

Development, characterization, and application of RNA catalysts for *in situ* labeling of target RNA molecules



Dissertation

for the award of the degree

“Doctor rerum naturalium”

of the Georg-August-Universität Göttingen

within the doctoral program IMPRS Molecular Biology

of the Georg-August University School of Science (GAUSS)

submitted by

Mohammad Ghaem Maghami

from Tehran, Iran

Göttingen, June 2020

Thesis Committee

Prof. Dr. Claudia Höbartner, Organic and Biomolecular Chemistry, Institute for Organic Chemistry, University of Würzburg.

Prof. Dr. Marina Rodnina, Biophysical Chemistry, Max Planck Institute for Biophysical Chemistry, Göttingen

Prof. Dr. Stefan Jakobs, Mitochondrial Structure and Dynamics, Max Planck Institute for Biophysical chemistry, Göttingen

Members of the Examination Board

Referee: Prof. Dr. Claudia Höbartner, Organic and Biomolecular Chemistry, Institute for Organic Chemistry, University of Würzburg.

2nd Referee: Prof. Dr. Marina Rodnina, Biophysical Chemistry, Max Planck Institute for Biophysical Chemistry, Göttingen.

Further members of the Examination Board

Prof. Dr. Stefan Jakobs, Mitochondrial Structure and Dynamics, Max Planck Institute for Biophysical Chemistry, Göttingen.

Prof. Dr. Patrick Cramer, Molecular Biology, Max Planck Institute for Biophysical Chemistry, Göttingen.

Prof. Dr. Markus Bohnsack, Molecular Biology, University Medical Center Göttingen.

Prof. Dr. Kai Tittmann, Molecular Enzymology, Göttingen Center for Molecular Biology (GZMB).

Date of oral examination: 06.07.2020

Herewith I declare that I prepared the doctoral thesis "**Development, characterization, and application of RNA catalysts for *in situ* labeling of target RNA molecules**" on my own and with no other sources and aids than quoted.

This thesis has not been submitted elsewhere for any academic award or qualification.

Mohammad Ghaem Maghami.

Göttingen, 27.05.2020

Table of content

Summary	1
1 Introduction	3
1.1 Non-catalytic role of RNA	3
1.2 RNA as catalyst.....	4
1.3 Strategies for development of artificial RNA catalysts.....	5
1.3.1 Selection using transition state analogs.....	6
1.3.2 Selection based on tagging of the active species.....	7
1.3.2.1 Addition of oligonucleotide tags	7
1.3.2.2 Tagging by circularization.....	8
1.3.2.3 Affinity tagging	8
1.3.2.3.1 Thiol-reactive capture of the active species	10
1.3.2.3.2 Biotin as a capture tag	11
1.3.2.4 Selection using other tags.....	15
1.3.3 Selection based on <i>in vitro</i> compartmentalization.....	16
1.4 RNA imaging methods.....	17
1.4.1 Non-covalent labeling methods.....	18
1.4.1.1 Hybridization based methods	18
1.4.1.1.1 FISH	18
1.4.1.1.2 Molecular beacon	18
1.4.1.2 RNA binding protein-based methods	19
1.4.1.2.1 MS2-tagging.....	19
1.4.1.2.2 Pumilio domain	20
1.4.1.2.3 CRISPR-Cas based systems	21
1.4.1.3 Fluorogenic aptamers	23
1.4.1.3.1 GFP fluorophore analog binding aptamers	24
1.4.1.3.2 Thiazole-orange binding aptamers	25
1.4.1.3.3 Silicon rhodamine binding aptamer	25
1.4.1.3.4 DNB aptamers.....	26
1.4.1.3.5 BHQ aptamers	26
1.4.1.3.6 SRB-2 aptamer	26
1.4.1.3.7 o-Coral.....	27
1.4.1.3.8 Riboglow	27
1.4.2 Covalent methods.....	28
1.4.2.1 Polymerase mediated incorporation of labeled nucleotides	28
1.4.2.2 RNA modifying enzymes.....	29
1.4.2.2.1 RNA methyltransferases	29

1.4.2.2.2	Other tRNA-modifying enzymes	31
1.4.2.3	Catalytic nucleic acid-based RNA labeling.....	32
1.4.2.3.1	Ribozyme based methods.....	32
1.4.2.4	Deoxyribozyme based RNA labeling methods	35
2	Aim of the study.....	37
3	Results	39
3.1	General design of the selection strategy.....	39
3.1.1	The design of the selection pool.....	39
3.1.2	Construction of the selection pool.....	39
3.1.3	Overview of the selection process.....	40
3.2	Selection of the nucleotidyltransferase ribozymes	44
3.2.1	Selection process	44
3.2.2	Activity assay on the enriched pool	45
3.2.2.1	Cis-activity assay.....	45
3.2.2.2	Trans-activity assay of the pool.....	46
3.2.3	Cloning of the enriched pool and activity assay of the individual clones	47
3.2.4	Sanger sequencing results	48
3.2.5	Kinetic comparison of the evolved species	49
3.2.6	Secondary structure prediction of the evolved variants	50
3.2.7	Characterization of the reaction product	51
3.2.7.1	Probing of the modification-site.....	51
3.2.7.2	ESI-Mass spectrometry of the reaction product.....	52
3.2.8	Characterization of the FH14 and FH31 ribozymes.....	53
3.2.8.1	Target sequence scope of the evolved variants	53
3.2.8.1.1	Substrate sequence specificity of FH14 and FH31.....	53
3.2.8.1.2	Modification site sequence-context requirements of FH14 and FH31.....	56
3.2.8.2	NTP substrate scope of the evolved variants.....	57
3.2.8.2.1	NTP analog specificity range of FH14.....	57
3.2.8.2.2	NTP substrate specificity range of FH31	61
3.2.8.3	MgCl ₂ dependency and NTP substrate concentration range	61
3.2.8.4	Effect of TbCl ₃ on FH14 reaction kinetics	63
3.2.9	FH14-mediated labeling of RNA from cellular origin.....	64
3.2.9.1	FH14-mediated labeling of 5S rRNA.....	65
3.2.9.1.1	Labeling of the pure transcript generated in vitro	65
3.2.9.1.2	Labeling 5S rRNA in the context of total cellular RNA	66
3.2.9.2	FH14-mediated labeling of RyhB small RNA	67
3.3	Tenofovir transferase ribozymes	69
3.3.1	Selection progress	69

3.3.2	Cis- and trans- activity of the enriched pool	69
3.3.3	Sequence identification of the evolved variants.....	70
3.3.3.1	Sanger sequence	70
3.3.3.2	Next-generation-sequencing of the enriched pool.....	71
3.3.4	Kinetic comparison of the evolved variants	72
3.3.5	Secondary structure prediction of FJ1, FJ8 and FJC9.....	73
3.3.6	Characterization of the reaction product	74
3.3.6.1	Probing of the modification site	74
3.3.6.2	ESI-mass analysis of the reaction product	75
3.3.6.3	Dbr1 resistance of the phosphonyl branch	76
3.3.7	Characterization of the chosen variants.....	77
3.3.7.1	Target sequence requirements	77
3.3.7.1.1	Modification-site context requirements of FJ1 and FJ8.....	77
3.3.7.1.2	Target sequence generality of FJC9	80
3.3.7.1.3	Effect of mismatch on FJ1 and FJC9 labeling efficiency.....	82
3.3.7.2	Tenofovir analog scope of the ribozymes	83
3.3.7.3	Orthogonality test of FH and FJ ribozymes	86
3.3.7.4	Dependency of FJ1 on Mg ²⁺	86
3.3.8	Selection of tenofovir transferase ribozymes using an N25-pool.....	87
3.3.8.1	Sanger sequencing of the selection outcome.....	87
3.3.8.2	Substrate sequence variability and tenofovir substrate scope of FK22.....	87
3.3.9	Application of the tenofovir transferase ribozymes	89
3.3.9.1	FJC9 mediated labeling of 5S rRNA.....	89
3.3.9.2	Dual-color orthogonal labeling of synthetic and cellular RNA.....	90
3.3.9.2.1	Mutually orthogonal double labeling of an in vitro transcribed model substrate.....	90
3.3.9.2.2	Dual-color simultaneous labeling of large cellular RNA	92
4	Discussion.....	97
4.1	Selection of trans-acting ribozymes for site-specific RNA modification	97
4.1.1	Analogies with deoxyribozyme selection methods.....	97
4.1.2	Basis of the modification site selectivity of the evolved ribozymes	98
4.1.3	Comparison with other trans-acting RNA modifying ribozymes.....	99
4.2	Properties of FH ribozymes.....	100
4.2.1	Sequence analysis and kinetic comparison of FH ribozymes	100
4.2.2	Potential of the FH ribozymes as RNA labeling tools	101
4.2.3	FH14-mediated labeling of long heavily structured RNA	103
4.2.4	FH ribozymes and perspective for RNA labeling in live cells.....	104
4.3	Tenofovir transferase ribozymes	105
4.3.1	Antiviral nucleotide analogs as substrate for ribozyme catalyzed RNA labeling	105

4.3.2	Target sequence generality of tenofovir transferase ribozymes	108
4.3.3	Tenofovir diphosphate specificity range of FJ ribozymes	109
4.3.4	Orthogonality of FH and FJ ribozymes: potential for simultaneous dual-color RNA labeling	110
4.3.5	FJ ribozymes and the path to RNA labeling in situ.....	111
5	Conclusion and Perspective.....	113
5.1	Possible strategies for FJ ribozyme optimization for cellular application	113
5.1.1	Reducing Mg ²⁺ -dependency by reselection <i>in vitro</i>	113
5.1.2	Possibility of selection <i>in vivo</i>	113
5.2	Potential future selection projects	115
6	Material and methods.....	117
6.1	Instruments.....	117
6.2	Materials.....	118
6.2.1	Chemicals and reagents.....	118
6.2.2	Buffers and solutions.....	118
6.3	Methods.....	120
6.3.1	General procedures.....	120
6.3.1.1	PAGE purification.....	120
6.3.1.2	Ethanol precipitation	120
6.3.1.3	In vitro transcription.....	121
6.3.1.4	Solid-phase oligonucleotide synthesis.....	121
6.3.1.5	Oligonucleotide labeling	122
6.3.1.5.1	3'-end labeling using periodate glycol-oxidation	122
6.3.1.5.2	CuAAC labeling of the synthetic 5'-hexynylated RNA	123
6.3.1.5.3	5'- ³² P labeling of DNA primers	123
6.3.1.6	Total cellular RNA extraction	123
6.3.2	Ribozyme selection	123
6.3.2.1	Preparation of the selection pool.....	123
6.3.2.2	Selection process	124
6.3.2.3	Incubation step	124
6.3.2.4	Capture, wash and elution	125
6.3.2.5	Monitoring of the selection progress.....	126
6.3.2.6	Amplification step.....	127
6.3.2.7	Activity assay on the enriched pool	128
6.3.2.7.1	Cis-activity assay.....	128
6.3.2.7.2	Trans-activity assay.....	129
6.3.3	Sanger sequencing.....	129
6.3.3.1	Cloning and of the enriched pool	129

6.3.3.2	Heat shock transformation.....	130
6.3.3.3	Colony PCR.....	130
6.3.3.4	Activity assay of the individual clones.....	131
6.3.3.5	Plasmid extraction	131
6.3.4	Next-generation sequencing sample preparation.....	132
6.3.5	Characterization of the chosen variants.....	132
6.3.5.1	General kinetic assay protocol	132
6.3.5.2	Characterization of the reaction product	133
6.3.5.2.1	Preparative scale modification of the reaction product for mass analysis.....	133
6.3.5.2.2	Preparative scale modification of the 5'-5FAM-R660.....	133
6.3.5.2.3	RNase T1 probing	133
6.3.5.2.4	Alkaline hydrolysis probing	133
6.3.5.2.5	Debranching resistance assay.....	134
6.3.6	Application of the ribozymes	134
6.3.6.1	FH14 catalyzed labeling of the in vitro transcribed 5S rRNA and RyhB RNA.....	134
6.3.6.2	Labeling 5S rRNA and RyhB RNA in total cellular RNA context.....	135
6.3.6.3	Mutually orthogonal double labeling of the synthetic model substrate.....	135
6.3.6.3.1	Labeling reaction.....	135
6.3.6.3.2	8-17NG mediated cleavage reaction	135
6.3.6.4	FH14 and FJ1 mediated labeling of 16S and 23S rRNA.....	136
6.3.6.5	Superscript III primer extension experiment.....	136
6.3.6.5.1	Preparation of the 5S rRNA modified using FH14 and FJC9 ribozymes	136
6.3.6.5.2	Preparation of the 16S and 23S rRNA modified using FJ1 and FH14.....	136
6.3.6.5.3	Primer extension protocol	137
6.3.6.5.4	Sequencing ladder preparation	137
7	Bibliography	138
8	Appendix.....	150
	Abbreviations.....	162
	Acknowledgements.....	165
	Curriculum Vitae	166

List of figures

Figure 1-1 Selection based on transition state analogs.....	6
Figure 1-2 Circularization based selection of Iso-6 ribozyme.	8
Figure 1-3 General scheme of the affinity tagging selection methods.....	9
Figure 1-4 Thiol-reactive capture methods.....	10
Figure 1-5 Biotinylated substrates used in selection of RNA-modifying ribozymes.....	12
Figure 1-6 Reactions catalyzed by ribozymes with non-nucleotide substrates.....	14
Figure 1-7 Unusual tagging strategies in ribozyme selection.	16
Figure 1-8 Hybridization-based RNA imaging methods.	19
Figure 1-9 MS2-tagging system in RNA imaging.	20
Figure 1-10 Pumilio-based RNA imaging system..	21
Figure 1-11 RNA guided CRISPR-Cas based imaging methods.....	22
Figure 1-12 Fluorogen activating aptamers.....	23
Figure 1-13 Fluorogenic aptamer-dye pairs	28
Figure 1-14 RNA methyltransferases in RNA labeling.	31
Figure 1-15 tRNA-modifying enzyme-mediated RNA labeling.....	32
Figure 1-16 Cis-reacting self-alkylating ribozymes as RNA labeling tools.....	33
Figure 1-17 Twin-ribozyme based RNA labeling.....	34
Figure 1-18 24-3 polymerase ribozyme mediated RNA labeling.	35
Figure 1-19 10DM24 mediated labeling of an RNA of interest.....	36
Figure 2-1 Schematic representation of RNA-Catalyzed site-specific RNA labeling.	38
Figure 3-1 Pool design and construction.....	40
Figure 3-2 Overview of the selection process.....	42
Figure 3-3 Priming scheme of the amplification process.....	43
Figure 3-4 Selection substrate and the desired reaction for adenylyl transferase ribozymes.....	44
Figure 3-5 Selection progress of the adenylyltransferase ribozyme.	45
Figure 3-6 Cis- and Trans-activity of the enriched pool..	47
Figure 3-7 Streptavidin gel-shift assay of ten exemplary, individual clones	48
Figure 3-8 Sequence alignment and kinetic comparison of the evolved variants..	50
Figure 3-9 Secondary structure prediction of the three variants.	50
Figure 3-10 Probing of the modification site.	52
Figure 3-11 ESI-mass analysis of the FH14 modification product..	53
Figure 3-12 Target sequence variability of FH14 and FH31.	55
Figure 3-13 Folded and misfolded states of the FH14 ribozyme.	55
Figure 3-14 Modification-site sequence context requirement of the FH14 and FH31.....	57
Figure 3-15 List of the ATP analogs used in this thesis project.....	58
Figure 3-16 NTP analog specificity of FH14 ribozyme.....	59

Figure 3-17 Further investigation of the NTP analog scope of FH14 ribozyme.....	60
Figure 3-18 ATP analog scope of FH31 ribozyme..	61
Figure 3-19 MgCl ₂ dependency and the ATP analog concentration range of FH14.....	62
Figure 3-20 Effect of Tb ³⁺ on FH14 reaction kinetics.....	64
Figure 3-21 FH14 mediated labeling of <i>E. coli</i> 5S rRNA.....	67
Figure 3-22 FH14-catalyzed labeling of RyhB RNA.....	68
Figure 3-23 Tenofovir transferase ribozyme selection substrate and desired reaction.	69
Figure 3-24 Selection progress and cis- and trans-activity of the enriched pool..	70
Figure 3-25 Abundance of individual sequences in reads per million (RPM) compared in round 7 and round 12.....	72
Figure 3-26 Partial alignment and kinetic comparison of the variants identified from Sanger and next-generation sequencing.....	73
Figure 3-27 Predicted secondary structures and base pair probability dot plots of the selected FJ ribozymes.	74
Figure 3-28 Probing of the FJ1 and FJC9 modification site	75
Figure 3-29 Mass analysis of FJ1 ribozyme product.	76
Figure 3-30 Resistance of FJ1 vs FH14 modified RNA to debranching.....	77
Figure 3-31 Target sequence variability of FJ1 and FJ8 ribozymes.	79
Figure 3-32 Substrate sequence flexibility of FJC9.....	81
Figure 3-33 Effect of mismatch on FJ ribozyme labeling efficiency.....	82
Figure 3-34 list of Tenofovir-diphosphate analogs used in this study.	83
Figure 3-35 Tenofovir-diphosphate analog specificity of FJ ribozymes.....	85
Figure 3-36 Racemic TenDP vs R-TenDP and their reactivity towards FJ1.	86
Figure 3-37 Orthogonality test of FH and FJ ribozymes.....	86
Figure 3-38 Mg ²⁺ dependency of FJ1.....	87
Figure 3-39 Selecting a tenofovir transferase ribozyme using an N25 pool.	88
Figure 3-40 FJC9 mediated labeling of the <i>E. coli</i> 5S rRNA.....	90
Figure 3-41 Mutually orthogonal double labeling of a synthetic transcript.....	92
Figure 3-42 Targeted positions on <i>E. coli</i> 16S and 23 rRNAs.....	94
Figure 3-43 FJ1 and FH14 mediated labeling of 16S and 23S rRNA at targeted positions.....	95
Figure 3-44 Reverse transcriptase primer extension assay on FJ1 or FH14 modified, 16S, and 23S rRNA.....	96
Figure 4-1 Potential FH14 mediated labeling substrates with bioorthogonal handles.	103
Figure 4-2 Tenofovir, an acyclic nucleotide phosphonate analog of ATP.....	107
Figure 4-3 Alternative fluorophore-TenDP conjugate.....	109
Figure 4-4 Debranching resistant NTP ^s analogs..	112
Figure 5-1 Possible three-hybrid system for reselection of FJ ribozymes in vivo.	116
Figure 8-1 ESI-mass spectrum and the deconvolution of the R549 modified using FH14.....	159
Figure 8-2 ESI-mass spectrum and the deconvolution of the R549 modified using FJ1	161

List of tables

Table 3-1 Details of the adenylyltransferase ribozyme selection rounds.....	45
Table 3-2 Sequence groups identified from Sanger sequencing of the isolated clones	49
Table 3-3 k_{obs} values of FH ribozymes, and FH14 with different substrates	61
Table 3-4 k_{obs} values of FH14 ribozyme-catalyzed labeling of 5FAM-R660 with N ⁶ -biotin-ATP, with varying Mg ²⁺ and ATP concentrations.....	63
Table 3-5 Details of the selection rounds of tenofovir transferase ribozymes.	70
Table 3-6 Sequence groups identified from the Sanger sequencing of isolated clones	71
Table 3-7 The chosen sequences from NGS data analysis with their respective log2 values.....	72
Table 3-8 Kinetic parameters of the FJ variants.....	73
Table 3-9 Kinetic Details of the investigation of the FJ1 modification context.	80
Table 3-10 Kinetic details of the investigation of the FJC9 modification context.....	82
Table 3-11 Details of the selection rounds of the N25 tenofovir transferase ribozyme selection....	87
Table 6-1 List of instruments used in this thesis project.....	117
Table 6-2 Details of each FH selection round.....	126
Table 6-3 Details of each FJ selection round.	126
Table 6-4 Cycling parameters of the one-pot, RT-PCR reaction	127
Table 6-5 Cycling parameters of the PCR reaction.....	128
Table 8-1 List of primers used in this thesi project.....	150
Table 8-2 List of transcription templates	151
Table 8-3 The selection pool and the transcribed substrate sequences	154
Table 8-4 List of the synthetic RNAsubstrates used for ESI-mass analysis and probing	155
Table 8-5 List of the transcribed ribozyme variants. Binding arms are underlined.	155
Table 8-6 Details of the ten main clusters identified from the NGS analysis of the round 12 pool160	
Table 8-7 First sequence of fist ten clusters ordered by log2 (enrichment).	160

Summary

RNA molecules play a diverse set of crucial roles in biological systems. Studying various aspects of RNA biology such as RNA transport, localization, degradation, and structural dynamics is therefore of utmost significance. A prerequisite for studying such phenomena is availability of methods for labeling and visualization of the RNAs of interest. A wide array of RNA labeling methods has been developed over the years with promising results. Nevertheless, versatile, and efficient tools for covalent site- and sequence-specific labeling of RNA in live cell (*in situ*) are non-existent.

One of the more recently emerged RNA labeling methods involves application of the 10DM24 deoxyribozyme. The recognition arms of this DNA catalyst recognize the desired labeling positions in the RNA of interest via Watson-Crick base pairing. Using a wide variety of 2'-modified GTP analogs, 10DM24 can label the target RNA at specific internal positions, efficiently and accurately. Despite the success achieved using this system, the deoxyribozyme has never been tested for RNA labeling within live cell. The dependency of the deoxyribozyme on metal ions and challenges regarding maintenance and delivery of DNA catalysts in the cell, has so far limited its application to labeling *in vitro*. Therefore, the aim of this study was to develop ribozymes for RNA labeling with the potential for cellular applications.

We established a selection method through which selection of trans-acting, site-selective RNA labeling ribozymes was made possible. The selection involved a partially structured RNA pool with a bulged adenosine nucleotide as a predetermined modification-site. For the development and optimization of the selection strategy, biotinylated ATP was used as the selection substrate. The selection using this substrate led to the identification of adenylyltransferase ribozymes, denoted as FH ribozymes. The characterization of these catalysts confirmed the success of our selection method in dictating the modification site of ribozymes. The selected variants were effectively directed to modify the bulged adenosine nucleotide by forming a 2'-5'-branched phosphodiester bond with labeled adenosines. The FH ribozymes could be readily converted to trans-acting variants, and the ribozyme showed a broad substrate scope based on Watson-Crick base-pairing between target RNA and ribozyme binding arms. The selected ribozymes, especially the most efficient variant (FH14), efficiently accepted a wide range of N⁶-modified-ATP analogs as labeling substrates, including fluorophore-conjugated derivatives. FH14 was successfully applied in

site-specific labeling of large, heavily structured cellular RNA such as 5S, 16S, and 23S ribosomal RNA in total cellular RNA.

In the next stage, we attempted to develop ribozymes that utilized a more bioorthogonal substrate such as the antiviral ATP analog tenofovir-diphosphate. Ligation of tenofovir analogs to the target RNA would also result in a more enzymatically stable linkage type, in which the branched 2'-5'-phosphodiester is replaced by a phosphonate ester. Biotinylated tenofovir-diphosphate was synthesized and applied in an *in vitro* selection process that led to the identification of tenofovir-transferase (FJ) ribozymes. These ribozymes showed similar properties to FH14 ribozyme, in terms of site-selectivity, substrate sequence generality, and broad labeling substrate scope. Moreover, these ribozymes were perfectly orthogonal to FH14 ribozyme, a feature that was exploited for dual-color RNA labeling. The ribozymes FJ1 and FH14 were successfully applied in simultaneous dual-color labeling of a synthetic transcript at two different positions. These ribozymes have also been used for simultaneous, site-specific labeling of 16S and 23S rRNAs using two different fluorophores. Although the cellular application of these ribozymes has not yet been demonstrated, they possess great potential for further optimization into efficient tools for RNA labeling *in situ*. Furthermore, the established selection process paves the way for future development of other RNA labeling tools, using non-nucleotide based selection substrates.

1 Introduction

1.1 Non-catalytic role of RNA

RNA is arguably the most versatile class of biomolecules in terms of the various roles it can assume in living systems. It is known to act as the main form of genetic information storage space in some life forms (Poltronieri et al., 2015). The three major classes of RNA (mRNA, tRNA, and rRNA) are key to converting genetic information into protein. Apart from these three types of RNA, there is a large variety of other classes of small and long RNA involved in regulating or maintaining different aspects of the central dogma of molecular biology.

siRNA and miRNA are some of the examples of small antisense RNA molecules found in eukaryotic cells (Tang, 2005). They regulate the translation and stability of their target mRNA with the aid of a protein complex known as RISC (Kaikkonen et al., 2011; O'Brien et al., 2018). Analogous small antisense RNA molecules such as the Hfq and ProQ dependent small RNAs (sRNA) that regulate the translation and stability of their target RNA, can also be found in bacteria (Dutta and Srivastava, 2018).

Various classes of long non-coding RNA (lncRNA) have been identified in higher multicellular eukaryotes (Fatica and Bozzoni, 2014; Laurent et al., 2015). The function and significance of many lncRNA are still unknown, however, some are known to regulate gene expression by various means (Dahariya et al., 2019; Kaikkonen et al., 2011). Xist, the most well-known instance of lncRNAs, is involved in the inactivation of X-chromosome in female mammals. It acts by recruiting chromatin remodeling complexes (Loda and Heard, 2019).

RNA can also affect gene expression via direct interaction with micro- and macromolecules. Riboswitches are a great example of such RNA functionality. These relatively small RNA elements can be found primarily in the leader sequence of a variety of bacterial RNA. Their binding to the target molecule and the subsequent conformational change can inhibit or enhance translation by masking or unmasking of the ribosome binding site. They can also prevent or promote transcription elongation via the formation or destabilization of terminator/antiterminator structures (Bédard et al., 2020).

1.2 RNA as catalyst

Catalytic potential of the RNA was appreciated in 1982 when Thomas Cech's research group discovered that class I introns can catalyze reactions necessary for RNA maturation. The term ribozyme was therefore coined to refer to catalytically active RNA sequences (Kruger et al., 1982). In 1983 the catalytic activity of the RNA subunit of RNase P was revealed (Guerrier-Takada et al., 1983). Class II introns were also demonstrated to catalyze splicing reactions on their own. (Peebles et al., 1986). In 1992, it was shown that an *E. coli* ribosome subjected to protein extraction retains a significant portion of its peptidyl transferase activity. The finding hinted to the involvement of rRNA in peptide synthesis (Noller et al., 1992). Crystallographic studies in the early 2000s confirmed that the peptidyl transferase center indeed lies in 23S rRNA (Nissen et al., 2000).

Self-cleaving ribozymes are another class of catalytic RNA abundantly found in nature. The first examples to be discovered were the hammerhead (Hutchins et al., 1986; Prody et al., 1986) and HDV ribozymes (Wu et al., 1989) that were originally discovered in the genome of viroids and RNA viruses. Further research revealed homologous sequences with self-cleaving activity in the genomic sequences of all domains of life (De la Peña et al., 2017; Jimenez et al., 2015). Hairpin and Varkud satellite ribozymes, also belong to the category of self-cleaving ribozymes that were discovered through biochemical means (Jimenez et al., 2015). With the advances in bioinformatic sequence analysis, ribozymes such as twister, pistol, and hatchet were also discovered in various sources (Müller et al., 2016; Roth et al., 2014).

Except for ribosomal peptidyl transferase centers, all naturally occurring ribozymes discovered so far catalyze trans-esterification reactions at the RNA phosphodiester backbone (Müller et al., 2016). The advent of the in vitro selection methods, however, facilitated the discovery of artificially evolved ribozymes with a broad range of activities. The known catalytic repertoire of RNA was therefore dramatically expanded (Fiammengo and Jäschke, 2005; Jäschke, 2001; Jäschke and Seelig, 2000). Discovery of such ribozymes provides further evidence for the existence of an "RNA world". According to the RNA world theory early life forms relied heavily on RNA for biocatalysis as well as storage of genetic information (Cech, 2012; Gilbert, 1986).

1.3 Strategies for development of artificial RNA catalysts

Artificial ribozymes are almost exclusively evolved from a random pool of RNA sequences. The development process of these catalysts involves an iterative accelerated evolution process also known as *in vitro* selection. One may likely find several sequences possessing a desired activity in a combinatorial random RNA pool of 10^{14} - 10^{16} unique sequences (Joyce, 2004; Lorsch and Szostak, 1996). The selection process is designed in a way that provides survival advantage for the active species of the pool while the large bulk of inactive sequences are rapidly eliminated. The frequency of the active sequences in an initial pool is extremely low. Thus, it is necessary to perform several rounds of functional selection and amplification. Throughout the selection rounds, the active sequences are exponentially enriched to the point that they comprise the majority of the pool.

To facilitate the amplification process, the random segments of the pool are often flanked by constant sequences. These constant regions serve as primer binding sites during the reverse transcription and PCR amplification. During the PCR process, a T7-promoter is generally introduced at the 5'-end of the DNA template, which is essential for *in vitro* transcription using T7 polymerase. Another amplification step occurs at the *in vitro* transcription step as every DNA template can be transcribed multiple times. The selection rounds are pursued with increasing stringency until the most desirable variants are remaining (Joyce, 2004). The pool is then subjected to sequence analysis either through classical means such as Sanger sequencing or is directly sequenced via next-generation sequencing. The identified sequences are then subjected to in-depth functional and structural characterization (Joyce, 2004).

Aptamers are another class of functional nucleic acids that can be evolved artificially (Dunn et al., 2017). They are developed to specifically recognize and bind to target molecules of interest. While aptamers also exist naturally as part of riboswitches (Bédard et al., 2020), a wide variety of them have also been developed using *in vitro* selection methods, generally regarded as SELEX (Dunn et al., 2017; Tuerk and Gold, 1990). For aptamer selection mere target binding is desired whereas for ribozymes substrate recognition and catalytic property must coincide in the same oligonucleotide sequence. Thus, *in vitro* selection process of ribozymes is more complicated than that of aptamers.

Various strategies exist for isolation of catalytic RNA from libraries of random RNA sequences, which will be discussed in the following section.

1.3.1 Selection using transition state analogs

Before application in ribozyme selection, this strategy was utilized for the development of catalytic antibodies also known as abzymes (Tramontano et al., 1986). This approach relies on the idea that enzyme active sites are complementary to the transition state of the reaction they catalyze. The stable transition state analogs of the reactions of interest, are therefore, designed, synthesized, and used as ligands in SELEX experiments. The RNA molecules with binding affinity for these analogs are then screened for catalytic properties (Jäschke and Seelig, 2000).

An example of RNA catalysts selected using this method is a cholesterol esterase ribozyme developed by Chun et al. The group first evolved strong binders to a phosphodiester analog of the cholesterol carbonate ester. This analog has similar geometry and charge distribution as the transition state of the carbonate ester being hydrolyzed (**Figure 1-1 A**). Two of the RNAs that bound to this analog were catalytically active, including the one with the highest binding affinity (Chun et al., 1999).

Prudent et al also utilized this approach to select for a ribozyme that catalyzes isomerization of a “bridged biphenyl” substrate. A planar transition state analog of the isomerization reaction was used as the ligand during the selection process in this instance (**Figure 1-1 B**). Prudent et al isolated three variants that bound to the selection ligand. Only one of these variants demonstrated catalytic activity (Prudent et al., 1994).

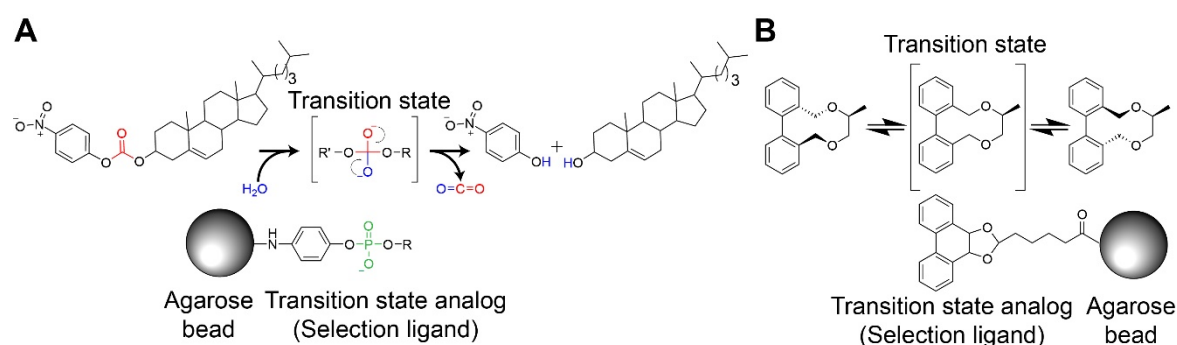


Figure 1-1 Selection based on transition state analogs. **(A)** Reaction catalyzed by the cholesterol esterase ribozyme and the transition state analog used for its selection. **(B)** Isomerization of the bridged biphenyl compound and the transition state analog used for the selection of the isomerase ribozyme.

Selection using transition state analogs has a low success rate and results in catalysts with low activities (Jäschke and Seelig, 2000). Therefore, the alternative approach in which the

oligonucleotides are directly selected for catalytic activity is the preferred strategy for ribozyme selection.

1.3.2 Selection based on tagging of the active species

In this strategy, the selection is designed in a way that catalytic activity leads to the tagging of the active sequences. The tagged sequences can then be separated using a variety of means depending on the type of tag. Some of the most common tagging and capture methods are described in detail in this section.

1.3.2.1 Addition of oligonucleotide tags

Addition of sequence tags that lead to amplification compatibility has been used for the selection of RNA ligase ribozymes. Through their activity, the catalysts in the pool ligate a sequence tag to their own 5'-end. The ligated sequence is then utilized as a primer binding site during the amplification process leading to selective amplification of active species. The active species can also be further purified using beads coupled to oligonucleotides complementary to the sequence tag. Szostak's group, for example, utilized this strategy to select for an RNA ligase ribozyme. The desired activity in this case was the ability of the ribozyme to catalyze 3'-5' ligation of a sequence tag to its own 5'-triphosphate end in a template-directed manner. The ligated sequence was first used as an affinity tag then as a primer binding site for specific amplification of active species (Bartel and Szostak, 1993). This selection experiment led to the identification of the Class I ligase ribozyme which was evolved further into various ligase or even polymerase ribozymes (Joyce, 2004). Other ligase ribozymes such as L1 (Robertson and Ellington, 1999) and R3 (Rogers and Joyce, 2001) were also evolved using a similar strategy. R3 ligase is a peculiar example as it was selected from an RNA pool that contained no cytidine nucleotides. Cytidine nucleotides were then incorporated during a reselection process which led to an improved version called the R3C ligase (Rogers and Joyce, 2001). This was further evolved and recently minimized to the smallest catalytic core of 18 nucleotides of a ligase ribozyme (Nomura and Yokobayashi, 2019).

1.3.2.2 Tagging by circularization

During their search for aminoacyl-AMP synthetase-like activity, Yarus' laboratory devised a strategy that was supposed to convert the active species to 5'-monophosphorylated form. The reaction product in the selection was a mixed anhydride between the free carboxyl group of an amino acid and the 5'-phosphate of the pool. Incubation under acidic conditions leads to hydrolysis of this mixed anhydride and formation of 5'-monophosphorylated ribozymes. The active variants could then be circularized using T4 RNA ligase (**Figure 1-2 A**). Considering the significantly lower electrophoretic mobility of circular RNA, they then used PAGE purification to isolate the active species (**Figure 1-2 A**). While the selection did not meet its original objective, it resulted in the isolation of a pyrophosphatase ribozyme called Iso-6. The ribozyme hydrolyzed its 5'-triphosphate to 5'-monophosphate (**Figure 1-2 B**) (Huang and Yarus, 1997a). Iso-6 was also shown to be capable of forming phosphoanhydride bonds by exchanging the 5'- $\beta\gamma$ -pyrophosphate with a terminal phosphoryl group (**Figure 1-2 A**). Various phosphorylated compounds can therefore be conjugated to the 5'-end of this ribozyme via a phosphoanhydride bridge. Iso-6-catalyzed conjugation of a GDP leads to formation of a structure analogous to the eukaryotic mRNA cap structure. The catalyst was thus described as a 5'-capping ribozyme (Huang and Yarus, 1997a, b).

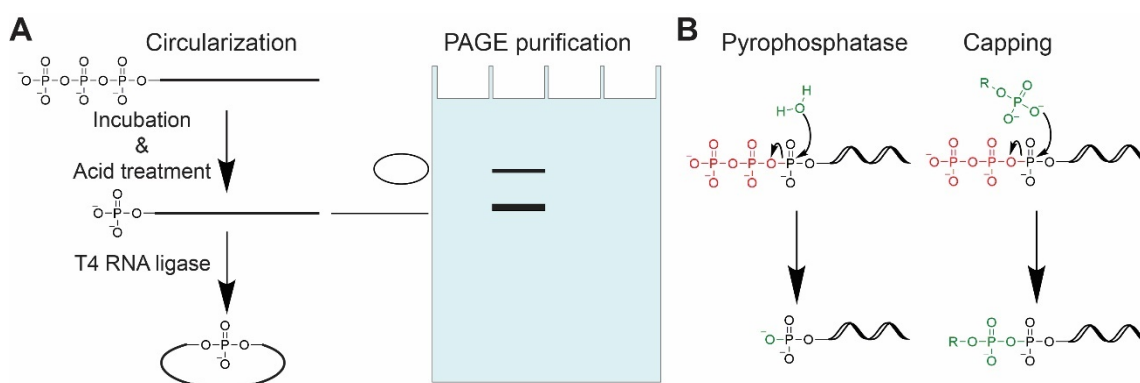


Figure 1-2 Circularization based selection of Iso-6 ribozyme. (A) Schematic representation of the incubation and capture steps of the capping ribozyme selection. (B) Pyrophosphatase vs capping functions of Iso-6.

1.3.2.3 Affinity tagging

These strategies involve activity-related addition of chemical tags, which can be captured through direct or indirect methods. Ribozymes selected using affinity capture strategies can be classified into two major groups (Jäschke and Seelig, 2000)(**Figure 1-3**). The first group includes RNA or self-modifying ribozymes. During the selection of these ribozymes, a segment of the catalyst is evolved to act as the modification target. In case of the ribozymes

evolved for tRNA 3'-aminoacylation, the tRNA substrate is also fused to the selection pool (Ishida et al., 2020; Saito et al., 2001). The typical reactions catalyzed by RNA/self-modifying ribozyme are nucleophilic substitutions. The affinity-tagged substrate usually contains an electrophilic center coupled to a leaving group which is displaced by a nucleophilic group from the ribozyme (**Figure 1-3 A**). Exceptions can be found for the reactions that involve 5'-triphosphate of the ribozyme. For these reactions, the nucleophilic group resides on the tagged selection substrate. The 5'- $\beta\gamma$ -pyrophosphate of the ribozyme assumes the role of the leaving group (Coleman and Huang, 2002; Kang and Suga, 2007).

Ribozymes acting on non-nucleotide substrates are the second group evolved using affinity tagging strategy. For the selection of these catalysts, one of the reactants must be covalently tethered to the RNA pool while the second reactant carries the selection tag. Ribozyme catalyzed reaction between the two substrates results in the formation of a ribozyme bound product-selection tag conjugate (**Figure 1-3 B**).

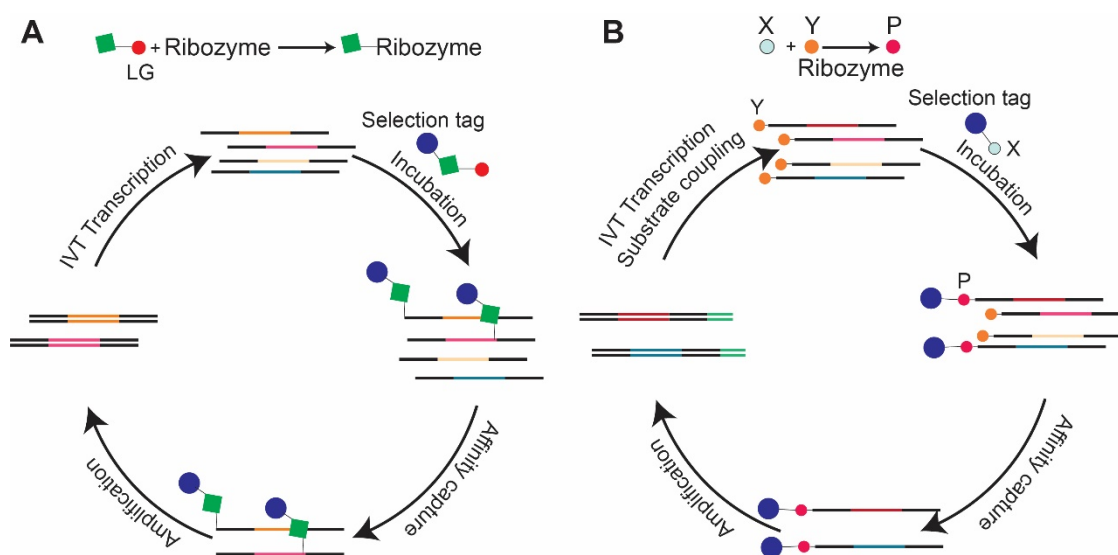


Figure 1-3 General scheme of the affinity tagging selection methods. **(A)** Selection of the RNA modifying ribozymes (LG= Leaving group). **(B)** Selection of ribozymes acting on non-nucleotide substrates.

Several different tags are used for the affinity tag-based ribozyme selection. Thiol and biotin tags, however, are the most frequently used in the reported literature. The selection strategies using various affinity tags are detailed in the following section.

1.3.2.3.1 Thiol-reactive capture of the active species

Modification of the active variants with thiol/thione groups facilitates direct isolation using thiol-reactive matrices. Pyridyl disulfide functionalized particles for example can immobilize thiol/thione-modified pool species via disulfide exchange. Addition of a reducing agent such as DTT or β -ME helps elute the attached RNA (**Figure 1-4 A**). Affinity gel electrophoresis is another method for isolating thiol/thione modified RNA. In this method a layer containing acrylaminothiophenyl mercuric acid (APM) is included in the gel which forms covalent bonds with thiol/thione groups (Igloi, 1988). The thiol/thione-modified RNA is therefore trapped in this layer of the gel during PAGE and can be extracted (**Figure 1-4 B**).

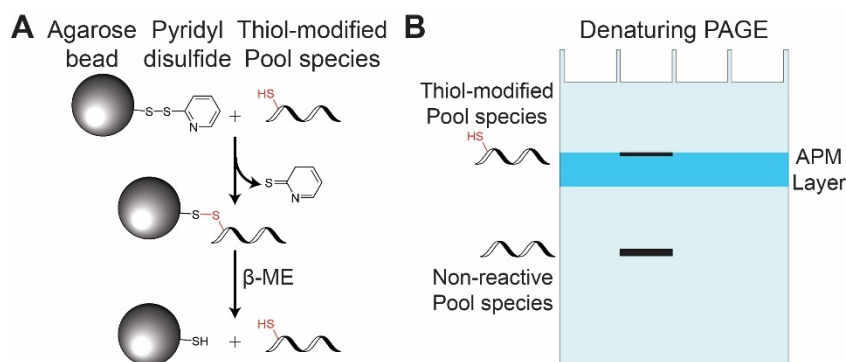


Figure 1-4 Thiol-reactive capture methods. **(A)** Pyridyl disulfide functionalized beads can form disulfide bonds with thiol/thione-modified RNA. Reducing agents such β -ME are then used to elute RNA. **(B)** Affinity gel-electrophoresis using an APM containing layer in denaturing PAGE.

Modification with a thiophosphate has been utilized by several groups in selection of ribozymes with kinase activity. The RNA pool was incubated with either ATP γ S (Lorsch and Szostak, 1994; Saran et al., 2005) or GTP γ S (Curtis and Bartel, 2005) and in one case both (Biondi et al., 2010). The catalytically active RNAs with the ability to transfer thiophosphoryl group to a nucleophilic position along their length were then captured. The selected kinases in most cases (thio)phosphorylate themselves either at a free 5'-OH or an internal 2'-OH. Ribozyme K28 (Biondi et al., 2010) is believed to thiophosphorylate itself at two internal positions (Biondi et al., 2012). Evidence exists that at least one of the thiophosphorylation reactions of K28 occurs on the 2-amino of an internal guanine nucleotide (Poudyal et al., 2017).

Thiol-based selection system has not however been limited to selection of kinase ribozymes. ATP γ S has also been used as substrate for the selection of a 5'-purine nucleotide transferase ribozyme (Kang and Suga, 2007). A second 5'-capping ribozyme was evolved with the

original objective of the selection of an RNA polymerase ribozyme, using 4sUTP (Zaher et al., 2006).

In 2001, Yarus et al successfully evolved aminoacyl-AMP synthetase like ribozymes using 3-mercaptopropionic acid as substrate (Kumar and Yarus, 2001). The isolated ribozyme catalyzes formation of a mixed anhydride linkage by facilitating a nucleophilic attack from the free carboxyl group of the 3-mercaptopropionic acid onto the α -phosphate of the 5'-triphosphate of the ribozyme itself.

1.3.2.3.2 Biotin as a capture tag

Biotin tagging is the most frequently reported selection system in literature (**Figure 1-5**). The system relies on the extremely high affinity of streptavidin for biotin ($K_d = 10^{-15}$ M) (Holmberg et al., 2005). The high affinity of interaction allows efficient retainment of the minuscule amounts of active variants in the initial selection rounds. The complex has also been shown to be extremely resistant to denaturants such as guanidinium hydrochloride and urea (González et al., 1999). Owing to this stability, harsh washing steps can then be performed to remove non-specific binders from the affinity matrix (Joyce, 2004).

Three self-alkylating ribozymes have been selected using biotinylated substrates containing reactive groups such as: iodoacetamide (Wilson and Szostak, 1995), chloroacetamide (Ameta and Jäschke, 2013) and disubstituted epoxide (McDonald et al., 2014) (**Figure 1-5 B**). Selection using these substrates led to identification of ribozymes that self-alkylated at specific internal guanine residues, at N⁷-position (**Figure 1-5 D**).

Self-aminoacylating or aminoacyl transferase ribozymes have also been evolved using biotinylated amino acids with various forms of α -carboxy activation (**Figure 1-5 C**). Activated amino acids used in these experiments were in the form of mixed anhydrides with AMP (Chumachenko et al., 2009), thioesters with CoA (Li and Huang, 2005), cyanomethyl esters (Ishida et al., 2020; Murakami et al., 2003; Saito et al., 2001), dinitrobenzyl esters (Murakami et al., 2006), 2'-aminoacyl-AMP (Jenne and Famulok, 1998) and amino acids esterified to the 3'-OH of short oligonucleotides (Lohse and Szostak, 1996). Most recently self-aminoacylating ribozymes have also been selected using a biotinylated oxazolone analog of tyrosine (**Figure 1-5 C**). The selected ribozymes are esterified at either the 3'-OH positions or at an internal 2'-OH (**Figure 1-5 E**). (Pressman et al., 2019). Biotin, in most cases, is attached to the α -amino group of the aminoacyl substrate, except for two cases

where an amino group in the side chain was biotinylated (Li and Huang, 2005; Murakami et al., 2006). The most famous group of tRNA-aminoacyltransferase ribozymes are flexizymes (Ohuchi et al., 2007). These artificial RNA catalysts use dinitrobenzyl-ester forms of amino acids (Murakami et al., 2006), regardless of the sidechain, to charge tRNA molecules. These ribozymes have been used for incorporation of unnatural amino acids into polypeptides through *in vitro* translation systems (Murakami et al., 2006).

Yarus' laboratory selected another capping ribozyme with similar activity to Iso-6, using biotin conjugated to a terminally phosphorylated linker (Huang et al., 2000). In 2014, an RNA 5'-hydroxyl triphosphorylating ribozyme was identified by Morretti et al. During the selection, the activity of the sequence converted it to a substrate for the R3C RNA ligase ribozyme. R3C then ligated a biotinylated capture oligonucleotide to the 5'-triphosphorylated ribozymes, allowing them to be isolated using streptavidin-coated magnetic beads (Moretti and Müller, 2014). In a subsequent report the ribozyme was minimized and improved via partial randomization and reselection under more stringent conditions, using the same capture strategy (Dolan et al., 2015). These reports are intriguing examples of an artificial ribozyme being used as a tool for *in vitro* selection of novel RNA catalysts.

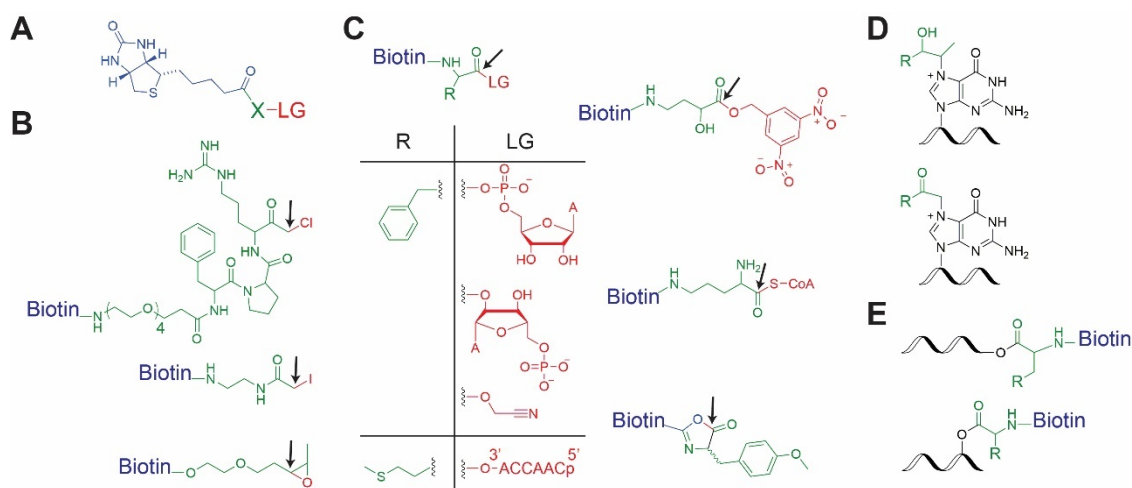


Figure 1-5 Biotinylated substrates used in selection of RNA-modifying ribozymes. Biotin tag is marked in blue, the transferred moieties and the leaving groups are marked in green and red, respectively. Arrows indicate the electrophilic centers. **(A)** General design of the selection substrate (LG= leaving group). **(B)** Selection substrates for reported self-alkylating ribozymes. **(C)** Selection substrates for aminoacyl transferase ribozymes. **(D)** Reaction products of self-alkylating ribozymes. (Top: epoxide substrate product, bottom: chloro/iodoacetamide substrate product. **(E)** 3'-end (top) and internal (bottom) aminoacylation by aminoacyl transferase ribozymes.

Ribozymes acting on non-nucleotide substrates have also been selected using biotin-streptavidin based capture systems (**Figure 1-6**). Instances of these ribozymes (**Figure 1-6 A**) include: A Diels-Alderase ribozyme (Seelig and Jäschke, 1999), A peptidyl synthetase like ribozyme (Zhang and Cech, 1997), thioester forming or acyl-CoA synthesizing ribozymes (Coleman and Huang, 2002; Jadhav and Yarus, 2002), a ribozyme catalyzing an aldol condensation reaction (Fusz et al., 2005) and a thiamine-dependent pyruvate decarboxylase like ribozyme (Cernak and Sen, 2013)

The examples mentioned so far involved reaction substrates directly tagged to biotin. Indirect tagging methods also exist in which a ribozyme-tethered functional group, orthogonal to RNA is formed, due to the catalytic activity. The reactive group is then derivatized using suitable forms of biotin. The NAD⁺-dependent alcohol-dehydrogenase ribozyme, for instance, was identified through indirect biotinylation. Active species in this selection converted a tethered benzyl alcohol to benzaldehyde. The resulting aldehyde was then reacted with biotin-hydrazide which led to biotinylation of the ribozymes through Schiff-base formation (**Figure 1-6 B**) (Tsukiji et al., 2003, 2004).

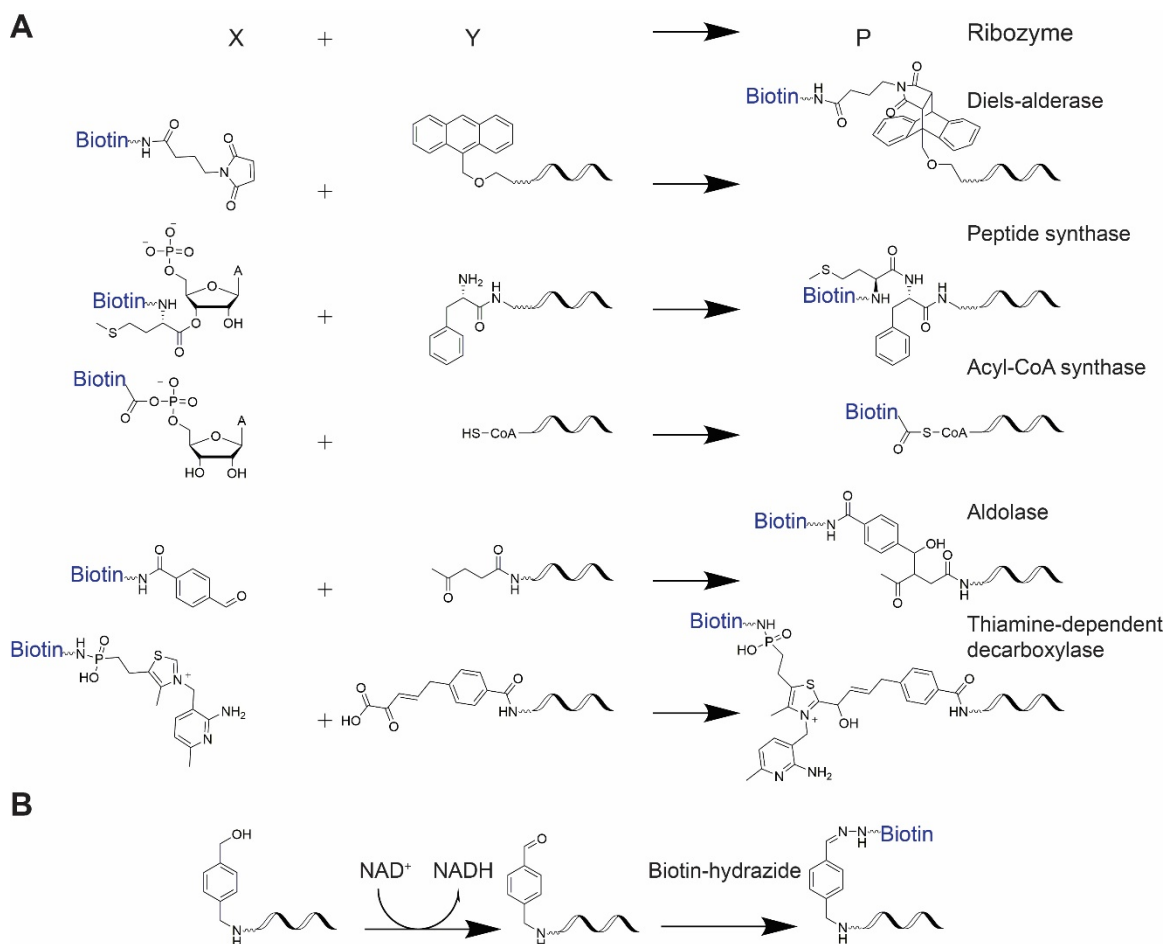


Figure 1-6 Reactions catalyzed by ribozymes with non-nucleotide substrates. **(A)** Direct capture strategy. X = the biotinylated substrate, Y= ribozyme tethered substrate and P = reaction product. **(B)** In direct capture strategy of the alcohol dehydrogenase ribozyme.

The method of choice for conjugation of the reactants to the RNA pool is transcriptional priming (Schlatterer and Jäschke, 2006). GMP nucleotides modified with orthogonal reactive groups on α -phosphate are used for transcriptional priming of the pool before each round. The substrate derivatized with the appropriate functionalities is then conjugated to the primed pool variants.

In another interesting example of artificial ribozymes being used as tools for ribozyme selection was reported by Jadhav et al. They utilized an improved Iso-6 ribozyme, called Iso-6-1, for 5'-conjugation of coenzyme-A to a selection pool. Acyl-CoA forming ribozymes were then evolved from the pool using biotinyl-AMP (Jadhav and Yarus, 2002).

1.3.2.4 Selection using other tags

Reports exist of ribozymes selected utilizing other types of tags. For example, a urea synthetase ribozyme was selected using a peptide-phosphonate. The selection substrate was activated at its N-terminus using weakly reactive N-hydroxysuccinimide carbamate ester. The ribozyme catalyzes formation of a urea bond between the peptide-phosphonate and the N⁴ of a cytidine nucleotide at its 3'-end. The selection substrate is a suicide inhibitor of human neutrophil elastase (HNE). It forms a covalent bond with a critical serine residue at the active site of the enzyme. HNE coupled beads were therefore used as the capture matrix for the active sequences (**Figure 1-7 A**). This capture method led to direct identification of urea synthase ribozymes with stereoselectivity towards the peptide substrate (Nieuwlandt et al., 2003).

Sharma et al developed another example of ribozymes selected using unusual affinity tags. The group reports using fluorescein-iodoacetamide as the selection substrate followed by immunocapture by anti-fluorescein antibody-coated magnetic beads (**Figure 1-7 B**). The approach was regarded as Immunoprecipitation (IP)-SELEX (Sharma et al., 2014).

Illangasekare and coworkers at Yarus' research group utilized a completely different approach for identification of a self-aminoacylating ribozyme. They reported selection of a 3'-selfaminoacylating ribozyme using phenylalanyl-AMP as substrate. The free α -amino group of 3'-phenylalanyl-ribozyme product was then conjugated to a hydrophobic naphthoxyacetyl group. The reacted ribozymes were then separated from the inactive species using reverse phase-high performance liquid chromatography (RP-HPLC) (**Figure 1-7 C**) (Illangasekare et al., 1995).

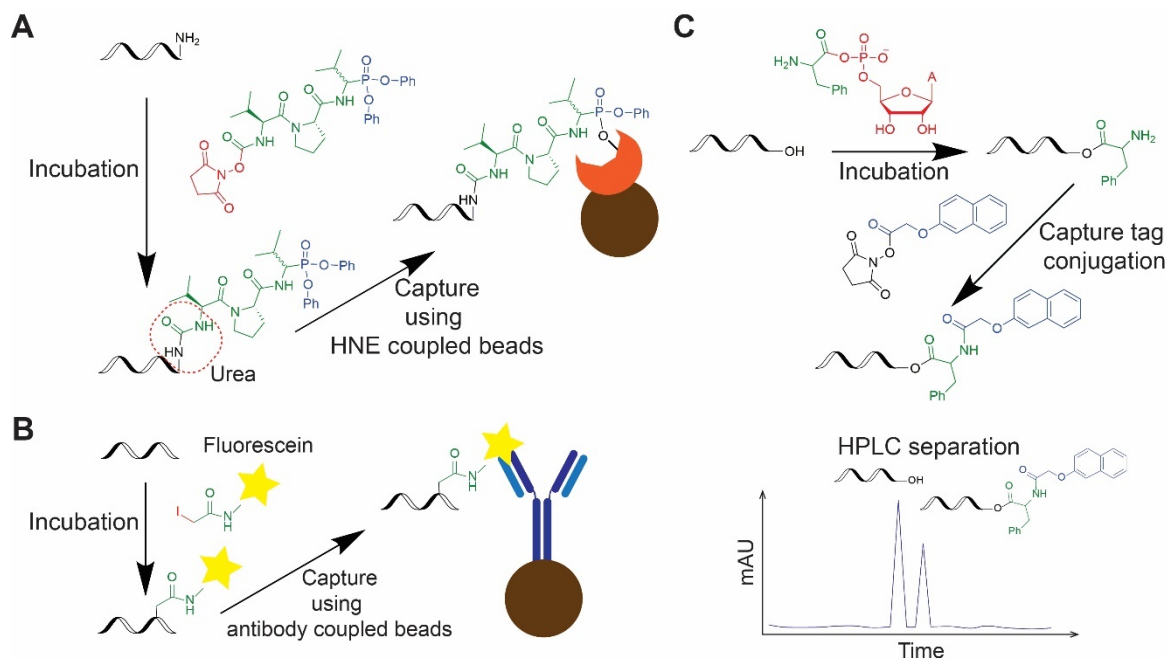


Figure 1-7 Unusual tagging strategies in ribozyme selection. **(A)** Selection of the urea synthase ribozyme using a suicide inhibitor of HNE. HNE coupled beads used for the capture step. **(B)** Selection of a self-alkylating ribozyme using fluorescein-iodoacetamide and IP-SELEX. **(C)** Capture by naphthoxyacetyl derivatization and HPLC purification.

1.3.3 Selection based on *in vitro* compartmentalization

The traditional selection methods have been used to evolve countless examples of RNA catalysts from large pools of RNA however, they are not without limitations. The ribozymes identified through these methods often fail to act as true enzymes due to the low turnover number or lack of it. The absence of trans-activity is also an issue regarding many ribozymes selected using these strategies (Griffiths and Tawfik, 2000).

In vitro compartmentalization methods facilitate direct selection of catalysts with high turnover numbers and improved activity in trans. These groups of strategies were initially applied to the directed evolution of protein enzymes (Griffiths and Tawfik, 2000) prior to RNA catalyst development. For ribozyme selection using these methods, the DNA templates coding for individual variants are trapped in separate compartments. The compartments are mostly in the form of droplets in a water-mineral oil emulsion. The templates are then amplified within the droplets and transcribed while the compartments are still maintained. The selection is designed in a way that the DNA template is tagged in-trans due to the activity of its corresponding transcript. The emulsion is then broken, and the tagged dsDNA templates are captured using an affinity matrix. The direct tagging of the dsDNA template has been used for improving turnover number of a Diels-Alderase ribozyme. One of the

reactants in this case was conjugated to the template rather than the RNA pool, before each round (Agresti et al., 2005). Compartmentalization based methods have also led to a dramatic increase in the processivity of an RNA polymerase ribozyme by direct tagging of the DNA template (Zaher and Unrau, 2007).

The dsDNA templates can also be tagged indirectly via co-immobilization, on beads with the tagged reaction product. This strategy was used for improving the trans-activity of an RNA ligase ribozyme. The capture step in this experiment was performed using FACS of the beads coated with individual ribozyme templates and the fluorescently labeled ligation product (Levy et al., 2005).

Due to the retainment of the DNA template in these techniques, the reverse transcription step is no longer needed. Omitting the reverse transcription step provides a higher likelihood for the selection of species with more complex structures. During the reselection of the Diels-Alderase ribozyme for example, a higher level of structural diversity was observed in isolated clones than the variants from the traditional selection (Agresti et al., 2005). This is due to the fact that reverse transcriptases are known to have difficulty reverse transcribing stably structured RNA. The RNA species with complex and stable structures are, therefore, at a selective disadvantage in conventional selection methods (Agresti et al., 2005; Joyce, 2004; Porter et al., 2017).

