Artificial Evolution of Nucleic Acid Catalysts and their Use for Studying RNA

Artifizielle Evolution von katalytischen Nucleinsäuren und deren Anwendung für die Untersuchung von RNA

Doctoral thesis for a doctoral degree
at the Graduate School of Life Sciences,
Julius-Maximilians-Universität Würzburg,
Section Biomedicine

submitted by

Anam Liaqat

from Pakistan

Würzburg 2022

Submitted on: ……..

Office stamp

Members of the Thesis Committee

Chairperson: ……....

Primary Supervisor: Prof. Dr. Claudia Höbartner

Supervisor (Second): Prof. Dr. Utz Fischer

Supervisor (Third): Prof. Dr. Elmar Wolf

Date of Public Defence: ……

Date of Receipt of Certificates: ……..

Contents

Summary

RNA molecules play diverse roles in biological systems. Post-transcriptional RNA modifications and dynamic structures enhance the functional diversity of RNA. A prerequisite for studying their biological significance is the availability of reliable methods for the detection of RNA modifications and structures. Several promising approaches have been developed in the last few decades; however, efficient, and versatile tools are still required to study the dynamic features of RNA. This thesis focuses on the development of nucleic acid catalysts as a tool to address the current needs in studying RNA.

The major part of this thesis aimed at the development of deoxyribozymes as a tool for the detection of RNA modifications. Recently, m^6A-sensitive RNA cleaving deoxyribozymes were reported to interrogate the methylation status of target RNAs and efficiently validate the m^6A sites in natural RNA sequences including lncRNA and a set of C/D box snoRNAs. However, due to the lack of detailed structural and mechanistic insights into deoxyribozymes, it is currently not possible to rationally re-engineer existing DNA enzymes for the detection of other RNA modifications. Therefore, we focused on the development of novel DNAzymes for the detection of RNA modification beyond m^6A. Using in vitro selection from a random DNA library, we found deoxyribozymes that are sensitive to N^6-isopentenyladenosine (i^6A), a native tRNA modification and structural analogue of m^6A. The in vitro evolution identified three classes of DNA enzymes: AA, AB08, and AC17 DNAzymes. AA deoxyribozymes cleaved unmodified RNA and were strongly inhibited by i^6A. AB08 cleaved i^6A-modified RNA at a significantly faster rate than unmodified RNA. Contrary to that, AC17 showed unique behavior by shifting the cleavage site one nucleotide upstream in the presence of i^6A modification. Furthermore, these three classes of deoxyribozymes showed strong discrimination between structural analogues, i.e., m^6A and i^6A.

In the continuation of the project, we attempted to develop RNA-cleaving deoxyribozymes that differentially respond to monomethylated cytidine isomers, 3-methylcytidine (m^3C), N^4-methylcytidine (m^4C), and 5-methylcytidine (m^5C). To evolve potentially more versatile DNA enzymes, we used RNA substrates containing a degenerate nucleotide upstream of the target cytidine. Several deoxyribozymes were identified from in vitro selection, which are selective for

a specific methylated cytidine isomer. The characterization of AL112, AM101, AN05, and AK104 catalysts confirmed the successful evolution of modification-specific and general deoxyribozymes that showed a broad substrate scope. Furthermore, the use of AL112, AN05 and AK104 was demonstrated as a programmable tool for validation of m^3C and m^5C modifications in human mitochondrial tRNAs.

In order to accelerate the DNAzymes discovery, a high throughput sequencing method (DZ-seq) was established that directly quantifies the RNA cleavage activity and cleavage site from deep sequencing data. Illumina sequencing libraries were prepared using active DNA pools ligated to distinct RNA substrates. After the cleavage reaction, both cleaved and uncleaved fractions were extended by poly(A) tailing followed by reverse transcription using oligo-dT primer. The libraries at this stage contained information about cleavage status, cleavage site and sequence of deoxyribozymes and RNA substrate. The fraction cleaved (FC) data obtained from Dz-seq was validated for a subset of deoxyribozmes using conventional gel based kinetic assay and showed a good linear correlation (R^2 = 0.91). Interestingly, Dz-seq revealed a poor correlation between cleaving activity and abundance of deoxyribozymes in the final rounds of in vitro selections. For example, AM301 displayed very low abundance but high FC values in the DZ-seq data set. This is only one examples that shows that Dz-seq possesses a great potential for the discovery of novel deoxyribozymes for the analysis of various RNA modifications in the future.

The second objective of the current study was the development of structure-specific RNA labeling ribozymes. This was inspired by the recently reported FH14 and FJ1 ribozymes, which are sequence-specific RNA labeling ribozymes that recognize their target RNA by Watson-Crick base pairing. Here, we attempted to develop ribozymes that targets RNA of interest by structure-specific interaction rather than base-pairing and focused on a specific RNA G-quadruplex as the target. The in vitro selection used an RNA pool covalently attached to a fragment of the 5'-UTR of BCL2 RNA, which has been described to fold into a G-quadruplex. Two subsequent selection experiments led to the identification of the adenylyltransferase ribozymes AO10.2 and AR9. The partial characterization of these catalysts showed that A010.2 was unable to recognize intact BCL2 structure, but it turned out as the first reported trans-active ribozyme that efficiently labeled uridine in a defined substrate RNA hybridized to the ribozyme. Thus, it could potentially be used

as a site-specific labeling tool for target RNA. The other ribozyme AR9 was shown to serve as a trans-active, self-labeling ribozyme that catalyzed adenylyl transferase reaction in the presence of the intact BCL2 sequence. Based on these preliminary findings, we envision that AR9 could potentially serve as a reporter RNA by self-labeling in the presence of an RNA G-quadruplex. However, both AO10.2 and AR9 still require more detailed characterization for their potential applications.

Zusammenfassung

RNA hat zahlreiche Funktionen in verschiedensten biologischen Systemen. Sowohl post-transkriptionelle Modifikationen als auch die Dynamik der dreidimensionalen Struktur von RNA trägt zu deren funktionalen Diversität bei. Eine Voraussetzung, um die biologische Bedeutung von RNA genauer zu untersuchen, ist die Verfügbarkeit zuverlässiger Methoden zur Detektion von RNA-Modifikationen und -Strukturen. In den letzten Jahrzenten wurden hierfür zahlreiche vielversprechende Ansätze entwickelt und berichtet. Allerdings besteht weiterhin der Bedarf an effizienten und vielseitig einsetzbaren Hilfsmitteln, um die Dynamik von RNA weiter zu erforschen. Diese Arbeit konzentriert sich auf die Entwicklung von Nucleinsäure basierten Katalysatoren, die in Zukunft als Werkzeug zur Untersuchung von RNA eingesetzt werden können.

Der Großteil dieser Arbeit strebte die Entwicklung von Desoxyribozymen als Werkzeug für die Detektion von RNA-Modifikation an. Vor kurzem wurden m^6A-sensitive DNA-Enzyme berichtet, die RNA schneiden können und damit Auskunft über deren Methylierungs-Status geben können. Diese sind auch in der Lage m^6A in natürlichen RNAs wie lncRNAs und C/D box snoRNAs zu detektieren. Allerdings fehlen detaillierten strukturelle und mechanistische Erkenntnissen darüber, wie Desoxyribozyme solche Modifikationen detektieren. Deshalb ist es noch nicht möglich bereits vorhandene DNA-Enzyme umzuarbeiten, damit diese auch andere RNA-Modifikationen erkennen können. Aus diesem Grund fokussierten wir uns hier auf die Entwicklung neuer DNA-Enzyme für die Detektion von RNA-Modifikationen über m^6A hinaus. Mit Hilfe von in vitro Selektion konnten wir ausgehend von einer randomisierten DNA-Bibliothek, Desoxyribozyme finden, die sensitiv gegenüber N^6-Isopentenyladenosin (i^6A) sind. Bei dieser Modifikation handelt es sich um ein strukturelles Analogon von m^6A, die natürlicherweise in tRNA vorkommt. Als Ergebnis der in vitro Selektion konnten drei Klassen an DNA-Enzymen identifiziert werden: AA, AB08 und AC17 Desoxyribozyme. AA DNA-Enzyme spalteten unmodifizierte RNA und wurden durch i^6A stark inhibiert. AB08 schnitten i^6A-modifizierte RNA signifikant schneller als unmodifizierte RNA. Im Gegensatz hierzu zeigte AC17 ein einzigartiges Verhalten, indem es die Schneide-Position innerhalb der RNA um ein Nukleotid Richtung 5‘-Ende verschob, wenn eine i^6A-Modifikation vorhanden war. Des weiteren konnten alle drei Klassen an DNA-Enzymen eindeutig zwischen m^6A und i^6A unterscheiden.

Im weiteren Verlauf des Projektes strebten wir an RNA-schneidente Desoxyribozyme zu entwickeln, die die mono-methylierten Cytidin-Isomere 3-Methylcytidin (m^3C), N^4-Methylcytidin (m^4C) und 5-Methylcytidine (m^5C) voneinander unterscheiden können. Um vielseitigere DNA-Enzyme zu erhalten, benutzten wir RNA-Substrate, die ein randomisiertes Nukleotid in 5'-Richtung neben dem methylierten Cytidin besaßen. Mehrere Desoxyribozyme konnten identifiziert werden, die selektiv und spezifisch für eines der methylierten Cytidin Isomere waren. Die Charakterisierung der DNA-Enzyme AL112, AM101, AN05 und AK104 bestätigte die erfolgreiche Evolution von einerseits modifikations-spezifischen sowie aber auch generellen DNA-Enzymen, die einen großen Substrat-Bereich abdecken. Zudem konnte gezeigt werden, dass AL112, AN05 und AK104 als programmierbare Werkzeuge zur Bestätigung von m^3C- und m^5C-Modifikationen in menschlicher mitochondrialen tRNA eingesetzt werden können.

Um die Entdeckung von DNA-Enzymen weiter zu beschleunigen, wurde eine Hochdurchsatz-Sequenzierungsmethode (DZ-seq) entwickelt. Diese nutzt die Sequenzierungsdaten, um direkt die Schneideaktivität einzelner Desoxyribozyme zu quantifizieren sowie die genaue Stelle der RNA-Spaltung festzustellen. Illumina Sequenzierungsbibliotheken wurden ausgehend von aktiven DNA-Pools hergestellt, welche an bestimmte RNA-Substrate ligiert wurden. Nachdem die Schneide-Reaktion stattgefunden hatte, wurden sowohl die geschnittenen als auch die ungeschnittenen Fraktionen mit Hilfe von Poly(A)-Polymerase verlängert, woraufhin eine reverse Transkription mit Oligo-dT Primern folgte. Zu diesem Zeitpunkt beinhalteten die Bibliotheken bereits Informationen über den Schneide-Status, die exakte Schnittstelle innerhalb der RNA sowie über die Sequenzen des entsprechenden DNA-Enzyms und der Substrat-RNA. Die Daten, die durch Dz-Seq über die Schneideaktivität der einzelnen Desoxyribozyme erhalten wurden, wurde für einen Teil der Enzyme anhand konventioneller, gel-basierter kinetischer Assays validiert. Diese zeigten eine gute lineare Korrelation ($R^2 = 0.91$). Interessanterweise zeigte Dz-Seq aber nur eine schwache Korrelation zwischen der Schneideaktivität und der Häufigkeit der Desoxyribozyme in der letzten Runde der in vitro Selektion. Zum Beispiel war AM301 nur zu einem geringen Anteil im DZ-seq Datensatz zu finden, war jedoch sehr aktiv. Dies ist nur ein Beispiel des großen Potentials von DZ-seq für die Entdeckung neuer DNA-Enzyme, die für die zukünftige Analyse zahlreicher RNA-Modifikationen angewendet werden könnten.

