UNIVERSITY OF CALIFORNIA, SAN DIEGO

Enzyme-Mediated Site-Specific Incorporation of a Fluorescent Nucleoside into RNA: Method and Applications

A Thesis submitted in partial satisfaction of the requirements for the degree Master of Science

in

Chemistry

by

Yao Li

Committee in charge:

Professor Yitzhak Tor, Chair
Professor Thomas Hermann
Professor Ulrich Muller

2016

The Thesis of Yao Li is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

_____

_____

_____

_____

Chair

University of California, San Diego

2016

EPIGRAPH

In my life, I had come to realize that, when things were going very well, indeed, it was just the time to anticipate trouble. And, conversely, I learned from pleasant experience that at the most despairing crisis, when all looked sour beyond words, some delightful "break" was apt to lurk just around the corner.

Amelia Earhart

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I acknowledge, with gratitude, my debt of thanks to Dr. Yitzhak Tor for giving me this opportunity to work in his lab, and for his advice and guidance that supported me throughout my graduate studies.

I would also like to thank Dr. Lisa McCoy and Dr. Andrea Fin, for not only being outstanding mentors, but also great friends and fun people to work with.

To the Tor lab members: Patrycja, Alex, Kaivin, Christina and Ryan, I would like to show my appreciation as every single one of them has always been extremely friendly and supportive.

Lastly, I would like to thank my parents for their moral support, and my friends for sharing and cheering for life with me.

Chapter one, as well as the abstract of thesis, in full, is currently being prepared for submission for publication for the material. Li, Y.; Fin, A.; McCoy, L.; Tor, Y. Polymerase-mediated site-specific incorporation of a synthetic fluorescent isomorphic G surrogate into RNA. The thesis author was the primary investigator and author of this material.

ABSTRACT OF THE THESIS

Enzyme-Mediated Site-Specific Incorporation of a Fluorescent Nucleoside into RNA: Method and Applications

by

Yao Li

Master of Science in Chemistry

University of California, San Diego, 2016

Professor Yitzhak Tor, Chair

An enzyme-mediated approach for the assembly of singly modified RNA constructs in which specific G residues are replaced with thG, an emissive isomorphic G surrogate is reported. Transcription initiation in the presence of thG and native nucleoside triphosphates enforces initiation with the unnatural analog, yielding 5'-end modified transcripts that can be mono-phosphorylated and ligated to provide longer

site-specifically modified RNA constructs. To illustrate the utility of this effective approach, we explore its scope and limitations via the assembly of several altered hammerhead (HH) ribozymes and a singly modified HH substrate. By strategically modifying key positions, a mechanistic insight into the ribozyme-mediated cleavage is gained. Additionally, the emissive features of the modified nucleoside and its responsiveness to environmental changes can be used to monitor cleavage in real time by steady state fluorescence spectroscopy.

Chapter One: Polymerase-Mediated Site-Specific Incorporation of a Synthetic Fluorescent Isomorphic G Surrogate into RNA

1. Introduction

Site-specific modification of RNA has served as a powerful tool for probing structure, function and mechanisms of biologically-relevant constructs.[1] The most commonly used approach involves the traditional solid phase synthesis, which, in principle, facilitates the incorporation of any modification at any position, assuming the necessary building blocks can be prepared and the modified nucleotide survives the obligatory cleavage and deprotection protocols.[2] Enzymatic methods, which are frequently more amenable for the synthesis of longer biologically-relevant sequences and for the incorporation of sensitive nucleoside replacements, tend to be more limiting but with different liabilities.[3] Synthetically, a nucleoside triphosphate (NTP) needs to be made and recognized as a substrate by the polymerase. Favorable recognition of a nucleotide surrogate by the polymerase leads, in turn, to multiple incorporations, thus limiting this approach to short non-statistical sequences.[4] To overcome such obstacles, orthogonal base-pairing schemes directing the site-specific incorporation of novel nucleosides have been advanced, which require, however, the synthesis of both modified templates and NTPs.[3c,5] Alternatively, short synthetic fragments, typically obtained by solid-phase synthesis, have been subjected to ligation reactions to afford the full-length, site-specifically modified constructs.[6] These methods have found unique applications, but have not yet been universally adopted.

We have previously introduced highly isomorphic nucleoside analogs, with favorable biophysical, biochemical and photophysical features, built around a

thieno[3,4-d]pyrimidine core, as a purine replacement (see Figure 1.1.1).[7] We demonstrated that thGTP, a GTP surrogate, successfully facilitated initiation of in vitro transcription reactions and elongation of the growing transcripts catalyzed by T7 RNA polymerase.[8] In the resulting transcripts all guanosine residues were replaced with thG. When applied to the minimal hammerhead (HH) ribozyme HH16, a small functional RNA, the impact of individual residues was obscured. To facilitate refined mechanistic studies of functional RNA molecules and shed light on the role of individual G residue, a method for site-specific enzymatic modifications was sought.

An enzyme-mediated approach for the assembly of singly modified RNA constructs in which specific G residues are replaced with their thG surrogate (Figure 1.1.) was disclosed. It relies on transcription initiation in the presence of thG and native nucleoside triphosphates, which enforces initiation with the unnatural analog. The resulting 5'-end modified transcripts are then mono-phosphorylated and ligated to provide longer site-specifically modified RNA constructs (Figure 1.1.). To critically assess the utility of this protocol, we explore its scope and limitations via the assembly of several altered HH enzymes, as well as a singly modified HH substrate. By strategically modifying key positions, a mechanistic insight into the ribozyme-mediated cleavage is gained. Additionally, we exploit the emissive features of the modified nucleoside and its responsiveness to environmental changes and demonstrate that cleavage can be monitored in real time by steady state fluorescence spectroscopy. The general method reported here can be exploited to address questions in RNA biochemistry and to facilitate the assembly of fluorescence-based assays for RNA structure/function.

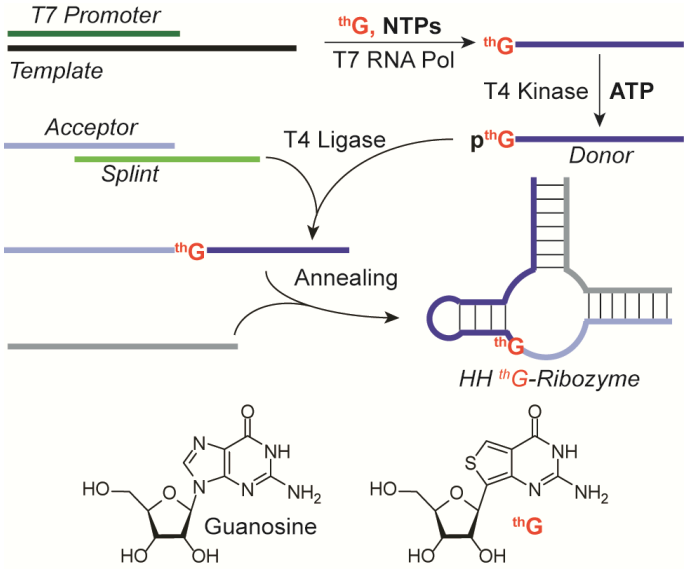

Figure 1.1. An enzyme-mediated approach for the assembly of singly modified RNA constructs, replacing a G residue with thG.

2. Result

2.1. Transcription Initiation with thG

Transcription reactions of template 1 with the analytically pure thG, all native NTPs and T7 RNA polymerase were performed under previously reported conditions (Figure 2.1a).[8,9] thG-initiated RNA constructs and the unmodified native products were successfully separated using gel electrophoresis (Figure 2.1.b, lane 2). Importantly, UV illumination (302 nm) shows the product and truncated transcripts, which are all highly fluorescent (figure 2.1.b). Following extraction and desalting, the full-length thG initiated transcript (transcript 1b) and native transcript (transcript 1a) were quantified and analyzed by mass spectrometry (Figure 2.1.c). The relative transcript yield (1b/1a) was 0.55 ± 0.02.

To identify the optimal thG/GTP ratio for the production of thG-initiated full-length transcripts, a series of reactions was carried out, where the concentration of thG was varied (1–13 mM), while keeping the concentration of all native NTPs,

including GTP, constant (1 mM). The relative yield of transcript 1b increased from 0.29 to 1.09 as thG concentrations were elevated, while the total yield of RNA remained approximately constant (Figure A1.). Scaled-up transcription reactions of template 1 were therefore carried out with 5 mM of thG. We note that the optimal conditions for thG-initiated transcription reactions were, however, template-dependent.

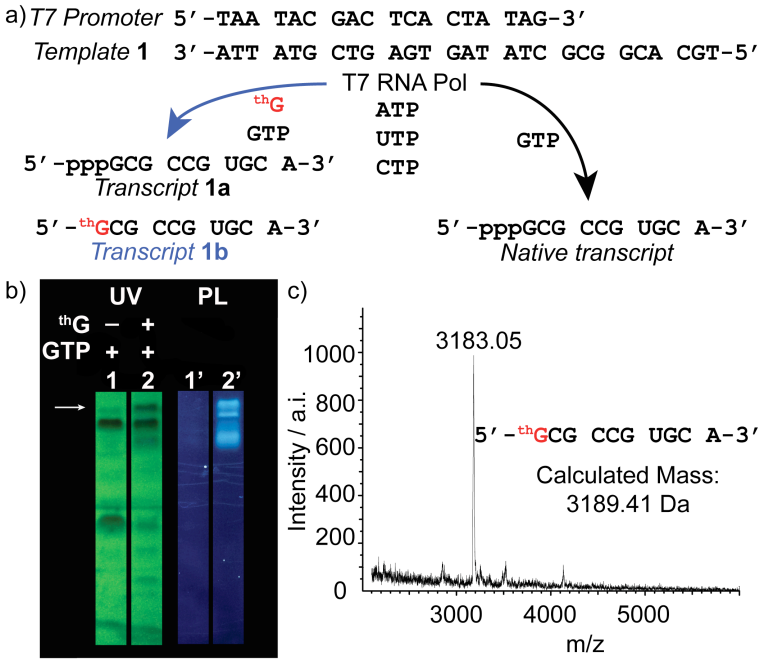

Figure 2.1. Transcription reactions with template **1**. a) T7 promoter and template **1** depicting the enzymatic incorporation reaction using natural NTPs with or without the presence of thG resulting in different transcripts. b) Transcription reaction using template **1** with 1 mM of all natural NTPs (lane 1 and 1'), 1 mM of all natural NTPs and 5 mM of thG (lanes 2 and 2'). c) MALDI analysis of transcript **1b**.

2.2. Preparation of thG-Modified Oligonucleotides via Transcription and Ligation

To facilitate the preparation of internally, singly-modified HH enzymes (schematically depicted in Figure 1.1.), thG-terminated oligonucleotides have been synthesizes as described above and then phosphorylated and ligated to the corresponding unmodified oligonucleotides. DNA templates bearing the sequence terminated at the 5' with the modification site, designated as donor strands

(corresponding to templates **2–5**), were designed and transcribed as described above. A thG/GTP ratio between 5 and 8 was used for the different templates (see Section 6.2.). The crude transcription mixtures were gel purified and the bands corresponding to the desired thG-terminated construct (Figure 2.2., transcripts **2b**, **3b**, **4b** and **5b**) and the undesired native transcripts (Figure 2.2., transcripts **2a**, **3a**, **4a**, **5a**) were resolved and isolated (Figure A2.). The relative yield of each thG-initiated full-length transcript was determined (0.88 ± 0.01 for transcript **2b**/**2a**, 1.16 ± 0.12 for transcript **3b**/**3a**, 0.82 ± 0.06 for transcript **4b**/**4a**, and 0.36 ± 0.06 for transcript **5b**/**5a**).

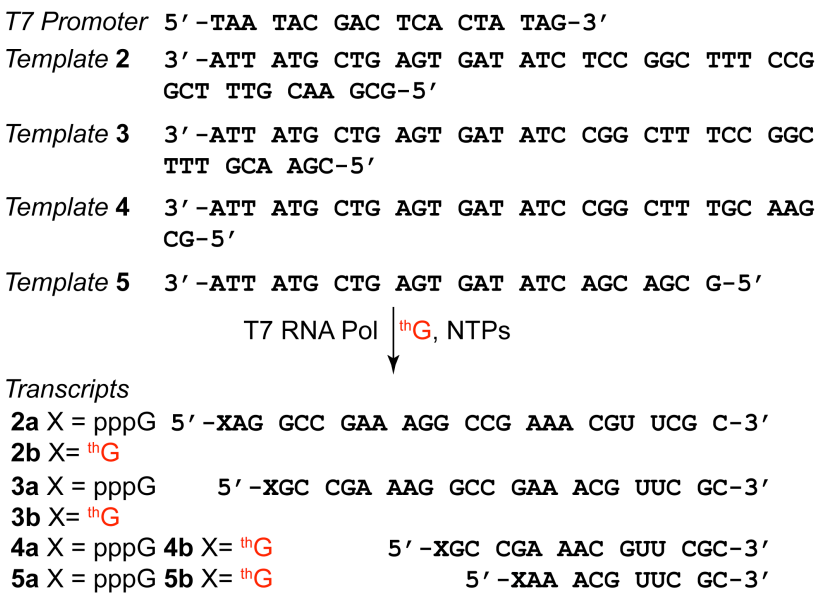

Figure 2.2. Transcription reaction of different 5'-thG terminated oligonucleotides (**2b**, **3b**, **4b** and **5b**) starting from template **2–5**.

T4 polynucleotide kinase (PNK) effectively phosphorylated the 5'-end of all thG-terminated RNA oligonucleotides (Figure A3.). The purified constructs were then ligated to oligonucleotides representing the HH ribozyme's enzyme strand sequence from its 5'-terminus to the modification site (designated acceptor strands), using T4 DNA ligase (Figure 2.3.). DNA strands, complementing 10 to 15 nucleotides in the donor/acceptor were used as splint to create a short nicked duplex (Figure A4.).

