Electronic Supplementary Information (ESI)

Site-Specific RNA Modification via Initiation of In vitro Transcription Reactions with m^6A and Isomorphic Emissive Adenosine Analogs

Deyuan Cong,[†] Kfir B. Steinbuch,[†] Ryosuke Koyama,[†] Tyler V. Lam,[†] Jaime Y. Lam,[†] and Yitzhak Tor*,[†]

[†]Department of Chemistry and Biochemistry, University of California, San Diego, La Jolla, California 92093-0358, USA, E-mail: ytor@ucsd.edu

Table of Contents

1. Materials and Methods

Reagents, buffers, and salts were purchased from Sigma-Aldrich, Fluka, TCI, Acros and Synchem, Inc. (Elk Grove, IL), 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 electrophoresis and subjected to standard desalting protocols. T7 polymerase (P266L) was expressed and purified as previously reported.[1]

Solvents were purchased from Sigma-Aldrich and Fisher Scientific and dried by standard techniques. NMR solvents were purchased from Cambridge Isotope Laboratories (Andover, MA). All reactions were monitored with analytical TLC (Merck Kieselgel 60 F_{254}). All experiments involving air and/or moisture sensitive compounds were carried out under an argon atmosphere. Column chromatography was carried out with silica gel particle size 40-63 μm. NMR spectra were obtained on a Bruker 300 MHz spectrometer. MALDI-TOF mass spectra were obtained on Bruker Autoflex Max MALDI-TOF-MS. ESI-TOF mass spectra were obtained on an Agilent 6230 HR-ESI-TOF MS at the Molecular Mass Spectrometry Facility at the UCSD Chemistry and Biochemistry Department.

1.1 Transcription reactions

^{tz}A, thA and m^{6}A (N^6-methyladenosine) modified RNA oligonucleotides were synthesized through T7 RNA polymerase transcription, T4 kinase phosphorylation and T4 DNA ligase splinted ligation.

For native transcription, each single DNA template was annealed to an 18-mer phi 2.5 T7 polymerase promoter (sequence: 5'-d-TAA TAC GAC TCA CTA TTA) in annealing buffer (10 mM Tris-HCl, 1 mM EDTA, 100 mM NaCl, pH 7.8). The 1:1 mixture (10 μM) was heated to 90 °C for 3 min and was then equilibrated at room temperature for 15 min. Transcription reactions were incubated at 37 °C for 4 h containing 1:1 annealed templated (500 nM), transcription buffer (40 mM Tris-HCl, pH 7.9), 10 mM NaCl, 2 mM

spermidine, 20 mM $MgCl_2$, 2.5 mM NTPs, 4 mM PEG-8000, 20% v/v DMSO, 10 mM DTT, RNase inhibitor (Ribolock, 1 U/µL) and RNA T7 P266L RNA polymerase (0.08 µg/µL) in a total volume of 100 µL.

For 5' modified RNA strands, **template 1** (sequence: 5'-d-GCA TAG CTG AAT TGT AAT AGT GAG TCG TAT TA) and various concentrations of thA (6.25–18.75 mM), ^{tz}A (6.25–15mM), m^6A (18.75–32 mM) were screened to get the optimal ratio for transcription, while keeping the concentration of all native NTPs constant (2.5 mM) (Figure S1–S3). Due to the limited solubility of ^{tz}A in water, concentrations above 15 mM were not screened. The relative yields of the thA and m^6A transcripts increased from 0.15 to 0.40 and 0.13 to 0.39, respectively, as the concentrations of the modified nucleosides were elevated. Ultimately, transcription reactions were supplemented with additional 15 mM ^{tz}A, 18.75 mM thA and 32 mM m^6A in DMSO and enforced the polymerase to form strands containing ^{tz}A, thA, or m^6A at their 5'-end.

To quench the transcription reactions, 1:1 v/v gel loading buffer was added. The content was denatured at 75 °C for 3 min and then each aliquot was loaded onto the 20% Urea-Polyacrylamide denaturing gel. The gel was UV-shadowed and the targeted band was excised and extracted in 0.5 M ammonium acetate solution overnight. The polyacrylamide was removed by centrifugation and further purification was done through Sep-Pak C18 column, with the desired product eluted with 50% ACN/H_2O (HPLC grade) and dried by lyophilization. The concentration of the RNA constructs was determined by UV absorption at 260 nm, using the following extinction coefficients: C = 7200, U = 9900, G = 11500, A = 15400, ^{tz}A = 2830, thA = 7106, m^6A = 14260 $L \cdot mol^{-1} \cdot cm^{-1}$. The molecular mass of each strand was confirmed by MALDI. The relative yield of ^{tz}A initiated strand, ^{tz}A initiated strand, and m^6A initiated strand was 0.4 (transcript **T3**/native), 0.4 (transcript **T4**/native), 0.39 (transcript **T5**/native).

1.2 Phosphorylation reactions

The 5' Initiated ^{tz}A, thA, and m^{6}A strands (6 µM) were 5'-phosphorylated using T4 kinase. The reaction was incubated at 37 °C for 2 h in kinase buffer (1× New England Biolabs) and included 5 mM DTT, 2 mM ATP, 0.6 U/µL T4 polynucleotide kinase (New England Biolabs). The phosphorylated product was precipitated with cold 200 proof ethanol (2.5 v/v), 100 mg/µL Glycoblue (Invitrogen) and 0.4 M ammonium acetate in dry ice bath for 1 h, followed by centrifugation (14000 cpm, 20 min). The supernatant was collected, 100 mg/µL Glycoblue was added and the mixture kept at –20 °C prior to centrifugation (14000 cpm, 20 min). The blue pellets were washed by cold 200 µL 75% ethanol and dried by gentle air flow. Autoclaved H_2O (25 µL) and 1:2 v/v gel loading buffer were added to dissolve the pellets and the content was loaded onto 20% Urea-Polyacrylamide denaturing gel. The gel was UV-shadowed and the targeted band was cut, extracted in 0.5 M ammonium acetate solution overnight. The gel was removed through centrifugation and further purification was run through Sep-Pak C18 column. The concentration of concentrated RNA was determined by UV absorption at 260 nm, as described above.