Der zweite Teil dieser Arbeit beschäftigte sich mit der Entwicklung von Ribozymen, die spezifisch eine Ziel-RNA anhängig von deren Struktur markieren können. Die Inspiration hierfür stammt von den kürzlich berichteten Ribozymen FH14 und FJ1. Hierbei handelt es sich um Ribozyme, die über Watson-Crick-Basenpaarung ihre Ziel-RNA erkennen und diese dann sequenz-spezifisch markieren. Unser Ziel war es nun Ribozyme zu entwickeln, die ihre Ziel-RNA über Struktur-spezifische Wechselwirkungen anstelle von Basenpaarung erkennen. Hierbei fokussierten wir uns auf einen RNA G-Quadruplex als entscheidendes Strukturelement. Bei der in vitro Selektion wurde eine RNA Bibliothek verwendet, die kovalent an ein Fragment der 5‘-UTR der BCL2 RNA gebunden war. Von dieser RNA ist bekannt, dass sie einen G-Quadruplex formt. Zwei aufeinanderfolgende Selektionen führten zur Identifikation der Adenylyltransferasen AO10.2 und AR9. Die vorläufige Charakterisierung dieser beiden Ribozyme zeigte, dass AO10.2 die intakte BCL2-Struktur nicht erkennen kann. Stattdessen stellte sich heraus, dass dies das erste trans-aktive Ribozym ist, das effizient Uridin markieren kann, welches sich in einer definierten RNA-Struktur befindet, die mit dem Ribozym hybridisiert ist. Demnach hat es großes Potential als Werkzeug für die spezifische Markierung von RNA eingesetzt werden zu können. Beim zweiten Ribozym AR9 stellte sich heraus, dass es sich um ein trans-aktives, selbst-markierendes RNA-Enzym handelt, welches die gewünschte Reaktion nur bei Vorhandensein der intakten BCL2-Sequenz katalysiert. Basierend auf diesen vorläufigen Ergebnissen könnte AR9 als Reporter-RNA dienen, die sich RNA G-Quadruplexes selbst markiert. Allerdings benötigen sowohl AO10.2 als auch AR9 noch eine detailliertere Charakterisierung, bevor sie für potenzielle Anwendungen eingesetzt werden können.

1. Introduction

1.1 Post-transcriptional RNA modifications

The “Central dogma of life” explains the process by which genetic information is transformed from DNA to protein through RNA. In addition to reversible chemical modifications of DNA and histone proteins, post-transcriptional RNA modifications also play regulatory roles in gene expression as an integral part of the central dogma. The spectrum of RNA modifications is much broader than in DNA. More than 170 distinct modifications have been identified in RNAs, compared to a limited number of epigenetic modifications in DNA[1]. This is most likely because RNA performs diverse functions beyond its classical role of transferring genetic information. The chemical modifications are known to occur in all classes of RNAs (tRNAs, mRNA, rRNAs, and ncRNAs) in all domains of life, but the most extensive array of modifications has been identified in tRNA and rRNA. The abundance of RNA modifications deciphers their vital role in biological systems such as structural components in ribosomal RNA (rRNAs), a carrier of genetic information in messenger RNA (mRNAs), translator in transfer RNA (tRNAs) as well as regulators in small interfering RNAs (siRNAs)[2]. To perform these diverse functions, RNA expands the diversity of its primary structure by introducing chemical modifications that directly influence at the site of modification or globally affect the structure of RNA. Despite hundreds of reported RNA modifications, few are well studied in terms of their biological roles. The research field of post-translational RNA modification is expanding continuously, and new modifications are being added to the repository. The following sections highlight only few examples of the most prevalent modifications and their functions in three main types of RNA (tRNA, mRNA, and rRNA).

1.1.1 tRNA modifications

Simple RNA modifications such as methylation (N1-methyladenosine), acetylation (N4-acetylcytosine), or isomerization (pseudouridine) are the most abundant and identified in all tRNAs. The hypermodified bases are less abundant presumably because of their multi-step biosynthesis and bulkier chemical groups containing heterocyclic structure, monosaccharides, and amino acid residues[3]. The biological functions of some of these modifications are well studied, for example, modifications at position 34 (wobble base) or position 37 (a nucleotide

downstream of the anticodon) directly affect the decoding[4]. Queuosine (Q) is a hypermodified guanosine analog (7-deaza-guanine core attached to an amino-methyl side chain and cyclo-pentanediol moiety) that occurs at the wobble position, which plays a key role in the regulation of translation through codon-anticodon interaction. Additionally, this modification protects cognate tRNAs from ribonucleolytic cleavage, thus altering the tRNA-derived small RNAs pool in the cell[5]l. Likewise, 3-methylcytidine (m^3C) modification is present at position 32 in tRNA and has a potential role in fine-tuning the translation[6].

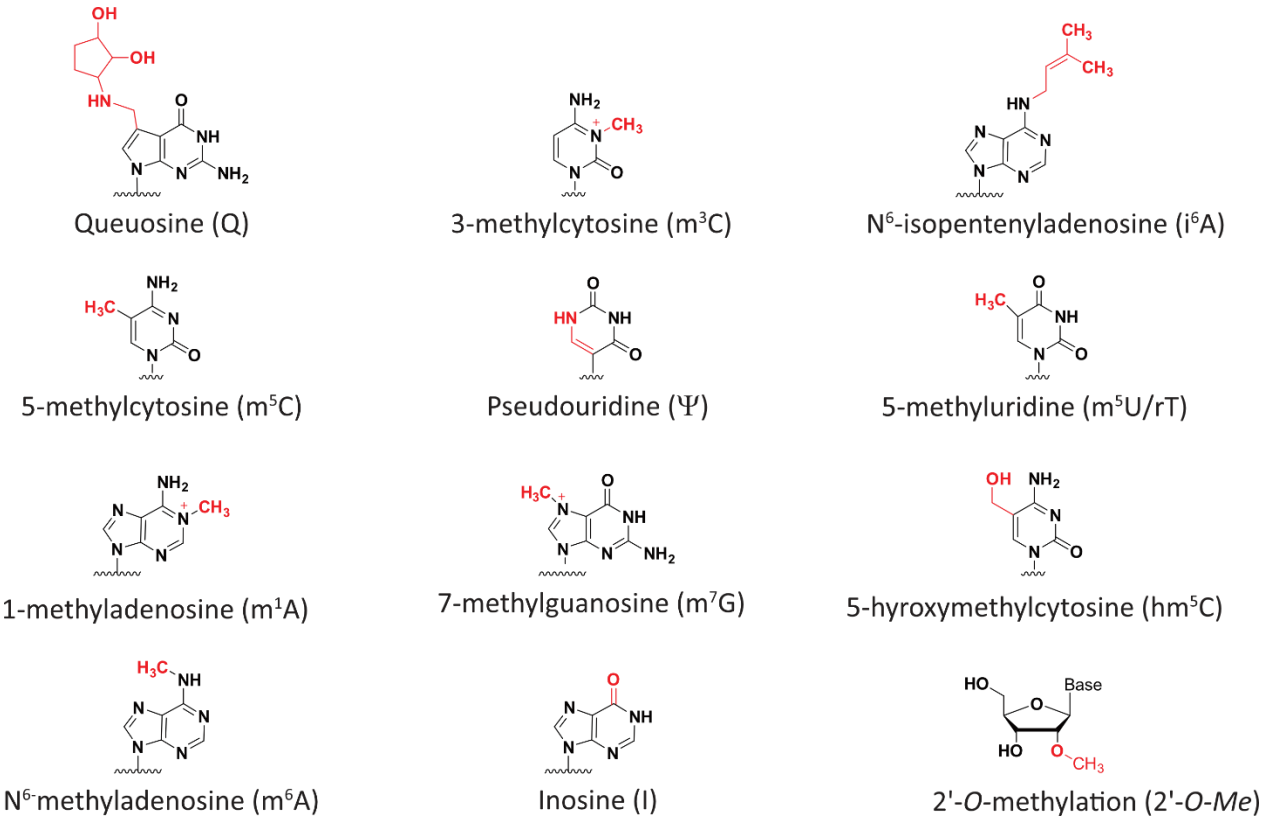


Figure 1 Chemical structure of few selected examples of abundant modifications in RNA. Sites of modifications in four canonical nucleotides is highlighted in red.

Position 37 is mostly occupied by modified purine bases such as N^6-threonylcarbamoyladenosine (t^6A), N^6-isopentenyladenosine (i^6A), 2-methylthio-N^6-isopentenyladenosine (ms_2i^6A) and the structurally complex wybutosine (yW). Generally, these modifications reduce the conformational flexibility of the anticodon loop by stabilizing the cognate codon-anticodon interaction and therefore optimizing the translation efficiency[7]. Additionally, bulkier side chains hindered the base-pairing between position 32 or 33 with position 38 resulting in an open anticodon loop structure[8]. The tRNA elbow (a region formed by the interaction between D and T-loop) and variable loop regions are known to have RNA modifications such as 5-methyluridine (m^5U), pseudouridine-isomerization of uridine (Ψ), 5-methylcytosine (m^5C), 1-methyladenosine (m^1A) and 7-methylguanosine (m^7G) that enforce the correct folding of tRNA and stabilize the characteristic L-shape three-dimensional structure[3, 9].

1.1.2 mRNA modifications

In addition to tRNA modifications, there has been a recent shift of attention in the research community towards post-transcriptional modifications in the coding and untranslated regions (UTRs) of mRNAs. In eukaryotes, mRNA 5'cap structure predominately has 7-methylguanosine (m^7G) linked through a 5'-5' triphosphate bridge to the first transcribed nucleotide which is methylated on ribose sugar at the 2'-hydroxyl group (Nm) followed by second methylated nucleotide (Nm). The first nucleotide was observed to have an additional methyl group and usually N^6,2'-O-dimethyladenosine (m^6Am)[10]. These modifications affect different cellular processes such as mRNA stability, export to the cytoplasm, mRNA splicing, and translation efficiency which have a profound impact on cellular transcriptome and proteome[11]. N^6-methyladenosine (m^6A) was identified in the coding region of mRNA as the most abundant, dynamic, and reversible modification, which is installed by methylation complex METTL3, METLL14, and WTAP proteins or by the METTL16 methyltransferase and erased by demethylases such as FTO and ALKBH5[12]. m^6A is known to play an important role in modulating mRNA folding, maturation, nuclear processing, translation, and degradation[13].

Unlike m^6A, 1-methyladenosine (m^1A) occurs on Watson-Crick interface and possess a positive charge which influence protein-RNA interaction and RNA secondary structure[14]. m^1A has been found near translation start site and splice site and was suggested to be involved in upregulation

of the translation process[15]. Additional methylation marks under investigation in mRNA are 5-methylcytosine (m^5C) and its oxidized derivative hydroxymethylcytosine (hm^5C), where m^5C plays a role in nuclear export of m^5C-containing mRNAs but the function of hm^5C is currently elusive[16]. Adenosine to inosine RNA editing is a commonly observed mechanism to modify RNA post-transcriptionally catalyzed by adenosine deaminase acting on RNA (ADAR) family of enzymes that bind to adenosine in double-stranded RNA (dsRNA)[17]. Inosine is recognized as guanosine by translational and splicing machinery, which leads to re-coding events during translation[18].