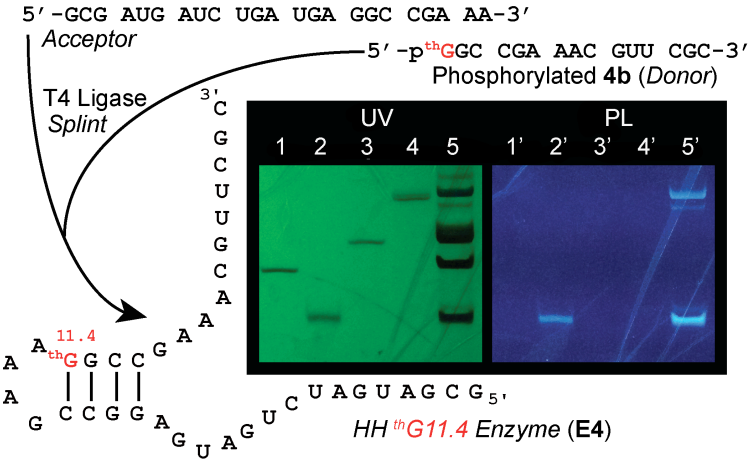

Figure 2.3. Splinted ligation of HH16 thG11.4 E donor (**4b**) and acceptor performed with T4 DNA ligase. HH16 thG11.4 E acceptor (lane 1 and 1'), phosphorylated **4b** (lane 2 and 2'), splint (lane 3 and 3'), native HH16 E (lane 4 and 4') and ligation reaction (lane 5 and 5').

Gel purifications provided an effective means to resolve the ligated product from the un-ligated starting material (Figure 2.3.). As before, the ligated product was fluorescent under long UV illumination, indicating the presence of thG in the oligonucleotides (Figure 2.3.). Following the same method, thG8 enzyme (E2), thG10.1 enzyme (E3), thG11.4 enzyme (E4) as well as thG12 enzyme (E5) were prepared (Figure 2.4.). The ligation reactions were found to be effective, providing the desired HH enzymes in 20–40% yield (Figure A4.). The native enzyme (E1) was prepared by standard in vitro transcription.

	X_8	$X_{10.1}$	$X_{11.4}$	X_{12}
E1	G	G	G	G
E2	thG	G	G	G
E3	G	thG	G	G
E4	G	G	thG	G
E5	G	G	G	thG

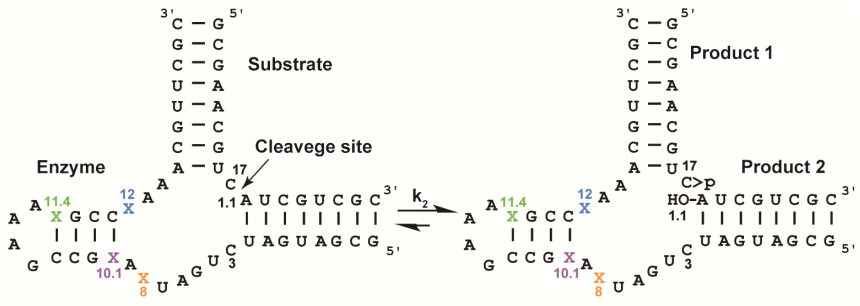

Figure 2.4. Hammerhead ribozymes and schematic cleavage reactions representation of natural substrate (**S1**) with native (**E1**), thG8 (**E2**), thG10.1 (**E3**), thG11.4 (**E4**) and thG12 (**E5**) enzymes.

All the thG-terminated donor strands and ligated oligonucleotides were characterized by mass spectroscopy (Figures A5.–A14.), and E5 was also digested using S1 nuclease and dephosphorylated before being subjected to HPLC analysis (Figure 2.5.a), confirming the presence and stoichiometry of the modified intact nucleoside. To validate the position of modifications, enzymatic digestion with T1 nuclease, which is an N-7 dependent RNases that cleaves single-stranded RNA 3' to G residues,[10] was also applied to all the modified HH16 enzyme strands (Figure A15.). As seen in Figure 2.5.b, a comparison of the T1 RNase cleavage pattern obtained for the native RNA E1 and the singly modified E5 shows a "footprint" at position 12, where thG replaces G (compare lanes 3 and 6), thus further substantiating the presence of thG at this position.

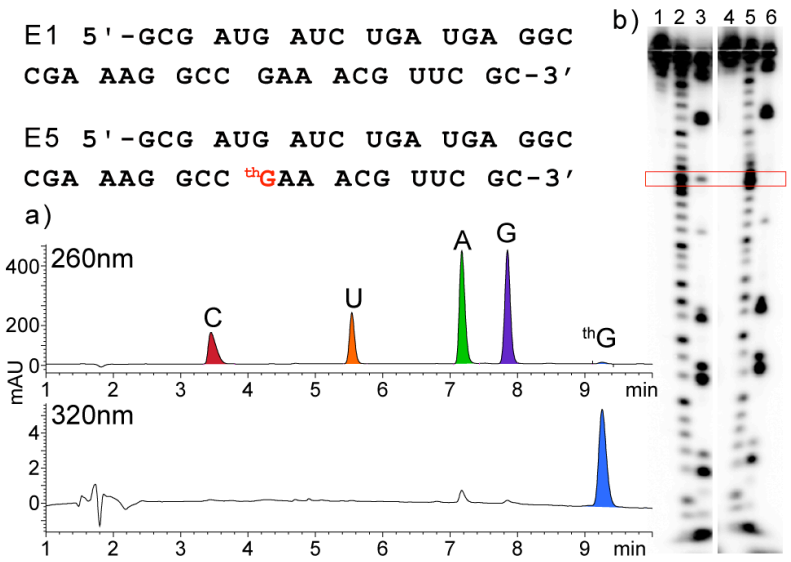

Figure 2.5. Characterization of **E5**. a) HPLC traces of S1 nuclease digestion of **E5**. b) T1 digestion of **E1** and **E5**. Lane 1, non-treated **E1**. Lane 2, alkaline hydrolysis of **E1**. Lane 3, T1 nuclease digestion of **E1**. Lane 4, non-treated **E5**. Lane 5, alkaline hydrolysis of **E5**. Lane 6, T1 nuclease digestion of **E1**. Position 12 is red boxed.

2.3 Cleavage Activity of thG-Modified Hammerhead Ribozymes

thG-modified enzymes (**E2–E5**) were assembled into the corresponding

hammerhead ribozymes (Figure 2.4.) with a native ^{32}P-labeled substrate (**S1**) and tested for strand cleavage (Figure 2.6.a), using conditions similar to those previously published.[8,11]

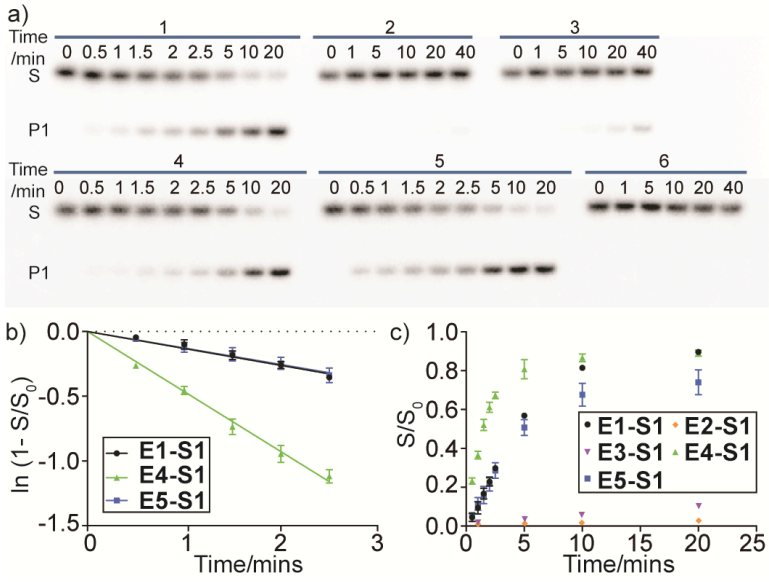

Figure 2.6. Cleavage of ^{32}P-labeled **S1** by HH enzyme strands with replacement of G for thG at different positions. a) **E5** (1), **E2** (2), **E3** (3), **E1** (4), **E4** (5) and fully thG-modified enzyme, **thG-E** (6) with native substrate (**S1**). **S1** and **P1** indicate the substrate and the product strands respectively. b) Initial kinetics of **S1-E1** (black), **S1-E5** (blue) and **S1-E4** (green). The pseudo-first-order rate constants (k_2) of the cleavage reactions are determined as the slope of semi-logarithmic plot of the fraction cleaved as function of time. c) Ribozyme-mediated cleavage curves as determined by ^{32}P data for **S1-E1** (black), **S1-E2** (orange), **S1-E3** (purple) **S1-E4** (green) and **S1-E5** (blue). Fraction cleaved (S/S_0) was determined by dividing the amount of cleaved substrate by the sum of the full length and cleaved substrate.

The rate constant obtained for **E5-S1** was 0.13 ± 0.02 min^{-1}, while that of the native **E1-S1** was 0.13 ± 0.01 min^{-1}, and 0.47 ± 0.02 min^{-1} for **E4-S1** (Figure 2.6.b, c). However, the cleavage of **S1** was prohibited to a large extent for **E2** and **E3** (Figure 2.6.a, c), with 4.9% ± 0.6% and 18.4% ± 0.8% in 40 minutes, respectively (Table 2.1.).

Table 2.1. Cleavage data for HH ribozymes

	E1-S1	E2-S1	E3-S1	E4-S1	E5-S1
$\mathbf{k_2}$[a]	0.13 ± 0.01	n.d.	n.d.	0.47 ± 0.02	0.13± 0.02
S/S$_0$ (20 min)[b]	0.90 ± 0.01	0.03 ± 0.01	0.10 ± 0.01	0.89 ± 0.02	0.74 ± 0.06
S/S$_0$ (40 min)[b]	-	0.05 ± 0.01	0.18 ± 0.01	-	-

[a] $\mathbf{k_2}$ is the pseudo-first-order rate constant reported in min^{-1} and it is equal to the slope of the semi-logarithmic plot in Figure 2.6b.
[b] Fraction cleaved (**S/S$_0$**) was determined after 20 and/or 40 minutes by dividing the amount of cleaved substrate by the sum of the full length and cleaved substrate.

To explore the impact of incorporating thG at the cleavage site (X1.1), modified substrates with either thG [thG1.1-S (**S3**)] or guanosine [G1.1-S (**S2**)] residues replacing A1.1 in **S1**, were prepared and assembled into Hammerhead ribozymes by hybridization to **E6** (Figure 2.7.). The cleavage kinetics of **E6-S2** as well as that of **E6-S3** was revealed by ^{32}P labeling of the substrate following the same conditions as previous experiments (Figure 2.7.). No major differences between the cleavage rates of **E6-S2** and **E6-S3** were seen (0.11 ± 0.01 and 0.08 ± 0.01 min^{-1}, respectively; Figure 2.8.a, b), indicating that the cleavage reaction was not significantly impacted by replacing G1.1 with thG at the ribozyme's cleavage site.

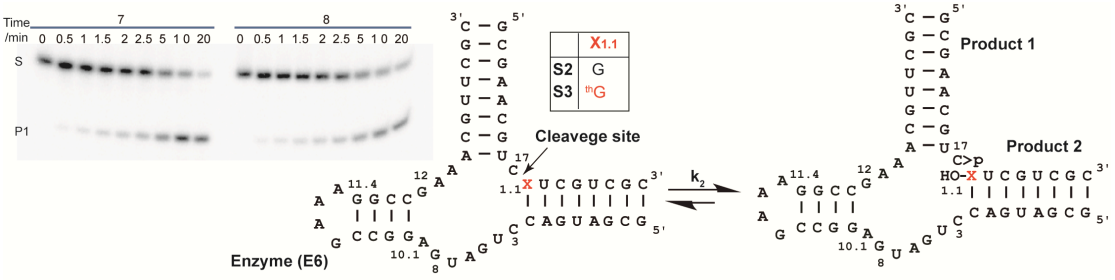

Figure 2.7. Hammerhead ribozymes and cleavage reactions of G1.1-S (**S2**), thG1.1-S (**S3**) and corresponding enzyme (**E6**). Hammerhead ribozyme cleavage reactions, **S2-E6** (7) and **S3-E6** (8), were followed by ^{32}P radioactive labeling of substrate strands. **S** and **P1** indicate the substrates and the product strands respectively.

The cleavage reaction of **E6-S3** (0.06 min^{-1}) was also followed with a

nonradiolabeled substrate by monitoring changes in fluorescence under the same experimental conditions as for the radiolabeled constructs but in a slightly larger scale (Figure 2.8.c, d). The initial increased fluorescence intensity (Figure A16., Figure 2.8.b) was likely due to the cleavage of **S3**, which resulted in the release of product **3** from the duplex with thG1.1 at its 5'-terminus. Good agreement between the radioactively monitored HH reaction and the fluorescence-monitored one was seen (Figure 2.8.d).

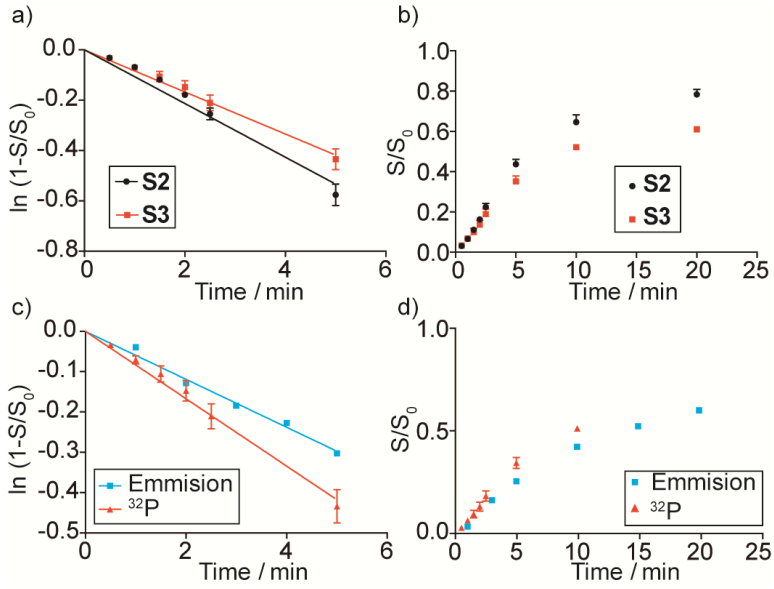

Figure 2.8. Cleavage reactions of **E6** with **S2** and **E6** with **S3**. a) Initial kinetics of **E6** with the native **S2** (black) and the thG-modified substrate (red). The pseudo-first-order rate constants (k_2) of the cleavage reactions are determined as the slope of semi-logarithmic plot of the fraction cleaved as function of time. b) Ribozyme-mediated cleavage curves as determined by ^{32}P data for **E6** with **S2** (black) and with **S3** (red). Fraction cleaved (S/S_0) was determined by dividing the amount of cleaved substrate by the sum of the full length and cleaved substrate. c) Initial kinetics of **E6** with **S3** monitored by radioactive assay (red) and fluorescence spectroscopy (cyan). d) Ribozyme-mediated cleavage curves as determined by ^{32}P data (red) as well as fluorescence spectroscopy (cyan).