1.3 T4-Ligase mediated ligation reactions

The extended singly-modified RNA oligonucleotides strands were obtained by T4 DNA ligase-mediated ligations. The reactions were performed with a complementary 24 DNA nucleosides splint. The phosphorylated ^{tz}A, thA and m^{6}A strands (10 µM), corresponding acceptor (10 µM) and splint (10 µM) were annealed in Tris-HCl buffer (40 mM, pH 7.8) at 95 °C for 3 min and then incubated room temperature for 30 min. The reaction buffer, containing $MgCl_2$ (10 mM), DTT (10 mM), 50% PEG-4000 (0.5 mM) and T4 DNA ligase (1 U/µL, Fermentas), was added, and the reaction was incubated at 37 °C for 2 h. The same purification protocol, as described above, was employed to precipitate the impure ligation product. Autoclaved H_2O (25 µL) and 1:1 v/v gel loading buffer were added to dissolve the precipitated pellets before loading onto 20% Urea-Polyacrylamide denaturing gel. The developed gel was UV-shadowed and

the targeted band was cut out and extracted in 0.5 M ammonium acetate solution overnight. The polyacrylamide was removed via centrifugation and the ligated product was desalted on a Sep-Pak C18 column. The concentration of concentrated RNA was determined by UV absorption at 260 nm. The molecular mass of the ligation product was confirmed by MALDI and ESI-TOF MS and nucleoside composition via HPLC as described below.

1.4 S1 digestion and HPLC trace

Digestion reactions of all singly modified oligomers (1 nmol) were carried out in S1 reaction buffer (1x, Promega) with S1 nuclease (1.2 U) in 30 µL. The content was incubated at 37 °C for 2 h. Consequently, alkaline phosphatase (CIAP, 5 U) and dephosphorylation buffer (1×, Promega) were added, and solutions were incubated at 37 °C for 2 h. The ribonucleoside mixture obtained was analyzed by HPLC (Agilent 1200) using reverse phase C18 column and a water/acetonitrile mobile phase system. All stock HPLC solvents were prepared by adding 0.1% formic acid (v/v) into MilliQ water and HPLC grade acetonitrile, then filtered through Millipore nylon net 0.2 µm filters. Each injection was subjected to gradient run where conditions were dependent on the modified nucleoside (20 min of run, flow rate of 1 mL/min, 0.5–2.5% acetonitrile in water for ^{tz}A, 24 min of run, flow rate of 0.5 mL/min, 0.5–8.5% acetonitrile in water for thA and 24 min of run, flow rate of 0.5 mL/min, 0.5–12.5% acetonitrile in water for m^{6}A-modified strands). The column temperature was controlled at 25.0 ± 0.1 °C. Each run was monitored at 260 nm using a calibration reference at 650 nm and slits set at 1 nm. Nucleoside mixtures containing thA or ^{tz}A were also monitored at 341 nm or 338 nm, respectively.

1.5 MALDI-TOF Mass Spectrometry

Mass spectra were taken by MALDI-TOF mass spectrometer in negative ion mode with Bruker Daltonics V Workstation software. All the experimental data was analyzed via Bruker FlexAnalysis. THAP matrix was prepared in 1:1 water and ACN containing 20 g/L 2,4,6-trihydroxyacetophenone (THAP) and 4 g/L diammonium hydrogencitrate dibasic.

1.6 MazF catalyzed cleavage reaction

1.6.1 Gel electrophoresis

The hairpin RNA 24-mer (10 μM) was denatured/renatured in 40 mM sodium phosphate buffer (pH 7.5, 0.01% Tween-20, 1 mM EDTA), by heating to 75 °C for 3 min and slowly cooling down to room temperature. The mixture was then placed on crushed ice for 10 min and then at room temperature for 5 min. The mixture was prewarmed up to 37 °C for 5 min before adding the MazF enzyme (1 U). The reaction was kept at 37 °C. Aliquots (1 μL) were quenched by adding to 1× TBE gel loading buffer (1 μL, 7 M Urea 89 mM Tris-HCl, 89 mM Boric Acid, 2 mM EDTA, pH 8.0, 7.5% w/v Ficoll) at designated time intervals. All aliquots were kept on dry ice. After all time points were sampled, the reaction mixtures were denatured at 90 °C for 3 min and were loaded onto a 20% Urea-Polyacrylamide denaturing gel. The gel was then stained with SYBR gold solution (1× SYBR gold, 89 mM Tris-HCl, 89 mM Boric Acid, 2 mM EDTA, pH 8.0) for 30 min and analyzed on Typhoon 5 for kinetics purpose.

1.6.2 Oligomers Ladder

All oligomer ladders share the sequence with the native 24-mer RNA (**M1**). The 16-mer and 15-mer were synthesized by in vitro transcription (See **Section** 1.1). The 9-mer and 8-mer were purchased from IDT and were 5'-end phosphorylated (See **Section** 1.2). Each oligomer (0.1 nmol) was dissolved in gel loading buffer (20 μL) and the solution was denatured at 75 °C for 3 min before loading onto a 20% Urea-

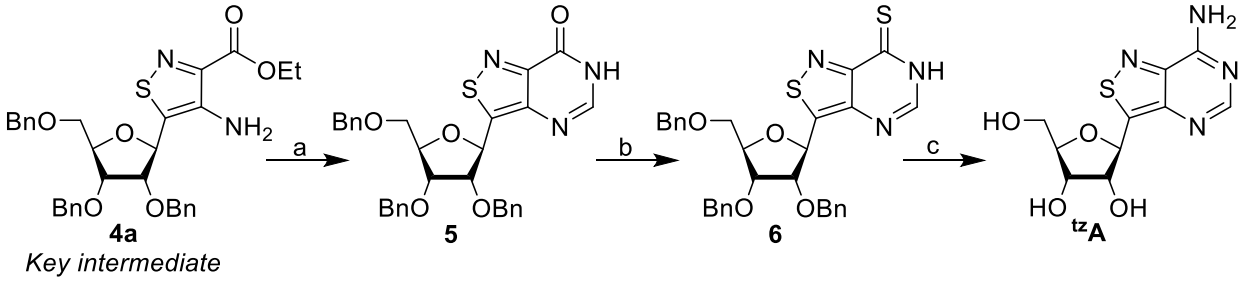
Polyacrylamide denaturing gel. The gel was visualized by staining with Stains-all solution for overnight and destained in water for analysis of the cleavage site of MazF-mediated reaction.