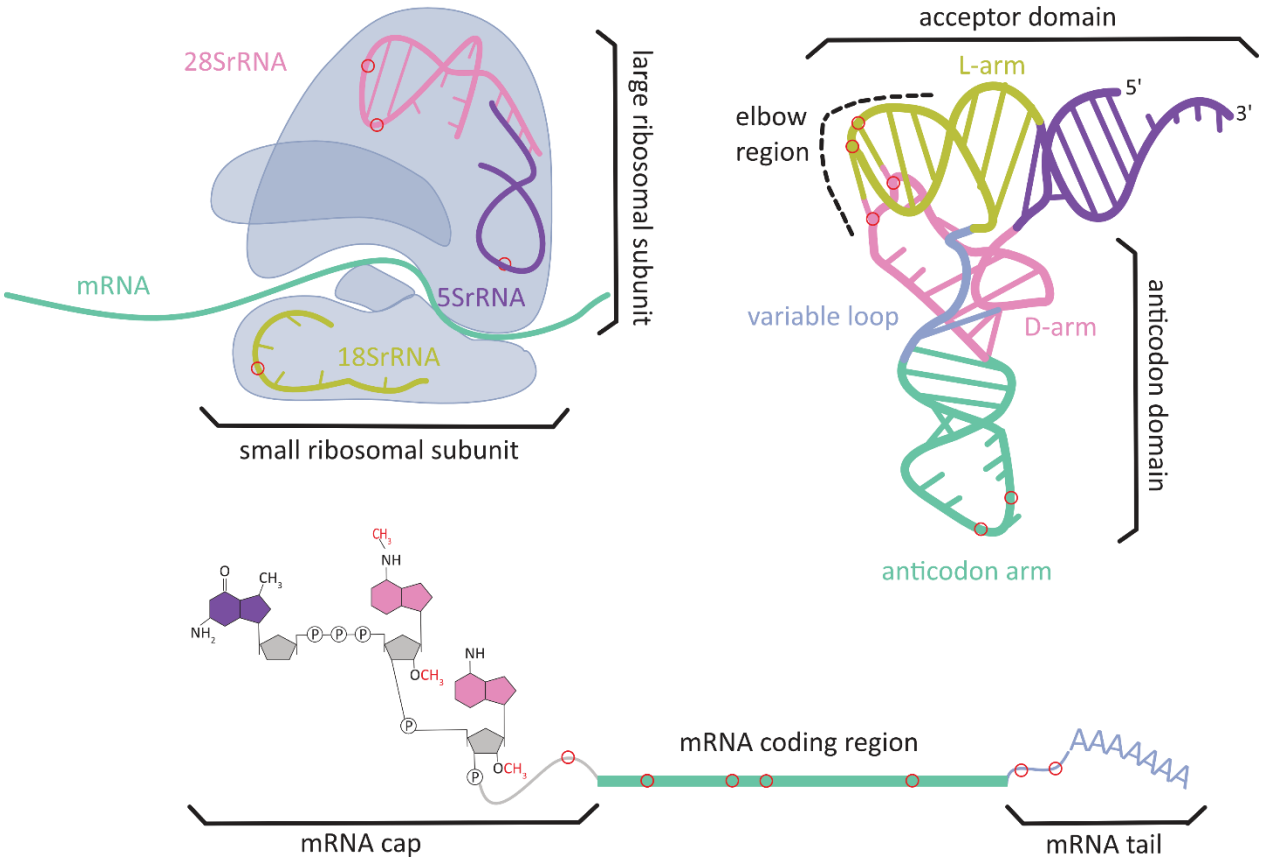


Figure 2 Three main types of RNA; rRNA, tRNA and mRNA. Modified positions highlighted in red circles.

1.1.3 rRNA modifications

Ribosomal RNA (rRNA) is the most abundant and functionally important RNA which is extensively modified during transcription and subsequent maturation processes. rRNA modifications are frequently clustered at functionally important, highly conserved and specific areas of ribosomes including peptidyl transferase center, tRNA-mRNA binding site, the subunit-subunit interaction sites and polypeptide exit tunnel[19]. The most abundant rRNA modifications are methylation of 2'-hydroxyl group of ribose, methylation of nucleobase and pseudouridine. These modifications collectively affect the hydrogen bonding potential, induce base stacking effect and increase structural rigidity[20].Methylation of 2'-OH prevents hydrolysis of phosphodiester bond, favors the 3' endo configuration, alters the hydrogen bonding pattern, and thus has a profound impact on RNA secondary structure[21]. m^5C in rRNA most likely stabilizes RNA folding by promoting base stacking and increasing the thermal stability of hydrogen bonding with guanine[22]. Similarly, Pseudouridine (Ψ) provides rigidity to the phosphodiester backbone of the RNA, alters base-pairing interactions and thermal stability that affects RNA conformation[23]. In addition to their functions in modulating RNA structure, these modifications play a crucial role in amino acid incorporation, and stop codon readthrough to regulate the translation efficiency[24]. The rRNA modifications also play a role in communication between distant regions, either by altering rRNA folding or by interacting with ribosomal proteins[25].

1.2 Existing methods for detection of RNA modifications

RNA modifications that are typically of low abundance with respect to the overall number of nucleotides may go unnoticed with standard detection methods of molecular biology, and in this case their physiological roles remain unexplored. The inadequacy of sensitive and reliable methods for analyzing RNA modifications is one of the major limitations in the field. In this section, various methods for the detection of nucleotide modifications are discussed based on their physicochemical properties, chemical reactivity, and the influence of modifications on the activity of protein enzymes.

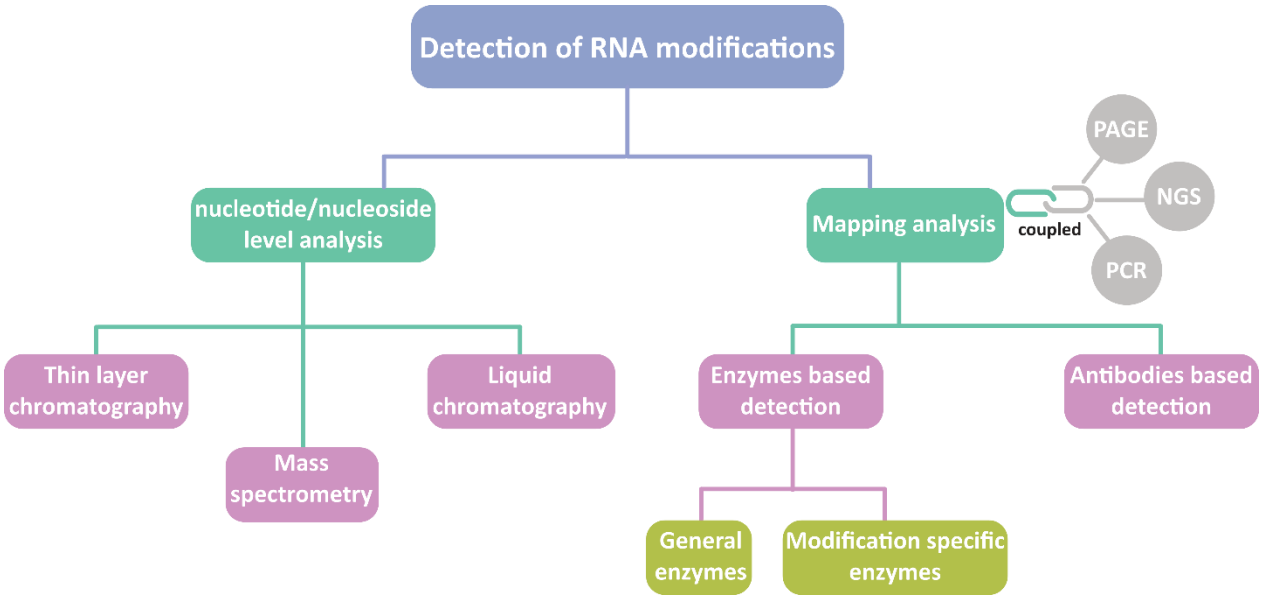


Figure 3 Flow chart of methods for detection of RNA modifications. The approaches are mainly categorized in two groups: overall detection of RNA modifications on nucleotide or nucleoside level, detection of RNA modifications at specific locations.

1.2.1 Nucleotide/Nucleoside level analysis

In general, the small modifications such as methylation, do not alter RNA's physicochemical properties enough to distinguish it from its unmodified counterpart. Therefore, RNA samples were digested to nucleotides or nucleosides and then analyzed by various methods including thin layer chromatography, capillary electrophoresis, and mass spectrometry.

Thin layer chromatography (TLC) is a low-tech method that identifies RNA modifications based on polarity, hydrophobicity, and net charge of nucleotides[26]. The long RNA chains are digested to mononucleotides with selected nucleases (RNase T2, RNase P1), and the modified nucleotides can then be determined by comparing their mobility with that of known (synthetic) standards followed by detection by UV spectrophotometry or autoradiography. Radioactive labeling of nucleotides with ^{32}P is more sensitive compared to UV-based detection approach and require less amount of RNA for analysis[27]. Two-dimensional (2D) TLC improves the separation capabilities and almost 100 chemically distinct RNA modifications have been identified and characterized using three standard solvent system[28].

Column liquid chromatography (LC) is a well-established technique for the analysis of nucleotides and nucleosides where RNA enzymatically digested to mononucleotides or nucleosides prior to separation by ion exchange or reverse phase column[29]. The eluted species were identified through UV detector for determination of base composition of RNA. Reverse phase liquid chromatography (RPLC) combined with UV detection is an advanced method that has capability to resolve approximately 65 nucleosides in a single analysis based on retention time and characteristic UV spectrum[30]. Similarly, high-performance liquid chromatography (HPLC) is a method with high sensitivity and specificity for detection of modified nucleotides of digested RNA. The occurrence of RNA modifications within a RNA sample of interest can be determined by comparison with the chromatogram of standard mixtures of nucleosides[31].

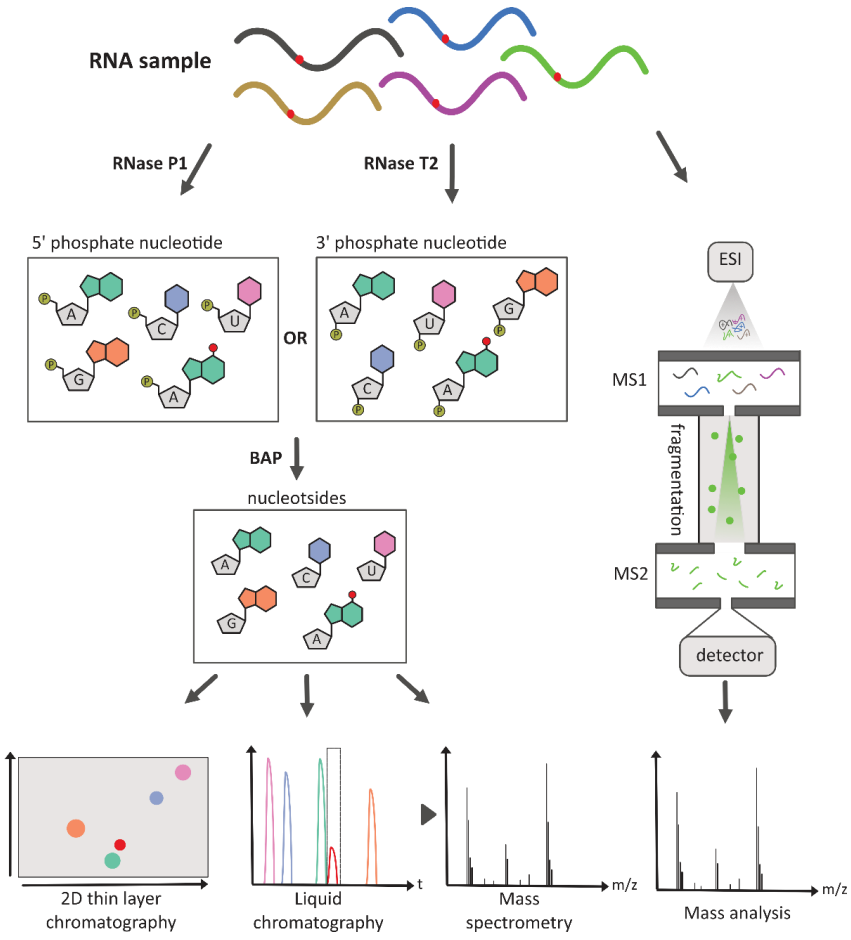


Figure 4 Detection of RNA modification at nucleotide or nucleoside level using thin layer chromatography (TLC), liquid chromatography coupled with mass spectrometry (LC-MS) and tandem MS-MS spectrometry.

Liquid chromatography coupled with mass spectrometry (LC-MS) is a powerful MS-based platform employed in several settings for investigating modifications in specific RNA at nucleoside level[32]. Reverse phase liquid chromatography (RPLC) coupled with electrospray ionization mass spectrometry (ECI-MS) is a widely used combination for characterizing and quantifying RNA modifications with high sensitivity. RNA samples are digested to nucleoside level using enzymes (nuclease P1, snake venom phosphodiesterase and alkaline phosphatase) followed by separation and analysis with LC-MS. Modified nucleosides are quantified based on comparison to the synthetic standards such as stable isotope labeling or spike in measurements. Tandem mass spectrometry (MS-MS) is the variation of technique where two or more mass analyzers are coupled to enhance the sensitivity and allows the detection of multiple substances co-eluting in a single peak[33]. Nucleosides are ionized in first spectrometer and separated based on mass-to-charge (m/z) ratio that spilt into smaller fragments ions (nucleobase and ribose moiety) followed by analyses in second detector. For example, m^6A and 2'-O-methyladenosine are coeluted in regular mass spectrometry, but the fragmentation step (i.e. glycosidic bond cleavage) makes it possible to distinguish the two species that have similar m/z-ratios. The major disadvantage of these techniques that rely on nucleotide analysis is the loss of sequence information.