1.4 RNA imaging methods

Modern high throughput transcriptomic techniques such as microarray (Jaksik et al., 2015) and next-generation sequencing (Kukurba and Montgomery, 2015) provide invaluable information regarding changes in the cellular level of various RNA. These analyses, however, provide no information regarding RNA transport, localization, stability, interactions, and conformational dynamics. The diverse roles of RNA are inseparable from these aspects of RNA biology. Studying these phenomena and the underlying mechanisms leads to a more profound understanding of the function of coding and especially non-coding RNA. The study of such facets of RNA function and metabolism has only been made possible with the advent of RNA labeling and visualization techniques (Baker, 2012; Muthmann et al., 2020; Urbanek et al., 2014).

A wide range of tools has been developed that facilitate RNA visualization and tracking. The RNA labeling tools and strategies can be classified into two major groups of covalent and

non-covalent methods. In the following section, some of these methods will be described in detail. A major focus, however, is placed on the vector-encodable systems.

1.4.1 Non-covalent labeling methods

1.4.1.1 Hybridization based methods

1.4.1.1.1 FISH

This method is one of the oldest ways of visualizing specific RNA sequences in fixed cells. It is based on the use of a fluorescently labeled oligonucleotide complementary to a segment of the RNA of interest (ROI). The cells are typically fixed, and their membranes are permeabilized allowing for the entry of the oligonucleotide probe. The excess of the unbound probe is then washed away allowing sequence-specific visualization of the targeted RNA molecules (**Figure 1-8 A**). Signal intensity can be increased in this method by increasing the number of probes annealed to a single RNA. The method also facilitates simultaneous multi-color imaging of several target RNA when multiple probes are used, with each probe conjugated to a different fluorescent dye (Lawrence and Singer, 1986). This method however can only be used for static imaging of RNA localization in the cell and does not provide information about the dynamics of RNA metabolism or transport in live cells.

1.4.1.1.2 Molecular beacon

Like FISH, molecular beacon imaging method is based on complementary oligonucleotide probes. In this method the probe is flanked by short self-complementary segments that allow formation of a stem. A fluorophore is conjugated to one end of the probe and a quencher to another. When unbound the fluorophore and quencher are brought to proximity due to formation of a short basal stem. Binding of the probe to the target sequence prevents formation of this stem, and the fluorophore is unquenched (**Figure 1-8 B**). This method reduces the background signal quite significantly compared to FISH, obviating the need for a washing step when imaging fixed samples. It also makes possible imaging of RNA in live cell, however, the probe is not readily cell-permeable and it requires special cellular delivery methods (Monroy-Contreras and Vaca, 2011).

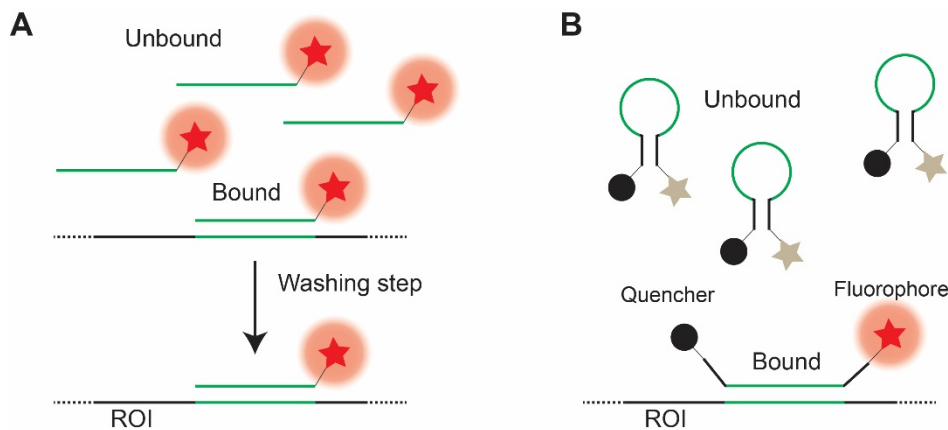


Figure 1-8 Hybridization-based RNA imaging methods. (A) Schematic representation of FISH. A washing step is needed to eliminate the background from the unbound probe. (B) Scheme of the molecular beacon imaging system. The unbound probe remains in a quenched state, thus, there is no need for the washing step.

1.4.1.2 RNA binding protein-based methods

1.4.1.2.1 MS2-tagging

Originally developed in 1998 by Singer and coworkers (Bertrand et al., 1998), MS2-tagging (aka. MS2-MCP system) has become the gold standard of live cell RNA imaging methods (**Figure 1-9**). Since its invention, this system has been extensively used in various types of cells including *S. cerevisiae* (Bertrand et al., 1998), *E. coli* (Golding et al., 2005), mammalian cell-lines (Ben-Ari et al., 2010; Darzacq et al., 2007; Grünwald and Singer, 2010) and even whole organisms such as *Drosophila* (Bothma et al., 2015) and mouse (Lionnet et al., 2011; Park et al., 2014). MS2-tagging is based on high affinity and specificity of the RNA bacteriophage MS2 coat protein (MCP) for its 19-nucleotide stem-loop binding site, in the bacteriophage genome (Valegård et al., 1997). 24 copies of the MS2 binding site (MBS) are typically inserted in the mRNA of interest. The MBS containing RNA is then co-expressed inside the cell along with the MCP fused to a fluorescent protein (MCP-FP). The MCP-FP fusion protein binds to each MBS, as a homodimer, therefore tagging the RNA of interest with several FP molecules (**Figure 1-9 A & B**). A nuclear localization signal (NLS) is also added to the MCP-FP fusion protein to ensure nuclear segregation of the unbound probe, therefore, reducing the cytosolic background (**Figure 1-9 A & B**). Apart from FPs, fluorogenic protein small molecule reactive/binding domains such as SNAP-tag and DHFR have also been used as MCP fusion (Carrocci and Hoskins, 2014). Optimized versions of this system have recently been reported in which the affinity of the MBS for MCP has been reduced. The linker sequence length between the stem-loops has also been optimized. These

systems allow more dynamic mRNA tracking with less interference on mRNA degradation rate (Tutucci et al., 2018a; Tutucci et al., 2018b).

Singer and coworkers also developed an analogous system based on bacteriophage PP7 coat protein (PCP) and its binding site (PBS) (Larson et al., 2011), allowing simultaneous two-color RNA tracking (Hocine et al., 2013). Combining these two systems facilitated engineering of a fluorogenic system based on split EGFP (Wu et al., 2014). These systems are denoted as trimolecular fluorescence complementation (TriFC) (**Figure 1-9 C**). More recently Park et al, have reported a system based on tripartite superfolder GFP that is only complemented and made fluorescent upon binding to the target RNA. Analogous to TriFC these systems are regarded as tetramolecular fluorescence complementation (TetFC) (Park et al., 2020). (**Figure 1-9 D**).

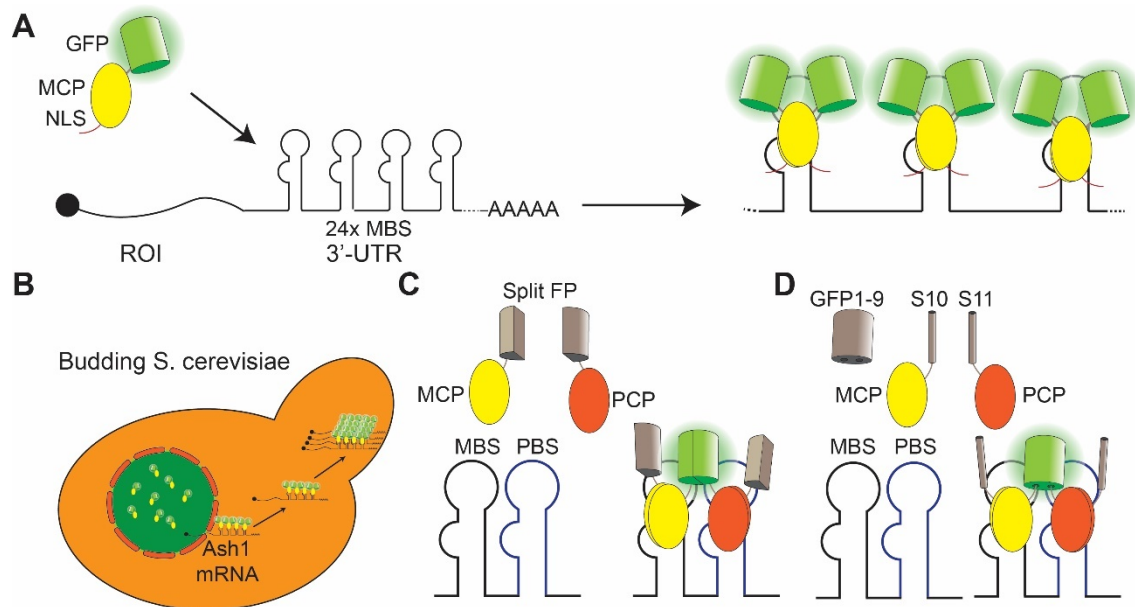


Figure 1-9 MS2-tagging system in RNA imaging. **(A)** MCP is fused to GFP. MBS is inserted in multiple copies in the 3'-UTR of the RNA of interest. Binding of the MCP-GFP to RNA facilitates imaging. NLS is to ensure accumulation, in the nucleus, of unbound MCP-GFP. **(B)** Schematic representation of MS2-tagging in tracking of the Ash1 mRNA in budding yeast (Bertrand et al., 1998). **(C)** And **(D)** Fluorogenic MS2-tagging systems based on trimolecular and tetramolecular complementation.

1.4.1.2.2 Pumilio domain

A drawback of the MS2-tagging is the requirement of exogenous expression of target RNA with multiple copies of MBS inserted in their 3'-UTR (George et al., 2018). An alternative system is based on the pumilio domain, which can be engineered for binding to any desired sequence. The RNA binding domain, pumilio, belongs to the PUF family of proteins. These proteins bind to their specific 8-nucleotide recognition motifs that lie in 3'-UTR of their

target mRNA. Upon binding they regulate translation and/or stability of their target mRNA. The pumilio domain consists of eight repetitive structural modules, each made of three α -helices. Three amino acid side chains in each module are responsible for recognition of a single nucleobase. Each nucleobase has a fixed binding code and therefore replacing these three amino acid residues alters the binding specificity towards a different sequence (Cheong and Hall, 2006). While recognition modules for A, G, and U naturally exist, the binding code for C nucleotide was evolved artificially (Dong et al., 2011). Various cellular and viral RNAs have been imaged in live mammalian and plant cells using pumilio based TriFC systems (**Figure 1-10 A**) (Adamala et al., 2016; Ozawa et al., 2007; Tilsner et al., 2009). Rentmeister's group has also engineered a TetFC strategy based on this domain (**Figure 1-10 B**) (Kellermann et al., 2013; Kellermann and Rentmeister, 2016). Binding specificity of the pumilio TetFC system has been confirmed using FACS on *E. coli* cells (Kellermann and Rentmeister, 2017).

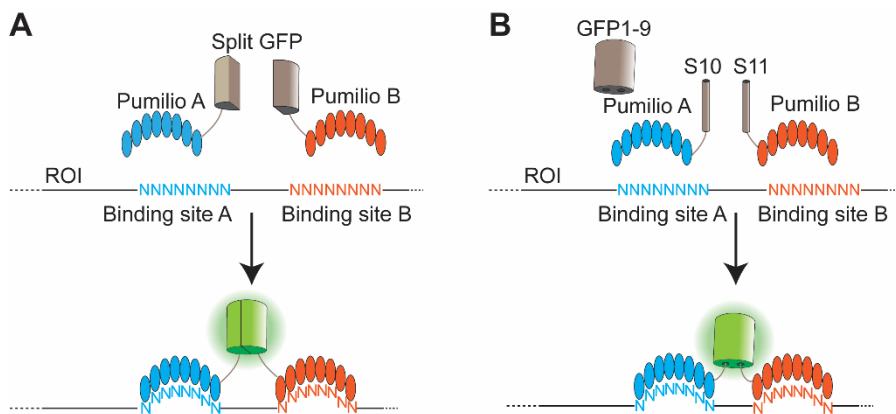


Figure 1-10 Pumilio-based RNA imaging system. **(A)** Scheme of the pumilio-based TriFC imaging system. **(B)** pumilio TetFC imaging system. Each pumilio domain is composed of eight modules. Each module is responsible for recognizing on nucleotide within the 8-nucleotide binding site.

1.4.1.2.3 CRISPR-Cas based systems

Cas proteins are a group of RNA guided endonucleases found in bacteria (Wiedenheft et al., 2012). They can be targeted towards any desired sequence simply by designing their guide RNA complementary to the sequence of interest. The originally discovered Cas proteins, such as *Streptococcus Pyogenes* Cas9 protein, were specific towards double-stranded DNA and therefore were extensively utilized for genomic editing applications (Savić and Schwank, 2016). Nuclease deficient mutant variants of Cas9 (dCas9) fused to various functional domains have been used for other purposes. The applications include regulation of gene expression (Qi et al., 2013) or introduction and removal of epigenetic marks (Lo and

Qi, 2017). dCas9, either in the form of an organic dye conjugate (Deng et al., 2015) or as a fusion with an FP (Chen et al., 2013), has been applied for imaging of the genomic loci.

Imaging and tracking RNA molecules in the cell using dCas9 based systems, is however more challenging for two main reasons: The first reason is that the dCas9-gRNA complex recognizes a double-stranded target sequence rather than the single-stranded RNA. The second reason is the requirement for the 5'-NGG-3' PAM motif on the non-target strand. Nelles and colleagues presented a solution to these problems by introducing PAM containing DNA oligonucleotides partially complementary to the target site. These oligonucleotides called PAMmers facilitated imaging of three different mRNAs using dCas9-GFP-gRNA complex (**Figure 1-11 A**). The absence of the PAM sequence in the genomic locus of the targeted RNA segment determines the specificity of the CRISP-dCas9-PAMmer system for RNA imaging (Nelles et al., 2016).

More recently Cas13a, formerly known as C2c2, was confirmed as a single-stranded RNA binding CRISPR-Cas complex (Abudayyeh et al., 2016). Cas13a allows selective cleavage or imaging of a target RNA without the need for the introduction of PAMmers (**Figure 1-11 B**) (Abudayyeh et al., 2017). dLwaCas13a has been successfully used by Abudayyeh et al, to image localization of abundant β -actin mRNA under stress conditions (Abudayyeh et al., 2017). In their more recent report, they tested Cas13a systems from various sources. As a result, dPspCas13 and dPguCas13 were revealed as more efficient RNA targeting systems (Yang et al., 2019). Since these Cas systems recognize mutually orthogonal gRNA sequences it is possible to use them for simultaneous dual-color imaging of RNA molecules in live cells (Yang et al., 2019).

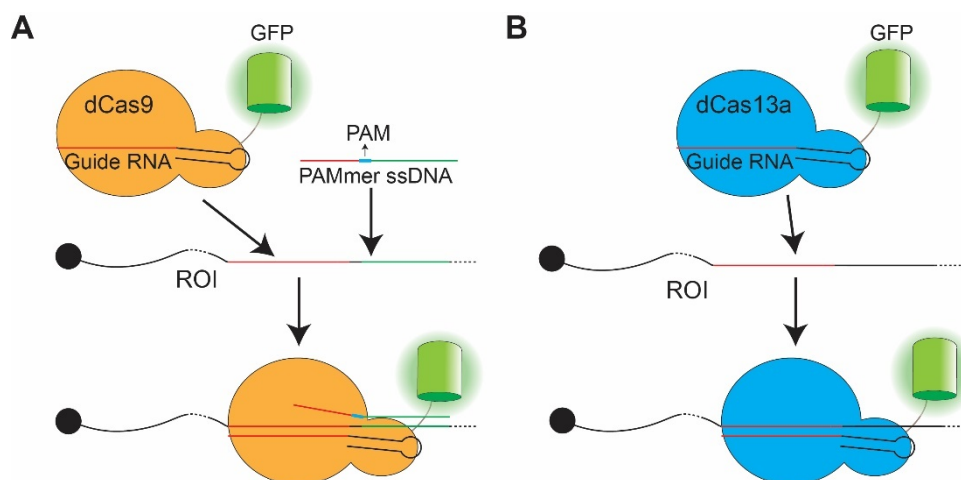


Figure 1-11 RNA guided CRISPR-Cas based imaging methods. (A) dCas 9 system requires addition of PAMmers for RNA binding. (B) dCAs13a based system allows direct RNA recognition.

1.4.1.3 Fluorogenic aptamers

While protein-based probes have been extensively studied and applied for imaging or tracking of various target RNAs the main limitation of these systems is the large size of probes and often-high background emission at the unbound state (Neubacher and Hennig, 2019; Rau and Rentmeister, 2017). The complexity of the system due to the number of the transfected vectors is also another disadvantage (George et al., 2018). Fluorogenic activating aptamers (FLAPs) have emerged as viable alternatives to the protein-based methods. Upon target binding, these aptamers activate or enhance fluorescent properties of their cognate ligand (Neubacher and Hennig, 2019). The fluorescent activation could be a direct result of the aptamer-dye complex formation or relieving of the quenching effect upon binding to dye-quencher conjugate (**Figure 1-12 A**) (Su and Hammond, 2020). Fluorogenic property of FLAPs turns them into ideal live-cell RNA imaging tools due to the low background of unbound dye. For live-cell applications, FLAPs are typically placed into a stably folded RNA scaffold. The FLAPs within the RNA scaffold are then inserted in multiple copies, into non-essential regions of the target RNA. (**Figure 1-12 B**) (Neubacher and Hennig, 2019). The scaffold minimizes the effect of the flanking sequences on aptamer folding and protects the aptamer from nuclease degradation (Martell et al., 2002; Ponchon and Dardel, 2007). The most frequently utilized scaffold is the human lysine tRNA (tRNA^{Lys3}) whose application is not limited to fluorogenic aptamers (Ponchon and Dardel, 2007). Aptamers are usually inserted into the anticodon loop of this tRNA. Filonov et al, demonstrated in 2015 that the tRNA scaffold undergoes endonucleolytic processing in human and *E. coli* cells. They therefore engineered an alternative RNA scaffold called F30. This scaffold promotes more stable levels of fluorogenic aptamer expression in cells (Filonov et al., 2015).

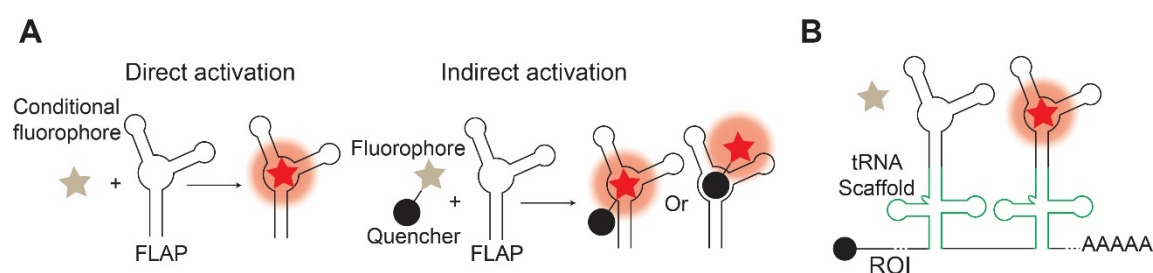


Figure 1-12 Fluorogen activating aptamers. **(A)** Direct vs indirect fluorogenic activation. **(B)** FLAPs are inserted into the RNA of interest (ROI), at multiple copies within the context of a stably folded scaffold.

1.4.1.3.1 GFP fluorophore analog binding aptamers

The malachite green binding aptamer, developed in 2003 (Babendure et al., 2003) is the first ever reported case of a FLAP. This class of functional RNAs however, became a major focus of research after Samie R Jaffrey's group published their report on FLAP development in 2011. The group revealed that certain RNA aptamers can activate fluorescence of synthetic GFP fluorophore (HBI) analogs upon binding (Paige et al., 2011). They used SELEX to evolve RNA aptamers with binding affinity towards various synthetic HBI analogs (**Figure 1-13**). The binders were then screened for their fluorogenic properties. These efforts led to the discovery of several FLAPs with binding specificity for various HBI analogs and distinct spectral properties. They named the most well characterized aptamer "Spinach". The aptamer emits green fluorescence when bound to its ligand, DFHBI (Paige et al., 2011). Other DFHBI derivatives have also been shown to bind this aptamer with slightly red-shifted fluorescence (Song et al., 2014). Spinach2 is a superfolder version of this aptamer developed via functional mutagenesis. This aptamer demonstrates improved fluorescence emission and thermal stability compared to the original Spinach (Strack et al., 2013).

X-ray crystallographic studies of Spinach revealed a binding core with a two-tiered G-quadruplex structure. Based on these studies Samie R. Jaffrey's group managed to rationally minimize Spinach by removing non-essential nucleotides (Warner et al., 2014). The minimized version was called baby Spinach and showed superior fluorogenic properties compared to Spinach and Spinach2.

Broccoli is another aptamer of the Spinach family, which was selected directly for fluorogenic activity. The selection was done using FACS on *E. coli* cells expressing a pre-enriched RNA pool of DFHBI-1T binders (**Figure 1-13**). Owing to this new selection strategy, Broccoli aptamer was directly selected for fluorogenic activation. Broccoli is less dependent on Mg^{2+} compared to spinach and was more stably folded and maintained in live cell. It has also been shown to be less dependent on the RNA scaffold when expressed in cell (Filonov et al., 2014).

Corn is an RNA analog of red fluorescent protein DsRed that was evolved in a similar fashion as broccoli (Song et al., 2017). It was selected against DFHO (Figure 1-13) and binds to its ligand as a homodimer (Warner et al., 2017).

Chili is another member of the HBI binding family of fluorogenic aptamers. This aptamer is based on the previously reported 13-2 aptamer (Paige et al., 2011) which was further minimized in Höbartner's group. Novel large Stokes shift HBI derivatives were tested for

binding to this aptamer and fluorescence activation. The aptamer produces a variety of emission colors upon binding to these ligands. The cationic nature of these HBI analogs resulted in improved affinity and less dependence on Mg^{2+} . Strongest fluorescence activation in the case of this aptamer is achieved with DMHBI⁺ dye (**Figure 1-13**) (Steinmetzger et al., 2019a; Steinmetzger et al., 2019b). Despite some promising features such as large Stokes shift and lower dependency on Mg^{2+} , this aptamer has not yet been tested in cellular systems.

1.4.1.3.2 Thiazole-orange binding aptamers

Standard thiazole orange dye, TO1, has a natural tendency to bind to double-stranded nucleic acids and become fluorescent (Nygren et al., 1998). Dolgosheina and coworkers synthesized TO1-biotin, based on TO1-acetate, which was previously known to show dramatically lower non-specific binding capacity for RNA or DNA. They then developed aptamer Mango which can bind and activate fluorescence of TO1-biotin (**Figure 1-13**) (Dolgosheina et al., 2014).

In 2018, Autour et al, reselected the round 12 pool from Mango aptamer selection through several rounds of a microfluidic in vitro compartmentalization (μ IVC) based approach (Autour et al., 2018). In this method, individual DNA templates of the pool were compartmentalized in single droplets. PCR amplification was then performed on these compartmentalized templates within the thermal cycling module of the microfluidic device. The droplets containing amplified templates were then fused to the ones containing the transcription mix and the TO1-biotin. After a certain incubation period the droplets were sorted based on their fluorescence emission. Mango II-IV were identified through these selection experiments with improved fluorescence and reduced dependency on salt concentration. Mango III was successfully used for live-cell RNA imaging within F30 and U6 scaffolds (Autour et al., 2018). Arrays of up to thirty-six mango II aptamers, without a scaffold, were used to localize β -actin mRNA and the non-coding NEAT-1 RNA in live cell (Cawte et al., 2020).

1.4.1.3.3 Silicon rhodamine binding aptamer

SiRA aptamer is one of the newest additions to the pallet of fluorogenic aptamers. The aptamer binds to the environment-sensitive silicon rhodamine dye (Wirth et al., 2019) (**Figure 1-13**). SirA binding stabilizes the zwitterionic state of the dye, which is significantly more fluorescent than the spirolactone form. A fluorescence activation of roughly 7-folds is

therefore observed. SirA can also bind to other 5-carboxy rhodamine dyes such as TAMRA and carborhodamine. Five copies of SirA within the tRNA scaffold, inserted into GFP mRNA, allowed successful localization of the mRNA using confocal and superresolution microscopy (Wirth et al., 2019).

1.4.1.3.4 DNB aptamers

Dinitroaniline (DN) acts as a general contact-mediated quencher when coupled to a wide range of fluorophores. Arora et al reported the selection of dinitroaniline binding aptamer called DNB (**Figure 1-13**). Since the aptamer is only specific for the quencher moiety, it can activate fluorescence of a wide range of fluorescent dyes albeit with varying efficiency. From the dye-DN conjugates tested, TMR-DN, SR-DN, and TR-DN showed the highest fluorescent enhancement. DNB transcribed in live *E. coli* cells within the context of a tRNA scaffold demonstrated fluorogenic activation of the tested Dye-DN conjugate. (Arora et al., 2015).

1.4.1.3.5 BHQ aptamers

Black Hole Quenchers (BHQs) have a wide absorption range throughout the visible light spectrum. As a result, they can strongly quench numerous fluorophores when physically coupled to them. BHQ binding aptamers have also been shown to activate the fluorescence of dye-BHQ conjugates (**Figure 1-13**) (Murata et al., 2011). These aptamers were designed for target RNA recognition in trans. In this system hybridization to the target RNA stabilizes the basal stem of the BHQ binding aptamer resulting in dye-quencher binding and fluorescence activation (Sato et al., 2015).

1.4.1.3.6 SRB-2 aptamer

SRB2 is a rhodamine binding aptamer originally selected for binding to sulforhodamine (**Figure 1-13**) (Holeman et al., 1998). The aptamer lacks direct fluorogenicity, however, when bound to dinitroaniline conjugated sulforhodamine, it could induce a fluorescent turn-on of roughly 100-folds (Sunbul and Jäschke, 2013). Binding affinity of this aptamer is not limited to sulforhodamine but rather it can bind to a variety of organic dyes with “xanthene-like” core structure. Coupling of each of these dyes to dinitroaniline quencher led to SRB-2

mediated fluorescence activation. SRB-2 aptamer has been used for RNA imaging in live *E. coli* cells and mammalian cell-lines (Sunbul and Jäschke, 2018).

1.4.1.3.7 o-Coral

Inspired by strong fluorogenic activation of SRB-2 aptamer, Bouhedda et al attempted a further improvement. They synthesized a dimeric self-quenched form of sulforhodamine B called Gemini-561. Through alternating rounds of PCR mutagenesis, SELEX, and functional μ IVC, o-Coral was isolated for binding to Gemini-561 and direct fluorescence activation. The o-Coral aptamer is essentially a dimerized SRB-2 aptamer with a few mutations. Binding of this aptamer to Gemini-561 results in roughly 10 times fluorescence enhancement presumably by relieving the self-quenching effect (**Figure 1-13**). Due to the exceptional brightness of the aptamer-dye pair, even a single copy of o-Coral facilitated imaging and tracking of Pol III and Pol II transcribed RNA in live mammalian cells (Bouhedda et al., 2020).

1.4.1.3.8 Riboglow

Riboglow is another example of quencher binding aptamers derived from the natural bacterial cobalamin (Cbl) riboswitch (Brasemann et al., 2018). Brasemann and colleagues reported that Cbl could act as an efficient quencher for a variety of synthetic fluorophores. The dye is conjugated via a linker, to the 5'-hydroxyl group of Cbl, which remains accessible when bound to the riboswitch (**Figure 1-13**). The binding of the aptamer to Cbl turns on the fluorescence of the conjugated dye. Riboglow system has several advantages over artificially selected aptamers. Since it is derived from a natural source, the aptamer can retain its active folded structure within the cell without the need for an additional scaffold. This system has been used successfully, to localize β -actin mRNA and U1 snRNA. The main disadvantage of this system is the cell impermeability of the Cbl-fluorophore conjugates and the rather low turn-on ratio. Special delivery methods such as bead-loading is therefore required to introduce the fluorescent probe into the cell (Brasemann et al., 2018).

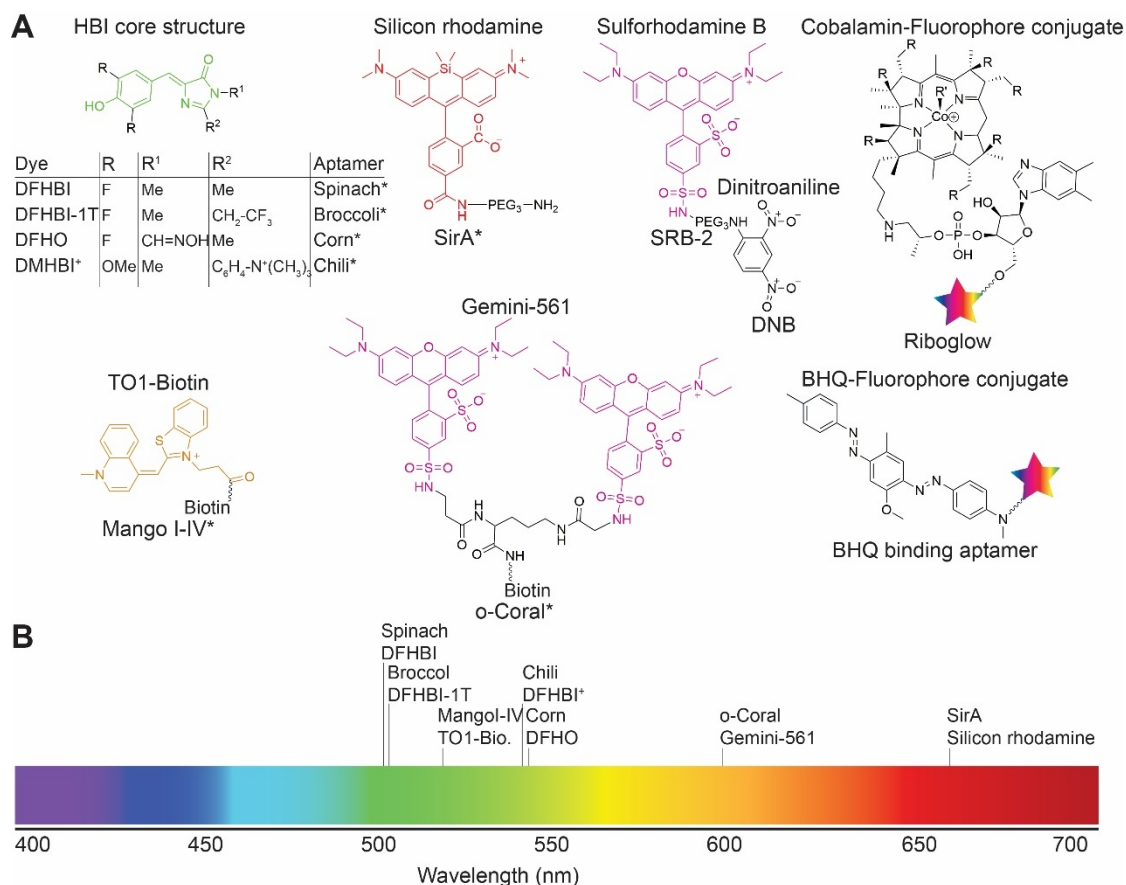


Figure 1-13 Fluorogenic aptamer-dye pairs (**A**) Chemical structure of the ligands recognized by FLAPs. Written on top of each structure is the name of the ligand and the name of the binding aptamer is written at the bottom of each structure. Asterix indicates direct fluorogenicity of the aptamer (**B**) approximate emission maxima of the direct FLAP-dye pairs.

1.4.2 Covalent methods

1.4.2.1 Polymerase mediated incorporation of labeled nucleotides

RNA polymerases can incorporate NTP analogs with certain modifications *in vivo* and *in vitro* (Anhäuser and Rentmeister, 2017; Holstein and Rentmeister, 2016). Pyrimidines with bioorthogonal modification at C5, such as alkyne or azide are some of the most frequently used examples of these types of NTPs (Asare-Okai et al., 2014; Qu et al., 2013; Sawant et al., 2016). The alkynyl and azido groups allow derivatization using copper(I)-catalyzed alkyne-azide cycloaddition reaction (CuAAC) (Muthmann et al., 2020). The azido moieties can also be derivatized using strain-promoted alkyne-azide cyclo-addition (SPAAC) (**Figure 1-14 A**) (Muthmann et al., 2020). 4-sUTP is another example of pyrimidines that has been successfully incorporated into RNA and derivatized by various means (Duffy et al., 2019). Purine analogs such as N⁶-propargyl-ATP have also been used to derivatize RNA transcripts in live cell. These analogs can be incorporated either by cellular polymerases or via the action

of the Poly-A polymerase (Grammel et al., 2012). 2'-Azido-ATP (Anhäuser et al., 2019) and 2-ethynyl-ATP (Curanovic et al., 2013) on the other hand have been incorporated specifically into the poly-A tail of mRNA via the action of poly-A polymerase.

Polymerase based approaches do not result in site or sequence-specific labeling of RNA, but rather result in random labeling of all actively transcribed RNAs. Unnatural orthogonal base-pairs have been used for site- and transcript-specific labeling of RNA *in vitro*, however, application of these systems for live-cell RNA imaging is extremely challenging (Hamashima et al., 2018).

1.4.2.2 RNA modifying enzymes

1.4.2.2.1 RNA methyltransferases

Numerous examples of methyltransferase enzymes have been applied for labeling of various target molecules such as proteins and DNA and more recently RNA (Holstein and Rentmeister, 2016; Lukinavičius et al., 2007; Muthmann et al., 2020). The common feature among these enzymes is the use of S-adenosyl-methionine as the methyl group donor. A number of these enzymes have been reported in literature, which can transfer extended alkyl chains from the various synthetic analogs of SAM to their target molecule. The most notable examples of such extensions include bioorthogonal functionalities and chemical probes (Muthmann et al., 2020). Motorin et al published the earliest report on application of such enzymes for RNA labeling. They reported that the RNA methyltransferases Trm1 and Trm11 could transfer beta-unsaturated carbon chains with terminal alkynyl functionality. Trm1 and Trm11 modify N²-groups of tRNA^{Phe} at positions 26 and 10, respectively (**Figure 1-14 B & C**). Alexa fluor 594 azide was then subsequently conjugated to these modified positions using CuAAC reaction (Motorin et al., 2011).

Protein engineering techniques can sometimes be employed to improve the reactivity of RNA methyltransferases towards the unnatural SAM analogs. Schulz et al, for example engineered *Giardia lamblia* trimethyl guanine synthase 2 (Gla-Tgs2) for efficient labeling of mRNA cap-structures (Schulz et al., 2013). The engineered enzyme Gla-TGs2-Var1 demonstrates three times higher turnover number, K_m and k_{cat} for *S*-AdoPropen compared to the wild type. Such improvement was achieved with only one single amino acid substitution (V34A). Gla-Tgs2-Var1 was then applied for functionalization of the N²-position in the m⁷G of the mRNA cap structure. The groups transferred using this enzyme include alkynyl (Schulz et al., 2013), azido (Holstein et al., 2014), and 4-vinylbenzyl (Holstein et al., 2015)

functionalities. The functionalized cap can subsequently be derivatized with chemical probes using CuAAC, SPAAC, or iEDDA (Inverse electron-demand Diels-Alder) reactions (**Figure 1-14**).

Ecm1 is another cap methyltransferase with an inherent ability to transfer bulky moieties from their corresponding SAM-analogs. This enzyme has been exploited by Rentmeister's research group, to transfer a wide variety of bioorthogonal functionalities to N⁷ of the cap guanine (Holstein et al., 2016; Muttach et al., 2017). Functional groups as bulky as a norbornene connected to a benzylic linker have been transferred using this enzyme (**Figure 1-14**). The modified cap-structure was then labeled using rapid and fluorogenic iEDDA reaction in cell lysate (Muttach et al., 2017).

Perhaps the most versatile example of methyltransferase enzymes reported so far is the C/D-box dependent archaeal methyltransferase. This RNA guided enzyme alkylates internal 2'-OH positions of RNA based on complementarity to the guide C/D box guide RNA. The approach shows great promise for versatile site-specific labeling of RNA. However the reactivity of the enzyme is extremely low towards the tested SAM analogs (**Figure 1-14**) (Tomkuvienė et al., 2012). This drawback may be circumvented using protein engineering methods, however, no further development has so far been reported on this system.

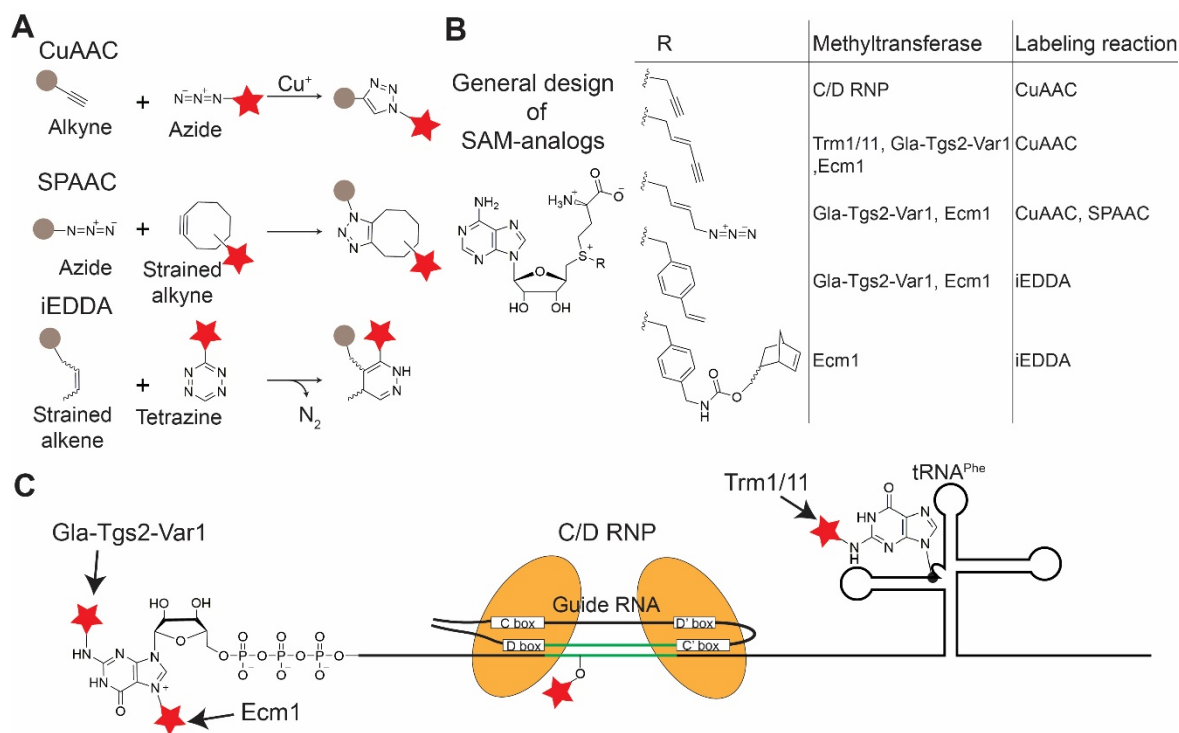


Figure 1-14 RNA methyltransferases in RNA labeling. **(A)** Common bioorthogonal reactions employed for indirect labeling of RNA, derivatized using RNA modifying enzymes. **(B)** General architecture of SAM analogs and the side chains transferred using various methyltransferases. **(C)** Schematic depiction of RNA methyltransferase-catalyzed RNA labeling. Gla-Tgs2-Var1 and Ecm1 are specific for mRNA cap structure. C/D RNP can be targeted towards different RNA sequences based on guide RNA complementarity. Trm1 and Trm11 methyltransferases modify specific G-nucleotides within their cognate tRNA.

1.4.2.2.2 Other tRNA-modifying enzymes

Except for RNA methyltransferases other classes of tRNA modifying enzymes also exist which have shown great potential for RNA labeling. The archaeal tRNA^{Ile2}- agmatidine synthetase (Tias) is one example of such enzymes. The enzyme naturally catalyzes ATP dependent substitution, with agmatine, of the O² at C34 of its cognate tRNA (**Figure 1-15 A**) (Osawa et al., 2011). Li et al demonstrated that the enzyme can readily accept a variety of alkynyl or azido functionalized agmatine analogs, including propargylamine (**Figure 1-15 A**). They then co-expressed Tias with a 5S rRNA-tRNA^{Ile2} fusion transcript in a mammalian cell-line. The cells were then treated with propargylamine. Following this treatment, the cells were fixed. CuAAC reaction was then used to label propargylamine modified RNA in situ with SulfoCy5-Azide (**Figure 1-15 B**) (Li et al., 2015).

Another tRNA modifying enzyme-based system, developed by Alexander et al, is called RNA-transglycosylation at guanine (RNA-TAG) (Alexander et al., 2015). In this method, bacterial tRNA guanine transglycosylase (TGT) is taken advantage of for RNA labeling.

This enzyme is responsible for the substitution of specific guanine nucleotides with the noncanonical guanine analog PreQ₁ (**Figure 1-15 C**). Various PreQ₁ analogs with modifications at their exocyclic primary amine have been shown to be efficiently accepted as TGT substrate. Direct attachment of bulky moieties such as Cy7, bodipy, thiazole-orange, or biotin has also been possible using this enzyme. In case of the thiazole-orange conjugated PreQ₁, a 40-fold increase in fluorescent intensity was observed upon RNA labeling. The enzyme recognizes a 17-nucleotide stem-loop called ECY-A1 as the core target sequence. Insertion of ECY-A1 into the 3'-UTR of the mCherry mRNA led to its TGT mediated labeling *in vitro* (**Figure 1-15 D**) (Alexander et al., 2015). Photocleavable groups have also been attached to 5'-UTR of an *in vitro* transcribed EGFP mRNA. The mRNA was transfected into a mammalian cell-line and its translation was induced upon photocleavage (Zhang et al., 2018).

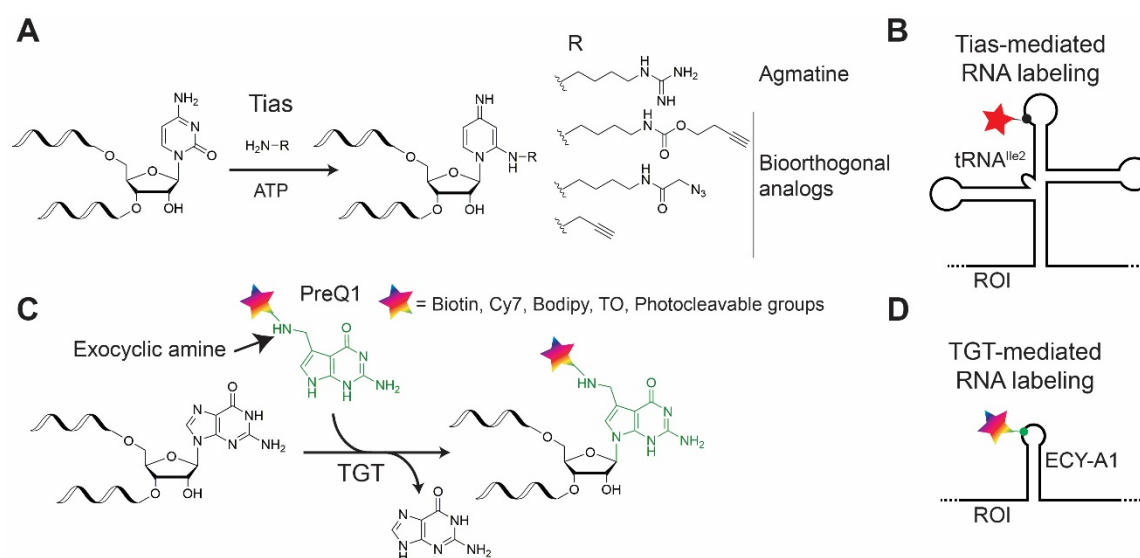


Figure 1-15 tRNA-modifying enzyme-mediated RNA labeling. **(A)** Agmatidine synthetase activity of Tias and the substrate scope of the enzyme. **(B)** Insertion of the tRNA^{Ile2} into the RNA of interest (ROI), allows Tias-mediated two-step labeling of target RNA. **(C)** TGT mediated RNA labeling. The exocyclic amine of PreQ₁ substrate can be derivatized directly with a variety of bulky moieties. **(D)** Insertion of the ECY-A1 stem-loop into ROI allows TGT mediated direct labeling of RNA.

1.4.2.3 Catalytic nucleic acid-based RNA labeling

1.4.2.3.1 Ribozyme based methods

Self-alkylating fluorescein-iodoacetamide reactive ribozyme selected by Sharma et al, is an example of a catalytic RNA that was directly selected for RNA labeling. The selection was performed using fluorescein-iodoacetamide and IP-SELEX (**Figure 1-7 B**) as mentioned in **1.3.2.4**. The outcome of this selection were two ribozymes denoted as 1FR1 and 5FR1. These

ribozymes were able to self-alkylate at an undetermined internal position, in the presence of fluorescein-iodoacetamide. The fusion of these two ribozymes to a transcript coding for a part of the mCherry protein resulted in the labeling of the fusion construct *in vitro* and in cell lysate (**Figure 1-16 A**) (Sharma et al., 2014). The high specificity of these ribozymes towards Fluorescein-IA limits their application to single-color and single-probe labeling reactions.

Another example of self-alkylating ribozymes with the potential for RNA labeling is a 42-nucleotide long RNA catalyst derived from *Aeropyrum pernix* genomic RNA sequences (McDonald et al., 2014). To identify this ribozyme, McDonald et al. used a mixture of eight biotinylated compounds with electrophilic functionalities, in combination with an RNA pool of transcriptomic fragments derived from 9 different species. The self-alkylating ribozyme was isolated within 6 rounds of selection. Further characterization on this ribozyme revealed its specific reactivity towards di-substituted cis-epoxides (**Figure 1-5 B**). The self-alkylation was still observed even when the biotin was replaced by other moieties containing alkynyl or azido functionalities or even fluorescent dyes such as TAMRA. 5S rRNA and ASH1 mRNA were transcriptionally fused to this catalyst and were shown to be successfully labeled as pure transcript or in the context of total cellular RNA (**Figure 1-16 B**) (McDonald et al., 2014).

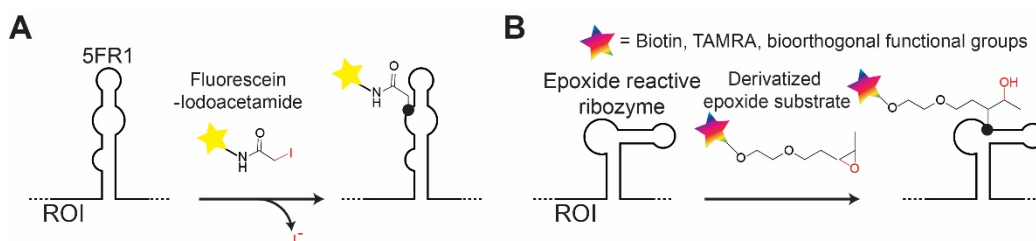


Figure 1-16 Cis-reacting self-alkylating ribozymes as RNA labeling tools. **(A)** Fluorescein iodoacetamide reactive ribozyme can be inserted into target RNA, allowing fluorescein labeling at internal positions. **(B)** Insertion of the epoxide reactive ribozyme into the ROI allows labeling using various epoxide containing chemical probes.

The Twin-ribozyme is an engineered form of the hairpin ribozyme, which has also been applied to RNA labeling *in trans*. As mentioned in 1.2, the hairpin ribozyme is a self-cleaving ribozyme. The cleavage reaction generates a 2'-3' cyclic phosphodiester and a 5'-OH at the ends of the RNA fragments, which can be rejoined, since the hairpin ribozyme can also catalyze the re-ligation reaction (Hampel and Tritz, 1989). The ribozyme can be engineered into a *trans*-acting variant consisting of a conserved catalytic core and a pair of binding arms.

The arms can be used to target the ribozyme to RNA sequences of interest (Walter and Burker, 1998). This engineered variant can catalyze a similar cleavage/re-ligation reaction on an independent RNA molecule in-trans. The twin-ribozyme was developed by tandem-connection of two copies of the hairpin ribozyme via a linker sequence (Schmidt et al., 2000; Welz et al., 2003). The arms and the linker are designed complementary to the target RNA. Upon hybridization, twin-ribozyme facilitates excision of an RNA segment and its replacement with a synthetic piece carrying a label at a desired position. For re-ligation to occur, the ligated patch needs to carry a 2'-3'cyclic phosphodiester at its 3'-end. Since the original piece can also be ligated back, the linker is designed to form an unpaired tetraloop when bound to target RNA. The synthetic patch, however, is made fully complementary to the linker sequence. This is to ensure that the synthetic piece can replace the excised fragment more efficiently (**Figure 1-17**) (Vauléon et al., 2005).

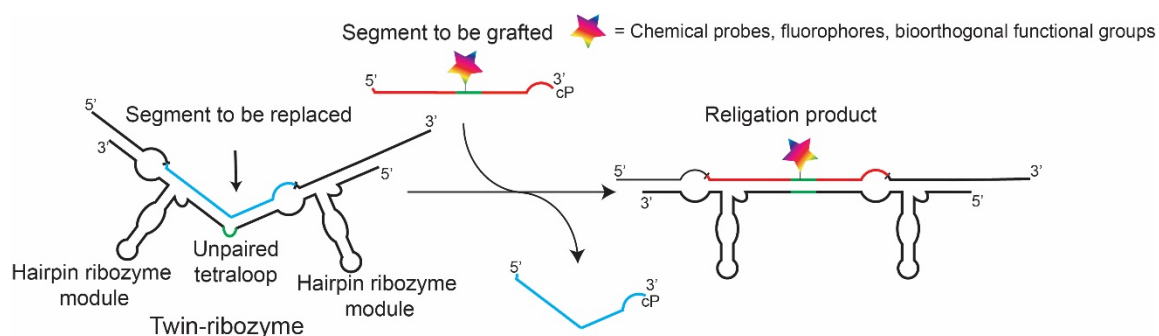


Figure 1-17 Twin-ribozyme based RNA labeling. Twin-ribozyme labels target RNA via strand exchange mechanism. A synthetic RNA piece carrying a variety of labels (red) can be used to replace a target segment (cyan). 3'-cyclic phosphodiester (cp) is essential to the targeting reaction. The tetraloop region (green) remains unpaired to the original segment while pairing to the synthetic piece to allow efficient strand exchange.

Samanta et al, report another RNA labeling method using a polymerase ribozyme (Samanta et al., 2018). Polymerase ribozymes are a class of artificial RNA catalysts that can catalyze template-based extension of DNA or RNA primers using NTPs (McGinness and Joyce, 2003). Samanta et al, exploited polymerase ribozyme 24-3, for template-dependent addition of labeled nucleotides to 3'-end of RNA or DNA (**Figure 1-18**). The RNA of interest in this system acts as a primer annealed to a template oligonucleotide. The template oligonucleotide and the ribozyme are designed complementary at their 5'-end (**Figure 1-18**). The segment of the ribozyme complementary to the template is called tag sequence. This complementarity facilitates primer-template pair recognition by the polymerase ribozyme. The template is designed in a manner that only a single nucleotide extension to the RNA of interest is permissible. The NTP analogs added using this strategy could bear a diverse set of

functionalities and modifications at various positions. This system has only so far been used *in vitro*. Live-cell application of this ribozyme is challenging due to the lack of orthogonality towards natural NTPs (Samanta et al., 2018). the application of this ribozyme is also only limited to the 3'-end labeling. This may cause additional complications particularly when it comes to specific targeting of 3'-polyadenylated RNA.

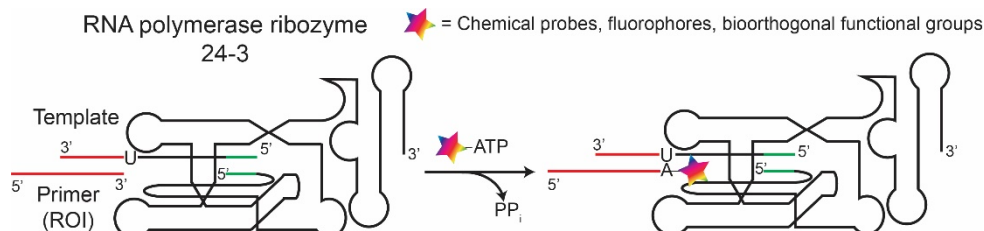


Figure 1-18 24-3 polymerase ribozyme mediated RNA labeling. Tag sequence (green) at 5'-end of the ribozyme is designed complementary to the 5'-end of the template. The ROI in this method acts as a primer. Annealing of ROI to the template allows one nucleotide extension in the presence of the appropriate NTP analog. The NTP analog can carry a variety of labels at various positions.

1.4.2.4 Deoxyribozyme based RNA labeling methods

DNA has also been shown to possess catalytic potential like RNA. No natural catalytic DNA sequence has so far been identified, and therefore all the existing deoxyribozymes have been artificially developed through *in vitro* selection efforts (Silverman, 2009). PAGE purification seems to be the prevalent method of choice for deoxyribozyme selection. Many deoxyribozymes with a wide variety of activities have been selected so far. The instances of these functions include RNA (Santoro and Joyce, 1997) or DNA (Carmi et al., 1996) cleavage, RNA ligation (Coppins and Silverman, 2004b), amide and ester bond hydrolysis (Brandsen et al., 2013), depurination (Höbartner et al., 2007), tyrosine phosphorylation (Dokukin and Silverman, 2014) or nucleotidylation (Wang and Silverman, 2016) and so on.

A 3'-5' RNA ligase deoxyribozyme, called 9DB1, the first deoxyribozyme whose crystal structure has been resolved (Ponce-Salvatierra et al., 2016), was applied for RNA labeling *in vitro*. The labeling is performed through ligation of a labeled synthetic RNA piece to a 5'-triphosphorylated *in vitro* transcribed RNA (Büttner et al., 2013). The more extensively utilized deoxyribozyme is 10DM24 with 2'-5'-ligation activity which leads to the formation of a branched RNA (Zelin et al., 2006). 10DM24 was used for RNA labeling at 2'-hydroxyl of specific internal adenosines, using short labeled oligonucleotide (Baum and Silverman, 2007). Further engineering of these catalysts led to its ability to ligate single GTP to the target RNA forming a 2'-5'-mononucleotidyl branch (Höbartner and Silverman, 2007). The

incoming GTP is recognized by Watson-Crick base-pairing to a C nucleotide in the 10DM24 catalytic core. Mutations to this nucleotide allows ligation of other NTPs. Purine based NTPs however are more efficiently ligated (Höbartner and Silverman, 2007). In 2014 Büttner et al showed that 10DM24 tolerates a wide range of modifications at 2'-position of GTP allowing efficient ligation of various derivatizations such as bioorthogonal functional groups or even bulky fluorescent dyes such as Cy3 and Cy5 via a (d)GMP linker (**Figure 1-19**) (Büttner et al., 2014). 10DM24 based RNA labeling is possibly the most versatile and efficient method for labeling of unmodified RNA. Targeting of the position to be labeled is based on Watson-crick base-pairing of a pair of binding arms to the target position. The arms flank an unpaired adenosine which serves as the modification site for the deoxyribozyme. 10DM24 allows site-specific and efficient installation of relatively small chemical probes onto the target RNA without the need for insertion of exogenous segments. The effects of this labeling technique on the structure and function of the target RNA is therefore minimal. This deoxyribozyme however has never been applied in live cell due to challenges regarding delivery and maintenance of single stranded DNA in cells. The other limitations of the catalyst is the requirement of a helper oligonucleotide for formation of a rather complicated 3-helix junction structure (**Figure 1-19**) and toxic Tb^{3+} ions for optimal activity.

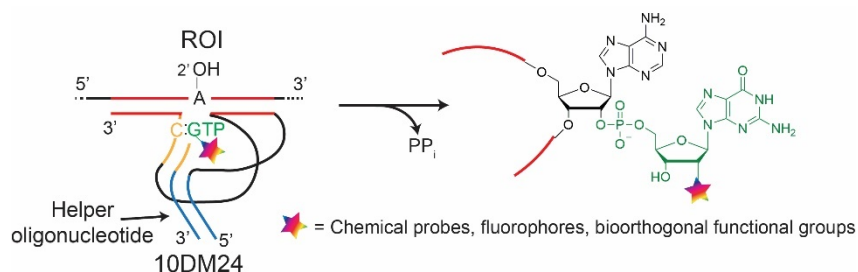


Figure 1-19 10DM24 mediated labeling of an RNA of interest. The target RNA is recognized by a pair of binding arms (red). The modified A-nucleotide remains unpaired and, therefore, bulges out. A helper oligonucleotide pairs to a third binding arm (light blue). The first four nucleotides of the helper oligonucleotide pair to a segment of 10DM24 catalytic core (light orange). This arrangement leads to the formation of a three-helix-junction surrounding the targeted A-nucleotide. GTP nucleotides carrying a vast range of functionalities and labeled at their 2'-positions can be ligated to the bulged A using this method. The recognition of GTP is made via W:C pairing to a C nucleotide in the catalytic core.

2 Aim of the study

Various RNA labeling and visualization methods have been developed over the last few decades. Despite the advances made in the field of RNA imaging, flexible and site-selective labeling strategies are still in high demand. The main limitation of most existing non-covalent and covalent methods is the need for insertion of exogenous sequences into the target RNA. On the other hand, low labeling efficiency is an obstacle in case of more generalizable covalent labeling methods such as the C/D RNP methyltransferase approach (Tomkuvienė et al., 2012).

Among the tools reviewed in 1.4, the 10DM24 deoxyribozyme RNA labeling system offers an attractive choice. This system provides one of the most versatile and flexible options for site-specific RNA labeling at internal adenosine nucleotides. The deoxyribozyme can be easily targeted to desired positions through Watson-Crick complementarity of its recognition arms to the target section (Büttner et al., 2014; Höbartner and Silverman, 2007). The deoxyribozyme also offers a wide range of possibilities regarding the choice and type of the label. 10DM24 readily accepts GTP substrates conjugated to various types of small and bulky modifications (Büttner et al., 2014). Additionally, the reaction efficiency of 10DM24 is exceptionally high, compared to other systems guided by complementarity such as the C/D RNP methyltransferase (Tomkuvienė et al., 2012) or the Twin-ribozyme (Vauléon et al., 2005). The labeling reaction does not require manipulation of the linear RNA sequence and the size of the label is quite small. The interference of such labeling system in function or folding of the target RNA is therefore expected to be minimal (Büttner et al., 2014). 10DM24 however has only been used for RNA labeling in vitro or in cellular extracts supplemented with catalyst's metal ion cofactors.

Tools with similar properties as 10DM24 for cellular applications have not yet been reported. Thus, as the main goal of this thesis, we decided to develop ribozyme analogs of 10DM24 with potential for future cellular application. The main advantage of ribozymes over deoxyribozymes is their vector encodability. RNA catalysts can be converted into their coding dsDNA sequences, which can be cloned into appropriate expression vectors. The resulting “ribozyme gene” can be transcribed into the active ribozyme within the cell (Park et al., 2019). Engineered self-cleaving ribozymes have been reported in numerous examples that have been encoded into DNA vectors and successfully expressed within the cell (Felletti et al., 2016; Ogawa and Maeda, 2008).

For ribozymes with similar function as 10DM24, we imagined a pair of recognition arms that recognize the target RNA sequence, based on complementarity. Analogous to 10DM24 the ribozymes should also possess catalytic cores with broad labeling substrate scope (**Figure 2-1**). In this thesis, we designed a selection process through which 10DM24-like ribozymes can be evolved. There are undesirable aspects to 10DM24 based system as well such as the dependency on Tb^{3+} and the complex three-helix junction structure (Büttner et al., 2014). We, therefore, focused on developing ribozymes that do not require such metal cofactor and have more simplified structures.

Our first objective was to establish and optimize the selection process. At first, as proof of principle, we initiated a selection process using ATP based substrates. This selection experiment led to the identification of adenylyltransferase ribozymes with many desirable properties. We then switched our focus to substrates more orthogonal to the biological system, such as antiviral nucleotide analogs. The FDA approved antiviral ATP analog, tenofovir-diphosphate (Gallant and Deresinski, 2003), was chosen in the next stage as a candidate for ribozyme development. Unlike ATP analogs, tenofovir is not enzymatically incorporated into RNA. Presence of a phosphonate instead of α -phosphate in this analog, would also result in formation of a less enzymatically labile linkage. Selection using tenofovir-diphosphate-based substrates was initiated to develop tenofovir transferase ribozymes.

The developed ribozymes were then characterized and tested for their RNA labeling potential *in vitro*.

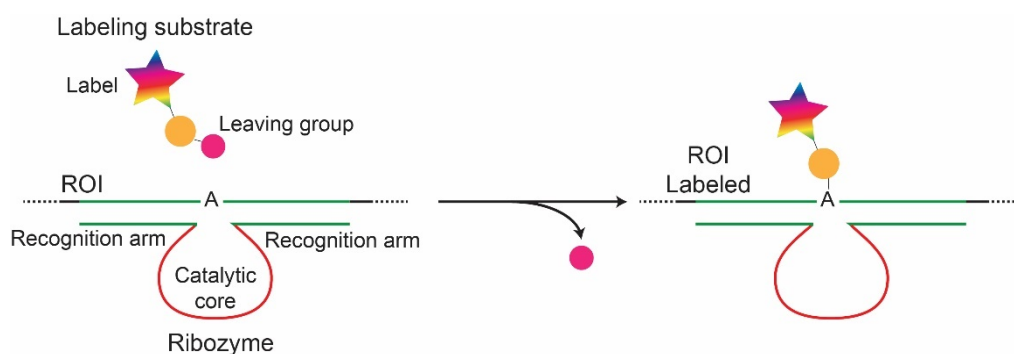


Figure 2-1 Schematic representation of RNA-Catalyzed site-specific RNA labeling.

3 Results

3.1 General design of the selection strategy

3.1.1 The design of the selection pool

The selection pool was designed to include five major segments (**Figure 3-1 A**): **1)** A hypothetical substrate sequence to which the modification was to be directed. **2)** A 14-nucleotide connecting loop. **3)** A pair of binding arms. **4)** A 40-nucleotide random region. **5)** An 8-nucleotide unpaired 3'-extension.

The 19-nucleotide substrate sequence was designed in a way that it could not form stable intramolecular secondary structures based on mfold predictions (Gallant and Deresinski, 2003). The substrate sequence was initiated with a 5'-GG, to serve as transcription start-site for T7 RNA polymerase. The recognition arms, flanking the 40-nucleotide random region, were designed complementary to the substrate sequence. They paired to the substrate sequence leaving the desired modification-site adenosine unpaired. The adenosine nucleotide was placed within a 5'-GAG-3' context. This design was analogous to the complex formed by 10DM24 and its target sequence, in which the modification site A is bulged out. GAG modification context is also labeled most efficiently by 10DM24 (Büttner et al., 2014). To facilitate priming during the reverse transcription step, the 3'-binding arm was further extended by an additional 8-nucleotides.

