel image. Right lanes show control reactions with Were-1 in the presence of 30 μM controls, as labeled above the gel image. Right, a K_D value of 4.8 µM was calculated based on the change in intensity with increasing amino-*t*SS at nucleotides A113 and U107. **d,** T1 digestion controls. Left lanes contain a control with no RNA, partial hydrolysis (OH), and RNase T1 digestion for sequence reference. Right lanes indicate no ligand (0), followed by Were-1 in the presence of 30 μM controls, as labeled above the gel image. **e,** In-line probing of Were-1 with 30 μM controls, as labeled above the gel image, followed by RNase T1 digestion for sequence reference, partial hydrolysis (OH), and a no ligand control (0). **f,** SHAPE of Were-1 in presence of control small molecules. Left lanes show RT stops for Were-1 in the presence of 30 µM amino-*c*SS, tDHS, tS, and SAM. Right lanes show RT stops due to ddT (“A”), ddC (“G”), ddG (“C”), and ddA (“U”) incorporation for sequence reference. **g,** T1 digestion of Were-1 in the presence of increasing amino-*c*SS. Left lanes show undigested RNA, partial hydrolysis, and a G-specific sequencing lane. Middle lanes show Were-1 digestion in the presence of increasing amino-*c*SS and the right lane show amino-*t*SS experiments at low and high concentrations, for direct comparison of the ligand-induced structural probing. At high concentration (100 µM) amino-*c*SS mimics the bound profile of amino-*t*SS. The apparent K_D calculation for *c*SS is shown below the gel image and in Fig. 6.1b, revealing a K_D of 108 μM.


A
trans stilbene
5'
3'
RBS
+
Q
F
Ribosome-mimicking DNA toehold
"OFF"
RBS
3'
5'
Q
F
3'
RBS
5'
Q
F
Q
RBS
3'
5'
F
"ON"
B
Toehold Fluorescence (a.u.)
1000
500
0
0
20
40
60
Time (min)
7.4 μM amino-tSS
0 μM amino-tSS
quenched toehold oligo control
tRNA control
fluorescein-oligo control
C
Normalized toehold fluorescence (a.u.)
1.0
0.8
0.6
0.4
0.001
0.01
0.1
1
10
amino-tSS (μM)

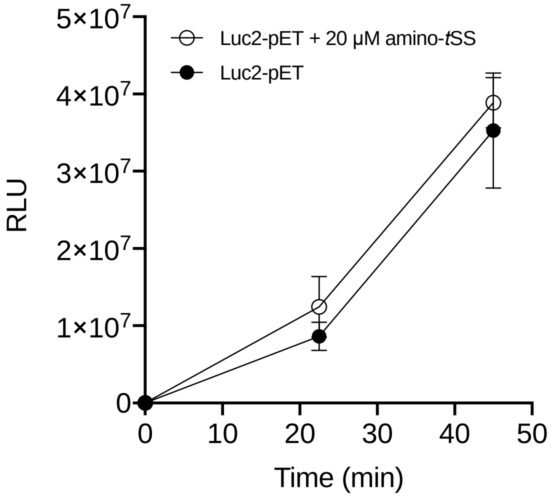
Supplementary Figure 6.5. Binding PAGE-purified Were-1 RNA to a ribosome-mimic. a, Schematic of strand displacement, where a DNA duplex containing a toehold complementary to the Shine-Dalgarno sequence of Were-1 is displaced when no amino-*t*SS is present, producing a fluorescence. **b**, The presence of amino-*t*SS prevents the quencher strand release, suppressing toehold fluorescence (± SD) over time. Control reactions were performed using the unquenched fluorophore oligo (green) and the quenched toehold oligo (red) for fully quenched fluorescence to provide a window for maximum fluorescence, and a tRNA (blue) lacking the Shine-Dalgarno sequence as a negative control. **c**, Dose-response curve of PAGE-purified Were-1 RNA in the presence of increasing concentrations of amino-*t*SS and analyzed by the toehold-fluorescence assay. Half maximum fluorescence is observed at 6.3 μM amino-*t*SS.