1.2.2 Mapping Analysis

As the location of an RNA modification is crucial to decipher its function, various technologies and methods have been developed over the past few decades to map RNA modifications. These approaches include the enzymes, chemical labeling, and immunoprecipitation in combination of subsequent gel electrophoresis, TLC, MS, PCR, and sequencing for analysis.

1.2.2.1 Enzyme based identification

1.2.2.1.1 General enzymes

Generally, cellular enzymes behave differently when exposed to RNA modifications, which can be exploited for detection. One such example is RNase H, a non-sequence-specific endonuclease that catalyzes the cleavage of RNA in an RNA/DNA hybrid duplex substrate and was used to locate 2'-O-methylation in RNA[34]. In this method, 2'-O-methylated RNA-DNA chimeric oligonucleotide is hybridized to target RNA followed by incubation with RNase H, which cleaves unmodified RNA

and is blocked by 2'-O-methylation at target site. The cleaved products are analyzed with gel electrophoresis. Site specific RNase H cleavage has been extended to the detection of other RNA modifications such as pseudouridine (Ψ) with minor alteration in described method[35]. First, an RNA is cleaved with RNase H at the 5' side of a nucleotide of interest as mentioned before. Next, the 5' end of the nucleotide is radioactively labeled with ^{32}P followed by nuclease P1 digestion. The modified nucleotide is then identified by thin layer chromatography (TLC).

RNase T1 is an endoribonuclease that cleaves single-stranded RNA at guanosine (G) and inosine (I) residues, reading I as G like most of the other enzymes and reagents. Morse et al, developed the method to specifically cleave RNA at inosine after treating RNA with glyoxal, stabilizing the guanosine-glyoxal adduct with borate followed by cleavage with RNase T1[36]. Glyoxal forms a stable adduct only with guanosine and is resistant to RNase T1 thus facilitating the cleavage of RNA specifically at inosine residues. This method provides the potential way to map inosine in RNA. Site-specific cleavage and radioactive-labeling followed by ligation-assisted extraction and thin-layer chromatography (SCARLET) is an advanced version to determine the sites of RNA modifications[37]. As the name implies, SCARLET combines site-specific digestion by RNase H, splint ligation, RNA digestion and TLC analysis. The method has high sensitivity due to additional

Sequence-depended analysis step. It was successfully employed to investigate m^6A in mRNA and lncRNA. Another approach is based on splint ligation of oligonucleotides using modified or unmodified RNA as a template[38].

T4 DNA ligase used in this method is sensitive to the presence and the type of modification. Two pairs of oligonucleotides are selected, one is discriminating oligos pair (D-oligos) prefers ligation using unmodified over modified RNA template or vice versa and second is non-discriminating oligos pair (N-oligos) has equal ligation efficiency for unmodified and modified RNA template. Two parallel ligations with each pair are performed to detect the presence of modification. Reaction with known amount of modification is performed to draw a calibration curve. This approach was successfully used for detection of 2`-O-methylation, m^6A and pseudouridine[39].

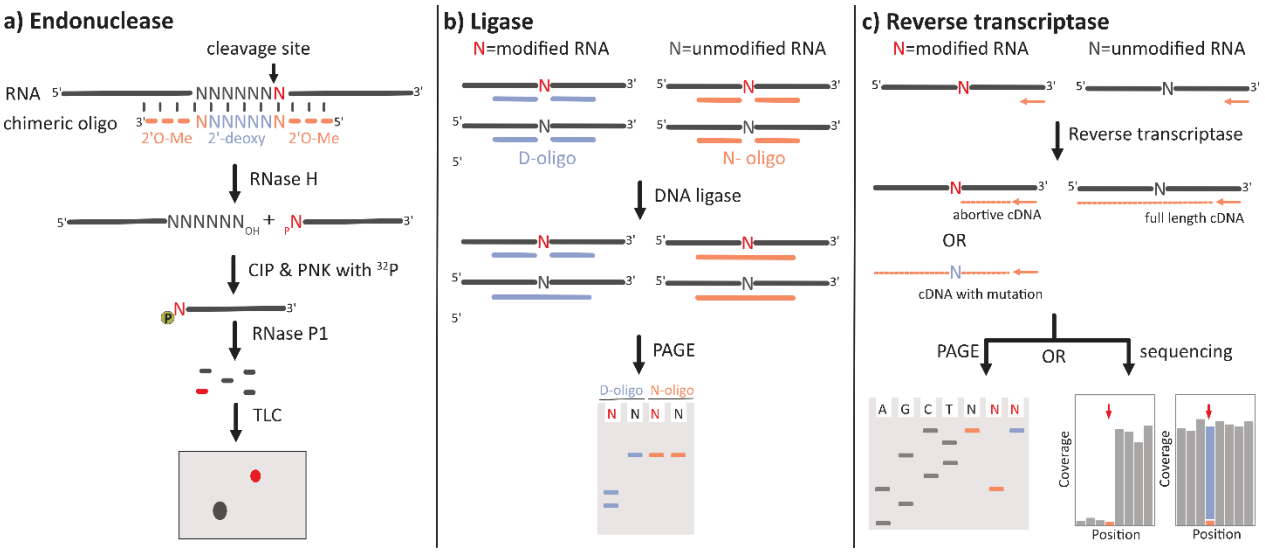


Figure 5 Schematic illustration of general enzyme-based approaches for mapping analysis of RNA modifications. a) endonuclease cleavage coupled with TLC. b) ligation followed by PAGE analysis. c) reverse transcription followed by PAGE and sequencing analysis.

Reverse transcriptase (known as RNA dependent DNA polymerase), a classical cellular enzyme that reverse-transcribes single stranded RNA into DNA (called cDNA) has been extensively used for the study of RNA modification. The enzyme is sensitive to a variety of RNA modifications usually on the Watson-Crick face of the nucleobase (m^1A, m^1G, m^3C, m^3U), which lead to RT-arrest or misincorporation of noncomplementary dNTP. RT-signatures (abortive cDNA, misincorporation of dNTP) are analyzed through gel electrophoresis or sequencing to map RNA modifications. The RT-based method does not require large amounts or purification of RNA samples as it uses target-specific primers for cDNA synthesis. The method was also successfully employed to identify A-to-I editing sites as inosine can base-pair with cytosine, during cDNA synthesis inosine was replaced with guanine and target sites can then directly be identified by comparing cDNA sequence with genome sequence[40].

Small modifications or modifications at the Hoogsteen edge are usually ignored by reverse transcriptase and result in incorporation of the correct complementary dNTP during polymerization. The enzyme's recognition properties have been exploited by modifying experimental conditions such as dNTPs concentration, engineering of reverse transcriptase or chemical labeling of modifications to generate an RT arrest signals.

Dong et al, developed RTL-P (reverse transcription at low dNTPs concentration coupled with PCR amplification) approach for relative quantification of 2'-O-methylation where site-specific primer extension by reverse transcriptase generates abortive cDNA strands at low dNTP concentration followed by semiquantitative PCR amplification (qRT-PCR)[41]. Later, an engineered KlenTaq RT-enzyme was developed as an alternative to low dNTPs conditions, that can discriminate 2'-O-methylated from unmethylated RNA and quantify 2'-O-methylation of individual nucleotides directly from total RNA extracts by a simple qRT-PCR [42]. Similarly, RT-active KlenTaq polymerase variant was evolved for detection of m^6A sites in target RNA. Engineered polymerase exhibits high misincorporation rate opposite to m^6A that can directly be identified by deep sequencing[43].

In chemical labeling approach, a specific chemical reacts with RNA modification that change their RT arrest or misincorporation for improved detection. For example, Inosine chemical erasing (ICE) was developed where acrylonitrile specifically reacts with inosine to generate N1-cyanoethylinosine followed by reverse transcription to generate RT arrest signal[44]. Abortive cDNA strand is unable to amplify, resulting only in cDNAs containing G because of A-to-I editing being erased from PCR product. Inosine chemical erasing (ICE) was combined with deep sequencing (ICE-seq) for the reliable identification of transcriptome-wide A-to-I editing sites[45]. This method eliminates the false positive results generated through single point mutations or inevitable mapping error during aforementioned direct sequencing approach for detection of inosine.

Bisulfite sequencing (BS-seq) is a well-stablished method for detecting m^5C in DNA based on the selective chemical deamination of cytosine (but not m^5C) to uracil by bisulfite treatment[46]. Cytosine form adducts, which are relatively unstable and rapidly decomposed to uracil at alkaline pH. The conversion of cytosine (but not m^5C) residues can then be detected by sequencing. However, the harsh reaction conditions (bisulfite treatment at 95°C, alkaline pH) are detrimental to the RNA stability. The bisulfite deamination method has recently been adapted to facilitate the analysis of RNA methylation patterns. The chemical deamination of cytosines in RNA under mild conditions(bisulfite treatment at 60°C), followed by PCR-based amplification of cDNA and DNA sequencing allowed the analysis of methylation status in tRNAs and rRNAs.[47].

Similarly, PSI-seq (Ψ-seq) is also a chemical labeling method that relies on unique stability of N_3-CMC-Ψ to alkaline hydrolysis and the ability to terminate reverse transcription[48]. CMCT (N-

cyclohexyl-N'-β-(4-methylmorpholinium) ethyl carbodiimide p-tosylate) modified guanosine (G), uridine (U), inosine (I) and pesudouridine (Ψ) where CMC-moieties from G, U and I can be removed under alkaline condition leaving N_3-CMC-Ψ which generate RT-arrest signal. The method is coupled with RNA-sequencing to map Ψ in mRNA and ncRNAs[49].

Ribo-Meth-seq and Nm-seq are the methods that benefit from the differential reactivity of 2'-OH and 2'-O-Me to alkaline hydrolysis and periodate oxidation respectively followed by analysis through next generation sequencing[50]. Using these strategies, a comprehensive analysis of rRNA and mRNA 2'-O-methylation sites turn out to be possible.