3. Discussion

Despite significant progress in solid phase synthesis of RNA oligonucleotides, in vitro transcription reactions frequently are the method of choice for the preparation

of RNA oligonucleotides with long and biologically-relevant sequences. The site-specific incorporation of modified nucleosides through enzymatic approach remains, however, challenging.[12] Paradoxically, this is particularly problematic with highly isomorphic nucleoside surrogates such as thG, since their high resemblance to their native counterparts, leads to total replacement in *in vitro* transcription reactions with the corresponding triphosphates.[8] To overcome such challenges, and generate singly labeled RNA with thG, a faithful and very useful G surrogate, which can be used both as a fluorescent and a mechanistic probe, we sought to advance a universal enzymatic approach. By enforcing transcription initiation with the synthetic nucleoside itself in the presence of all native triphosphates, transcripts labeled at their 5'-end were generated, which could be then phosphorylated and ligated to provide long and singly modified RNA constructs. In this contribution we outline the optimization of this protocol for short and medium size oligonucleotides and then demonstrate that the terminally-modified oligonucleotides can be phosphorylated and ligated by commonly used enzymes. We further then illustrate the utility of such singly labeled RNA oligonucleotides to gain mechanistic insight into the HH ribozyme folding and cleavage mechanism. Furthermore, we show that a HH substrate, singly labeled next to the cleavage position, is cleaved as effectively as the corresponding native RNA construct. Its fluorescence changes upon cleavage can be used in real time to monitor the ribozyme-mediated reaction.

Modifying the 5'-end of RNA oligonucleotides by initiating transcription with GTP alternatives, such as those with modifications on the ribose and conjugated guanosine, including some initiator dinucleotides, have been explored.[9,13] Little has been reported, however, with guanine nucleobase surrogates, where alternative heterocycles are effectively recognized by the enzyme and initiate transcription in

preference over GTP.[13] This has inspired our study here, in which a singly thG-labeled RNA constructs were sought through enzymatic transcription initiation by thG, followed by phosphorylation and ligation to provide internally labeled RNA oligonucleotides.

Transcription initiation with thG, using a short model template, was first explored. T7 RNA polymerase-mediated transcription reactions using template **1** in the presence of ATP, UTP, CTP, GTP and thG yielded not only the full-length thG-initiated transcript but, as expected, also the undesired native transcript (Figure 2.1.). Optimization reactions with different thG concentrations were then carried out to identify the optimal thG/GTP ratio for the production of the modified transcript, eventually settling on 5–8 mM of thG and 1 mM of the native NTPs, depending on the template.

Importantly, the presence of thG at the 5'-terminus of the oligonucleotides did not hamper T4 polynucleotide kinase-mediated phosphorylation and T4 DNA ligase-mediated ligation. This thus allowed us to generate internally modified long RNA constructs. To illustrate the utility of this method and its potential for enhancing mechanistic studies and the fabrication of real-time fluorescence-based assays, we explored the role of individual G residues in HH16, a small catalytic RNA.

We have previously demonstrated that neither the overall yield of a HH16 substrate S1 cleavage nor the kinetics were affected by the replacement of all the G residue with the emissive thG analogs in the substrate strand (thG-S1), illustrating the isomorphic nature of this emissive G sorrogate.[8] In contrast, the activity of the fully modified thG-containing enzyme (thG-E1) was severely diminished, suggesting that the substitution of G for thG interferes with either the folding and/or catalysis of the HH enzyme. Due to the total substitution of all G residues, the impact on individual

residues could not be assessed. The method described in this contribution, facilitating single-site modification, allows one to probe this very question and, in principle, shed light on mechanistic intricacies that emanated from the molecular differences that distinguish thG from G. As thG is typically extremely well tolerated in duplexed regions (see below),[7e] we presumed that individual thG residues in or near active site were responsible for the obliteration of cleavage in the fully modified enzyme.[7a] We thus primarily focused on G residues within the catalytic core, including G8, G10.1, G12 that had been demonstrated to be important for the cleavage reaction.[14] Each residue has been therefore replaced, one at a time, with thG, using the new approach outlined above.

To visualize the impact of substituting each one of these residues, substrate **S1** was radioactively labeled and subjected to ribozyme-mediated cleavage with the site-specifically modified HH enzymes **E1**–**E5** (Figure 2.4.). Comparing to the native enzyme **E1**, the cleavage ability of **E2** and **E3** significantly diminished, while that of **E4** was greatly elevated, and **E5** cleaved **S1** at approximately the same rate as **E1**. As articulated below, some of these results can be easily rationalized, while others appear more challenging to explain, suggesting a rather high susceptibility of certain HH position to nucleobase alterations.

In the proposed ribozyme cleavage mechanism,[14] a divalent metal ion, which plays an important role in stabilizing the tertiary structure of the folded HH ribozyme, coordinates G10.1 through its N7, which is missing in thG. This could therefore explain the low activity of **E3** (Figure 3.1.).

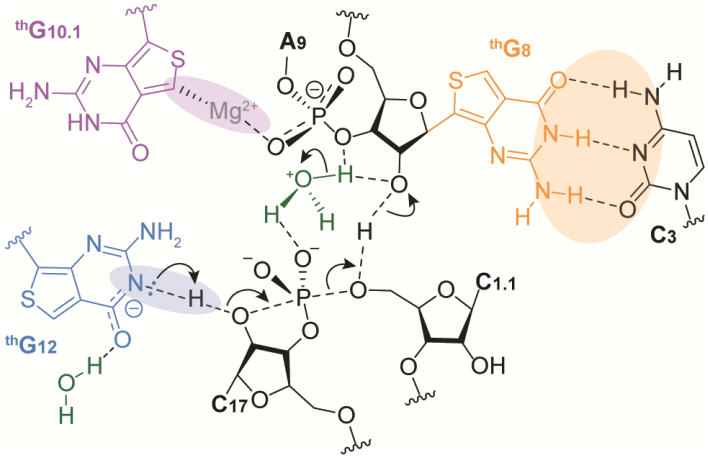

Figure 3.1. Schematic representation of the HH cleavage reaction with highlighted the specific role and interactions of each thG nucleotide.

To study the influence of replacing G residue with thG at positions that could potentially be more tolerant to variations, we replaced G11.4, which is part of helix II, and thus is not directly involved in the catalysis process, with thG. This substitution significantly enhanced the cleavage rate. Previous studies have shown that improved cleavage rate could be obtained by specific mutations at auxiliary sites of HH ribozymes,[14, 15] but not, to our knowledge, with unnatural nucleosides. The HH ribozyme undergoes a conformational rearrangement, which, based on previous studies, requires an optimal combination of sufficient stability and conformational flexibility of the stem-loop II.[16, 17] This might provide a clue towards understanding the high cleavage rate of **E4**. It is plausible that the difference in the aromaticity and the higher hydrophobic characters of thG in comparison to G affect the dynamic of **E4**, favoring its catalytically active conformations.

Interestingly, **E5** cleaves the native substrate at a comparable rate to that of **E1**, the native enzyme (Figure 2.6.b, c). The nucleobase of G12 is involved in the cleavage reaction where the putatively deprotonated N1 acts as a base to abstract the proton on the O2' of C17, which as a nucleophile attacks the adjacent 3' phosphate, leading to

strand cleavage (Figure 3.1.). Previously reported theoretical calculation[18] and our experimental determination (Figure A17.) suggest high similarity between the N1 acidity of thG compared to the native G (pK_a 10.1 and 9.2–9.6 respectively).[19] Thus, the HH cleavage process is not significantly disturbed by the replacement of G12 with a synthetic thG surrogate.

Perhaps surprisingly, the substitution of G8 with thG in **E2** severely reduced the cleavage rate of the native substrate strand, although only the ribose of this nucleotide has been proposed to directly impact the catalytically active conformation.[14] We speculate that substituting the invariant G8 for thG might subtly impact the tertiary WC base-pairing with the invariant C3. This tertiary pair appears to be the "Achilles heel" of HH16, as any modification significantly diminishes cleavage, and the "compensatory" G•C to C•G double mutation has been shown to only partially restored activity.[20,21] Hammann and coworkers have indicated that tolerance to the exchange of WC base pair between position 3 and 8 can depend on the respective sequence context.[22] While thG has been shown to form highly stable WC pairs in duplexes, tertiary pairing has not yet been explored. Inferring from the elevated stability of thG•C vs. G•C in double stranded constructs, which is likely due to the higher "stackability" of the former, such a replacement may again impact the dynamics of the HH16 fold, thus impacting its active conformer accessibility.[7a,21] It is plausible that minor differences in the structure and stability of the modified and native base pair influence the HH ribozyme dynamic equilibrium and lower the population of the active conformation.[21,22]

The results described above reflect the complex and intricate molecular interactions involved in a ribozyme-mediated cleavage (Figure A18.). We appreciate the fact that any substitution of a native residue for a synthetic one, regardless how

"isomorphic" the modification might be, clearly impacts multiple molecular and supramolecular features (as a result of different hydrogen bonding strengths, pK_a values, stackability, etc.). Yet, while one might question the actual insight gained by such isomorphic replacement, we stress that a HH16 substrate, where thG replaces G at the cleavage site, undergoes the expected cleavage reaction at essentially the same rate as a native RNA.

While this thG-containing substrate can be used to monitor the reaction by fluorescence, avoiding radioactive labeling, it serves as a critical control illustrating the true functionality of this G surrogate. This, in turn, suggests, that the differences seen in the various thG-modified enzymes, while not always fully decipherable, indeed reflect subtle molecular features that impact the HH dynamics and conformation.

4. Conclusions

Our results show that thG, as a free nucleoside, can be used to initiate in vitro transcription reactions, thus generating 5'-end modified RNA constructs. These emissive strands can be easily phosphorylated and ligated to provide longer internally modified RNA strands. The modified RNAs containing single thG substitution, in addition to potentially serving as fluorescence probes, provide mechanistic insight into the HH folding and dynamics. We anticipate this approach can be employed to address additional mechanistic questions in regarding the impact of specific G residues on RNA function and recognition.

Chapter one, as well as the abstract of thesis, in full, is currently being prepared for submission for publication for the material. Li, Y.; Fin, A.; McCoy, L.; Tor, Y. Polymerase-mediated site-specific incorporation of a synthetic fluorescent isomorphic

G surrogate into RNA. The thesis author was the primary investigator and author of this material.

5. Future Directions

Recently in the Tor group an evolved isomorphic emissive guanosine analog (^{tz}G), derived from isothiazolo[4,3-d]pyrimidine core (Figure 5.1.), was introduced.[23] The restoration of the nitrogen in the N7 position is of crucial importance considering the key role played by the N7 in the formation of Hoogsteen base pairing, RNA/RNA, RNA/protein interactions and in enzymatic processes by coordinating metal ions.

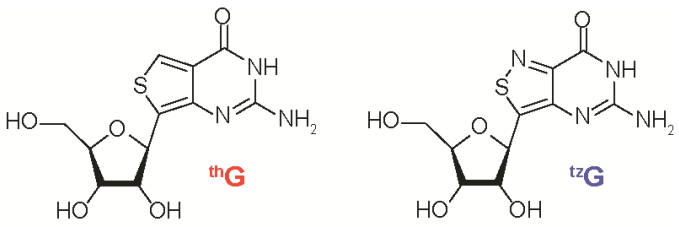

Figure 5.1. Chemical structure of thG and ^{tz}G nucleosides.

More specifically, the new guanosine analog could be employed to shed light on the HH ribozyme substituted at the position 10.1. Different from thG labeled HH ribozyme 10.1, the ^{tz}G analog should be able to coordinate the magnesium ion, as in the proposed mechanism, leading to a cleavage kinetic comparable to the native ribozyme. This would strengthen and confirm the role, at molecular level, of the G10.1 in the proposed enzymatic mechanism.

6. Experimental

6.1. Materials

Solvents, buffers and salts were purchased from Sigma-Aldrich, Fluka, TCI, and Acros and were used without further purification unless otherwise specified. NTPs were purchased from Fisher, the enzymes were purchased from New England Biolabs or Promega. The oligonucleotides were purchased from IDT and further purified by gel electrophoreris and subjected to standard desalting protocols.

6.2. Synthetic procedures

Ribonucleoside, thG, was synthesized as previously published.[7a]

6.2.1. RNA oligonucleotides synthesis

Transcripts **E1**, **S1**, **S2**, **S3** as well as donor strands **1b**, **2b**, **3b**, **4b** and **5b** were transcribed using T7 RNA polymerase. Each single strand DNA template was annealed to a consensus 18-mer T7 RNA polymerase promoter in TE buffer (10 mM Tris-HCl, 1 mM EDTA, 100 mM NaCl, pH 7.8) by heating a 1:1 mixture (10 μM) at 90 °C for 3 min and cooling the solution slowly down to room temperature. Large-scale transcription reactions were performed in buffer (40 mM Tris-HCl, pH 7.9) containing annealed templates (500 nM), dithiothreitol (10 mM DTT), NaCl (10 mM), spermidine (2 mM), RNase inhibitor (RiboLock, 1 U / μL), and T7 RNA polymerase (0.1μg / μL) in a total volume of 250 μL. For the native transcripts **E1**, **S1** and **S2**, ATP (2 mM), GTP (2 mM), CTP (2 mM), UTP (2 mM) and $MgCl_2$ (20 mM) were used.

Various concentrations of NTPs and thG as well as $MgCl_2$ were used for different templates to prepare 5'-thG transcripts, as indicated below. For template **1**: ATP (1 mM), GTP (1 mM), CTP (1 mM), UTP (1 mM), thG (5 mM) and $MgCl_2$ (10

mM); template **2**: ATP (2.5 mM), GTP (2.5 mM), CTP (2.5 mM), UTP (2.5 mM), **thG** (18.75 mM) and $MgCl_2$ (20 mM); template **3** and **4**: ATP (2 mM), GTP (2 mM), CTP (2 mM), UTP (2 mM), thG (15 mM) and $MgCl_2$ (20 mM); template **5**: ATP (1 mM), GTP (1 mM), CTP (1 mM), UTP (1 mM), thG (7.5 mM) and $MgCl_2$ (10 mM).