To ensure the physical connection of the modified substrate sequence to the evolved catalytic core a 14-nucleotide connecting loop was introduced. The sequence of the connecting loop was also designed arbitrarily. The mfold secondary structure predictions confirmed that the loop does not interfere with the secondary structure of the constant region of the pool.

3.1.2 Construction of the selection pool

The DNA template for the starting pool was 119 nucleotides long, including the T7 promoter. A DNA sequence of this size is challenging to synthesize with high efficiency and accuracy using standard solid-phase synthesis method, and was thus constructed from two synthetic fragments (**Figure 3-1 B**). The first piece (forward pool primer), was synthesized in the sense direction, and included the T7 promoter, the substrate sequence, and the connecting loop. The other piece (reverse pool primer), was synthesized in the antisense direction, coded for the binding arms and the random region and the connecting loop. The

two pieces were complementary at their 3'-end in the connecting loop segment (**Figure 3-1 B**). This complementarity allowed overlap synthesis of the full-length pool template using Klenow fragment. The 14-nucleotide length of the connecting loop was chosen to provide enough priming range for efficient overlap synthesis. The substrate sequence and the binding arms were not included in the overlap region to avoid 3'-self-complementarity of the primers. After Klenow extension, 450 pmol of the double-stranded DNA pool (corresponding to $\sim 2.7 \times 10^{14}$ unique variants) was transcribed in vitro. The transcription yielded ~ 17 nmol of the starting pool used in selection experiments (**Figure 3-1 B**).

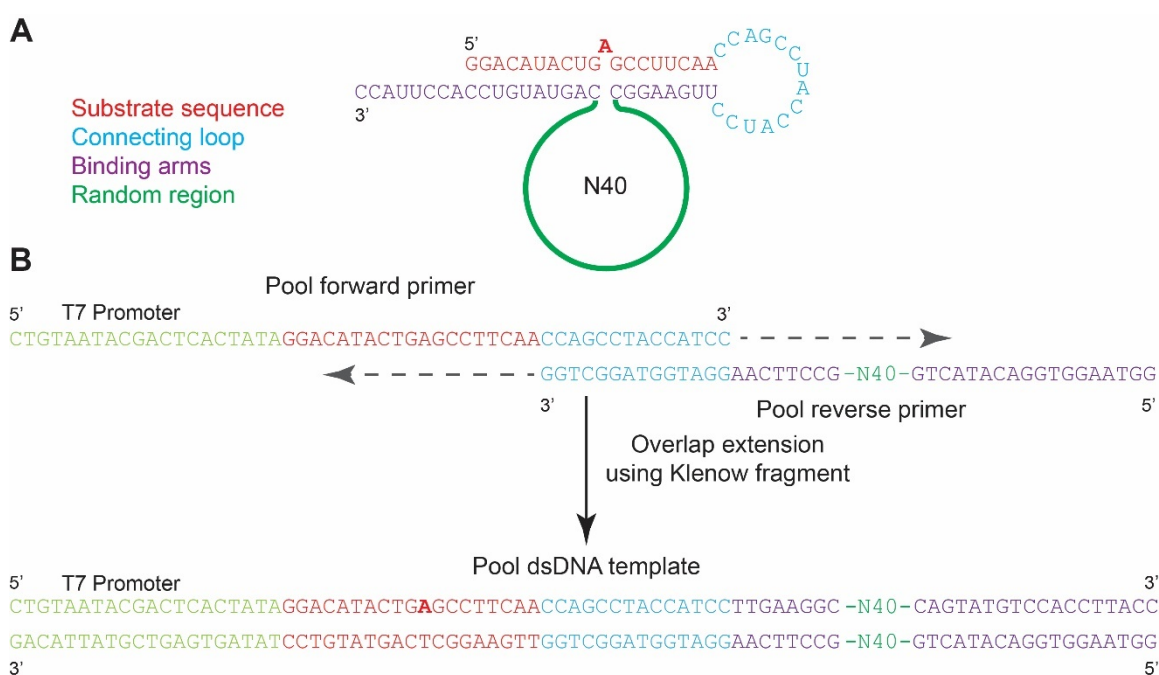


Figure 3-1 Pool design and construction. **(A)** Schematic representation of the selection pool. Red shows hypothetical substrate sequence. Predetermined modification site shown in bold font. Cyan signifies the connecting loop and purple indicates the binding arms. The green line represents the 40 nucleotide (N40) random region. **(B)** Construction of the transcription template of the selection pool. The complementarity of the connecting loop sequence at 3'-end of each primer made possible the assembly of the two pieces via overlap extension using Klenow fragment.

3.1.3 Overview of the selection process

The selection involved four major steps: incubation, capture, and amplification. The main features of each step are explained in more detail in the section below. The schematic representation of the selection process is depicted in (**Figure 3-2**)

During the incubation step, the RNA pool was dissolved in the selection reaction along with the biotinylated selection substrate. At this stage, the active pool variants react with the selection substrate. The reaction leads to “self-biotinylation” of the active pool species. All

the incubation rounds were performed at 37°C. For the first few rounds of selection, the incubation was done overnight to provide enough time for all the active species in the pool to be tagged. After a certain enrichment level was reached, the selection was made more stringent by gradually reducing the incubation time. This is to ensure only the fastest and most efficient variants survive the selection.

The selection buffer contained 120 mM KCl and 5 mM NaCl. The monovalent K⁺ and Na⁺ are known to neutralize the negative charge of the RNA phosphate backbone, leading to secondary or tertiary structure stabilization (Woodson, 2005). They are commonly added to functional nucleic acid selection buffers to promote efficient folding of the active pool species.

Bivalent ions play more critical roles in the activity of the functional nucleic acids, particularly ribozymes (Woodson, 2005). Apart from structural stability, these metal-ions have also been known to be directly involved in catalysis. Different types of bivalent metal ions are included in ribozyme selection reactions. The isolated variants often heavily rely on these metal ions for their activity. Examples of these metal ions are Mg²⁺, Ca²⁺, Zn²⁺, Mn²⁺, and Ba²⁺. Among the listed bivalent ions, Mg²⁺ is the most abundant in living systems (Maguire and Cowan, 2002). Natural ribozymes rely on this metal ion for catalysis. In many cases of such ribozymes, Mg²⁺ is directly involved in the catalytic process (Bowman et al., 2012). Since our final goal was to develop ribozymes with potential for cellular applications, we decided to include only the Mg²⁺ as the bivalent metal-ion cofactor.

The active pool species that managed to get “self-biotinylated” during the incubation step were separated from the rest of the pool at the capture step. To capture the biotinylated pool members, we employed streptavidin or neutravidin coated magnetic beads. Neutravidin is another biotin-binding protein with similar properties to streptavidin (Hiller et al., 1990). To avoid selection of streptavidin or neutravidin binders, the affinity matrix was alternated between neutravidin and streptavidin-coated beads every two rounds. To avoid non-specific binding of pool species, the beads were treated with *E. coli* tRNA before every capture step. This was to ensure that all the potential sites for nonspecific interactions would be blocked by tRNA.

After binding of the biotinylated pool species, the beads were washed extensively using the denaturing wash buffer. The wash buffer was supplemented with 8 M urea as the denaturant. At this stage, species of the pool with binding affinity for streptavidin/neutravidin were denatured and released. Biotinylated variants, on the other hand, remained tightly bound.

After the bulk of the inactive species were removed through the washing step, the active species were eluted. Extreme denaturing conditions were therefore applied to disrupt biotin-streptavidin/neutravidin interactions. For this purpose, a combination of heat (95°C) and an elution buffer containing 95% formamide was used.

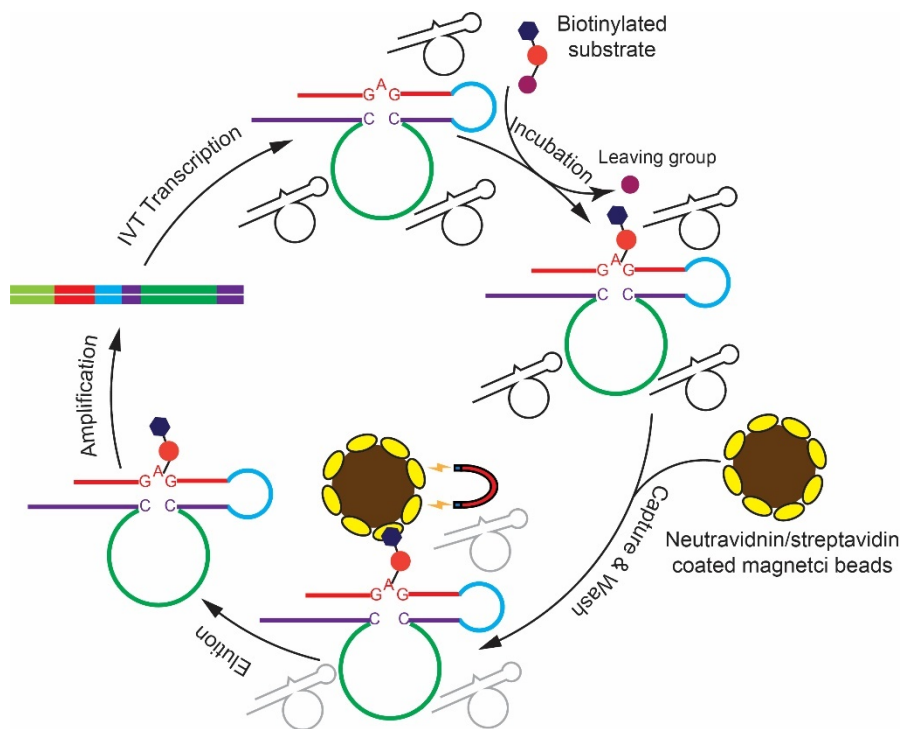


Figure 3-2 Overview of the selection process. During the incubation step the active species react with the biotinylated substrate. The “self-biotinylated” reactive species are immobilized onto streptavidin/neutravidin magnetic beads. The inactive species are washed extensively, and the active species are eluted from the beads. The eluted RNA is then subjected to reverse transcription and PCR, in the amplification step. The double-stranded DNA produced is then transcribed to generate the enriched pool of the next selection round.

The recovered RNA from elution was ethanol-precipitated and subjected to a two-step amplification protocol. The first step involved a one-pot RT-PCR reaction during which the recovered RNA was reverse transcribed to cDNA, followed by 10 cycles of PCR. This number of cycles was chosen to ensure complete second strand synthesis and limited amplification to avoid introducing amplification bias. The number of cycles was maintained for the first few rounds of selection where the amount of eluted RNA was vanishingly small. As the selection progressed and the amount of eluted RNA was increased, the number of cycles was also reduced from 10 to 5 and eventually to 2 cycles. A portion of the one-pot RT-PCR product was then used in a second PCR reaction and the rest of the sample was stored as the backup of the selection round. The second PCR amplification generates the

full-length template, from which RNA entering the subsequent selection round is transcribed.

Two separate forward primers were designed to ensure the amplification of ribozymes that modify the substrate sequence, while ribozymes that modify internal nucleotides in the core would not be amplified efficiently. The forward primer used in the RT-PCR step covers the connecting loop while only partially covering the substrate sequence (Starting at 2 nucleotides 3' to the predetermined modification site). In the second PCR reaction the forward pool primer was used to restore the substrate sequence to its full-length. This primer also carried the T7 promoter sequence at its 5'-end. The reverse primer was the same in both amplification steps. The priming scheme during the amplification process is depicted in (Figure 3-3)

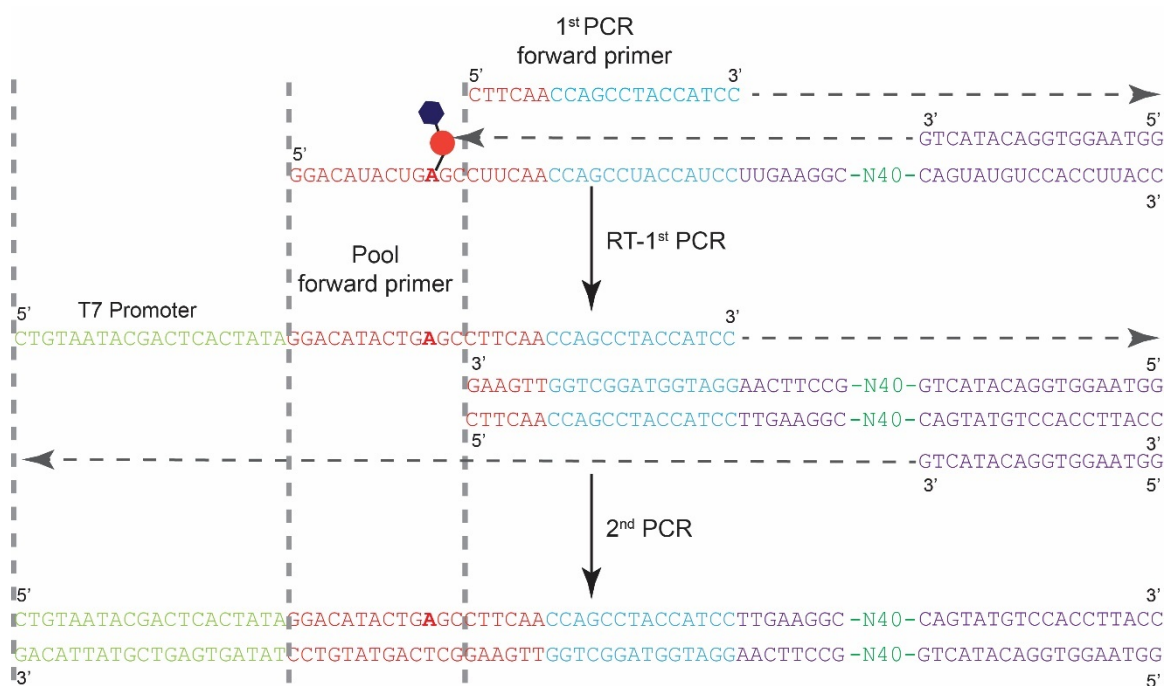


Figure 3-3 Priming scheme of the amplification process. Alternative forward primers were included in the first PCR reaction. This was to provide equal amplification chance for the incomplete cDNA of the modified species and the full-length cDNA of the carry-over RNA. A 2nd amplification cycle restores the cDNA to its full-length and adds the T7 promoter.

To monitor the progress of the selection procedure, a fraction of the pool before each round, was labeled at 3'-end using lucifer yellow. By measuring the fluorescence intensity of the eluted sample and comparing it to the total intensity of the selection sample, we estimated the percentage of active species of the pool, at the end of each capture step.

3.2 Selection of the nucleotidyltransferase ribozymes

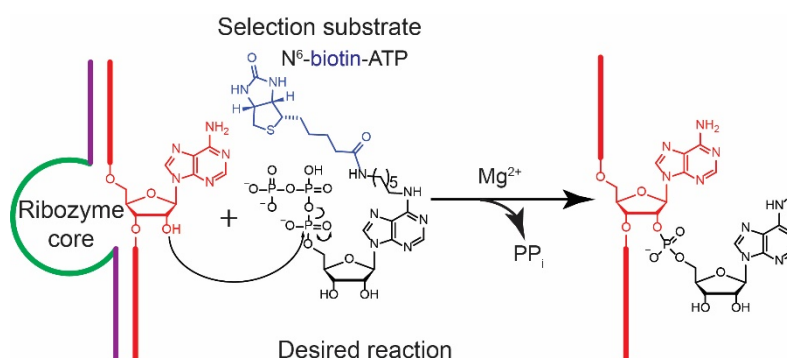


Figure 3-4 Selection substrate and the desired reaction for adenylyl transferase ribozymes. N⁶-biotin-ATP was used as the selection substrate in this selection. The desired reaction is the formation of a 2'-5' phosphodiester linkage between the 2'-OH of the predetermined modification site and the AMP moiety of the selection substrate.

3.2.1 Selection process

The selection started with $\sim 2.7 \times 10^{14}$ unique variants with an average copy number of 7. The incubation step of each selection round was performed in the selection buffer including 200 μM concentration of the N⁶-aminohexyl-ATP-biotin (N⁶-biotin-ATP) (**Figure 3-4**) and 40 mM MgCl₂. The first sign of enrichment was observed at round 4 when about $\sim 3.5\%$ of the pool RNA was eluted from the streptavidin beads. This amount was almost doubled in the next round and reached around 8%. This enrichment level did not change substantially at round 6 and 7, presumably due to saturation of the beads. Up to this point, the concentration of the RNA pool in the reaction was maintained at $\sim 50 \mu\text{M}$ (Including 150-300 pmol of the LY-labeled pool RNA). At round 8, the total amount of RNA was reduced by half, while the amount of the labeled fraction remained the same. The amount of eluted RNA at this round reached approximately 16%. Therefore, we decided to increase the selection stringency by reducing the incubation with the biotinylated substrate in subsequent rounds. This was done to provide survival advantage for the faster reacting, more efficient species in the pool. The duration of incubation was overnight until round 8, while at round 9, it was reduced to 3.5 hours. This led to an initial dramatic reduction in the amount of eluted RNA presumably due to the slower reacting variants being eliminated. The incubation time was further reduced to 1.5 hours, throughout the selection rounds 9 to 13. Despite the reduction in the incubation time, the amount of eluted fraction rose steadily throughout these rounds. This was an indication that faster-reacting ribozymes being enriched. At round 12 and 13, the amount of RNA was further reduced to only 150-300 pmol of the LY-labeled fraction. After 13 rounds of selection, the pool was cloned, and selected representatives were

sent for sequencing. The summary of the selection progress can be seen in (Figure 3-5). Details of the selection rounds are listed in (Table 3-1)

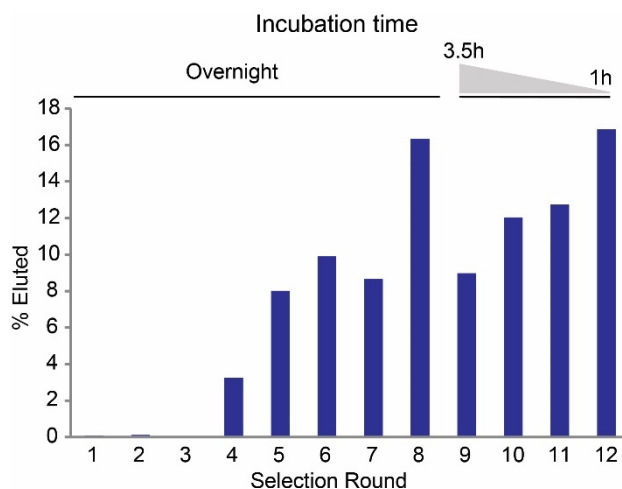


Figure 3-5 Selection progress of the adenylyltransferase ribozyme. The percentage of the eluted RNA after each round was estimated using fluorescence measurement. The selection stringency was increased by reducing the incubation time after a certain enrichment level was achieved.

Table 3-1 Details of the adenylyltransferase ribozyme selection rounds.

Selection round	1	2	3	4	5	6	7	8	9	10	11	12	13	
Pool concentration (μM)	~ 50						~25			~12.5				
Incubation time (h)	Overnight								3.5		2		1	
Eluted %	-	-	-	3.5	8	10	8.7	16.4	9	12	12.7	17	-	

3.2.2 Activity assay on the enriched pool

3.2.2.1 Cis-activity assay

Estimation of the enrichment level using fluorescence emission of the eluted and initial fractions was prone to underestimation due to several reasons. Factors such as saturation of the streptavidin/neutralavidin beads, loss of beads during washing steps, and competition between the LY labeled and unlabeled fraction of the pool may lead to underestimation of the active portion. Therefore, we utilized streptavidin gel-shift to estimate the Cis-activity level directly and more accurately (Figure 3-6 A). To do so, the 3'-lucifer yellow labeled round 12 pool, was incubated with the biotinylated substrate under selection conditions. Time-point samples were taken at 0, 1, 2, and 6 hours. After ethanol precipitation, the pellet from the time point samples was incubated with a large excess of streptavidin, followed by resolution on 10% native PAGE. Fluorescent imaging of the gel revealed that over 75 % of

the pool species in the 6-hour time-point were shifted. The majority of the pool species were, therefore, determined to be active and efficient enough to self-biotinylate within 6 hours (**Figure 3-6 B**). The presence of more than one shifted band is due to the oligomerization of streptavidin which can be observed on native gels (Humbert et al., 2005).

3.2.2.2 Trans-activity assay of the pool

Up to this point, it was established that the evolved variants can self-biotinylate efficiently under the selection conditions and the presence of the N⁶-Biotin-ATP. It was not however clear whether the ribozymes modify the substrate sequence and if so, whether they can modify it in trans. (**Figure 3-6 A**). To answer these questions, the RNA pool from round 13 was reverse transcribed. The resulting cDNA was PCR amplified using a new forward primer. The primer was designed to remove the substrate sequence and add a T7 promoter along with a 5'-GG start site to the beginning of the connecting loop (**Figure 3-6 C**). The PCR product was ethanol precipitated and *in vitro* transcribed. Thus, the final product was a ribozyme pool, which lacks the substrate sequence segment. This transcript was mixed with 3'-lucifer yellow labeled selection substrate at an approximate ratio of 1:2 and incubated in 10 μ L of the selection buffer. The reaction mixture included 40 mM MgCl₂ and 200 μ M N⁶-Biotin-ATP. The excess of the substrate sequence was to ensure that all pool species were annealed to it. Timepoint samples were taken at 0, 30 minutes, 2 hours, and after overnight incubation. In this case due to the small size of the substrate sequence, the branching reaction creates a significant enough mass shift that can be resolved using gel electrophoresis. The samples were resolved on a 20% denaturing PAGE and subjected to fluorescent imaging. The results of this experiment clearly showed that the modification site has successfully been localized to the substrate sequence. The results also confirmed that the enriched pool from round 13 was trans-active (**Figure 3-6 D**).

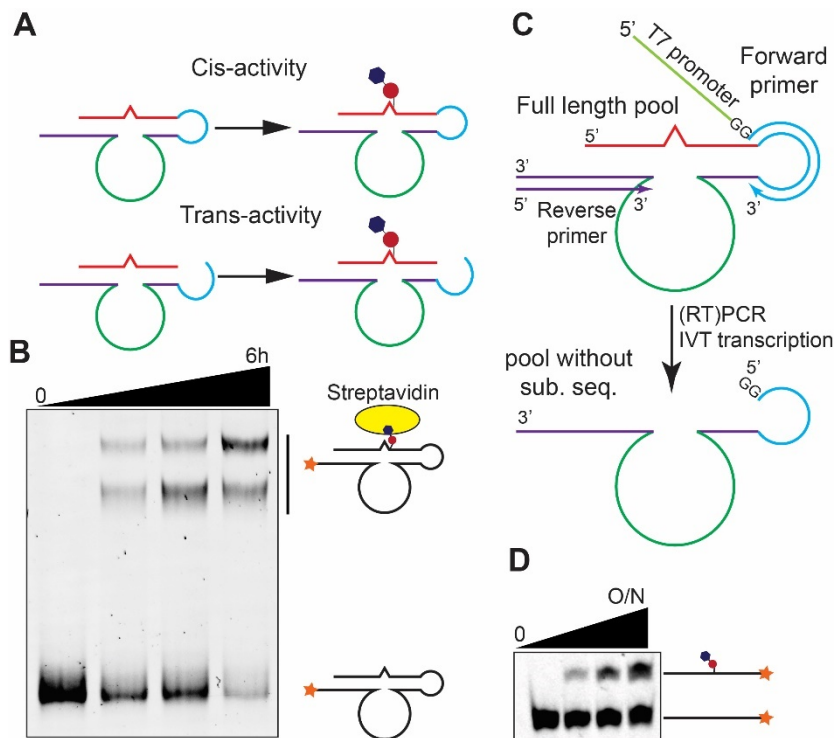


Figure 3-6 Cis- and Trans-activity of the enriched pool. **(A)** Cis- activity, defined as ribozyme modifying a part of itself, whereas trans-activity means the ribozyme modifies a target RNA as a separate entity. **(B)** Cis-activity assay of the enriched pool. Streptavidin gel-shift assay of the round 12 pool. Reaction conditions: 15 μ M Round 12 RNA Pool labeled with LY-CH in selection buffer including 200 μ M biotinylated ATP and 40 mM $MgCl_2$. Timepoints taken at 0, 1, 2, and 6 hours, and subjected to streptavidin gel-shift on 10% PAGE. **(C)** Generating the trans-active enriched pool from round 13. **(D)** Trans-activity assay of the enriched pool. Reaction conditions: LY- labeled substrate sequence (24 μ M) and trans-active round 13 pool (12 μ M) in 10 μ L of the selection buffer containing 200 μ M of the biotinylated ATP and 40 mM of $MgCl_2$. Timepoints taken at 0, 0.5, 2 hours and after overnight incubation, were resolved on a 20 % denaturing PAGE.

3.2.3 Cloning of the enriched pool and activity assay of the individual clones

Since the majority of the pool was proven active in cis and trans, the pool was cloned using TOPO-TA system and was submitted to Sanger sequencing. The transformed bacteria were plated and subjected to blue-white screening. A stop-codon cassette was included in the forward primer to decrease the number of false-negative colonies. To confirm that the clones were harboring the pool insert, we performed colony PCR on 40 randomly picked colonies (35 white, 5 blue). All tested colonies, including the blue ones, proved positive. To test whether the obtained clones contained active sequences, transcripts were generated from the colony PCR samples. Briefly, a portion from each colony PCR product was amplified in a larger PCR reaction. The reaction was then precipitated, and the resulting dsDNA was transcribed *in vitro*. Streptavidin gel-shift assay was performed to test the individual variants

for activity. The results of this experiment confirmed that all colonies were active. However, they were not equally efficient (**Figure 3-7**).

3.2.4 Sanger sequencing results

After the activity assay, the plasmids extracted from 22 of the colonies were submitted for Sanger sequencing. Twenty one out of 22 clones were successfully sequenced. Analysis of these sequencing results revealed three major sequence groups denoted as FH14, FH20, and FH31. The sequence groups identified through Sanger sequencing are listed in (**Table 3-2**). The sequences belonging to FH14 and FH20 groups shared remarkable similarity (**Figure 3-8 A**) and both groups proved more efficient in streptavidin gel-shift assay. The FH31 group shared little similarity to the other two groups except for a 5'-GAYA-3' motif shared across all variants.

The FH31 sequence group demonstrated the lowest efficiency of the three groups, despite having the highest frequency (12 out of 21 sequences) (**Figure 3-7**). A closer inspection of the FH31 group revealed the presence of a point mutation in the left binding arm. This mutation may have been the reason behind the lower efficiency of the FH31 sequence group. The left binding arm was the only part of the constant region not being included in any primer. Thus, this arm of the pool could mutate during the PCR reactions.

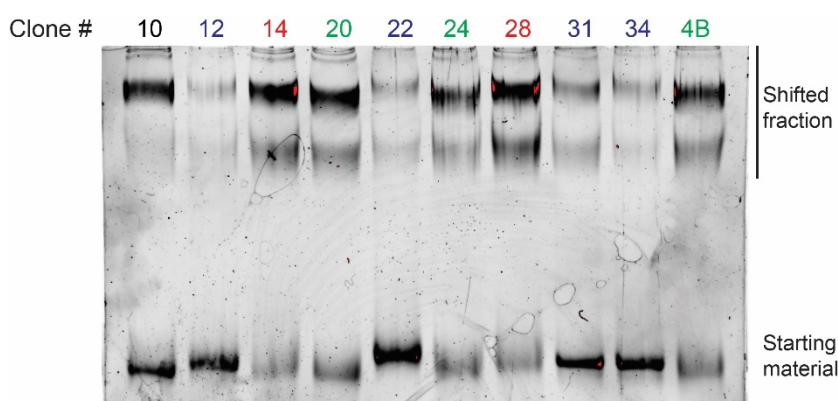


Figure 3-7 Streptavidin gel-shift assay of ten exemplary, individual clones indicate varying levels of efficiency. Clone numbers 14 and 28 (highlighted in red) belong to the sequence group FH14 and the clones 20, 24, and 4B (highlighted in green) belong to the sequence group FH20. Clones 12, 22, 31, and 34 (highlighted in blue) which belong to the sequence group FH31 show drastically lower efficiency compared to the other two groups.

Table 3-2 Sequence groups identified from Sanger sequencing of the isolated clones

Group	5'- Sequence-3'	Times repeated
FH14	CACGCTGAGTAGACATACTTGCAAACGCTGCAAACATAGA*	3
	CACGCTGAGTAGACATACTTGCAAAACGCTGCAAACATAGA	1
FH20	CTCGCTGCGTAGACATACCTGGCTCCCTCCAGCCATAGA*	5
FH31	ACCCCATCACTGATAGTATAGAAAAAACAGGGCATGTTA*	2
	ACCCCATCACTGATAGTATAGAAAAAACAGGGCATGTTA	7
	ACCCCATCACTGATAGTATAGAAAAAACAGGGCATGTTA	2
	ACCCCATCACTGATAGTATAGAAAAAACAGGGCATGTTA	1

* Denotes: Sequences subjected to kinetic experiments

3.2.5 Kinetic comparison of the evolved species

DNA templates for FH14, FH20 and FH31 were ordered in which the loop and the 3'-extension was removed, and a 5'-GG was added to the right arm to ensure transcribability. The mutation in the FH31 binding arm was also restored in these templates. Single turnover kinetic experiments were then performed in triplicates using 3'-fluorescein-labeled selection substrate sequence (Parent sequence). The tests were performed to compare the reaction rate and efficiency of the FH variants. The ratio between the substrate sequence and the ribozyme was adjusted at 1:10. The reaction conditions were identical to the selection process. The pseudo-first order kinetic curves derived from these experiments determined k_{obs} values of $\sim 1.8 \times 10^{-2}$, $\sim 0.14 \times 10^{-2}$, and $\sim 0.78 \times 10^{-2} \text{ min}^{-1}$ for FH14, FH20 and FH31, respectively (**Figure 3-8 B & C**) (**Table 3-3**). These values indicate that FH14 is the fastest reacting and the most efficient variant followed by FH31. FH20 however, falls far behind. Ribozymes FH14 and FH31 were therefore characterized, while FH14 was studied more thoroughly.

Despite remarkable sequence similarity between FH14 and FH20, and high activity of FH20 in cis-, the results of kinetic experiments in trans were quite unexpected. On the other hand, FH31, which performed poorly in the streptavidin gel shift assay, demonstrated significantly higher efficiency and reaction yield compared to FH20. The discrepancy between the cis- and trans-activity test results of the FH31 could be attributed to the presence of the mutation in the right binding arm, which was restored for the trans-activity kinetic experiments.

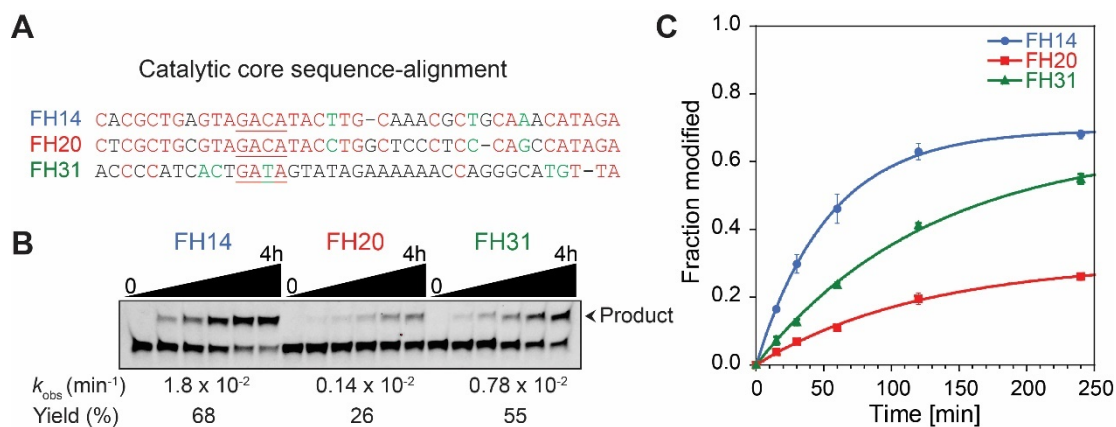


Figure 3-8 Sequence alignment and kinetic comparison of the evolved variants. **(A)** Partial alignment of the FH14, FH20, and FH31 catalytic core sequences. Red: identity, Green: similarity, Black: difference. GAYA motif shared among variants is underlined. **(B)** Fluorescent gel image of the single-turnover kinetic assay performed on the evolved variants. Reaction conditions: 3'-fluorescein parent substrate sequence (1 μ M), ribozyme transcript (10 μ M), 1x selection buffer, including 40 mM MgCl₂ and 200 μ M N⁶-biotin-ATP. Incubation at 37°C. time-points taken at 0, 15, 30, 60, 120 and 240 minutes, resolved on a 20% analytical denaturing PAGE. **(C)** The bands were quantified and plotted using Kaleidagraph software. Curve-fitting was performed using pseudo-first order kinetic function ($Y=Y_{max}(1-e^{-k_{obs}t})$).

3.2.6 Secondary structure prediction of the evolved variants

Possible secondary structures of the evolved variants were predicted using the Vienna RNAfold server (Hofacker, 2003). FH14 and FH20 showed similar predicted secondary structures, while FH31 was largely unstructured. The predicted secondary structures and the base-pair probability dot plots of the three variants can be seen in **(Figure 3-9)**

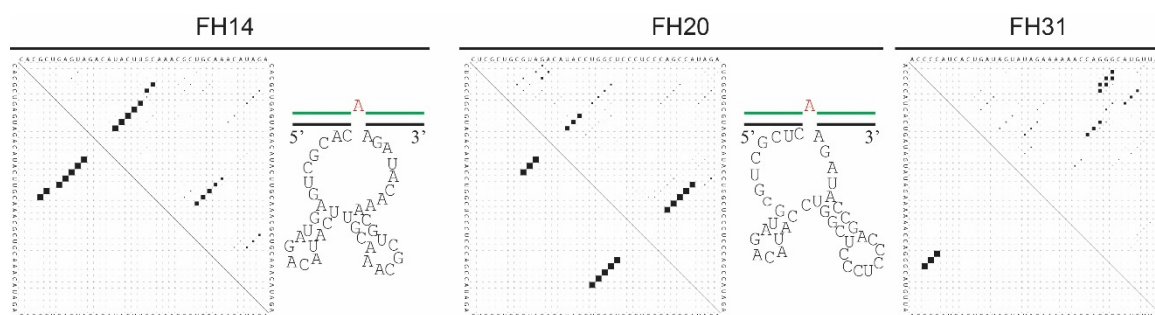


Figure 3-9 Secondary structure prediction of the three variants. Structures and the dot plot of the base pair probabilities generated using the Vienna RNAfold server.

3.2.7 Characterization of the reaction product

3.2.7.1 Probing of the modification-site

The trans-activity assays confirmed the modification site to be located in the substrate sequence. The exact modified position, however, could not be determined from those experiments.

To pinpoint the modification site, a synthetic 21-nucleotide analog of the substrate sequence (R660) was synthesized with a hexynyl modification at the 5'-end. This sequence was labeled using 5-FAM-azide and CuAAC reaction resulting in 5'-5-FAM-R660. The labeled oligonucleotide was then modified on a preparative scale, with an N⁶-biotinylated-adenylyl branch using the ribozymes FH14 and FH31. The reaction mixture was resolved after overnight incubation on a 20% denaturing PAGE and subjected to fluorescence imaging. The conversion efficiency based on fluorescence imaging was estimated to be around 75 to 80% in each case. The modified band was then PAGE purified with an isolated yield of roughly 50% (**Figure 3-10 A**).

RNase T1 digestion and alkaline hydrolysis was performed on the ribozyme modified and unmodified 5'-FAM-R660:

In case of the T1 lanes, the bands resulting from cleavage after G1 and G9 lack mobility shift in the modified lanes relative to the unmodified lane (**Figure 3-10 A & B**). The bands pertaining to the cleavage after G11 and the full-length oligonucleotide in the lanes with the modified sample however are shifted compared to their counterparts in the unmodified lane (**Figure 3-10 B**).

In the alkaline hydrolysis lanes of the modified sample, a clear shift is visible in case of the bands corresponding to cleavage after A10. This shift is lost in the bands resulting from cleavage before A10. A further observation, which can be made from this experiment, is that the band corresponding to the cleavage immediately 3'- to A10 is missing from the lanes containing the modified sample (**Figure 3-10 B**). Since the cleavage reaction in alkaline hydrolysis relies on deprotonation of the 2'-OH, lack of the A10 cleavage band indicated that the 2'-position of this nucleotide is blocked.

These results confirm that our selection process has guided the modification to the bulged A nucleotide without the need for exerting additional selection pressure.

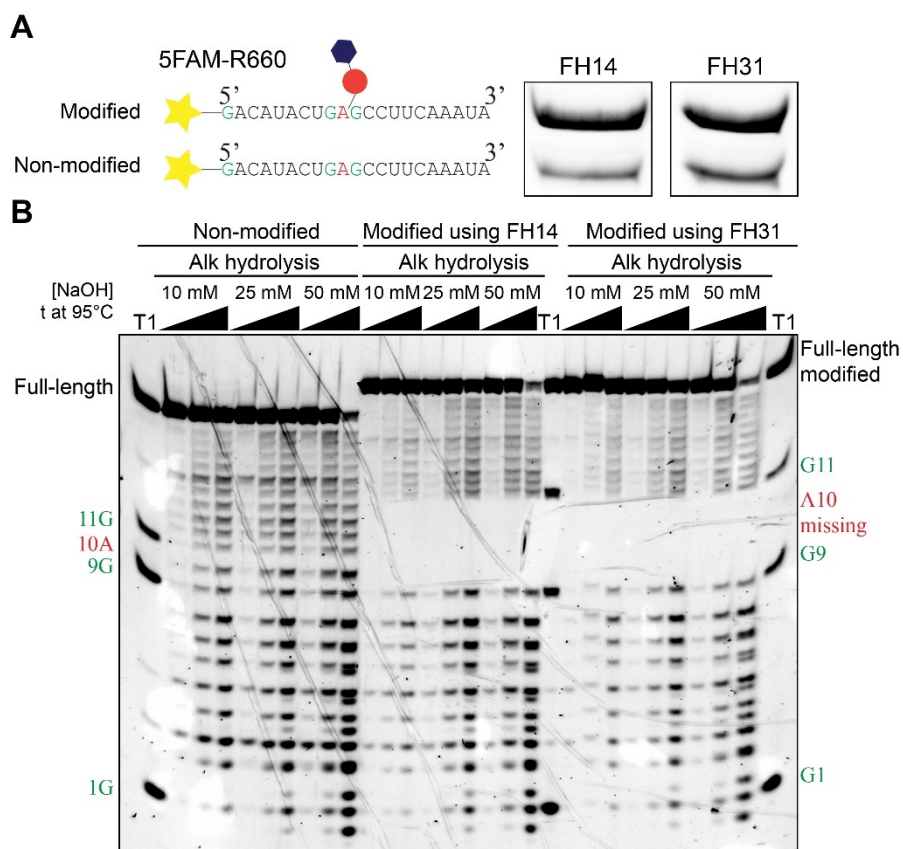


Figure 3-10 Probing of the modification site. **(A)** Preparative scale modification of the 3'-5FAM-R660. (5FAM-R660 (1 nmol), ribozyme (1.3 nmol) mixed in 25 μ L of the selection buffer including 40 mM $MgCl_2$ and 400 μ M N^6 -biotin-ATP. Incubated overnight at 37°C and isolated on a 20% analytical denaturing PAGE. **(B)** Gel image of the alkaline hydrolysis and RNase T1 probing of the modification site. The RNase T1 digestion (T1) performed using 5FAM-R660 modified or unmodified (1 μ M), 50 mM Tris buffer pH= 7.5, and 0.2 U/ μ L of RNase T1, 37°C for 30 seconds. The alkaline hydrolysis was performed using 5-FAM-R660 modified or unmodified (1 μ M) in 10 μ L of 10, 25, or 50 mM NaOH. Samples were incubated at 95°C. Timepoint samples were taken at 1, 2.5, and 5 minutes. Based on the results of this experiment, both FH14 and FH31 appear to modify the predetermined A.

3.2.7.2 ESI-Mass spectrometry of the reaction product

The desired activity in the FH selection was 2'-5' adenylyl transferase reaction that leads to the addition of a single adenylyl branch to the modification site. This branch was expected to be connected via a 2'-5' phosphodiester bond. However, ribozymes forming other types of linkages might also arise from our selection. It was therefore important to confirm the mass of the product to gain further proof of its identity. A short synthetic analog of the original substrate sequence was used for this purpose. The synthetic oligonucleotide was shorter than the original substrate sequence by two nucleotides from the 5'-end. The removed nucleotides were the 3'-GG that was necessary for in vitro transcription. The synthetic oligonucleotide

also contained a 3'-aminoethyl modification, which added the potential for further derivatization. This oligonucleotide was called R594.

R594 was modified, using FH14 ribozyme, in a semi-preparative manner under the selection conditions including 400 μM N⁶-aminoethyl-ATP and 40 mM MgCl₂. PAGE purification of the modified shifted band resulted in approximately 50% isolated product yield. The product was submitted to ESI-mass spectrometry. The expected mass was calculated to be 5959.9909 Da, based on the monoisotopic mass of the N⁶-aminoethyl-adenylylated R594. The observed mass was found to be 5959.07598 Da, which confirms the formation of the desired product (**Figure 3-11**).

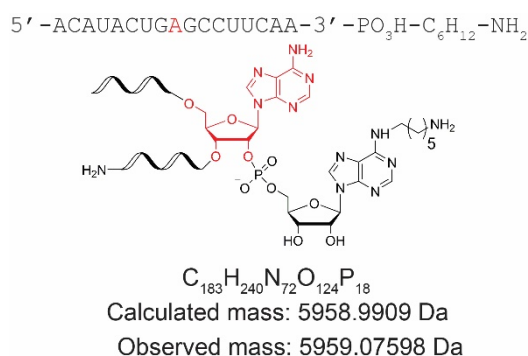


Figure 3-11 ESI-mass analysis of the FH14 modification product. Sequence of the R594 and the modification product of FH14. Calculated mass vs observed mass confirms the identity of the product. (Full spectrum shown in (**Figure 8-1**) of the appendix).

3.2.8 Characterization of the FH14 and FH31 ribozymes

3.2.8.1 Target sequence scope of the evolved variants

3.2.8.1.1 Substrate sequence specificity of FH14 and FH31

After it was discovered that the modification has been successfully directed to the desired position, the sequence generalizability of the evolved variants was tested. For this purpose, a set of target sequences were designed in which particular sets of mutations were made to the parent substrate sequence. In our initial tests, we decided to leave the GAG modification context intact while heavily mutating the other positions. Since the 5'-GG dinucleotide was required for efficient *in vitro* transcription, this dinucleotide was also left unchanged. All the other positions were either transition mutated (TM substrate sequence) or subjected to two types of transversion mutations. In case of the TV1 substrate sequence, all the positions were mutated into their complementary nucleotides. For the TV2 substrate sequence, the transversions were made to the non-complementary counterparts. FH14 and FH31 ribozymes were also designed with their binding arms co-varied corresponding to each of

the designed substrate sequences (**Figure 3-12 A**). The target sequences were transcribed *in vitro* and 3'-fluorescein-labeled using periodate oxidation method. Single-turnover kinetic experiments were performed using these substrate sequences and their corresponding ribozymes in the selection conditions. The results revealed a high degree of flexibility regarding the positions outside the GAG context for both FH14 and FH31 ribozymes (**Figure 3-12 B-E**). A rather surprising observation, in the case of the FH14 ribozyme is that all the mutated substrate sequence variants show remarkably higher modification rate and efficiency compared to the parent sequence (**Figure 3-12 D**). All three sequences demonstrated a yield of over 75% within 1 hour. The k_{obs} values for the mutant substrate sequence/ribozyme pairs however could not be derived from these experiments for the FH14 ribozyme, due to fast reaction kinetics. Kinetic experiments with more data points between 0-30 minutes are needed for to determine the k_{obs} values for these mutants. A stretch of 8-nucleotides within the catalytic core of the FH14 has an identical sequence to the parent substrate. A possible explanation for the lower efficiency of the parent sequence might, therefore, be due to the formation of misfolded FH14: parent sequence complex (**Figure 3-13**).

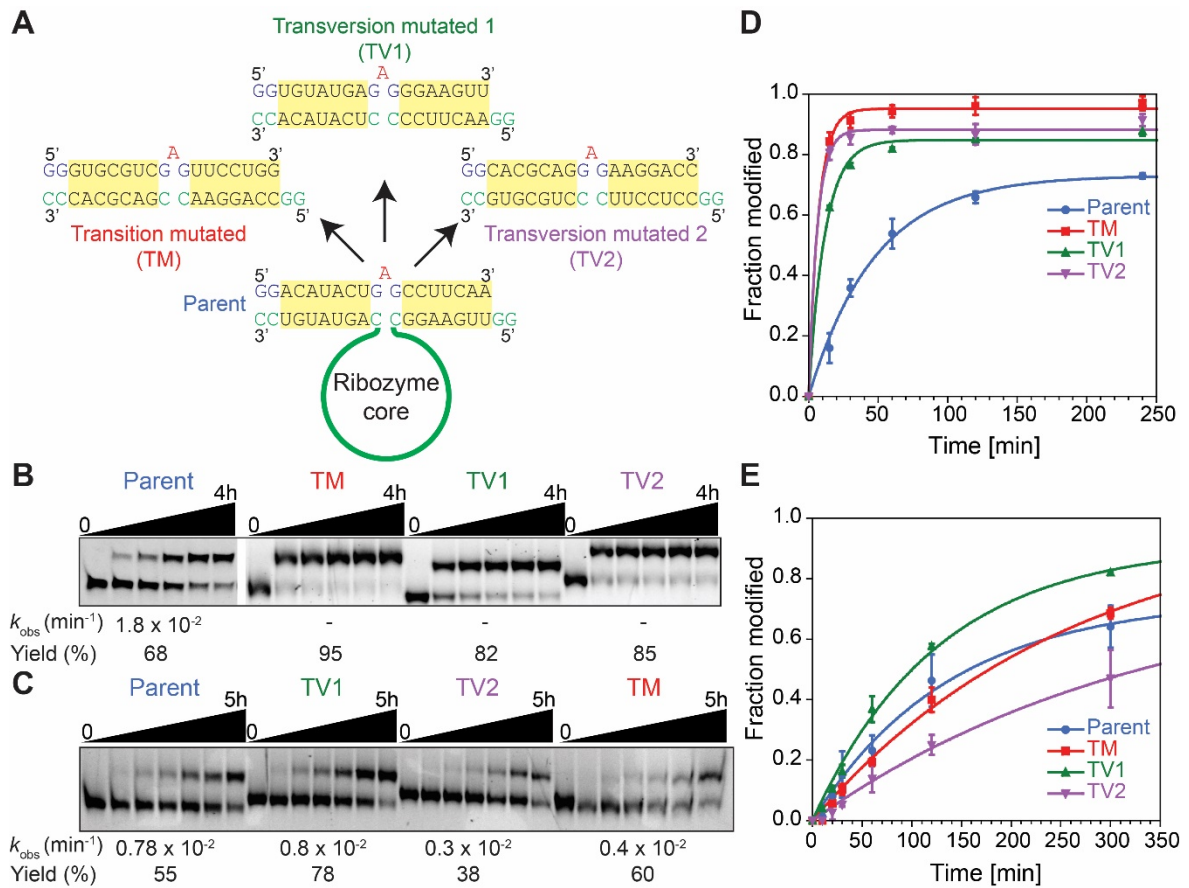


Figure 3-12 Target sequence variability of FH14 and FH31. **(A)** Positions outside the GAG context on the substrate sequence were subjected to three sets of mutations. Modification-site is signified in Red. Blue signifies positions unchanged from the parent sequence. Yellow boxes indicate co-varied positions on the substrate sequence and the ribozyme arm. Green shows positions unchanged in ribozyme arms. Gel image of the kinetic assays of the ribozyme: substrate sequence pairs for FH14 **(B)** and FH31 **(C)**. Reaction conditions: 3'-substrate fluorescein-substrate (1 μM) sequence and corresponding ribozyme (10 μM) in standard selection conditions. Time points: 0, 15, 30, 60, 120 and 240 minutes for FH14 and 0, 10, 20, 30, 60, 120 and 300 minutes for FH31. **(D)** And **(E)** Pseudo-first order kinetic curve fits of the FH14 and FH31 respectively.

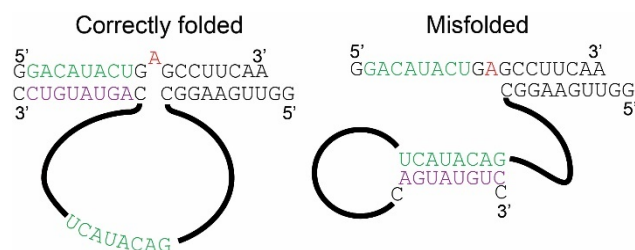


Figure 3-13 Folded and misfolded states of the FH14 ribozyme paired with the parent substrate sequence. An 8-nucleotide segment in the catalytic core is identical to a part of the substrate sequence (green). This segment might pair to the ribozyme's left-arm (purple) and prevent the formation of the active complex.

3.2.8.1.2 Modification site sequence-context requirements of FH14 and FH31

To determine the essentiality of the GAG modification context, substrate sequences were designed in which the two G-nucleotides flanking A were mutated. Mutations were made to one G-nucleotide at a time, to the three other possible substitutions (**Figure 3-14 A**). These transcripts were based on the TM substrate sequence since it proved to be the most efficiently modified by their corresponding FH14 ribozymes (**Figure 3-12 B & D**). Kinetic assays were performed using these substrate sequence variants and their corresponding ribozymes. The results proved that FH14 ribozyme tolerates all the mutations to either of the G-nucleotides. FH31 ribozyme, however, demonstrates a strong preference towards the G nucleotide located 3' to the modification site A, while the mutations to the G located at the 5' are better tolerated (**Figure 3-14 B**). Analysis of the combined effects of these mutations to the modification context sequence is yet to be performed.

An analog of the parent substrate sequence containing an A to G mutation (**Figure 3-14 C**) at the modification site was not modified by either of the ribozymes (**Figure 3-14 D**). The modification site A has therefore been determined as essential for the ribozyme activity. The modification site sequence context required for the FH14 ribozyme is therefore determined as 5'-NAN-3'. For the FH31 ribozyme it is evident that 5'-NAG-3' is required for the optimal labeling efficiency.

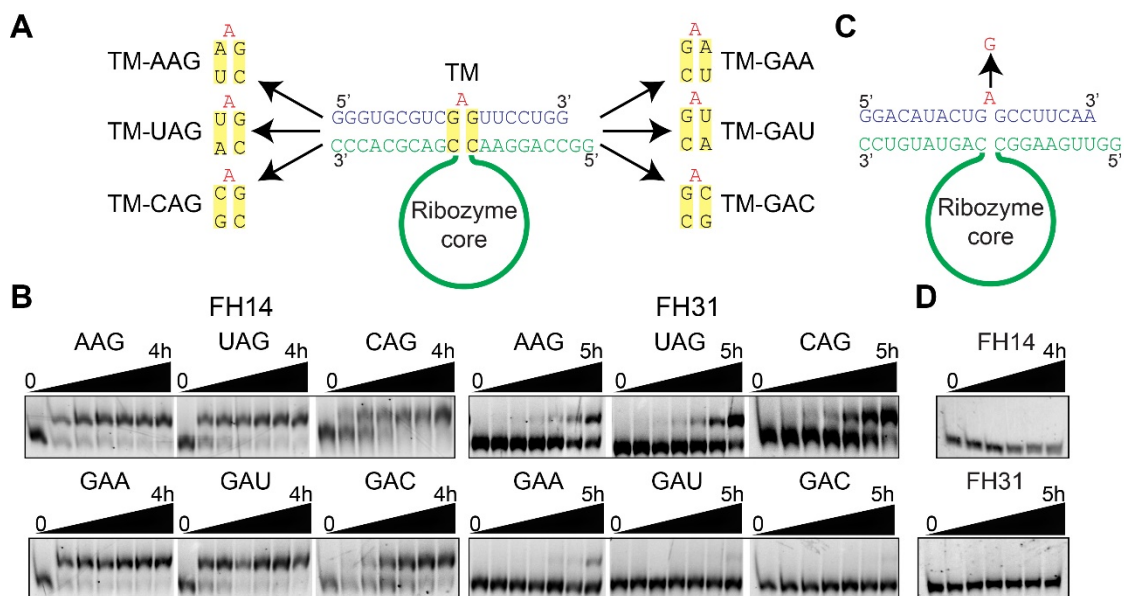


Figure 3-14 Modification-site sequence context requirement of the FH14 and FH31. **(A)** The TM substrate was used for this purpose with mutations in the 5'-GAG-3' context, ribozyme arms were co-varied to maintain base pairing. **(B)** Gel-image of the single-turnover kinetic experiments on FH14 and FH31 using TM-GAG mutant: ribozyme pairs. All mutations appeared to be well tolerated by FH14 ribozyme, whereas FH31 strongly prefers the presence of a G nucleotide 3' to the modification site A (red). **(C)** Mutation of the modification site A to G. **(D)** Fluorescent gel-image of the single-turnover kinetic experiments on FH14 and FH31 using the substrate with A to G mutation in the modification site.

3.2.8.2 NTP substrate scope of the evolved variants

3.2.8.2.1 NTP analog specificity range of FH14

The NTP analog used in this selection was ATP, biotinylated at N⁶ via an aminohexyl linker. Thus, it was important to determine whether biotin and/or the aminohexyl linker are essential substrate requirements of the ribozymes. Single turnover kinetic assays were performed using N⁶-aminohexyl-ATP, which is highly similar to the selection substrate except that it lacks the biotin (**Figure 3-15**). Unmodified ATP was also used to examine the necessity of the aminohexyl linker. The results of these tests determined that biotin is not an essential substrate sequence requirement of FH14. The N⁶-aminohexyl-ATP demonstrated similar ligation rates and overall efficiency as the biotinylated ATP (**Figure 3-16 A & C**) (**Table 3-3**). On the other hand, a dramatic drop, in reaction rate and efficiency, was observed when unmodified ATP was used as the modification substrate. These results indicate that the presence of the linker at the N⁶- position is preferred for efficient labeling (**Figure 3-16 A & C**) (**Table 3-3**).

The ligation efficiency of FH14 was also examined with ATP analogs in which biotin in the selection substrate was replaced by a fluorophore. For this purpose, N⁶-aminohexyl-ATP

conjugated to 5-FAM, ATTO550, or sulfoCy3 (**Figure 3-15**) were used as modification substrate in single turnover-kinetic experiments. All three fluorophore conjugates were efficiently ligated to the target RNA (**Figure 3-16 B & C**). The reaction yield for ATTO550 and 5FAM conjugated ATP was 75% in 5 hours. SulfoCy3 however, resulted in a yield of roughly 50% in the same duration of time (**Table 3-3**). The reduced efficiency of the sulfoCy3-conjugated ATP might be attributed to the presence of the negatively charged sulfonate groups. These groups may be electrostatically repelled by the similarly charged phosphate backbone of the ribozyme. Nevertheless, the results of these kinetic assays established the FH14 ribozyme as a potent tool for RNA labeling using fluorescent dyes or biotin.

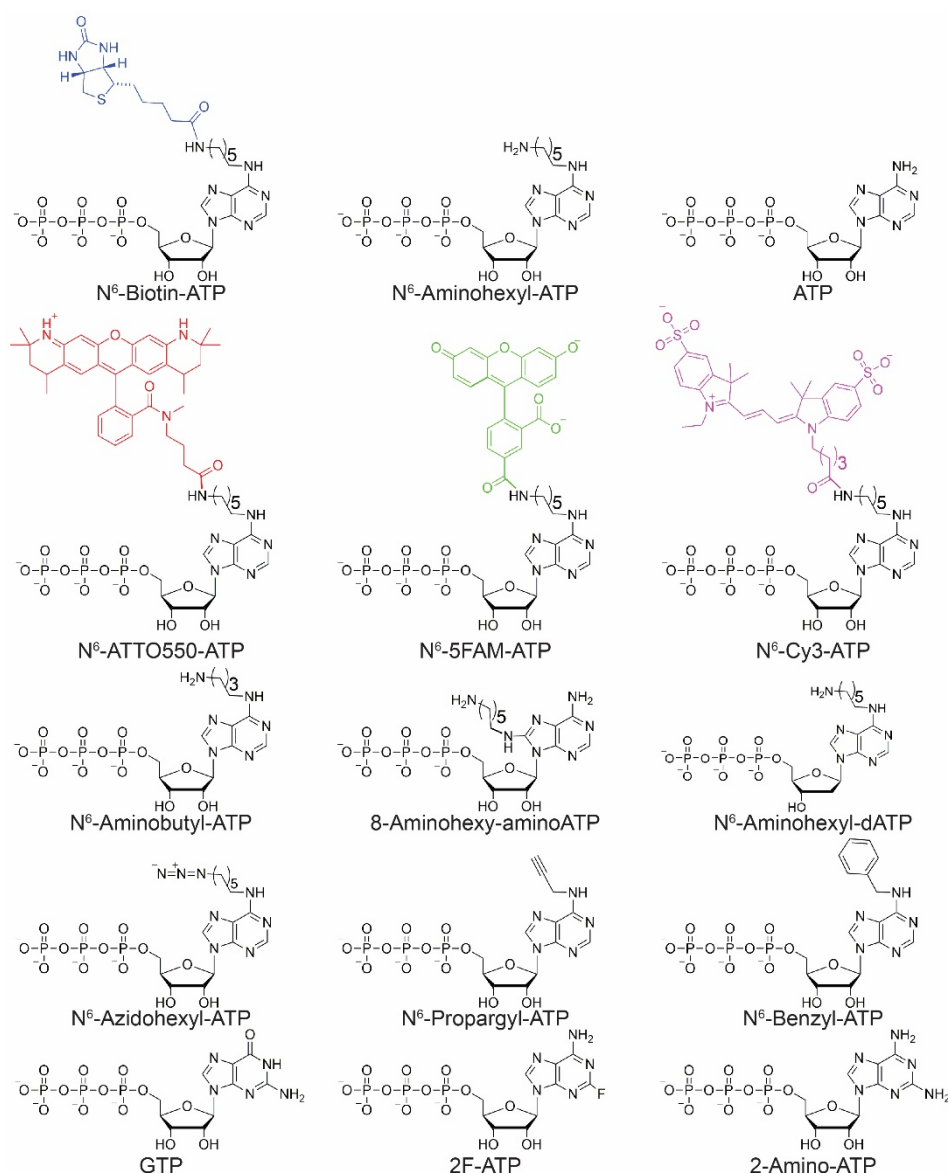


Figure 3-15 List of the ATP analogs used in this thesis project.

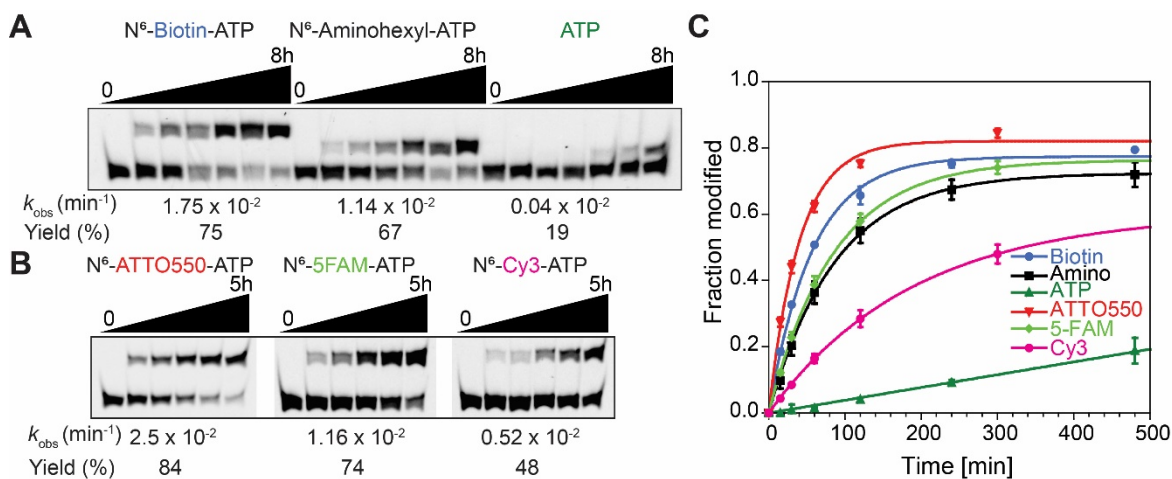


Figure 3-16 NTP analog specificity of FH14 ribozyme. **(A)** Kinetic gel of FH14 ribozyme tested using N^6 -biotin and N^6 -aminohexyl-ATP and unmodified ATP. Reaction conditions (5FAM-R660 (1 μM), FH14 (10 μM) in 1x selection buffer including 40 mM MgCl_2 and 200 μM of the ATP analog. 37°C, time points: 0, 30, 60, 120, 240, 360 and 480 minutes) **(B)** Kinetic gel of FH14 ribozyme tested with fluorophore-conjugated ATP. Reaction conditions: (Cy5-R660 (1 μM) and FH14 (10 μM) in 1x selection buffer including 40 mM MgCl_2 and 200 μM ATP analog.) Time points taken at 0, 15, 30, 60, 120 and 300 minutes. **(C)** Pseudo-first order kinetic curve fits of the quantified bands from **(A)** and **(B)**.

Necessity of the 2'-OH of the ATP analog was also tested using N^6 -aminohexyl-dATP as modification substrate. The effect of removal of the 2'-OH on reaction yield was negligible. In another experiment, 8-aminohexyl-aminoATP was used in which the position of the linker is moved from N^6 to C8. No reactivity was observed in case of this substrate. These results may imply that the C8 modified substrate may not be recognized by the ribozyme possibly due to the steric hindrance caused by modification at this position (**Figure 3-17 A**).

Further investigation was performed into the NTP substrate scope of FH14 using various N^6 -modified ATP analogs (**Figure 3-15**). For example, the effect of reducing the length of the aminohexyl linker was examined by using N^6 -aminobutyl-ATP as substrate. The substrate resulted in an 8-hour yield of ~48% compared to ~80% for the substrate for aminohexyl linker. Despite this drop in the reaction efficiency, the yield remained significantly higher than unmodified ATP (~23% after 8 hours of incubation). The results indicate a strong preference for N^6 -modified analogs (**Figure 3-17 B**).