1.6.3 Fluorescence monitored reactions

The hairpin RNA 24-mer (10 μM) was denatured/renatured in 40 mM sodium phosphate buffer (pH 7.5, 0.01% Tween-20, 1 mM EDTA), by heating to 75 °C for 3 min and slowly cooling down to room temperature. The mixture was then placed on crushed ice for 10 min and then at room temperature for 5 min. The mixture was prewarmed up to 37 °C for 5 min before adding MazF (1 U). The reaction was kept at 37 °C. Aliquots (1 μL) were quenched by adding to 1× TBE gel loading buffer (1 μL, 7 M Urea 89 mM Tris-HCl, 89 mM Boric Acid, 2 mM EDTA, pH 8.0) at designated time intervals. All aliquots were kept to dry ice. After all time points were sampled, the reaction mixtures were denatured at 90 °C for 3 min and flash cooled on ice. Each sample was diluted to 125 μL. Steady-state fluorescence spectra (360–650 nm) were obtained upon excitation at 341 nm (excitation and emission slit widths were 5 and 10 nm, respectively).

1.7 Synthesis of ^{tz}A

Scheme S1[a]



[a]*Reagents and conditions*: (a) Formamidine acetate, Et_3N, EtOH, reflux, 70%; (b) P_2S_5, Pyridine, 90 °C, 2 h; (c) i. NH_3, MeOH, 70 °C, 16 h; ii. $HSCH_2CH_2SH$, $BF_3{\cdot}OEt_2$, DCM, rt, 72 h, 72% (for steps b and c)

The introduced reaction to form **5** from **4a** made the latter the common intermediate for the ^{tz}**G** and ^{tz}**A** synthetic pathways. Conditions for previously reported reactions b and c were slightly modified resulting in higher yields.[2] The procedure for reaction a and modifications to reactions b and c are described below.

3-((2R,3S,4S,5R)-3,4-bis(benzyloxy)-5-((benzyloxy)methyl)tetrahydrofuran-2-yl)isothiazolo[4,3-d]pyrimidin-7(6H)-one (5)

Key intermediate **4a**[2] (4.83 g, 8.40 mmol), formamidine acetate (8.31 g, 79.8 mmol) and triethylamine (3.51 mL, 25.2 mmol) were refluxed in EtOH (70 mL) overnight. The reaction mixture was then evaporated, redissolved twice more in DCM and evaporated to yield a solid residue. The crude solid was dissolved in 100 mL of DCM and washed twice with $NaHCO_3$ (satd.), water, and brine, dried over Na_2SO_4, and evaporated. The residue was then purified by column chromatography (70% EtOAc/Hexanes) to give **5** as a white solid (3.27 g, 70 %). The ^{1}H-NMR spectrum was in agreement with the previously reported data for this compound (Figure S13). ^{1}H NMR (300 MHz, $CDCl_3$) δ 11.81 (d, *J* = 3.3 Hz, 1H), 8.15 (d, *J* = 3.0 Hz, 1H), 7.44–7.22 (m, 15H), 5.86 (d, *J* = 3.6 Hz, 1H), 4.83 (d, *J* = 12.1 Hz, 1H), 4.78 (d, *J* = 12.1 Hz, 1H), 4.68 (d, *J* = 12.0 Hz, 1H), 4.59 (d, *J* = 12.0 Hz, 1H), 4.58 (d, *J* = 11.9 Hz, 1H), 4.52–4.36 (m, 2H), 4.31 (dd, *J* = 4.8, 3.6 Hz, 1H), 4.09 (dd, *J* = 6.7, 4.8 Hz, 1H), 3.84 (dd, *J* = 10.8, 2.8 Hz, 1H), 3.69 (dd, *J* = 10.8, 3.8 Hz, 1H).

Modifications to the synthetic procedure to form 6.

The thianation reaction was performed in pyridine using P_2S_5 as previously reported[2], with the following modifications: The temperature was kept at 90 °C instead of 125 °C; the 2 eq of P_2S_5 were added

in portions, 1 eq, at the beginning of the reaction and 1 eq after 1 h. The reaction was evaporated to dryness after 2 h as previously reported, TLC at 70% EtOAc/Hexanes indicated completion of the reaction.

Modifications for final benzyl ethers deprotection to form ^{tz}A

The benzyl ether deprotection reaction was performed as previously reported[2], using 1,2-ethandithiol and $BF_3{\cdot}OEt_2$. Upon completion, instead of evaporation, the reaction was Buchner filtered and the solid was washed with DCM and diethyl ether, to remove excess of $BF_3{\cdot}OEt_2$ and thiols, then dissolved in MeOH, and evaporated. The residue was then purified using column chromatography (18% MeOH/DCM) to give **^{tz}A** as a white solid (1.21 g, 72% over 3 steps). The ^{1}H-NMR spectrum was in agreement with the previously reported data for this compound (Figure S14). ^{1}H NMR (300 MHz, DMSO-*d6*) δ 8.26 (s, 1H), 8.21 (s, 1H), 8.20 (s, 1H), 5.48 (s, 1H), 5.39 (d, J = 5.3 Hz, 1H), 5.05 (s, 1H), 4.24 (t, J = 4.9 Hz, 1H), 3.95 (dt, J = 8.2, 5.0 Hz, 2H), 3.66 (dd, J = 11.9, 3.2 Hz, 1H), 3.55 (dd, J = 11.9, 3.9 Hz, 1H), 3.35 (s, 1H).

2. Supplementary Figures

2.1 Urea-PAGE Gels

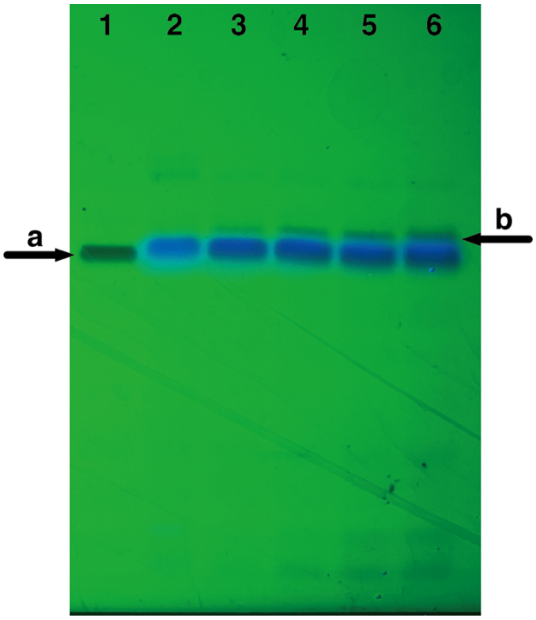


Figure S1. PAGE of transcription reactions using various concentration of thA. Lane 1: native NTPs only. Lane 2–6: transcription initiation with thA (6.25mM–18.75 mM). Black arrows indicate the expected product native transcript (arrow a) and modified transcript (arrow b). UV shadowing done upon illumination at 254 nm.