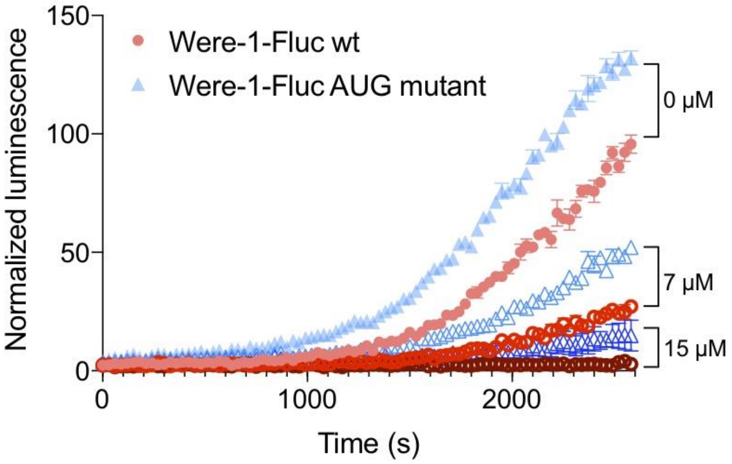
Supplementary Figure 6.6. Co-transcriptional binding of a ribosome-mimic *in vitro* under various Mg^{2+} conditions. Whereas the rate of transcription increases with magnesium (black), the bound state (gray) does not change significantly, suggesting that co-transcriptional binding is minimally affected by magnesium concentration.



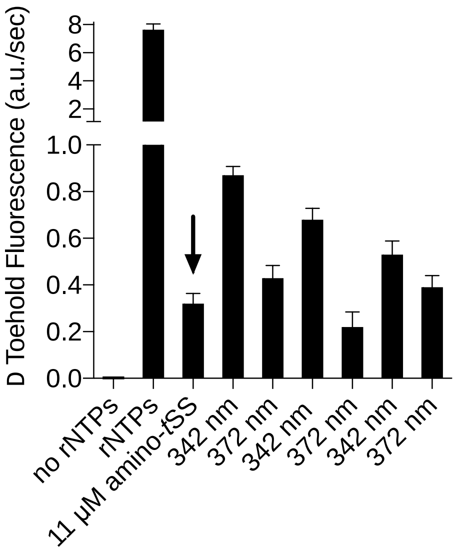
Supplementary Figure 6.7. Co-transcriptional Were-1 binding of a ribosome-mimic *in vitro* in the presence of *t*SS (in 10 % DMSO). A minor dose-response is observed in the presence of increasing *t*SS concentrations, but full inhibition, as seen for amino-*t*SS, is not observed, possibly due to low solubility of *t*SS.



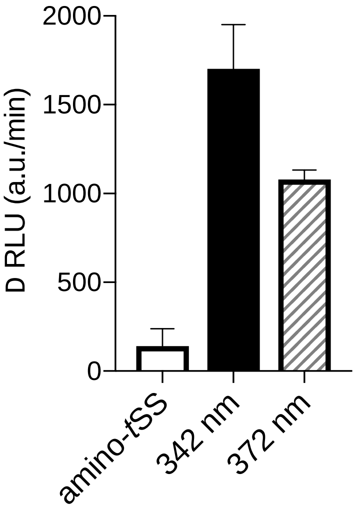
Supplementary Figure 6.8. Translation of a control plasmid, Luc2-pET, lacking the Were-1 riboswitch is not inhibited *in vitro* by amino-*t*SS. There is no significant difference in Luc2-pET luminescence (± SD) in the presence and absence of amino-*t*SS, suggesting that amino-*t*SS does not inhibit the *in vitro* translation system.