Since it was established that FH14 strongly prefers N^6 -modified ATP substrates, other types of modifications at this position were also tested. The prospect of FH14 ribozyme was examined for indirect RNA labeling via means such as CuAAC or SPAAC. For this purpose, kinetic experiments were performed using N^6 -azidohexyl-ATP and N^6 -propargyl-ATP (**Figure 3-15**). Despite similarity between the N^6 -aminohexyl and N^6 -azidohexyl ATP

analog, replacement of the amino group with an azide functionality drastically reduced ribozyme reactivity. The 8-hour yield of this reaction was ~17%, which is lower than that of the unmodified ATP. The N⁶-propargyl analog also performed poorly with only ~12% conversion after 8-hours (**Figure 3-17 C**). The results from the azido and propargyl functionalized analogs, therefore, indicate that these substrates are not suitable for FH14 mediated indirect RNA labeling.

N⁶-benzyl-ATP was also tested as an analog with a bulky modification at N⁶. This substrate demonstrated surprisingly high reactivity towards FH14 ribozyme. Despite being radically different from the N⁶-aminohexyl analog, N⁶-benzyl-ATP resulted in 61% yield in 8-hours (**Figure 3-17 D**). The reason behind this high yield might be formation of stacking interactions with nucleotides positioned in the ribozyme's substrate-binding pocket.

No product was formed in kinetic experiments using GTP (**Figure 3-15**) as modification substrate, within the time course of 4 hours (**Figure 3-17 E**). 2-amino-ATP and 2F-ATP (**Figure 3-15**) however were accepted albeit weakly by the ribozyme (**Figure 3-17 D E**). The results indicate the significance of the 6-amino group of adenine for FH14 mediated recognition.

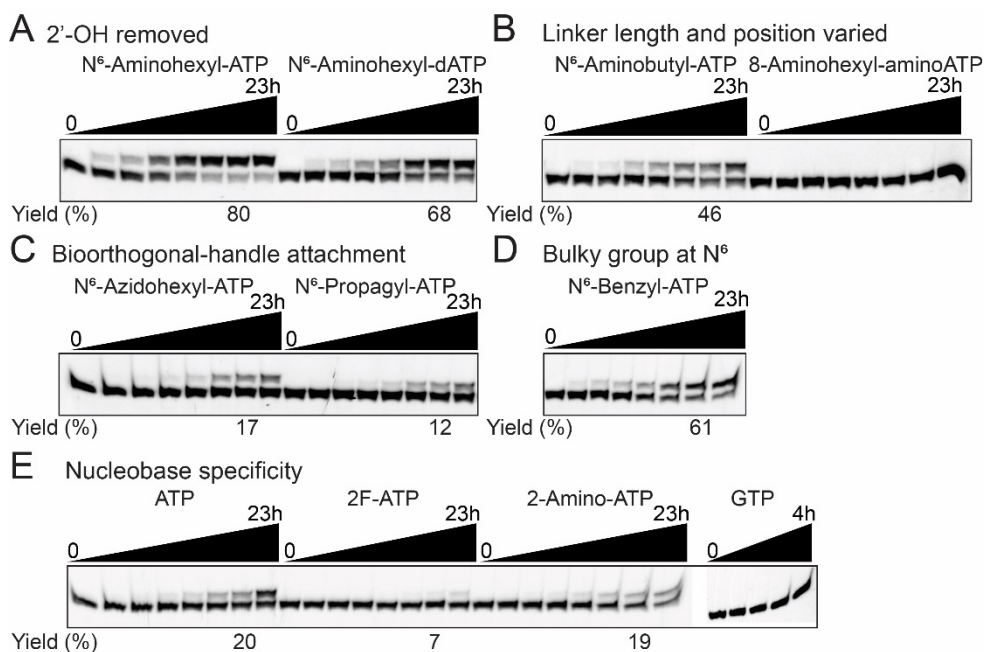


Figure 3-17 Further investigation of the NTP analog scope of FH14 ribozyme. Reported yields are from 8-hour time-point samples Single-turnover kinetic experiments using 5-FAM-R660 and FH14 at 1:10 ratio. 40 mM MgCl₂ and 200 200 μM of the NTP analog.

3.2.8.2.2 NTP substrate specificity range of FH31

Although generally less efficient than FH14, FH31 demonstrated a similar pattern of reactivity towards the analogs tested. FH31 showed no dependence on biotin and proved to be specific for ATP and its N⁶-modified analogs (**Figure 3-18**). Carolin Scheitl, as a practical course student, performed these sets of kinetic experiments.

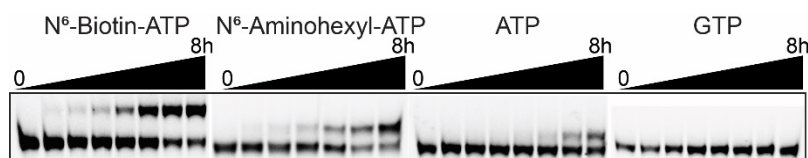


Figure 3-18 ATP analog scope of FH31 ribozyme. FH31 demonstrates a similar reactivity pattern towards tested NTP analogs as FH14.

Table 3-3 k_{obs} values of FH ribozymes, and FH14 with different substrates

Ribozyme	Substrate RNA	N ⁶ -modified ATP	$k_{\text{obs}} \times 10^{-2} \text{ (min}^{-1}\text{)}$	Yield (%) *
FH14	Parent-Flu	Biotin-ATP	1.8 ± 0.2	68 ± 1
FH20	Parent-Flu	Biotin-ATP	0.14 ± 0.01	26 ± 1
FH31	Parent-Flu	Biotin ATP	0.78 ± 0.06	55 ± 1
FH14	5FAM-R660	Biotin-ATP	1.75 ± 0.03	75 ± 1
FH14	5FAM-R660	Aminohexyl-ATP	1.14 ± 0.1	67 ± 3
FH14	5FAM-R660	ATP	0.04 ± 0.005	19 ± 4
FH14	Cy5-R660	ATTO550-ATP	2.5 ± 0.1	84 ± 1
FH14	Cy5-R660	5-FAM-ATP	1.16 ± 0.06	74 ± 2
FH14	Cy5-R660	Cy3-ATP	0.53 ± 0.03	48 ± 3

* For ATTO550, Cy3 and 5-FAM modified substrates the reported yield is after 5 h while other values pertain to 4 h reaction time.

3.2.8.3 MgCl₂ dependency and NTP substrate concentration range

The selection process and the kinetic assays have so far all been performed under 200 μM concentration of the N⁶-aminohexyl-biotin and in the presence of 40 mM MgCl₂. We decided to examine the dependency of FH14 ribozyme on Mg²⁺ and to determine the range of ATP analog concentrations under which FH14 can efficiently modify its target sequence. Two sets of kinetic experiments were prepared. In one group of experiments, the N⁶-aminohexyl-ATP-biotin concentration was kept constant (200 μM) while the MgCl₂ concentration was varied (20, 10, 5, and 1 mM). At ATP analog concentration of 200 μM , reducing MgCl₂ by half did not affect the final yield of the reaction after 8 hours compared to the reaction performed previously at 40 mM MgCl₂. The reaction kinetics however turned significantly slower as the k_{obs} value was reduced by roughly 50%. Further reduction of the MgCl₂ concentration to 10 mM and 5 mM, had a more severe effect on k_{obs} value and the final yield

of the reaction. At MgCl_2 concentration of 1 mM no product was formed during the time course of the reaction (**Figure 3-19 B & D**).

Another group of reactions were performed at constant MgCl_2 concentration (40 mM) and different concentrations of the N^6 -biotinylated ATP analog (100, 50, 20, and 10 μM). The reduction of the ATP analog concentration from 200 μM to 100 μM , at 40 mM concentration of MgCl_2 , had no effect whatsoever on the final reaction yield. However, the k_{obs} value of the reaction was reduced by ~34%. Further reduction of the ATP analog concentration led to a more dramatic reduction in reaction yield and rate constant. These effects however, were not as severe as observed in the case of the lowered MgCl_2 concentration experiments (**Figure 3-19 A & C**). Combined effect reducing ATP analog and MgCl_2 concentration, simultaneously, has not yet been examined. All the k_{obs} values and the reaction yields under different MgCl_2 and ATP analog concentrations are listed in (**Table 3-4**)

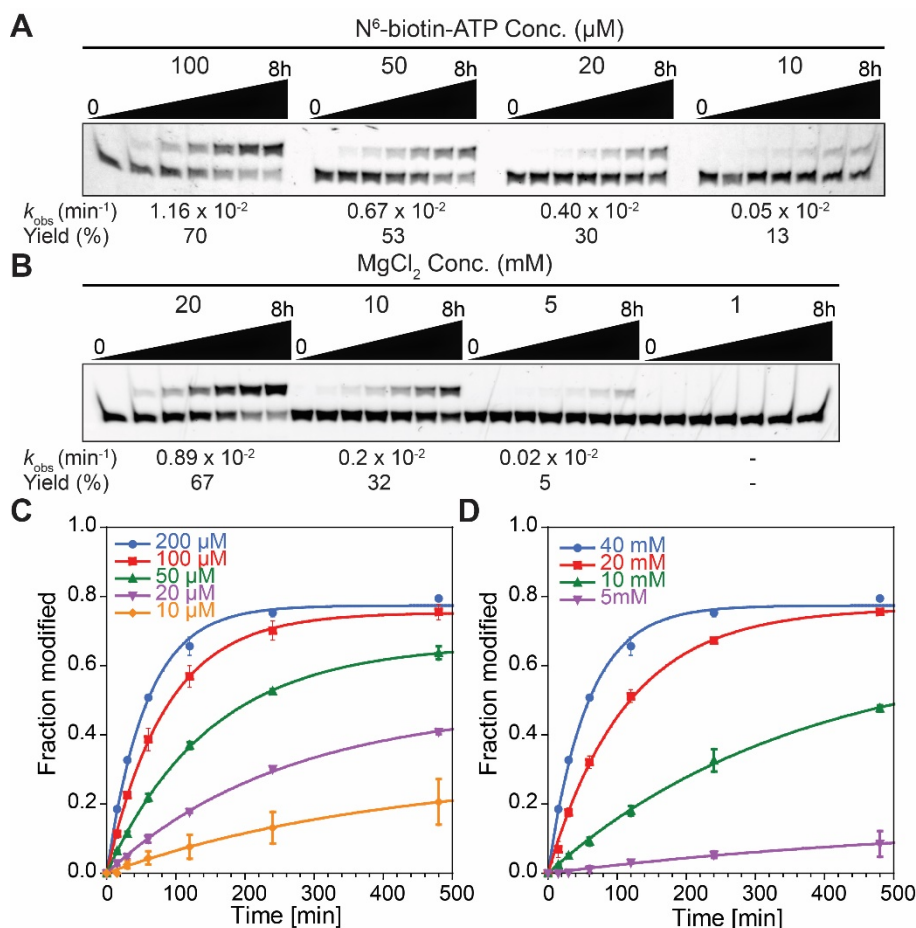


Figure 3-19 MgCl_2 dependency and the ATP analog concentration range of FH14. **(A)** Effect of reducing ATP analog concentration on FH14 efficiency and reaction rate. **(B)** Effect of reducing MgCl_2 concentration on FH14 reaction rate and efficiency **(C)** Pseudo-first order curve fit of FH14 reaction kinetics under varying ATP analog concentrations. **(D)** Pseudo-first order curve fit of FH14 reaction kinetics under varying MgCl_2 concentrations.

Table 3-4 k_{obs} values of FH14 ribozyme-catalyzed labeling of 5FAM-R660 with N⁶-biotin-ATP, with varying Mg²⁺ and ATP concentrations.

MgCl ₂ Conc. (mM)	ATP Conc. (μM)	$k_{\text{obs}} \times 10^{-2}$ (min ⁻¹)	Yield (%) 4h	Yield (%) 8h
40	200	1.75 ± 0.03	75 ± 1	79 ± 0.3
20	200	0.89 ± 0.07	67 ± 1	76 ± 0.5
10	200	0.2 ± 0.02	32 ± 3	48 ± 1
5	200	0.02 ± 0.006	5 ± 1	8 ± 0.4
40	100	1.16 ± 0.06	70 ± 3	76 ± 2
40	50	0.67 ± 0.07	53 ± 0.6	64 ± 2
40	20	0.4 ± 0.03	30 ± 0.3	41 ± 0.6
40	10	0.05 ± 0.02	13 ± 5	21 ± 7

3.2.8.4 Effect of TbCl₃ on FH14 reaction kinetics

It was previously shown that the presence of TbCl₃ remarkably augments the reaction efficiency and the rate constant of 10DM24. The effect is largely dependent on the NTP substrate concentration and can reach a rate enhancement up to four orders of magnitude (Büttner et al., 2014). We were interested to examine whether a similar effect can be observed in the case of the ribozyme FH14. Single turnover kinetic experiments were then performed on FH14 ribozyme in the presence of 0, 10, and 100 μM concentrations of TbCl₃. MgCl₂ and N⁶-biotin-ATP concentrations were identical to selection conditions. TbCl₃ at 100 μM, did indeed dramatically enhance the reaction rate of FH14, causing it to behave in a bi-exponential manner. The rate constant of the fast component of this reaction was improved by roughly 8 times corresponding to the originally determined k_{obs} value in the absence of TbCl₃. At 10 μM concentration however, TbCl₃ showed no effect on the reaction efficiency or rate constant (**Figure 3-20**). These results of these experiments suggest a possible similarity between the catalytic mechanism of FH14 and 10DM24. TbCl₃, however, is known to induce RNA degradation, which may affect the quality of the labeled sample. For this reason, TbCl₃ was not included in any of the labeling experiments.

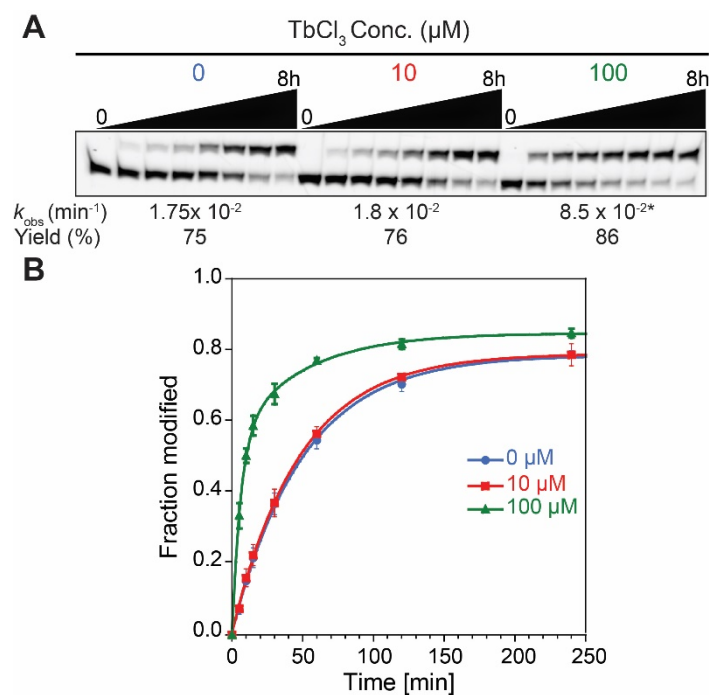


Figure 3-20 Effect of Tb³⁺ on FH14 reaction kinetics. **(A)** Single turnover kinetics of FH14 ribozyme under 0, 10, and 100 μM concentration of TbCl₃. Reaction conditions: 5-FAM-R660: FH14 ratio of 1:10. 40 mM MgCl₂ and 200 μM N6-biotin-ATP. (*) indicates that the value is the fast component of a biexponential kinetic curve fitting. **(B)** Kinetic plot of the FH14 reaction under varying TbCl₃ concentrations. Curve fitting for 100 μM TbCl₃ condition was done using biexponential function. The other two conditions showed pseudo-first order behavior.

3.2.9 FH14-mediated labeling of RNA from cellular origin

Up to this point, all the kinetic experiments using various target sequences or ATP analogs were performed using short (17 to 21 nucleotides) synthetic or in vitro transcribed RNA. These model RNA substrates were used to simplify kinetic analysis using PAGE and fluorescence imaging. The results obtained using these sequences, however, may not directly be translatable to large RNA targets. Larger RNA molecules tend to be more heavily structured compared to short model substrates. Their modification target sites may, therefore, not be readily accessible to the ribozyme. Thus, we decided to test the efficacy of the FH14 ribozyme for labeling of cellular RNA. 5S rRNA (Noller, 1984) and RyhB (Massé and Gottesman, 2002) small RNA from *E. coli* were chosen for this purpose. These transcripts were significantly larger than the model RNA substrates that had been tested. At the same time, they were in a size range that could be easily analyzed using established techniques in our laboratory such as PAGE and fluorescent imaging. The other reason behind this choice was the structured nature of these RNA molecules especially in case of 5S rRNA, which is heavily structured. Moreover, due to its role in ribosomal structure and function, 5S rRNA is one of the most abundantly found RNA molecules in cells, including *E. coli*.

Various potential modification sites were chosen in these RNA molecules to be labeled by FH14 ribozyme.

3.2.9.1 FH14-mediated labeling of 5S rRNA

3.2.9.1.1 Labeling of the pure transcript generated *in vitro*

Three adenine nucleotides were chosen in *E. coli* 5S rRNA (**Figure 3-21 A**). One of the selected positions, A49, resides in a context within a loop region, presumably accessible to the FH14 binding arms. The second position, A66, lies in a single nucleotide bulge in a region where its neighboring nucleotides are mainly paired. The last position, A99, is located in a bubble, where its flanking nucleotides, except for a few, are engaged in base pairing.

FH14 variants were designed with binding arms complementary to 8-10 nucleotides flanking these positions from each side. We also produced *in vitro* transcribed 5S rRNA and labeled it at 3'-end using fluorescein thiosemicarbazide. We first attempted to label each of the chosen positions individually, using their corresponding FH14 ribozyme and N⁶-ATTO550-ATP. Time points were taken from each reaction at 0, 2, and 4 hours and resolved on a 10% analytical denaturing PAGE. Multichannel fluorescent imaging of the gel revealed successful labeling of all three positions, within 2 to 4 hours. The ATTO550 emission intensity of the 2-hour and 4-hour time-points were roughly similar meaning the labeling reaction reached a plateau in under 2 hours (**Figure 3-21 B**). The efficiency of this reaction however could not be estimated, due to the poor resolution of PAGE, for differentiating the labeled from unlabeled RNA, as a result of the large size of the transcript.

We also tested whether the three chosen positions can be labeled simultaneously. We, therefore, set up reactions with only one ribozyme (FH14_5S-A45 or two ribozymes (FH14_5S-A45 and FH14_5S-A66) or all three ribozymes. Non-labeled *in vitro* transcribed 5S rRNA was used in this case and the ATTO550 conjugate ATP was added as the labeling substrate. The reactions were quenched after 4 hours and aliquots taken from them were resolved on 10% denaturing analytical PAGE next to the 3'-fluorescein-labeled 5S rRNA as size marker. As expected, the signal intensity consistently increased from the single-labeled lane to the double-labeled and the triple labeled one (**Figure 3-21 C**).

3.2.9.1.2 Labeling 5S rRNA in the context of total cellular RNA

To test whether the FH14 ribozyme can specifically recognize its cognate substrate when there is a large number of nonspecific sequences present, we attempted to label 5S rRNA in the context of total cellular RNA. Hence, three reactions were set up using each of the 5S rRNA targeting FH14 variants along with total cellular RNA extracted from TOP 10 E. coli cells. N⁶-ATTO550-ATP was used as the labeling substrate. The reaction mixture was precipitated after 4 hours to remove the excess unreacted ATTO550 conjugated ATP. The pellets were then resolved on a 10% denaturing analytical PAGE along with 3'-fluorescein-labeled in vitro transcribed 5S rRNA as a size marker. Multichannel fluorescent imaging of the gel revealed the presence of a strong sharp ATTO550 labeled band, with similar electrophoretic mobility to the 5S rRNA size marker in all three ribozyme lanes (**Figure 3-21 D**). We, therefore, concluded that the designed FH14 variants successfully recognized and labeled 5S rRNA from a large bulk of nonspecific cellular RNA. The gel was then stained with SYBR Gold to reveal the rest of the cellular RNA (**Figure 3-21 D**).

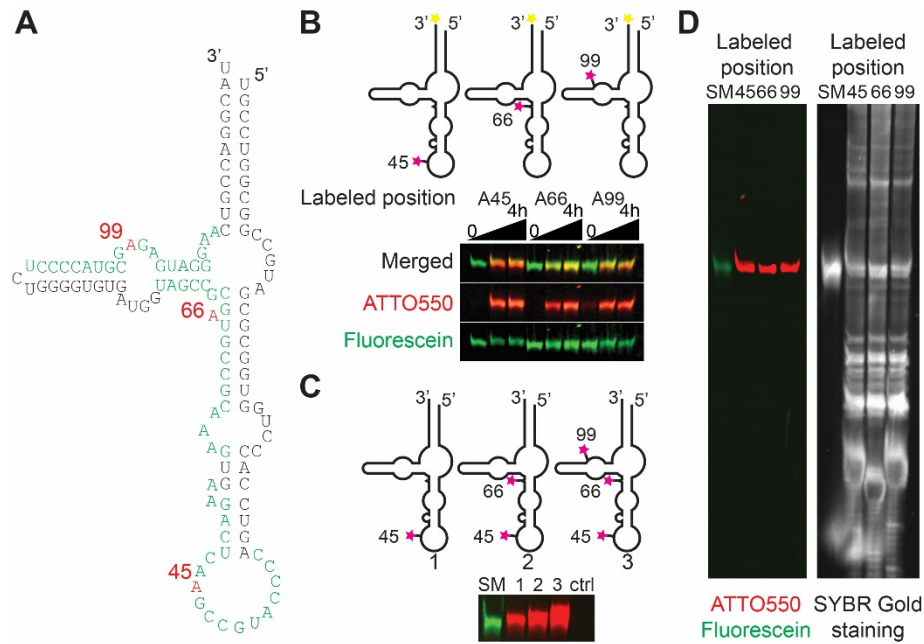


Figure 3-21 FH14 mediated labeling of *E. coli* 5S rRNA. (A) Sequence and secondary structure of the *E. coli* 5S rRNA (Noller, 1984). Positions targeted for modification are highlighted in red. Nucleotides marked in green, represent the ribozyme binding regions. (B) Multichannel fluorescent gel image of FH14 mediated ATTO550 labeling of in vitro transcribed 3'-fluorescein-labeled 5S rRNA at three different positions. Reaction conditions: (5S rRNA-Flu (2 μ M), corresponding ribozyme (20 μ M), 1x selection buffer including 40 mM MgCl₂ and 200 μ M N⁶-ATTO550-ATP, 37°C. Time points taken at 0, 1h and 2h.) (C) Multichannel gel-image of the simultaneous FH14-catalyzed labeling of 5S rRNA at multiple positions, using ATTO550. Reaction conditions: (5S rRNA transcript (1 μ M), one (1), two (2) or all three (3) FH14 ribozymes (10 μ M) in 1x selection buffer including 200 μ M N⁶-ATTO550 and 40 mM MgCl₂, 37°C) 3'-fluorescein-labeled 5S rRNA used as size marker (SM). (D) Left: multichannel gel image of FH14 mediated 5S rRNA labeling in total cellular RNA context. (50 ng *E. coli* total cellular RNA, 50 pmol of the respective ribozyme in 5 μ L of the 1x selection buffer including 200 μ M N⁶-ATTO550 and 40 mM MgCl₂, 4 hours at 37°C. 3'-fluorescein-labeled 5S rRNA as size marker. Right: SYBR Gold staining of the same gel to reveal the rest of the cellular RNA.

3.2.9.2 FH14-mediated labeling of RyhB small RNA

RyhB is an Hfq binding small RNA found in *E. coli* (Massé and Gottesman, 2002). Its main role is to regulate the translation efficiency and stability of the mRNA of proteins involved in RNA metabolism. In iron-rich media, it inhibits translation of its cognate mRNA via recruitment of RNase E (Massé et al., 2005). However, RNase E also cleaves RyhB ensuring coregulation of the level of RyhB with its target mRNA (Massé et al., 2005). In iron-poor media where the target mRNA levels are low RyhB can accumulate at higher levels. We chose two sites (A23 and A62) within this RNA and designed two FH14 variants to target these positions (Figure 3-22 A). The experiments were performed as done for 5S rRNA in the way that RyhB was labeled both as pure transcript and within total cellular RNA. Labeling of the pure transcript was done using 3'-fluorescein-labeled in vitro transcribed

RyhB with ATTO550 conjugated ATP as labeling substrate. Both positions were successfully labeled based on these results within 2 to 4 hours (**Figure 3-22 B**).

For the cellular RNA experiments, TOP10 cells were grown in the presence or absence of the iron chelator 2-2'-Bipyridine (BiP). Iron depletion using BiP is known to stabilize RyhB (Massé et al., 2005) (**Figure 3-22 C**). RNA extracted from each of these cultures were incubated with either of the RyhB targeting FH14 variants. ATTO550 conjugated ATP was used as the labeling substrate and 3'-fluorescein-labeled RyhB in vitro transcript as size marker. No band of the correct size was observed when the labeling reaction was performed on RNA extracted from cells grown in the absence of BiP. A strong truncated band, however, was visible only when the A62 was targeted for labeling. In the case of RNA extracted from BiP treated cells, faint bands of the correct size were visible for both labeled positions. The truncated band however was still visible in the lane pertaining to the A62 labeling product (**Figure 3-22 D**).

The signal intensity in case of the FH14 mediated RyhB labeling experiments was dramatically lower compared to the experiments performed on 5S rRNA. The reason might be that the optimal conditions, in the growth medium, for maximum stability and level of RyhB had not been achieved. The nature of the truncated band appearing only when A62 is labeled is not clear. This position is situated in the HfQ binding site of the RyhB (Tjaden et al., 2006) and therefore one explanation for the appearance of this band is that, this segment of the RNA is protected from further nuclease degradation after cleavage by RNase E.

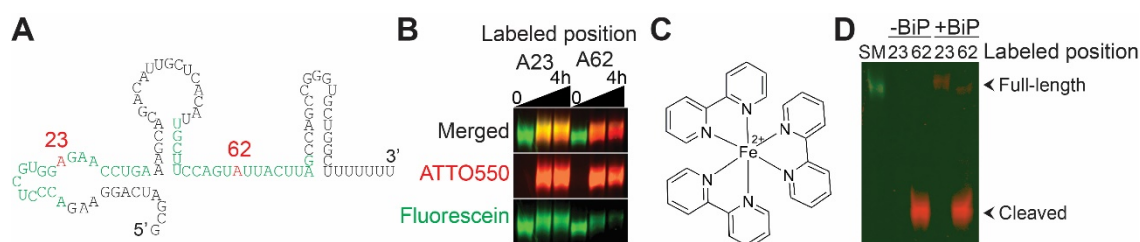


Figure 3-22 FH14-catalyzed labeling of RyhB RNA. **(A)** Sequence and secondary structure of *E. coli* RyhB RNA. Targeted positions shown in red. Ribozyme binding arms designed complementary to the positions shown in green. **(B)** Multichannel fluorescent gel-image of the FH14 catalyzed labeling of the 3'-fluorescein-labeled RyhB in vitro transcript, at two different positions using ATTO550. Reaction conditions: RyhB-Flu (2 μ M), corresponding FH14 (20 μ M), in 1x selection buffer, 40 mM MgCl₂ and 200 μ M N⁶-ATTO550-ATP, 37°C. timepoints: 0, 2, and 4 hours. **(C)** 2-2'-bipyridine (BiP)-iron complex. BiP was added to deplete the culture medium from iron **(D)** Multichannel fluorescent gel image of FH14-catalyzed labeling of RyhB in total cellular RNA. RNA extracted from *E. coli* cells in the presence (+ BiP) or absence (- BiP) of 250 mM 2-2'-bipyridine, was subjected to labeling using FH14 targeted to A23 or A62 or RyhB. 3'-fluorescein-labeled RyhB in vitro transcript was used as size marker (SM). Reaction conditions: 50 ng cellular RNA, 50 pmol of the respective ribozyme, 5 μ L of the selection buffer, 40 mM MgCl₂, 200 μ M N⁶-ATTO550-ATP.

3.3 Tenofovir transferase ribozymes

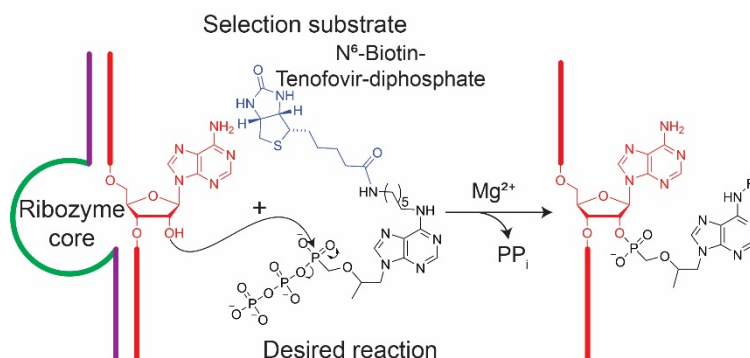


Figure 3-23 Tenofovir transferase ribozyme selection substrate and desired reaction. N⁶-Biotin-Tenofovir-diphosphate (N⁶-Biotin-TenDP) was used for this selection, the desired reaction was a nucleophilic attack by the 2'-OH of the predetermined modification site A onto the phosphonate center pushing out pyrophosphate. The product is, therefore, formation of a phosphonomonoester bridge between the target RNA and tenofovir.

3.3.1 Selection progress

The selection of the tenofovir transferase ribozymes was performed analogous to that of adenylyltransferase ribozymes. The first round started with similar complexity ($\sim 2.7 \times 10^{14}$ unique variants at an average copy number of 7). The incubation step of each round was performed using $\sim 50 \mu\text{M}$ concentration of the RNA pool (including the 3'-LY labeled fraction) in selection buffer including 40 mM MgCl₂ and 300 μM of N⁶-Biotin-Tenofovir-diphosphate (synthesized by a postdoctoral fellow of the group, Dr. Surjendu Dey) (**Figure 3-23**) at 37°C. The first sign of enrichment became apparent at the end of the 4th selection round where $\sim 0.44\%$ of the pool was retained. A rise in the activity level to $\sim 1.42\%$ in the next round further confirmed the success of the enrichment process. The incubation time was reduced from overnight to 4 hours at round 8 and was further decreased to 1 hour by the 12th round (**Figure 3-24 A**). Details of the selection rounds are listed in (**Table 3-5**).

3.3.2 Cis- and trans- activity of the enriched pool

Similar to the adenylyltransferase ribozyme selection, the enriched pool from the 12th round was subjected to cis- and trans-activity assays. The cis-activity assay was performed using streptavidin gel-shift assay on 3'-lucifer yellow labeled RNA pool. Trans-activity assay was performed using roughly 2:1 ratio of the 3'-lucifer yellow-labeled substrate sequence over the trans-acting pool transcript generated by PCR and in vitro transcription as shown in (**Figure 3-6 C**). Both reactions were performed under identical conditions as the selection

experiment. According to the results of the cis-activity assay, over 80% of the pool was biotinylated after overnight incubation under selection conditions (**Figure 3-24 B**). Trans-activity assay also confirmed that the pool was compatible for intermolecular target sequence modification (**Figure 3-24 C**).

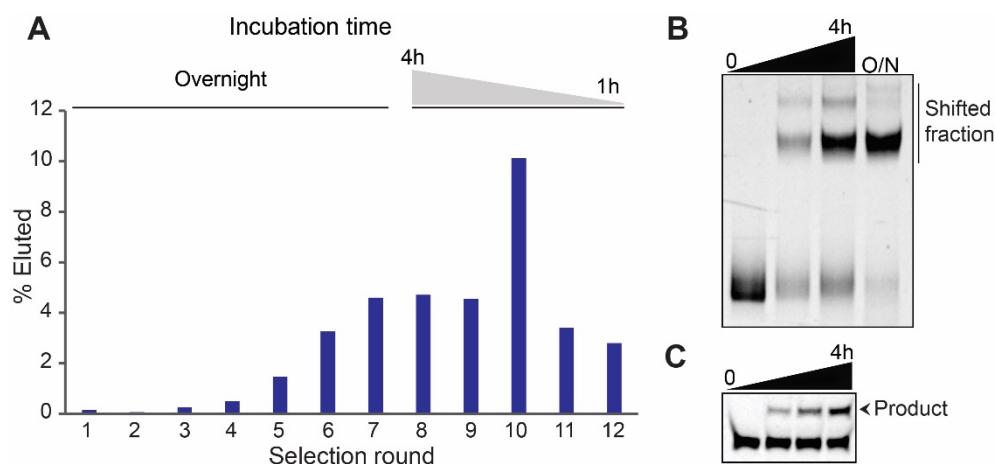


Figure 3-24 Selection progress and cis- and trans-activity of the enriched pool. **(A)** The enrichment level at each selection round was estimated by fluorescence measurement of the eluted RNA. Incubation time was reduced after significant enrichment was observed to allow the selection of faster-reacting ribozymes. **(B)** Streptavidin gel-shift assay of the round 12 pool. reaction conditions: 15 μM Round 12 RNA Pool labeled with LY-CH in selection buffer including, 300 μM N^6 -biotin-TenDP and 40 mM MgCl_2 . Time points taken at 0, 1, 6 hours, and overnight were subjected to streptavidin gel-shift on 10% PAGE. **(C)** Trans-activity assay of the enriched pool. LY-CH labeled substrate sequence (24 μM) and the trans-active round 13 pool (12 μM) were dissolved in the selection buffer containing 300 μM of the N^6 -biotin-TenDP and 40 mM of MgCl_2 . Time points were taken at 0, 1, 2, and 4 hours and overnight.

Table 3-5 Details of the selection rounds of tenofovir transferase ribozymes.

Selection round	1	2	3	4	5	6	7	8	9	10	11	12
Pool concentration (μM)	~ 50											
Incubation time (h)	Overnight							4	3.5	2.5	2	1
Eluted %	-	-	-	0.4	1.4	3.2	4.5	4.7	4.5	10	3.4	2.8

3.3.3 Sequence identification of the evolved variants

3.3.3.1 Sanger sequence

The outcome of the 12th selection round was subjected to cloning and the resulting colonies was confirmed using colony PCR. 10 colonies were randomly chosen for submission to Sanger sequencing. The sequencing results revealed the presence of only two variants in 10 colonies that were denoted as FJ1 with eight copies and FJ8 with two copies. The sequence groups identified through Sanger sequencing are listed in **Table 3-6**

Table 3-6 Sequence groups identified from the Sanger sequencing of isolated clones

Group	5'-Sequence-3'	Number repeated
FJ1	CCACCCTCATAAAACTGAAGATCCTTTGGCAAGGGTCTA	8
FJ8	ACGGTGTTACAAAACGAAACCTTAGGACATTACCTTCCCC	2

3.3.3.2 Next-generation-sequencing of the enriched pool

To study the selection pool more deeply and investigate the enrichment patterns of different variants, we performed Illumina sequencing. The samples from round 7 and round 12 were chosen for submission to NGS analysis. Round 7 was picked since it was the last low-stringency round and presumably had higher complexity than the round 12 samples. Comparison of the change in frequency of each variant from round 7 to round 12 also provided a measure of enrichment or de-enrichment of the variants through the selection rounds. After quality control and trimming of the adapters, the sequencing data were analyzed using fastaptamer tools (Alam et al., 2015). Ten major sequence clusters were identified through these analyses (**Figure 3-25**) (Details of the read numbers in (**Table 8-6**) of the appendix). Surprisingly, the clusters to which FJ1 and FJ8 belonged, were not the most abundant groups. These ribozymes belonged to cluster 4 and cluster 2 respectively. These two sequences however, had the highest log₂ values (8.67 for FJ1 and 9.74 for FJ8) (**Table 3-7**) based on the comparison of their read per million (RRM) values between round 7 and round 12. These values are calculated using $\log_2(\text{RPM}_{\text{round7}}/\text{RPM}_{\text{round12}})$ and indicate that these ribozymes were highly enriched between round 7 to round twelve possibly due to their superior reaction kinetics. The next highest enrichment level belonged to the first sequence of cluster 9 (FJC9) with a log₂ value of 4.32 (**Table 3-7**) (full details in **Table 8-7** of the appendix). The first sequence of cluster 1 (FJC1), demonstrated the next highest log₂ value of 1.95. Log₂ values between 0 and 2 indicate no meaningful enrichment or de-enrichment. Other clusters mostly showed negative log₂ values, which indicates that they were gradually lost from round 7 to round 12. FJC9 and FJC1 were, therefore, chosen for further investigation along with FJ1 and FJ8. From the de-enriched clusters, FJC3 from cluster 3 (log₂= -0.35) was also chosen for comparison. A partial alignment of the core sequence of these ribozymes can be seen in (**Figure 3-26 A**). FJ1 and FJ8 are remarkably similar and share a resemblance with FJC1 as well. FJC9 and FJC3, however, seem to be completely different.

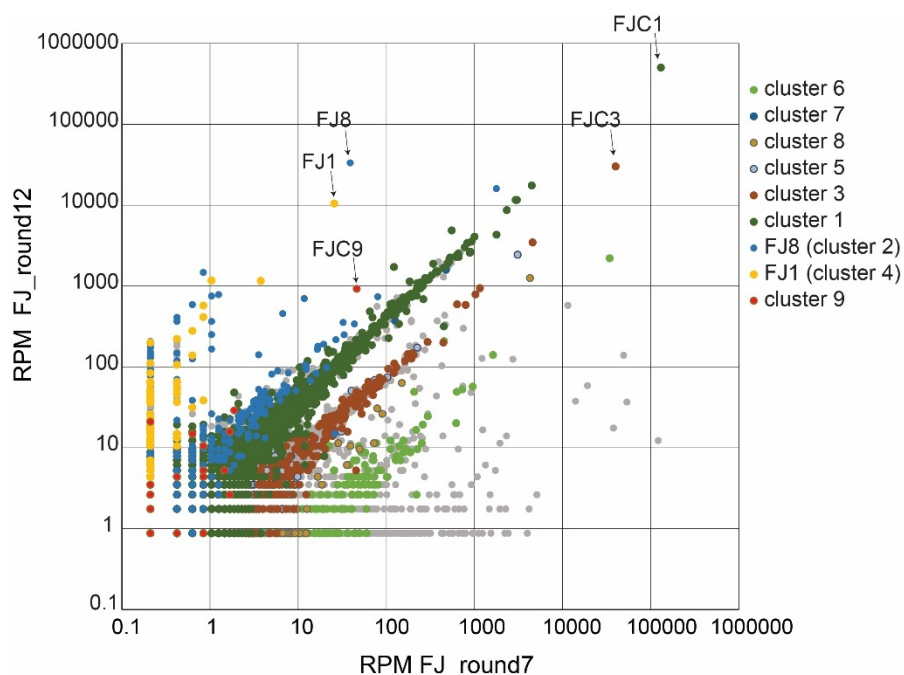


Figure 3-25 Abundance of individual sequences in reads per million (RPM) compared in round 7 and round 12. Sequence clusters identified in round 12 are shown in the same color. The dots representing the chosen sequences are marked with an arrow.

Table 3-7 The chosen sequences from NGS data analysis with their respective \log_2 values

Variant	5'-Sequence-3'	Cluster#	$\log_2(X/Y)^*$
FJ8	ACGGTGTACAAAACGAAACCTTAGGACATTACCTTCCCC	2	9.74
FJ1	CCACCCTCATAAACTGAAGATCCTTTGGCAAGGGTCTA	4	8.67
FJC9	ACGAGATATGTTGCACTACACTTTAGCGAATTGGGCATCC	9	4.32
FJC1	GAAACGTGTCACATAAGAAAACGGTAAACTAGCAAGTTCC	1	1.95
FJC3	CGAATGCCACCGAACCCTATAATTGCCGCTCCAATTTC	3	-0.35

3.3.4 Kinetic comparison of the evolved variants

The DNA templates coding for the selected variants, with loop and 3'-extension removed, were ordered and transcribed. These transcripts were subjected to single-turnover kinetic experiments, using 10:1 ratio of the ribozyme over 3'- fluorescein-labeled parent substrate sequence. Buffer and reagent concentrations were identical to the selection reaction. Based on the results of these experiments, FJ1 was clearly determined as the fastest reacting variant ($k_{\text{obs}} = 0.27 \text{ h}^{-1}$). FJ8 and FJC9 showed similar reaction kinetics with k_{obs} value of 0.09 and 0.1 h^{-1} respectively. The overall efficiency of all three variants were roughly the same (over 70 % after 22 hours). FJC1 showed significantly slower reaction kinetics compared to FJC9 however, the final yield of FJC1 was similar to that of FJC9. As expected, FJC3 was the

least efficient variant, demonstrating almost linear kinetics and a final yield of ~ 39% (**Figure 3-26 B & C**). The details of the reaction yields and k_{obs} values of the tested variants are listed in (**Table 3-8**). These results are correlated to the \log_2 values calculated for each variant from round 7 to round 12.

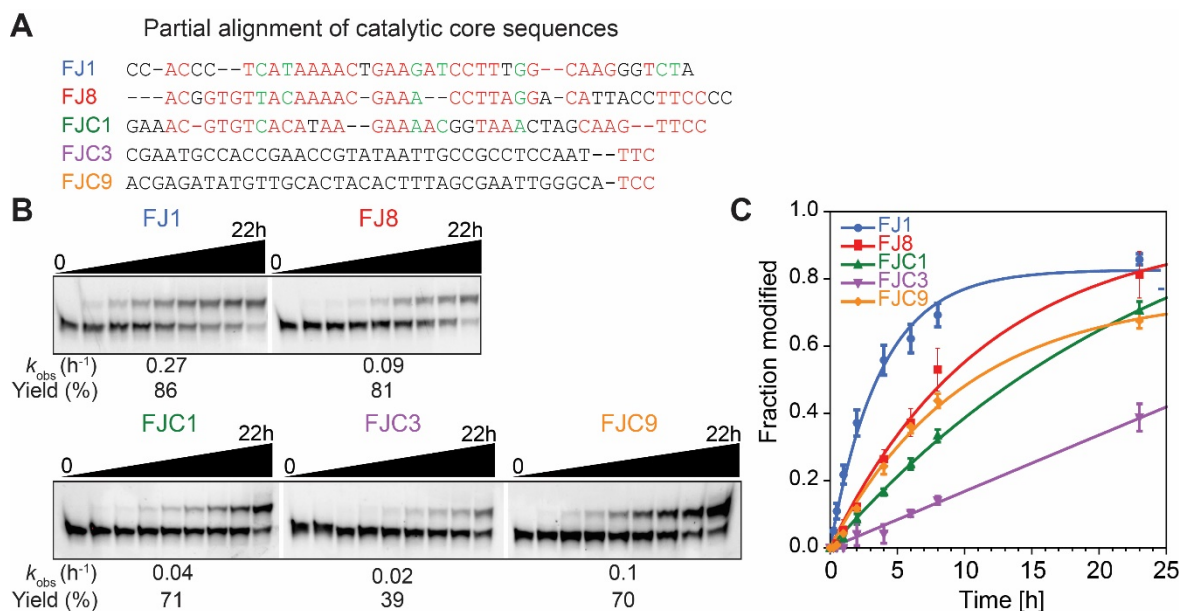


Figure 3-26 Partial alignment and kinetic comparison of the variants identified from Sanger and next-generation sequencing. (A) Partial alignment of the catalytic core sequences of the chosen variants. Red: identity, Green: similarity, Black: difference. (B) Gel image of the single turnover kinetic assay performed on the chosen FJ variants. 3'-fluorescein-labeled parent sequence: ribozyme at 1:10 ratio in selection buffer including 40 mM MgCl₂, 300 μM N⁶-Biotin-TenDP, 37°C. Time points: 0, 0.25, 0.5, 1, 2, 4, 6, 8 and 22 hours. (C) Pseudo-first order kinetic curve fitting of the quantified bands.

Table 3-8 Kinetic parameters of the FJ variants

Ribozyme	Substrate sequence	k_{obs} (h ⁻¹)	Final yield (%)
FJ1	3'-fluorescein-Parent	0.27 ± 0.04	86 ± 2
FJ8	3'-fluorescein-Parent	0.09 ± 0.01	81 ± 7
FJC1	3'-fluorescein-Parent	0.04 ± 0.004	71 ± 3
FJC3	3'-fluorescein-Parent	0.02 ± 0.002	39 ± 4
FJC9	3'-fluorescein-Parent	0.10 ± 0.005	70 ± 2

3.3.5 Secondary structure prediction of FJ1, FJ8 and FJC9

Secondary structure prediction was performed on the three most efficient variants, using the Vienna RNAfold server. The predicted structures of the FJ1 and FJ8 ribozymes are very different despite the high degree of similarity in their sequence. Schematic representation of

the predicted structures and the corresponding base pair probability dot plots are depicted in (Figure 3-27).

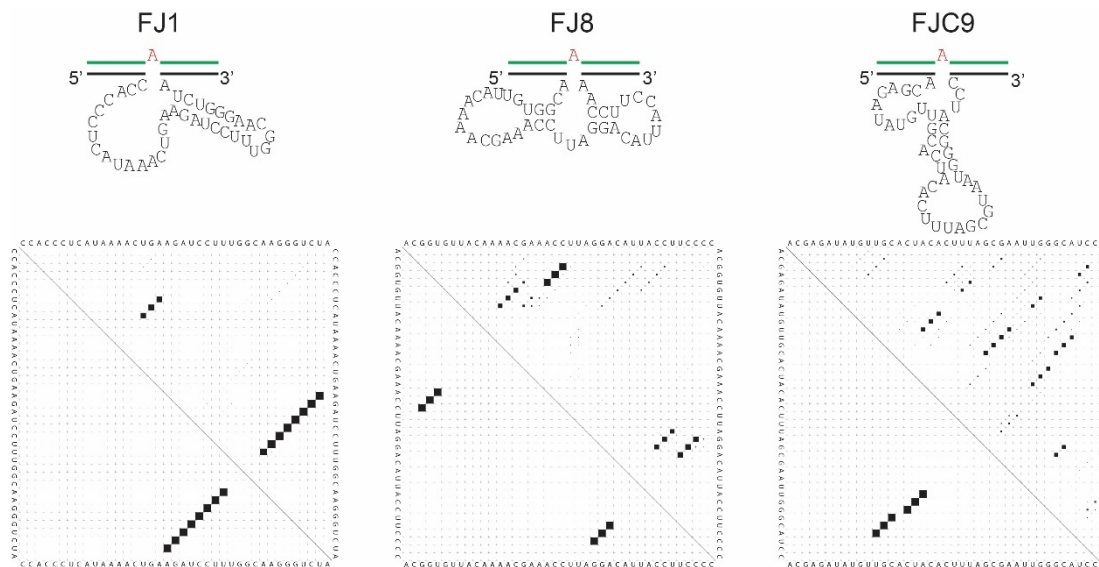


Figure 3-27 Predicted secondary structures and base pair probability dot plots of the selected FJ ribozymes, generated by the Vienna RNAfold server.

3.3.6 Characterization of the reaction product

3.3.6.1 Probing of the modification site

Trans-activity assay of the enriched pool and the kinetic experiments confirmed that the FJ ribozymes are active in trans. As in the case of adenylyltransferase ribozymes, these experiments offer no information regarding the exact positions of modification. RNase T1 and alkaline hydrolysis experiments were then performed to identify the modification site of two ribozymes, FJ1 and FJC9. We, therefore, set-up semi-preparative reactions using either of these ribozymes, to modify the 5FAM-R660 and isolated the modified product (**Figure 3-28 A**).

RNase T1 and alkaline hydrolysis experiments on the oligonucleotide modified using FJ1, revealed a similar pattern of bands as in the case of FH14 and FH31 ribozymes (**Figure 3-28 B**). This means that FJ1 modifies the desired position. The results from the FJC9 ribozyme however were different. The bands corresponding to cleavage at G9 were missing in all RNase T1 digestion and alkaline hydrolysis lanes from the modified RNA. The A10 cleavage band, in this case, was present, however, it was shifted compared the unmodified samples. The shifted pattern of bands was lost when the cleavage happened before G9. These

results, therefore, determine 2'-OH of the G9 rather than A10, as the modification target of the FJC9 ribozyme (**Figure 3-28 B**).

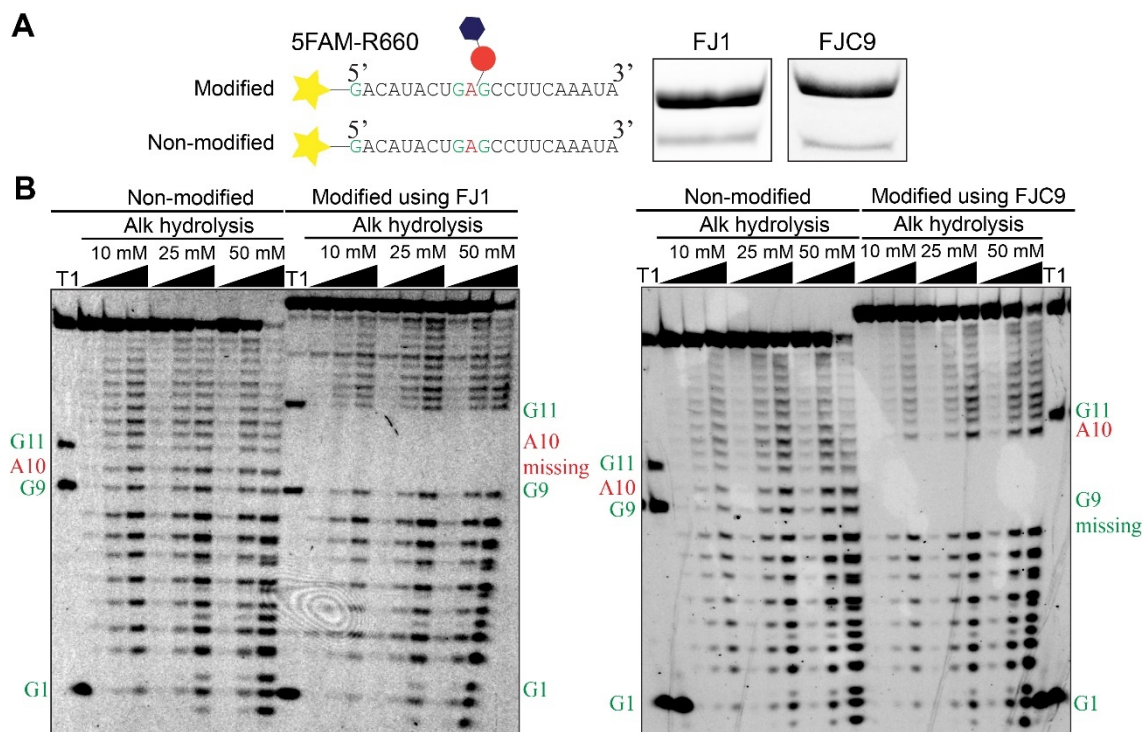


Figure 3-28 Probing of the FJ1 and FJC9 modification site. **(A)** Preparative scale modification of 5FAM-R660 using FJ1 and N⁶-Biotin-TenofovirDP. **(B)** Alkaline hydrolysis and RNase T1 probing gel image of the 5FAM-R660 modified using FJ1 and FJC9. The results indicate the modification site for FJ1 as A10, whereas FJC9 modification site is revealed as G9.

3.3.6.2 ESI-mass analysis of the reaction product

The desired reaction in this selection experiment was the formation of a phospho-monoester bond between 2'-OH of the modification site and tenofovir. To confirm this linkage type we semi-preparatively modified the R594 using FJ1 and N⁶-biotin-TenDP used in selection. The reaction product was isolated at a yield of roughly 50%. The product was then submitted to ESI-mass spectrometry. The observed mass was found to be 6125.14 Da vs the calculated mass of 6125.08371 Da (**Figure 3-29**) (Full mass spectrum is shown in **Figure 8-2** of the appendix).

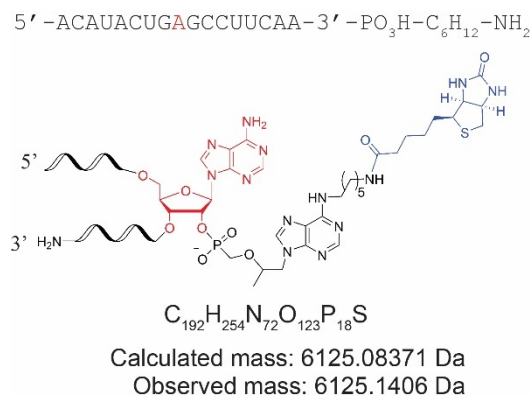


Figure 3-29 Mass analysis of FJ1 ribozyme product. Sequence of R594 and the structure of the product of the FJ1-catalyzed reaction. Calculated mass of the product was 6125.08371 Da vs the observed mass of 61205.1406 Da.

3.3.6.3 Dbr1 resistance of the phosphonyl branch

One of the reasons behind choosing tenofovir-diphosphate analogs as selection substrate was formation of the phosphonomonoester linkage. We assumed this type of linkage to be more enzymatically stable than the phosphodiester branch, formed by the adenylyltransferase. Debranching enzymes in particular are responsible for efficient hydrolysis of 2'-5' phosphodiester linkages in cell (Chapman and Boeke, 1991). We decided to compare the stability of the phosphonomonoester and phosphodiester linkages against yeast debranching enzyme (Dbr1). Debranching assay experiments with recombinant Dbr1 (kindly provided by A. Hoskins, U of Wisconsin) were therefore set up on 5FAM-R550 substrate sequence carrying either N⁶-biotinylated tenofovir or N⁶-biotinylated adenylyl branch structures. These modifications were introduced using FJ1 and FH14 ribozymes respectively (**Figure 3-30**). At 1 μ M MnCl₂ concentrations, the phosphonomonoester branch strongly resisted the debranching by Dbr1 with only ~17% of the RNA, debranched after 1-hour incubation. The phosphodiester counterpart, however, was quantitatively debranched under the same conditions, within the same duration of time. At 1 μ M MnCl₂, between 15 to 30 minutes time points the fraction of the debranched RNA rises from ~30% to ~65% in the case of the phosphodiester branch. The half-life ($t_{1/2}$) of the phosphodiester bond under these conditions, therefore, can be estimated to be roughly 25 minutes. On the other hand, the $t_{1/2}$ of the phosphonomonoester bond, introduced by FJ1, can be estimated to be over 200 minutes under the same conditions. At 5 μ M MnCl₂ concentrations, the FJ1 modification product is still significantly more resistant to debranching than the FH14 modified RNA. The half-lives of these two products are estimated to be ~20 minutes for the FJ1 product and less than 5 minutes for the FH14 modified RNA. At 10 and 100 μ M MnCl₂ however, both branch

structures are rapidly degraded. These conditions clearly demonstrate higher stability of the phosphonomonoester bond against debranching compared to the phosphodiester branch. The FJ1 labeling product, however, was not completely resistant to debranching. Moreover, the conditions tested in these debranching assays, do not reflect the debranching rate of these products under cellular conditions.

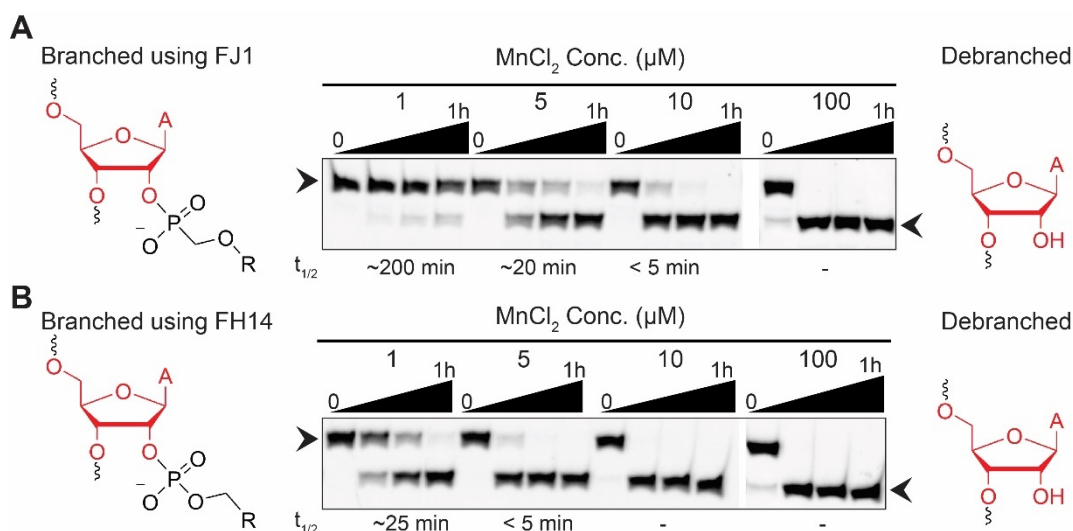


Figure 3-30 Resistance of FJ1 vs FH14 modified RNA to debranching. Gel images of the Dbr1 catalyzed debranching of the (A) FJ1- and (B) FH14- modified 5FAM-R660. Debranching was done using the modified RNA (1 μM), 50 ng/μL of Dbr1 in 50 mM Tris pH = 7.4, 25 mM NaCl, 2.5 mM DTT, 0.01% (V/V) Tween-20, 0.15% (V/V) glycerol, 1, 5, 10 and 100 μM MnCl₂ at 37°C. Time points taken at 0, 15, 30, and 60 minutes.

3.3.7 Characterization of the chosen variants

3.3.7.1 Target sequence requirements

3.3.7.1.1 Modification-site context requirements of FJ1 and FJ8

Substrate sequence generalizability of FJ ribozymes was investigated in the same manner as it was done for FH ribozymes. The same transcripts with transition and transversion mutations (**Figure 3-12 A**) outside the GAG context were utilized for FJ1 ribozymes with co-varied binding arms. FJ1 was unable to modify any of the three extensively mutated variants. Even after 22 hours of incubation, no product was formed (**Figure 3-31 A**). TV1 mutant of the parent substrate sequence was also tested with the corresponding FJ8 ribozyme. No modification was observed when testing this ribozyme: substrate sequence pair either (**Figure 3-31 E**).

Since FJ1 and FJ8 ribozymes were unable to modify the mutated substrate sequence variants, new mutants were designed to further investigate their modification context requirement

(**Figure 3-31 B**). In one variant, the constant GAG window from the parent substrate sequence was extended by one nucleotide from each side and the rest of the sequence was transition mutated. The result of these mutations was called TM-UGAGC. The 5'-UGAGC-3' context in this transcript is identical to the parent substrate sequence. Another sequence was also designed in which the segment 5' to the GAG was transversion mutated while the 3'-segment was kept unchanged resulting in 5'TV1-3'Parent substrate sequence. Both the TM-UGAGC and 5'TV1-3'Parent substrates were successfully modified by their corresponding FJ1 ribozyme. The reaction rate and efficiency for both these substrate sequences were nearly identical to the parent sequence (**Figure 3-31 C & D**) (**Table 3-9**). The common feature in these sequences was that the GAG context was followed by a C nucleotide. At this point, the modification context sequence of the FJ1 was determined to be 5'-GAGC-3'. Based on this information, a new mutant was designed in which 5'-GAGC-3' context was left intact. The rest of the sequence was transversion mutated to complementary nucleotides with respect to the parent sequence. This mutant, denoted as TV1-GAGC, was indeed modified by its respective FJ1 variant, however, the reaction efficiency was extremely poor (**Figure 3-31 C & D**) (**Table 3-9**). We, therefore, concluded that for efficient FJ1 mediated labeling to occur, it is best if the 5'-GAGC-3' context is followed by at least one pyrimidine nucleotide.

We tested whether mutations to the GAG context are tolerated by mutating each G on either side of modification site A of the parent substrate sequence. The rest of the sequence was kept the same as the parent. Mutational tests within the GAG context revealed the mutations to the G situated 5' the bulged A are tolerated however, when pyrimidines are substituted in this position, reaction rate and efficiency fall drastically. Replacing this G with an A however, has no effect on either of those parameters (**Figure 3-31 C & D**) (**Table 3-9**). Mutating the 3'-G to an A on the other hand, completely abolishes the reactivity. Mutating the modification site A to a G, reduced the efficiency severely, however, it did not completely abolish the reactivity (**Figure 3-31 C**).

Considering all the data from substrate sequence specificity tests, it can be concluded that 5'-AGC-3' is the minimal essential required context for FJ1 modification. For optimal labeling efficiency, however, 5'-RAGCY-3' is preferred.

FJ8 ribozyme could also modify the parent substrate sequence with AAG modification context as well as the 5'TV1-3'Parent sequence. The original substrate sequence with

modification site A mutated to G was not modified by this ribozyme. Other mutated target sequence variants were not examined for this ribozyme (**Figure 3-31 E**).