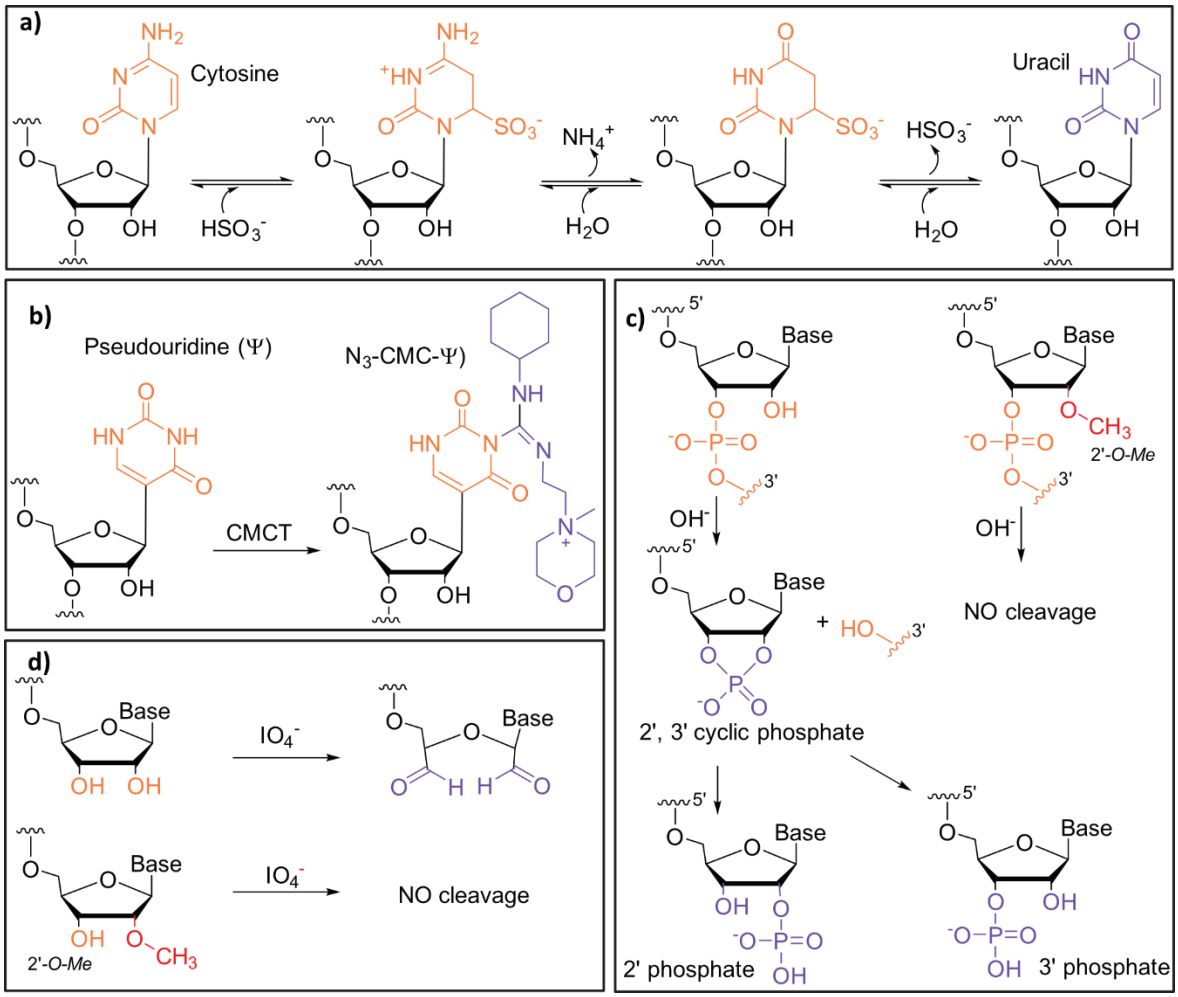


Figure 6 Chemical labeling of RNA modifications **a)** Bisulfite mediated deamination of cytosine **b)** N3-CMC-Ψ adduct formation of pseudouridine after treatment with CMCT **c)** resistance of internal 2'-O-Me to nucleolytic cleavage under alkaline conditions d) resistance of 2'-O-Me at 3'-end to periodate oxidation.

1.2.2.1.2 Modification specific enzymes

As an alternative to general enzymes (ligase, endonuclease, reverse transcriptase), modification specific enzymes could be leveraged to detect modifications. It is also useful for mapping modifications such as m^6A which is difficult to identify due to its inert chemical properties and similar response to general enzymes as adenosine. For example, DART-seq (deamination adjacent to RNA modification targets), fused YTH binding domain (a family of m^6A reader proteins) with the APOBEC1 (cytidine deaminase) to induce C-to-U deamination at sites adjacent to m^6A residues, which are subsequently identified using RNA sequencing[51]. Similarly, m^6A-SEAL is FTO-assisted selective chemical labeling method for transcriptome-wide identification of m^6A[52]. FTO-mediated oxidative demethylation of m^6A generates the unstable intermediate hm^6A (N^6-hydroxymethyladenosine) that forms a thioacetal-containing adduct (called dm^6A) when treated with dithiothreitol (DTT). The free sulfhydryl group on dm^6A was then used to install affinity tags such as biotin followed by the sequencing of enriched species having m^6A marks.

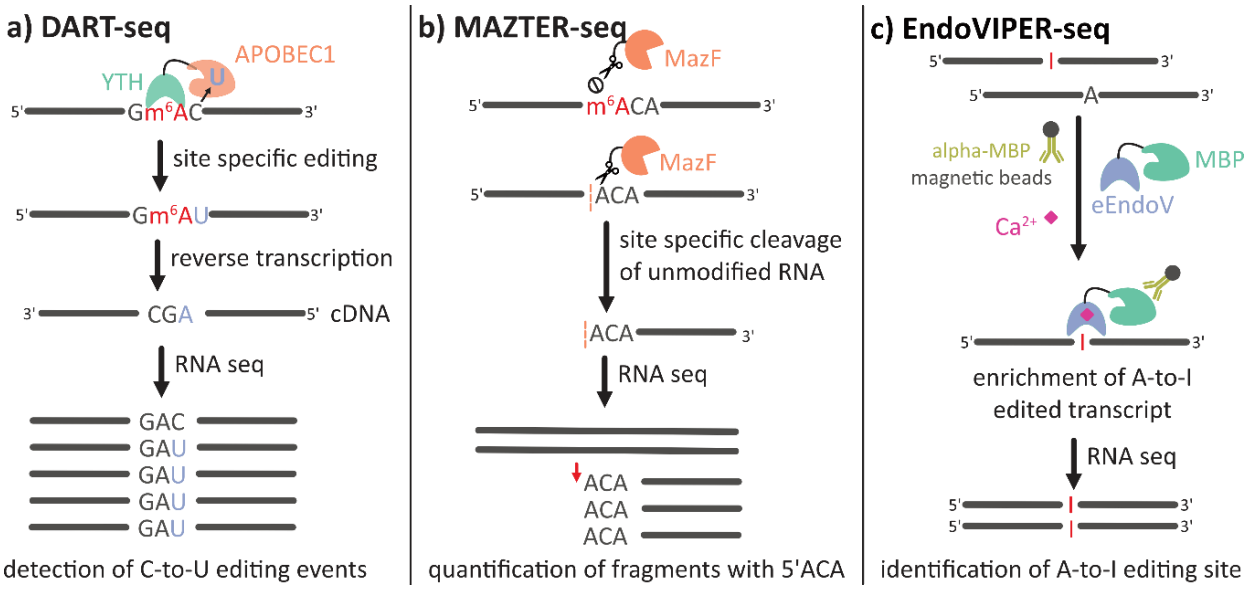


Figure 7 Schematic illustration of modification specific enzymes-based approaches for mapping analysis of RNA modifications. a) DART seq; used YTH domain fused with APOBEC1 enzyme. b) MAZTER-seq; MazF endonuclease cleave at ACA junction. c) EndoVIPER-seq; endonuclease V only binds with inosine in presence of Ca2+ to pull down inosine modified RNA.

Furthermore, m^6A-REF-seq/MAZTER-seq has been established based on bacterial endoribonuclease MazF, sensitive to m^6A and cleave immediately upstream of unmethylated ACA motif[53]. Deep sequencing of the fragments where 5' end starts with ACA or 3' end ends at ACA (unmethylated sites) and fragments cross over the site (methylated site) followed by quantification of these fragments gives a measure of methylation status at particular site. As MazF is a ribonuclease for single stranded RNAs and require ACA recognition motif, its activity can be influenced by secondary structure and sequence context, therefore methyltransferase knockouts are necessary for normalization of data to obtain absolute methylation level.

Along the same lines, EndoVIPER-seq (Endonuclease V inosine precipitation enrichment sequencing) has been established where *E.coli* Endonuclease V (eEndoV), an inosine-cleaving enzyme was repurposed to bind and isolate A-to-I edited RNA[54]. In nature, eEndoV catalyzes RNA cleavage in presence of Mg^{+2} whereas Ca^{2+} facilitates binding to inosine without cleavage. This enables high affinity capture of inosine in transcript prior to RNA-seq.

1.2.2.2 Immunoprecipitation

Enrichment of RNA containing modification before analysis is advantageous to map low-abundant modifications. RIP-seq (RNA immunoprecipitation followed by next-generation sequencing) offers a straightforward platform for transcriptome-wide analysis of RNA modifications. MeRIP-seq is the well-known approach to reveal m^6A abundance at 3'UTR and near stop codon of mRNA[55]. However, RIP-seq cannot detect the modification site at single nucleotide resolution. Modified protocols for RIP-seq involve covalent cross-linking of antibodies to the RNA modifications that generate specific RT-signature which is analyzed at single base resolution by subsequent deep sequencing. miCLIP is such an approach that combined UV light-induced antibody-RNA cross-linking with reverse transcription to capture specific mutations at m^6A sites[56]. The antibodies to study other RNA modifications such as m^1A, m^5C, Ψ are available. However, immunoprecipitation suffers from several limitations such as i) Antibody-based approaches are unable to quantify the stoichiometry of modified sites in a transcriptome-wide manner ii) Require relatively large amounts of RNA samples iii) The specificity of antibodies is not high enough to distinguish structurally similar RNA modifications in some cases as m^6A and m^6Am.

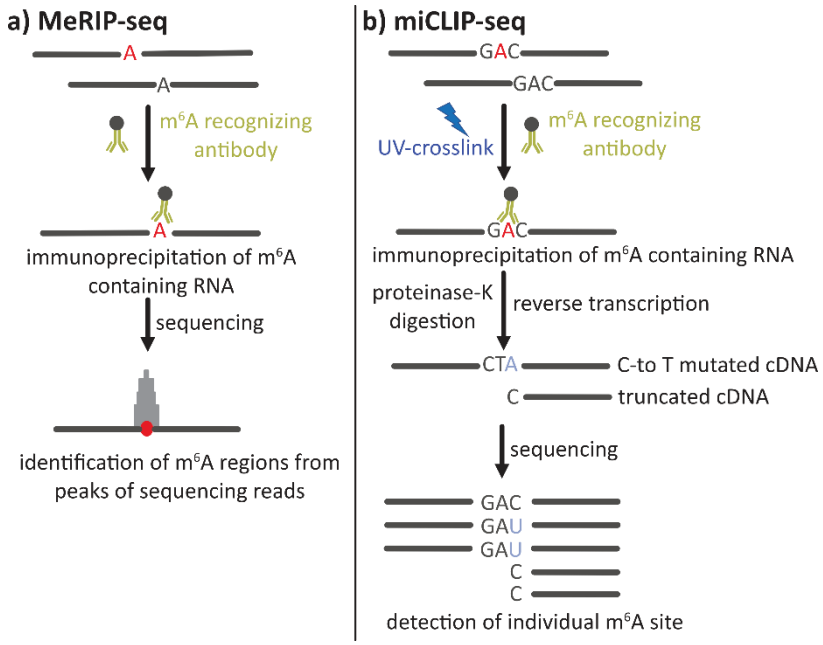


Figure 8 Schematic illustration of enrichment of modified RNA by specific antibodies followed by sequencing analysis. a) enrichment of modified RNA using specific antibody b) enrichment of modified RNA using specific antibody UV-crosslinked to RNA, followed by reverse transcription and sequencing.

1.3 Evolution of deoxyribozymes for detection of RNA modification

Naturally, deoxyribonucleic acid (DNA) is the molecule that store genetic information and ribonucleic acid (RNA) translates that coded information into functional proteins. In early 1980s, the catalytic ability of RNA was discovered by Cech and Altman, and the ribonucleases made of RNA were called ribozymes[57]. The discovery of natural ribozymes inspired the scientist to develop in vitro selection technique (also known as SELEX) for evolution of artificial ribozymes from a random pool of RNA sequences[58]. The iterative process of selection involves the enrichment of active species and elimination of inactive sequences from a random pool of ~10^{14}-10^{16} unique sequences. The frequency of active sequences gradually increases in successive rounds of functional selection and amplification. Eventually, the population converges on the most active species in the pool [59]. Several ribozymes have been exclusively identified through SELEX and performed variety of chemical reactions including RNA cleavage, ligation, and polymerization[60].

Since DNA has never been identified in nature as a catalyst, the discovery of artificial deoxyribozymes (DNAzymes or DNA enzymes) broadened the concept of enzymatic catalysis and introduce DNAzymes as new contributors in the field. In 1994, Breaker and Joyce identified the first Pb^{2+} dependent RNA-cleaving deoxyribozymes[61]. In the following years, hundreds of deoxyribozymes were discovered that can perform a range of reactions. The reactions catalyzed by DNAzymes include RNA hydrolysis, RNA and DNA ligation, peptide modification, carbon-carbon bond formation, hydrolysis of amides, azide-alkyne cycloaddition and many more[62].

1.3.1 In vitro selection

As the catalytically active DNA are single stranded molecules, the selection procedure begins with synthesis of single stranded DNA library by solid-phase synthesis. The random region is usually 40-80 nucleotide in length depending on the type of reaction being catalyzed. In some cases, the random region can be shortened to twenty nucleotides. To facilitate PCR amplification, random region of pool is flanked by constant sequences called primer binding site (PBS). The substrate sequence is introduced during PCR step or covalently attached to the random pool via ligation. After the construction of the desired library, it follows three major steps of the selection process: selection reaction, capture, and amplification.