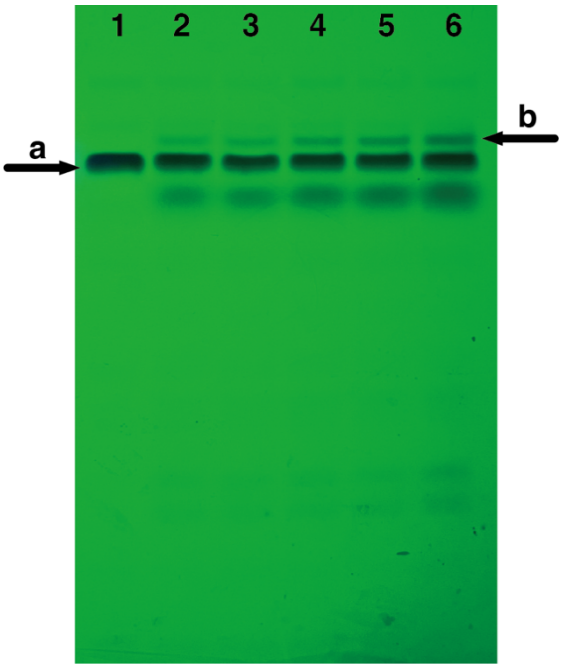


Figure S2. PAGE of transcription reactions using various concentration of m^6A. Lane 1: native NTPs only. Lane 2–6: transcription initiation with m^6A (18.75mM–32 mM). Black arrows indicate the expected product native transcript (arrow a) and modified transcript (arrow b). UV shadowing done upon illumination at 254 nm.

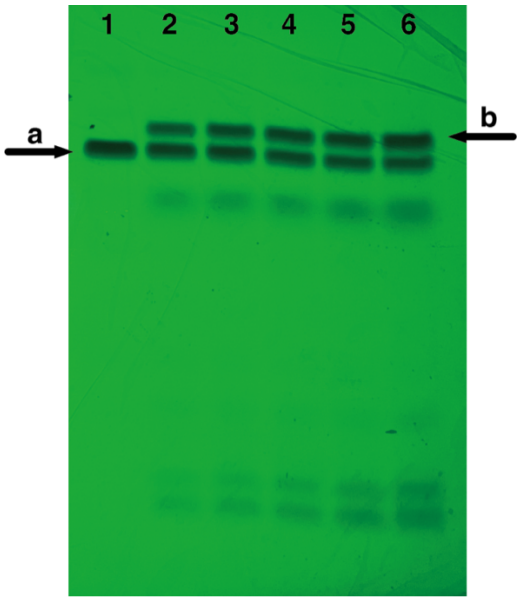


Figure S3. PAGE of transcription reactions using various concentration of ^{tz}A. Lane 1: native NTPs only. Lane 2–6: transcription initiation with ^{tz}A (6.25mM–15 mM). Black arrows indicate the expected product native transcript (arrow a) and modified transcript (arrow b). UV shadowing done upon illumination at 254 nm.

2.2 RNA characterization

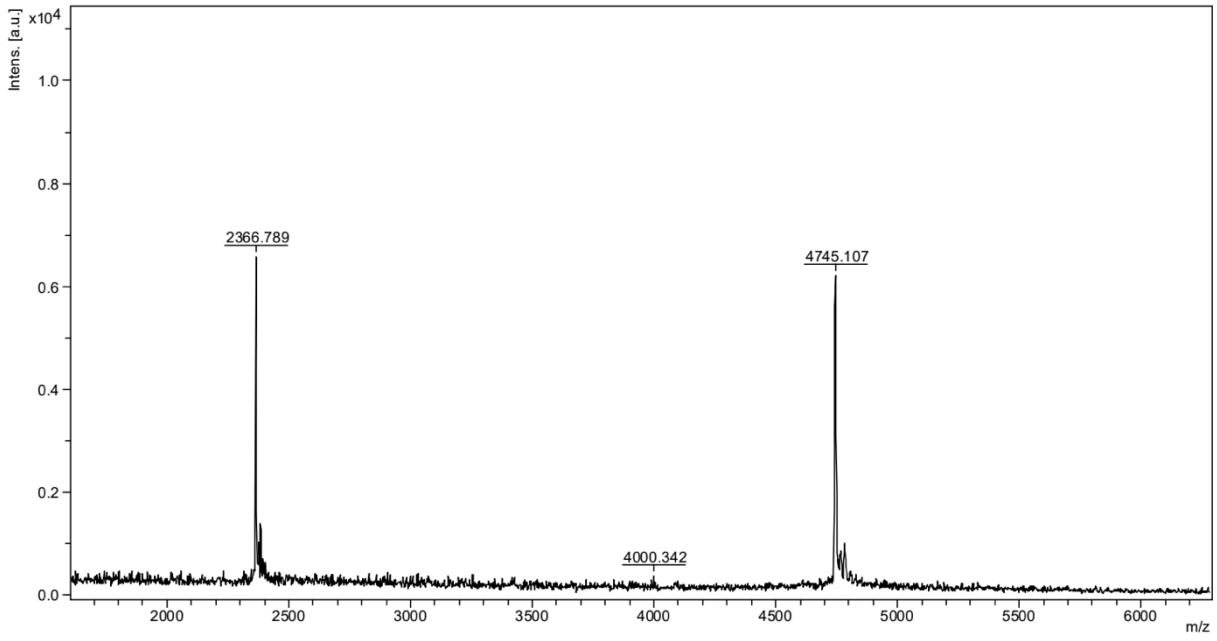


Figure S4. MALDI-TOF mass spectrum of transcript **T3** initiated with thA. The expected molar mass was 4734.87 Da and the observed molar mass was 4733.56 Da after calibration.