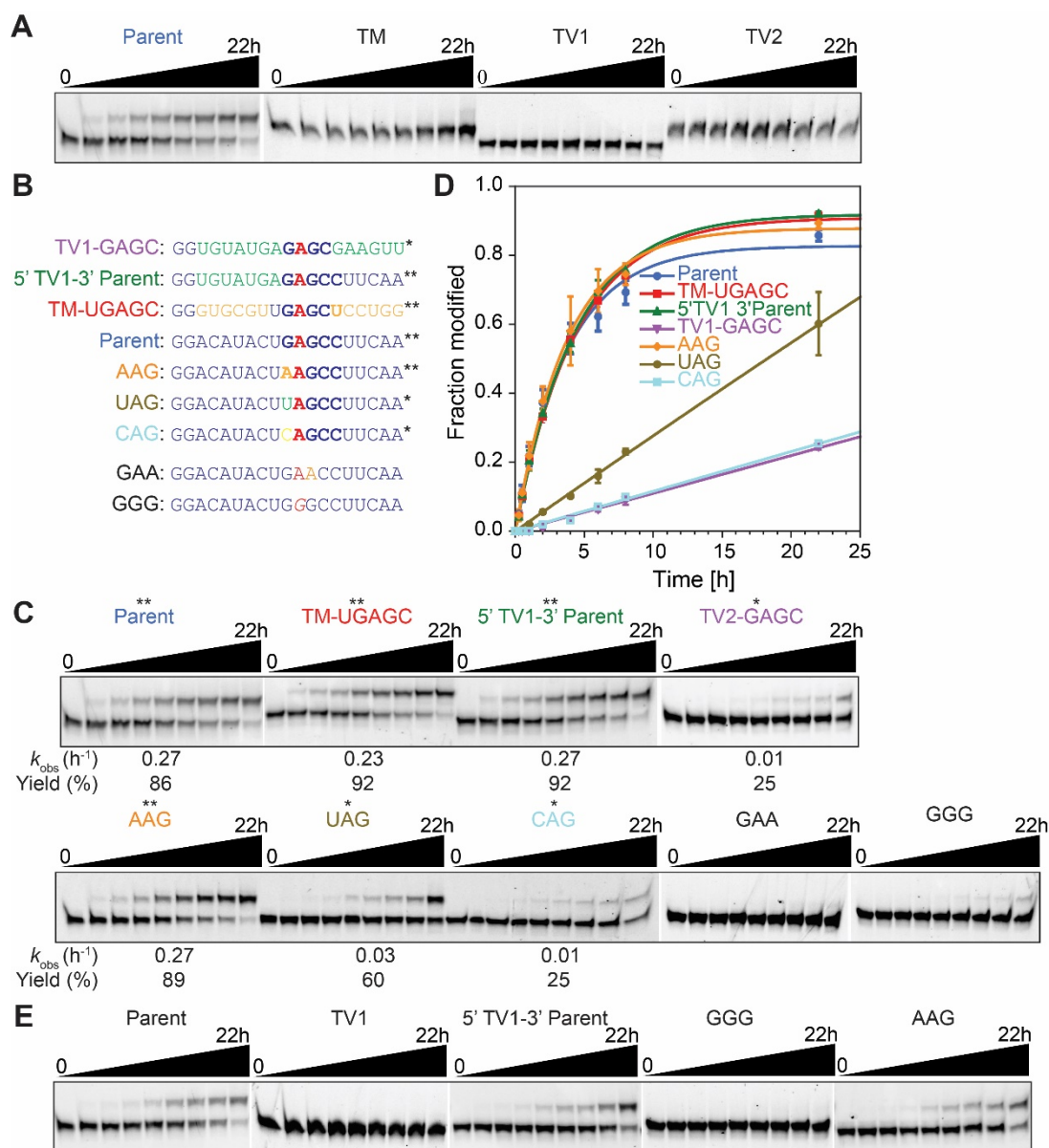


Figure 3-31 Target sequence variability of FJ1 and FJ8 ribozymes. **(A)** Gel image of the single turnover kinetic experiments using parent, TM, TV1 and TV2 substrate sequences. **(B)** Additional substrate sequence variants designed for investigation of the modification context sequence requirement of FJ1 and FJ8. Ribozyme arms were also co-varied. Blue: same as parent, Orange: TM mutation, Green: TV1 mutation, Yellow: TV2 mutation. Red: modification site A, Italic: mutation of the modification site. Two asterisks signify similar efficiency to parent, and one asterisk indicate severely reduced efficiency. No asterisk: no modification. Common features of the reactive sequences shown in bold. **(C)** gel image and **(D)** kinetic plot of the kinetic experiments performed using sequence variants in **(B)** and their corresponding FJ1 ribozyme. **(E)** Gel image of the kinetic experiments performed using substrate sequence variants shown in **(B)** and their corresponding FJ8 ribozyme. Reaction conditions in all cases: 3'-Fluorescein labeled substrate sequence and corresponding ribozyme at 1:10 ratio. Selection buffer, 40 mM MgCl₂, 300 μM N⁶-Biotin-TenDP, 37°C. time points: 0, 0.25, 0.5, 1, 2, 4, 6, 8 and 22h.

Table 3-9 Kinetic Details of the investigation of the FJ1 modification context.

Substrate sequence	Modification context*	k_{obs} (h ⁻¹)	Final yield (%)
Parent	5' - GAGCC -3'	0.27 ± 0.04	86 ± 2
TM-UGAGC	5' - GAGCU -3'	0.23 ± 0.01	92 ± 1
5'TV1-3'Parent	5' - GAGCC -3'	0.23 ± 0.02	92 ± 1
TV1-GAGC	5' - GAGCG -3'	0.011 ± 0.001	25 ± 0.1
Parent-AAG	5' - AAGCC -3'	0.27 ± 0.06	89 ± 1
Parent-UAG	5' - UAGCC -3'	0.031 ± 0.001	60 ± 9
Parent-CAG	5' - CAGCC -3'	0.012 ± 0.002	25 ± 3

* Modification site is shown in bold.

3.3.7.1.2 Target sequence generality of FJC9

FJC9 ribozyme demonstrated a greater level of flexibility than FJ1 and FJ8, regarding target sequence generalizability. This ribozyme was able to modify all three mutated target sequences (TM, TV1, TV2), albeit at varying efficiencies. The most efficiently modified target sequence, by FJC9, was the parent sequence. The other substrate sequences were significantly slower; however, the final reaction yield was over 60 %, except for TV2 mutant (**Figure 3-32 A**).

Mutating the G nucleotide, which happens to be the modification site for this particular ribozyme, significantly reduces the reaction rate (**Figure 3-32 B-D**). When this position is mutated to A the final reaction yield is roughly the same as in the case of G. The k_{obs} value, on the other hand, is reduced by roughly an order of magnitude. Mutations of the modification site G to U or C, have much more drastic effects on yield and reaction rate (**Table 3-10**).

Mutating the G at the 3' of the GAG context enhances the reaction rate with the best results achieved by mutation to C. This mutation increases the k_{obs} value by 20% and the final yield is increased. Mutations of this G to A or U do not change the rate constant significantly, however, the final yields are slightly increased (**Figure 3-32 B-D**)(**Table 3-10**). When A in the **GAG** context is mutated to G, no product formation can be observed even after overnight incubation (**Figure 3-32 B & C**). The best modification context tested for the FJC9 is determined to be 5'-**GAC**-3'. The combined effect of mutating both Gs at the same time, however, has not been studied. The surprising observation in the case of this ribozyme is that the bulged A seems to be even more essential than the modification site G. The reason behind this fact can only be known upon deeper structural and mechanistic studies on this ribozyme.

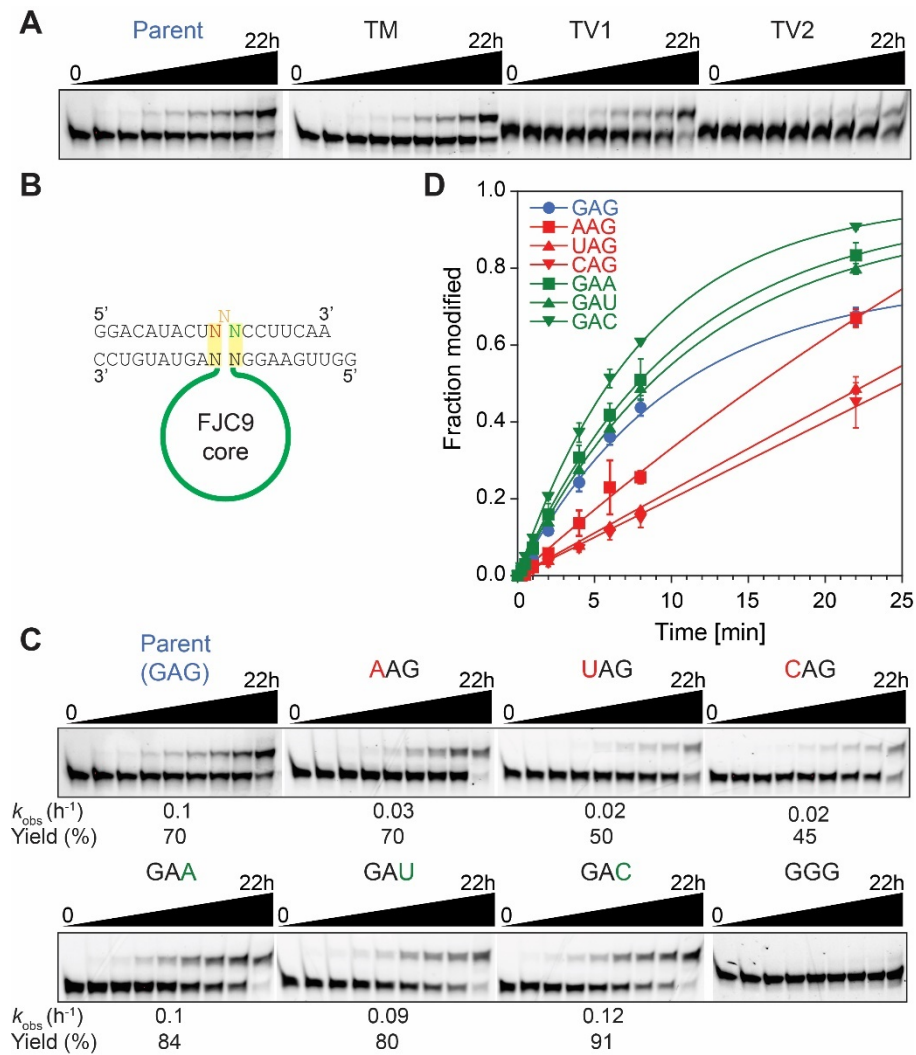


Figure 3-32 Substrate sequence flexibility of FJC9. **(A)** Gel image of kinetic experiments performed using four main substrate sequence mutants and their corresponding FJC9 ribozyme. **(B)** Mutations into the GAG modification context of the parent substrate sequence. Red: modification site, Orange: bulging nucleotide. Positions placed in yellow boxes were co-varied in the ribozyme and the substrate sequence. **(C)** gel image and **(D)** kinetic plot of the single turnover kinetic experiments performed using GAG context mutated substrate sequence variants and their corresponding FJC9 ribozymes. Reaction conditions in all cases: 3'-Fluorescein labeled substrate sequence and corresponding ribozyme at 1:10 ratio. Selection buffer, 40 mM $MgCl_2$, 300 μM N^6 -Biotin-TenDP, 37°C. time points: 0, 0.25, 0.5, 1, 2, 4, 6, 8 and 22h.

Table 3-10 Kinetic details of the investigation of the FJC9 modification context.

Substrate sequence	Modification context	k_{obs} (h ⁻¹)	Final yield (%)
Parent	5' -GAG-3'	0.10 ± 0.005	70 ± 2
Parent-GAA	5' -GAA-3'	0.1 ± 0.02	84 ± 3
Parent-GAU	5' -GAU-3'	0.09 ± 0.01	80 ± 1
Parent-GAC	5' -GAC-3'	0.12 ± 0.01	91 ± 1
Parent-AAG	5' -AAG-3'	0.031 ± 0.001	70 ± 2
Parent-UAG	5' -UAG-3'	0.022 ± 0.001	50 ± 1
Parent-CAG	5' -CAG-3'	0.020 ± 0.003	45 ± 7

3.3.7.1.3 Effect of mismatch on FJ1 and FJC9 labeling efficiency

We examined the effect of introducing mismatch on either side of the bulging A-nucleotides. Substrates Parent-AAG and Parent-GAA were tested for modification with FJ1 and FJC9 specific for Parent sequence (GAG). Parent-AAG and Parent-GAA substrates combined with original ribozymes result in A-C mismatches 5' or 3' to the bulging A, respectively (**Figure 3-33 A**). For FJC9 ribozyme, introduction of the A-C mismatch at 5' side of the A, resulted in an almost complete loss of activity. The same mutations at 3', resulted in slightly more efficient labeling with a final yield of 19% after 22 hours of incubation (**Figure 3-33 B**). This number is still drastically lower than the matched combination, which yields in 70% modification in the same amount of time. For FJ1, A-C mutation at 5' reduced the labeling yield after 22 hours to 17% from 86% observed with the matched combination. Mismatch at the 3' resulted in complete loss of activity (**Figure 3-33 B**).

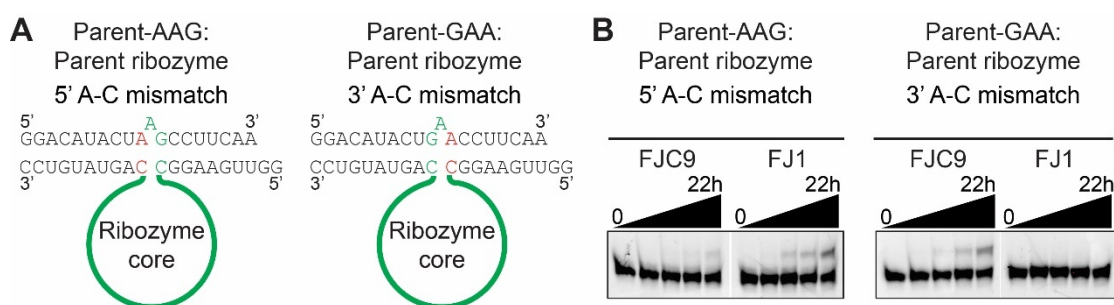


Figure 3-33 Effect of mismatch on FJ ribozyme labeling efficiency. (A) Mismatched substrate:ribozyme combinations tested. Red: Mismatched positions. Green: unchanged with regards to the GAG-context. (B) Kinetic gel of the mismatched substrate:ribozyme pairs. Reaction conditions: 3'-Fluorescein labeled substrate sequence and corresponding ribozyme at 1:10 ratio. Selection buffer, 40 mM MgCl₂, 300 μM N⁶-Biotin-TenDP, 37°C. time points: 0, 2, 4, 8 and 22h).

3.3.7.2 Tenofovir analog scope of the ribozymes

The ribozymes were selected using a TenDP conjugated to biotin via an N⁶-aminohexyl linker. Thus, a risk exists that the evolving variants during selection may develop a dependency on biotin or the aminohexyl linker. To test whether the linker or biotin are essential, we performed single-turnover kinetic experiments using commercially available unmodified tenofovir-diphosphate as substrate (**Figure 3-34**). All three ribozymes accepted the unmodified TenDP, showing no dependency whatsoever on biotin or the aminohexyl linker. Surprisingly, FJ1 and FJ8 ribozymes demonstrated significantly higher labeling rates and efficiencies in the presence of unmodified tenofovir than the selection substrate, although the yields were difficult to quantify due to the poor resolution (**Figure 3-35 A & C**).

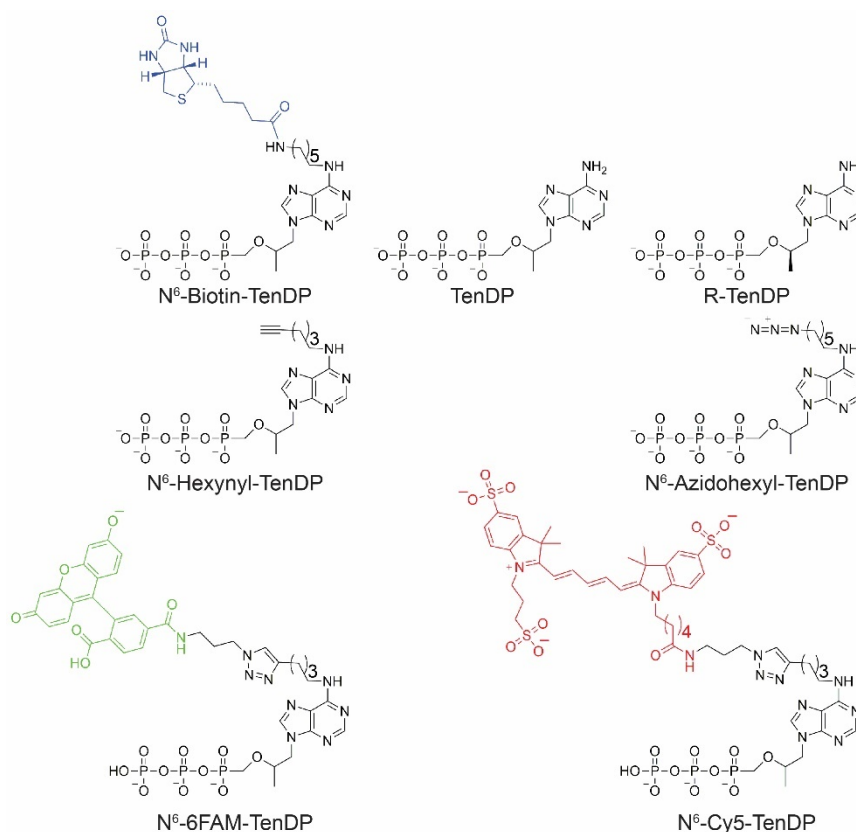


Figure 3-34 list of Tenofovir-diphosphate analogs used in this study. R-TenDP is the only commercially available analog. The rest of the analogs were synthesized in our lab by Dr. Surjendu Dey.

In the next step, we replaced biotinamidohexyl moiety in the selection substrate, with hexynyl or azidohexyl chains (**Figure 3-34**). These two substrates were also accepted, at an even higher rate and efficiency than N⁶-biotin-TenVP by all three ribozymes (**Figure 3-35 A & C**). This observation establishes these ribozymes as valuable tools for indirect labeling of RNA using CuAAC or SPAAC techniques.

We also tested the potential of FJ1 and FJC9 for direct coupling of fluorophore-conjugated tenofovir to target RNA. We synthesized 5-FAM and sulfoCy5 conjugated tenofovir-diphosphate analogs (**Figure 3-34**), through CuAAC reaction, using the azido functionalized dyes and N⁶-hexynyl-TenDP. These substrates however, showed poor reactivity towards either of the ribozymes. The reaction rates were extremely slow but the final yields after 22 hours were between 40%-60% depending on the ribozyme and dye-conjugated substrate (**Figure 3-35 B & C**).

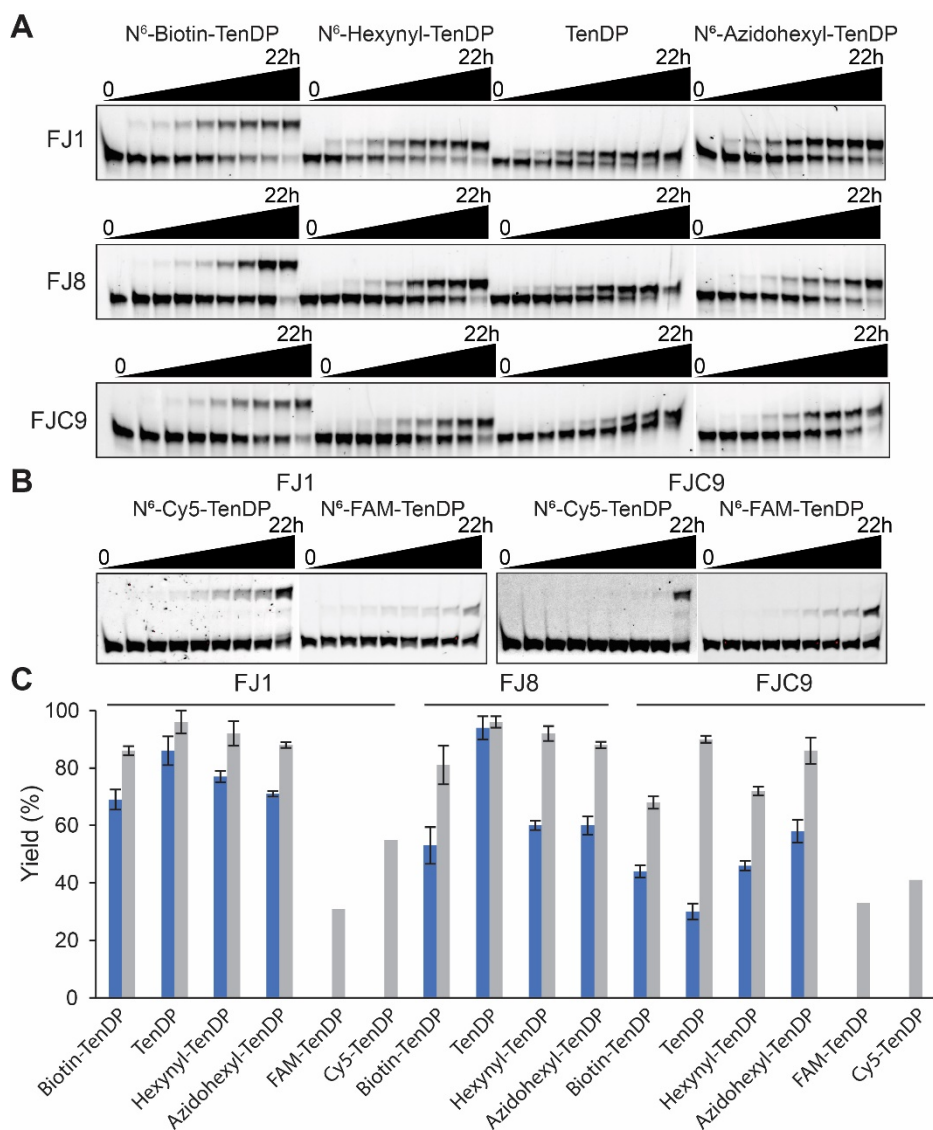


Figure 3-35 Tenofovir-diphosphate analog specificity of FJ ribozymes. **(A)** Gel images of the kinetic experiments performed on FJ ribozymes using TenDP analogs with and without N⁶-modification. **(B)** Gel images of the kinetic experiments on FJ1 and FJC9, using N⁶- Cy5 and FAM conjugated TenDP. **(C)** Reaction yields of the ribozymes with different TenDP analogs after 8 hours (blue) and 22 hours (grey). Reaction conditions in all cases: 3'-Fluorescein labeled substrate sequence and ribozyme at 1:10 ratio. Selection buffer, 40 mM MgCl₂, 300 μM TenDP analog, 37°C. time points: 0, 0.25, 0.5, 1, 2, 4, 6, 8 and 22h.

The selection substrate and all the modified variants were synthesized as racemic mixtures while the commercially provided TenDP was the pure R-enantiomer. Therefore, we speculated that higher reactivity of the FJ1 and FJ8 towards unmodified TenDP could be due to the stereospecificity of the ribozyme. To test this hypothesis, unmodified TenDP was also synthesized as a racemic mixture and tested along with the R-TenDP (**Figure 3-36 A**). Both reactions showed similar rate and efficiency (**Figure 3-36 B**), thus excluding the preference for one of the enantiomers.

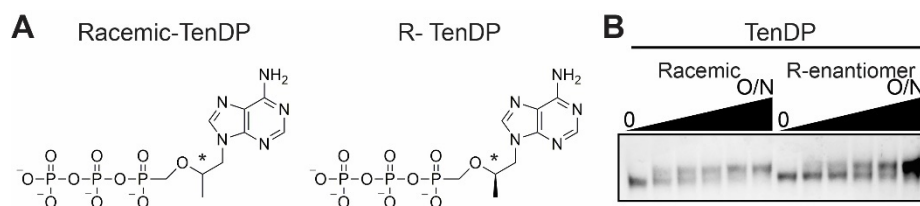


Figure 3-36 Racemic TenDP vs R-TenDP and their reactivity towards FJ1. **(A)** Asterisk indicates the steric center. **(B)** Both substrates show a similar rate and efficiency.

3.3.7.3 Orthogonality test of FH and FJ ribozymes

To test the possibility of cross-reaction between the FH and FJ ribozymes, their ability for utilizing one another's substrates were put to test. The FH ribozymes were tested for the tenofovir transferase activity. Using N⁶-biotin-TenDP. FH14 showed no reactivity towards this substrate even at 800 μM concentration. FH31, on the other hand revealed residual ligating activity towards N⁶-TenDP (**Figure 3-37 A**). When FJ1 and FJC9 were tested for reactivity towards N⁶-biotin-ATP, no product formation could be detected even after overnight incubation (**Figure 3-37 B**). The FJ ribozymes and the FH14 ribozyme can thus be considered as orthogonal.

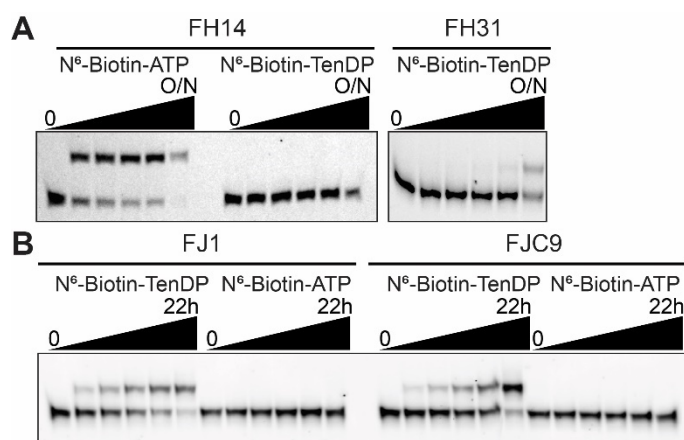


Figure 3-37 Orthogonality test of FH and FJ ribozymes. **(A)** Reactivity of FH ribozymes towards N⁶-TenDP. **(B)** Orthogonality of FJ1 and FJC9 towards N⁶-modified ATP analogs.

3.3.7.4 Dependency of FJ1 on Mg²⁺

To determine the range of Mg²⁺ concentration required for efficient FJ1-mediated labeling, we set up single-turnover kinetic experiments under different MgCl₂ concentrations. Experiments were set up using 3'-fluorescein-parent substrate sequence and N⁶-Biotin-TenDP. Timepoint samples were taken after 8, 24 and 36 hours of incubation at 37°C. PAGE resolution followed by fluorescent imaging of the gel revealed the FJ1 shows no activity at

1 mM MgCl₂ concentration and severely reduced activity at 5 mM concentration of this salt. The reaction yield was slightly reduced at 20 mM MgCl₂ compared to 40 mM. The reaction yield at 10 mM was drastically reduced, however, nearly half of the parent substrate sequence was modified within 24 hours (**Figure 3-38**).

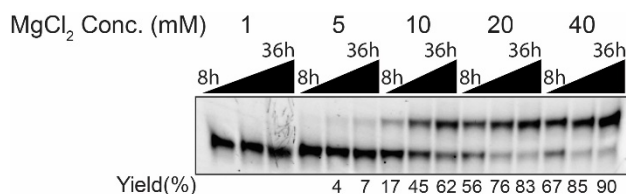


Figure 3-38 Mg²⁺ dependency of FJ1. Reaction conditions: 1 μM 3'-fluorescein-parent substrate sequence, 10 μM ribozyme, 300 μM N⁶-TenDP-biotin. 37°C. timepoints: 8, 24 and 36 hours.

3.3.8 Selection of tenofovir transferase ribozymes using an N25-pool

Another selection experiment was performed using a pool with similar design as the N40-pool, but with a 25-nucleotide random region. The details of this selection are summarized in (**Table 3-11**).

Table 3-11 Details of the selection rounds of the N25 tenofovir transferase ribozyme selection.

Selection round	1	2	3	4	5	6	7	8	9	10
Pool concentration (μm)	~50									
Incubation time (h)	Overnight							4	3	2
Eluted %	-	-	-	-	0.3	1.6	1.4	1.5	1.7	0.72

3.3.8.1 Sanger sequencing of the selection outcome

After 10 rounds of selection, the pool was subjected to cloning and sequencing (**Figure 3-39 A**). The 10 clones that were submitted to sequencing all harbored the same insert. A single variant was identified from this selection experiment. This ribozyme was denoted as FK22 (**Figure 3-39 B**).

3.3.8.2 Substrate sequence variability and tenofovir substrate scope of FK22

To test the target sequence generalizability of the FK22 ribozyme, we tested the TV1 substrate sequence for modification by the corresponding version of the ribozyme. The ribozyme successfully modified this target sequence albeit at a lower rate and efficiency compared to the original selection substrate (**Figure 3-39 C & D**). Further investigation into

this ribozyme was not performed. The exact modification-site of this ribozyme has also not been probed.

The FK22 ribozyme shows greater dependency on biotin and the aminohexyl linker as it does not accept unmodified TenDP as a substrate. TenDP analogs harboring azidohexyl and hexynyl at N⁶ of the nucleobase, also acted as poor substrates (**Figure 3-39 E**). These substrates demonstrated extremely slow reaction kinetics however, 22h yields of over 50% was achieved.

The fact that this ribozyme has such a strong dependency on biotinyl-moiety is quite unexpected. We assumed that a ribozyme with a smaller catalytic core might provide fewer contact points with the substrate. The dependency of FK22 on the selection tag, however, proves that this assumption may not be necessarily true.

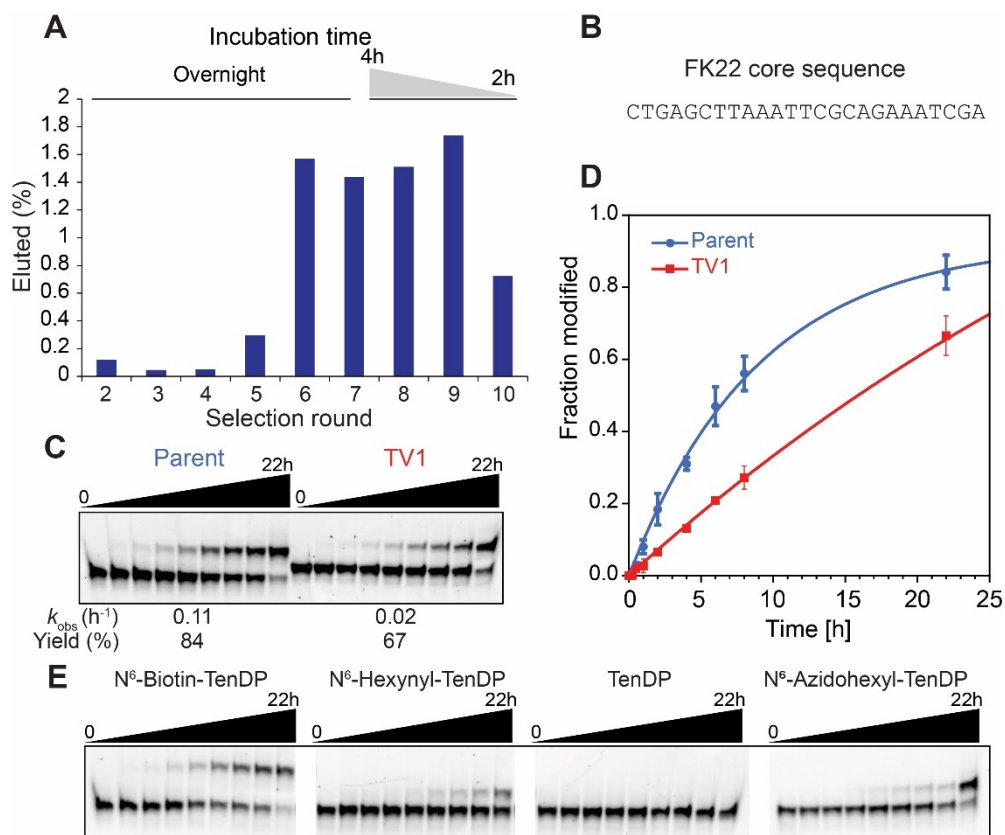


Figure 3-39 Selecting a tenofovir transferase ribozyme using an N25 pool. **(A)** Selection progress monitored using fluorescence measurement of the eluted RNA after each round. **(B)** Sequence of the catalytic core of the single variant identified in the selection of the N25 tenofovir transferase ribozyme. **(C)** Gel image and **(D)** Kinetic plot of the single turnover kinetic assay performed using parent and TV1 substrate sequences and their corresponding FK22 ribozyme. **(E)** Gel image of the single turnover kinetic experiments performed on FK22 using TenDP analogs with and without N⁶-modification. Reaction conditions in all cases: 3'-Fluorescein labeled substrate sequence and ribozyme at 1:10 ratio. Selection buffer, 40 mM MgCl₂, 300 μM TenDP analog, 37°C. Time points: 0, 0.25, 0.5, 1, 2, 4, 6, 8 and 22h.

3.3.9 Application of the tenofovir transferase ribozymes

3.3.9.1 FJC9 mediated labeling of 5S rRNA

Position G98, in 5S rRNA was targeted for labeling (**Figure 3-40 A**), using FJC9 and 5FAM-TenDP. The reactions were performed analogously to 5S rRNA labeling using the FH14 ribozyme. 5S rRNA was effectively recognized and labeled by FJC9 in total cellular RNA (**Figure 3-40 B**). While the reaction rate of this ribozyme with 5FAM-TenDP was shown to be slow (**Figure 3-35 B & C**), a 5-hour incubation time was enough for 5S rRNA to be labeled and visualized on PAGE. 3'-fluorescein-labeled 5S rRNA in vitro transcript was used as size marker.

To provide additional proof for the specificity of this ribozyme for the modified position, we performed reverse transcriptase-based primer extension assay. These experiments are based on the fact that certain modified positions stall reverse transcriptase during cDNA synthesis, resulting in abortive products. These products can be detected using end-labeled primers for reverse transcription followed by resolution of the product on denaturing analytical PAGE. A 5'-³²P-labeled primer was synthesized complementary to the last 14 nucleotides of the 5S rRNA. This primer was annealed to the 5S rRNA in vitro transcript, modified either using FH14 at each individual FH14 target site or FJC9 and subjected to primer extension using superscript III reverse transcriptase. Unmodified 5S rRNA was used as control. Abortive bands were observed in the case of the modified 5S rRNA lanes when the primer extension product was resolved on 15% analytical PAGE and subjected to phosphorus imaging. To determine the exact positions at which the reverse transcriptase is stalled, RNA sequencing reactions were performed based on dideoxynucleotide chain-termination. Since the sequence of 5S rRNA was known, we only performed sequencing using ddCTP and ddTTP to determine the G and A positions, respectively. The product of the sequencing reaction was also resolved next to the samples from the primer extension performed on the modified and non-modified 5S rRNA. The abortive bands appeared for positions modified by FH14 at exactly one nucleotide 3' to the modification site, which indicates the presence of a steric block, at the modified positions. Since FJC9 modified G98 rather than A99, the abortive band therefore appears at A99 (**Figure 3-40 C**).

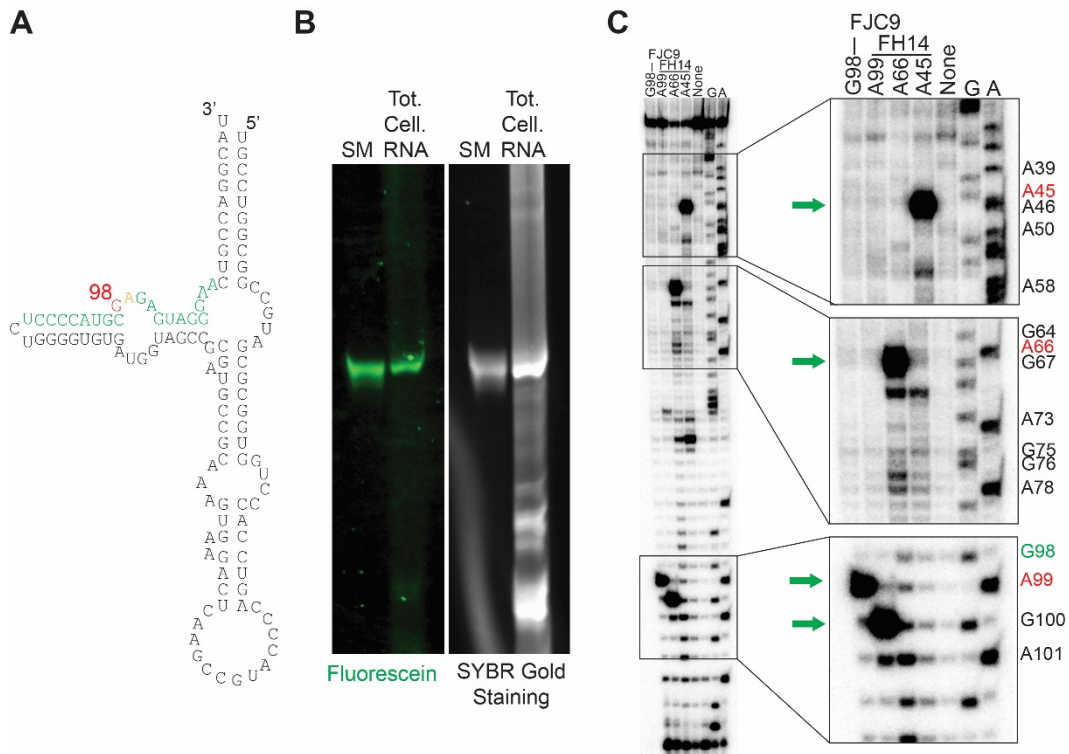


Figure 3-40 FJC9 mediated labeling of the *E. coli* 5S rRNA. (A) Sequence and secondary structure of the *E. coli* 5S rRNA. The position highlighted in red indicates the modification site (G98), the bulging A is shown in orange. The green positions including G98 were paired with ribozyme arms. (B) Left: gel image of the labeling of 5S rRNA in total cellular RNA using FJC9 and N⁶-FAM-TenDP. Conditions: 50 ng cellular RNA, 100 pmol of the corresponding FJC9 in selection buffer including 40 mM MgCl₂ and 200 μM N⁶-FAM-TenDP. 37°C, 6 hours. Right: SYBR Gold staining of the same gel. (C) Reverse transcriptase primer extension experiments on in vitro transcribed 5S rRNA modified using FJC9 and FH14 ribozymes. Biotinylated TenDP and ATP were used for FJC9 and FH14-mediated 5S rRNA modification, respectively. ³²P-labeled primer was used for this reaction. The reaction was performed on non-modified and modified 5S rRNA using each ribozyme. A and G sequence ladders were made using primer extension in the presence of ddTTP and ddCTP, respectively.

3.3.9.2 Dual-color orthogonal labeling of synthetic and cellular RNA

3.3.9.2.1 Mutually orthogonal double labeling of an in vitro transcribed model substrate

Ribozymes FH14 and FJ1 were established as perfectly orthogonal with no cross-reactivity to one another's substrates (**Figure 3-37**). These ribozymes, therefore, hold the potential for simultaneous dual-color, mutually exclusive labeling of a single RNA molecule at different targeted positions. To put this potential to test, we designed a synthetic transcript of 41 nucleotides with two potential sites for FJ1 and FH14 mediated labeling. The transcript was the result of the fusion of the parent substrate sequence via a 4-nucleotide linker to an 18-nucleotide sequence containing a GAG modification context. The segment corresponding to the parent sequence was targeted using FJ1 ribozyme while the other segment was targeted

using an FH14 ribozyme with appropriate binding arms (**Figure 3-41 A**). Three reactions were set up using this transcript with both ribozymes present in the reaction. The reaction mixtures were incubated in the presence of either Cy5-TenDP or 5FAM-ATP individually, or both at the same time (**Figure 3-41 A**). Aliquots were taken from time-points 0, 5h, and after overnight incubations. The samples were analyzed using a 15% analytical denaturing PAGE and multichannel fluorescent imaging. As expected, only one labeled band was observed in time points 5h and overnight, the reaction in which only one labeled NTP analog was included. The emitted signal corresponded to the dye that was conjugated to the NTP analog in each reaction. Cy5-tenofovir branched and 5FAM-AMP branched RNA demonstrated shifted electrophoretic mobility due to the difference in mass and charge of the added moieties. Three labeled bands were visible in the lanes pertaining to the 5h and overnight samples of the reaction, in which both analogs were included. The lightest band emitted FAM signal, the slightly heavier band emitted Cy5 signal, and for the heaviest band both signals overlapped. These results demonstrate that the synthetic transcript was successfully labeled with both dyes using both ribozymes (**Figure 3-41 C**).

To confirm the positional specificity of the labeling reaction, we utilized the RNA cleaving 8-17NG deoxyribozyme (Schlosser et al., 2008). The deoxyribozyme was designed to cleave the model substrate between the nucleotides 24 and 25. The uneven cleavage results in two fragments with the larger fragment (24 nt) harboring the FJ1 modification site and the smaller fragment (17nt) containing the FH14 modification site (**Figure 3-41 B**). Cleaving the singly and doubly labeled model substrate with this 8-17NG and PAGE analysis of the cleavage products, confirmed the mutually orthogonal nature of the labeling reaction. The larger fragments from the reactions in which Cy5-TenDP had been included, always emitted with Cy5 signal and the smaller fragments from the reactions, in which only 5FAM-ATP had been included, always emitted with FAM signal (**Figure 3-41 C**).

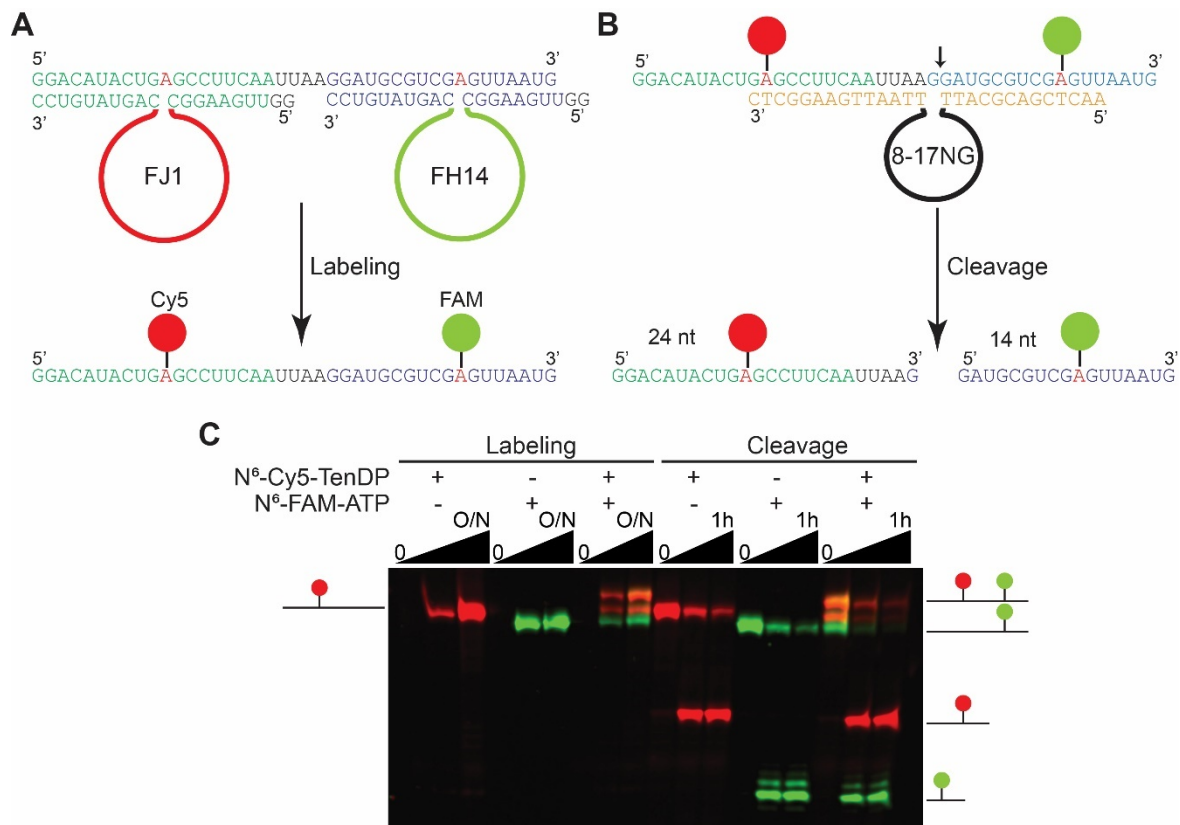


Figure 3-41 Mutually orthogonal double labeling of a synthetic transcript. **(A)** Double labeling of the synthetic transcript using FJ1 and FH14 ribozymes. Segment recognized and modified by FJ1 is highlighted in green and the segment targeted by FH14 is shown in blue. Modified As are shown in red. **(B)** Asymmetric 8-17NG deoxyribozyme cleavage of the double-labeled target sequence. **(C)** Multichannel fluorescent gel image of the double and single labeling reactions and the cleavage of the labeled products. Labeling reaction conditions: target sequence (4 μM), corresponding FJ1 and FH14 ribozymes (10 μM each) were dissolved in selection buffer including 200 μM of N^6 -Cy5-TenDP or N^6 -FAM-ATP or both, and incubated at 37°C. Time-points taken at 0, 5 hours, and after overnight incubation. Cleavage conditions: 20 μM 8-17NG deoxyribozyme, 50 mM HEPES (pH=7.5), 400 mM KCl, 100 mM NaCl, 10 mM MnCl_2 , 10 mM MgCl_2 at 37°C. Time points taken at 0, 0.5, and 1 hour.

3.3.9.2.2 Dual-color simultaneous labeling of large cellular RNA

We established the high specificity and efficiency of FH14 ribozyme in labeling highly structured long RNA such as *E. coli* 5S rRNA. Thus, we attempted to test labeling potential of the FH14 and FJ1 ribozymes for labeling even larger cellular RNAs such as 16S and 23S rRNA (Montpetit et al., 1998; Petrov et al., 2013) (**Figure 3-42**). Ultimately, we sought to address both ribosomal RNAs, simultaneously, using different fluorescent dyes and both ribozymes.

These two ribosomal RNAs were chosen due to their size, structural complexity, and abundance in *E. coli* cells. In fact, these RNAs are so abundant that they can be visualized as two distinct bands using standard staining methods, during agarose gel electrophoresis of

total cellular RNA (Smith et al., 1988). Two potential target sites in 16S rRNA (A325 and A412), and one in 23S rRNA (A637) were targeted using FH14 alone. Three sites within 5'-RAGCY-3' context were chosen to be addressed by FJ1. Positions A272 and A1572 also situated in 5'-RAGCY-3' context, where targeted by both ribozymes (**Figure 3-42**). ATTO550-ATP or 5FAM-ATP were used as labeling substrates for FH14 ribozyme and Cy5-TenDP for FJ1 ribozyme.

These two rRNAs were targeted by FH14 and FJ1 at chosen positions individually or simultaneously within the context of total cellular RNA. Due to the large size of these RNAs, agarose gel electrophoresis followed by multichannel fluorescent imaging was used for the analysis of the labeled products. All targeted positions were successfully labeled by each individual ribozyme using corresponding dye-conjugated substrate analog. In one of the samples, we tried simultaneous targeting of A649 from 16S rRNA and A272 from 23S rRNA, using FJ1 and FH14 ribozymes, respectively. As a result, 16S rRNA was labeled with Cy5 and the 23S rRNA was labeled by ATTO550, as revealed by agarose gel electrophoresis and multichannel fluorescent imaging (**Figure 3-43 A**).

The size difference between the two ribosomal rRNAs was enough to confirm the specificity of each ribozyme for its corresponding target sequence and cognate NTP analog. Nevertheless, an additional confirmatory test was performed. We chose to target A1572 of 23S rRNA with FJ1 for Cy5 labeling and A412 of the 16S rRNA with FH14 for 5FAM labeling. Five reactions were prepared all containing total cellular RNA. Two of the reactions contained FJ1 ribozyme alone, with either Cy5-TenDP or 5FAM-ATP. Two other reactions were set up similarly using only the FH14 ribozyme. The 5th reaction contained both ribozymes and both substrates. This experiment further approved the strong specificity of the ribozymes for their corresponding target sequence and their orthogonality relative to one another (**Figure 3-43 B**). Either 23S or 16S rRNA were labeled by their corresponding ribozymes when the correct ribozyme: dye-conjugated substrate combination was present. In the 5th reaction where both ribozymes and substrates were present, both ribosomal RNA were labeled with the dye assigned to their corresponding ribozyme.

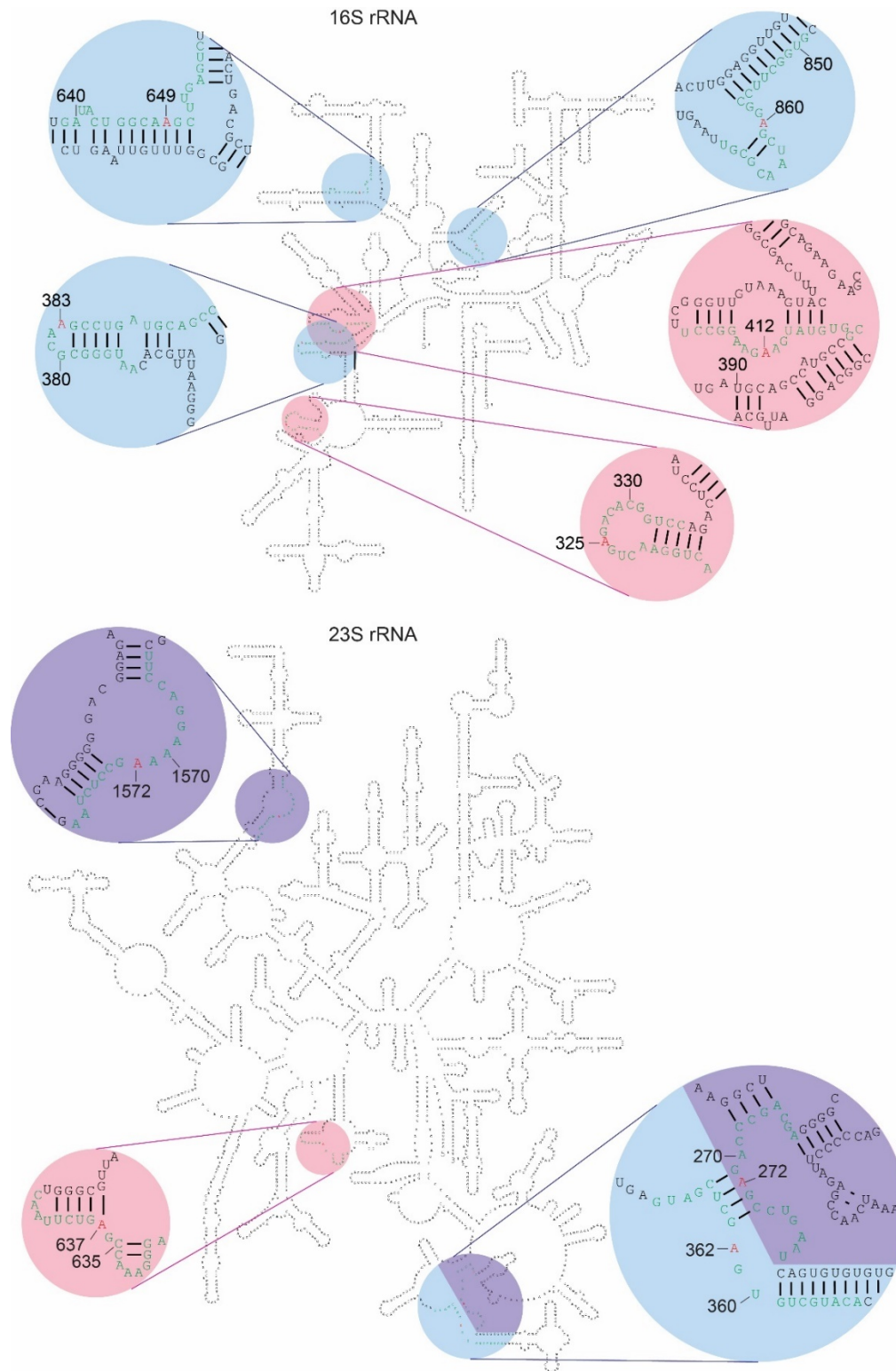


Figure 3-42 Targeted positions on *E. coli* 16S and 23S rRNAs. Positions shaded in blue were targeted to be labeled using FJ1. Positions shaded in pink were targeted for labeling using FH14. Positions shaded in purple were targeted by both ribozymes individually. Positions marked in red, correspond to the modification site. Positions marked in green were paired to ribozyme arms. 16S and 23S rRNA secondary structures were generated by “rnacentral.org” using PDB IDs 3J9Z for 16S rRNA and 4V56 for 23S rRNA.

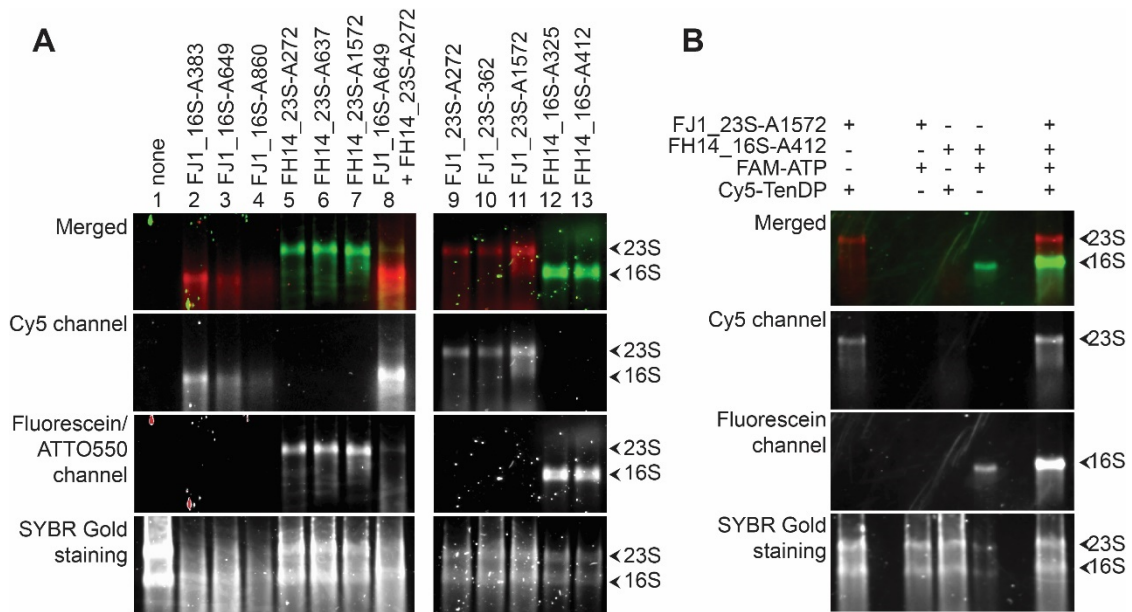


Figure 3-43 FJ1 and FH14 mediated labeling of 16S and 23S rRNA at targeted positions. **(A)** Lanes 2-4: FJ1 mediated labeling of 16S rRNA at three different positions using N⁶-Cy5-TenDP. Lanes 5-8: FH14 mediated labeling of 23S rRNA at different positions using N⁶-ATTO550-ATP. Lane 8: corresponds to the simultaneous labeling of 16S and 23S rRNA using FJ1 and FH14 respectively. N⁶-Cy5-TenDP was used as the labeling substrate for FJ1, whereas N⁶-ATTO550-ATP was used for FH14. Lanes 9-12: FJ1 mediated labeling of 23SrRNA at three different positions using N⁶-Cy5-TenDP. Lanes 12 and 13: FH14-catalyzed labeling of 16S rRNA at 2 different positions using N⁶-FAM-ATP as substrate. Lane 1 included unmodified RNA as control. The top three images were acquired using single- and multichannel fluorescent imaging of the agarose gel while the bottom gel image was acquired after SYBR Gold staining. **(B)** Orthogonality test of FJ1 and FH14. Four first lanes from the left include cellular RNA with either FJ1 or FH14 with cognate or orthogonal substrates. The last lane includes both ribozymes and both substrates. Successful labeling was only observed when matching ribozyme and substrate were present in the reaction. N⁶-Cy5-TenDP and N⁶-FAM-ATP were used as labeling substrates. The top three images were taken using multichannel fluorescent imaging while the bottom image was acquired after SYBR gold staining of the same gel.

Primer extension test using Superscript III reverse transcriptase and appropriate 5'-³²P-labeled primers were performed for some of the targeted sites in 16S or 23S rRNA, as was done in the case of 5S rRNA modified by FH14 and FJC9 ribozymes. For 16S and 23S rRNA however, total cellular RNA was used for modification reactions and primer extension assays. These tests confirmed the site-specificity of the FJ1 and FH14 labeling reaction for the targeted adenine nucleotides (**Figure 3-44**).

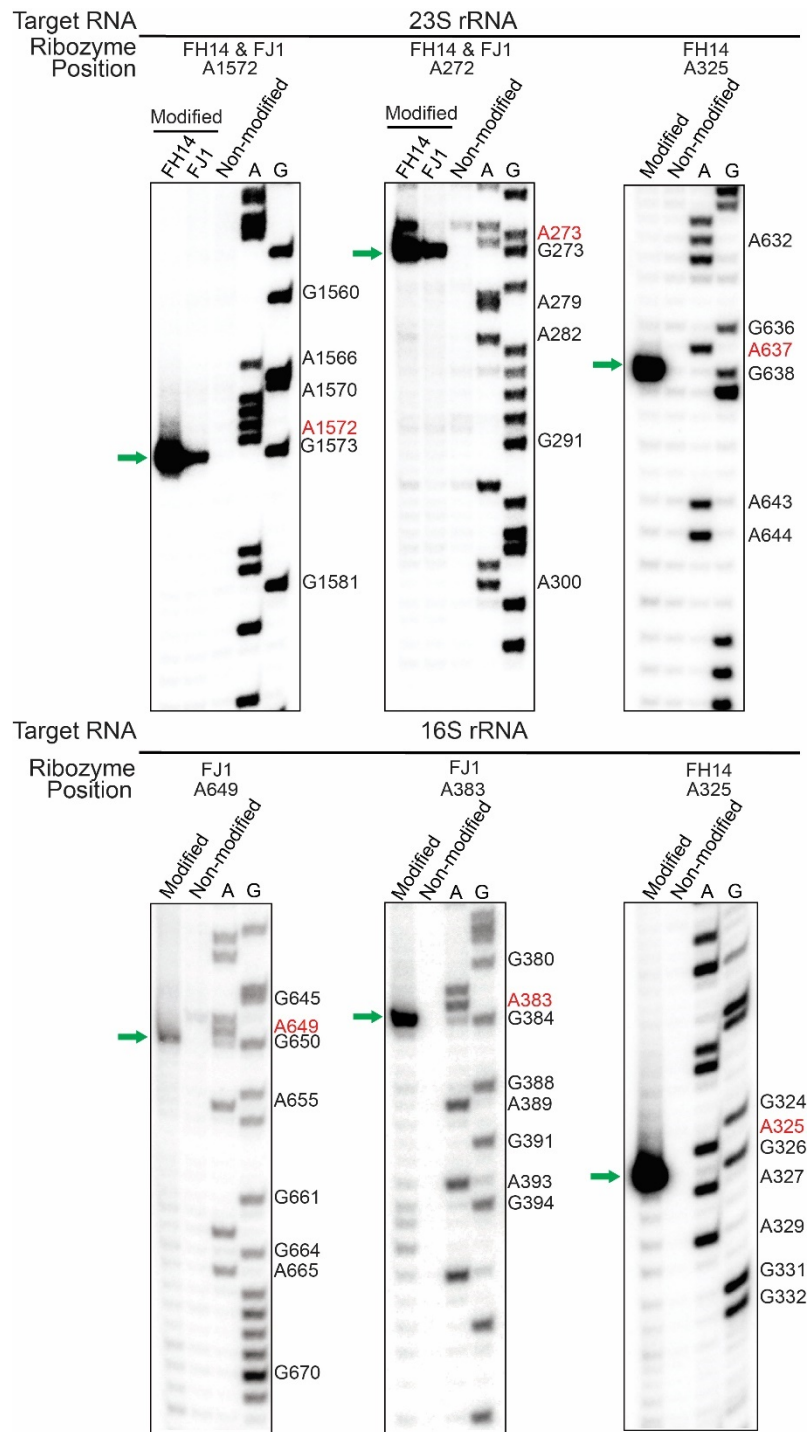


Figure 3-44 Reverse transcriptase primer extension assay on FJ1 or FH14 modified, 16S, and 23S rRNA. The 16S and 23S rRNA subjected to FJ1 and FH14-mediated modification in the context of total cellular RNA. N⁶-Biotin-TenDP and N⁶-biotin-ATP were used as modification substrates for FJ1 and FH14, respectively. 5'-³²P labeled primers, binding up to 40 nucleotides downstream of the modification site were used for the extension assay. A and G sequencing ladders were also included to determine the exact position of the modification site. The abortive bands appeared in every case at exactly one nucleotide 5'-to the modified position confirming site-specificity of FJ1 and FH14 ribozymes. These bands were missing in the lanes pertaining to the none-modified RNA samples.

4 Discussion

4.1 Selection of trans-acting ribozymes for site-specific RNA modification

4.1.1 Analogies with deoxyribozyme selection methods

Our partially structured pool and selection strategy enabled the selection of efficient ribozymes for site-specific modification of target RNA in trans. The modification occurs in most cases at a predetermined position. The pool was designed to include a hypothetical substrate sequence, which was paired to left and right recognition arms, leaving one bulged nucleotide as the desired modification site. The hypothetical substrate sequence was physically connected to the pool via a 14-nucleotide connecting loop. The ribozymes were therefore selected essentially as “self-biotinylating” catalysts. Removal of the connecting loop, however, was enough to convert these ribozymes into efficient trans-acting tools.

The design of our selection library was a simplified variation of the 10DM24 deoxyribozyme selection pool. The design of the 10DM24 selection library was rather complex with four binding arms, recognizing two RNA substrates. Pairing of the arms to the substrate sequences led to formation of a three-helix-junction (Zelin et al., 2006). The deoxyribozyme was selected for 2'-5' ligation of two RNA oligonucleotides and was further engineered for 2'-5' branching of an RNA target using a single nucleotide (Höbartner and Silverman, 2007). The engineering process was done by removing the 5'-triphosphorylated guanine nucleotide from the 5'-end of the right hand oligonucleotide substrate and supplying it as a free GTP. The deoxyribozyme recognizes the GTP analog through Watson-Crick base-pairing to a cytosine in one of its binding arms (Höbartner and Silverman, 2007). By mutating the cytosine in this position, the specificity of 10DM24 can be switched towards other complementary NTPs (Höbartner and Silverman, 2007).

Our selection was performed directly for the single nucleotide branching activity using natural and non-natural ATP analogs. Hence, a “preorganized three-helix junction” setting was not necessary. The number of binding arms was therefore reduced to two recognizing a single substrate RNA. Through this selection strategy, ribozymes have been evolved that could be easily engineered for targeting various RNA sequences. The mode of NTP analog recognition of these ribozymes, however, may be more complex than 10DM24. In-depth investigation should thus be performed to elucidate the mechanism behind substrate specificity of these ribozymes, in order to engineer NTP analog specificity.

Analogies can also be made between our selection strategy and other deoxyribozyme selection methods. The constant regions of the deoxyribozyme selection pools are in most cases designed as binding arms complementary to the substrate sequence(s). Deoxyribozymes acting on non-oligonucleotide substrates have also been selected following similar design principles. In these instances of DNA catalysts, oligonucleotides complementary to the constant regions of the pool are conjugated to the reaction substrates. This pairing facilitates the localization of reactants in the vicinity of the random region, improving the likelihood of the isolation of the desired catalyst. The oligonucleotide (tethered) substrate is required to be ligated to the single-stranded DNA pool, before each selection round, as it is lost during the PCR amplification cycles. The evolved variants are also generally compatible with activity in trans when the target oligonucleotide or the oligonucleotide tethered substrate is not covalently attached to the catalyst (Silverman, 2009).

For selection of RNA catalysts acting on unmodified RNA substrates the physical linkage between the substrate sequence and the pool was made transcriptionally. The extra ligation step, necessary in case of 10DM24 deoxyribozyme was therefore not needed.