The reaction was incubated for 4 hours at 37 °C. The precipitated magnesium pyrophosphate was removed by centrifugation. Loading buffer (125 μL) was added after the reaction was concentrated to half its volume. The mixture was heated up for 3 minutes at 75 °C and loaded onto a preparative 20% denaturing polyacrylamide gel. The gel was UV shadowed; appropriate bands were excised, extracted overnight with ammonium acetate (0.5 M), and desalted on a Sep-Pak C18 column. Concentrations of the RNA transcripts were determined using UV absorption spectroscopy at 260 nm, using the following molar extinction coefficients: C (ε_{260} = 7200 L·mol^{-1}cm^{-1}), U (ε_{260} = 9900 L·mol^{-1}cm^{-1}), G (ε_{260} = 11500 L·mol^{-1}cm^{-1}), A (ε_{260} = 15400 L·mol^{-1}cm^{-1}) and thG (ε_{260} = 5517 L·mol^{-1}cm^{-1}).

6.2.2. 5' Phosphorylation

RNA oligonucleotides containing thG at their 5'-end (transcripts **2b**, **3b**, **4b**, **5b** and donor **S3**) were phosphorylated with T4 polynucleotide kinase. Large-scale phosphorylation reactions containing thG-initialed RNA oligonucleotides (1-2 nmol, 6 μM) were performed in kinase buffer (1X New England Biolabs), additional DTT (5 mM), ATP (1 mM) and T4 polynucleotide kinase (0.2 U/ μL New England Biolabs). The reaction was incubated for 2 hours at 37 °C and precipitated with ammonium acetate (0.4 M), Glycoblue (100 μg/mL), and cold ethanol (2.5−3 v/v) in dry ice bath for 1 hour, followed by centrifugation (14,000 rpm) for 20 min and removal of the supernatant. The pellet was washed with cold ethanol (200 μL, 70%) and air-dried for 30 minutes before getting dissolved in gel loading buffer (50 μL). The mixture was

heated for 3 minutes at 75 °C and loaded onto a preparative 20% denaturing polyacrylamide gel. The gel was UV shadowed; appropriate bands were excised, extracted overnight with ammonium acetate (0.5 M), and desalted on a Sep-Pak C18 column. Concentrations of the RNA transcripts were determined using UV absorption spectroscopy at 260 nm as described above.

6.2.3. Splinted ligation

Each 5'-phosphorylated donor strand (10 μM) was mixed with the corresponding acceptor strands (10 μM) and splint and annealed in Tris-HCl buffer (40 mM, pH 7.8) by heating up the solution for 3 minutes at 90 °C and cooling down slowly to room temperature. $MgCl_2$ (10 mM), DTT (10 mM), PEG 4000 (0.5 mM, 0.1 v/v of 50%) and T4 DNA ligase (0.7-1 U/μL, Fermentas) were then added to the mixture. The reaction was incubated at 37 °C for 2 hours before precipitated as described above. The pellet was washed with cold ethanol (200 μL, 70%) and air-dried for 30 minutes before getting dissolved in the gel-loading buffer (50 μL). The RNA was resolved by gel electrophoresis and isolated and desalted as described above.

6.2.4. 5'-^{32}P-Labeling

Native transcript **S1** was first dephosphorylated with calf intestinal alkaline phosphatase (CIP). Transcript **S1** (26 pmol) was mixed with dephosphorylation buffer (6 μL 10X) and CIP (2 μL) in a total volume of 60 μL and incubated at 37 °C for 2 hours. Water (70 μL) was added and the reaction mixture was extracted with phenol:chloroform ($CHCl_3$):isoamyl alcohol (iAA) (100 μL, 25:24:1). The water layer was then extracted with chloroform (100 μL). The RNA in the aqueous layer was precipitated with ammonium acetate (20 μL, 10M), glycoblue (5 μL), EtOH (400 μL) and put in dry ice bath for 1 hour, followed by centrifugation (14,000 rpm) for 20 minutes and removal of the supernatant. The pellet was washed with cold ethanol (4 ·

50 µL, 70%). The pellet was air-dried for 30 minutes and then dissolved in water (38 µL). Kinase buffer (5 µL, 10X), γ-^{32}P ATP (5 µL), of DTT (1 µL), and T4 polynucleotide kinase (1 µL), were added and the reaction was heated to 37 °C for 2 hours. The RNA was then precipitated with ammonium acetate (10 µL, 10 M), glycoblue (2 µL), ethanol (200 µL) and washed with cold ethanol (4 · 25 µL, 70%). The pellet was dissolved in loading buffer (1X TBE, 7M urea), and then the RNA was resolved by gel electrophoresis on a denaturing 20% polyacrylamide gel. The RNA was cut out and extracted with water overnight, filtered, and then concentrated using a speed vac.

6.3. RNA oligonucleotides characterization

MALDI spectra of the oligonucleotides were run either on Applied Biosystems Voyager-DETM Pro or Bruker biflex IV MALDI-TOFMS. The corresponding measuring matrices are described in the figure caption for each oligonucleotide. ESI of the oligonucleotides were run on ThermoFinnigan LCQ DECA XP for ESI.

6.3.1. Digestion with S1 nuclease

E5 (1.5 nmol) was incubated in S1 reaction buffer (1X, Promega) with S1 nuclease for 2 hours at 37 °C, followed with incubation in dephosphorylation buffer (Promega) with alkaline phosphatase for 2 hours at 37 °C. The ribonucleoside mixture obtained was analyzed by reverse-phase analytical HPLC. HPLC analysis was carried out with an Agilent 1200 series system with an Eclipse XDB-C18 5µm, 4.5 · 150 mm column. 0.1% formic acid stock solutions were prepared by dissolving 1 mL of formic acid (Acros, 99%) in 999 mL HPLC grade acetonitrile (Sigma) or MilliQ water and filtered using Millipore type GNWP 0.2 µm filters before use. The injection was subjected to a gradient (12 minutes, from 0 to 6% acetonitrile 0.1% formic acid in water

0.1% formic acid). A flow rate of 1 mL / minute was used and the run was carried out at 25.00 ± 0.10 °C. Each run was monitored at 260 and 321 nm with calibrated references at 650 nm and slit set at 1 nm.

6.3.2. Digestion with T1 nuclease

Each ^{32}P-labeled **E1**, **E2**, **E3**, **E4**, **E5**, **S2** and **S3** was digested with T1 nuclease. Each RNA (2 μL, about 60,000 cpm) was first denatured in T1 denaturing buffer (pH 5) containing of sodium citrate (20 mM), urea (7M) and 1 mM EDTA (1 mM) by heating up the solution at 55 °C for 3 minutes. Yeast tRNA (2 μg) and T1 nuclease (1 U) were added after the mixture was cooled down to room temperature, resulting in a total volume of 10 μL. The reaction mixture was incubated at room temperature for 20 minutes before precipitated and washed with ethanol as described above. The pellet was dissolved in stop buffer (8 μL) containing formamidine (95%), EDTA (20 mM), bromphenol blue (0.05%) and xylene cyanol (0.05%) and then loaded on 15% polyacrylamide with urea (7M) gel. Corresponding bands were imaged with Personal Molecular Imager.

6.3.3. Alkaline hydrolysis

Each ^{32}P-labeled RNA (2 μL, about 60,000 cpm) was incubated in alkaline hydrolysis buffer (pH 9.2) containing sodium carbonate (50 mM) and EDTA (1 mM) in a total volume of 15 μL at 90 °C for 20 minutes before precipitated and washed with ethanol as described above. The pellet was dissolved in stop buffer (8 μL) containing formamide (97%), EDTA (20 mM), bromphenol blue (0.05%) and xylene cyanol (0.05%) and then loaded on 15% polyacrylamide with urea (7M) gel. Corresponding bands were imaged with Personal Molecular Imager.

6.4. Ribozyme cleavage reaction conditions

Cleavage reactions were conducted in a total reaction volume of 34 μL for **E1**, **E4**, **E5** and 22 μL for **E2**, **E3** with **S1**, and 34 μL for **E6** with **S2** or **S3** for radiography. For the fluorescence-based experiments, a total volume of 125 μL was used. The reactions were carried out at 31 °C in a buffer containing Tris-HCl (50 mM, pH 7.0) and NaCl (200 mM). Buffered solutions of the substrate **S3** (0.6 μM) and enzyme **E6** (6 μM) were denatured separately by heating to 90 °C for 3 minutes and cooled down to room temperature over 10 minutes to allow for refolding. $MgCl_2$ was added to both the enzyme and substrate to make a final concentration of 10 mM, and both were equilibrated at 31 °C for 10 minutes. The cleavage reaction was then initiated by manually mixing equal volumes of the modified or natural substrate (0.6 μM) with the enzyme or modified enzyme (6 μM) in a heat block at 31 °C, to give final concentrations of 0.3 μM of the substrate and 3 μM of the enzyme and 10 mM $MgCl_2$.

6.4.1. Ribozyme cleavage assay

For initial data points (time = 0), an aliquot of the substrate **S1**, **S2** or **S3** (2 μL) was removed immediately prior to starting the reaction. Following initiation of the reaction, aliquots (4 μL) were removed at designated time periods and quenched with urea containing loading buffer (15 μL, 7 M urea, 1· TBE buffer, 0.05% bromophenol blue, and 0.05% xylene cyanol). Each tube was heated up to 90 °C for 1.5 minutes and loaded on a 20% polyacrylamide with urea (7M) gel. Corresponding bands were quantified on a Personal Molecular Imager and analyzed with Quantity One software (Bio-Rad).

6.4.2. Ribozyme cleavage data analysis

Enzymatic process rate constants (k_2) were calculated as the slope of the semi-logarithmic plot of ln $(1 - S/S_0)$ versus time, where S/ S_0 is the fraction of cleaved

substrate. For experiments utilizing a radioactively labeled substrate, S/S_0 was determined by dividing the amount of cleaved substrate by the sum of the full length and cleaved substrates.

6.5. Absorption and emission spectroscopy assays and data

Absorption spectra were measured on a Shimadzu UV-2450 spectrophotometer setting the slit at 1 nm and using a resolution of 0.5 nm. All the spectra were corrected for the blank. Steady state emission spectra were measured on a Horiba Fluoromax-4 equipped with a cuvette holder with a stirring system setting both the excitation and the emission slits at 3 nm, the resolution at 1 nm and the integration time 0.1 s if not otherwise described. The steady state fluorescence spectra were performed upon excitation at a wavelength according to the following equation.

$$\lambda_{exc}(nm) = \lambda_{abs}(nm) - 5\ (nm)$$

All the spectra were corrected for the blank. Both instruments were equipped with a thermostat controlled ethylene glycol-water bath fitted to specially designed cuvette holder and the temperature was kept at 25.00 ± 0.10 °C if not otherwise described. thG was dissolved in DMSO (3.47 mM) to prepare highly concentrated stock solution.

6.5.1. pK_a determination

Aqueous stock solutions (100 mL) were prepared by mixing aqueous sodium phosphate monobasic (0.5 M), aqueous sodium phosphate dibasic (0.5 M) and aqueous sodium chloride (2 M) to have a final concentration of 100 mM NaCl and 10 mM phosphate ions. The pH of each solution was adjusted to the desire value by adding aliquots of 2 M aqueous HCl or 2 M aqueous NaOH prior to spectral measurements. thG

was dissolved in DMSO (3.5 mM) to prepare highly concentrated stock solution.

In a typical experiment, aliquots (10 μL) of the concentrated DMSO solution were diluted with air-saturated solvents (3 mL). The solution was mixed with a pipette for 10 seconds and placed in the cuvette holder at 25.00 ± 0.10 °C for 3 minutes before spectra were recorded. All sample contain 0.3 v/v % of DMSO.

The absorption (λ_{abs}) and the emission (λ_{em}) maxima were plotted versus the pH and fitted using a Boltzmann sigmoidal curve using Kaleida Graph 3.5. The pK_a values were determined by interpolation of the fitting curves. The reported pK_a values represent the average of three independent sets of measurements. The values and the relative standard deviations are reported in table 6.1.

Table 6.1. thG pK_a values

	pK_a	
Absorbance	5.06 ± 0.06 (N3)	10.08 ± 0.08 (N1)
Emission	5.17 ± 0.05 (N3)	7.95 ± 0.09 (N1)

6.5.2. Ribozyme cleavage emission assay

The ribozyme cleavage experiment was carried out on a total volume of 125 μL in a 125 μL cuvette. The reactions were carried out at 31 °C in a buffer containing Tris-HCl (50 mM, pH 7.0) and NaCl (200 mM). Buffered solutions of the substrate **S3** (0.6 μM) and enzyme **E6** (6 μM) were denatured separately by heating to 90 °C for 3 minutes and cooled down to room temperature over 10 minutes to allow for refolding. $MgCl_2$ was added to both the enzyme and substrate to make a final concentration of 10 mM, and both were equilibrated at 31 °C for 10 minutes. The cleavage reaction was then initiated by manually mixing equal volumes of the modified or natural substrate (0.6 μM) with the enzyme or modified enzyme (6 μM) in the cuvette 31 ± 0.10 °C, to give final concentrations of 0.3 μM of the substrate and 3 μM of the enzyme and 10 mM

$MgCl_2$. The reaction was followed by recording emission spectra (430 –480 nm) upon excitation at 360 nm over time. The excitation and the emission slits were set at 10 nm, the resolution at 1 nm and the integration time 0.1 s.

Appendix

Polyacrylamide gels

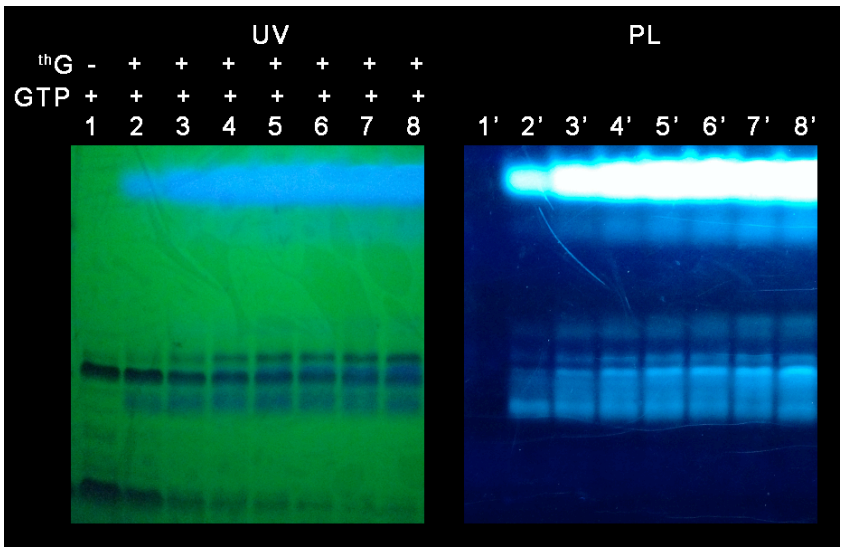

Figure A 1. Transcription reactions of template 1 with various thG/GTP ratios. Lane 1, 1 mM of each NTP (ATP, GTP, CTP, UTP); lane 2, 1 mM of each NTP, 1 mM thG; lane 3, 1 mM each NTP, 3 mM thG; lane 4, 1 mM each NTP, 5 mM thG; lane 5, 1 mM each NTP, 7 mM thG; lane 6, 1 mM each NTP, 9 mM thG; lane 7, 1 mM each NTP, 11 mM thG; lane 8, 1 mM each NTP, 13 mM thG.