Selection reaction: A portion of initial library (~10^{14}-10^{16} different sequences) is incubated under appropriate conditions where active species can perform the desired reaction. Chemical transformation is usually coupled with the physical transformation of active sequences that allow the separation in the next step.

Capture: The active sequences are separated from inactive species during capture step depending on the nature of catalysis being performed by library. For example, in case of an RNA cleavage reaction, sized-based separation is performed to capture active species. During denaturing gel electrophoresis, active species being smaller in size run faster, thereby allowing their separation followed by isolation from gel.

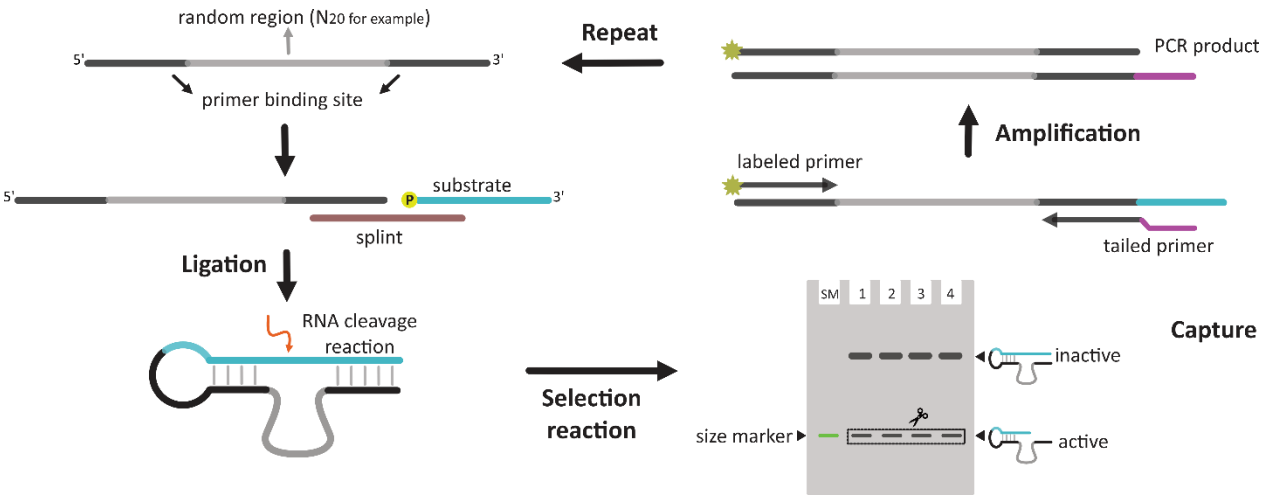


Figure 9 Schematic presentation of in vitro selection scheme for evolution of RNA-cleaving deoxyribozymes.

Amplification: After the active sequences have been separated, they are amplified by PCR using primers that bind to the binding sites that flank the random region. As PCR amplification generates double stranded DNA product, modified primers (tailed primers or 5' phosphorylated primers) are used to generate single stranded DNA library for next round of selection.

The selection process is then repeated for 5 – 10 rounds to enrich the library with most active species. As the selection proceeds, more stringent conditions are introduced to ensure the evolution of most efficient candidates. After a certain enrichment level, the active pool is subjected to sequencing to identify and characterize individual deoxyribozymes.

1.3.2 Deoxyribozymes for detection of RNA modifications

The deoxyribozymes 8-17 and 10-23 are the most studied RNA-cleaving DNA enzyme that facilitate the site-specific cleavage of RNA substrate[63]. 10-23 has the ability to cleave any purine-pyrimidine junction with robust activity for P-U and significantly reduced activity for R-C (R= A or G)[64]. 8-17 originally was selected for the cleavage at A-G junction but numerous variants of the 8-17 motif have been repeatedly isolated in different independent in-vitro selections. Neither of the variant was able to cleave all the possible junctions. However, the collective ability of 8-17 variants showed cleavage for all 16 dinucleotide junctions according to the order NG > NA > NC > NU[65]. As these variants have been selected using chimeric substrate (DNA substrate containing

single rN), it is unclear if the cleavage site versatility is effective against all-RNA substrates. The activity of the 10-23 and 8-17 deoxyribozymes depend on divalent ions (Mg^{2+} and Mn^{2+}). These endonucleases catalyze a site-specific transesterification reaction, where the 2'-hydroxyl group attacks on the adjacent phosphodiester linkage, forming 2'-3' cyclic phosphate and 5'-hydroxyl termini on the two RNA fragments.

Deoxyribozymes usually have a conserved catalytic core flanked by two variable binding arms, which can be designed to target any desired RNA substrate. DNAzymes exhibiting versatile substrate scope have a great potential for biological applications such as potential therapeutics, and biosensors[66]. DNAzymes have been employed for the identification of RNA modifications. Buchhaupt et al, manipulated 10-23 and 8-17 DNAzymes for site-specific analysis of 2'-O-Me and pseudouridine[67]. As 2'-OH group in RNA is involved in ribonuclease reaction mechanism, its methylation at the target site prevents DNAzyme mediated cleavage. However, this method is only applicable if modification is directly interacting with functional groups involved in the key cleavage reaction.

In another approach, tRNA modifications have been analyzed using DNAzyme-directed cleavage, followed by radioactive labeling of 5'OH of cleaved modified fragment. The labeled fragment was further analyzed by enzymatic digestion to mononucleotide and thin layer chromatography[68]. Importantly, DNAzymes must have equal cleavage efficiency for modified and unmodified RNA. As the influence of RNA modification on DNAzyme activity is not extensively studied, this approach is limited to study the modification present at non-Watson-Crick edges (e.g., m^5C, m^5U, Ψ). An alternative approach has been established where modified region of RNA was excised from full-length RNA using a pair of deoxyribozymes[68]. These two DNAzymes cleave at unmodified positions several nucleotides upstream and downstream of the modification. The excised fragment was further analyzed using conventional methods such as 2D-TLC. However, these deoxyribozymes are not able to directly report the presence of RNA modifications as the cleavage activity is independent of the modified nucleotide.

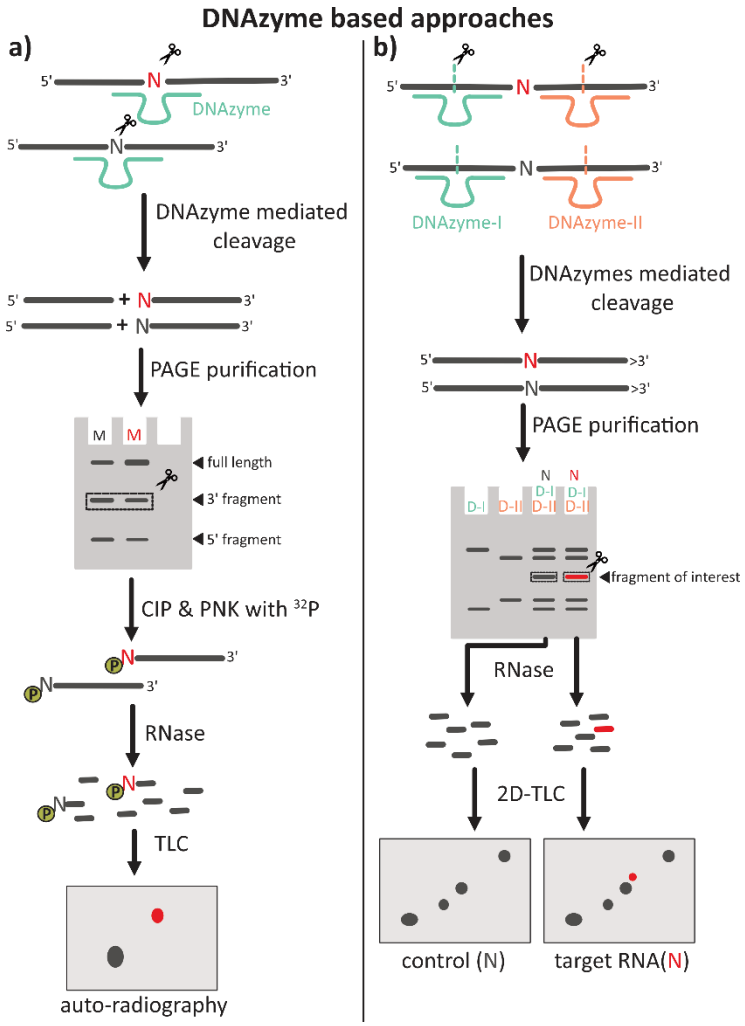


Figure 10 Schematic presentation of previously reported DNAzyme-based approaches for detection of RNA modifications. a) Cleavage of RNA at the target site. b) excision of modified fragments using two DNAzymes

Deoxyribozymes that can directly distinguish modified from unmodified RNA have only recently been added to the toolbox. The first example was a family of RNA-cleaving deoxyribozymes have been identified for the site-specific interrogation of m^6A in endogenous RNA such as lncRNA and a set of C/D box snoRNAs[69]. VMC10 differentiates N^6-methyladenosine from adenosine and preferentially cleaves unmodified RNA. The methyl group presumably can affect the active conformation of the catalytic core, consequently inhibiting its activity. On the contrary, VMA8 and VMA15 were the DNA enzyme that exhibits accelerated rate of reaction for m^6A modified RNA. However, absence of methyl group does not completely inactivate the enzyme thus makes it less specific. This was the starting point for the work performed in this thesis, as no DNAzymes for the examination of any other modified nucleotides have been known.

1.4 RNA structure

The distinct ability of RNA molecules to form secondary and complex tertiary structures encodes an additional layer of information. Indeed, the three-dimensional structure of RNA is the prerequisite for all functions beyond the transfer of the sequence information. The RNA sequence is called the primary structure, it is a direct copy of the complementary DNA sequence and encodes the sequence that is translated into the amino acid sequence of a protein. Besides serving as the binding site for interacting proteins, the RNA primary sequence is also guiding the formation of RNA-RNA interactions by Watson-Crick base pairing. Intramolecular base-pair formation gives rise to RNA secondary structures as the basis for functional motifs. Hairpin loops, stem loops and bulges are widely known RNA secondary structure motifs, where short double-stranded regions are interrupted by single-stranded regions. RNA secondary structure formation mostly involves Watson-Crick base pairing, but a large diversity of non-canonical base pairs also plays a crucial role in RNA architectures. These include complex H-bonding patterns involving several nucleobases, as for example in base triples, but and simple motifs are known. For example, G:U wobble base pairing in tRNA allows effective aminoacylation and decoding of mRNA in ribosomes[70].