4.1.2 Basis of the modification site selectivity of the evolved ribozymes

Out of four ribozymes probed for modification site determination, three modified the predetermined adenosine. The other ribozyme labeled the target sequence at one nucleotide 5' to the bulged A. This proves that our selection strategy facilitates rapid evolution of ribozymes with the ability to modify this particular position. Two reasons may contribute to the site-selectivity of our selection process:

First, adenine nucleotides in a bulged structural context have been shown to have higher tendency to participate in branching reactions compared to other nucleotides. This conclusion was made by Silverman et al based on their experience of selecting 2'-5' ligase deoxyribozymes. The 7S11 deoxyribozyme developed by this group is the earliest example of 2'-5' RNA ligase deoxyribozymes (Coppins and Silverman, 2004a). The selection process of 7S11 was initially aimed at development of a linear 3'-5' ligase deoxyribozyme. 7S11, however, evolved in a way that it formed a triple-helix junction surrounding a bulged adenine nucleotide that served as the branch-site (Coppins and Silverman, 2005). This structural context was the inspiration for the selection of the 10DM24 ribozyme. The deoxyribozyme was initially selected for branching a bulged uridine, however, replacement of the uridine

with an adenine remarkably improved the ligation rate. These findings along with strong prevalence of bulged A-nucleotides at branch-sites of class II and spliceosomal introns, led to the conclusion that the bulged adenines are “inherently favored” for 2'-5' branching reactions. This conclusion was the basis of our choice for the modification-site adenine nucleotide (Zelin et al., 2006). Furthermore, Büttner et al discovered that the 5'-GAG-3' is the optimal modification context for 10DM24 (Büttner et al., 2014). For that reason, the bulged A for our ribozyme selection was also placed within a GAG context.

Additionally, the desired reactions for which the ribozymes were selected, involve the introduction of bulky groups to the ribose backbone of the RNA. These types of modifications are known to stall reverse transcriptase. Thus, the variants that self-modify at positions outside the priming range (i.e., in the random region) are eliminated from the pool due to failure in full-length cDNA synthesis.

4.1.3 Comparison with other trans-acting RNA modifying ribozymes

To our knowledge, no 2'-5' nucleotidyltransferase ribozymes had been developed prior to our work. However, several other artificial RNA-modifying ribozymes exist that possess trans-activity (Jadhav and Yarus, 2002; Kang and Suga, 2007; Lorsch and Szostak, 1995; Poudyal et al., 2017; Saran et al., 2005). These ribozymes were mostly selected using unstructured RNA libraries. Except for some 5'- or 3'- modifying ribozymes, no predetermined modification site was considered during the selection process of most of these ribozymes. The modification sites were, therefore, randomly evolved. To turn these ribozymes into trans-acting variants, an in-depth investigation had to be performed to probe the modified position and predict or determine the secondary structure of the ribozyme. Mutational studies are then often performed to minimize the ribozyme and determine the critical positions often resulting in loss of efficiency. The ribozyme is then split into the substrate segment and the catalyst segments. In some cases, the substrate and the catalyst segment interact via partially complementary regions. The complementary parts can sometimes be covaried to allow flexibility regarding the target sequence, however, in many cases these ribozymes have a narrow range of substrate sequences they can modify. The required modification context in those instances is complex and with a low probability of appearing in any given RNA sequence. For example, the kinase ribozyme 2PTmin3.2 is a trans-acting ribozyme that recognizes its target sequence via partial complementarity. This

ribozyme phosphorylates the 2'-OH of the first purine nucleotide within a 5'-**R**AAAANCG-3' context (Saran et al., 2005).

Our selection strategy using a partially structured pool allows the selection of ribozymes that can readily be adopted for trans-activity simply by removing the connecting loop and varying the binding arms based on the desired target. These changes do not cause any loss of efficiency. Moreover, the variants selected mostly demonstrated remarkable target sequence variability.

4.2 Properties of FH ribozymes

The adenylyl transferase ribozymes identified in this thesis project were denoted as FH ribozymes. These ribozymes were selected for site-specific RNA labeling and as RNA analogs of 10DM24 deoxyribozyme. The selection process involved N⁶-Biotin-ATP as the selection substrate and led to the identification of three major sequence groups. These groups were named based on their representative clone number as FH14, FH20, and FH31.

4.2.1 Sequence analysis and kinetic comparison of FH ribozymes

Upon sequence alignment, a remarkable degree of similarity was observed between FH14 and FH20. FH31 however had very few sequence signatures with similarity to FH14 and FH20, particularly in 5'-GAYA-3' motif that was shared among all variants. The functional significance of this motif is still unknown and may become apparent through in-depth mutational and structural studies.

Single-turnover kinetic experiments of the trans-acting versions of these variants revealed the FH14 as the most efficient of the three, followed by FH31. FH20, on the other hand, was the least efficient variant in these experiments, despite its high activity level in cis and the strong sequence resemblance to FH14. The reasons behind this disparity is not yet understood. Structural and mutational comparison of FH14 and FH20 may facilitate minimization of FH14 through the identification of the critical and non-critical positions. Furthermore, the data acquired from these studies may provide further insight into the factors and structural features that contribute to the efficient trans-activity of the ribozyme. Additionally, modern techniques based on high-throughput sequence analysis allow simultaneous testing of a large number of mutants. Nokumura and Yokobayashi, have

recently utilized these techniques to determine critical positions in R3C ligase (Nomura and Yokobayashi, 2019). Their work led to minimization of the catalytic core of this ribozyme to only 18 nucleotides. Such techniques can also be utilized to minimize ribozymes evolved during this thesis project.

4.2.2 Potential of the FH ribozymes as RNA labeling tools

Tools that are applied for site-specific post-transcriptional labeling of RNA at internal positions should possess two main features:

First, the tool should be modular and easily targeted towards various RNA sequences. In other words, it should possess the capability to be engineered for targeting RNA sequences of interest through simple means. While the tool can be flexibly designed for various targets, it should maintain high specificity towards targeted sequences. Secondly, the tool should provide the possibility for attachment of a wide range of chemical probes, and bioorthogonal functionalities, to the target RNA.

The FH14 and FH31 ribozymes demonstrate these features to a large extent. We first established that these ribozymes are trans-active and modify the hypothetical substrate sequence at the expected position. We then proceeded by derivatizing the binding arms of these ribozymes to target extensively mutated variants of the substrate sequence used in selection. All the sequences were efficiently modified by these ribozymes. The FH14 ribozyme even revealed enhanced modification rate and efficiency for these mutated target sequences, compared to the selection substrate it was evolved to modify. FH14 was able to modify all three mutated substrate sequences (TM, TV1, and TV2) with over 70% yield, in less than 15 minutes.

Investigation into the modification context of FH14 and FH31 ribozymes, revealed their modification-site sequence context as a single A nucleotide for FH14 ribozyme and 5'-AG-3' for FH31. Assuming a random distribution of nucleotides in an RNA of interest, the probability of finding a potential modification-site is on average 1 in every 4 nucleotides for FH14, and 1 in every 16 nucleotides for FH31 ribozyme. In reality however, targeting the potential labeling positions may not be so simple. The desired position for example, can be situated in a stable secondary or tertiary structure or bound by proteins. In those cases, the modification sites may not be accessible to the ribozyme and therefore not efficiently labeled. Both FH14 and FH31 ribozymes require simple modification contexts. A large

number of potential FH14/31 labeling positions can therefore be found in most RNAs of interest. Thus, it may not be particularly challenging to find easily accessible labeling positions for these ribozymes.

The FH ribozymes, and specially FH14, demonstrated remarkable flexibility regarding the diversity of ATP analogs they could accept. Our initial concern regarding the use of biotinylated ATP substrate was the selection of variants with biotin as an essential substrate requirement. Reports of artificially developed biotin-binding functional nucleic acids exist in literature. Biotin binding aptamer and the self-alkylating biotin-ligase ribozymes are two examples of such sequences (Starikov and Nilsson, 2002; Wilson and Szostak, 1995). Fluorogenic Mango aptamer, selected using TO1-biotin, also showed dependency on biotin for optimal binding and fluorescence activation (Dolgosheina et al., 2014). These studies provide evidence that RNA can be evolved to recognize biotin. It was therefore likely that some of our selected variants might bind to the selection substrate, via the biotinyl moiety. Fortuitously, this was not the case for the FH ribozymes. ATP analogs of the selection substrate, in which biotin was removed or replaced with fluorophores were efficiently ligated to the substrate sequence by FH14 ribozyme.

Further investigation into ATP analog specificity of the FH14 ribozyme revealed that the presence of the N⁶-substituent of the ATP analog is essential for ribozyme binding. This conclusion was reached when we observed that the GTP is not accepted by the ribozyme, while ATP, 2F-ATP or 2-aminoATP can still be ligated. The reaction rate and efficiency with these ATP analogs, however, were severely reduced compared to some ATP analogs modified at N⁶. Analogs derivatized at N⁶ with linear aminoalkyl chains of 4-6 carbons, were ligated efficiently with final yields of 50-80%. Azidohexyl or propargyl group substitution at this position however, reduced the reaction rate and efficiency drastically. Nevertheless, other forms of azido or alkynyl groups, such as the types conjugated via an amide bond to the aminohexyl-ATP analog, may be ligated more efficiently (**Figure 4-1 A**).

Another interesting observation was the relatively high yield of FH14 ribozyme mediated labeling when N⁶-benzyl-ATP was used. Thus, the ribozyme may also be able to accept bioorthogonal functionalities similar to the benzyl group, such as styrene or tetrazine (**Figure 4-1 B**). Both types of modifications can act as bioorthogonal handles for IEDDA reactions.

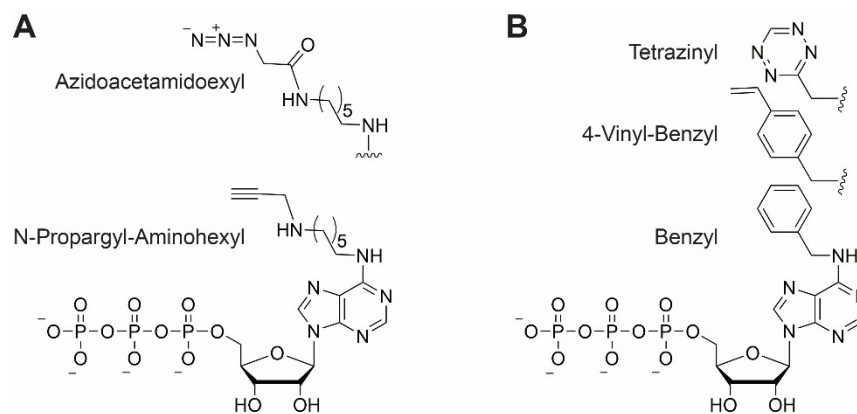


Figure 4-1 Potential FH14 mediated labeling substrates with bioorthogonal handles. **(A)** Azido and alkyne containing analogs **(B)** Bioorthogonal functionalities with similarity to benzyl substitution.

4.2.3 FH14-mediated labeling of long heavily structured RNA

Reliance of the ribozymes on Watson-Crick base-pairing for target recognition may be a disadvantage when it comes to labeling of heavily structured RNA. This could be due to the competition between the ribozyme arms and the internal sequences of the substrate RNA for pairing to the target site. The potential of FH14 ribozyme for labeling large and heavily structured RNAs was tested by targeting three target sites in *E. coli* 5S rRNA. 5S rRNA is significantly larger than the synthetic model substrate sequences. The second reason was the strongly structured nature of this RNA. The 5S rRNA targeting FH14 variants successfully labeled all three target positions individually and simultaneously on in vitro transcribed pure 5S rRNA. While it was impossible to determine the yield of the labeling reaction, the intensity of the labeled band signal after 4 hours of reaction was roughly similar across all labeled positions. FH14 also successfully labeled all three positions of 5S rRNA within the context of cellular RNA.

The results were reproducible when another class of cellular RNA was labeled in 2 different positions using the FH14 ribozyme. RyhB was a second RNA of cellular origin that was successfully labeled using FH14 as a pure transcript. Labeling within the total RNA cellular context of RyhB resulted in significantly lower intensity bands compared to 5S rRNA. The reason behind this difference could be attributed to the instability of the RyhB RNA in cell under the growth conditions used.

FH14 was also successfully applied to the labeling of even larger cellular RNA such as *E. coli* 16S and 23S rRNA, at multiple positions. The site-specificity was confirmed by both

fluorescent imaging of the agarose gel and using reverse transcriptase primer extension assay.

4.2.4 FH ribozymes and perspective for RNA labeling in live cells

FH14 ribozyme demonstrated immense potential for RNA labeling *in vitro*. The ribozyme shows excellent substrate sequence variability and a broad range of specificity towards ATP analogs. Various synthetic and cellular RNAs could be efficiently labeled using this ribozyme, even the positions we considered difficult to access. Successful labeling of 5S, 16S and 23S rRNAs and the RyhB sRNA in total cell RNA confirms the potential of this ribozyme for application in live cell. Nonetheless, some limitations make the cellular application of this ribozyme impossible, at least in its current form.

First and foremost, the nature of the substrate utilized for RNA labeling is not entirely orthogonal to the cellular RNA synthesis machinery. The N⁶-modified ATP analogs are expected to be incorporated into actively transcribed RNA sequences through the action of the cellular RNA polymerases (Grammel et al., 2012). Poly-A polymerase is another cellular enzyme that can use N⁶-modified ATP analogs for extending the poly-A tail of the polyadenylated cellular RNA (Grammel et al., 2012). Both these events would result in strong non-specific background emission, making the signal emitted from the desired target indistinguishable.

Another inherent limitation of the labeling reaction is the formation of a 2'-5'-phosphodiester branch. Such structures are recognized by cellular debranching enzymes and rapidly removed, leading to loss of label (Chapman and Boeke, 1991). This has been demonstrated in case of the 10DM24 labeling products in yeast cell lysate (Carrocci et al., 2017).

A third limitation may be the cellular entry of the N⁶-modified ATP analogs. The triphosphate forms of these analogs are not cell-permeable, due to the high density of negative charges. One way to solve this problem is to express NTP transporters in the cell. These transporters have been reported to facilitate entry of unnatural bioorthogonal NTP analogs into bacterial cells (Feldman et al., 2018). Examples of NTPs attached to bulky fluorophores, however, have not been reported. An alternative way is introducing these NTPs in the nucleoside form in the culture media. Treatment of mammalian cells with N⁶-propargyl-adenosine, for example, facilitated RNA polymerase and poly-A polymerase

mediated RNA labeling (Grammel et al., 2012). Bulkier modifications on the N⁶, however, may reduce cell permeability even in the form of a nucleoside.

Another major issue of the FH14 ribozyme is its dependence on high concentrations of free Mg²⁺ ions. Our Mg²⁺ dependency test revealed that the ribozyme's efficiency drops severely at Mg²⁺ concentrations below 10 mM. At 1 mM, which is the closest condition tested to the free Mg²⁺ level in living cell (Murphy, 2000; Romani, 2011), the ribozyme shows no activity. Molecular crowding effect in the cell has been shown to allow natural ribozymes to operate under significantly lower salt concentrations *in vivo* than required *in vitro* (Paudel et al., 2018). Despite this fact however, artificial nucleic acids with lower Mg²⁺-dependency are known to act more efficiently in living cells (Filonov et al., 2014; Huang et al., 2019). It is conceivable that FH14 can be reselected for reduced Mg²⁺ dependency or evolved towards usage of new bioorthogonal ATP analogs. In this thesis, we followed another route, namely the selection with bioorthogonal nucleotide analogs.

4.3 Tenofovir transferase ribozymes

4.3.1 Antiviral nucleotide analogs as substrate for ribozyme catalyzed RNA labeling

Due to the lack of bioorthogonality of the N⁶-labeled ATP substrates, we decided to explore unnatural ATP analogs as potential substrates for ribozyme catalyzed RNA labeling. Antiviral ATP analogs were chosen as potential substrates that may solve the orthogonality issue of the analogs based on natural ATP. Nucleotide analogs that act as inhibitors of the viral polymerases while demonstrating low reactivity towards cellular polymerases were considered as suitable choices. Many different classes of antiviral nucleotides exist with these properties (Poudyal et al., 2017). While studying various options as potential substrate choices, we decided that acyclic nucleotide phosphonates might present an attractive option.

Acyclic nucleotide phosphonate analogs are an unusual class of bioactive molecules that can be considered as analogs to natural nucleotide monophosphates. Tenofovir (**Figure 4-2**) is an analog of AMP that belongs to this group of analogs. It has been approved by FDA as an anti-HIV and anti-HBV drug (De Clercq and Holý, 2005). The most striking feature of tenofovir, and acyclic nucleotide phosphonates in general, is the presence of a phosphonate group instead of the α -phosphate. Due to this negatively charged phosphonate group, tenofovir in its free form, cannot permeate the cellular membrane. It is therefore

administered in more cell-permeable, lipophilic prodrug forms in which the phosphonate group has been masked. Tenofovir disoproxil (Gallant and Deresinski, 2003) and tenofovir alafenamide (Ray et al., 2016) are the two most commonly used forms of the drug wherein the phosphonate group is masked by isopropyl methyl carbonate esters or in the form of a phosphoramidate respectively. Upon cellular entry, natural cellular enzymes remove the protective groups. The free phosphonate group is then phosphorylated by cellular nucleotide kinases that convert the drug to its active diphosphate form (**Figure 4-2**). Tenofovir-diphosphate in turn, acts as an ATP analog and is incorporated by viral HIV reverse transcriptase or HBV DNA polymerase into the nascent DNA chain. Upon incorporation, tenofovir acts as an obligatory chain terminator blocking replication or reverse transcription of the viral genome (Naesens et al., 1998; van Hemert et al., 2014). Owing to its extremely low reactivity towards mammalian cellular DNA polymerases, tenofovir is essentially non-toxic to mammalian cells (Cihlar et al., 2002). The incorporation rate of tenofovir-diphosphate, into RNA via cellular RNA polymerases has not been measured. Nevertheless, due to the lack of a functional group analogous to the ribose 2'-OH, it is safe to assume that the incorporation rate is extremely low (Wang et al., 2006). Moreover, upon incorporation of tenofovir an abortive transcript is formed, which may be degraded rapidly through the RNA quality control mechanism in the cell (Kilchert and Vasiljeva, 2013).

Features such as cellular activation and orthogonality made tenofovir an attractive candidate as a substrate for ribozyme catalyzed RNA labeling *in situ*. Furthermore, polymerase-mediated incorporation of tenofovir into DNA results in formation of a phosphonomonoester linkage. Ribozyme-catalyzed branching of a target sequence using tenofovir diphosphate, results in a phosphonomonoester branch. We speculated that this branch type would be more resistant to debranching compared to the 2'-5'- phosphodiester linkage.

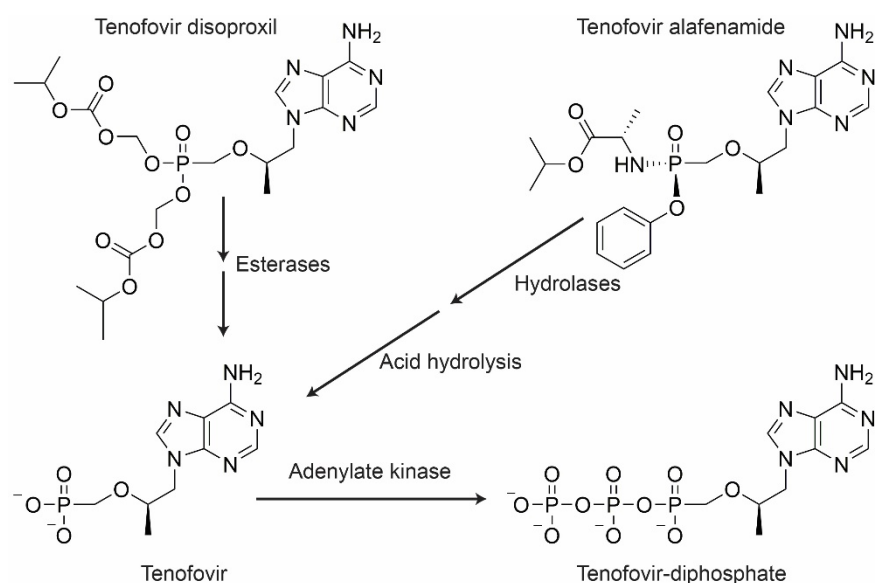


Figure 4-2 Tenofovir, an acyclic nucleotide phosphonate analog of ATP. Prodrug forms of tenofovir and their cellular activation pathway.

To develop tenofovir transferase ribozymes, a selection experiment was initiated using N⁶-Biotin-TenDP as the selection substrate. This analog of tenofovir is not commercially available and was therefore synthesized in our lab. Selection using this substrate led to the successful identification of two closely related variants through traditional Sanger sequencing, denoted as FJ1 and FJ8.

Several other species were later identified through Illumina next-generation sequencing from which FJC1, FJC3, and FJC9 were chosen for further investigation. The variants were tested in single turnover kinetic experiments. Through these tests, FJ1 was determined as the most efficient catalyst of the group. FJ8 and FJC9 demonstrated similar reaction rates while the final yield of FJ8 was higher. FJC1 and FJC3 were the least efficient variants. These two variants had the lowest log₂ values from the comparison of the sequence reads obtained from NGS analysis of the round 7 and round 12 pool samples. Log₂ offers a measure for the enrichment of ribozymes throughout the selection process (Alam et al., 2015). Log₂ values of above 2, mean a positive enrichment from an earlier round to a later selection round. Values between 0-2 mean that sequence was neither enriched nor actively eliminated during the selection rounds. The log₂ value of FJC1 was 1.95 meaning that the effect of increasing selection pressure was neutral on this variant. The log₂ value of FJC3 was -0.35 meaning a slight de-enrichment from round 7 to round 12. The low log₂ values of these ribozymes are directly translated into their lower catalytic efficiency. The other three ribozymes had log₂

values significantly higher than 2. They, therefore, benefited from the increasing selection stringency due to their faster reaction kinetics.

4.3.2 Target sequence generality of tenofovir transferase ribozymes

The ribozymes FJ1, FJ8, and FJC9 were chosen for further characterization due to their higher efficiency. The modification site of FJ1 was determined to be the bulged A nucleotide in the substrate sequence. We then tested whether FJ1 can be flexibly engineered for targeting other substrate sequences. This ribozyme was unable to modify the extensively mutated substrates (TM, TV1, and TV2), tested for the FH ribozymes. Therefore, we wondered whether this inability is due to the more complex sequence requirement of FJ1 compared to FH14 or FH31. Further experiments proved that the modification contexts of FJ1 and FJ8 were indeed more complicated than that of FH ribozymes. The minimal modification site sequence requirement of FJ1 was found to be 5'-AGC-3' with 5'-RAGCY-3' as the optimal context. Thus, the probability of finding an optimal labeling site for FJ1 in any given target RNA can be calculated as one in every 256 nucleotides. Thus, it is more challenging to find FJ1 modification sites in desired positions in an RNA of interest compared to FH14. A rather surprising observation regarding FJ1 was that the mutation of the modification site A to G, did not result in complete loss of activity although the reaction efficiency was severely diminished. Mutations of the GC dinucleotide following the A led to complete loss of labeling. It was concluded from these results that the 5'-GC-3' in the 5'-AGC-3' context, may have a more critical role than the modification site A-nucleotide. The reason behind the indispensability of these nucleotides is not yet understood.

FJC9, on the other hand, proved far less restrictive compared to FJ1, regarding its target sequences. All three mutated substrate sequences (TM, TV1, and TV2) were successfully modified using this ribozyme. Mutations to the G nucleotides flanking the presumptive modification site A, were also tolerated. Mutating 5'-G however especially to a pyrimidine did result in significant loss of activity whereas mutations of the 3'-G enhanced the reaction rate and efficiency. The labeling activity was fully lost when A in the context was mutated to G. These findings initially led us to believe that FJC9, shares the same modification site as other ribozymes probed in this thesis. Nevertheless, probing of the labeling products of this ribozyme using primer extension, RNase T1 digestion, and alkaline hydrolysis disclosed the modification site as 5'-G within the 5'-GAG-3' context. Hence, the identity of the modified nucleotide in the case of this ribozyme proved less essential than the unpaired A

that follows it. The modification site sequence context of this ribozyme was thus determined as 5'-NA-3' with a strong preference for 5'-RA-3'.

4.3.3 Tenofvir diphosphate specificity range of FJ ribozymes

All the tested FJ ribozymes demonstrated a broad range of TenDP analog specificity. No dependency on biotin or the aminohexyl linker was observed in case of FJ1 and FJ8 ribozymes. These ribozymes ligated the unmodified TenDP to the substrate sequence, at an even enhanced rate compared to N⁶-Biotin-TenDP. Similar results were observed when N⁶-azidohexyl and hexynyl TenDP analogs were used for all three FJ ribozymes. These ribozymes, therefore, demonstrate great potential for indirect RNA labeling using CuAAC and SPAAC chemistries.

Fluorophore-conjugated TenDP analogs were ligated to the target RNA at reduced rates and efficiencies by FJ1 and FJC9. The conjugates were prepared by conjugating the azido-functionalized fluorophore to the N⁶-Hexynyl-TenDP by CuAAC. As a result, the aliphatic linker length between the dye and TenDP is reduced to four nucleotides and a bulkier and more rigid triazole linkage is formed (**Figure 4-3 A**). The combination of the short, rigid linker along with the inherent bulkiness of the fluorophore may have contributed to the lower reaction efficiency due to steric hindrance. Perhaps fluorophore-TenDP conjugates prepared via different chemistries can be ligated more efficiently. For example, NHS-ester functionalized dyes can be conjugated to the N⁶-aminohexyl-TenDP (**Figure 4-3 B**). The formed product in this instance, will share greater similarity to the selection substrate as well as being more akin to the fluorophore-ATP conjugates used for FH14.

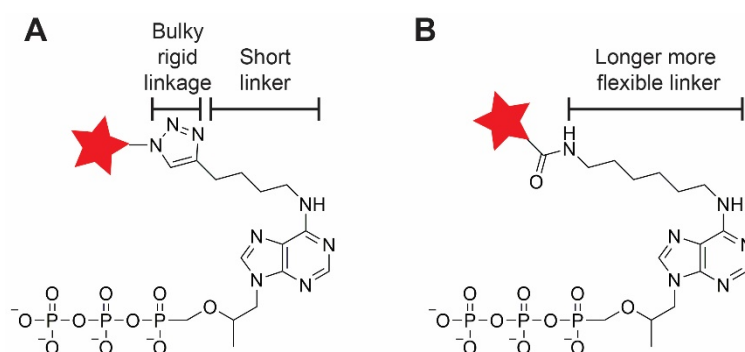


Figure 4-3 Alternative fluorophore-TenDP conjugate. (A) Conjugates tested in this thesis project. (B) Conjugates that may be ligated more efficiently by FJ ribozymes.

4.3.4 Orthogonality of FH and FJ ribozymes: potential for simultaneous dual-color RNA labeling

The ribozymes FJ1 and FJC9 demonstrated no reactivity with the FH14 ribozyme substrate N⁶-Biotin-ATP. FH14 and FH31 ribozymes were also tested for reactivity with N⁶-Biotin-TenDP. FH14 showed no activity with the TenDP analog whereas FH31 managed to ligate this substrate to the target sequence, albeit at low efficiency. Thus, FJ1 and FJC9 can be considered as perfectly orthogonal to the FH14 ribozyme but not to FH31. The orthogonality of the ribozymes allowed labeling of a synthetic RNA transcript at two different positions simultaneously, using FJ1 and FH14. Cy5-TenDP and FAM-ATP were used as labeling substrates for FJ1 and FH14, respectively. Dual-color labeling of target RNA at two different positions is significant, as it allows the installation of FRET pairs in strategic positions. Subsequently, the conformational dynamics of labeled RNA can be studied using the FRET signal. The main problem with large RNA molecules, in this case, is the lack of suitable techniques for separation of the doubly labeled RNA from the unlabeled or singly labeled products.

We have also managed to label large cellular RNA such as *E. coli* 16S or 23S rRNA, using FJ1 and FH14 ribozymes, both individually and at the same time. The results of these reactions once again asserted the orthogonality of these ribozymes as well as their high degree of specificity. Our method involved labeling of the target sequence RNA in total cellular RNA. The treated cellular RNA was then resolved on a gel and visualized via fluorescent imaging. This strategy can be compared to blotting techniques such as northern blotting. Our strategy, however, requires fewer steps, as in northern blotting the RNA is typically transferred and fixed onto a nylon membrane. Synthetic probes that are labeled either radioactively or fluorescently are then annealed to the target RNA followed by visualization on the membrane after washing off the unbound probe. In our strategy, RNA is visualized directly in the gel, without the need for additional labeled probes. For this strategy to replace northern blotting however, it is important to investigate the sensitivity of our technique. The lowest range of RNA that can be detected using this strategy needs to be determined. The smallest amount of RNA detected in northern blotting using near infra-red fluorescent dyes was determined to be around 0.05 fmol (Köhn et al., 2010; Miller et al., 2018). These systems however use multiple dyes per probe. The sensitivity of our system may also be improved by targeting the same RNA at several positions and attaching multiple labels, as demonstrated for 5S rRNA that was labeled at three sites.

4.3.5 FJ ribozymes and the path to RNA labeling in situ

The tenofovir-diphosphate analog used for selection of FJ ribozymes solves the orthogonality issue of the ATP based substrates used by FH ribozymes. The background signal produced as a result of FJ ribozyme labeling in live cell, may be lower as cellular polymerases do not incorporate tenofovir.

We have also tested resistance to debranching of the RNA samples branched using FJ1 or FH14. As expected, the phosphonomonoester linkage introduced by FJ1 was more resistant towards debranching by highly active recombinant Dbr1 compared to their phosphodiester counterparts. Although less enzymatically labile, the phosphonomonoester branch was not fully resistant towards debranching. The debranching rate of this product at Mn^{2+} concentrations above 10 μM was similar to the phosphodiester linkage. The results obtained in our debranching experiments by no means reflect the complex cellular conditions. Debranching assay in environments such as cellular extracts should still be performed to provide us with a more realistic view.

There may be a way to make phosphonomonoester linkages even more resistant to debranching. Carrocci et al discovered that using *Sp* diastereomer of GTP- αS , in 10DM24 catalyzed branching reaction results in the formation of the *Rp*-thiophosphodiester (**Figure 4-4 A**). This branch type showed exceptional resistance to debranching even in 4 mM Mn^{2+} concentration (Carrocci et al., 2017). Similarly, thiophosphonate analogs can be synthesized and tested for FJ ribozyme mediated ligation (**Figure 4-4 B**). Ultimately, the thiophosphonate-branched RNA substrate can be examined for debranching resistance. There is a likelihood that FJ ribozymes may not accept the thiophosphonate analog as substrate. In that case it may be necessary to evolve new variants that utilize such tenofovir analogs. Rescue using thiophilic metal ions such as Cd^{2+} may also be another option (Basu and Strobel, 1999), however such ions are highly toxic for cells.

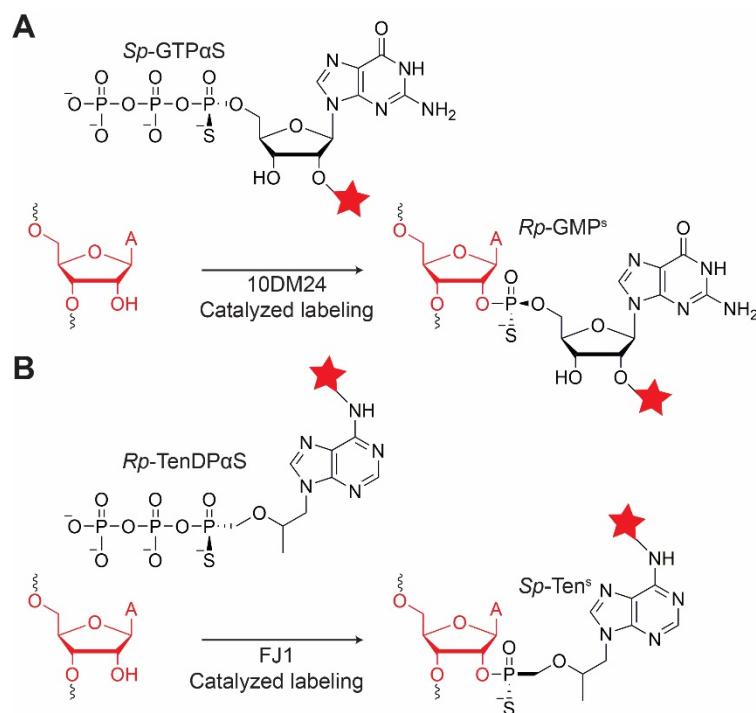


Figure 4-4 Debranching resistant NTP β s analogs. **(A)** GTP α S analog used by Carocci et al **(B)** Potentially debranching resistant substrate for FJ1 labeling.

Apart from substrate orthogonality and stability of the introduced label, FJ ribozymes share one major limitation with FH ribozymes. The dependency on high Mg^{2+} concentration is still a major issue that needs to be resolved prior to any cellular experiments. Folding in live cell and sensitivity to nuclease degradation is also an issue facing artificially developed functional nucleic acids (Filonov et al., 2015). This problem may be resolved by placing the FJ ribozymes within the context of stably folded RNA scaffold (Filonov et al., 2015) as mentioned in 1.4.1.3, for the case of fluorogenic aptamers.

Another limitation of both FH and FJ ribozymes, like most other covalent labeling methods, is the lack of fluorogenicity. It is important to design a fluorogenic substrate that only becomes fluorescent after it is attached to RNA. Otherwise, the background emission from the unligated substrate will interfere with the analysis. Absence of fluorogenicity limits the application of these ribozymes to fixed cells where the excess of unbound substrate can be washed out of the cell (Muthmann et al., 2020). The problem may be solved by using internally quenched nucleotide analogs, in which a fluorophore is conjugated to the nucleobase and a quencher to the γ -phosphate (Hacker et al., 2015). These analogs, however, are not naturally produced in the cell. The cellular entry of these types of substrates will also be challenging.

5 Conclusion and Perspective

During this thesis project, a selection method was established, which consistently resulted in the isolation of trans-acting RNA-modifying ribozymes using two different types of selection substrates. Our selection process leads to evolution of variants that mostly modify a predetermined position within the substrate sequence. Moreover, the identified variants demonstrate a great deal of sequence versatility and flexibility regarding their target sequence and cognate nucleotide analogs. They have been easily designed to target small and large RNA molecules at various potential labeling sites. All the targeted sites have successfully been labeled using various fluorophore-conjugated forms of labeling substrates.

The FH14 and the FJ ribozymes are perfectly orthogonal regarding their labeling substrates, which made possible simultaneous mutually exclusive dual-color labeling of RNA. Furthermore, due to orthogonality of the labeling substrate and resistance to enzymatic cleavage, FJ ribozymes have great potential for cellular application. Although the current ribozymes have not been tested for cellular applications, due to the limitations discussed, they may be further optimized for that purpose.

5.1 Possible strategies for FJ ribozyme optimization for cellular application

5.1.1 Reducing Mg^{2+} -dependency by reselection *in vitro*

One commonly used strategy for improving existing activities or evolving new ones, is partial randomization and reselection. The FJ ribozyme catalytic core can be randomized partially either through error-prone PCR or by doping during chemical synthesis of the template. The selection process can then be re-initiated at reduced $MgCl_2$ concentrations. After signs of enrichment become apparent, the $MgCl_2$ concentration can be further reduced until substantial activity at concentrations below 1 mM is achieved. Since cellular free Mg^{2+} concentrations are typically between 0.25 to 1 mM, a ribozyme that functions efficiently within this range, might be more suitable for RNA labeling *in situ*.

5.1.2 Possibility of selection *in vivo*

Functional nucleic acids evolved *in vivo* are known to function more efficiently under cellular conditions. Fluorogen-activating Broccoli and Corn aptamers for example,

demonstrate significantly lower $MgCl_2$ dependency compared to the *in vitro* selected HBI-analog binding aptamers. They have also been shown to demonstrate significantly improved folding and enzymatic stability in cell, as they do not necessarily require a stable RNA scaffold (Filonov et al., 2014; Song et al., 2017).

More recently an *in vivo* selection system was applied to improve the trans-activity of the Hammerhead ribozyme *in vivo* by reducing its dependence on Mg^{2+} (Huang et al., 2019). Analogously, FJ ribozymes may be further reselected *in vivo* for improved activity under cellular conditions. It is, therefore, necessary to define conditions under which the survival of cells harboring various mutant forms of FJ ribozymes (FJ1 or FJC9) can be linked to the labeling activity of these ribozymes.

A bacterial three-hybrid system can be imagined, through which cell survival and FJ1 mediated activity can be linked under certain growth conditions. Bacterial three hybrid systems are designed for the study of RNA-protein interactions (Berry and Hochschild, 2018). In these systems, a DNA-binding protein such as λ CI is fused to an RNA-binding protein such as MCP. The binding site of the λ CI protein (λ operator) is inserted upstream of a weak promoter with an extremely low basal transcription level. The promoter drives the transcription of a reporter gene. From a second plasmid, a hybrid RNA is transcribed containing the MBS stem-loop fused to an RNA library that may contain sequences with binding affinity for a particular RNA-binding protein. The RNA-binding protein of interest is fused to the α -subunit of bacterial RNA polymerase, replacing the natural α -subunit of the C-terminal domain (α CTD). λ CI-MCP fusion binds to the λ operator upstream of the reporter gene. MCP, in turn, binds to the MBS stem-loop of the chimeric RNA. In case the RNA segment fused to MBS harbors the binding motif of the RNA binding protein of interest, the bacterial RNA polymerase is recruited to the promoter, dramatically enhancing reporter gene expression (Berry and Hochschild, 2018).

Similarly, three-hybrid systems can be designed for selecting FJ ribozymes in *cis*- or *trans*-reacting settings (**Figure 5-1**):

For the *Cis*-reactive version, FJ1 or FJC9 with a partially randomized core, along with a substrate sequence, inserted into a stably folded RNA scaffold are fused to the MBS stem-loop. A DHFR-fused α -subunit is then expressed from another plasmid. N^6 -Trimethoprim conjugated tenofovir is supplied in the culture medium in disoproxil or alafenamide form. Trimethoprim is a DHFR inhibitor with strong binding affinity and has been shown to bind to this protein even when conjugated to other tags (Carrocci and Hoskins, 2014; Gallagher

et al., 2007). The tenofovir analog enters the cell and is activated by cellular machinery. The FJ ribozyme mutants that can function under cellular conditions will be conjugated to trimethoprim allowing the chimeric RNA to bind to the DFHR conjugated α -subunit, recruiting the *E. coli* RNA polymerase to the promoter and activating expression of an antibiotic resistance gene (**Figure 5-1 A**). The medium is also supplied with the corresponding antibiotic, allowing only cells that harbor highly active FJ ribozyme constructs to survive. Alternatively, the gene coding for a fluorescent protein can be placed instead of the antibiotic resistance gene, and the cells with desirable levels of FJ ribozyme activity can be sorted using FACS.

The selection for FJ ribozymes under the trans-settings can be performed following the same principles. In this system, only the substrate sequence is fused to MBS stem-loop. The FJ ribozyme in its trans-acting form, with a partially randomized core, is inserted in an RNA scaffold. Modification of the substrate sequence in trans- using trimethoprim-TenDP conjugate, allows for the activation of a reporter gene (**Figure 5-1 B**). The design for inserting a trans-acting FJ ribozyme in a stable scaffold was inspired by the work of Huang et al, in which trans-acting hammerhead ribozyme was transcribed intracellularly within the context of a tRNA scaffold (Huang et al., 2019).

5.2 Potential future selection projects

The selection process established during the work on this project offers endless possibilities for ribozyme-based RNA labeling tools. The focus in this thesis was the development of ribozymes utilizing NTP analogs based on ATP. The selection, however, can be performed using other types of biotinylated natural and unnatural NTPs. Those selection experiments can result in the development of more ribozyme variants with specificities towards various NTP-based substrates. This will allow us to expand our labeling potential beyond dual-color imaging of RNA.

By changing the bulged nucleotide adenosine to other nucleotides, it might be possible to evolve ribozyme with specificities towards other internal nucleotides. Ultimately, randomizing this position may provide the possibility of selecting ribozymes that can modify any type of nucleotide. Such ribozymes can be targeted to virtually any accessible position within the RNA of interest.

Moreover, non-nucleotide-based selection substrates can be applied in the selection process to develop other classes of RNA-modifying ribozymes. Other possibilities include the development of ribozymes such as RNA alkyltransferases or acyltransferases. The modification sites in those cases may not be limited to the 2'-OH of the ribose backbone but the nucleobases may also be involved. Non-nucleotide based, uncharged, fluorogenic substrates can be used in the selection to evolve RNA labeling ribozyme with fluorescent turn-on potential. The uncharged nature of the substrates would also resolve the issue of cell permeability.

Countless other selection scenarios can be imagined or realized using the selection system established and optimized during the work of this thesis project. Thus, the search for ideal RNA labeling ribozymes goes on.

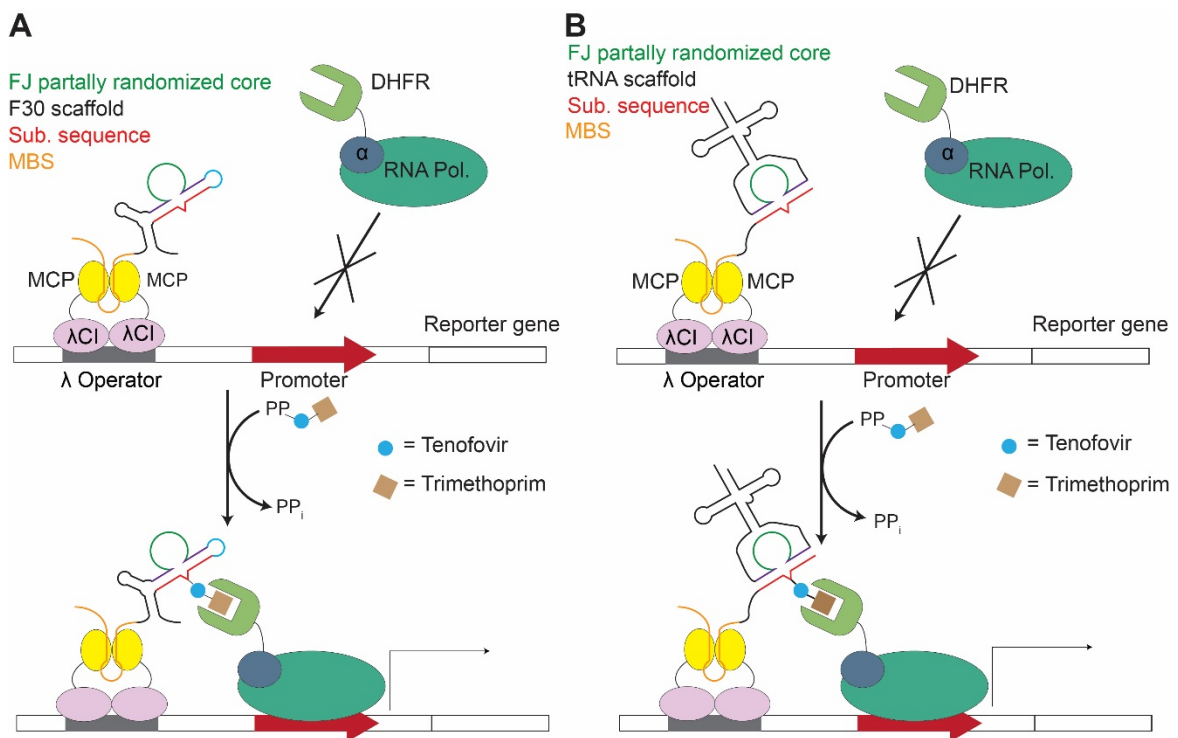


Figure 5-1 Possible three-hybrid system for reselection of FJ ribozymes in vivo. **(A)** Reselection based on Cis-activity. **(B)** reselection based on Trans-activity.

6 Material and methods

6.1 Instruments

Table 6-1 List of instruments used in this thesis project

Instrument	Model	Company
Centrifuges	Centrifuge 5415R Centrifuge 5424 Centrifuge 5804R	Eppendorf
Fluorescence imaging	ChemiDoc MP with epi illumination	Biorad
Fluorescence spectrophotometer	FP-8300	Jasco
Gel dryer	Model 583 gel dryer	Biorad
HPLC	ÄKTApurifier 10 with autosampler ÄKTAmicro	GE Healthcare
Mass spectrometer	micrOTOF-Q III	Bruker
Nanophotometer UV/Vis	Nanodrop spectrophotometer with Lid10 and Lid50	Implen
Oligonucleotide desalting	ÄKTA start	GE Healthcare
PCR	T100 Thermal cycler	Biorad
Phosphorimager	Amersham Typhoon	Amersham Bioscience
Speed Vac	Savant Speedvac Concentrator	Thermo Scientific
Synthesizer	LKB Gene Assembler Plus	Pharmacia
UV imaging system	DIAS-II	SERVA

6.2 Materials

6.2.1 Chemicals and reagents

All common chemicals and reagents were purchased from ROTH, Applichem, Sigma Aldrich, and Roth with no further purification.

All the ATP analogs used in this study, including the N⁶-Biotin-ATP and the N⁶-fluorophore-conjugated analogs were purchased from Jena bioscience.

Fluorescein-NHS ester was purchased from ThermoFisher Scientific, fluorescein-thiosemicarbacide and lucifer yellow carbohydrazide were purchased from Sigma. 6-FAM-PEG₃-azide was purchased from Jena Bioscience.

Enzymes were purchased from ThermoFisher Scientific. T7 RNA polymerase was a lab prepared stock. Yeast Dbr1 was kind gift from A. Hoskins (Dept. of Biochemistry, U. Madison-Wisconsin).

MicroSpin G-25 Columns and Ready-To-Go RT-PCR Beads were purchased from GE Healthcare. QIAquick[®] Gel Extraction Kit and QIAprep[®] Spin Miniprep *Kit* were purchased from Qiagen. TOPO[®] TA Cloning[®] Kit for Sequencing with One Shot[®] TOP10 Chemically Competent *E. coli* was purchased from New England Biolab.

γ -³²P-ATP was purchased from Hartmann Analytic GmbH.

Transcription templates and primers were purchased from Microsynth and purified on denaturing PAGE prior to use.

Water used for all experiments was purified with a Milli-Q-unit of Sartorius.

6.2.2 Buffers and solutions

TBE buffer (10x): 890 mM Tris, 890 mM boric acid, 200 mM EDTA (pH= 8.3)

Acrylamide-gel stock solution (10%): 10% acrylamide, 7 M urea, 89 mM Tris, 89 mM boric acid, 20 mM EDTA (pH= 8.3)

Acrylamide-gel stock solution (15%): 15% acrylamide, 7 M urea, 89 mM Tris, 89 mM boric acid, 20 mM EDTA (pH= 8)

Acrylamide-gel stock solution (20%): 20% acrylamide, 7 M urea, 89 mM Tris, 89 mM boric acid, 20 mM EDTA (pH= 8)

High dye solution: 80% formamide (99%), 89 mM Tris (pH = 8), 89 mM boric acid, 70 mM EDTA (pH = 8), 0.25 mg mL⁻¹ bromophenol blue, 0.25 mg mL⁻¹ xylene cyanol

STOP solution: 80% formamide (99%), 89 mM Tris (pH = 8), 89 mM boric acid, 50 mM EDTA (pH = 8)

TEN buffer: 100 mM Tris (pH = 8), 1 mM EDTA, 300 mM NaCl

Transcription annealing buffer (10x): 100 mM Tris (pH = 8), 50 mM NaCl, 0.1 mM EDTA

Selection buffer (5x): 250 mM HEPES (pH = 7.5), 600 mM KCl, 25 mM NaCl

Binding-Washing (BW) buffer (2x): 2 M NaCl, 10 mM Tris (pH = 7.5), 1 mM EDTA (pH = 8)

Denaturing washing buffer: 8 M urea, 10 mM Tris (pH = 7.5), 1 mM EDTA (pH = 8), 0.01% tween-20

Formamide elution buffer: 95% formamide, 1 mM EDTA (pH = 8)

TBS buffer (5x): 250 mM Tris (pH = 7.5), 750 mM NaCl

X-gal solution: 40 mg/ml X-gal in DMSO

8-17NG cleavage buffer (5x): 250 mM HEPES (pH = 7.5), 2 M KCl, 500 mM NaCl

Dbr1 reaction buffer (5x): 250 mM Tris pH = 7.4, 125 mM NaCl, 12.5 mM DTT, 0.05% (V/V) Tween-20, 0.75% (V/V) glycerol

RNA extraction solution: 95% formamide, 18 mM EDTA (pH = 8), 0.025% SDS, 1% 2-mercaptoethanol.

Detritylation solution: 3% Dichloroacetic acid in 1,2-dichloroethane

Activation solution: 0.25 M ethylthiotetrazole in acetonitrile

Oxidation solution: 10 mM Iodine in acetonitrile/ *sym*-collidine/ H₂O: 10/1/5

Capping A solution: 0.5 M 4-(Dimethylamino)pyridine in acetonitrile.

Capping B solution: Aceticanhydride/ *sym*-collidine/ acetonitrile: 2/3/5

6.3 Methods

6.3.1 General procedures

6.3.1.1 PAGE purification

To ensure maximum purity of the oligonucleotides used in the experiments, purchased samples were purified using denaturing polyacrylamide gel electrophoresis (PAGE) prior to use. Depending on the size of the oligonucleotides, they were resolved on 10%, 15% or 20% acrylamide gel with a thickness of 0.75 mm. For gel preparation, 50 mL acrylamide gel stock solution of the corresponding percentage was mixed with 45 μL of *N,N,N',N'*-tetramethylethylene-1,2-diamine (TEMED) and 150 μL of ammonium peroxodisulfate (APS, 25% (*w/v*) in water). The polymerized gel was put into a gel electrophoresis apparatus filled with 1x TBE buffer and pre-equilibrated at a constant power of 35 W for 30 minutes. The dissolved samples were mixed with an equal amount of high dye solution and loaded on the pre-equilibrated gel. After application of the samples, it was run for another two hours using the same voltage and current conditions as before. The gel was then transferred between cling film and imaged using UV light (265 nm). The bands corresponding to the desired oligonucleotides were cut out, chopped up, and transferred into 1.5 mL microtubes. TEN buffer was added until the gel pieces were completely covered (usually 300 - 450 μL). The mixture was then incubated at 37 °C for 1.5 - 3 hours or overnight at 4 °C. The gel extract was aspirated and transferred into a fresh microtube. A second fraction of TEN buffer was added to the gel pieces and treated as the previous fraction. The gel extracts were EtOH precipitated, and the pellets were dissolved in water.

6.3.1.2 Ethanol precipitation

Absolute ethanol with a temperature of -20 °C was added to the solution containing oligonucleotides at a ratio of 3:1. The minimum volume of oligonucleotide solution was 100 μL including a final concentration of 300 mM NaCl as precipitant. After freezing the mixture in liquid nitrogen for 3-5 minutes, it was centrifuged in a refrigerated tabletop centrifuge at maximum speed, at 4 °C for 30 minutes. The supernatant was aspirated, and the pellet was desalted by adding 75 μL of ice-cold 70% EtOH. The sample was spun down at top speed and 4 °C for another 10 minutes. Afterwards, the supernatant was discarded, the pellet was dried in vacuum and dissolved in milliQ water.

6.3.1.3 In vitro transcription

The annealing mixture was prepared by dissolving 100 pmol of the oligonucleotide D008 (T7 promoter) (Milligan et al., 1987) and 100 pmol of the single-stranded DNA template in 50 μ L of 1x annealing buffer. Annealing was performed by incubating the reaction mixture for 5 minutes at 95 °C followed by incubation at room temperature for 10 minutes. Afterwards, 4 μ L of 1M Tris buffer (pH= 8), 4 μ L of 250 mM DTT, 4 μ L of each NTP (100 mM stock concentrations), 3 μ L of 1 M MgCl₂, 2 μ L of 100 mM spermidine was added to the annealed mixture and the total volume was brought to 96 μ L using deionized water. After mixing properly, 4 μ L of T7 RNA polymerase (4 mg/mL) was added to the reaction and the sample was incubated at 37°C for 3-6 hours. The final concentration of each reagent was therefore 40 mM Tris, 10 mM DTT, 4 mM of each NTP, 30 mM MgCl₂, 2 mM spermidine. After the incubation, the reaction was then stopped by adding 20 μ L of 0.5M EDTA and 50 μ L of high dye gel loading buffer. The final transcript was then purified from the mixture using PAGE purification.

6.3.1.4 Solid-phase oligonucleotide synthesis

RNA oligonucleotides were synthesized on amino-ON controlled pore glass (CPG) with a pore size of 500 Å and a loading of 30-40 μ mol g⁻¹. The synthesis scale was 0.85 μ mol (25 mg of CPG), with a final detritylation step to remove the 4,4'-dimethoxytrityl group. For each coupling step, 75 μ L of a 100 mM phosphoramidite solution (in MeCN) was used, which was dried over activated 3 Å molecular sieves prior to synthesis. The same MeCN was also used for the rinsing procedure and preparation of other solutions for the solid-phase synthesis. The detritylation was performed using dichloroacetic acid (3% in DCE). ETT (0.25 M in MeCN) was used for the activation of the phosphoramidite building blocks. The coupling time for unmodified RNA phosphoramidites was 4 minutes and 12 minutes for modified phosphoramidites. For the capping of unreacted nucleosides, a 1:1 mixture of Capping A and Capping B solution was used. Oxidation was performed by treatment with an iodine solution (10 mM in *sym*-collidine/H₂O/MeCN 1:5:11) for 1 minute. After the synthesis, the CPG support with the bound oligonucleotide was dried under vacuum and subsequently subjected to deprotection.

Cleavage of the oligonucleotides from the solid support and deprotection of the nucleobases and phosphate groups was achieved by incubating the CPG with 0.5 mL of MeNH₂ (40% in water) and 0.5 mL of NH₃ at 65 °C for 3 hours. The reaction mixture was centrifuged, the

supernatant was collected, and the CPG was washed with 0.5 mL of water. Both solutions were combined, and the solvent was evaporated. For 2'-deprotection, the RNA was re-dissolved in 0.5 mL of a 1 M TBAF solution (in THF). The reaction was performed overnight at 37 °C while shaking (ca. 200 rpm). The reaction was quenched by adding 0.5 mL of 1 M TRIS-HCl solution (pH = 8.0) and THF was removed under reduced pressure. Subsequently, the oligonucleotides were desalted by size exclusion chromatography using HiTrap desalting columns (3 x 5 mL) and water as eluent. After desalting, the solvent was removed under reduced pressure and the RNA was re-dissolved in 0.5 mL of water. The crude RNA samples obtained were purified on denaturing PAGE.

6.3.1.5 Oligonucleotide labeling

6.3.1.5.1 3'-end labeling using periodate glycol-oxidation

For the in vitro transcribed substrate sequences, 3-5 nmol of RNA was dissolved to a final concentration of 100 μ M in 50 mM sodium phosphate buffer pH=7.4, NaIO₄ was added to a final concentration of 20 mM and the reaction was incubated at 37°C for 10 minutes. The excess of unreacted periodate was quenched by addition of Na₂SO₃ to a final concentration of approximately 100 mM and the reaction was incubated for an additional 5 minutes at 37°C. Fluorescein-thiosemicarbazide was added to a final concentration of approximately 10 mM and the reaction was incubated in the dark for 1 hour. Stop solution was then added to the reaction at an amount equal to the final reaction volume and the sample was resolved on a 0.75 mm, 20% denaturing PAGE under constant power of 35 W for 2 hours.

For longer transcripts (i.e. the selection pool and in vitro transcribed *E. coli* 5S rRNA and RyhB RNA) a slightly modified protocol was used. The RNA was dissolved in the reaction mixture to a final concentration 30-60 μ M in 50 mM sodium phosphate buffer pH= 7.4, NaIO₄ was added to a final concentration of 20 mM and the reaction was incubated at 37°C for 10 min followed by quenching at 37°C for 5 min using a final concentration of ~100 mM Na₂SO₃. Lucifer yellow-carbohydrazide or fluorescein-thiosemicarbazide was then added to the sample (~10 mM final concentration) and the reaction was incubated at 37°C for 1 hour. Instead of gel purification, Illustra G-25 spin columns (GE-healthcare) were used to remove the salts and the excess of the unreacted dye from the samples.

6.3.1.5.2 CuAAC labeling of the synthetic 5'-hexynylated RNA

5 nmol of the synthetic 5'-hexynyl-RNA was dissolved in 10 μ L of H₂O: DMSO: tBuOH (4:3:1) containing 5 mM 5-FAM-azide or sulfocyanine 5-Azide, 5 mM CuBr and 10 mM TBTA. The reaction mixture was incubated at 37°C for 3 hours and then quenched by adding 10 μ L of high dye gel-loading buffer. The sample was resolved on a 20% denaturing gel. The band corresponding to the labeled RNA was then cut out and subjected to gel extraction and ethanol precipitation.

6.3.1.5.3 5'-³²P labeling of DNA primers

100 pmol of the DNA primer was dissolved in 10 μ L of PNK buffer A including 5 μ Ci of γ -³²P-ATP, and 5 units of PNK. The reaction was incubated at 37°C for 1 h, followed by two rounds of ethanol precipitation. The pellet was dissolved in 10 μ L of milliQ water, resulting in a sample with 2000-3500 IPS.

6.3.1.6 Total cellular RNA extraction

Total RNA from *E.coli* was obtained by RNA extraction using a modified RNA snap protocol (Stead et al., 2012). Briefly, TOP10 cells (NEB) were grown in LB medium to an OD₆₀₀ of 0.4-0.8. 1 mL of this culture was then spun down, and the supernatant was discarded. The pellet was then resuspended in 100 μ L of RNA extraction solution and the sample was placed at 95°C for 7 minutes. The sample was then immediately centrifuged at 16000g for 5 minutes. The supernatant containing total cellular RNA was then carefully removed and brought to 400 μ L by adding milliQ water. The sample was then subjected to P/CI extraction and ethanol precipitation. This procedure usually yielded in isolation of 0.7-1.5 μ g of total cellular RNA.

6.3.2 Ribozyme selection

6.3.2.1 Preparation of the selection pool

The DNA template of the pool was prepared by mixing 600 pmol of each of the forward and reverse pool primers in a total volume of 100 μ L of milliQ water. The sample was annealed by placing at 95°C for 5 minutes and slow cooling to room temperature for 10 minutes. The rest of the reagents were then added to this mixture, resulting in a 200 μ L final volume of 1x

Klenow buffer, 200 μM of each dNTP, and 0.2 Unit/ μL of Klenow exo^- . The reaction was then incubated at 37°C for 1 hour. The sample was then subjected to P/CI extraction and ethanol precipitation to obtain 600 pmol of the double-stranded full-length DNA template.

Approximately 450 pmol of the double-stranded pool template was dissolved in 400 μL of standard *in vitro* transcription mix (See 6.3.1.3). The sample was incubated at 37°C for a minimum of 6 hours. The reaction was then stopped by adding 80 μL of 0.5 M EDTA pH = 8 and 320 μL of stop solution. The pool transcript was then subjected to PAGE purification.

6.3.2.2 Selection process

6.3.2.3 Incubation step

Each incubation step of the FJ selection, and until round 7 of the FH selection, was performed with ~ 50 μM of the pool RNA. This approximation is made by taking into consideration the ~ 150 -300 pmol of the 3'-lucifer yellow labeled pool. The 1st selection round in each selection experiment was performed using ~ 3.3 nmol of the RNA pool (including the lucifer yellow labeled fraction) in 60 μL of the selection reaction. For the subsequent rounds, the amount of RNA pool used in the reaction was adjusted based on the transcription yield of the previous round. The volume of the selection reaction was kept at a minimum of 12 μL . During the FH selection, the RNA concentration was reduced to half after the 7th round. After the 11th round of the FH selection only the labeled fraction was added.

For FH selection the concentration of the selection substrate, N⁶-biotin-ATP was 200 μM throughout the selection rounds. The concentration of the selection substrate N⁶-TenDP in FJ selection was maintained at 300 μM in all selection rounds. Both selections were performed in 1x selection buffer and the presence of 40 mM MgCl₂. At every selection round, an annealing step was performed prior to the addition of the selection substrate and MgCl₂. To do so the samples were placed at 95°C for 3 minutes followed by 10 minutes incubation at room temperature. The incubation temperature for each round was 37°C for both selections. The incubation period was overnight in the initial rounds of both selection reactions which was gradually reduced after significant enrichment was observed. The details of FH and FJ selection rounds are listed in (**Table 6-2**) and (**Table 6-3**), respectively.

6.3.2.4 Capture, wash and elution

To separate the active (biotinylated) species from the large bulk of inactive pool, streptavidin or neutravidin coated magnetic beads were used. To avoid selecting for neutravidin/streptavidin binders, the affinity matrix was alternated between neutravidin and streptavidin every 2 rounds. For the first selection round of all selection experiments, a volume from the bead suspension corresponding to 1 mg of magnetic beads was transferred into a 1.5 mL microcentrifuge tube. The beads were equilibrated by washing three times, each time using 1 mL of 1x BW buffer. The beads were then resuspended to a final concentration of 2.5 mg/mL in 1X BW buffer containing 5 mg/mL *E. coli* tRNA (Sigma-Aldrich) and incubated at room temperature for 20 minutes while shaking. The supernatant was then removed after placing the beads on the magnetic stand for 1 min. The beads were washed once with 1mL of 1x BWT buffer (1x BW + 0.1% Tween-20) and twice, each time with 0.5 mL of 1X BW buffer. The binding of the biotinylated species was performed, after tRNA treatment, by resuspending the beads in 2x BW buffer. The biotinylated pool RNA, dissolved in an equal volume of milliQ water, was then added to the suspension. The sample was then shaken at RT for 1 hour.

At this point the beads were placed on the magnetic stand and the supernatant containing the unbound pool RNA was removed. The beads were then washed 5 times, each time with 1 mL of denaturing wash buffer and twice, each time 1 mL of milliQ water. The elution was done by resuspending the beads in 100 μ L of the elution buffer and placing the sample at 95°C for 10 minutes. The sample was placed on a magnetic stand and the supernatant was transferred to a fresh microcentrifuge tube. The beads were further washed twice with a total volume of 590 μ L of TEN buffer and the supernatants from these washing steps were pooled in the same tube as the initial eluent. 10 μ L of 3M NaCl was then added to the sample and the eluted RNA was then precipitated by addition of ice-cold ethanol (2 mL). The pellet was dissolved in 50 μ L H₂O and the enrichment was quantified by fluorescence measurement (See 6.3.2.5).

In the subsequent selection rounds 0.1 mg of beads were used for each 100 pmol of the pool RNA. The minimum amount of beads used was 0.6 mg. The details of bead equilibration, tRNA treatment, binding, washing, and elution were the same as for the first round. See (Table 6-2) and (Table 6-3).