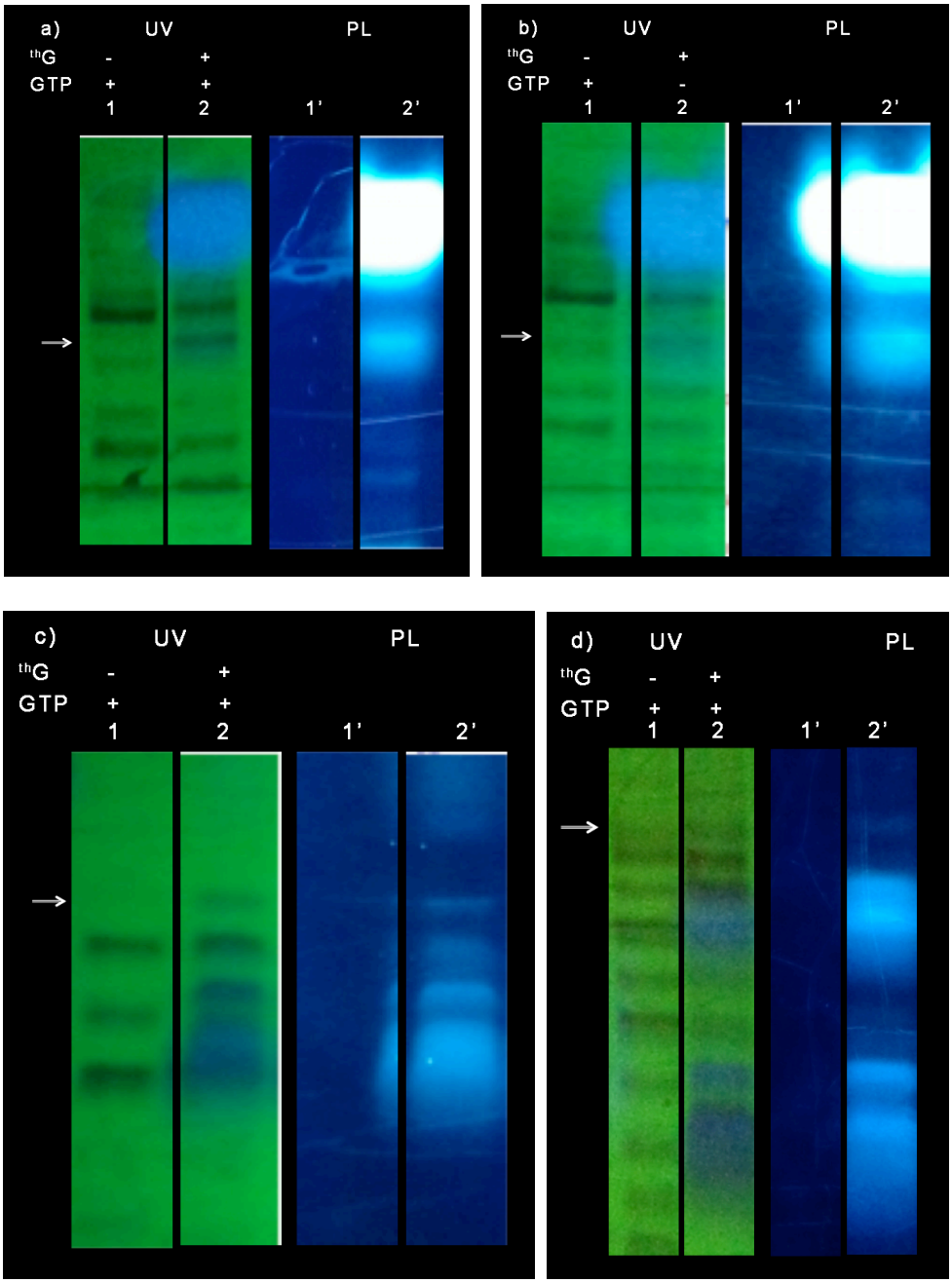

Figure A 2. Transcription reactions of HH Ribozyme donor strands. The arrows indicate positions of full-length 5'-thG transcripts. a) transcription reactions of template **3**; b) transcription reactions of template **2**; c) transcription reactions of template **5**; d) transcription reactions of template **4**.

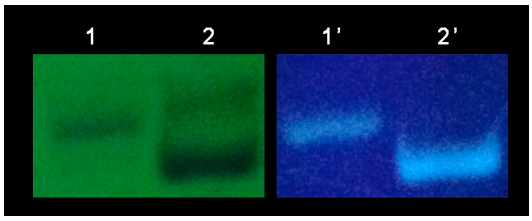

Figure A 3. Phosphorylation of transcript **3b**. Lane 1, non-treated transcript **3b**; lane 2, phosphorylation of transcript **3b**.

Acceptor	5′-G CGA UGA UCU GAU-3′
Donor (phosporilated **2b**)	5′-pthGAG GCC GAA AGG CCG AAA CGU UCG C-3′
Splint	3′-C GCT ACT AGA CTA CTC CGG CTT TCC GGC TTT GC-5′
E2 (24%)	5′-G CGA UGA UCU GAU thGAG GCC GAA AGG CCG AAA CGU UCG C-3′
Acceptor	5′-GCG AUG AUC UGA UGA-3′
Donor (phosporilated **3b**)	5′-pthGGC CGA AAG GCC GAA ACG UUC GC-3′
Splint	3′-GC TAC TAG ACT ACT CCG GCT TTC CGG CTT TGC-5′
E3 (35%)	5′-GCG AUG AUC UGA UGA thGGC CGA AAG GCC GAA ACG UUC GC-3′
Acceptor	5′-GC GAU GAU CUG AUG AGG CCG AAA-3′
Donor (phosporilated **4b**)	5′-pthGGC CGA AAC GUU CGC-3′
Splint	3′-CTA GAC TAC TCC GGC TTT CCG GCT TTG CAA GCG-5′
E4 (18%)	5′-GC GAU GAU CUG AUG AGG CCG AAA thGGC CGA AAC GUU CGC-3′
Acceptor	5′-GCG AUG AUC UGA UGA GGC CGA AAG GCC-3′
Donor (phosporilated **5b**)	5′-pthG AAA CGU UCG C-3′
Splint	3′-C TAG ACT ACT CCG GCT TTC CGG C TTT GCA AGC G-5′
E5 (40%)	5′-GCG AUG AUC UGA UGA GGC CGA AAG GCC thGAA ACG UUC GC-3′

Figure A 4. Design and yield of splinted ligation of **E2**, **E3**, **E4** and **E5**.

MALDI and ESI Spectra

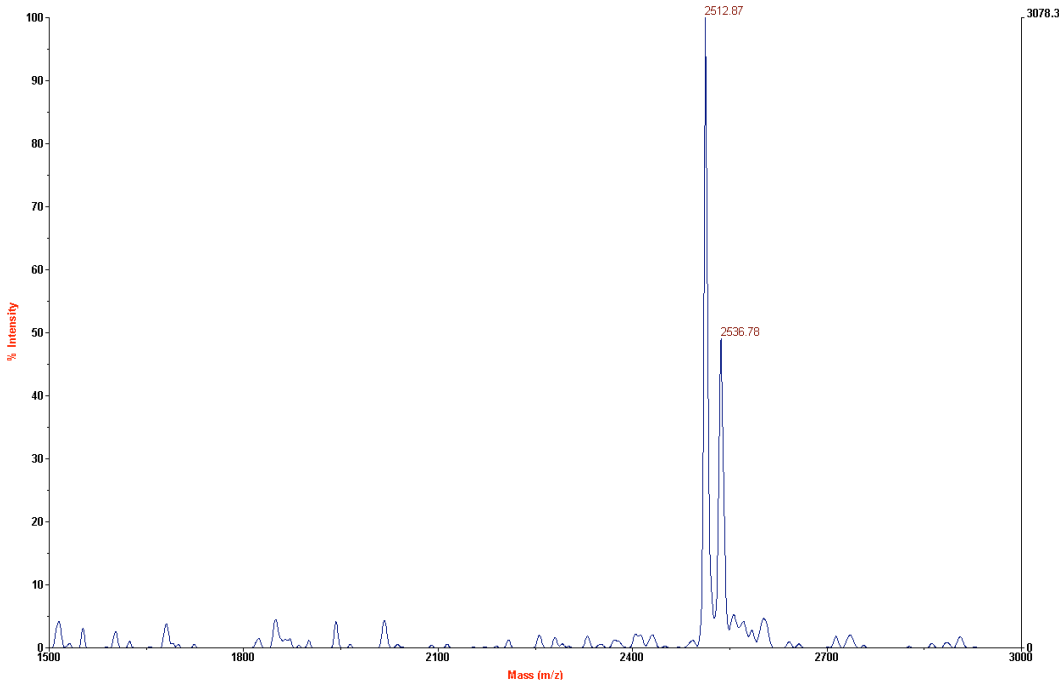

Figure A 5. MALDI-TOF mass spectrum of donor strand of **S3**. Expected molar mass: 2517.56 Da. Taken by MALDI-TOF mass spectrometer in megetive ion mode with Voyager^{TM} Biospectrometry^{TM} Workstation software. The raw mass spectra data was analyzed with Data Explorer version 4.0.0.0. The matrix was 20 g/L 2,4,6-trihydroxyacetophenone (THAP) and 4 g/L diammonium hydrogencitrate dibasic in 50% ACN.

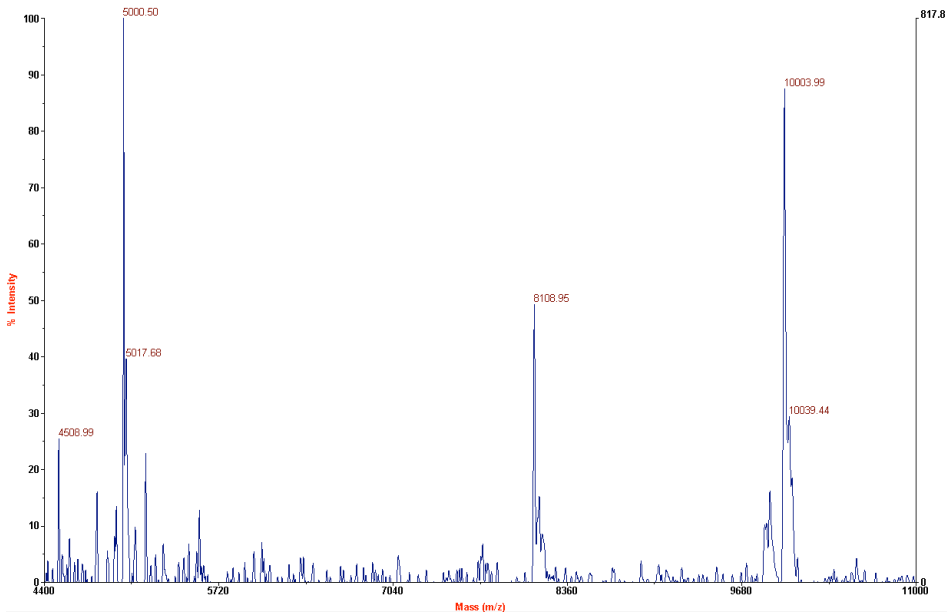

Figure A 6. MALDI-TOF mass spectrum of transcript **2b** with standard. Expected molar mass of transcript **2b** was 8113.95 Da, and expected molar mass of the standard oligonucleotide was 10007.90 Da. Taken by MALDI-TOF mass spectrometer in negative ion mode with Voyager™ Biospectrometry™ Workstation software. The raw mass spectra data was analyzed with Data Explorer version 4.0.0.0. The matrix was 20 g/L 2,4,6-trihydroxyacetophenone (THAP) and 4 g/L diammonium hydrogencitrate dibasic in 50% ACN.

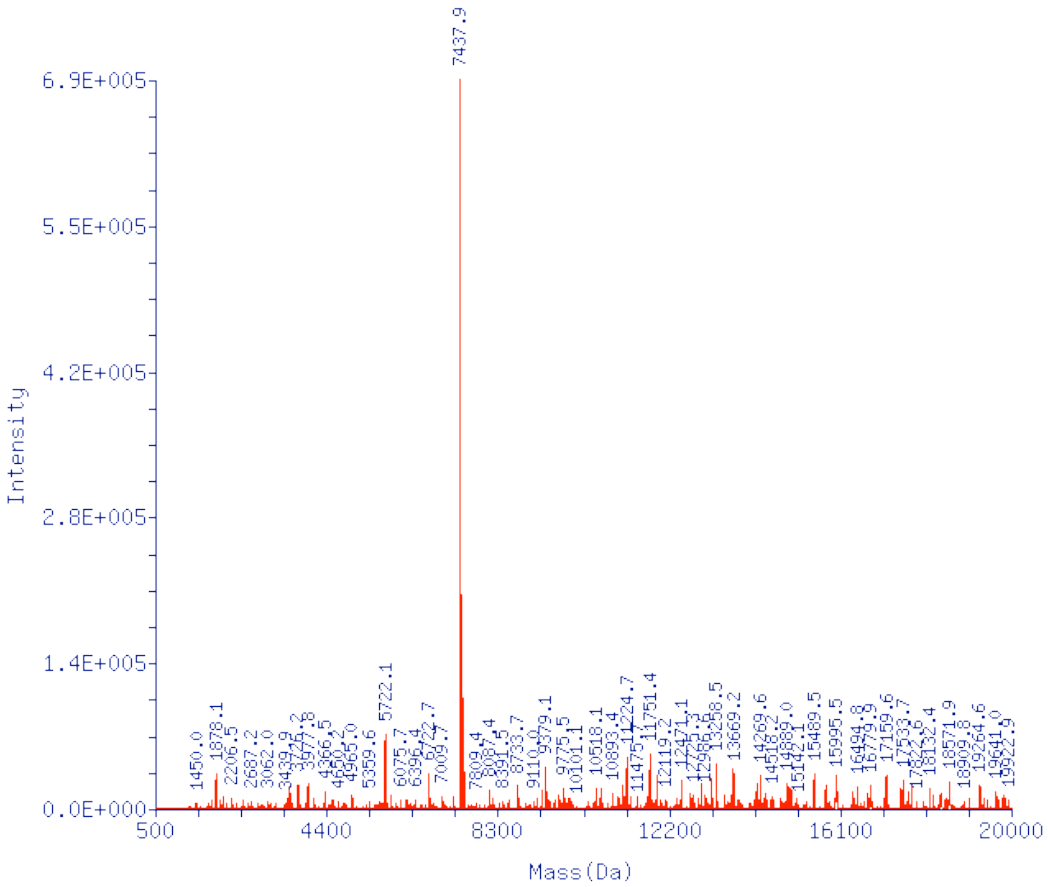

Figure A 7. ESI mass spectrum in negative ion mode transcript **3b** with Xcalibur software version 1.3, and the raw ESI-MS m/z data were deconvoluted by ProMass for Xcalibur Version 2.5 SR-1. The running buffer was 10 mM *tert*-butylamine in 70% acetonitrile in water.