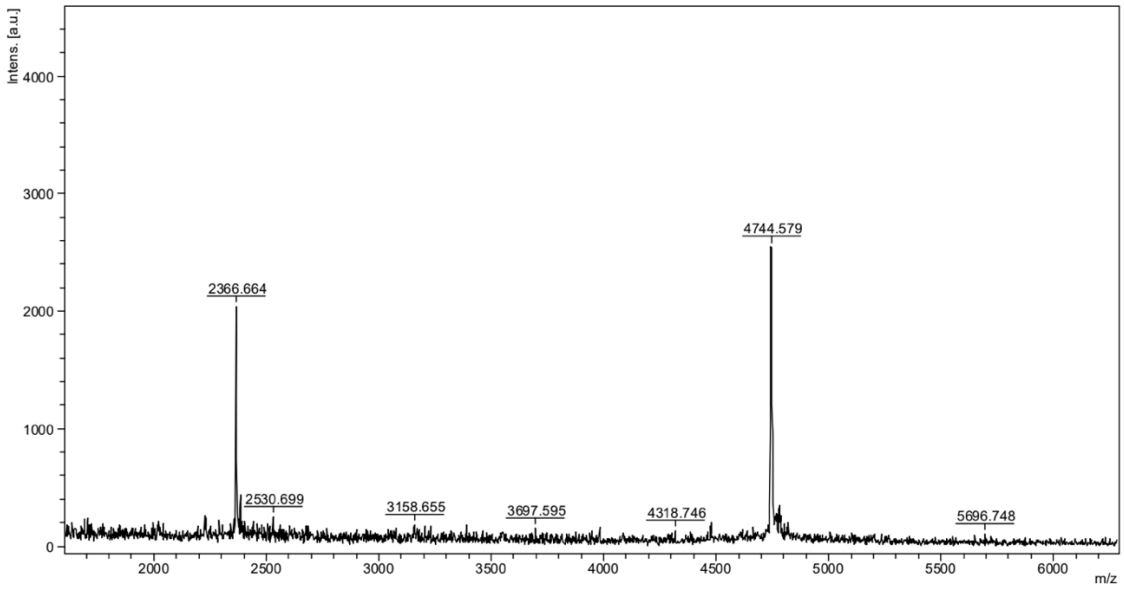


Figure S5. MALDI-TOF mass spectrum of transcript **T4** strand initiated with ^{tz}A. The expected molar mass was 4735.87 Da and the observed molar mass was 4733.32 Da after calibration.

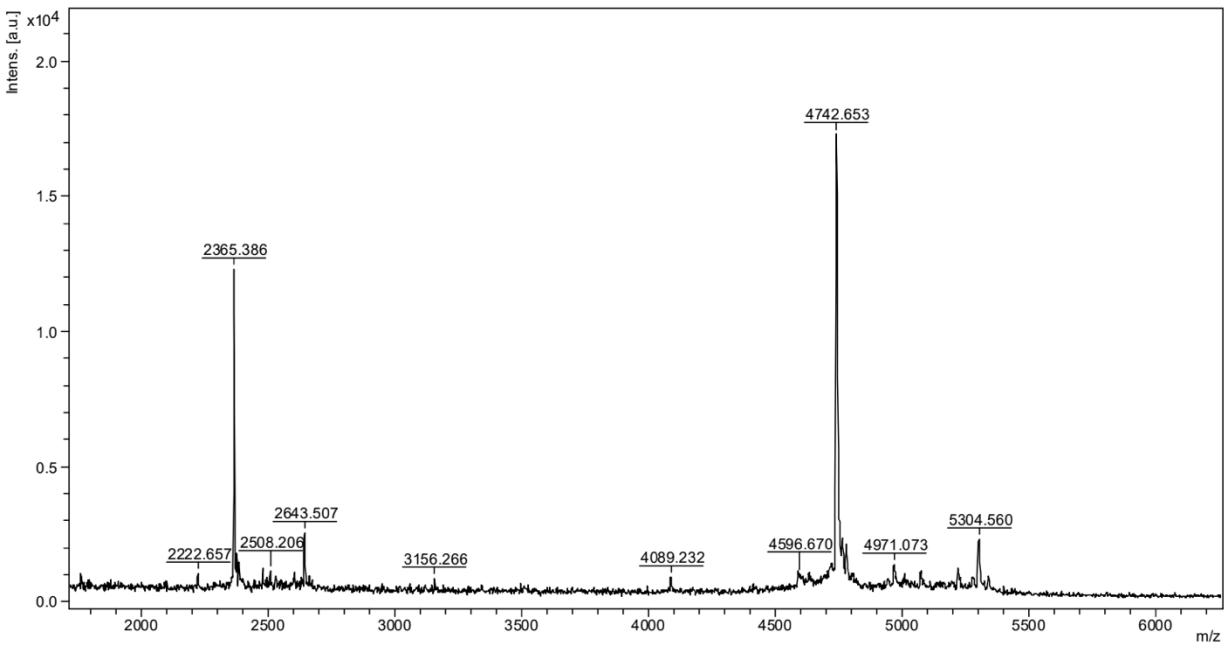


Figure S6. MALDI-TOF mass spectrum of transcript **T5** strand initiated with m^6A. The expected molar mass was 4732.87 Da and the observed molar mass was 4730.76 Da after calibration.