6.3.2.5 Monitoring of the selection progress

A portion of the selection pool, entering each selection round was labeled at 3'-end using LY-CH. After each incubation step the sample was subjected to ethanol precipitation and the resulting pellet was dissolved in an amount of milliQ water corresponding to twice the volume of the bead slurry taken for the capture step (or equal to the volume of the 2x BW buffer used for resuspending the beads after equilibration and blocking). 10 μ L of this sample was taken and diluted to a total volume of 50 μ L. This sample was then subjected to fluorescent measurement using Jasco fluorescent spectrophotometer. The following parameters were used during fluorescence measurement: Excitation wavelength 428 nm, Emission range 440-700 nm, emission excitation bandwidth 5nm, scanning speed 100 nm/min. The resulting pellet from the elution step was also dissolved in 50 μ L of milliQ water and subjected to fluorescent measurement using the same parameters. The estimation of the percentage of RNA pool recovered after each round was made using the following formula:

$$\% \text{ Eluted} = \frac{A_e}{A_i \left(\frac{2 * \text{Bead volume}}{10} \right)} * 100$$

(Ae = integrated peak area of the fluorescent emission of the eluted sample, Ai = integrated peak area of the fraction taken from the sample prior to addition of the bead slurry)

Table 6-2 Details of each FH selection round.

Selection round	1	2	3	4	5	6	7	8	9	10	11	12	13
Amount of RNA(nmol)	3	1	1	0.6	0.6	0.6	0.6	0.3	0.3	0.3	0.3	0.15	0.15
Reaction volume (μ L)	60	20	20	12	12	12	12	12	12	12	12	12	12
Bead type	N	N	S	S	N	N	S	S	N	N	S	S	N
Amount of beads (mg)	1	1	1	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6

Table 6-3 Details of each FJ selection round.

Selection round	1	2	3	4	5	6	7	8	9	10	11	12
Amount of RNA*(nmol)	3	1	0.8	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6
Reaction volume (μ)	60	20	16	12	12	12	12	12	12	12	12	12
Bead type**	N	N	S	S	N	N	S	S	N	N	S	S
Amount of beads (mg)	1	1	0.8	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6	0.6

6.3.2.6 Amplification step

The capture step was followed by a two-step amplification protocol. The first step involved a one-pot RT-PCR reaction during which the recovered RNA at the end of each round was reverse transcribed to cDNA, followed by 10 cycles of PCR to result in the back-up sample of the selection. A portion of this sample was then used in a standard PCR reaction to generate a sufficient amount of DNA template required for transcription of the enriched RNA pool entering, next selection round (**Figure 3-3**). The rest of the sample was stored as a back-up of the selection round.

The RT-PCR amplification scheme is depicted in (**Figure 3-3**). Two separate forward primers were designed to ensure the amplification of ribozymes that modify the substrate sequence, while ribozymes that modify internal nucleotides in the core would not be amplified efficiently. The forward primer used in the RT-PCR step covers the connecting loop while only partially covering the substrate sequence (Starting at two nucleotides 3' to the predetermined modification site). The forward pool primer was then used in the second PCR reaction to restore the substrate sequence to its full-length. This primer also carried the T7 promoter sequence at its 5' end. The reverse primer was the same in both amplification steps (**Figure 3-3**).

RT- (1st) PCR

Following fluorescent measurement, 30 pmol of the forward and 50 pmol of the reverse primers were added to the 50 μ L eluted RNA sample. The resulting solution was transferred into a 0.2 μ L microcentrifuge tube containing one Illustra Ready-to-Go RT-PCR bead and the sample was placed in a thermal cycler. Cycling parameters are listed in **Table 6-4**.

Table 6-4 Cycling parameters of the one-pot, RT-PCR reaction

Reaction step	Temperature ($^{\circ}$ C)	Incubation time	Number of cycles
Reverse transcription	45	30 min	1
Initial melting	95	4 min	1
Melting	95	30 sec	10-2*
Annealing	65	30 sec	
Extension	72	25 sec	
Final extension	72	4 min	1
Infinite 4 $^{\circ}$ C hold			

* For the initial selection rounds the number of cycles was maintained at 10. After the enrichment was observed the cycle number was reduced due to the increasing amount of template.

2nd PCR

A portion of the initial RT-PCR sample was used as a template in a subsequent PCR reaction for making sufficient transcription template. The rest of the sample was stored as a back-up of the selection round. In general, 10 μ l of the first PCR reaction was used with 100 pmol of forward and 150 pmol of reverse primer in 200 μ l of the 1x DreamTaq buffer, including 200 μ M of each dNTP, 10 % DMSO and 1 unit of the DreamTaq DNA polymerase. The cycling parameters are listed in **Table 6-5**.

Table 6-5 Cycling parameters of the PCR reaction

Reaction step	Temperature ($^{\circ}$ C)	Incubation time	Number of cycles
Initial melting	95	4 min	1
Melting	95	30 sec	32-15*
Annealing	65	30 sec	
Extension	72	25 sec	
Final extension	72	4 min	1
Infinite 4 $^{\circ}$ C hold			

* In the initial rounds 32 cycles were used, after the enrichment level increased the number of required cycles was also reduced.

6.3.2.7 Activity assay on the enriched pool

6.3.2.7.1 Cis-activity assay

Prior to cloning and sequencing, the enriched pool from round 12 of FH and FJ selection experiments were examined for activity using streptavidin gel-shift assay. 150 pmol of each pool was labeled at 3'-end using lucifer yellow. The labeled samples were then incubated with their cognate biotinylated substrate under selection condition in a total reaction volume of 10 μ L. An annealing step, in this case, was also performed prior to addition of the biotinylated substrate and MgCl₂. Aliquots (2 μ L) were taken at certain time-points and were quenched by adding to 98 μ L of TEN buffer. The samples were then subjected to ethanol precipitation to remove the excess of the unreacted biotinylated substrate. The pellet was subsequently dissolved in 5 μ L 1x TBS buffer including 2 μ g of streptavidin. The samples were then incubated at room temperature for 5 minutes and mixed with 1 μ L of the 6x glycerol loading dye. The samples were then resolved on a 10% PAGE (minigel 10 x 8 x 0.1 cm) without urea for 45 min at 200 V. The gel was then subjected to fluorescent imaging

using the biorad chemidoc gel documentation device. The shifted bands were quantified using image-lab software.

6.3.2.7.2 Trans-activity assay

The trans-acting ribozymes were generated according to the scheme depicted in (**Figure 3-6 C**). The PCR reactions were performed in 100 μ L of the 1x DreamTaq buffer including 200 μ M of each dNTP, 0.5 unit of the DreamTaq DNA polymerase, 100 pmol of the selection reverse primer and the alternative forward primer. The forward primer was designed to remove the substrate sequence and add the T7 promoter and transcription start site directly to the connecting loop. For the sample pertaining to FH selection, 2 μ L of the backup sample from round 13 was used as template. In the case of the FJ selection 2 μ L of the backup sample from round 12 was added to the reaction as template. Cycling parameters in case both reactions were as listed in **Table 6-5**, however, the annealing temperature was reduced to 62°C. The reaction was then ethanol precipitated and the dsDNA template was transcribed in vitro under standard conditions.

About 120 pmol of the RNA transcribed from the generated templates were then mixed with 240 pmol of the in vitro transcribed, 3'-LY labeled substrate sequence in 10 μ L of a reaction with the same conditions as the selection rounds. 1 μ L aliquots were taken from the reaction at certain time-points and quenched by transferring into 4 μ L of stop solution. 2.5 μ L of this sample was then resolved on a 20% denaturing PAGE under 35 W for 1 hour. The gel was then subjected to fluorescent imaging.

6.3.3 Sanger sequencing

6.3.3.1 Cloning and of the enriched pool

In the case of FH selection, 2 pmol of the transcript generated after the round 13 was subjected to RT followed by 15 cycles of PCR, using the pool forward primer and the selection reverse primer. Parameters were set as described in **Table 6-4**. 2 μ L of this reaction mixture was then transferred into a 200 μ L PCR reaction using the forward cloning primer and the selection reverse primer (Reaction conditions: 1x DreamTaq buffer including, 200 μ M of each dNTP, 10% DMSO, 1 unit DreamTaq DNA polymerase and 1 μ M of each primer. cycling conditions in **Table 6-5**). The product was then cleaned up using PCR clean-up kit (Qiagen). The insert was then used in a TOPO-TA cloning reaction according to the

manufacturer's protocol. Briefly, 16 ng of the insert was mixed with 1 ng of the vector in a total volume of the 1x salt solution provided by the manufacturer. The reaction was incubated at room temperature for 30 minutes before transformation.

For FJ selection 2 μL of the back-up RT-PCR sample was first PCR amplified using the pool forward primer and the selection reverse primer. 2 μL of this sample was then used as template in a 200 μL PCR reaction in which the cloning forward primer and the selection reverse primer were included (Reaction conditions: 1x DreamTaq buffer including, 200 μM of each dNTP, 10% DMSO, 1 unit DreamTaq DNA polymerase and 1 μM of each primer. cycling conditions in **Table 6-5**). The Product was cleaned-up using Qiagen kit and 30 ng of this product was used in TOPO-TA cloning reaction.

6.3.3.2 Heat shock transformation

2 μL of the ligation mixture were added into a vial of OneShot chemically competent *E. coli* cells (NEB). The sample was mixed by gentle tapping on the tube and placed on ice for 30 minutes. The mixture was then subjected to heat shock by placing at 45°C for 30 seconds and back on ice immediately. Afterwards 250 μL of room temperature S.O.C.-medium was added to the mixture. The tubes were capped tightly and shaken horizontally (ca. 200 rpm) for 1 hour at 37°C. 10 μL and 50 μL fraction from each suspension were then plated onto LB agar containing 150 $\mu\text{g}/\text{mL}$ Ampicillin. The surface of the solid culture medium was coated evenly with 40 μL of X-gal solution prior to plating of the cells. The X-gal coating was done for blue-white screening.

6.3.3.3 Colony PCR

To confirm the success of the cloning process forty PCR reaction mixtures were prepared. Each mixture contained 25 pmol of the cloning forward primer and 25 pmol of the selection reverse primer dissolved in 20 μL of 1x Taq buffer including 200 μM of each dNTP, 0.3 units of the enzyme and 5% DMSO. The randomly chosen colonies were picked using sterilized wooden toothpicks and spotted onto a fresh LB agar containing 150 $\mu\text{g}/\text{mL}$ ampicillin as the backup of the chosen colonies. After spotting onto the backup plates the toothpicks were dipped inside the individual PCR reactions a few times. The toothpicks were discarded, and the reactions were sealed. The tubes were cycled using the same parameters as in **Table 6-5**. 5 μL of each PCR reaction was mixed with 1 μL of the glycerol loading dye

and resolved on 10% native PAGE (minigel) under 200 V for 45 minutes. The gel was then subjected to SYBR green staining and fluorescent imaging.

6.3.3.4 Activity assay of the individual clones

The clones that harbored the insert were subjected to streptavidin gel-shift based activity assay. To generate transcripts from the individual clones for the activity assay, 2 μL of the colony PCR reaction products were added into PCR reactions with total volumes of 25 μL . reaction conditions and cycling parameters were identical to the colony PCR reactions. The reaction products of these PCR reactions were ethanol precipitated and the dried pellets were dissolved in 30 μL of standard in vitro transcription mix (Tris 50 mM pH= 8, 10 mM DTT, 4 mM each NTP, 30 mM MgCl_2 , 2 mM spermidine, 4 $\mu\text{g}/\text{mL}$ T7 RNA polymerase). The samples were then incubated at 37°C overnight. Afterwards the reactions were quenched by adding 30 μL of high-dye loading buffer and the crude transcripts were purified using preparative 10% denaturing PAGE.

The activity assay was performed using 75 pmol of transcript from each clone in 5 μL of the 1x selection buffer including 40 mM MgCl_2 and 200 μM . The samples were incubated at 37°C, for 2 hours and quenched by adding 95 μL of TEN buffer and subjected to ethanol precipitation, the dried pellet was dissolved in 25 μL of milliQ water. About 3 pmol (1 μL) of the reacted RNA was added in 5 μL of 1x TBS buffer containing 1 μg of streptavidin and incubated at RT for 5 minutes. 1 μL of the 6x glycerol-based loading dye was added to the solution and the sample was resolved on a 10% PAGE, for 45 minutes under 200 V. After the gel was run, it was stained using SYBR green II and subjected to fluorescent imaging.

6.3.3.5 Plasmid extraction

Chosen colonies were grown overnight at 37°C, in 5-7 mL of LB medium containing 150 $\mu\text{g}/\text{mL}$ ampicillin. The cells were then pelleted and subjected to plasmid extraction using Qiagen miniprep plasmid extraction kit, according to the manufacturer's procedure. 0.5-2 μg of each plasmid in 25 μL of milliQ buffer, were submitted to Microsynth for sanger sequencing. The results were analyzed with the aid of the Snapgene Viewer software.

6.3.4 Next-generation sequencing sample preparation

50 μ L PCR reactions were setup using 50 pmol of the extended NGS reverse primer and the forward primer containing the random 8 nucleotide unimolecular identifier sequence. 2 μ L of round 7 or round 12 backup samples were added to each reaction as template. 2 μ L of these PCR reactions were used as templates in 100 μ L PCR reaction each containing 100 pmol of the forward primer with the common NGS index. The reverse primer of each selection rounds had a unique index sequence and was added to the respective reaction (100 pmol). The conditions and cycling parameters of each of these PCR steps were as described in 6.3.2.6. The 100 μ L PCR reactions were then precipitated and purified using 2% agarose gel and Qiagen QIAquick Gel Extraction Kit. The samples were submitted to Core Unit Systems Medicine (Uni Würzburg) for amplicon sequencing (NextSeq-500 HighOutput 75nt single end).

6.3.5 Characterization of the chosen variants

6.3.5.1 General kinetic assay protocol

Kinetic assays were performed under single turn-over conditions with a 10:1 ratio of ribozyme: substrate sequence. Generally, 10 pmol of the fluorescently labeled RNA substrate was mixed with 100 pmol of the ribozyme in a total volume of 10 μ L of the selection buffer including 40 mM $MgCl_2$. For FH ribozymes NTP analog concentration in these experiments was 200 μ M. for FJ ribozymes TenDP analog concentrations were adjusted to 300 μ M. To ensure proper folding of the ribozyme and efficient pairing of the recognition arms with the substrate sequence, an annealing step (95°C for 3 min, room temperature for 10 min) was performed before addition of $MgCl_2$ and the NTP substrate. The reaction mixtures were incubated at 37°C. 1 μ L aliquots were taken from the samples at certain time-points and quenched immediately by adding to 4 μ L of stop solution. 2.5 μ L of each time-point sample was resolved on 0.4 mm, 20 % denaturing analytical gel under constant power of 35 W for 32 cm gel plates, and 25 W for 20 cm gel plates. The running times were generally 1 hour. For experiments in which unmodified tenofovir was included the running time was increased to 1.5 hours. The gel was then imaged using a Bio-Rad Chemidoc gel documentation device. The settings for imaging were chosen based on the fluorophore. The shifted bands were quantified using Imagelab software. Data-points were plotted using kaleidagraph software and the curve-fits were made using the pseudo-first order kinetics function: $Y = Y_{max} (1 - \exp(-k_{obs}t))$.

6.3.5.2 Characterization of the reaction product

6.3.5.2.1 Preparative scale modification of the reaction product for mass analysis

To prepare the sample for mass spectrometry, 1 nmol of the R 594 and 1.2 nmol of the ribozyme was dissolved in a final volume of 25 μ L of the selection buffer including 40 mM $MgCl_2$. For FH14 ribozyme N^6 -aminoheptyl ATP (400 μ M) was used as the modification substrate and for the FJ1 N^6 -TenDP-biotin (500 μ M) was used as the modification substrate. After incubation (37°C, overnight) the reaction was resolved on a 20%, 0.4 mm denaturing PAGE for 2 hours under constant power of 35 W. The band corresponding to the modified substrate sequence was then extracted from the gel and ethanol precipitated. 200 pmol of this sample was then dissolved in 8 μ L of water and was submitted to ESI-mass spectroscopy.

6.3.5.2.2 Preparative scale modification of the 5'-5FAM-R660

For each reaction 1 nmol of the 5FAM-R660 and 1.3 nmol of the ribozyme were dissolved in 25 μ L of the 1x selection buffer including 400-500 μ M of the ribozyme's cognate biotinylated substrate and 40 mM $MgCl_2$. The reaction was incubated at 37°C overnight. The reaction was then quenched by adding 25 μ L of the stop solution and resolved on a 20% analytical denaturing PAGE. The band corresponding to the modified RNA was excised and isolated.

6.3.5.2.3 RNase T1 probing

RNase T1 probing was performed by dissolving 5 pmol of the modified or unmodified substrate sequence in a final reaction volume containing 50 mM Tris pH=7.5 and 1 Unit of RNase T1. The sample was incubated at 37°C for approximately 30 seconds before quenching by addition of 5 μ L of stop solution. A volume corresponding to 1.25 pmol of the substrate sequence, from each of these samples were resolved on a 20% analytical denaturing PAGE for 1 hour under constant power of 35W and subjected to fluorescent imaging.

6.3.5.2.4 Alkaline hydrolysis probing

Alkaline hydrolysis was performed under three different concentrations of NaOH with samples taken at different time points. 10 pmol of modified or unmodified substrate sequence was dissolved in 10 μ L of 10, 25, or 50 mM NaOH. The samples were incubated

at 95°C and 2 µL aliquots were taken from each sample at 1, 2.5- and 5-minutes time-points and quenched by adding to 3 µL of the stop solution and placing on ice. 2.5 µL of each time-point sample were resolved on a 20% analytical denaturing PAGE for 1 hour, under constant power of 35W, and subjected to fluorescent imaging.

6.3.5.2.5 Debranching resistance assay

5 pmol of 5'-fluorescein-labeled substrate sequence modified using FJ1 or FH14 was dissolved in 5 µL of Dbr1 reaction buffer including 1, 5, 10, or 100 µM MnCl₂ and 0.5 ng/µL of recombinant Dbr1. Reaction conditions were adopted from (Carrocci et al., 2017) with slight modifications. The sample was incubated at 37°C and 1 µL timepoint samples were taken at 0, 15, 30, and 60 minutes and quenched by adding to 4 µL of stop solution and placing in liquid nitrogen. 2.5 µL of each of these samples were then resolved on a 20% analytical denaturing PAGE for 1 hour under constant power of 25W. The gel was then subjected to fluorescent imaging.

6.3.6 Application of the ribozymes

6.3.6.1 FH14 catalyzed labeling of the in vitro transcribed 5S rRNA and RyhB RNA

Reactions were set up with 100 pmol of either of the FH14 variants and 10 pmol of their 3'-fluorescein-labeled of their cognate transcripts (5S rRNA or RyhB). The reactions were performed in 5 µL total volume of the selection buffer including 40 mM MgCl₂ and 200 µM N⁶-ATTO550-ATP. The reactions were incubated at 37°C. 1 µL aliquots were taken at 0, 2, and 4 hours and quenched by adding to 4 µL of stop solution. 2.5 µL of each of these samples was resolved on a 10% denaturing PAGE under constant power of 25 W for 1hour. The gel was then subjected to dual-channel fluorescent imaging.

For simultaneous labeling of target positions in 5S rRNA, reactions were set up with 5 pmol of the non-labeled 5S rRNA in vitro transcript and 50 pmol of FH14_5S-A45 alone, FH14_5S-A45 and FH14_5S-A66 or all three ribozymes together. The reactions were performed in 5 µL of the 1x selection buffer including 200 µM N⁶-ATTO550-ATP and 40 mM MgCl₂. The reactions were quenched by adding 20 µL of stop solution after 4h incubation at 37°C. 2.5 µL of each of the quenched samples were then resolved on a 0.4 mm,

10% denaturing PAGE next to the 3'-fluorescein-labeled in vitro transcribed 5S rRNA as size marker.

6.3.6.2 Labeling 5S rRNA and RyhB RNA in total cellular RNA context

50 ng of cellular RNA was therefore mixed with 50 pmol of either of the three FH14 ribozymes in a final volume of 5 μ L of the 1x selection buffer including MgCl₂ and 200 μ M N⁶-ATTO550-ATP. The reactions were incubated at 37°C for 4 hours before quenching by addition of 95 μ L of TEN buffer followed by ethanol precipitation. The pellet was then dissolved in 2 μ L of milliQ water and 2 μ L of stop solution. The whole sample from all three experiments were resolved on 10% analytical PAGE for 1 hour. 1 pmol of the in vitro transcribed 3'-fluorescein-labeled 5S rRNA was also resolved on a neighboring well as a size marker. The gel was subjected to dual-channel fluorescent imaging to reveal the bands.

6.3.6.3 Mutually orthogonal double labeling of the synthetic model substrate

6.3.6.3.1 Labeling reaction

20 pmol of the double labeling substrate sequence was mixed with 50 pmol of each ribozyme in a total volume of 5 μ L of selection buffer containing 40 mM MgCl₂ and 200 μ M of either Cy5-TenDP or 6-FAM-ATP or both. The FJ1 ribozyme was targeted towards the parent substrate sequence while the FH14 was designed to target the mutated segment. Time-point samples (0.5 μ L) were taken 0, 5 h, and after overnight incubation and quenched by adding to 99.5 μ L of TEN buffer. The samples were then subjected to ethanol precipitation, and the pellet was dissolved in 5 μ L of the stop solution. 2.5 μ L of these samples were then resolved on a 15% analytical denaturing PAGE for 45 minutes and subjected to dual-channel fluorescent imaging.

6.3.6.3.2 8-17NG mediated cleavage reaction

To perform the 8-17NG mediated cleavage experiment, an aliquot (3.5 μ L) of the labeling reactions explained above was subjected to ethanol precipitation. The pellet was then dissolved in 10 μ L of 1x 8-17NG reaction buffer including 10 mM MnCl₂, 10 mM MgCl₂, and 200 pmol of the corresponding 8-17NG deoxyribozyme. The reaction was then incubated at 37°C and 1.4 μ L timepoint samples were taken at 0, 0.5, and 1 hour and mixed

with 3.57 μL of the stop solution. 2.5 μL of these samples were resolved on the same gel as the double labeling reaction. And subjected to dual-channel fluorescent imaging.

6.3.6.4 FH14 and FJ1 mediated labeling of 16S and 23S rRNA

20 ng of total cellular RNA from Top10 *E. Coli* cells (NEB) was mixed with 50 pmol of the corresponding ribozyme(s). The reaction was performed in 5 μL of the selection buffer including 40 mM MgCl_2 and 200 μM of the fluorescently labeled TenDP and/or ATP analogs. After 6h incubation at 37°C, the reaction was precipitated twice using isopropanol. The resulting pellet was washed using ice cold 70% ethanol. The dried pellet was dissolved in 2 μL of stop solution and 1 μL of milliQ water. The sample was then resolved on 1.3% agarose gel under 90 V for 45 minutes. The gel was then subjected to dual-channel fluorescent imaging. SYBR Gold staining was then performed to reveal the rest of the total cellular RNA.

6.3.6.5 Superscript III primer extension experiment

6.3.6.5.1 Preparation of the 5S rRNA modified using FH14 and FJC9 ribozymes

20 pmol of the 5S rRNA transcript was mixed with 30 pmol of either of the three 5S rRNA targeting FH14 ribozymes or the 5S rRNA targeting FJC9 ribozyme in 5 μL of 1x selection buffer including 40 mM MgCl_2 . In the samples in which FH14 ribozymes were present N^6 -biotin-ATP was included at 200 μM and in the reaction in which FJC9 ribozyme was present N^6 -biotin-TenDP was used at 300 μM . The samples were incubated overnight at 37°C. Afterwards the reactions were ethanol precipitated and the dried pellet was dissolved in 5 μL of milliQ water.

6.3.6.5.2 Preparation of the 16S and 23S rRNA modified using FJ1 and FH14

200 ng of *E. coli* total cellular RNA and 20 pmol of individual ribozymes were dissolved in a total volume of 10 μL of the selection buffer including 40 mM MgCl_2 and the ribozyme's cognate substrate. For FH14 type ribozyme N^6 -biotin-ATP (200 μM final concentration) was used, and for the FJ1 type ribozyme Biotin-Tenofovir-DP (340 μM final concentration). The reactions were incubated at 37°C for 5 h and were subjected to ethanol precipitation. The pellet was dried and directly used in the primer extension reaction.

6.3.6.5.3 Primer extension protocol

For primer extension experiments on the in vitro transcribed 5S rRNA, 5 pmol of the ³²P-labeled primer and 5 pmol of the modified or non-modified transcript were annealed in 5 µL of the were dissolved in 5 µL of the annealing buffer (Tris 5 mM pH= 7.5, EDTA 0.1 mM). The sample was then placed at 95°C for 3 minutes followed by incubation at room temperature for 10 minutes. For the experiments performed on 16S and 23S rRNA, 5 pmol of the ³²P-labeled primer and 200 ng of the ribozyme-modified or non-modified total cellular RNA were annealed in 1x annealing buffer.

2 µL of the 5x First-strand synthesis buffer (Invitrogen), 0.5 µL of 0.1 M DTT, 0.5 µL of 10 mM dNTP mix and 50 units of superscript III reverse transcriptase was then added to the reactions and the final reaction volumes of were then adjusted to 10 µL by adding milliQ water. The reactions were then incubated at 55°C for 1 hour. 1 µL of 2 N NaOH was then added to each reaction and the samples were then incubated for 5 minutes at 95°C. The reactions were then ethanol precipitated. The dried pellets were dissolved in 10 µL of high dye loading buffer. 2.5 µL of each sample was resolved using 15%, analytical 45 cm long PAGE, under constant power of 45 W for 2 hours and 15 minutes. The gel was dried at 80°C under vacuum for 30 minutes and exposed to the phosphorus screen overnight. Scanning of the exposed screen was performed using Typhoon phosphorimager.

6.3.6.5.4 Sequencing ladder preparation

The sequencing reactions were performed according to (Shahn et al., 1989) with some modification. Reactions were set up by annealing 5 pmol of the ³²P-labeled primers with 5 pmol of the non-modified 5S rRNA transcript or with 200 ng of the non-modified total cellular RNA in 5 µL of the annealing buffer. for the A sequencing ladder ddTTP, was added to the reaction to a final concentration of 0.5 mM and dTTP at 0.05 mM. The rest of the dNTPs were adjusted to a final concentration of 0.5 mM. For G sequencing ladder ddCTP was included in the reaction at a concentration of 0.5 mM and dCTP at 0.05 mM. The rest of the dNTPs had a final concentration of 0.5 mM. All the other reagents such as the buffer conditions and enzyme units were the same as described for the primer extension protocol.

7 Bibliography

- Abudayyeh, O.O., Gootenberg, J.S., Essletzbichler, P., Han, S., Joung, J., Belanto, J.J., Verdine, V., Cox, D.B., Kellner, M.J., and Regev, A. (2017). RNA targeting with CRISPR–Cas13. *Nature* *550*, 280-284.
- Abudayyeh, O.O., Gootenberg, J.S., Konermann, S., Joung, J., Slaymaker, I.M., Cox, D.B., Shmakov, S., Makarova, K.S., Semenova, E., and Minakhin, L. (2016). C2c2 is a single-component programmable RNA-guided RNA-targeting CRISPR effector. *Science* *353*, aaf5573.
- Adamala, K.P., Martin-Alarcon, D.A., and Boyden, E.S. (2016). Programmable RNA-binding protein composed of repeats of a single modular unit. *Proceedings of the National Academy of Sciences* *113*, E2579-E2588.
- Agresti, J.J., Kelly, B.T., Jäschke, A., and Griffiths, A.D. (2005). Selection of ribozymes that catalyze multiple-turnover Diels–Alder cycloadditions by using in vitro compartmentalization. *Proceedings of the National Academy of Sciences* *102*, 16170-16175.
- Alam, K.K., Chang, J.L., and Burke, D.H. (2015). FASTAptamer: a bioinformatic toolkit for high-throughput sequence analysis of combinatorial selections. *Molecular Therapy–Nucleic Acids* *4*, e230.
- Alexander, S.C., Busby, K.N., Cole, C.M., Zhou, C.Y., and Devaraj, N.K. (2015). Site-specific covalent labeling of RNA by enzymatic transglycosylation. *Journal of the American Chemical Society* *137*, 12756-12759.
- Ameta, S., and Jäschke, A. (2013). An RNA catalyst that reacts with a mechanistic inhibitor of serine proteases. *Chemical Science* *4*, 957-964.
- Anhäuser, L., Hüwel, S., Zobel, T., and Rentmeister, A. (2019). Multiple covalent fluorescence labeling of eukaryotic mRNA at the poly (A) tail enhances translation and can be performed in living cells. *Nucleic acids research* *47*, e42-e42.
- Anhäuser, L., and Rentmeister, A. (2017). Enzyme-mediated tagging of RNA. *Current opinion in biotechnology* *48*, 69-76.
- Arora, A., Sunbul, M., and Jäschke, A. (2015). Dual-colour imaging of RNAs using quencher-and fluorophore-binding aptamers. *Nucleic acids research* *43*, e144-e144.
- Asare-Okai, P., Agustin, E., Fabris, D., and Royzen, M. (2014). Site-specific fluorescence labelling of RNA using bio-orthogonal reaction of trans-cyclooctene and tetrazine. *Chemical Communications* *50*, 7844-7847.
- Autour, A., Jeng, S.C., Cawte, A.D., Abdollahzadeh, A., Galli, A., Panchapakesan, S.S., Rueda, D., Ryckelynck, M., and Unrau, P.J. (2018). Fluorogenic RNA Mango aptamers for imaging small non-coding RNAs in mammalian cells. *Nature communications* *9*, 1-12.
- Babendure, J.R., Adams, S.R., and Tsien, R.Y. (2003). Aptamers switch on fluorescence of triphenylmethane dyes. *Journal of the American Chemical Society* *125*, 14716-14717.
- Baker, M. (2012). RNA imaging in situ. *nature methods* *9*, 787-790.
- Bartel, D.P., and Szostak, J.W. (1993). Isolation of new ribozymes from a large pool of random sequences [see comment]. *Science* *261*, 1411-1418.
- Basu, S., and Strobel, S.A. (1999). Thiophilic metal ion rescue of phosphorothioate interference within the Tetrahymena ribozyme P4–P6 domain. *Rna* *5*, 1399-1407.
- Baum, D.A., and Silverman, S.K. (2007). Deoxyribozyme-catalyzed labeling of RNA. *Angewandte Chemie International Edition* *46*, 3502-3504.
- Bédard, A.-S.V., Hien, E.D., and Lafontaine, D.A. (2020). Riboswitch regulation mechanisms: RNA, metabolites and regulatory proteins. *Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms*, 194501.
- Ben-Ari, Y.a., Brody, Y., Kinor, N., Mor, A., Tsukamoto, T., Spector, D.L., Singer, R.H., and Shav-Tal, Y. (2010). The life of an mRNA in space and time. *Journal of cell science* *123*, 1761-1774.

- Berry, K.E., and Hochschild, A. (2018). A bacterial three-hybrid assay detects *Escherichia coli* Hfq-sRNA interactions in vivo. *Nucleic acids research* *46*, e12-e12.
- Bertrand, E., Chartrand, P., Schaefer, M., Shenoy, S.M., Singer, R.H., and Long, R.M. (1998). Localization of ASH1 mRNA particles in living yeast. *Molecular cell* *2*, 437-445.
- Biondi, E., Maxwell, A.W., and Burke, D.H. (2012). A small ribozyme with dual-site kinase activity. *Nucleic acids research* *40*, 7528-7540.
- Biondi, E., Nickens, D.G., Warren, S., Saran, D., and Burke, D.H. (2010). Convergent donor and acceptor substrate utilization among kinase ribozymes. *Nucleic acids research* *38*, 6785-6795.
- Bothma, J.P., Garcia, H.G., Ng, S., Perry, M.W., Gregor, T., and Levine, M. (2015). Enhancer additivity and non-additivity are determined by enhancer strength in the *Drosophila* embryo. *Elife* *4*, e07956.
- Bouhedda, F., Fam, K.T., Collot, M., Autour, A., Marzi, S., Klymchenko, A., and Ryckelynck, M. (2020). A dimerization-based fluorogenic dye-aptamer module for RNA imaging in live cells. *Nature chemical biology* *16*, 69-76.
- Bowman, J.C., Lenz, T.K., Hud, N.V., and Williams, L.D. (2012). Cations in charge: magnesium ions in RNA folding and catalysis. *Current opinion in structural biology* *22*, 262.
- Brandsen, B.M., Hesser, A.R., Castner, M.A., Chandra, M., and Silverman, S.K. (2013). DNA-catalyzed hydrolysis of esters and aromatic amides. *Journal of the American Chemical Society* *135*, 16014-16017.
- Braselmann, E., Wierzba, A.J., Polaski, J.T., Chromiński, M., Holmes, Z.E., Hung, S.-T., Batan, D., Wheeler, J.R., Parker, R., and Jimenez, R. (2018). A multicolor riboswitch-based platform for imaging of RNA in live mammalian cells. *Nature chemical biology* *14*, 964-971.
- Büttner, L., Javadi-Zarnaghi, F., and Höbartner, C. (2014). Site-specific labeling of RNA at internal ribose hydroxyl groups: terbium-assisted deoxyribozymes at work. *Journal of the American Chemical Society* *136*, 8131-8137.
- Büttner, L., Seikowski, J., Wawrzyniak, K., Ochmann, A., and Höbartner, C. (2013). Synthesis of spin-labeled riboswitch RNAs using convertible nucleosides and DNA-catalyzed RNA ligation. *Bioorganic & medicinal chemistry* *21*, 6171-6180.
- Carmi, N., Shultz, L.A., and Breaker, R.R. (1996). In vitro selection of self-cleaving DNAs. *Chemistry & biology* *3*, 1039-1046.
- Carrocci, T.J., and Hoskins, A.A. (2014). Imaging of RNAs in live cells with spectrally diverse small molecule fluorophores. *Analyst* *139*, 44-47.
- Carrocci, T.J., Lohe, L., Ashton, M.J., Höbartner, C., and Hoskins, A.A. (2017). Debranchase-resistant labeling of RNA using the 10DM24 deoxyribozyme and fluorescent modified nucleotides. *Chemical Communications* *53*, 11992-11995.
- Cawte, A.D., Unrau, P.J., and Rueda, D.S. (2020). Live cell imaging of single RNA molecules with fluorogenic Mango II arrays. *Nature Communications* *11*, 1-11.
- Cech, T.R. (2012). The RNA worlds in context. *Cold Spring Harbor perspectives in biology* *4*, a006742.
- Cernak, P., and Sen, D. (2013). A thiamin-utilizing ribozyme decarboxylates a pyruvate-like substrate. *Nature chemistry* *5*, 971.
- Chapman, K.B., and Boeke, J.D. (1991). Isolation and characterization of the gene encoding yeast debranching enzyme. *Cell* *65*, 483-492.
- Chen, B., Gilbert, L.A., Cimini, B.A., Schnitzbauer, J., Zhang, W., Li, G.-W., Park, J., Blackburn, E.H., Weissman, J.S., and Qi, L.S. (2013). Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system. *Cell* *155*, 1479-1491.
- Cheong, C.-G., and Hall, T.M.T. (2006). Engineering RNA sequence specificity of Pumilio repeats. *Proceedings of the National Academy of Sciences* *103*, 13635-13639.

- Chumachenko, N., Novikov, Y., and Yarus, M. (2009). Rapid and simple ribozymic aminoacylation using three conserved nucleotides. *Journal of the American Chemical Society* *131*, 5257-5263.
- Chun, S.-M., Jeong, S., Kim, J.-M., Chong, B.-O., Park, Y.-K., Park, H., and Yu, J. (1999). Cholesterol esterase activity by in vitro selection of RNA against a phosphate transition-state analogue. *Journal of the American Chemical Society* *121*, 10844-10845.
- Cihlar, T., Birkus, G., Greenwalt, D.E., and Hitchcock, M.J. (2002). Tenofovir exhibits low cytotoxicity in various human cell types: comparison with other nucleoside reverse transcriptase inhibitors. *Antiviral research* *54*, 37-45.
- Coleman, T.M., and Huang, F. (2002). RNA-catalyzed thioester synthesis. *Chemistry & biology* *9*, 1227-1236.
- Coppins, R.L., and Silverman, S.K. (2004a). A DNA enzyme that mimics the first step of RNA splicing. *Nature structural & molecular biology* *11*, 270-274.
- Coppins, R.L., and Silverman, S.K. (2004b). Rational Modification of a Selection Strategy Leads to Deoxyribozymes that Create Native 3' -5' RNA Linkages. *Journal of the American Chemical Society* *126*, 16426-16432.
- Coppins, R.L., and Silverman, S.K. (2005). A deoxyribozyme that forms a three-helix-junction complex with its RNA substrates and has general RNA branch-forming activity. *Journal of the American Chemical Society* *127*, 2900-2907.
- Curanovic, D., Cohen, M., Singh, I., Slagle, C.E., Leslie, C.S., and Jaffrey, S.R. (2013). Global profiling of stimulus-induced polyadenylation in cells using a poly (A) trap. *Nature chemical biology* *9*, 671.
- Curtis, E.A., and Bartel, D.P. (2005). New catalytic structures from an existing ribozyme. *Nature structural & molecular biology* *12*, 994-1000.
- Dahariya, S., Paddibhatla, I., Kumar, S., Raghuwanshi, S., Palapati, A., and Gutti, R.K. (2019). Long non-coding RNA: Classification, biogenesis and functions in blood cells. *Molecular immunology* *112*, 82-92.
- Darzacq, X., Shav-Tal, Y., De Turrís, V., Brody, Y., Shenoy, S.M., Phair, R.D., and Singer, R.H. (2007). In vivo dynamics of RNA polymerase II transcription. *Nature structural & molecular biology* *14*, 796-806.
- De Clercq, E., and Holý, A. (2005). Acyclic nucleoside phosphonates: a key class of antiviral drugs. *Nature Reviews Drug Discovery* *4*, 928-940.
- De la Peña, M., García-Robles, I., and Cervera, A. (2017). The hammerhead ribozyme: a long history for a short RNA. *Molecules* *22*, 78.
- Deng, W., Shi, X., Tjian, R., Lionnet, T., and Singer, R.H. (2015). CASFISH: CRISPR/Cas9-mediated in situ labeling of genomic loci in fixed cells. *Proceedings of the National Academy of Sciences* *112*, 11870-11875.
- Dokukin, V., and Silverman, S.K. (2014). A modular tyrosine kinase deoxyribozyme with discrete aptamer and catalyst domains. *Chemical Communications* *50*, 9317-9320.
- Dolan, G.F., Akoopie, A., and Müller, U.F. (2015). A faster triphosphorylation ribozyme. *PLoS One* *10*.
- Dolgosheina, E.V., Jeng, S.C., Panchapakesan, S.S.S., Cojocar, R., Chen, P.S., Wilson, P.D., Hawkins, N., Wiggins, P.A., and Unrau, P.J. (2014). RNA mango aptamer-fluorophore: a bright, high-affinity complex for RNA labeling and tracking. *ACS chemical biology* *9*, 2412-2420.
- Dong, S., Wang, Y., Cassidy-Amstutz, C., Lu, G., Bigler, R., Jezyk, M.R., Li, C., Hall, T.M.T., and Wang, Z. (2011). Specific and modular binding code for cytosine recognition in Pumilio/FBF (PUF) RNA-binding domains. *Journal of Biological Chemistry* *286*, 26732-26742.
- Duffy, E.E., Schofield, J.A., and Simon, M.D. (2019). Gaining insight into transcriptome-wide RNA population dynamics through the chemistry of 4-thiouridine. *Wiley Interdisciplinary Reviews: RNA* *10*, e1513.

- Dunn, M.R., Jimenez, R.M., and Chaput, J.C. (2017). Analysis of aptamer discovery and technology. *Nature Reviews Chemistry* *1*, 1-16.
- Dutta, T., and Srivastava, S. (2018). Small RNA-mediated regulation in bacteria: a growing palette of diverse mechanisms. *Gene* *656*, 60-72.
- Fatica, A., and Bozzoni, I. (2014). Long non-coding RNAs: new players in cell differentiation and development. *Nature Reviews Genetics* *15*, 7-21.
- Feldman, A.W., Fischer, E.C., Ledbetter, M.P., Liao, J.-Y., Chaput, J.C., and Romesberg, F.E. (2018). A tool for the import of natural and unnatural nucleoside triphosphates into bacteria. *Journal of the American Chemical Society* *140*, 1447-1454.
- Felletti, M., Stifel, J., Wurmthaler, L.A., Geiger, S., and Hartig, J.S. (2016). Twister ribozymes as highly versatile expression platforms for artificial riboswitches. *Nature communications* *7*, 1-8.
- Fiamengo, R., and Jäschke, A. (2005). Nucleic acid enzymes. *Current opinion in biotechnology* *16*, 614-621.
- Filonov, G.S., Kam, C.W., Song, W., and Jaffrey, S.R. (2015). In-gel imaging of RNA processing using broccoli reveals optimal aptamer expression strategies. *Chemistry & biology* *22*, 649-660.
- Filonov, G.S., Moon, J.D., Svendsen, N., and Jaffrey, S.R. (2014). Broccoli: rapid selection of an RNA mimic of green fluorescent protein by fluorescence-based selection and directed evolution. *Journal of the American Chemical Society* *136*, 16299-16308.
- Fusz, S., Eisenführ, A., Srivatsan, S.G., Heckel, A., and Famulok, M. (2005). A ribozyme for the aldol reaction. *Chemistry & biology* *12*, 941-950.
- Gallagher, S.S., Miller, L.W., and Cornish, V.W. (2007). An orthogonal dexamethasone-trimethoprim yeast three-hybrid system. *Analytical biochemistry* *363*, 160.
- Gallant, J.E., and Deresinski, S. (2003). Tenofovir disoproxil fumarate. *Clinical Infectious Diseases* *37*, 944-950.
- George, L., Indig, F.E., Abdelmohsen, K., and Gorospe, M. (2018). Intracellular RNA-tracking methods. *Royal Society Open Biology* *8*, 180104.
- Gilbert, W. (1986). Origin of life: The RNA world. *nature* *319*, 618-618.
- Golding, I., Paulsson, J., Zawilski, S.M., and Cox, E.C. (2005). Real-time kinetics of gene activity in individual bacteria. *Cell* *123*, 1025-1036.
- González, M.n., Argaraña, C.E., and Fidelio, G.D. (1999). Extremely high thermal stability of streptavidin and avidin upon biotin binding. *Biomolecular engineering* *16*, 67-72.
- Grammel, M., Hang, H., and Conrad, N.K. (2012). Chemical reporters for monitoring RNA synthesis and poly (A) tail dynamics. *ChemBioChem* *13*, 1112-1115.
- Griffiths, A.D., and Tawfik, D.S. (2000). Man-made enzymes—from design to in vitro compartmentalisation. *Current opinion in biotechnology* *11*, 338-353.
- Grünwald, D., and Singer, R.H. (2010). In vivo imaging of labelled endogenous β -actin mRNA during nucleocytoplasmic transport. *Nature* *467*, 604-607.
- Guerrier-Takada, C., Gardiner, K., Marsh, T., Pace, N., and Altman, S. (1983). The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme. *Cell* *35*, 849-857.
- Hacker, S.M., Welter, M., and Marx, A. (2015). Synthesis of γ -Phosphate-Labeled and Doubly Labeled Adenosine Triphosphate Analogs. *Current protocols in nucleic acid chemistry* *60*, 13.14. 11-13.14. 25.
- Hamashima, K., Kimoto, M., and Hirao, I. (2018). Creation of unnatural base pairs for genetic alphabet expansion toward synthetic xenobiology. *Current opinion in chemical biology* *46*, 108-114.
- Hampel, A., and Tritz, R. (1989). RNA catalytic properties of the minimum (-) sTRSV sequence. *Biochemistry* *28*, 4929-4933.
- Hiller, Y., Bayer, E.A., and Wilchek, M. (1990). [6] Nonglycosylated avidin. In *Methods in enzymology* (Elsevier), pp. 68-70.

- Höbartner, C., Pradeepkumar, P., and Silverman, S.K. (2007). Site-selective depurination by a periodate-dependent deoxyribozyme. *Chemical communications*, 2255-2257.
- Höbartner, C., and Silverman, S.K. (2007). Engineering a Selective Small-Molecule Substrate Binding Site into a Deoxyribozyme. *Angewandte Chemie International Edition* 46, 7420-7424.
- Hocine, S., Raymond, P., Zenklusen, D., Chao, J.A., and Singer, R.H. (2013). Single-molecule analysis of gene expression using two-color RNA labeling in live yeast. *Nature methods* 10, 119-121.
- Hofacker, I.L. (2003). Vienna RNA secondary structure server. *Nucleic acids research* 31, 3429-3431.
- Holeman, L.A., Robinson, S.L., Szostak, J.W., and Wilson, C. (1998). Isolation and characterization of fluorophore-binding RNA aptamers. *Folding and Design* 3, 423-431.
- Holmberg, A., Blomstergren, A., Nord, O., Lukacs, M., Lundeborg, J., and Uhlén, M. (2005). The biotin-streptavidin interaction can be reversibly broken using water at elevated temperatures. *Electrophoresis* 26, 501-510.
- Holstein, J.M., Anhäuser, L., and Rentmeister, A. (2016). Modifying the 5'-cap for click reactions of eukaryotic mRNA and to tune translation efficiency in living cells. *Angewandte Chemie International Edition* 55, 10899-10903.
- Holstein, J.M., and Rentmeister, A. (2016). Current covalent modification methods for detecting RNA in fixed and living cells. *Methods* 98, 18-25.
- Holstein, J.M., Schulz, D., and Rentmeister, A. (2014). Bioorthogonal site-specific labeling of the 5'-cap structure in eukaryotic mRNAs. *Chemical Communications* 50, 4478-4481.
- Holstein, J.M., Stummer, D., and Rentmeister, A. (2015). Enzymatic modification of 5'-capped RNA with a 4-vinylbenzyl group provides a platform for photoclick and inverse electron-demand Diels–Alder reaction. *Chemical science* 6, 1362-1369.
- Huang, F., Bugg, C.W., and Yarus, M. (2000). RNA-catalyzed CoA, NAD, and FAD synthesis from phosphopantetheine, NMN, and FMN. *Biochemistry* 39, 15548-15555.
- Huang, F., and Yarus, M. (1997a). 5'-RNA Self-Capping from Guanosine Diphosphate. *Biochemistry* 36, 6557-6563.
- Huang, F., and Yarus, M. (1997b). Versatile 5' phosphoryl coupling of small and large molecules to an RNA. *Proceedings of the National Academy of Sciences* 94, 8965-8969.
- Huang, X., Zhao, Y., Pu, Q., Liu, G., Peng, Y., Wang, F., Chen, G., Sun, M., Du, F., and Dong, J. (2019). Intracellular selection of trans-cleaving hammerhead ribozymes. *Nucleic acids research* 47, 2514-2522.
- Humbert, N., Zocchi, A., and Ward, T.R. (2005). Electrophoretic behavior of streptavidin complexed to a biotinylated probe: a functional screening assay for biotin-binding proteins. *Electrophoresis* 26, 47-52.
- Hutchins, C.J., Rathjen, P.D., Forster, A.C., and Symons, R.H. (1986). Self-cleavage of plus and minus RNA transcripts of avocado sunblotch viroid. *Nucleic acids research* 14, 3627-3640.
- Igloi, G.L. (1988). Interaction of tRNAs and of phosphorothioate-substituted nucleic acids with an organomercurial. Probing the chemical environment of thiolated residues by affinity electrophoresis. *Biochemistry* 27, 3842-3849.
- Illangasekare, M., Sanchez, G., Nickles, T., and Yarus, M. (1995). Aminoacyl-RNA synthesis catalyzed by an RNA. *Science* 267, 643-647.
- Ishida, S., Terasaka, N., Katoh, T., and Suga, H. (2020). An aminoacylation ribozyme evolved from a natural tRNA-sensing T-box riboswitch. *Nature Chemical Biology*, 1-8.
- Jadhav, V.R., and Yarus, M. (2002). Acyl-CoAs from coenzyme ribozymes. *Biochemistry* 41, 723-729.
- Jaksik, R., Iwanaszko, M., Rzeszowska-Wolny, J., and Kimmel, M. (2015). Microarray experiments and factors which affect their reliability. *Biology direct* 10, 46.

- Jäschke, A. (2001). Artificial ribozymes and deoxyribozymes. *Current opinion in structural biology* *11*, 321-326.
- Jäschke, A., and Seelig, B. (2000). Evolution of DNA and RNA as catalysts for chemical reactions. *Current opinion in chemical biology* *4*, 257-262.
- Jenne, A., and Famulok, M. (1998). A novel ribozyme with ester transferase activity. *Chemistry & biology* *5*, 23-34.
- Jimenez, R.M., Polanco, J.A., and Lupták, A. (2015). Chemistry and biology of self-cleaving ribozymes. *Trends in biochemical sciences* *40*, 648-661.
- Joyce, G.F. (2004). Directed evolution of nucleic acid enzymes. *Annual review of biochemistry* *73*, 791-836.
- Kaikkonen, M.U., Lam, M.T., and Glass, C.K. (2011). Non-coding RNAs as regulators of gene expression and epigenetics. *Cardiovascular research* *90*, 430-440.
- Kang, T.J., and Suga, H. (2007). In vitro selection of a 5'-purine ribonucleotide transferase ribozyme. *Nucleic acids research* *35*, 4186-4194.
- Kellermann, S.J., Rath, A.K., and Rentmeister, A. (2013). Tetramolecular fluorescence complementation for detection of specific RNAs in vitro. *ChemBioChem* *14*, 200-204.
- Kellermann, S.J., and Rentmeister, A. (2016). A genetically encodable system for sequence-specific detection of RNAs in two colors. *Chembiochem* *17*, 895-899.
- Kellermann, S.J., and Rentmeister, A. (2017). A FACS-based screening strategy to assess sequence-specific RNA-binding of Pumilio protein variants in *E. coli*. *Biological chemistry* *398*, 69-75.
- Kilchert, C., and Vasiljeva, L. (2013). mRNA quality control goes transcriptional (Portland Press Ltd.).
- Köhn, M., Lederer, M., Wächter, K., and Hüttelmaier, S. (2010). Near-infrared (NIR) dye-labeled RNAs identify binding of ZBP1 to the noncoding Y3-RNA. *Rna* *16*, 1420-1428.
- Kruger, K., Grabowski, P.J., Zaug, A.J., Sands, J., Gottschling, D.E., and Cech, T.R. (1982). Self-splicing RNA: autoexcision and autocyclization of the ribosomal RNA intervening sequence of *Tetrahymena*. *cell* *31*, 147-157.
- Kukurba, K.R., and Montgomery, S.B. (2015). RNA sequencing and analysis. *Cold Spring Harbor Protocols* *2015*, pdb. top084970.
- Kumar, R.K., and Yarus, M. (2001). RNA-catalyzed amino acid activation. *Biochemistry* *40*, 6998-7004.
- Larson, D.R., Zenklusen, D., Wu, B., Chao, J.A., and Singer, R.H. (2011). Real-time observation of transcription initiation and elongation on an endogenous yeast gene. *science* *332*, 475-478.
- Laurent, G.S., Wahlestedt, C., and Kapranov, P. (2015). The Landscape of long noncoding RNA classification. *Trends in Genetics* *31*, 239-251.
- Lawrence, J.B., and Singer, R.H. (1986). Intracellular localization of messenger RNAs for cytoskeletal proteins. *Cell* *45*, 407-415.
- Levy, M., Griswold, K.E., and Ellington, A.D. (2005). Direct selection of trans-acting ligase ribozymes by in vitro compartmentalization. *Rna* *11*, 1555-1562.
- Li, F., Dong, J., Hu, X., Gong, W., Li, J., Shen, J., Tian, H., and Wang, J. (2015). A covalent approach for site-specific RNA Labeling in mammalian cells. *Angewandte Chemie International Edition* *54*, 4597-4602.
- Li, N., and Huang, F. (2005). Ribozyme-catalyzed aminoacylation from CoA thioesters. *Biochemistry* *44*, 4582-4590.
- Lionnet, T., Czaplinski, K., Darzacq, X., Shav-Tal, Y., Wells, A.L., Chao, J.A., Park, H.Y., De Turris, V., Lopez-Jones, M., and Singer, R.H. (2011). A transgenic mouse for in vivo detection of endogenous labeled mRNA. *Nature methods* *8*, 165.
- Lo, A., and Qi, L. (2017). Genetic and epigenetic control of gene expression by CRISPR-Cas systems. *F1000Research* *6*.

- Loda, A., and Heard, E. (2019). Xist RNA in action: Past, present, and future. *PLoS genetics* *15*.
- Lohse, P.A., and Szostak, J.W. (1996). Ribozyme-catalysed amino-acid transfer reactions. *Nature* *381*, 442-444.
- Lorsch, J.R., and Szostak, J.W. (1994). In vitro evolution of new ribozymes with polynucleotide kinase activity. *Nature* *371*, 31-36.
- Lorsch, J.R., and Szostak, J.W. (1995). Kinetic and thermodynamic characterization of the reaction catalyzed by a polynucleotide kinase ribozyme. *Biochemistry* *34*, 15315-15327.
- Lorsch, J.R., and Szostak, J.W. (1996). Chance and necessity in the selection of nucleic acid catalysts. *Accounts of chemical research* *29*, 103-110.
- Lukinavičius, G., Lapienė, V., Staševskij, Z., Dalhoff, C., Weinhold, E., and Klimašauskas, S. (2007). Targeted labeling of DNA by methyltransferase-directed transfer of activated groups (mTAG). *Journal of the American Chemical Society* *129*, 2758-2759.
- Maguire, M.E., and Cowan, J.A. (2002). Magnesium chemistry and biochemistry. *Biometals* *15*, 203-210.
- Martell, R.E., Nevins, J.R., and Sullenger, B.A. (2002). Optimizing aptamer activity for gene therapy applications using expression cassette SELEX. *Molecular Therapy* *6*, 30-34.
- Massé, E., and Gottesman, S. (2002). A small RNA regulates the expression of genes involved in iron metabolism in *Escherichia coli*. *Proceedings of the National Academy of Sciences* *99*, 4620-4625.
- Massé, E., Vanderpool, C.K., and Gottesman, S. (2005). Effect of RyhB small RNA on global iron use in *Escherichia coli*. *Journal of bacteriology* *187*, 6962-6971.
- McDonald, R.I., Guilinger, J.P., Mukherji, S., Curtis, E.A., Lee, W.I., and Liu, D.R. (2014). Electrophilic activity-based RNA probes reveal a self-alkylating RNA for RNA labeling. *Nature chemical biology* *10*, 1049.
- McGinness, K.E., and Joyce, G.F. (2003). In search of an RNA replicase ribozyme. *Chemistry & biology* *10*, 5-14.
- Miller, B.R., Wei, T., Fields, C.J., Sheng, P., and Xie, M. (2018). Near-infrared fluorescent northern blot. *RNA* *24*, 1871-1877.
- Milligan, J.F., Groebe, D.R., Witherell, G.W., and Uhlenbeck, O.C. (1987). Oligoribonucleotide synthesis using T7 RNA polymerase and synthetic DNA templates. *Nucleic acids research* *15*, 8783-8798.
- Monroy-Contreras, R., and Vaca, L. (2011). Molecular beacons: powerful tools for imaging RNA in living cells. *Journal of nucleic acids* *2011*.
- Montpetit, A., Payant, C., Nolan, J., and Brakier-Gingras, L. (1998). Analysis of the conformation of the 3'major domain of *Escherichia coli* 16S ribosomal RNA using site-directed photoaffinity crosslinking. *RNA* *4*, 1455-1466.
- Moretti, J.E., and Müller, U.F. (2014). A ribozyme that triphosphorylates RNA 5'-hydroxyl groups. *Nucleic acids research* *42*, 4767-4778.
- Motorin, Y., Burhenne, J., Teimer, R., Koynov, K., Willnow, S., Weinhold, E., and Helm, M. (2011). Expanding the chemical scope of RNA: methyltransferases to site-specific alkynylation of RNA for click labeling. *Nucleic acids research* *39*, 1943-1952.
- Müller, S., Appel, B., Balke, D., Hieronymus, R., and Nübel, C. (2016). Thirty-five years of research into ribozymes and nucleic acid catalysis: where do we stand today? *F1000Research* *5*.
- Murakami, H., Ohta, A., Ashigai, H., and Suga, H. (2006). A highly flexible tRNA acylation method for non-natural polypeptide synthesis. *Nature Methods* *3*, 357-359.
- Murakami, H., Saito, H., and Suga, H. (2003). A versatile tRNA aminoacylation catalyst based on RNA. *Chemistry & biology* *10*, 655-662.
- Murata, A., Sato, S.-i., Kawazoe, Y., and Uesugi, M. (2011). Small-molecule fluorescent probes for specific RNA targets. *Chemical Communications* *47*, 4712-4714.
- Murphy, E. (2000). Mysteries of magnesium homeostasis (*Am Heart Assoc*).

- Muthmann, N., Hartstock, K., and Rentmeister, A. (2020). Chemo-enzymatic treatment of RNA to facilitate analyses. *Wiley Interdisciplinary Reviews: RNA 11*, e1561.
- Muttach, F., Muthmann, N., Reichert, D., Anhäuser, L., and Rentmeister, A. (2017). A benzylic linker promotes methyltransferase catalyzed norbornene transfer for rapid bioorthogonal tetrazine ligation. *Chemical science 8*, 7947-7953.
- Naesens, L., Bischofberger, N., Augustijns, P., Annaert, P., Van den Mooter, G., Arimilli, M.N., Kim, C.U., and De Clercq, E. (1998). Antiretroviral efficacy and pharmacokinetics of oral bis (isopropoxyloxycarbonyloxymethyl) 9-(2-phosphonylmethoxypropyl) adenine in mice. *Antimicrobial agents and chemotherapy 42*, 1568-1573.
- Nelles, D.A., Fang, M.Y., O'Connell, M.R., Xu, J.L., Markmiller, S.J., Doudna, J.A., and Yeo, G.W. (2016). Programmable RNA tracking in live cells with CRISPR/Cas9. *Cell 165*, 488-496.
- Neubacher, S., and Hennig, S. (2019). RNA Structure and Cellular Applications of Fluorescent Light-Up Aptamers. *Angewandte Chemie International Edition 58*, 1266-1279.
- Nieuwlandt, D., West, M., Cheng, X., Kirshenheuter, G., and Eaton, B.E. (2003). The first example of an RNA urea synthase: selection through the enzyme active site of human neutrophil elastase. *ChemBioChem 4*, 651-654.
- Nissen, P., Hansen, J., Ban, N., Moore, P.B., and Steitz, T.A. (2000). The structural basis of ribosome activity in peptide bond synthesis. *Science 289*, 920-930.
- Noller, H.F. (1984). Structure of ribosomal RNA. *Annual review of biochemistry 53*, 119-162.
- Noller, H.F., Hoffarth, V., and Zimniak, L. (1992). Unusual resistance of peptidyl transferase to protein extraction procedures. *Science 256*, 1416-1419.
- Nomura, Y., and Yokobayashi, Y. (2019). Systematic minimization of RNA ligase ribozyme through large-scale design-synthesis-sequence cycles. *Nucleic acids research 47*, 8950-8960.
- Nygren, J., Svanvik, N., and Kubista, M. (1998). The interactions between the fluorescent dye thiazole orange and DNA. *Biopolymers: Original Research on Biomolecules 46*, 39-51.
- O'Brien, J., Hayder, H., Zayed, Y., and Peng, C. (2018). Overview of microRNA biogenesis, mechanisms of actions, and circulation. *Frontiers in endocrinology 9*, 402.
- Ogawa, A., and Maeda, M. (2008). An artificial aptazyme-based riboswitch and its cascading system in *E. coli*. *ChemBioChem 9*, 206-209.
- Ohuchi, M., Murakami, H., and Suga, H. (2007). The flexizyme system: a highly flexible tRNA aminoacylation tool for the translation apparatus. *Current opinion in chemical biology 11*, 537-542.
- Osawa, T., Kimura, S., Terasaka, N., Inanaga, H., Suzuki, T., and Numata, T. (2011). Structural basis of tRNA agmatinylation essential for AUA codon decoding. *Nature structural & molecular biology 18*, 1275.
- Ozawa, T., Natori, Y., Sato, M., and Umezawa, Y. (2007). Imaging dynamics of endogenous mitochondrial RNA in single living cells. *Nature methods 4*, 413-419.
- Paige, J.S., Wu, K.Y., and Jaffrey, S.R. (2011). RNA mimics of green fluorescent protein. *Science 333*, 642-646.
- Park, H.Y., Lim, H., Yoon, Y.J., Follenzi, A., Nwokafor, C., Lopez-Jones, M., Meng, X., and Singer, R.H. (2014). Visualization of dynamics of single endogenous mRNA labeled in live mouse. *Science 343*, 422-424.
- Park, S.V., Yang, J.-S., Jo, H., Kang, B., Oh, S.S., and Jung, G.Y. (2019). Catalytic RNA, ribozyme, and its applications in synthetic biology. *Biotechnology advances*, 107452.
- Park, S.Y., Moon, H.C., and Park, H.Y. (2020). Live-cell imaging of single mRNA dynamics using split superfolder green fluorescent proteins with minimal background. *RNA 26*, 101-109.
- Paudel, B.P., Fiorini, E., Börner, R., Sigel, R.K., and Rueda, D.S. (2018). Optimal molecular crowding accelerates group II intron folding and maximizes catalysis. *Proceedings of the National Academy of Sciences 115*, 11917-11922.
- Peebles, C.L., Perlman, P., Mecklenburg, K., Petrillo, M., Tabor, J., Jarrell, K., and Cheng, H.-L. (1986). A self-splicing RNA excises an intron lariat. *Cell 44*, 213-223.