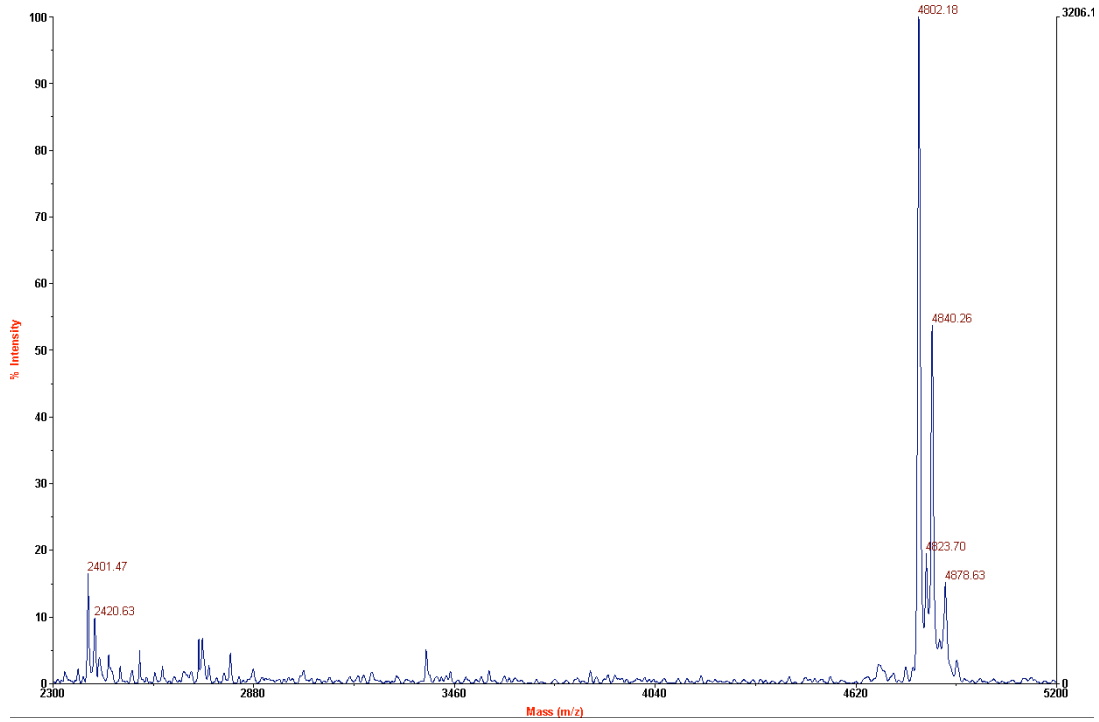

Figure A 8. MALDI-TOF mass spectrum of transcript **4b**. Expected molar mass: 4805.95 Da. Taken by MALDI-TOF mass spectrometer in negative ion mode with Voyager™ Biospectrometry™ Workstation software. The raw mass spectra data was analyzed with Data Explorer version 4.0.0.0. The matrix was 20 g/L 2,4,6-trihydroxyacetophenone (THAP) and 4 g/L diammonium hydrogencitrate (dibasic) in 50% ACN.

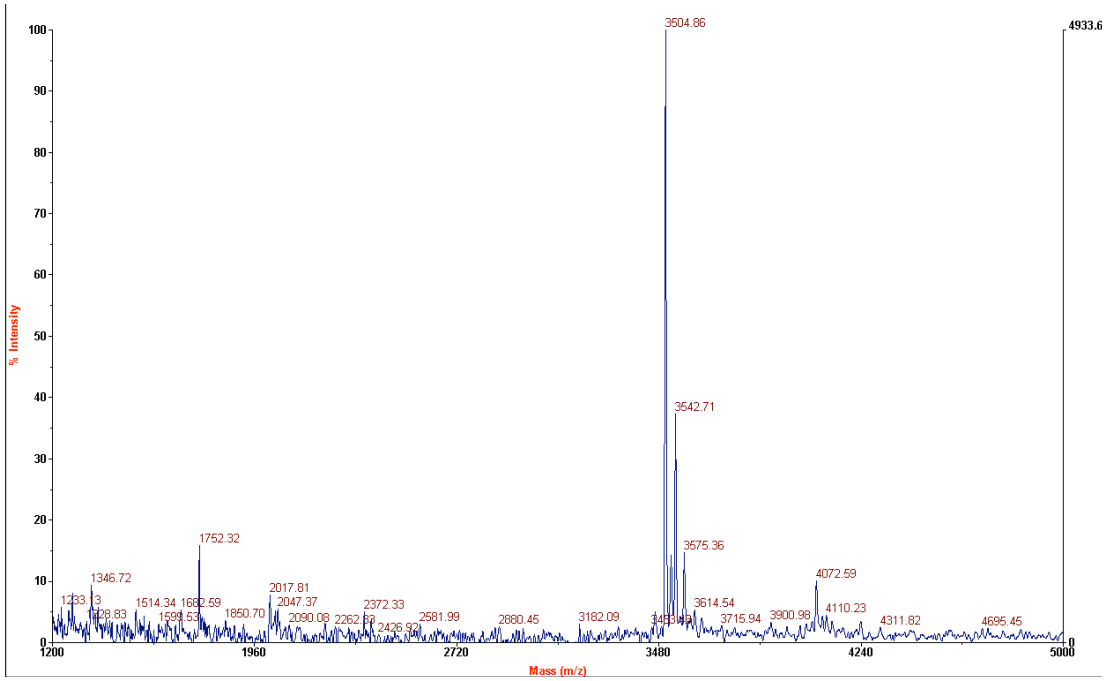

Figure A 9. MALDI-TOF mass spectrum of transcript **5b**. Expected molar mass: 3505.19 Da. Taken by MALDI-TOF mass spectrometer in negative ion mode with Voyager™ Biospectrometry™ Workstation software. The raw mass spectra data was analyzed with Data Explorer version 4.0.0.0. The matrix was 20 g/L 2,4,6-trihydroxyacetophenone (THAP) and 4 g/L diammonium hydrogencitrate dibasic in 50% ACN.

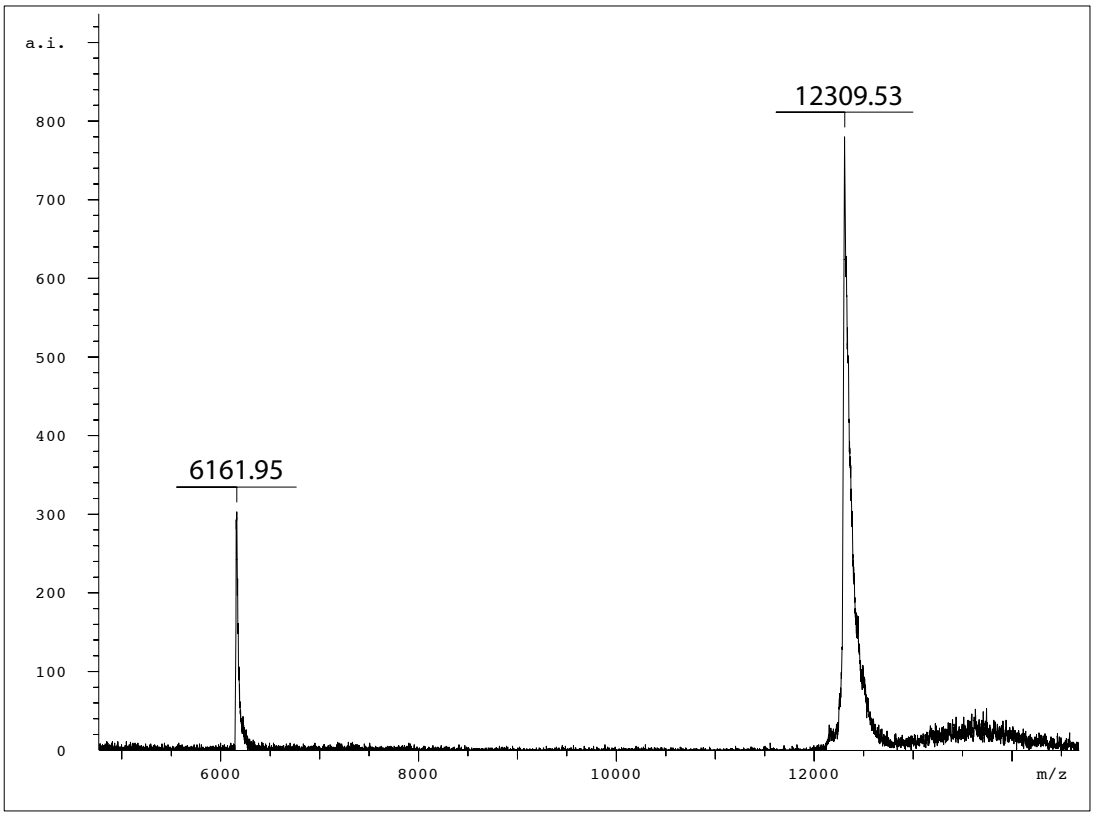

Figure A 10. MALDI-TOF mass spectrum of **E2**. Expected molar mass: 12317.38 Da. Taken by Bruker MALDI-TOF mass spectrometer in negative ion mode with FLEX Control software. The raw mass spectra data was analyzed with X-TOF. The matrix was 45.4 g/L 3-Hydroxypicolinic acid (3-HPA) and 10 g/L diammonium hydrogencitrate dibasic in 50% ACN.

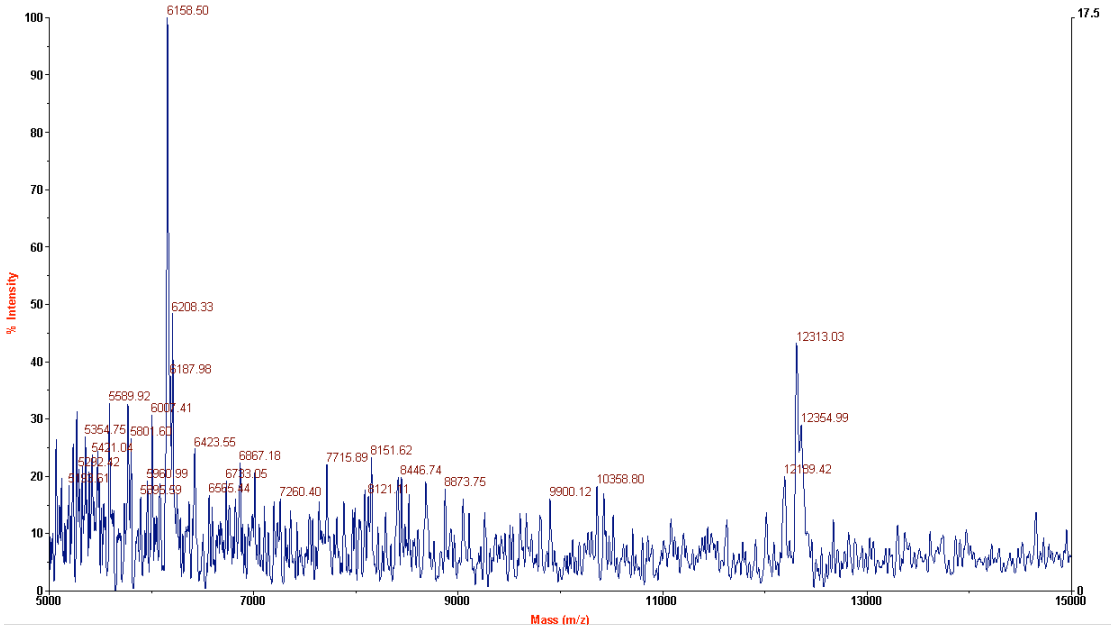

Figure A 11. MALDI-TOF mass spectrum of **E3**. Expected molar mass: 12317.38 Da. Taken by MALDI-TOF mass spectrometer in negative ion mode with Voyager™ Biospectrometry™ Workstation software. The raw mass spectra data was analyzed with Data Explorer version 4.0.0.0. The matrix was 20 g/L 2,4,6-trihydroxyacetophenone (THAP) and 4 g/L diammonium hydrogencitrate dibasic in 50% ACN.