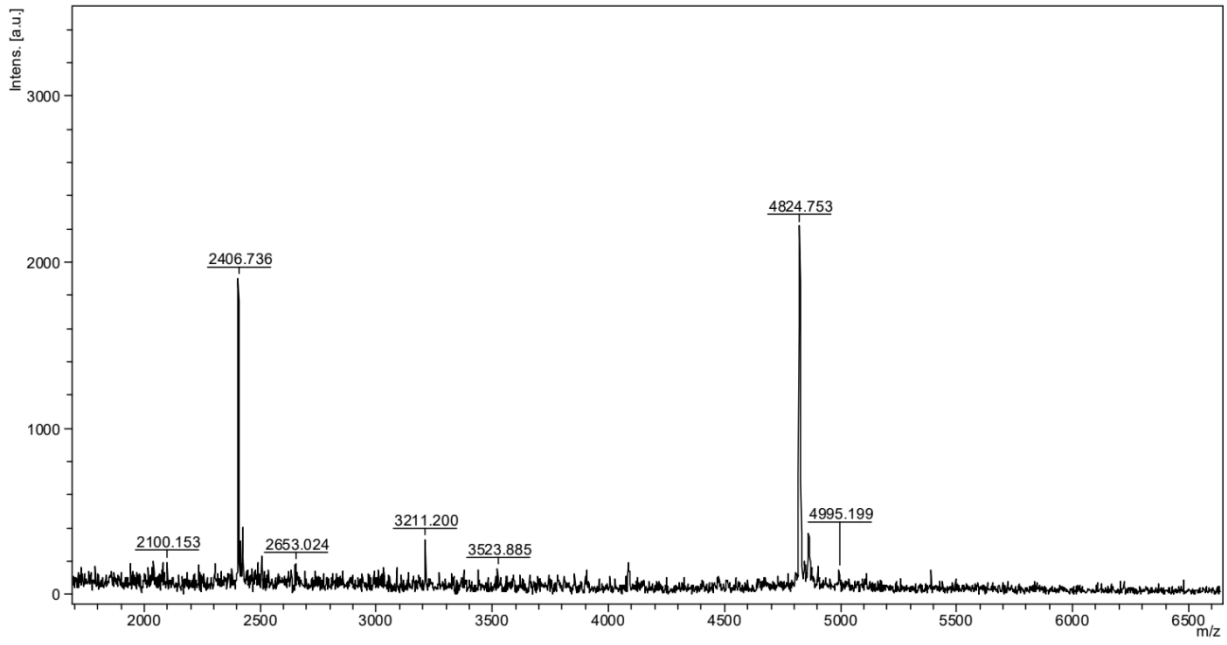


Figure S7. MALDI-TOF mass spectrum of phosphorylated transcript **T3** strand. The expected molar mass was 4815.87 Da and the observed molar mass was 4813.46 Da after calibration.

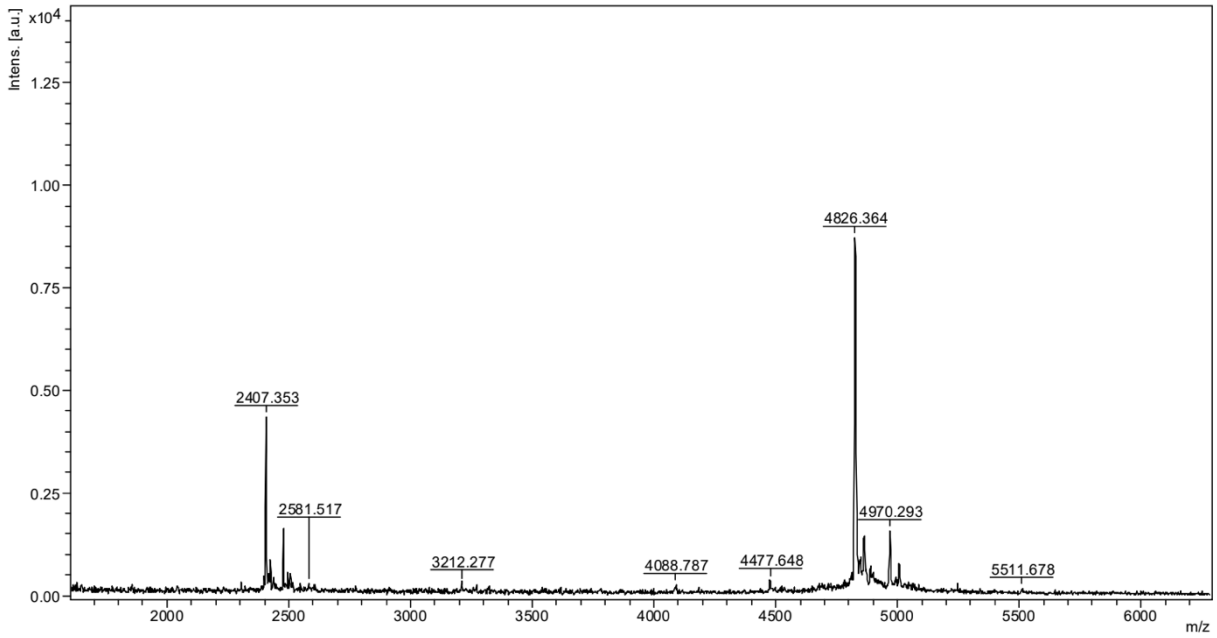


Figure S8. MALDI-TOF mass spectrum of phosphorylated transcript **T4** strand. The expected molar mass was 4816.87 Da and the observed molar mass was 4814.70 Da after calibration.

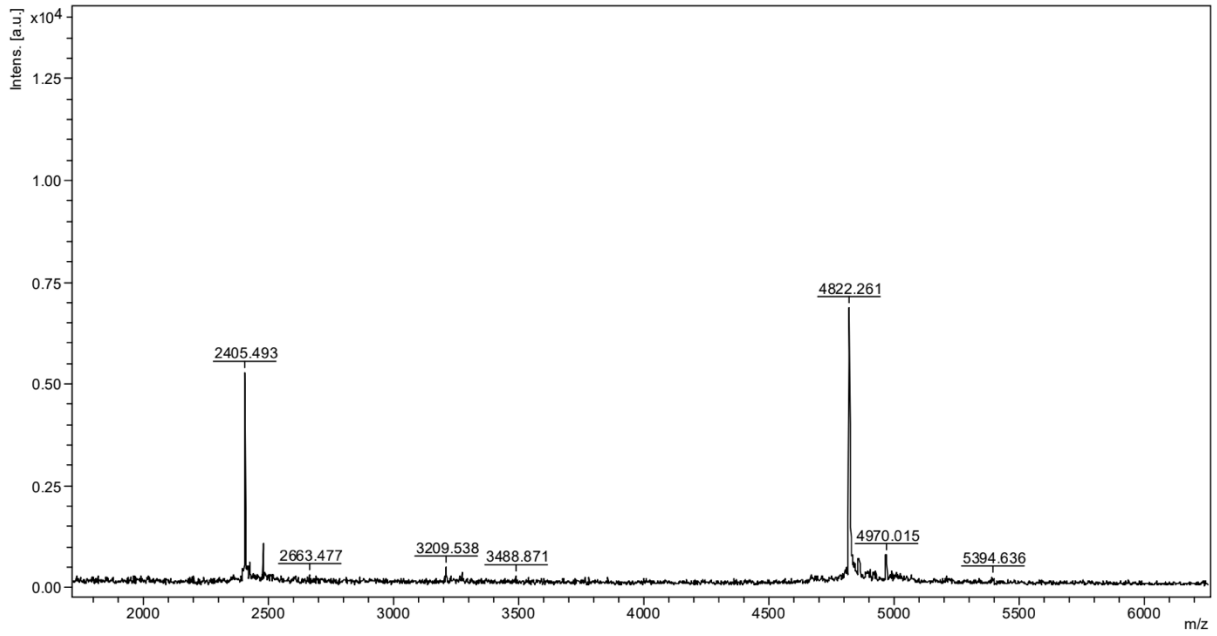


Figure S9. MALDI-TOF mass spectrum of phosphorylated transcript **T5** strand. The expected molar mass was 4813.87 Da and the observed molar mass was 4810.98 Da after calibration.