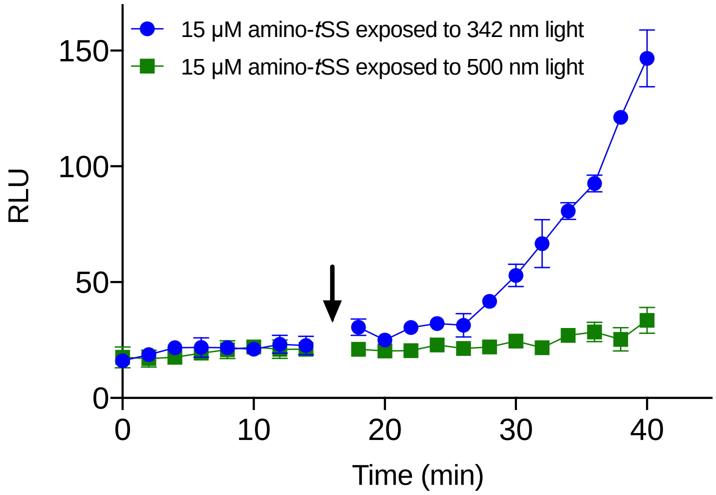
Supplementary Figure 6.9. The effect of the canonical start codon on Were-1-Fluc expression *in vivo*. Comparison of luciferase expression (± SEM) of Were-1-Fluc wild-type (wt; red) and Were-1-Fluc containing an AUG start codon (Were-1-Fluc AUG mutant; blue), in place of the pool-derived UUG minor start codon, with increasing amino-*t*SS concentrations. Bioluminescence was overall higher in the Were-1-Fluc AUG mutants for all conditions but retained the wt dose-dependent response.



Supplementary Figure 6.10. Co-transcriptional binding of a ribosome-mimic *in vitro* under amino-*t*SS photoisomerization conditions. Initial transcription of Were-1 without ligand was induced by adding ribonucleoside triphosphates (rNTPs) to a transcription mix, producing robust toehold fluorescence increase (see Fig. 6.2b for examples of co-transcriptional toehold binding). Upon adding 11 μM amino-*t*SS (arrow), fluorescence growth decreased, suggesting that Were-1 is bound to the ligand and prevents toehold binding. The sample was then exposed to 342 nm ($\Phi_q = 6.8*10^{-5}$ W/cm^2) for 60 s, causing photoisomerization of the ligand, and resulting in increased toehold fluorescence slope. Were-1 transcription reaction was subsequently exposed to 372 nm light ($\Phi_q = 1*10^{-4}$ W/cm^2) for 60 s, to switch the ligand from *cis* to *trans* conformation, which resulted in decrease of toehold fluorescence slope. This was repeated two more times, as indicated, until the transcription reaction plateaued, and all toehold was expended. The data suggest that the system can be regulated multiple times in one reaction. Error bars represent fluorescence slope error.



Supplementary Figure 6.11. Were-1 regulates translation *in vitro*. *In vitro* transcription-translation reaction of Were-1-Fluc mRNA in presence of amino-*t*SS (left column) was irradiated at 342 nm for 60 s, resulting in an increase in luciferase production (± SEM). The reaction was then irradiated at 372 nm for 60 s, resulting in a decrease in luciferase expression rate (± SEM) over time.



Supplementary Figure 6.12. Were-1 regulates protein expression *in vivo*. Were-1 *E. coli* bioluminescence (± SEM) was measured in the presence of amino-*t*SS. Initial Fluc expression showed identical increases in bioluminescence for all samples. In the presence of 15 μM amino-*t*SS, samples were either exposed to (arrow) 342 nm light (blue) to isomerize amino-*t*SS, or 500 nm (green), a wavelength that does not affect amino-*t*SS isomerization. 342-nm–exposed samples showed significantly higher bioluminescence compared to those exposed to 500 nm light.

Supplementary Table 6.1. DNA sequences used for *in vitro* and *in vivo* analysis of the Were-1 riboswitch.