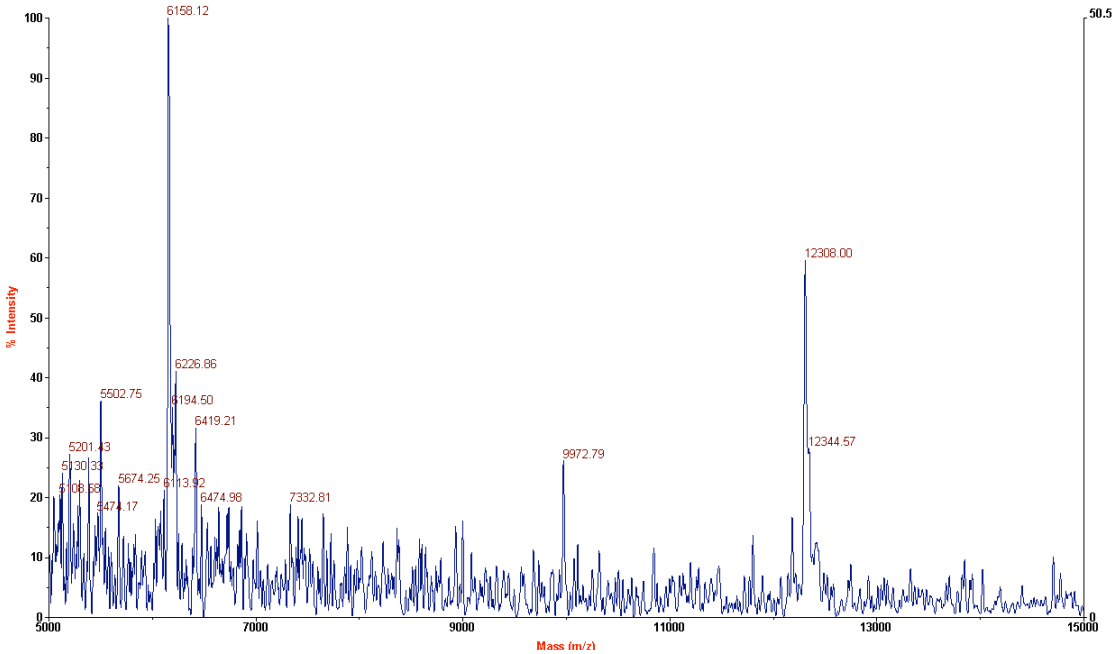

Figure A 12. MALDI-TOF mass spectrum of **E4**. Expected molar mass: 12317.38 Da. Taken by MALDI-TOF mass spectrometer in negative ion mode with Voyager™ Biospectrometry™ Workstation software. The raw mass spectra data was analyzed with Data Explorer version 4.0.0.0. The matrix was 20 g/L 2,4,6-trihydroxyacetophenone (THAP) and 4 g/L diammonium hydrogencitrate dibasic in 50% ACN.

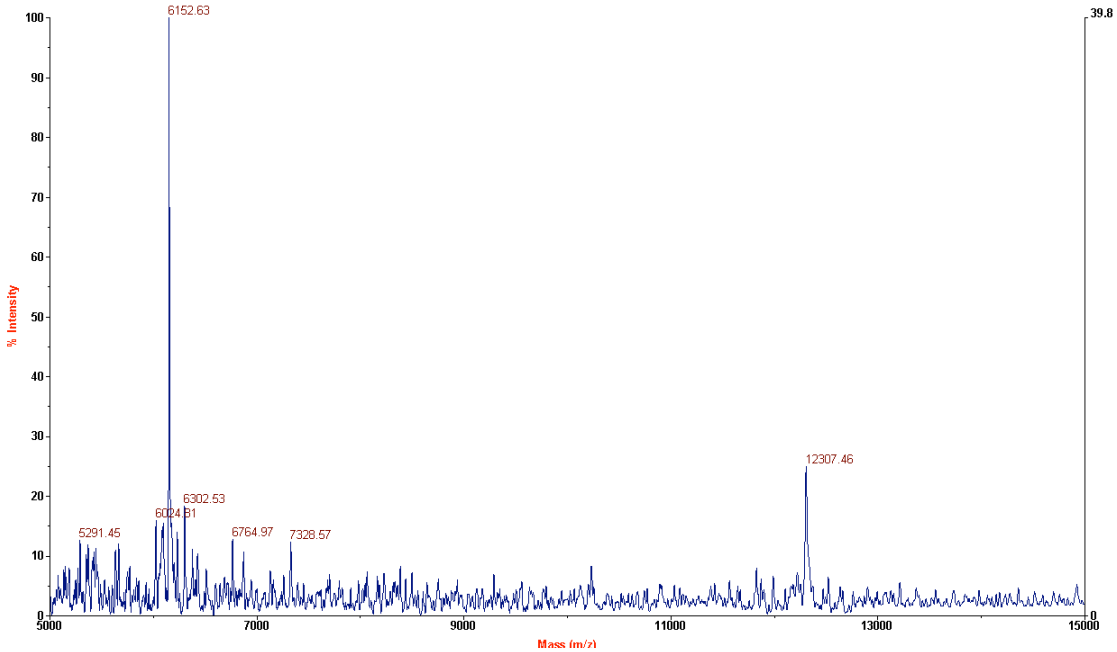

Figure A 13. MALDI-TOF mass spectrum of **E5**. Expected molar mass: 12317.38 Da. Taken by MALDI-TOF mass spectrometer in negative ion mode with Voyager™ Biospectrometry™ Workstation software. The raw mass spectra data was analyzed with Data Explorer version 4.0.0.0. The matrix was 20 g/L 2,4,6-trihydroxyacetophenone (THAP) and 4 g/L diammonium hydrogencitrate dibasic in 50% ACN.

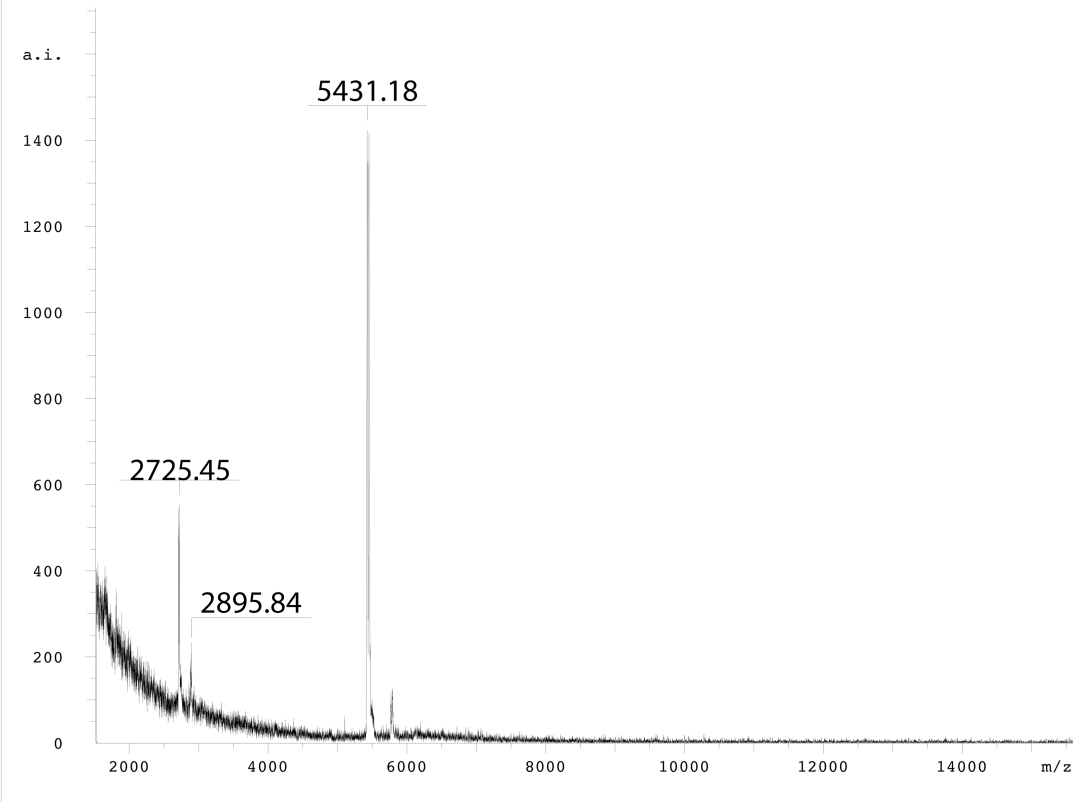

Figure A 14. MALDI-TOF mass spectrum of **S3**. Expected molar mass: 5433.28 Da. Taken on a Bruker MALDI-TOF mass spectrometer in negative ion mode with FLEX Control software. The raw mass spectra data was analyzed with X-TOF. The matrix was 45.4 g/L 3-Hydroxypicolinic acid (3-HPA) and 10 g/L diammonium hydrogencitrate dibasic in 50% ACN.

Digestion of ^{32}P-labeled RNA/polyacrylamide gel

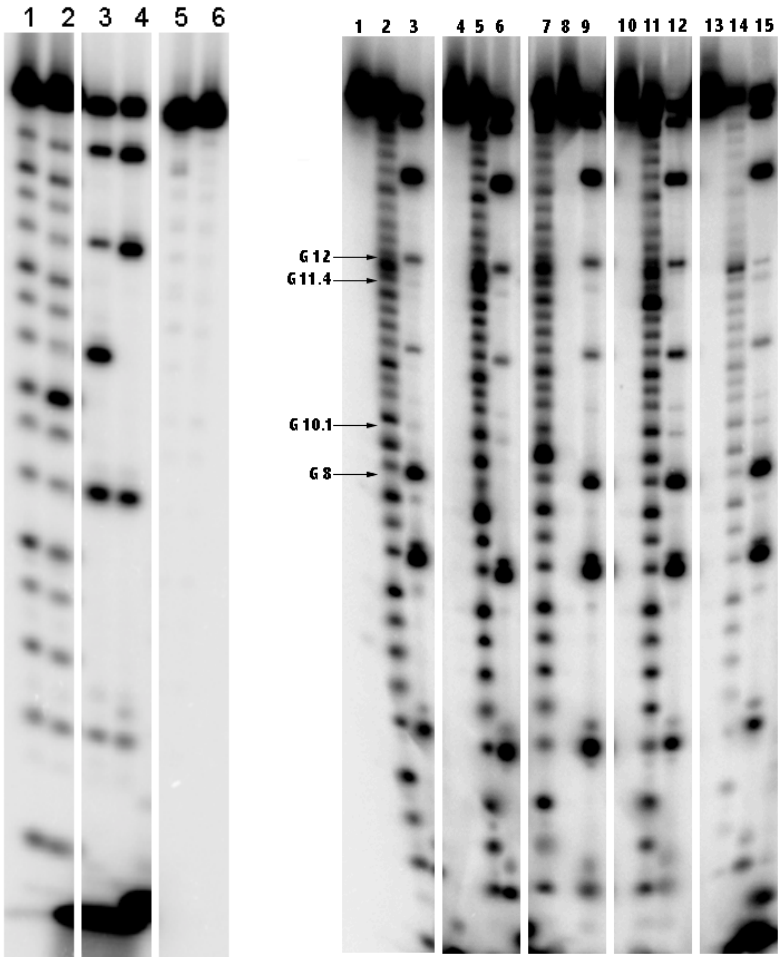

Figure A 15. Alkaline hydrolysis and T1 nuclease digestion of thG-modified RNA oligonucleotides. a) Lane 1 and 2, alkaline hydrolysis of substrates **S2** and **S3**; lane 3 and 4, T1 nuclease digestion of **S2** and **S3**; lane 5 and 6, non-treated **S2** and **S3**. b) Lane 1, non-treated enzyme strand **E1**; lane 2, alkaline hydrolysis of **E1**; lane 3, T1 nuclease digestion of **E1**; **E4**, non-treated enzyme strand **E2**; lane 5, alkaline hydrolysis of **E2**; lane 6, T1 nuclease digestion of **E2**; lane 7, alkaline hydrolysis of enzyme strand **E3**; lane 8, non-treated **E3**; lane 9, T1 nuclease digestion of **E3**; lane 10, non-treated enzyme strand **E4**; lane 11, alkaline hydrolysis of **E4**; lane 12, T1 nuclease digestion of **E4**; lane 13, non-treated enzyme strand **E5**; lane 14, alkaline hydrolysis of **E5**; lane 15, T1 nuclease digestion of **E5**. Position of G8, G10.1, G11.4 and G12 were indicated with arrows. T1 digestion at guanosine residues of helix II of the enzyme strands were not observed even for the native enzyme strand **E1**, which was most likely due to the formation of the duplex structure of helix II.

Fluorescence spectra

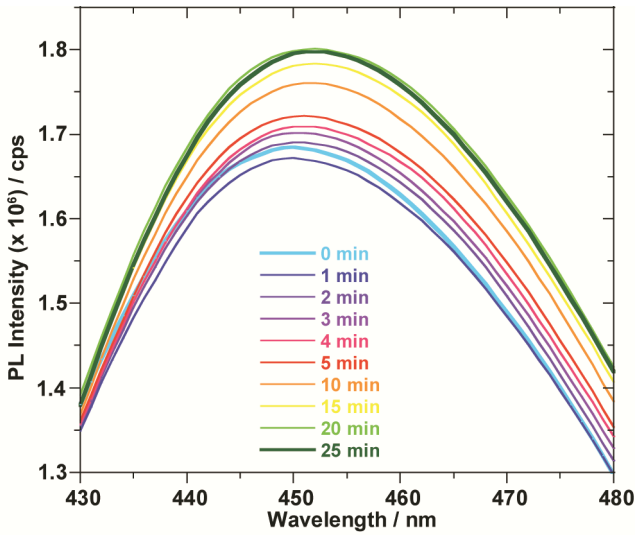

Figure A 16. Emission spectra over time for the enzymatic cleavage of substrate S3 and enzyme **E6**.

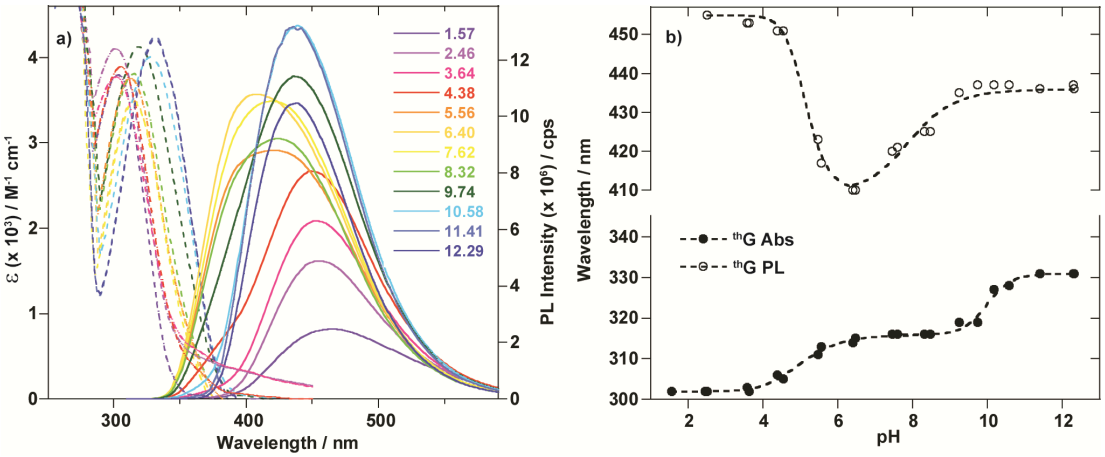

Figure A 17. a) Absorption (dashed lines) and emission (solid lines) traces in buffer solutions at different pH for thG. The emission spectra were normalized to 0.1 intensity at the excitation wavelength. b) Absorption (black) and emission (white) maxima variation versus pH for thG.