- Petrov, A.S., Bernier, C.R., Hershkovits, E., Xue, Y., Waterbury, C.C., Hsiao, C., Stepanov, V.G., Gaucher, E.A., Grover, M.A., and Harvey, S.C. (2013). Secondary structure and domain architecture of the 23S and 5S rRNAs. *Nucleic acids research* *41*, 7522-7535.
- Poltronieri, P., Sun, B., and Mallardo, M. (2015). RNA viruses: RNA roles in pathogenesis, coreplication and viral load. *Current genomics* *16*, 327-335.
- Ponce-Salvatierra, A., Wawrzyniak-Turek, K., Steuerwald, U., Höbartner, C., and Pena, V. (2016). Crystal structure of a DNA catalyst. *Nature* *529*, 231-234.
- Ponchon, L., and Dardel, F. (2007). Recombinant RNA technology: the tRNA scaffold. *Nature methods* *4*, 571-576.
- Porter, E.B., Polaski, J.T., Morck, M.M., and Batey, R.T. (2017). Recurrent RNA motifs as scaffolds for genetically encodable small-molecule biosensors. *Nature chemical biology* *13*, 295.
- Poudyal, R.R., Nguyen, P.D., Lokugamage, M.P., Callaway, M.K., Gavette, J.V., Krishnamurthy, R., and Burke, D.H. (2017). Nucleobase modification by an RNA enzyme. *Nucleic acids research* *45*, 1345-1354.
- Pressman, A.D., Liu, Z., Janzen, E., Blanco, C., Müller, U.F., Joyce, G.F., Pascal, R., and Chen, I.A. (2019). Mapping a systematic ribozyme fitness landscape reveals a frustrated evolutionary network for self-aminoacylating RNA. *Journal of the American Chemical Society* *141*, 6213-6223.
- Prody, G.A., Bakos, J.T., Buzayan, J.M., Schneider, I.R., and Bruening, G. (1986). Autolytic processing of dimeric plant virus satellite RNA. *Science* *231*, 1577-1580.
- Prudent, J.R., Uno, T., and Schultz, P.G. (1994). Expanding the Scope of RNA Catalysis. *SCIENCE* *264*, 24.
- Qi, L.S., Larson, M.H., Gilbert, L.A., Doudna, J.A., Weissman, J.S., Arkin, A.P., and Lim, W.A. (2013). Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell* *152*, 1173-1183.
- Qu, D., Zhou, L., Wang, W., Wang, Z., Wang, G., Chi, W., and Zhang, B. (2013). 5-Ethynylcytidine as a new agent for detecting RNA synthesis in live cells by “click” chemistry. *Analytical biochemistry* *434*, 128-135.
- Rau, K., and Rentmeister, A. (2017). Making the message clear: concepts for mRNA imaging. *ACS central science* *3*, 701-707.
- Ray, A.S., Fordyce, M.W., and Hitchcock, M.J. (2016). Tenofovir alafenamide: a novel prodrug of tenofovir for the treatment of human immunodeficiency virus. *Antiviral research* *125*, 63-70.
- Robertson, M.P., and Ellington, A. (1999). In vitro selection of an allosteric ribozyme that transduces analytes to amplicons. *Nature biotechnology* *17*, 62-66.
- Rogers, J., and Joyce, G.F. (2001). The effect of cytidine on the structure and function of an RNA ligase ribozyme. *Rna* *7*, 395-404.
- Romani, A.M. (2011). Cellular magnesium homeostasis. *Archives of biochemistry and biophysics* *512*, 1-23.
- Roth, A., Weinberg, Z., Chen, A.G., Kim, P.B., Ames, T.D., and Breaker, R.R. (2014). A widespread self-cleaving ribozyme class is revealed by bioinformatics. *Nature chemical biology* *10*, 56.
- Saito, H., Kourouklis, D., and Suga, H. (2001). An in vitro evolved precursor tRNA with aminoacylation activity. *The EMBO Journal* *20*, 1797-1806.
- Samanta, B., Horning, D.P., and Joyce, G.F. (2018). 3'-End labeling of nucleic acids by a polymerase ribozyme. *Nucleic acids research* *46*, e103-e103.
- Santoro, S.W., and Joyce, G.F. (1997). A general purpose RNA-cleaving DNA enzyme. *Proceedings of the national academy of sciences* *94*, 4262-4266.
- Saran, D., Nickens, D.G., and Burke, D.H. (2005). A trans acting ribozyme that phosphorylates exogenous RNA. *Biochemistry* *44*, 15007-15016.

- Sato, S.i., Watanabe, M., Katsuda, Y., Murata, A., Wang, D.O., and Uesugi, M. (2015). Live-cell imaging of endogenous mRNAs with a small molecule. *Angewandte Chemie International Edition* *54*, 1855-1858.
- Savić, N., and Schwank, G. (2016). Advances in therapeutic CRISPR/Cas9 genome editing. *Translational Research* *168*, 15-21.
- Sawant, A.A., Tanpure, A.A., Mukherjee, P.P., Athavale, S., Kelkar, A., Galande, S., and Srivatsan, S.G. (2016). A versatile toolbox for posttranscriptional chemical labeling and imaging of RNA. *Nucleic acids research* *44*, e16-e16.
- Schlatterer, J.C., and Jäschke, A. (2006). Universal initiator nucleotides for the enzymatic synthesis of 5'-amino- and 5'-thiol-modified RNA. *Biochemical and biophysical research communications* *344*, 887-892.
- Schlosser, K., Gu, J., Sule, L., and Li, Y. (2008). Sequence-function relationships provide new insight into the cleavage site selectivity of the 8–17 RNA-cleaving deoxyribozyme. *Nucleic acids research* *36*, 1472-1481.
- Schmidt, C., Welz, R., and Müller, S. (2000). RNA double cleavage by a hairpin-derived twin ribozyme. *Nucleic acids research* *28*, 886-894.
- Schulz, D., Holstein, J.M., and Rentmeister, A. (2013). A chemo-enzymatic approach for site-specific modification of the RNA cap. *Angewandte Chemie International Edition* *52*, 7874-7878.
- Seelig, B., and Jäschke, A. (1999). A small catalytic RNA motif with Diels-Alderase activity. *Chemistry & biology* *6*, 167-176.
- Shahn, C., Strauss, E.G., and Strauss, J.H. (1989). [11] Dideoxy sequencing of RNA using reverse transcriptase. In *Methods in enzymology* (Elsevier), pp. 121-130.
- Sharma, A.K., Plant, J.J., Rangel, A.E., Meek, K.N., Anamisis, A.J., Hollien, J., and Heemstra, J.M. (2014). Fluorescent RNA labeling using self-alkylating ribozymes. *ACS chemical biology* *9*, 1680-1684.
- Silverman, S.K. (2009). Deoxyribozymes: selection design and serendipity in the development of DNA catalysts. *Accounts of chemical research* *42*, 1521-1531.
- Smith, N., Crichton, P.B., Old, D., and Higgins, C. (1988). Ribosomal-RNA patterns of *Escherichia coli*, *Salmonella typhimurium* and related Enterobacteriaceae. *Journal of medical microbiology* *26*, 223-228.
- Song, W., Filonov, G.S., Kim, H., Hirsch, M., Li, X., Moon, J.D., and Jaffrey, S.R. (2017). Imaging RNA polymerase III transcription using a photostable RNA–fluorophore complex. *Nature chemical biology* *13*, 1187.
- Song, W., Strack, R.L., Svensen, N., and Jaffrey, S.R. (2014). Plug-and-play fluorophores extend the spectral properties of Spinach. *Journal of the American Chemical Society* *136*, 1198-1201.
- Starikov, E., and Nilsson, L. (2002). Structural basis of biotin–RNA aptamer binding: a theoretical study. *Chemical physics letters* *363*, 39-44.
- Stead, M.B., Agrawal, A., Bowden, K.E., Nasir, R., Mohanty, B.K., Meagher, R.B., and Kushner, S.R. (2012). RNA snapTM: a rapid, quantitative and inexpensive, method for isolating total RNA from bacteria. *Nucleic acids research* *40*, e156-e156.
- Steinmetzger, C., Bessi, I., Lenz, A.-K., and Höbartner, C. (2019a). Structure–fluorescence activation relationships of a large Stokes shift fluorogenic RNA aptamer. *Nucleic acids research* *47*, 11538-11550.
- Steinmetzger, C., Palanisamy, N., Gore, K.R., and Höbartner, C. (2019b). A multicolor large Stokes shift fluorogen-activating RNA aptamer with cationic chromophores. *Chemistry–A European Journal* *25*, 1931-1935.
- Strack, R.L., Disney, M.D., and Jaffrey, S.R. (2013). A superfolder Spinach2 reveals the dynamic nature of trinucleotide repeat-containing RNA. *Nature methods* *10*, 1219.
- Su, Y., and Hammond, M.C. (2020). RNA-based fluorescent biosensors for live cell imaging of small molecules and RNAs. *Current Opinion in Biotechnology* *63*, 157-166.

- Sunbul, M., and Jäschke, A. (2013). Contact-mediated quenching for RNA imaging in bacteria with a fluorophore-binding aptamer. *Angewandte Chemie International Edition* 52, 13401-13404.
- Sunbul, M., and Jäschke, A. (2018). SRB-2: a promiscuous rainbow aptamer for live-cell RNA imaging. *Nucleic acids research* 46, e110-e110.
- Tang, G. (2005). siRNA and miRNA: an insight into RISCs. *Trends in biochemical sciences* 30, 106-114.
- Tilsner, J., Linnik, O., Christensen, N.M., Bell, K., Roberts, I.M., Lacomme, C., and Oparka, K.J. (2009). Live-cell imaging of viral RNA genomes using a Pumilio-based reporter. *The Plant Journal* 57, 758-770.
- Tjaden, B., Goodwin, S.S., Opdyke, J.A., Guillier, M., Fu, D.X., Gottesman, S., and Storz, G. (2006). Target prediction for small, noncoding RNAs in bacteria. *Nucleic acids research* 34, 2791-2802.
- Tomkuvienė, M., Clouet-d'Orval, B., Černiauskas, I., Weinhold, E., and Klimašauskas, S. (2012). Programmable sequence-specific click-labeling of RNA using archaeal box C/D RNP methyltransferases. *Nucleic acids research* 40, 6765-6773.
- Tramontano, A., Janda, K.D., and Lerner, R.A. (1986). Catalytic antibodies. *Science* 234, 1566-1570.
- Tsukiji, S., Pattnaik, S.B., and Suga, H. (2003). An alcohol dehydrogenase ribozyme. *Nature Structural & Molecular Biology* 10, 713-717.
- Tsukiji, S., Pattnaik, S.B., and Suga, H. (2004). Reduction of an aldehyde by a NADH/Zn²⁺-dependent redox active ribozyme. *Journal of the American Chemical Society* 126, 5044-5045.
- Tuerk, C., and Gold, L. (1990). Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *science* 249, 505-510.
- Tutucci, E., Vera, M., Biswas, J., Garcia, J., Parker, R., and Singer, R.H. (2018a). An improved MS2 system for accurate reporting of the mRNA life cycle. *Nature methods* 15, 81.
- Tutucci, E., Vera, M., and Singer, R.H. (2018b). Single-mRNA detection in living *S. cerevisiae* using a re-engineered MS2 system. *Nature protocols* 13, 2268-2296.
- Urbanek, M.O., Galka-Marciniak, P., Olejniczak, M., and Krzyzosiak, W.J. (2014). RNA imaging in living cells—methods and applications. *RNA biology* 11, 1083-1095.
- Valegård, K., Murray, J.B., Stonehouse, N.J., van den Worm, S., Stockley, P.G., and Liljas, L. (1997). The three-dimensional structures of two complexes between recombinant MS2 capsids and RNA operator fragments reveal sequence-specific protein-RNA interactions. *Journal of molecular biology* 270, 724-738.
- van Hemert, F.J., Berkhout, B., and Zaaijer, H.L. (2014). Differential binding of tenofovir and adefovir to reverse transcriptase of hepatitis B virus. *PLoS One* 9.
- Vauléon, S., Ivanov, S.A., Gwiazda, S., and Müller, S. (2005). Site-specific fluorescent and affinity labelling of RNA by using a small engineered twin ribozyme. *ChemBioChem* 6, 2158-2162.
- Walter, N.G., and Burker, J.M. (1998). The hairpin ribozyme: structure, assembly and catalysis. *Current opinion in chemical biology* 2, 24-30.
- Wang, D., Bushnell, D.A., Westover, K.D., Kaplan, C.D., and Kornberg, R.D. (2006). Structural basis of transcription: role of the trigger loop in substrate specificity and catalysis. *Cell* 127, 941-954.
- Wang, P., and Silverman, S.K. (2016). DNA-Catalyzed Introduction of Azide at Tyrosine for Peptide Modification. *Angewandte Chemie International Edition* 55, 10052-10056.
- Warner, K.D., Chen, M.C., Song, W., Strack, R.L., Thorn, A., Jaffrey, S.R., and Ferré-D'Amaré, A.R. (2014). Structural basis for activity of highly efficient RNA mimics of green fluorescent protein. *Nature structural & molecular biology* 21, 658.
- Warner, K.D., Sjekloća, L., Song, W., Filonov, G.S., Jaffrey, S.R., and Ferré-D'Amaré, A.R. (2017). A homodimer interface without base pairs in an RNA mimic of red fluorescent protein. *Nature chemical biology* 13, 1195.

- Welz, R., Bossmann, K., Klug, C., Schmidt, C., Fritz, H.J., and Müller, S. (2003). Site-directed alteration of RNA sequence mediated by an engineered twin ribozyme. *Angewandte Chemie International Edition* 42, 2424-2427.
- Wiedenheft, B., Sternberg, S.H., and Doudna, J.A. (2012). RNA-guided genetic silencing systems in bacteria and archaea. *Nature* 482, 331-338.
- Wilson, C., and Szostak, J.W. (1995). In vitro evolution of a self-alkylating ribozyme. *Nature* 374, 777-782.
- Wirth, R., Gao, P., Nienhaus, G.U., Sunbul, M., and Jäschke, A. (2019). SiRA: a silicon rhodamine-binding aptamer for live-cell super-resolution RNA imaging. *Journal of the American Chemical Society* 141, 7562-7571.
- Woodson, S.A. (2005). Metal ions and RNA folding: a highly charged topic with a dynamic future. *Current opinion in chemical biology* 9, 104-109.
- Wu, B., Chen, J., and Singer, R.H. (2014). Background free imaging of single mRNAs in live cells using split fluorescent proteins. *Scientific reports* 4, 3615.
- Wu, H.-N., Lin, Y.-J., Lin, F.-P., Makino, S., Chang, M.-F., and Lai, M. (1989). Human hepatitis delta virus RNA subfragments contain an autocleavage activity. *Proceedings of the National Academy of Sciences* 86, 1831-1835.
- Yang, L.-Z., Wang, Y., Li, S.-Q., Yao, R.-W., Luan, P.-F., Wu, H., Carmichael, G.G., and Chen, L.-L. (2019). Dynamic Imaging of RNA in Living Cells by CRISPR-Cas13 Systems. *Molecular cell* 76, 981-997. e987.
- Zaher, H.S., and Unrau, P.J. (2007). Selection of an improved RNA polymerase ribozyme with superior extension and fidelity. *Rna* 13, 1017-1026.
- Zaher, H.S., Watkins, R.A., and Unrau, P.J. (2006). Two independently selected capping ribozymes share similar substrate requirements. *RNA* 12, 1949-1958.
- Zelin, E., Wang, Y., and Silverman, S.K. (2006). Adenosine is inherently favored as the branch-site RNA nucleotide in a structural context that resembles natural RNA splicing. *Biochemistry* 45, 2767-2771.
- Zhang, B., and Cech, T.R. (1997). Peptide bond formation by in vitro selected ribozymes. *Nature* 390, 96-100.
- Zhang, D., Zhou, C.Y., Busby, K.N., Alexander, S.C., and Devaraj, N.K. (2018). Light-activated control of translation by enzymatic covalent mRNA Labeling. *Angewandte Chemie International Edition* 57, 2822-2826.

8 Appendix

Table 8-1 List of primers used in this thesis project

Description	5'-Sequence-3'
T7 Promoter	CTGTAATACGACTCACTATA
Pool forward primer /2 nd PCR forward primer	CTGTAATACGACTCACTATAGGACATACTGAGCCTTCAACCAGCCTACCATCC
Pool reverse primer	GGTAAGGTGGACATACTG-N40-GCCTTCAAGGATGGTAGGCTGG
Selection reverse primer	GGTAAGGTGGACATACTG
1 st PCR forward primer	CTTCAACCAGCCTACCATCC
Cloning forward primer	TAAATAAAATAACTGTAATACGACTCACTATAGGACATACTGAGC
Forward primer for substrate sequence deletion	CTGTAATACGACTCACTATAGGCCAGCCTACCATCC
<i>E. coli</i> 5S rRNA forward primer	CTGTAATACGACTCACTATAGGTGCCTGGCGGCCGTAGCGCGGTGGTCCCACCTGACCCCATGCCGAAGCTCAGAAGTGAA
<i>E. coli</i> 5S rRNA reverse primer	ATGCCTGGCAGTTCCCTACTCTCGCATGGGGAGACCCACACTACCATCGGCGCTACGGCGTTTCACTTCTGAGTTCGG
<i>E. coli</i> RyhB forward primer	CTGTAATACGACTCACTATAGGCGATCAGGAAGACCCTCGCGGAGAACCTGAAAGCACGACA
<i>E. coli</i> RyhB reverse primer	AAAAAAAAAGCCAGCACCCGGCTGGCTAAGTAATACTGGAAGCAATGTGAGCAATGTCTGCTTTTCAGGTT
1st PCR NGS N8 UMI forward primer	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGNNNNNNNNNAACCA GCCTACCATCC
1st PCR NGS reverse primer	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGAAGGTCGACATACTG
Indexed forward primer I5_S502	AATGATACGGCGACCACCGAGATCTACACCTCTCTATTCTCGTCGGCAGCGTC
Indexed reverse primer I7_N703 (Round 12)	CAAGCAGAAGACGGCATAACGAGATTTCTGCCTGTCTCGTGGGCTCGG
Indexed reverse primer I7_N704 (Round 7)	CAAGCAGAAGACGGCATAACGAGATGCTCAGGAGTCTCGTGGGCTCGG
RT primer 16S 344-361 (A325)	CCCCTGCTGCCTCCCGT
RT primer 16S 413-430 (A383)	TACAACCCGAAGGCCTTC
RT primer 16S 687-704 (A649)	TCTACGCATTTACCCGCT
RT primer 23S 322-339 (A272)	ACGGGGCTGTCAACCCTGT
RT primer 23S 653-672 (A637)	GGTCTATACCCTGCAACT
RT primer 23S 1599-1618 (A1572)	TGTGTCCGTTTGGGGTACGA

Table 8-2 List of transcription templates

Description	5'-Sequence-3'
Parent	TTGAAGGCTCAGTATGTCCTATAGTGAGTCGTATTACAG
TM	CCAGGAACTCGACGCACCCTATAGTGAGTCGTATTACAG
TV1	AACTTCCCTCTCATAACACCTATAGTGAGTCGTATTACAG
TV2	GGTCCTTCTCCTGCGTGCCTATAGTGAGTCGTATTACAG
Parent-A11G	TTGAAGGCCAGTATGTCCTATAGTGAGTCGTATTACAG
TM-AAG	CCAGGAACTTGACGCACCCTATAGTGAGTCGTATTACAG
TM-UAG	CCAGGAACTAGACGCACCCTATAGTGAGTCGTATTACAG
TM-CAG	CCAGGAACTGGACGCACCCTATAGTGAGTCGTATTACAG
TM-GAA	CCAGGAATTCGACGCACCCTATAGTGAGTCGTATTACAG
TM-GAU	CCAGGAAATCGACGCACCCTATAGTGAGTCGTATTACAG
TM-GAC	CCAGGAAAGTCGACGCACCCTATAGTGAGTCGTATTACAG
Parent-AAG	TTGAAGGCTTAGTATGTCCTATAGTGAGTCGTATTACAG
Parent -UAG	TTGAAGGCTAAGTATGTCCTATAGTGAGTCGTATTACAG
Parent -CAG	TTGAAGGCTGAGTATGTCCTATAGTGAGTCGTATTACAG
Parent -GAA	TTGAAGGTTTCAGTATGTCCTATAGTGAGTCGTATTACAG
Parent -GAU	TTGAAGGATCAGTATGTCCTATAGTGAGTCGTATTACAG
Parent -GAC	TTGAAGGGTCAGTATGTCCTATAGTGAGTCGTATTACAG
TM-UGAGC	CCAGGAGCTCAACGCACCCTATAGTGAGTCGTATTACAG
5'TV1-3'parent	TTGAAGGCTCTCATAACACCTATAGTGAGTCGTATTACAG
TV1-GAGC	AACTTCGCTCTCATAACACCTATAGTGAGTCGTATTACAG
Double labeling model substrate	CATTAAGTTCGACGCATCCTTAATTGAAGGCTCAGTATGTCCTATAGTGAGTCGTATTACAG
FH14	GGACATACTGTCTATGTTTGCAGCGTTTGCAGTATGTCTACTCAGCGTGGCCTTCAACCTATAGTGAGTCGTATTACAG
FH20	GGACATACTGTCTATGGCTGGGAGGGAGCCAGGTATGTCTACGCAGCGAGGCCTTCAACCTATAGTGAGTCGTATTACAG
FH31	GGACATACTGTAACATGCCCTGGTTTTTTTCTATACTATCAGTGATGGGGTGCCTTCAACCTATAGTGAGTCGTATTACAG
FH14_TM	GGGTGCGTCGTCTATGTTTGCAGCGTTTGCAGTATGTCTACTCAGCGTGGTTCCTGGCCTATAGTGAGTCGTATTACAG
FH14_TV1	GGTGTATGAGTCTATGTTTGCAGCGTTTGCAGTATGTCTACTCAGCGTGGGGAAGTTCCTATAGTGAGTCGTATTACAG
FH14_TV2	GGCACGCAGGTCTATGTTTGCAGCGTTTGCAGTATGTCTACTCAGCGTGGGGAAGGACCCTATAGTGAGTCGTATTACAG
FH31_TM	GGGTGCGTCGTAAACATGCCCTGGTTTTTTTCTATACTATCAGTGATGGGGTGTTCCTGGCCTATAGTGAGTCGTATTACAG
FH31_TV1	GGTGTATGAGTAAACATGCCCTGGTTTTTTTCTATACTATCAGTGATGGGGTGGGGAAGTTCCTATAGTGAGTCGTATTACAG
FH31_TV2	GGCACGCAGGTAAACATGCCCTGGTTTTTTTCTATACTATCAGTGATGGGGTGAAGGACCCTATAGTGAGTCGTATTACAG

FH14_TM-AAG	GGGTGCGTCATCTATGTTTGCAGCGTTTGCAAGTATGTCTACTCAGCGTGGTTCCTGG CCTATAGTGAGTCGTATTACAG
FH14_TM-UAG	GGGTGCGTCTTCTATGTTTGCAGCGTTTGCAAGTATGTCTACTCAGCGTGGTTCCTGG CCTATAGTGAGTCGTATTACAG
FH14_TM-CAG	GGGTGCGTCCTCTATGTTTGCAGCGTTTGCAAGTATGTCTACTCAGCGTGGTTCCTGG CCTATAGTGAGTCGTATTACAG
FH14_TM-GAA	GGGTGCGTCGTCTATGTTTGCAGCGTTTGCAAGTATGTCTACTCAGCGTGATTCCTGG CCTATAGTGAGTCGTATTACAG
FH14_TM-GAU	GGGTGCGTCGTCTATGTTTGCAGCGTTTGCAAGTATGTCTACTCAGCGTGGTTCCTGG CCTATAGTGAGTCGTATTACAG
FH14_TM-GAC	GGGTGCGTCGTCTATGTTTGCAGCGTTTGCAAGTATGTCTACTCAGCGTGCTTCCTGG CCTATAGTGAGTCGTATTACAG
FH31_TM-AAG	GGGTGCGTCATAACATGCCCTGGTTTTTCTATACTATCAGTGATGGGGTGGTTCCTGG CCTATAGTGAGTCGTATTACAG
FH31_TM-UAG	GGGTGCGTCTTAACATGCCCTGGTTTTTCTATACTATCAGTGATGGGGTGGTTCCTGG CCTATAGTGAGTCGTATTACAG
FH31_TM-CAG	GGGTGCGTCCTAACATGCCCTGGTTTTTCTATACTATCAGTGATGGGGTGGTTCCTGG CCTATAGTGAGTCGTATTACAG
FH31_TM-GAA	GGGTGCGTCGTAACATGCCCTGGTTTTTCTATACTATCAGTGATGGGGTATTCCTGG CCTATAGTGAGTCGTATTACAG
FH31_TM-GAU	GGGTGCGTCGTAACATGCCCTGGTTTTTCTATACTATCAGTGATGGGGTTTTTCCTGG CCTATAGTGAGTCGTATTACAG
FH31_TM-GAC	GGGTGCGTCGTAACATGCCCTGGTTTTTCTATACTATCAGTGATGGGGTCTTCCTGG CCTATAGTGAGTCGTATTACAG
FH14_5S-A45	CCCCATGCCGTCTATGTTTGCAGCGTTTGCAAGTATGTCTACTCAGCGTGACTCAGAA CCTATAGTGAGTCGTATTACAG
FH14_5S-A65	GAAACGCCGTTCTATGTTTGCAGCGTTTGCAAGTATGTCTACTCAGCGTGGCGCCGAT CCTATAGTGAGTCGTATTACAG
FH14_5S-A99	TCCCCATGCCGTCTATGTTTGCAGCGTTTGCAAGTATGTCTACTCAGCGTGGAGTAGGG AACCTATAGTGAGTCGTATTAC
FH14_16S-A325	ACTGGAAGTGTCTATGTTTGCAGCGTTTGCAAGTATGTCTACTCAGCGTGGACACGGT CCTATAGTGAGTCGTATTACAG
FH14_16S-A412	GCGTGATGATCTATGTTTGCAGCGTTTGCAAGTATGTCTACTCAGCGTGGAAAGCCT CCTATAGTGAGTCGTATTACAG
FH14_23S-A272	AGGAGCCAGTCTATGTTTGCAGCGTTTGCAAGTATGTCTACTCAGCGTGGCCTGAAT CCTATAGTGAGTCGTATTACAG
FH14_23S-A637	AGGGAAACCGTCTATGTTTGCAGCGTTTGCAAGTATGTCTACTCAGCGTGGTCTTAAC CTATAGTGAGTCGTATTACAG

FH14_23S-A1572	TTCCAGGAAATCTATGTTTGCAGCGTTTGCAAGTATGTCTACTCAGCGTGGCCTCTAACCTATAGTGAGTCGTATTACAG
FH14_A34_D LRNA	GGATGCGTCGTCTATGTTTGCAGCGTTTGCAAGTATGTCTACTCAGCGTGGTAAATGCCTATAGTGAGTCGTATTACAG
FJ1	GGACATACTGTAGACCCTTGCCAAAGGATCTTCAGTTTTATGAGGGTGGGCCTTCAACCTATAGTGAGTCGTATTACAG
FJ8	GGACATACTGGGGGAAGGTAATGTCCTAAGGTTTCGTTTTGTAACACCGTGCCTTCAACCTATAGTGAGTCGTATTACAG
FJC1	GGACATACTGGGAACTTGCTAGTTTACCGTTTTCTTATGTGACACGTTTCGCCTTCAACCTATAGTGAGTCGTATTACAG
FJC3	GGACATACTGGAAATTTGGAGGCGGCAATTATACGGTTCGGTGGCATTCGGCCTTCAACCTATAGTGAGTCGTATTACAG
FJC9	GGACATACTGGGATGCCCAATTCGCTAAAGTGTAGTGCAACATATCTCGTGCCTTCAACCTATAGTGAGTCGTATTACAG
FJ1_TM	GGGTGCGTCGTAGACCCTTGCCAAAGGATCTTCAGTTTTATGAGGGTGGGTTCCTGGCCTATAGTGAGTCGTATTACAG
FJ1_TV1	GGTGTATGAGTAGACCCTTGCCAAAGGATCTTCAGTTTTATGAGGGTGGGGGAAGTTCCTATAGTGAGTCGTATTACAG
FJ1_TV2	GGCACGCAGGTAGACCCTTGCCAAAGGATCTTCAGTTTTATGAGGGTGGGAAGGACCCCTATAGTGAGTCGTATTACAG
FJ1_TM-UGAGC	GGGTGCGTTGTAGACCCTTGCCAAAGGATCTTCAGTTTTATGAGGGTGGGCTCCTGGCCTATAGTGAGTCGTATTACAG
FJ1_5'TV1-3'Par	GGTGTATGAGTAGACCCTTGCCAAAGGATCTTCAGTTTTATGAGGGTGGGCCTTCAACCTATAGTGAGTCGTATTACAG
FJ1_TV1-GAGC	GGTGTATGAGTAGACCCTTGCCAAAGGATCTTCAGTTTTATGAGGGTGGGCGAAGTTCCTATAGTGAGTCGTATTACAG
FJ1_AAG	GGACATACTATAGACCCTTGCCAAAGGATCTTCAGTTTTATGAGGGTGGGCCTTCAACCTATAGTGAGTCGTATTACAG
FJ1_GAA	GGACATACTGTAGACCCTTGCCAAAGGATCTTCAGTTTTATGAGGGTGGACCTTCAACCTATAGTGAGTCGTATTACAG
FJC9_TM	GGGTGCGTCGGGATGCCCAATTCGCTAAAGTGTAGTGCAACATATCTCGTGTTCCTGGCCTATAGTGAGTCGTATTACAG
FJC9_TV1	GGTGTATGAGGGATGCCCAATTCGCTAAAGTGTAGTGCAACATATCTCGTGGGAAGTTCTATAGTGAGTCGTATTACAG
FJC9_TV2	GGCACGCAGGGGATGCCCAATTCGCTAAAGTGTAGTGCAACATATCTCGTGAAGGACCCTATAGTGAGTCGTATTACAG
FJC9_AAG	GGACATACTAGGATGCCCAATTCGCTAAAGTGTAGTGCAACATATCTCGTGCCTTCAACCTATAGTGAGTCGTATTACAG
FJC9_UAG	GGACATACTTGGATGCCCAATTCGCTAAAGTGTAGTGCAACATATCTCGTGCCTTCAACCTATAGTGAGTCGTATTACAG
FJC9_CAG	GGACATACTCGGATGCCCAATTCGCTAAAGTGTAGTGCAACATATCTCGTGCCTTCAACCTATAGTGAGTCGTATTACAG
FJC9_GAA	GGACATACTGGGATGCCCAATTCGCTAAAGTGTAGTGCAACATATCTCGTACCTTCAACCTATAGTGAGTCGTATTACAG
FJC9_GAU	GGACATACTGGGATGCCCAATTCGCTAAAGTGTAGTGCAACATATCTCGTTCCTTCAACCTATAGTGAGTCGTATTACAG
FJC9_GAC	GGACATACTGGGATGCCCAATTCGCTAAAGTGTAGTGCAACATATCTCGTCCCTTCAACCTATAGTGAGTCGTATTACAG
FJC9_5S-A99	TCCCCATGCGGGATGCCCAATTCGCTAAAGTGTAGTGCAACATATCTCGTGAGTAGGGAACCTATAGTGAGTCGTATTAC
FJ1_23S-A272	AGGAGCCCAGTAGACCCTTGCCAAAGGATCTTCAGTTTTATGAGGGTGGGCCTGAATCCTATAGTGAGTCGTATTACAG
FJ1_23S-A362	ACATGCTGTGTAGACCCTTGCCAAAGGATCTTCAGTTTTATGAGGGTGGGCTCGATGCCTATAGTGAGTCGTATTACAG

FJ1_23S-A1572	TTCCAGGAAATAGACCCTTGCCAAAGGATCTTCAGTTTTATGAGGGTGGGCCTCTAAC CTATAGTGAGTCGTATTACAG
FJ1_16S-A383	AATGGGCGCATAGACCCTTGCCAAAGGATCTTCAGTTTTATGAGGGTGGGCCTGATGC AGCCTATAGTGAGTCGTATTAC
FJ1_16S-A649	GATACTGGCATAGACCCTTGCCAAAGGATCTTCAGTTTTATGAGGGTGGGCCTGAGTC CTATAGTGAGTCGTATTAC
FJ1_16S-A860	GTGGCTTCCGGTAGACCCTTGCCAAAGGATCTTCAGTTTTATGAGGGTGGGCTAACGC GTCCTATAGTGAGTCGTATTAC
FK22	GGACATACTGTCGATTTCTGCGAATTTAAGCTCAGGCCTCAACCTATAGTGAGTCGT ATTACAG
FK22_TV1	GGTGTATGAGTCGATTTCTGCGAATTTAAGCTCAGGGGAAGTTCCCTATAGTGAGTCGT ATTACAG

Table 8-3 The selection pool and the transcribed substrate sequences (modified positions shown in bold)

Description	5'-Sequence-3'
Selection pool	GGACAUACUGAGCCUUAACCAGCCUACCAUCCUUGAAGGC-N40- CAGUAUGUCCACCUUACC
Parent	GGACAUACUG A GCCUUCAA
TM	GGGUGCGUCG A GUUCCUGG
TV1	GGUGUAUGAG A GGGAAGUU
TV2	GGCACGCAGG A GAAGGACC
Parent-A11G	GGACAUACUG G GCCUUCAA
TM-AAG	GGGUGCGUCA A GUUCCUGG
TM-UAG	GGGUGCGUCU A GUUCCUGG
TM-CAG	GGGUGCGUCC A GUUCCUGG
TM-GAA	GGGUGCGUCG A AUUCCUGG
TM-GAU	GGGUGCGUCG A UUUCCUGG
TM-GAC	GGGUGCGUCG A CUUCCUGG
TM-UGAGC	GGGUGCGUUG A GCUCCUGG
5'TV1-3'parent	GGUGUAUGAG A GCCUUCAA
TV1-GAGC	GGUGUAUGAG A GCGAAGUU
Parent-AAG	GGACAUACUA A GCCUUCAA
Parent -UAG	GGACAUACUU A GCCUUCAA
Parent -CAG	GGACAUACUC A GCCUUCAA
Parent -GAA	GGACAUACUG A ACCUUCAA
Parent -GAU	GGACAUACUG A UCCUUCAA
Parent -GAC	GGACAUACUG A CCCUUCAA

Double labeling model substrate	GGACAUACUG <u>AG</u> CCUCAAUUAAGGAUGCGUCG <u>AG</u> UUAAUG
<i>E. coli</i> 5S rRNA (In vitro transcribed)	GGUGCCUGGGCGGCCGUAGCGCGGUGGUCCCACCUGACCCCA UGCCG <u>AA</u> CUCAGAAGUGAAAACGCCGU <u>AG</u> CGCCGAUGGUAGU GUGGGGUCUCCCCAUGCG <u>AG</u> AGUAGGGAACUGCCAGGCAU
<i>E. coli</i> RyhB RNA (In vitro transcribed)	GGCGAUCAGGAAGACCCUCGUGG <u>AGA</u> ACCUGAAAAGCACG ACAUUGCUCACAUUGC <u>U</u> CCAGU <u>AAU</u> ACUUAGCCAGCCG GGUGCUGGCUUUUUUUU

Table 8-4 List of the synthetic RNA substrates used for ESI-mass analysis and probing

Description	5'-Sequence-3'
R594	ACAUACUG <u>AG</u> CCUCAA-C6-NH2
R660	Hexyne-GACAUACUG <u>AG</u> CCUCAAUA

Table 8-5 List of the transcribed ribozyme variants. Binding arms are underlined.

Description	
FH14	GGUUGAAGGCC <u>CACGCUGAGUAGACAUACUUGCAAACGCUGCAAACAUAGACAG</u> <u>UAUGUCC</u>
FH20	GGUUGAAGGCC <u>CUCGCUGCGUAGACAUACCUGGCUCCUCCAGCCAUAGACAG</u> <u>UAUGUCC</u>
FH31	GGUUGAAGGCC <u>ACCCCAUCACUGAUAGUAUAGAAAAACCAGGGCAUGUUACAG</u> <u>UAUGUCC</u>
FH14_TM	GGCCAGGA <u>ACCACGCUGAGUAGACAUACUUGCAAACGCUGCAAACAUAGACGA</u> <u>CGCACCC</u>
FH14_TV1	GGAACU <u>UCCCCACGCUGAGUAGACAUACUUGCAAACGCUGCAAACAUAGACUC</u> <u>AUACACC</u>
FH14_TV2	GGUUGAAGGCC <u>CACGCUGAGUAGACAUACUUGCAAACGCUGCAAACAUAGACAG</u> <u>UAUGUCC</u>
FH31_TM	GGCCAGGA <u>ACACCCCAUCACUGAUAGUAUAGAAAAACCAGGGCAUGUUACGA</u> <u>CGCACCC</u>
FH31_TV1	GGAACU <u>UCCCCACCCCAUCACUGAUAGUAUAGAAAAACCAGGGCAUGUUACUC</u> <u>AUACACC</u>
FH31_TV2	GGGGUCCU <u>UACCCCAUCACUGAUAGUAUAGAAAAACCAGGGCAUGUUACCU</u> <u>GCGUGCC</u>
FH14_TM-AAG	GGCCAGGA <u>ACCACGCUGAGUAGACAUACUUGCAAACGCUGCAAACAUAGAUGA</u> <u>CGCACCC</u>
FH14_TM-UAG	GGCCAGGA <u>ACCACGCUGAGUAGACAUACUUGCAAACGCUGCAAACAUAGAAGA</u> <u>CGCACCC</u>
FH14_TM-CAG	GGCCAGGA <u>ACCACGCUGAGUAGACAUACUUGCAAACGCUGCAAACAUAGAGGA</u> <u>CGCACCC</u>

FH14_TM-GAA	<u>GGCCAGGAAUCACGCUGAGUAGACAUACUUGCAAACGCUGCAAACAUAGACGA</u> <u>CGCACCC</u>
FH14_TM-GAU	<u>GGCCAGGAAAACACGCUGAGUAGACAUACUUGCAAACGCUGCAAACAUAGACGA</u> <u>CGCACCC</u>
FH14_TM-GAC	<u>GGCCAGGAAGCACGCUGAGUAGACAUACUUGCAAACGCUGCAAACAUAGACGA</u> <u>CGCACCC</u>
FH31_TM-AAG	<u>GGCCAGGAACACCCCAUCACUGAUAGUAUAGAAAAAACAGGGCAUGUUAUGA</u> <u>CGCACCC</u>
FH31_TM-UAG	<u>GGCCAGGAACACCCCAUCACUGAUAGUAUAGAAAAAACAGGGCAUGUUAAGA</u> <u>CGCACCC</u>
FH31_TM-CAG	<u>GGCCAGGAACACCCCAUCACUGAUAGUAUAGAAAAAACAGGGCAUGUUAUGGA</u> <u>CGCACCC</u>
FH31_TM-GAA	<u>GGCCAGGAAUACCCCAUCACUGAUAGUAUAGAAAAAACAGGGCAUGUUAACGA</u> <u>CGCACCC</u>
FH31_TM-GAU	<u>GGCCAGGAAAACCCCAUCACUGAUAGUAUAGAAAAAACAGGGCAUGUUAACGA</u> <u>CGCACCC</u>
FH31_TM-GAC	<u>GGCCAGGAAGACCCCAUCACUGAUAGUAUAGAAAAAACAGGGCAUGUUAACGA</u> <u>CGCACCC</u>
FH14_5S-A45	<u>GGUUCUGAGUCACGCUGAGUAGACAUACUUGCAAACGCUGCAAACAUAGACGG</u> <u>CAUGGGG</u>
FH14_5S-A65	<u>GGAUCGGCGCCACGCUGAGUAGACAUACUUGCAAACGCUGCAAACAUAGAACG</u> <u>GCGUUUC</u>
FH14_5S-A99	<u>GGUUCCCUACUCCACGCUGAGUAGACAUACUUGCAAACGCUGCAAACAUAGAC</u> <u>GCAUGGGGA</u>
FH14_16S-A325	<u>GGAGGCCUUCACGCUGAGUAGACAUACUUGCAAACGCUGCAAACAUAGAUCA</u> <u>UACACGC</u>
FH14_16S-A412	<u>GGUGCGGGACCACGCUGAGUAGACAUACUUGCAAACGCUGCAAACAUAGAUAA</u> <u>CCCAACA</u>
FH14_23S-A272	<u>GGAUUCAGGCCACGCUGAGUAGACAUACUUGCAAACGCUGCAAACAUAGACUG</u> <u>GGCUCCU</u>
FH14_23S-A637	<u>GGUUAAGACCACGCUGAGUAGACAUACUUGCAAACGCUGCAAACAUAGACGGU</u> <u>UUCCCU</u>
FH14_23S-A1572	<u>GGUUAAGAGGCCACGCUGAGUAGACAUACUUGCAAACGCUGCAAACAUAGAUUU</u> <u>CCUGGAA</u>
FH14_A34-Double labeling model substrate	<u>GGCAUUAACCACGCUGAGUAGACAUACUUGCAAACGCUGCAAACAUAGACGAC</u> <u>GCAUCC</u>
FJ1	<u>GGUUGAAGGCCACCCUCAUAAAACUGAAGAUCUUGGCAAGGGUCUACAGU</u> <u>AUGUCC</u>

FJ8	GGUUGAAGGCGAAACGUGUCACAUAAGAAAACGGUAAACUAGCAAGUU CCCAGUAUGUCC
FJC1	GGUUGAAGGCCGAAUGCCACCGAACCGUAUAUUGCCGCCUCCAAUUC <u>CAGU</u> AUGUCC
FJC3	GGUUGAAGGCACGAGAUUGUUGCACUACACUUUAGCGAAUUGGGCAUCC <u>CAG</u> UAUGUCC
FJC9	GGCCAGGAACCCACCCUCAUAAAACUGAAGAUCUUGGGCAAGGGUCUAC <u>GAC</u> GCACCC
FJ1_TM	GGAA <u>CUUCCCC</u> ACCCUCAUAAAACUGAAGAUCUUGGGCAAGGGUCUAC <u>UCA</u> UACACC
FJ1_TV1	GGGGUCCU <u>UCCCC</u> ACCCUCAUAAAACUGAAGAUCUUGGGCAAGGGUCUA CCUGCGUGCC
FJ1_TV2	GGAA <u>CUUCCCC</u> ACGGUGUUACAAAACGAAACCUUAGGACAUUACCUU <u>CCCCUC</u> AUACACC
FJ1_TM-UGAGC	GGUUGAAGGCCACCCUCAUAAAACUGAAGAUCUUGGGCAAGGGUCUAC <u>UCA</u> UACACC
FJ1_5'TV1-3'par	GGAA <u>CUUCGCC</u> ACCCUCAUAAAACUGAAGAUCUUGGGCAAGGGUCUAC <u>UCA</u> UACACC
FJ1_TV1-GAGC	GGUUGAAGGCCACCCUCAUAAAACUGAAGAUCUUGGGCAAGGGUCUAU <u>AGU</u> AUGUCC
FJ1_AAG	GGUUGAAGGUCCACCCUCAUAAAACUGAAGAUCUUGGGCAAGGGUCUAC <u>AGU</u> AUGUCC
FJ1_GAA	GGUUGAAGGCACGGUGUUACAAAACGAAACCUUAGGACAUUACCUU <u>CCCCUC</u> AUACACC
FJC9_TM	GGAA <u>CUUCCCC</u> ACGAGAUUGUUGCACUACACUUUAGCGAAUUGGGCAUCC <u>CUC</u> AUACACC
FJC9_TV1	GGGGUCCU <u>UAC</u> CGAGAUUGUUGCACUACACUUUAGCGAAUUGGGCAUCC <u>CU</u> GCGUGCC
FJC9_TV2	GGUUGAAGGCACGAGAUUGUUGCACUACACUUUAGCGAAUUGGGCAUCC <u>UAG</u> UAUGUCC
FJC9_AAG	GGUUGAAGGCACGAGAUUGUUGCACUACACUUUAGCGAAUUGGGCAUCC <u>AAG</u> UAUGUCC
FJC9_UAG	GGUUGAAGGCACGAGAUUGUUGCACUACACUUUAGCGAAUUGGGCAUCC <u>GAG</u> UAUGUCC
FJC9_CAG	GGUUGAAGGUACGAGAUUGUUGCACUACACUUUAGCGAAUUGGGCAUCC <u>CAG</u> UAUGUCC
FJC9_GAA	GGUUGAAGGAACGAGAUUGUUGCACUACACUUUAGCGAAUUGGGCAUCC <u>CAG</u> UAUGUCC
FJC9_GAU	GGUUGAAGGGACGAGAUUGUUGCACUACACUUUAGCGAAUUGGGCAUCC <u>CAG</u> UAUGUCC

FJC9_ GAC	GG <u>AUUCAGGCC</u> CCACCCUCAUAAAACUGAAGA <u>UCCUUUGGCAAGGGUCUACUGG</u> GCUCCU
FJC9_5S-A99	GGU <u>UCCUACUC</u> CACGAGAU <u>AUGUUGCACUACACUUUAGCGAAUUGGGCAUCC</u> GCAUGGGGA
FJ1_23S-A272	GG <u>CAUCGAGCC</u> CCACCCUCAUAAAACUGAAGA <u>UCCUUUGGCAAGGGUCUACACA</u> GCAUGU
FJ1_23S-A362	GGU <u>UAGAGGCC</u> CCACCCUCAUAAAACUGAAGA <u>UCCUUUGGCAAGGGUCUAUUUC</u> CUGGAA
FJ1_23S-A1572	GG <u>AUCAAGGCC</u> CCACCCUCAUAAAACUGAAGA <u>UCCUUUGGCAAGGGUCUACAAU</u> GUUCAG
FJ1_16S-A383	GG <u>ACUCAAGCC</u> CCACCCUCAUAAAACUGAAGA <u>UCCUUUGGCAAGGGUCUAUGCC</u> AGUAUC
FJ1_16S-A649	GG <u>ACGCGUUAGCC</u> CCACCCUCAUAAAACUGAAGA <u>UCCUUUGGCAAGGGUCUACC</u> GGAAGCCAC
FJ1_16S-A860	GG <u>ACCGUGUCC</u> CACGCUGAGUAGACAUACUUGCAAACGCUGCAAACAUAGAC <u>CAG</u> UUCAGU
FK22	GGU <u>UGAAGGCC</u> CUGAGCUUAAA <u>UUCGCAGAAAUCGACAGUAUGUCC</u>
FK22_TV1	GGA <u>ACUUC</u> CC <u>CCUGAGCUUAAA</u> UUCGCAGAAA <u>UCGACUCAUACACC</u>

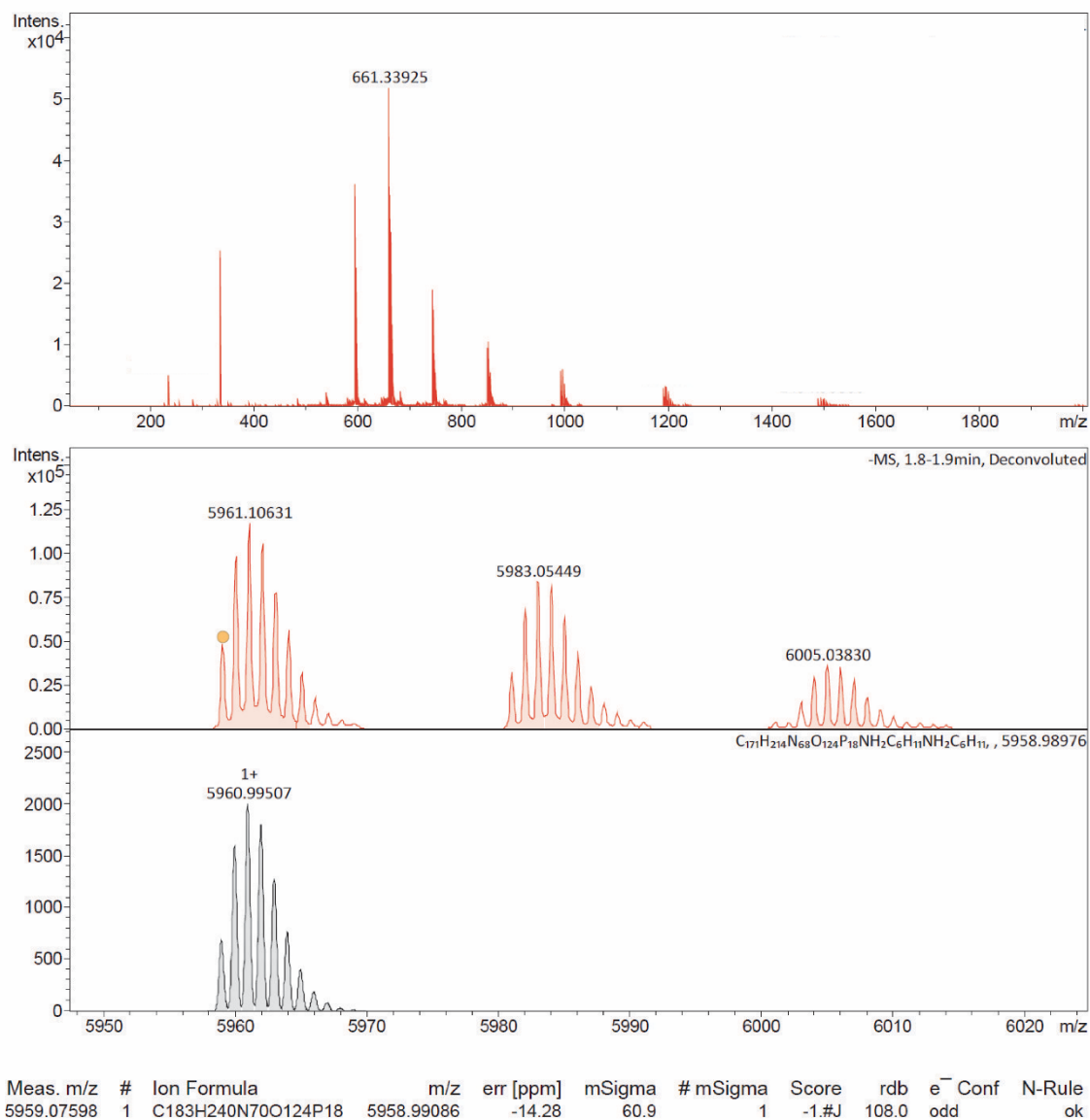


Figure 8-1 ESI-mass spectrum and the deconvolution of the R549 modified using FH14 and N⁶-aminohexyl-ATP. (Grey: Simulated mass spectrum)

Table 8-6 Details of the ten main clusters identified from the NGS analysis of the round 12 pool

Cluster	Unique	Reads	RPM
1	1407	86635	761012.5
2	407	84781	74472.2
3	259	55156	48449.4
4	158	21523	18905.9
5	28	3697	3247
6	38	3943	3463
7	20	2733	2400
8	26	2401	2109
9	10	1268	1113.8
10	6	720	632.4

Table 8-7 First sequence of fist ten clusters ordered by log₂ (enrichment).

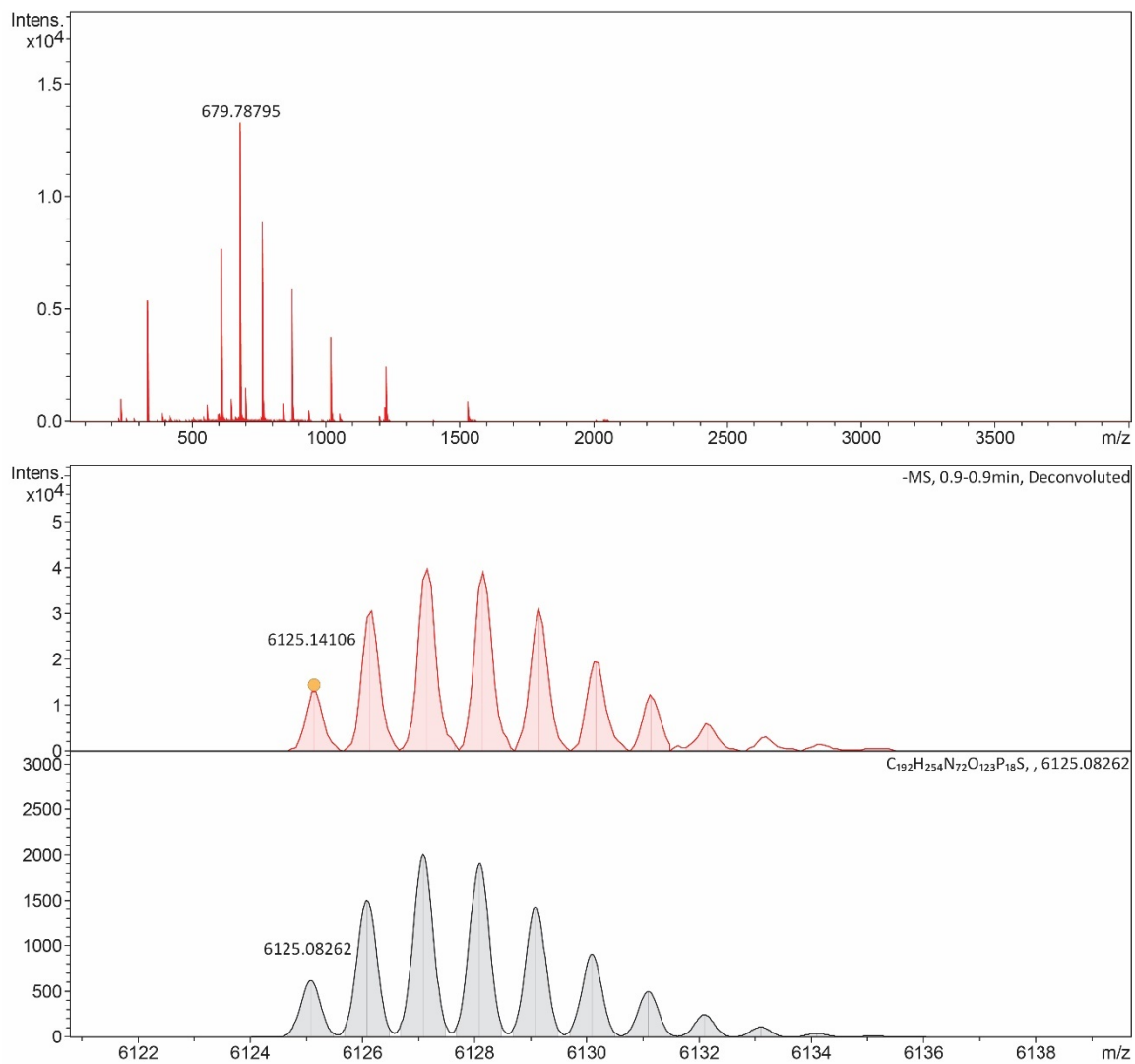
Cluster	sequence	RPM(x)*	RPM(Y)*	log ₂ x/y
>2-2-1-0=FJ8	AACCAGCCTACCATCCTTGAAGGC ACGGTGTACAAAACGAAACCTTAGGACATTACCTTCCCC CAGT	38.8	33325.9	9.74
>8-4-1-0=FJ1	AACCAGCCTACCATCCTTGAAGGC CCACCCTCATAAACTGAAGATCCTTTGGCAAGGGTCTA CAGTA	25.7	10479.4	8.67
>82-9-1-0=FJC9	AACCAGCCTACCATCCTTGAAGGC ACGAGATATGTTGCACTACACTTTAGCGAATTGGGCATCC CAGT	46.2	924.1	4.32
>1-1-1-0=FJC1	AACCAGCCTACCATCCTTGAAGGC GAAACGTGTCACATAAGAAAACGGTAAACTAGCAAGTTCC CAGT	130185	502593	1.95
>48-7-1-0 (a)	AACCAGCCTACCATCCTTGAAGGC T CAGTATGTC GACCTTCTGTCTCTTATACACATCTCCGAGCCCA	474.2	1586.4	1.74
>26-5-1-0 (b)	AACCAGCCTACCATCCTTGAAGGC CGAATGCCACCGAA CAGTATGTC GACCTTCTGTCTCTTATACAC	3096.4	2437.6	-0.35
>3-3-1-0=FJC3	AACCAGCCTACCATCCTTGAAGGC CGAATGCCACCGAACCCTATAATTGCCGCTCCAATTTTC CAGTA	39964.6	30094.3	-0.41
>56-8-1-0	AACCAGCCTACCATCCTTGAAGGC ACTGGTGTACAAAGTAAACGCAACCTTAATTTTACGATCC CAGT	4264.5	1258.8	-1.76
>32-6-1-0 (c)	AACCAGCCTACCATCCTTGAAGGC ATCGGTGTAACATAAACGAAACCTTAGGTCCTTTGGCCCC CAGT	34139.0	2206.6	-3.95
>126-10-1-0	AACCAGCCTACCATCCTTGAAGGC ATAGTTTCACAATAGCCACTTTAAGCTTATCCAAAGCTCA CAGT	11488.7	577.1	-4.32

* RPM(X)= Round 7 and RPM (Y)= Round2

(a) extended complement to the substrate sequence

(b) reveals a fully complementary binding arm after 14 nt in the center of the N40 core

(c) related to FJ8



Meas. m/z	#	Ion Formula	m/z	err [ppm]	mSigma	# mSigma	Score	rdb	e ⁻ Conf	N-Rule
6125.14106	1	C ₁₉₂ H ₂₅₄ N ₇₂ O ₁₂₃ P ₁₈ S	6125.08371	-9.36	36.2	1	100.00	111.0	odd	ok

Figure 8-2 ESI-mass spectrum and the deconvolution of the R549 modified using FJ1 and N⁶-biotin-TenDP. (Grey: Simulated mass spectrum)

Abbreviations

μ IVC	microfluidic in vitro compartmentalization
μ L	microliter
μ M	micromolar
A	adenine/adenosine
Alk	alkaline
AMP	adenosine monophosphate
APM	acrylaminophenyl mercuric acid
ATP	adenosine triphosphate
BHQ	black hole quencher
BiP	2-2'- bipyridine
bp	base pair
BW	binding-washing
C	cytosine/cytidine
Cas	CRISPR-associated complex
Cbl	cobalamin
cDNA	complementary DNA
CoA	coenzyme A
CRISPR	clustered regularly interspaced short palindromic repeats
CTP	cytidine triphosphate
CuAAC	Copper(I)- catalyzed alkyne-azide cycloaddition
Cy	cyanine
dATP	deoxyadenosine triphosphate
dCTP	deoxycytidine triphosphate
ddCTP	dideoxycytidine triphosphate
ddNTP	dideoxynucleoside triphosphate
ddTTP	dideoxythymidine triphosphate
dGTP	deoxyguanosine triphosphate
DHFR	dihydrofolate reductase
DMF	dimethylformamide
DMSO	dimethylsulfoxide
DN	dinitroaniline
DNA	deoxyribonucleic acid
dNTP	deoxynucleotide triphosphate
dsDNA	double-stranded DNA
DTT	dithiothreitol
dUTP	deoxyuridine triphosphate
EDTA	ethylenediaminetetraacetic acid
ESI-mass	electrospray ionization mass spectrometry
et al	et alii (english: and others)
FAM	carboxy fluorescein
FISH	fluorescent in situ hybridization

FLAP	fluorogen activating aptamer
Flu	fluorescein
FP	fluorescent protein
G	guanine/guanosine
GFP	green fluorescent protein
GMP	guanosine monophosphate
gRNA	guide RNA
GTP	guanosine triphosphate
h	hour
HBI	hydroxybenzilydene imidazole
HEPES	4-(2-Hydroxyethyl)-1-piperazineethanesulfonic acid
HNE	human neutrophil elastase
HPLC	high-performance liquid chromatography
iEDDA	inverse electron demand Diels-Alder reaction
IVT	in vitro transcription
L	liter
LG	leaving group
lncRNA	long non-coding RNA
LY-CH	lucifer-yellow carbonylhydrazide
M	molar
MBS	MS2 coat-protein binding-site
MCP	MS2 coat-protein
mg	milligram
min	minute
miRNA	microRNA
mL	milliliter
mM	millimolar
mRNA	messenger RNA
NGS	next-generation sequencing
NHS	N-hydroxysuccinimide
nM	nanomolar
nt	nucleotide
NTP	nucleoside triphosphate
O/N	overnight
PAGE	polyacrylamide gel electrophoresis
PAM	protospacer adjacent motif
PBS	PP7 coat-protein binding-site
PCP	PP7 coat-protein
PCR	polymerase chain-reaction
pH	potential of hydrogen
R	purine nucleotide
RISC	RNA-induced silencing complex
RNA	ribonucleic acid
ROI	RNA of interest
RPM	read per million

rRNA	ribosomal RNA
RT	reverse transcriptase/reverse transcription
SAM	S-adenosyl-methionine
sec	second
SELEX	systematic evolution of ligands by exponential enrichment
SiR	silicon rhodamine
siRNA	small-interfering RNA
SPAAC	strain promoted alkyne-azide cycloaddition
SRB	sulforhodamine B
ssDNA	single-stranded RNA
TAMRA	tetramethyl rhodamine
Taq	<i>Thermus aquaticus</i>
tBuOH	tert-butanol
TEN	Tris/EDTA/NaCl
TenDP	tenofovir diphosphate
TetFC	tetramolecular fluorescence complementation
TGT	tRNA guanine transglucosylase
Tias	tRNA(ile2)-agmatidine synthetase
TM	transition mutation
TO1	thiazole orange 1
TriFC	trimolecular fluorescence complementation
Trm	tRNA methyltransferase
tRNA	transfer RNA
TV	transversion mutation
U	uracil/uridine
UTP	uridine triphosphate
UTR	untranslated region
UV	ultraviolet
W: C	Watson-Crick
X-gal	5-bromo-4-chloro-3-indolyl- <i>b</i> -D-galactopyranoside
Y	pyrimidine nucleotide
β-ME	β-mercaptoethanol

Acknowledgements

First and foremost, I would like to thank Prof. Claudia Höbratner for giving me the opportunity to work at her interdisciplinary research group. It was an utterly fulfilling experience to work at the interphase of molecular biology and organic chemistry. I learned countless new things working under her supervision. I enjoyed every scientific discussion and idea exchange I have had with her, during this time. Never once did I feel that my potentials were left unused.

I would like to extend my gratitude to the coordinators of the Molecular biology program, especially, Dr. Steffen Burkhardt and Kerstin Grüniger for this amazing international experience. I am thankful for their unwavering support from my very first day in Germany.

I thank my lab mates, particularly Ann-Kathrin, Caro and Anam for providing such a nice and friendly environment.

I am deeply thankful to Dr. Surjendu Dey for the organic synthesis of tenofovir analogs used in this study.

I am also grateful to my friends from back home, Peyman and Sajad for making my vacations fantastic. Special thanks go to my good friend Mehrnoush for the moral support and help with proofreading this thesis.

Last but not least, I offer my deepest gratitude to my family, especially my parents, without whose efforts I would not have made it this far. It is to them that I dedicate this thesis.

Curriculum Vitae

Name: Mohammad Ghaem Maghami

Date of birth: 22.09.1987

Place of birth: Tehran, Iran

Contact: m.ghaemmaghams@stud.uni-goettingen.de

mohammad.ghaem-maghams@uni-wuerzburg.de

Education

June.2016-May.2019. PhD in Molecular biology. Georg-August University Göttingen. Göttingen, Germany.

October.2014-March.2016. MSc in Molecular biology. Georg-August University Göttingen. Göttingen, Germany.

September.2009-Feb.2012. MSc in Medical biotechnology. Tarbiat Modares University. Tehran, Iran.

September.2005-August 2009. BSc in Biotechnology. Shahed University, Tehran, Iran.

List of publications

Ghaem Maghami M., Lenz A.K., Dey S., Höbartner C. Repurposing Antiviral drugs for orthogonal RNA-catalyzed labeling of RNA. *Angew. Chem. Int. Ed.* 2020, DOI: 10.1002/anie.202001300.

Ghaem Maghami M., Scheitl C.P.M., Höbartner Claudia. Direct in vitro selection of Trans-acting ribozymes for posttranscriptional, site-specific, and covalent fluorescent labeling of RNA. *J. Am. Chem. Soc.* 2019, 141, 50, 19546-19549.