Toehold-Fluorophore Reporter	Rep F: 5'– 5FluorT/TA CCTGCAAGCTTCCCTTTTCAAAATAAAAACCCCT –3' Rep Q: 5'– ATTTTGAAAAGGGAAGCTTGCAGGTA/3IABkFQ –3'
Forward Overhang Primer (AL2909)	5'– CCCgaattcTAATACGACTCACTATAGGGATCCGAAAATTGAAATAGAC – 3'
Reverse Overhang Primer (AL2912)	5'– atggcgccgggcctttctttatgtttttggcgtcttCCCTGCAAGCTTCCCTTTTCAAAATAAAAAC –3'
Forward Werewolf Primer (AL2881)	5'– GGGATCCGAAAATTGAAATAGACTCC –3'
Reverse Nested Fluc Primer (AL1649)	5'– gattctgtgatttgtattcagcccata –3'
Pool Forward Primer (AL2045)	5'– TAATACGACTCACTATAGGGATCCGAAAATTGAAATAGAC –3'
Pool Reverse Primer (AL2049)	5'– TACCTGCAAGCTTCCCTTTTC –3'
AUG mutant Forward Primer (AL3267)	5'– 5Phos/ATGAAAAGGGAAGCTTGCAGGgaagac –3'
AUG mutant Reverse Primer (AL3268)	5'– /5Phos/AA TAA AAA CCC CTT CAA GGT TGG CTG AAG –3'
C89G mutant Forward Primer (AL3330)	5'– /5Phos/GACATCTTCAGCCAACCTTG – 3'
C89G mutant Reverse Primer (AL3331)	5'– /5Phos/GTTTGATGCTTCTGGTCATCTG –3'
G69C mutant Forward Primer (AL3328)	5'– /5Phos/CATGACCAGAAGCATCAAACCAC –3'
G69C mutant Reverse Primer (AL3329)	5'– /5Phos/TGGTAATTCACGGTGCTTACTGC –3'
C91G mutant Forward Primer (AL3326)	5'– /5Phos/GATCTTCAGCCAACCTTGAAGAAGG –3'
C91G mutant Reverse Primer (AL3327)	5'– /5Phos/TGGTTTGATGCTTCTGGTCATCTG –3'
C84G/G69C compensatory mutant Forward Primer (AL3384)	5'– /5Phos/GAAACCACATCTTCAGCCAACCTT–3'
C84G/G69C compensatory mutant Reverse Primer (AL3385)	5'– /5Phos/ATGCTTCTGGTCATGTGGTAATTCAC –3'
Were-1 DNA	5'– GGGATCCGAAAATTGAAATAGACTCCGATGGACGGTTGCGTGTATGCAGTAAGCACCGTGAATTACCAGATGACCAGAAGCATCAAACCACATCTTCAGCCAACCTTGAAGAAGGGGTTTTTATTTTGAAAAGGGAAGCTTGCAGGTA –3'

Amino-*t*SS pool	5'–TAATACGACTCACTATAGGGATCCGAAAATTGAAATAGAC111111111111111111111111111111111111111111111452554554554342255533332232231111111111GAAGAAGGGGTTTTTATTTTGAAAAGGGAAGCTTGCAGGTA–3'
Were-1-Fluc construct	5'–TAATACGACTCACTATAGGGATCCGAAAATTGAAATAGACTCCGATGGACGGTTGCGTGTATGCAGTAAGCACCGTGAATTACCAGATGACCAGAAGCATCAAACCACATCTTCAGCCAACCTTGAAGAAGGGGTTTTTATTTTGAAAAGGGAAGCTTGCAGGgaagacgccaaaaacataaagaaaggcccggcgccattctatccgctggaagatggaaccgctggagagcaactgcataaggctatgaagagatacgccctggttcctggaacaattgcttttacagatgcacatatcgaggtggacatcacttacgctgagtacttcgaaatgtccgttcggttggcagaagctatgaaacgatatgggctgaatacaaatcacagaatcgtcgtatgcagtgaaaactctcttcaattctttatgccggtgttgggcgcgttatttatcggagttgcagttgcgcccgcgaacgacatttataatgaacgtgaattgctcaacagtatgggcatttcgcagcctaccgtggtgttcgtttccaaaaaggggttgcaaaaaattttgaacgtgcaaaaaaagctcccaatcatccaaaaaattattatcatggattctaaaacggattaccagggatttcagtcgatgtacacgttcgtcacatctcatctacctcccggttttaatgaatacgattttgtgccagagtccttcgatagggacaagacaattgcactgatcatgaactcctctggatctactggtctgcctaaaggtgtcgctctgcctcatagaactgcctgcgtgagattctcgcatgccagagatcctatttttggcaatcaaatcattccggatactgcgattttaagtgttgttccattccatcacggttttggaatgtttactacactcggatatttgatatgtggatttcgagtcgtcttaatgtatagatttgaagaagagctgtttctgaggagccttcaggattacaagattcaaagtgcgctgctggtgccaaccctattctccttcttcgccaaaagcactctgattgacaaatacgatttatctaatttacacgaaattgcttctggtggcgctcccctctctaaggaagtcggggaagcggttgccaagaggttccatctgccaggtatcaggcaaggatatgggctcactgagactacatcagctattctgattacacccgagggggatgataaaccgggcgcggtcggtaaagttgttccattttttgaagcgaaggttgtggatctggataccgggaaaacgctgggcgttaatcaaagaggcgaactgtgtgtgagaggtcctatgattatgtccggttatgtaaacaatccggaagcgaccaacgccttgattgacaaggatggatggctacattctggagacatagcttactgggacgaagacgaacacttcttcatcgttgaccgcctgaagtctctgattaagtacaaaggctatcaggtggctcccgctgaattggaatccatcttgctccaacaccccaacatcttcgacgcaggtgtcgcaggtcttcccgacgatgacgccggtgaacttcccgccgccgttgttgttttggagcacggaaagacgatgacggaaaaagagatcgtggattacgtcgccagtcaagtaacaaccgcgaaaaagttgcgcggaggagttgtgtttgtggacgaagtaccgaaaggtcttaccggaaaactcgacgcaagaaaaatcagagagatcctcataaaggccaagaagggcggaaagatcgccgtgtaa–3'