Schematic mechanism

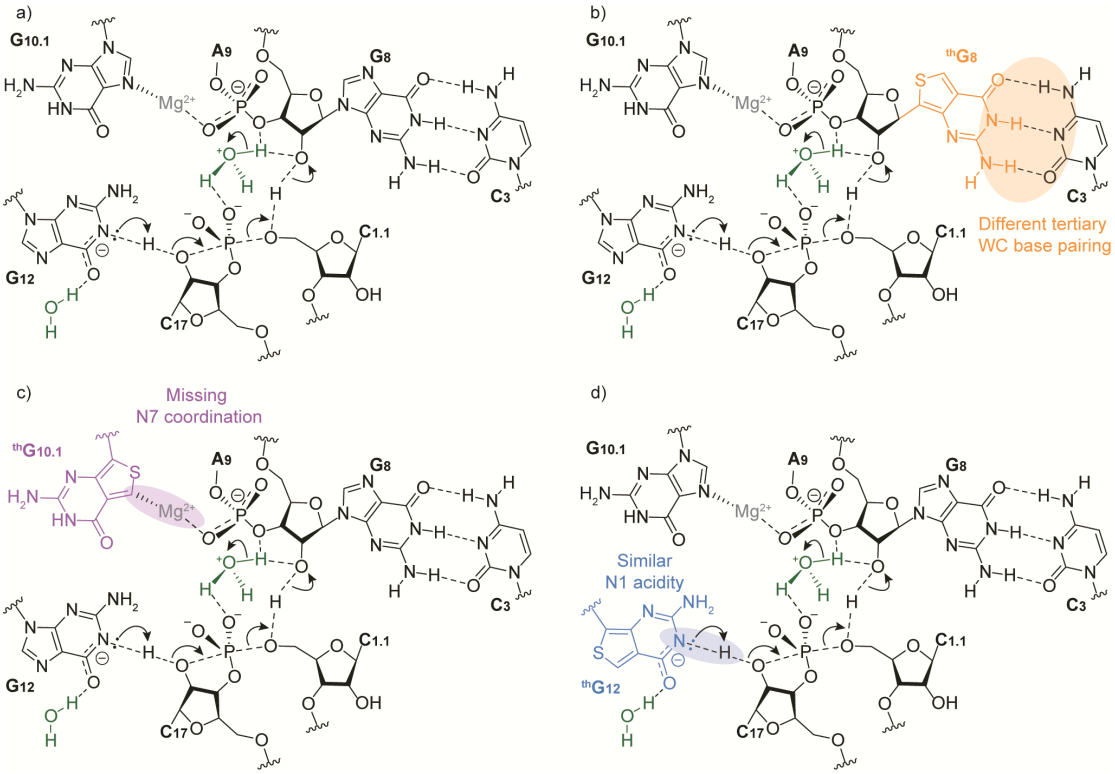

Figure A 18. Schematic representation of the enzymatic cycle for the native **E1** (a), **E2** (b), **E3** (c) and **E4** (d) HH enzymes.

References

1. (a) Allerson, C. R.; Chen, S. L.; Verdine, G. L. *J. Am. Chem. Soc*. **1997**, *119*, 7423–7433;

(b) Earnshaw, D. J.; Gait, M. J. *Biopolymer* **1998**, *48*, 39–55;

(c) Silverman, S. K.; Cech, T. R. Biochem, **1999**, *38*, 14224–14237;

(d) Vrma, S.; Vaish, N. K.; Eckstein, F. *Applications of Ribonucleotide Analogues in RNA Biochemistry*. In RNA; Söll, D.; Nishimura, S.; Moore, P. Ed; Elsevier Ltd., The Netherlands, 2001; 259–275;

(e) Chow, C. S.; Mahto, S. K.; Lamichhane, T. N. *ACS Chem. Biol.* **2008**, *3*, 30–37;

(f) Onizuka, K.; Taniguchi, Y.; Sasaki, S. Bioconjugate Chem. **2009**, *20*, 799–803;

(g) Onizuka, K.; Taniguchi, Y.; Nishioka, T.; Sasaki, S. *Nucleic Acid Symposium Series* **2009**, *53*, 67–68;

(h) Solomatin S. Herschlag, D. *Methods Enzymol.* **2009**, *469*, 47–68;

(i) Bramsen, J. B.; Malgorzata M. Pakula, M. M.; Hansen, T. B.; Bus, C.; Langkjær, N.; Odadzic, D.; Smicius, R.; Wengel, S. L.; Chattopadhyaya, J.; Engels, J. W.; Herdewijn, P.; Wengel, J.; Kjems, J. *Nucleic Acids Res.* **2010**, *38*, 5761–5773;

(j) Sasaki, S.; Onizuka, K.; Taniguchi, Y. *Chem. Soc. Rev*., **2011**, *40*, 5698–5706;

(k) Schulz, D.; Holstein, J. M.; Rentmeister, A. *Angew. Chem. Int. Ed*. **2013**, *52*, 7874–7878;

(l) Edwards, T. E.; Sigurdsson, S. T. Modified RNA as tools in RNA biochemistry. In *Handbook of RNA biochemistry 2nd Edition*; Hartmann, R. K.; Bindereif, A.; Schön, A.; Westhof, E. Ed.; Wiley-VCH: Weinheim, 2014; 151–171;

(m) Oshiro, I.; Jitsuzaki, D.; Onizuka, K.; Nishimoto, A.; Taniguchi, Y.; Sasaki, S. *CheBioChem* **2015**, *16*, 1199–1204.

2. (a) Gait, M. J.; Matthes, H. W. D.; Singh, M.; Sproat, B. S.; Titmas, R. C. *Nucleic Acids Res.* **1982**, *10*, 6243–6254;

(b) Caruthers M. H.; Beaton G.; Wu J. V.; Wiesler W. *Methods Enzymol.* **1992**, *211*, 3–20;

(c) Beaucage, S. L.; Iyer, R. P. *Tetrahedron* **1992**, *48*, 2223–2311;

(d) Beaucage S. L. *Methods Mol. Biol.* **1993**, *20*, 33–61;

(e) Hurley, D. J.; Tor, Y. *J. Am. Chem. Soc.* **1998**, *120*, 2194–2195;

(f) Komatsu, Y; Ohtsuka, E. Chemical RNA Synthesis (Including RNA with Unusual Constituents). In *RNA*; Söll, D.; Nishimura, S.; Moore, P. Ed; Elsevier Ltd., The Netherlands, 2001; 91–107;

(g) Hou, X.; Wang, G.; Gaffney, B. L.; Jones, R. A. *Nucleosides, Nucleotides and Nucleic Acids* **2009**, *28*, 1076–1094;

(h) Zlatev, I.; Lavergne, T.; Debart, F.; Vasseur, J–J.; Manoharan, M.; Morvan, F. *Org. Lett.* **2010**, *12*, 2190–2193;

(i) Höbartner, C. Wachowius, F. Chemical Synthesis of Modified RNA. In *The*

Chemical Biology of Nucleic Acids; G. Mayer Ed; John Wiley & Sons, Ltd, Chichester, UK, 2010; 1–37.

3. (a) Feix, G.; Pollet, R.; Weissmann, C. *Proc. Natl. Acad. Sci. U. S. A.* **1968**, *59*, 145–152;

(b) Cramer, F. *Acc. Chem. Res*. **1969**, *2*, 338–344;

(c) Tor, Y.; Dervan, P. B. *J. Am. Chem. Soc.* **1993**, *115*, 4461–4467;

(d) Hocek, M.; Fojta, M. *Org. Biomol. Chem.* **2008**, *6*, 2233–2241.

4. Srivatsan, S. G.; Tor, Y. *J. Am. Chem. Soc.* **2007**, *129*, 2044–2053.

5. (a) Piccirilli, J. A.; Krauch, T.; Moroney, S. E.; Benner, S. A. *Nature* **1990**, *343*, 33–37;

(b) Seo, Y. J.; Matsuda, S.; Romesberg, F. E. *J. Am. Chem. Soc.* **2009**, *131*, 5046–5047;

(c) Hirao, I.; Kimoto, M.; Yamashige, R. *Acc. Chem. Res.* **2012**, *45*, 2055–2065.

6. Lang, K.; Micura, R. *Nat. Protoc.* **2008**, *3*, 1457–1466.

7. (a) Shin, D.; Sinkeldam, R. W.; Tor, Y. *J. Am. Chem. Soc.* **2011**, *133*, 14912–14915;

(b) Liu, W.; Shin, D.; Tor, Y.; Cooperman, B. S. *ACS Chem. Biol.* **2013**, *8*, 2017–2023;

(c) Sinkeldam, R. W.; McCoy, L. S.; Shin, D.; Tor, Y. *Angew. Chem. Int. Ed.* **2013**, *52*, 14026–14030;

(d) Mizrahi, R. A.; Shin, D.; Sinkeldam, R. W.; Phelps, K. J.; Fin, A.; Tantillo, D. J.; Tor, Y; Beal, P. A. *Angew. Chem. Int. Ed.* **2015**, *54*, 8713–8716;

(e) Sholokh, M.; Sharma, R.; Shin, D.; Das, R.; Zaporozhets, O. A.; Tor, Y.; Mély, Y. *J. Am. Chem. Soc.* **2015**, *137*, 3185–3188.

8. McCoy, L. S.; Shin, D.; Tor, Y. *J. Am. Chem. Soc.* **2014**, *136*, 15176–15184.

9. Milligan, J. F.; Groebe, D. R.; Witherell, G. W.; Uhlenbeck, O. C. *Nucleic Acids Res*. **1987**, *15*, 8783–8798.

10. Pace, C. N.; Heinemann, U.; Hahn, U.; Saenger, W. *Angew. Chem. Int. Ed.* **1991**, *30*, 343–360.

11. (a) Hertel, K. J.; Herschiag, D.; Uhlenbeck, O. C. *Biochemistry* **1994**, *33*, 3374–3385;

(b) Kirk, S. R.; Luedtke, N. W.; Tor, Y. *Bioorg. Med. Chem*. **2001**, *9*, 2295–2301.

12. (a) Sinkeldam, R. W.; Greco, N. J.; Tor, Y. *Chem. Rev.* **2010**, *110*, 2579.

(b) Wilhelmsson, L. M. Q. *Rev. Biophys.* **2010**, *43*, 159–183.

13. (a) Sampson, J. R.; Uhlenbeck, O. C. Proc. *Natl. Acad. Sci. U. S. A.* **1988**, *85*, 1033–10337;

(b) Seeling, B.; Jäschke A. *Tetrahedron Lett*. **1997**, *38*, 7729–7732;

(c) Fiammengo, R.; Musilek, K.; Jaschke, A. *J. Am. Chem. Soc*. **2005**, *127*, 9271–9276;

(d) Williamson, D.; Cann, M. J.; Hodgson, D. R. W. *Chem. Commun.* **2007**, 5096–5098;

(e) Huang, F. Q.; He, J.; Zhang, Y. L.; Guo, Y. L. *Nat. Protoc*. **2008**, *3*, 1848–1861;

(f) Wolf, J.; Dombos, V.; Appel, B.; Muller, S. *Org. Biomol. Chem.* **2008**, *6*, 899–907;

(g) Paredes, E.; Das, S. R. *ChemBioChem* **2011**, *12*, 125–131;

(h) Lee, G. H.; Lim, H. K.; Jung, W.; Hah, S. S. *Bull. Korean Chem. Soc.* **2012**, *33*, 3861–3863;

(i) Samanta, A.; Krause A.; Jäschke A. *Chem. Commun.* **2014**, *50*, 1313–1316.

14. Martick, M.; Lee, T.–S.; York, D. M.; Scott, G. W. *Chem. Biol.* **2008**, *15*, 332–342.

15. (a) Burgin, A. B. Jr.; Gonzalez, C.; Matulic-Adamic, J.; Karpeisky, A. M.; Usman, N.; McSwiggen, J. A.; Beigelman, L. *Biochemistry* **1996**, *35*, 14090–14097;

(b) Clouet-d'Orva, B.; Uhlenbeck, O. C. *Biochemistry* **1997**, *36*, 9087–9092.

16. Wang, S.; Karbstein, K.; Peracchi, A.; Beigelman, L.; Herschlag, D. *Biochemistry* **1999**, *43*, 14363–14278.

17. (a) Long, D. M.; Uhlenbeck, O. C. *Proc. Natl. Acad. Sci. U. S. A.* **1994**, *91*, 6977–6981;

(b) Persson, T.; Hartmann, R. K.; Eckstein, F. *ChemBioChem* **2002**, *3*, 1066–1071.

18. (a) Bundari, S. *The Merck Index*, 12th ed.; Merck and Co., Inc.: Whitehouse Station, NJ, 1996;

(b) Sigel, H.; Massoud, S. S.; Corfù, N. A. *J. Am. Chem. Soc.* **1994**, *116*, 2958–2971;

(c) Kampf, G; Kapinos, L. E.; Gries-ser, R.; Lippert, B.; Sigel, H. J. *Chem. Soc., Perkin Trans. 2* **2002**, 1320–1327;

(d) Thapa, B.; Schlegel, H. B. *J. Phys. Chem. A* **2015**, *119*, 5134-5144.

19. (a) Samanta, P. K.; Manna, A. K.; Pati, S. K. *J. Phys. Chem. B* **2012**, *116*, 7618−7626;

(b) Lee, Y.-J.; Jang, Y. H.; Kim, Y.; Hwang, S. *Bull. Korean Chem. Soc.* **2012**, *33*, 4255–4257;

(c) Samanta, P. K.; Pati, S. K. *New J. Chem*. **2013**, *37*, 3640–3646;

(d) Gedik, M.; Brown, A. *J. Photochem. Photobiol. A: Chemistry* **2013**, *259*, 25–32;

(e) Samanta, P. K.; Pati, S. K. *Phys.Chem.Chem.Phys*. **2015**, *17*, 10053–10058.

20. Martick, M.; Scott, G. W. *Cell* **2006**, *126*, 309–320.

21. Nelson, J. A.; Uhlenbeck, O. C. *RNA* **2008**, *14*, 43–54.

22. Przybilski, R.; Hammann, C. *RNA* **2007**, *13*, 1625–1630.

23. Rovira, A. R.; Fin, A.; Tor, Y. *J. Am. Chem. Soc.* **2015,** *137*, 14602-14605.