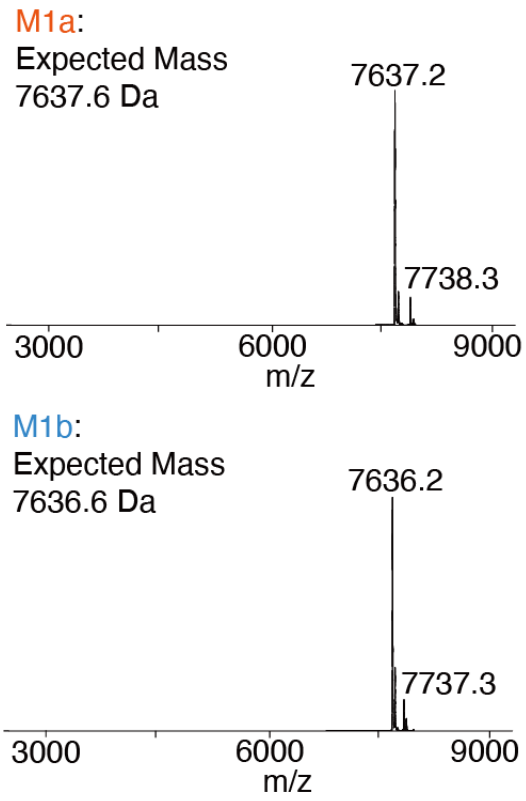


Figure S10. ESI-TOF deconvolution mass spectrum of **M1a** and **M1b** strand in negative ion mode.

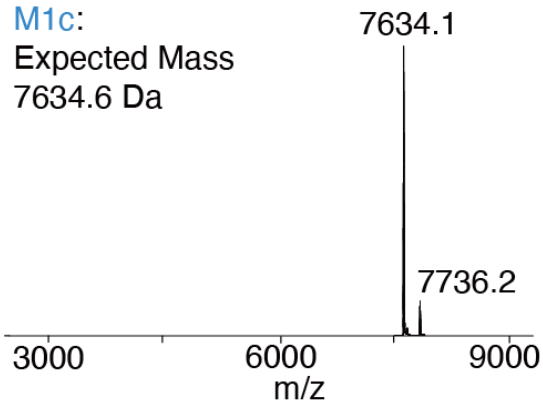


Figure S11. ESI-TOF deconvolution mass spectrum of **M1c** strand in negative ion mode.

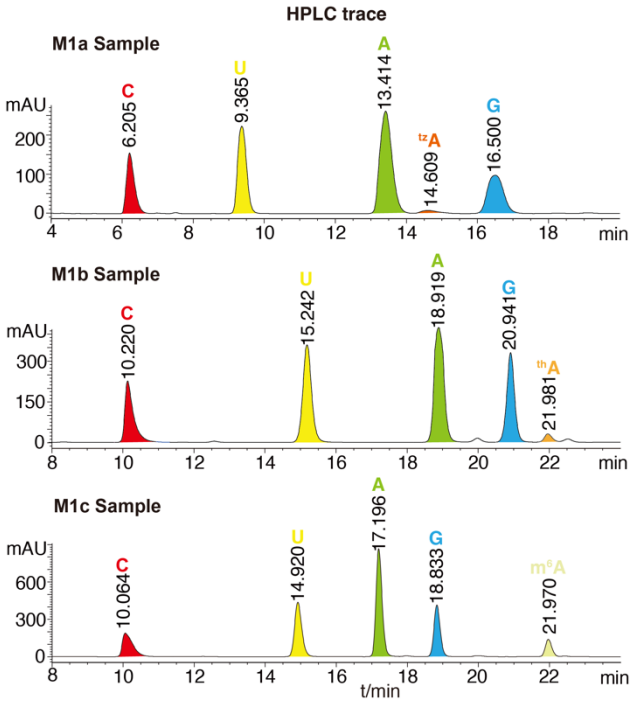


Figure S12. Characterization of strand **M1a**, **M1b** and **M1c** via digestion and HPLC analyses. All five nucleosides were separated after S1 digestion and dephosphorylation. The ratio of the concentration among five nucleosides was calculated by extinction coefficient and peak area. Note: different gradients were used for elution (See **Section** 1.4).