Multiple secondary structures assemble to form a high-order tertiary RNA structure stabilized by cations, osmolytes, molecular crowding, and interacting protein in the cellular environment[71]. The local conformation of secondary structure, the interaction of helices at junctions , and sequence-specific interactions are the key determinants of overall RNA tertiary structure[72]. Coaxial stacking (also known as helical stacking) is usually observed when two RNA helices are stabilized by base stacking at their interface to form a coaxial arrangement. Kissing loops and pseudoknots are well-known motifs that exhibit coaxial stacking. RNA junctions are additional examples of coaxial stacking; Y-shaped three-way junctions are widely distributed and play an essential role in assembling large RNA complexes such as riboswitches and ribozymes. Similarly, long-range interactions between hairpin stem-loops (kissing loop) or loop nucleotides and complementary RNA regions (pseudoknots) through Watson-Crick base pairing have been identified in many RNAs. The classical example of such interaction is the folding of tRNA cloverleaf

secondary structure into L-shaped tertiary structure due to long-range base pairing between D-loop and T-loop[73]

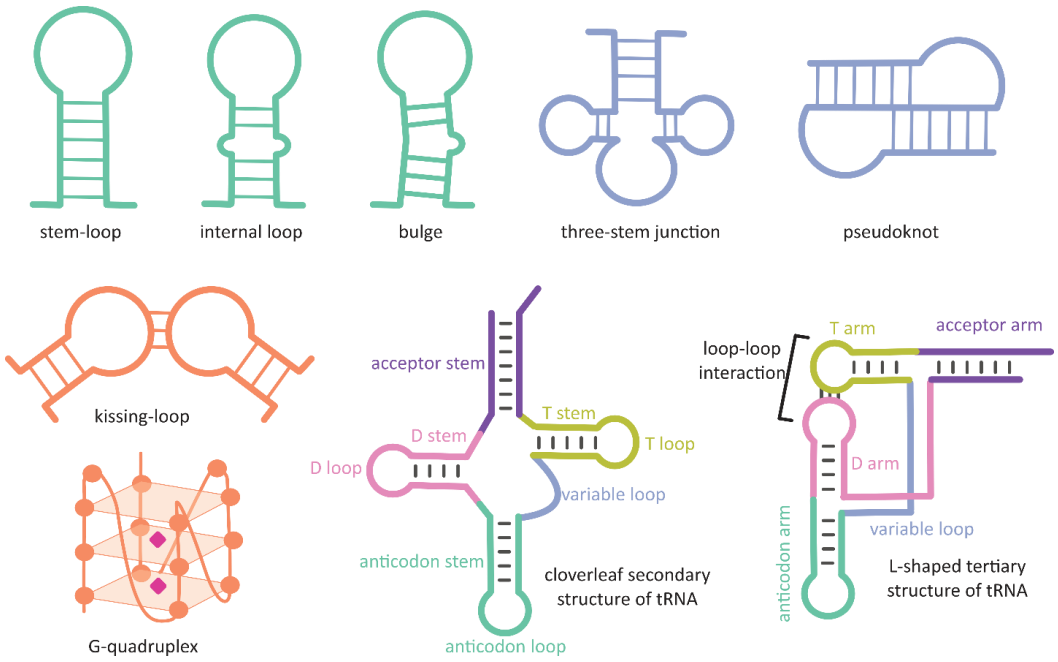


Figure 11 Most commonly observed RNA secondary structures in RNA; stem-loop, internal loop, bulge, three stem junction, pseudoknot, kissing loop and non-canonical RNA secondar structure i.e., RNA G-quadruplex. secondary structures in tRNA and their interactions to stabilize three-dimensional L-shaped tRNA.

A complex three dimensional architecture is observed in functional RNAs such as riboswitches, ribozymes, and regulatory elements. Riboswitches are structured RNA domains comprised of an aptamer and an expression platform to serve as intracellular signal-sensing genetic switches[74]. The signals include metal ions, nucleobases, amino acids, and cofactors[75]. The untranslated regions (UTRs) of mRNA are well known to exhibit these genetic control domains where ligand binding modulates the mRNA structure resulting in the gene-expression process such as transcription termination and translation initiation[76].

Ribozymes are RNA molecules with distinct tertiary structures that enable them to catalyze biochemical reactions. A primary class of such ribozymes is RNA-cleaving enzymes that direct cleavage in cis (self-cleavage) or in trans (cleavage of other RNA). The ribonucleoprotein (RNP) RNase P was identified to cleave tRNA molecules while cis-acting ribozymes have self-cleaving RNA domains such as hammerhead, hairpin satellite, and class II introns[57b, 77] . During protein

synthesis, ribosomal RNA (rRNA) plays a role as catalytic RNA molecule that perform peptidyl transferase reaction.[78].

RNA being a single stranded molecule has privilege to adopt a complex high-order architecture such as RNA G-quadruplex (rG4s), a non-canonical secondary structure that can be predicted from its primary sequence. Guanine repeats separated by loop regions form a stable four-stranded conformation that is assembled from stacks of guanine tetrads (called G-quartets) connected by Hoogsteen base pairing[79]. G4s is stabilized by cations such as K^+ and Na^+ which fits into the space between two quartets[80]. Additionally, the 2' hydroxyl group in the ribose sugar enhances the rG4s stability by forming intramolecular interactions within the structure[81]. RNA G-quadruplex is a structured element observed in various RNAs such as 5S ribosomal RNA (rRNA), messenger RNA (mRNA), transfer RNA fragments (tiRNA), and non-coding RNA (such as lncRNA, miRNA, telomere-associated RNA) [79, 82].

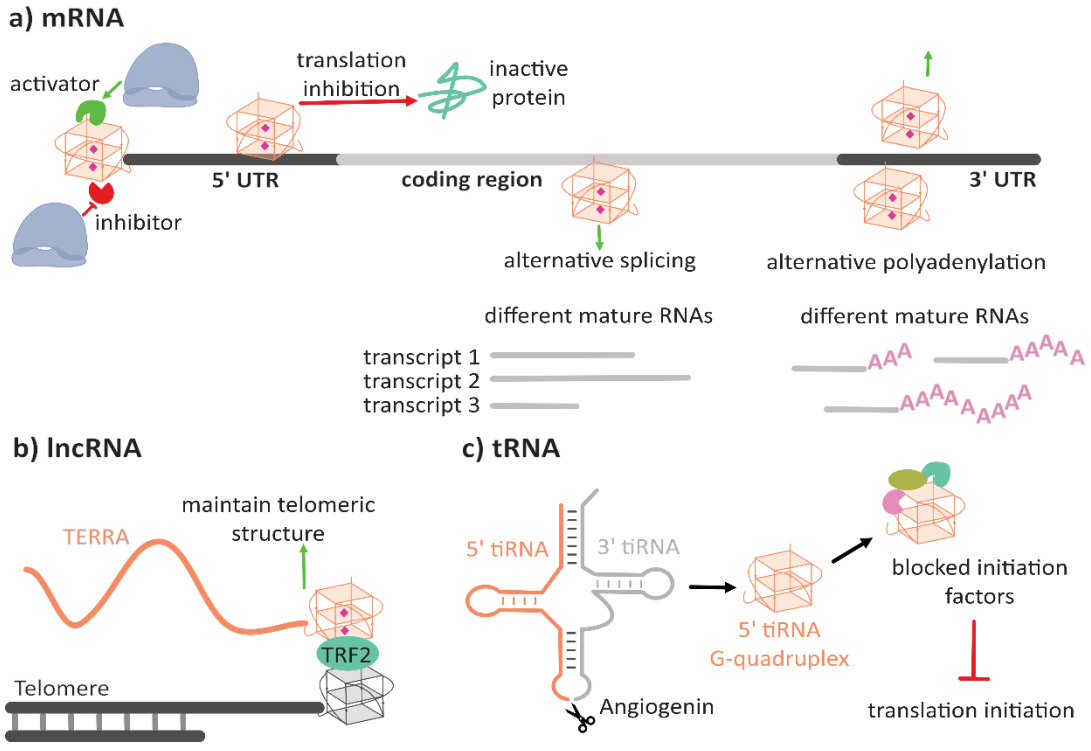


Figure 12 Role of RNA G-quadruplex in post-transcriptional regulation **a)** rG4s present in coding and non-coding regions controlling translation at different levels. **b)** rG4s in lncRNA play a role in maintaining telomeric structure. **c)** rG4s in tRNA form 5' tiRNA G-quadruplex which control translation initiation under stress conditions.

In mRNA, rG4s are found in functional regulatory regions, primarily 5' UTR and 3' UTR, where they function as either repressors or enhancers to modulate gene expression at various stages of protein synthesis[83]. Additionally, rG4s in mRNA are known to be involved in mRNA processing, stability, transport, localization, and alternative splicing[84]. tRNA-derived fragments assembled into intermolecular rG4s are required for the regulation of mRNA translation by the formation of stress granules during stress conditions[85]. Similarly, G-rich TERRA (telomeric repeat-containing RNA) has been shown to form rG4s structure and are considered to be important for the maintenance of telomere length, telomeric structure via interaction with telomere associated proteins and telomerase activity[86].

1.5 Existing method for detection of RNA G-quadruplex (rG4s)

The emerging role of the RNA G-quadruplex in regulating gene expression has driven strong interest in mapping rG4s in the transcriptome. Biophysical methods such as UV melting, UV spectrometry, and circular dichroism are the techniques usually performed in buffers containing monovalent cations (K+/ Na+) for detection of quadruplex formation in target sequence[87]. The characteristic signature signals are the key determinants of rG4s structure such as a large variation in absorbance at 295 nm compared to 260 nm during UV melting and a strong positive peak at 260 nm and a shallow negative peak at 240 nm during CD scan[87c]. Nevertheless, these assays might not reflect the actual folding of rG4s as they measure homogenous and short RNA oligonucleotides.

Chemical ligands are advantageous to study the structure and folding dynamics of the G-quadruplex with high sensitivity. Small synthetic molecules typically interact with G4s through π-stacking or carry a side chain that enhances contact with loops or inner quartets. In contrast to DNA G-quadruplex, a few ligands are shown to be specific for RNA G-quadruplex. For example, acridine binding to TERRA with additional interactions such as involvement of 2' hydroxyl group of ribose sugar and loops, not observed in DNA counterpart[88]. These unique chemical properties of rG4s then lead to the development of carboxypyridostatin that exhibit high specificity for RNA compared to DNA G-quadruplex[89]. The ligands are usually engineered to carry molecular tags such as fluorophores that changes fluorescence properties like emission maximum and intensity to study the quadruplex formation[90]. Moreover, affinity tags attached to specific ligands

facilitate the pull-down of G4s targets followed by sequencing to study the G4s transcriptomic landscape[91].

Reverse transcriptase (RT) foot-printing is the technique based on reverse transcriptase sensitive to covalent modifications or stable structures on template RNA. RT-stop signals can be identified as abortive cDNA bands on denaturing polyacrylamide gel electrophoresis. These methods involved the development of chemical reagents which selectively modified unstructured regions of RNA. The well-established techniques in these genera are SHAPE (2'-hydroxyl acylation), DMS probing (N7-methylation of guanine), and kethoxal foot-printing (N1 and N-NH2 alkylation of guanine). As a result, the structure and unstructured regions of RNA can be identified according to the rate of chemical modification indicated by the frequency of RT stops[92]. Later, these methods were coupled with RNA-seq technology such as rG4 seq, SHALiPE-seq, Keth -seq, RT-stop profiling, and G4RP-seq for high throughput profiling of rG4s [93]. The methods may be ambiguous when the targeted footprint is not unique to rG4s or ligands that bind to rG4s to stabilize the structure that does not occur under physiological conditions.

1.6. Site-specific RNA labeling

RNA labeling and visualization is important to study RNA dynamics both in vitro and in living cells. A wide range of RNA labeling methods has been developed, and these primarily rely on covalent or non-covalent binding of fluorophores to the RNA of interest. Non-covalent RNA labeling approaches usually involves the hybridization of fluorescently labeled oligonucleotide or binding of a fluorescent protein to the target RNA[94]. These methods however suffer from challenges such as high background fluorescence, cell permeability, probe instability and the large size of probe. Fluorogenic activating aptamers have emerged as an alternative approach relies on small-molecule recognition properties of RNA[95]. Several RNA aptamers have been reported where aptamer sequence binds to a small molecule, causing activation or increase in fluorescent properties of the ligand. RNA aptamers especially that mimics fluorescent proteins (GFP) and binds to analogue of the GFP fluorophore have certainly made significant advancement in RNA imaging technology[96]. However, noncovalent interaction of small molecules limits the scope of their potential application.

An alternative approach for site-specific, covalent labeling of RNA is the use of RNA catalyst. Self-alkylating fluorescein-iodoacetamide reactive ribozyme 1FR1 and 5FR1 were identified to label themselves at an internal position in the presence of fluorescein iodoacetamide[97]. The epoxide reactive ribozyme derived from *Aeropyrum prenix* genomic RNA is another self-alkylating ribozyme that reacts with a biorthogonal small molecule for selective and irreversible RNA modification[98]. However, these ribozymes are self-alkylating and need to be inserted into the target RNA. In contrast, twin-ribozyme has been developed to label RNA in-trans. Twin ribozyme is an engineered version of hairpin ribozymes where two copies of hairpin ribozymes are connected through a linker sequence[99]. Ribozyme recognize RNA of interest through complementary base pairing and catalyze the excision of an RNA fragment which is then replaced by the labeled RNA piece at desired position.