Abbreviations/Key

3IABkFQ – 3' Iowa black fluorophore quencher

5FluorT – 5' Fluorescein

1 A:G:C:T=1:1:1:1

2 A:G:C:T=17:1:1:1

3 A:G:C:T=1:17:1:1

4 A:G:C:T=1:1:17:1

5 A:G:C:T=1:1:1:17

6.7 References

1. Fenno, L., Yizhar, O. & Deisseroth, K. The Development and Application of Optogenetics. *Annu. Rev. Neurosci.* **34,** 389–412 (2011).
2. Möglich, A. & Moffat, K. Engineered photoreceptors as novel optogenetic tools. *Photochem. Photobiol. Sci.* **9,** 1286 (2010).
3. Motta-Mena, L. B. *et al.* An optogenetic gene expression system with rapid activation and deactivation kinetics. *Nat. Chem. Biol.* **10,** 196–202 (2014).
4. Cambridge, S. B. *et al.* Doxycycline-dependent photoactivated gene expression in eukaryotic systems. *Nat. Methods* **6,** 527–531 (2009).
5. Dhamodharan, V., Nomura, Y., Dwidar, M. & Yokobayashi, Y. Optochemical control of gene expression by photocaged guanine and riboswitches. *Chem. Commun.* **54,** 6181–6183 (2018).
6. Wulffen, B., Buff, M. C. R., Pofahl, M., Mayer, G. & Heckel, A. Caged glucosamine-6-phosphate for the light-control of riboswitch activity. *Photochem. Photobiol. Sci.* **11,** 489–492 (2012).
7. Walsh, S., Gardner, L., Deiters, A. & Williams, G. J. Intracellular Light-Activation of Riboswitch Activity. *ChemBioChem* **15,** 1346–1351 (2014).
8. Young, D. D., Garner, R. A., Yoder, J. A. & Deiters, A. Light-activation of gene function in mammalian cells via ribozymes. *Chem. Commun.* **5**, 568–70 (2009).
9. Chaulk, S. & MacMillan, A. M. Caged RNA: photo-control of a ribozyme reaction. *Nucleic Acids Res.* **26,** 3173–3178 (1998).
10. Lucas, T. *et al.* Light-inducible antimiR-92a as a therapeutic strategy to promote skin repair in healing-impaired diabetic mice. *Nat. Commun.* **8,** 15162 (2017).
11. Lee, H.-W., Robinson, S. G., Bandyopadhyay, S., Mitchell, R. H. & Sen, D. Reversible Photo-regulation of a Hammerhead Ribozyme Using a Diffusible Effector. *J. Mol. Biol.* **371,** 1163–1173 (2007).
12. Hayashi, G., Hagihara, M. & Nakatani, K. RNA Aptamers That Reversibly Bind Photoresponsive Azobenzene-Containing Peptides. *Chem. - A Eur. J.* **15,** 424–432 (2009).
13. Lotz, T. S. *et al.* A light-responsive RNA aptamer for an azobenzene derivative. *Nucleic Acids Res.* **47**, 2029-2040 (2018).
14. Young, D. D. & Deiters, A. Light-Regulated RNA-Small Molecule Interactions. *ChemBioChem* **9,** 1225–1228 (2008).
15. Tuerk, C. & Gold, L. Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science* **249,** 505–10 (1990).
16. Ellington, A. D. & Szostak, J. W. In vitro selection of RNA molecules that bind specific ligands. *Nature* **346,** 818–822 (1990).
17. Martini, L. *et al. In Vitro* Selection for Small-Molecule-Triggered Strand Displacement and Riboswitch Activity. *ACS Synth. Biol.* **4,** 1144–1150 (2015).
18. Quick, M. *et al.* Photoisomerization Dynamics of Stiff-Stilbene in Solution. *J. Phys. Chem. B* **118,** 1389–1402 (2014).