2.3 NMR Spectra

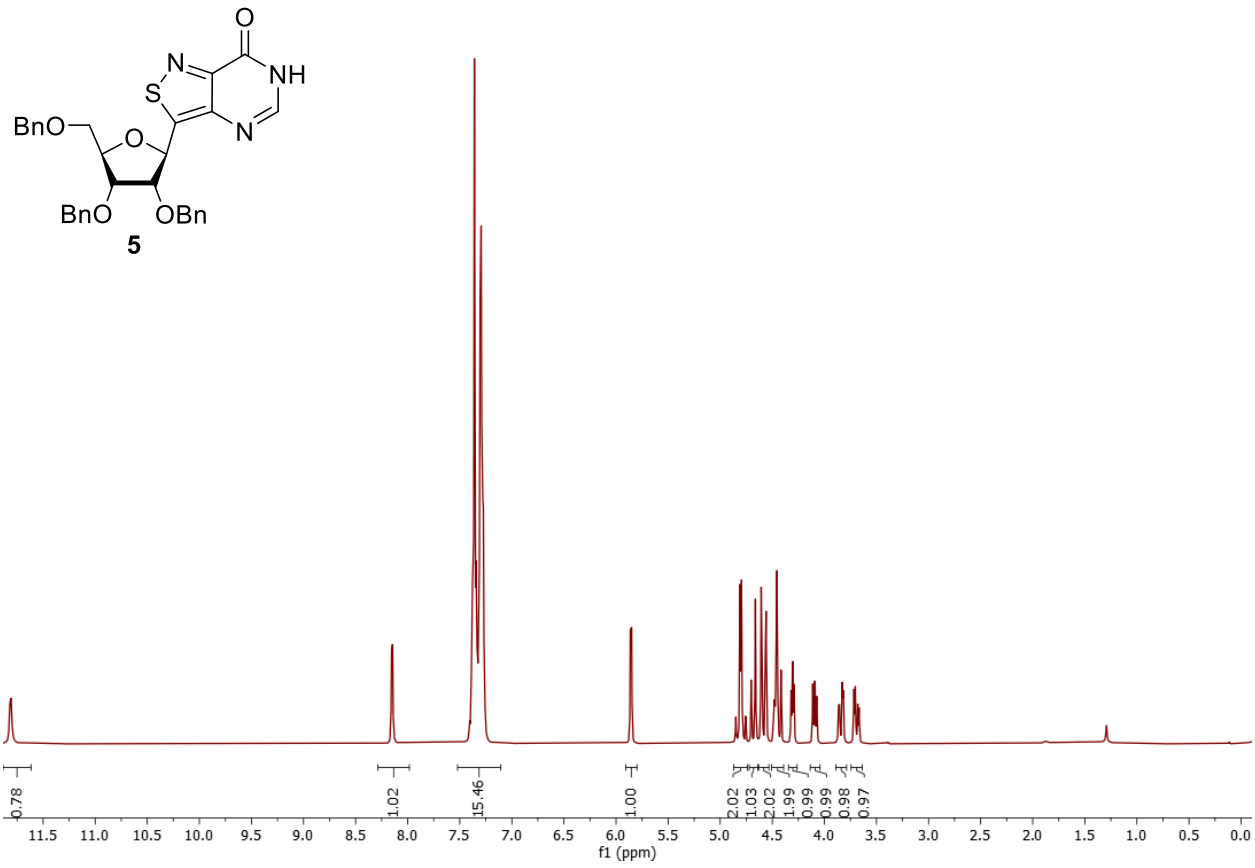


Figure S13. ^{1}H-NMR spectrum (300 MHz, $CDCl_3$) of previously reported compound **5**, synthesized according to a different procedure.

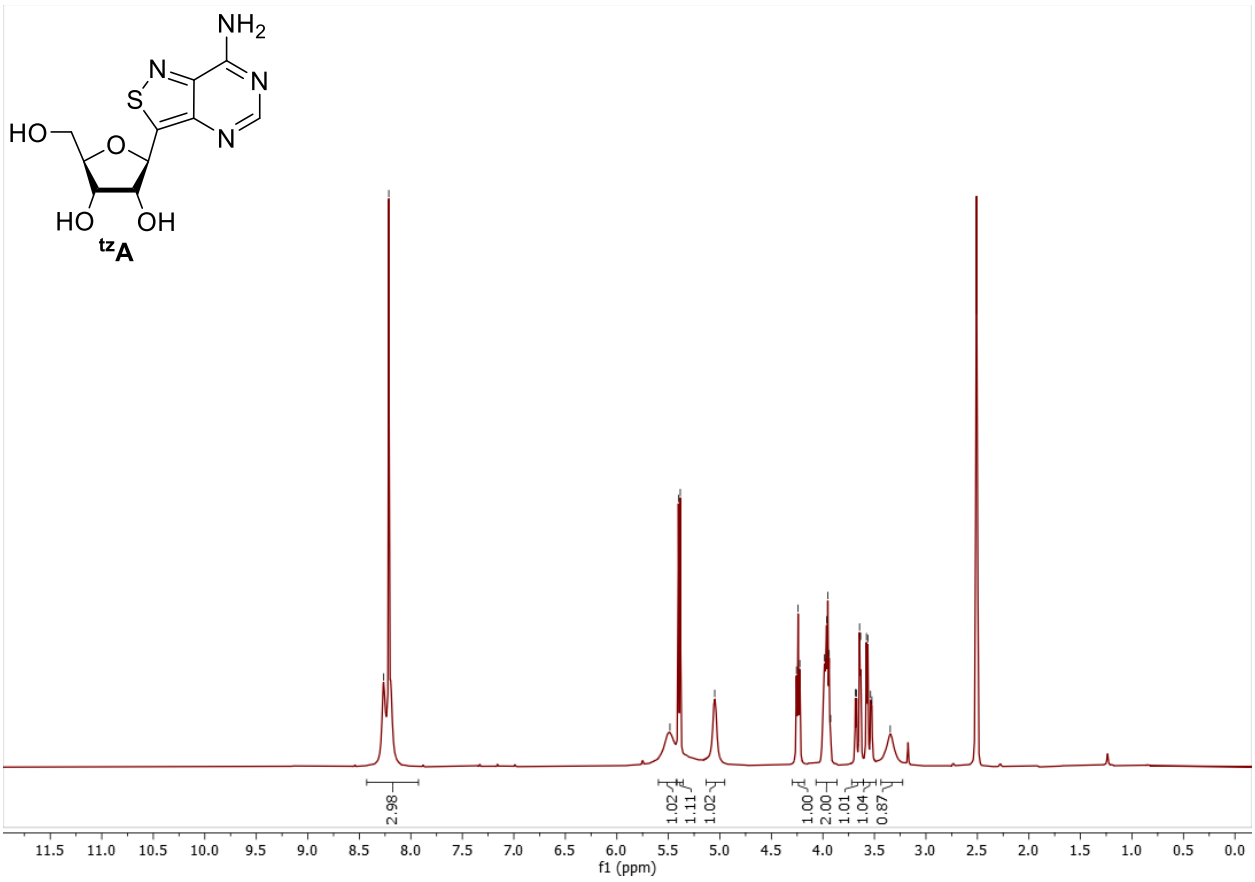


Figure S14. ^{1}H-NMR spectrum (300 MHz, DMSO-*d6*) of previously reported ^{tz}**A** synthesized according to the improved synthetic route.

References

[1] S. Lyon, V. Gopalan, *ChemBioChem* 2018, **19**, 142–146.
[2] A. R. Rovira, A. Fin, Y. Tor, *J. Am. Chem. Soc.* 2015, **137**, 14602–14605.