19. Waldeck, D. H. Photoisomerization Dynamics of Stilbenes. *Chem. Rev* **91,** 415–436 (1991).
20. Szymanski, W., Beierle, J. M., Kistemaker, H. A. V, Velema, W. A. & Feringa, B. L. Reversible Photocontrol of Biological Systems by the Incorporation of Molecular Photoswitches. *Chem. Rev.* **113,** 6114–6178 (2013).
21. You, M. & Jaffrey, S. R. Designing optogenetically controlled RNA for regulating biological systems. *Ann. N. Y. Acad. Sci.* **1352,** 13–19 (2015).
22. Liang, J. C., Bloom, R. J. & Smolke, C. D. Engineering Biological Systems with Synthetic RNA Molecules. *Mol. Cell* **43,** 915–926 (2011).
23. Jäschke, A. Genetically encoded RNA photoswitches as tools for the control of gene expression. *FEBS Lett.* **586,** 2106–2111 (2012).
24. Winkler, W. C., Nahvi, A., Sudarsan, N., Barrick, J. E. & Breaker, R. R. An mRNA structure that controls gene expression by binding S-adenosylmethionine. *Nat. Struct. Biol.* **10,** 701–707 (2003).
25. Peng, Y., Soper, T. J. & Woodson, S. A. RNase footprinting of protein binding sites on an mRNA target of small RNAs. *Methods Mol. Biol.* **905,** 213–24 (2012).
26. Reynolds, W. F. & Gottesfeld, J. M. Torsional stress induces an S1 nuclease-hypersensitive site within the promoter of the Xenopus laevis oocyte-type 5S RNA gene. *Proc. Natl. Acad. Sci. U. S. A.* **82,** 4018 (1985).
27. Harris, D.A., Todd, G.C. and Walter, N.G. Terbium(III) footprinting as a probe of RNA structure and metal binding sites. *In Handbook of RNA Biochemistry* **Part II**, 255-268 (2014).
28. Regulski, E. E. & Breaker, R. R. In-line probing analysis of riboswitches. *Methods Mol Biol.* **419**, 53-67 (2008).
29. Wilkinson, K. A., Merino, E. J. & Weeks, K. M. Selective 2′-hydroxyl acylation analyzed by primer extension (SHAPE): quantitative RNA structure analysis at single nucleotide resolution. *Nat. Protoc.* **1,** 1610–1616 (2006).
30. Martini, L., Ellington, A. D. & Mansy, S. S. An in vitro selection for small molecule induced switching RNA molecules. *Methods* **106,** 51–57 (2016).
31. Zhang, D. Y. & Winfree, E. Control of DNA Strand Displacement Kinetics Using Toehold Exchange. *J. Am. Chem. Soc.* **131,** 17303–17314 (2009).
32. Frieda, K. L. & Block, S. M. Direct observation of cotranscriptional folding in an adenine riboswitch. *Science* **338,** 397–400 (2012).
33. Grundy, F. J. & Henkin, T. M. The S box regulon: a new global transcription termination control system for methionine and cysteine biosynthesis genes in gram-positive bacteria. *Mol. Microbiol.* **30,** 737–49 (1998).
34. Band, L. & Henner, D. J. Bacillus subtilis requires a ‘stringent’ Shine-Dalgarno region for gene expression. *DNA* **3,** 17–21 (1984).
35. Spitale, R. C. *et al.* RNA SHAPE analysis in living cells. *Nat. Chem. Biol.* **9,** 18–20 (2013).