Recently, FH14 ribozyme was reported as highly versatile and efficient trans-acting ribozyme for site-specific labeling of RNA[100]. The ribozyme catalyzes the covalent binding of N^6-labeled adenosine to the internal 2'-OH group of bulged adenosine in target RNA through 2',5'-phosphodiester bond. FH14 has shown good scope for a broad range of N^6-labeled ATP derivatives and catalyzed labeling of 5S rRNA in total RNA cellular context extract.

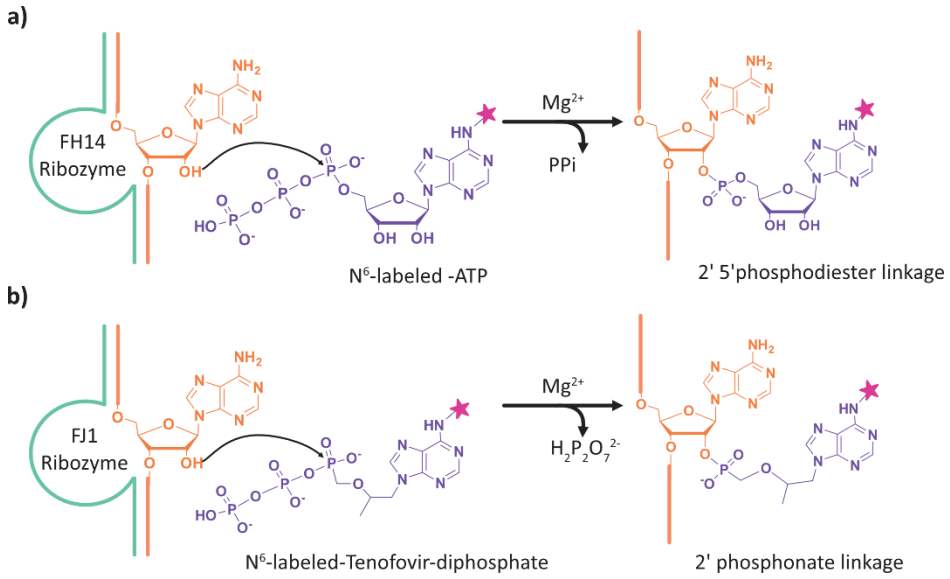


Figure 13 Ribozyme-catalyzed RNA labeling reaction a) FH14 ribozyme catalyze the labeling via 2'–5' phosphodiester linkage using N^6-labeled ATP b) FJ1 ribozyme catalyze the labeling via 2' phosphonate linkage using N^6-labeled-Tenofovir-diphosphate.

Although FH14 is a quite efficient in vitro RNA labeling tool, it has certain limitations for potential cellular applications. First, N^6-labeld ATP analogues have the propensity to compete with cellular ATP and ATP-dependent enzymatic processes that could result in ribozyme-independent background labeling of RNA. Secondly, the 2',5'- phosphodiester linkage in the labeled RNA is easily cleaved by natural debranching enzymes. To address this problem, a new ribozyme FJ1 was evolved that use orthogonal substrates, derived from the antiviral drug tenofovir, and attach biorthogonal functional groups, through a hydrolytically more stable phosphonate ester linkage[101].

The existing methods recognize target RNA based on sequence rather than structure, therefore we aimed to develop structure-specific ribozymes that label the target RNA structure.

2. Aims of the study

This thesis follows two general aims that address current needs in the fields of RNA modifications and RNA structure and labeling outlined in the introduction. The first part aims at the development of deoxyribozymes as tools for the detection of RNA modifications, and the second part aims at the development of structure-specific ribozymes for the recognition and labeling of RNA quadruplex structure motifs.

The epitranscriptomics is extensively flourishing field that requires sophisticated tools for identification of RNA modifications to better understand the underlying mechanism of their action. Deoxyribozymes are the recent addition to the toolbox for the site-specific identification and validation of ribose modification. Therefore, the goal of our study is to expand the DNAzyme toolbox for the detection of RNA modifications other than m^6A. We decided to start a new in vitro selection for the evolution of deoxyribozymes that could identify i^6A (isopentenyladenosine) modification, conserved in certain tRNAs. In addition to its biological significance, we are interested in addressing this modification due to its structural similarity to m^6A that would facilitate the isolation of DNA enzyme that could differentiate analogues or derivatives of modified nucleotides. This selection experiments led to the identification of highly specific deoxyribozymes with desirable properties. We then switched our focus to the RNA modifications that are positional isomers of each other and more challenging to address. Monomethylated isomers of cytidine (m^3C, m^4C and m^5C) were chosen as the next target modifications. As three modifications present in different sequence context in cellular RNAs, we modified the in vitro selections design to evolve potentially universal DNAzymes that could target various RNA sequencing. As the typical SELEX process requires several rounds of selection and subsequent characterization of candidate catalysts by gel shift-based kinetic assays, it is a laborious and time-consuming process. We aimed to develop a robust method for the discovery and activity profiling of modification-specific deoxyribozymes.

The dynamic RNA structures play a regulatory role in cellular processes. The development of new methods to identify RNA structures is in high demand in order to understand their potential role and dynamics during different cellular events. In this regard, a distinct approach was developed

to identify target hairpin structure using RNA aptamer through loop-loop interaction [102]. The method employed inherent behavior of RNA to develop RNA-RNA interactions for recognition of specific structure and has shown to be a more specific binder than the antisense oligonucleotide[103]. Inspired by this approach, we aimed to develop structure-specific ribozymes that could recognize RNA-quadruplexes through tertiary interactions, followed by labeling using small cofactors. Ribozyme is the molecule of choice in this study since they can attach labels covalently and site-specifically under near-physiological conditions and can easily be evolved using in vitro selection approach [100, 104]. Additionally, catalytic RNA has the advantage over DNA enzymes that it can easily be transferred into cells using a vector and transcribed in vivo by cellular machinery. Thus, ribozymes could potentially be used for recognizing and labeling rG4 under cellular conditions.

References

[1] P. Boccaletto, B. Bagiński, in *RNA Bioinformatics*, Springer, **2021**, pp. 481-505.

[2] aS. Delaunay, M. Frye, *Nature cell biology* **2019**, *21*, 552-559; bN. Jonkhout, J. Tran, M. A. Smith, N. Schonrock, J. S. Mattick, E. M. Novoa, *Rna* **2017**, *23*, 1754-1769.

[3] C. Lorenz, C. E. Lünse, M. Mörl, *Biomolecules* **2017**, *7*, 35.

[4] P. F. Agris, *EMBO reports* **2008**, *9*, 629-635.

[5] aF. Tuorto, C. Legrand, C. Cirzi, G. Federico, R. Liebers, M. Müller, A. E. Ehrenhofer-Murray, G. Dittmar, H. J. Gröne, F. Lyko, *The EMBO journal* **2018**, *37*, e99777; bX. Wang, Z. Matuszek, Y. Huang, M. Parisien, Q. Dai, W. Clark, M. H. Schwartz, T. Pan, *Rna* **2018**, *24*, 1305-1313.

[6] aC. T. Chan, W. Deng, F. Li, M. S. DeMott, I. R. Babu, T. J. Begley, P. C. Dedon, *Chemical research in toxicology* **2015**, *28*, 978-988; bS. D'Silva, S. J. Haider, E. M. Phizicky, *Rna* **2011**, *17*, 1100-1110.

[7] aP. C. Durant, A. C. Bajji, M. Sundaram, R. K. Kumar, D. R. Davis, *Biochemistry* **2005**, *44*, 8078-8089; bJ. W. Stuart, Z. Gdaniec, R. Guenther, M. Marszalek, E. Sochacka, A. Malkiewicz, P. F. Agris, *Biochemistry* **2000**, *39*, 13396-13404.

[8] J. W. Stuart, K. M. Koshlap, R. Guenther, P. F. Agris, *Journal of molecular biology* **2003**, *334*, 901-918.

[9] aS. K.-L. Schultz, U. Kothe, *RNA* **2020**, *26*, 1131-1142; bM. Helm, R. Giegé, C. Florentz, *Biochemistry* **1999**, *38*, 13338-13346.

[10] aY. Furuichi, *Proceedings of the Japan Academy, Series B* **2015**, *91*, 394-409; bM. Werner, E. Purta, K. H. Kaminska, I. A. Cymerman, D. A. Campbell, B. Mittra, J. R. Zamudio, N. R. Sturm, J. Jaworski, J. M. Bujnicki, *Nucleic acids research* **2011**, *39*, 4756-4768.

[11] A. Galloway, V. H. Cowling, *Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms* **2019**, *1862*, 270-279.

[12] aJ. Liu, Y. Yue, D. Han, X. Wang, Y. Fu, L. Zhang, G. Jia, M. Yu, Z. Lu, X. Deng, *Nature chemical biology* **2014**, *10*, 93-95; bK. E. Pendleton, B. Chen, K. Liu, O. V. Hunter, Y. Xie, B. P. Tu, N. K. Conrad, *Cell* **2017**, *169*, 824-835. e814; cG. Zheng, J. A. Dahl, Y. Niu, P. Fedorcsak, C.-M. Huang, C. J. Li, C. B. Vågbø, Y. Shi, W.-L. Wang, S.-H. Song, *Molecular cell* **2013**, *49*, 18-29; dG. Jia, Y. Fu, X. Zhao, Q. Dai, G. Zheng, Y. Yang, C. Yi, T. Lindahl, T. Pan, Y.-G. Yang, *Nature chemical biology* **2011**, *7*, 885-887.

[13] aX. Zhao, Y. Yang, B.-F. Sun, Y. Shi, X. Yang, W. Xiao, Y.-J. Hao, X.-L. Ping, Y.-S. Chen, W.-J. Wang, *Cell research* **2014**, *24*, 1403-1419; bW. Xiao, S. Adhikari, U. Dahal, Y.-S. Chen, Y.-J. Hao, B.-F. Sun, H.-Y. Sun, A. Li, X.-L. Ping, W.-Y. Lai, *Molecular cell* **2016**, *61*, 507-519; cK. D. Meyer, D. P. Patil, J. Zhou, A. Zinoviev, M. A. Skabkin, O. Elemento, T. V. Pestova, S.-B. Qian, S. R. Jaffrey, *Cell* **2015**, *163*, 999-1010; dX. Wang, B. S. Zhao, I. A. Roundtree, Z. Lu, D. Han, H. Ma, X. Weng, K. Chen, H. Shi, C. He, *Cell* **2015**, *161*, 1388-1399.

[14] aD. Dunn, *Biochimica et biophysica acta* **1961**, *46*, 198-200; bR. H. Hall, *Biochimica et Biophysica Acta (BBA)-Specialized Section on Nucleic Acids and Related Subjects* **1963**, *68*, 278-283.

[15] aH. Grosjean, H. Grosjean, *Fine-tuning of RNA functions by modification and editing, Vol. 24*, Springer, **2005**; bD. Dominissini, S. Nachtergaele, S. Moshitch-Moshkovitz, E. Peer, N.

Kol, M. S. Ben-Haim, Q. Dai, A. Di Segni, M. Salmon-Divon, W. C. Clark, *Nature* **2016**, *530*, 441-446.
[16] aY. Motorin, F. Lyko, M. Helm, *Nucleic acids research* **2010**, *38*, 1415-1430; bX. Yang, Y. Yang, B.-F. Sun, Y.-S. Chen, J.-W. Xu, W.-Y. Lai, A. Li, X. Wang, D. P. Bhattarai, W. Xiao, *Cell research* **2017**, *27*, 606-625; cL. Fu, C. R. Guerrero, N. Zhong, N. J. Amato, Y. Liu, S. Liu, Q. Cai, D. Ji, S.-G. Jin, L. J. Niedernhofer, *Journal of the American Chemical Society* **2014**, *136*, 11582-11585.
[17] Y. A. Savva, L. E. Rieder, R. A. Reenan, *Genome biology* **2012**, *13*, 1-10.
[18] aK. Nishikura, *Annual review of biochemistry* **2010**, *79*, 321-349; bM. J. Palladino, L. P. Keegan, M. A. O'connell, R. A. Reenan, *Cell* **2000**, *102*, 437-449; cA. Tariq, M. F. Jantsch, *Frontiers in neuroscience* **2012**, *6*, 99.
[19] S. Sharma, D. L. Lafontaine, *Trends in biochemical sciences* **2015**, *40*, 560-575.
[20] M. Charette, M. W. Gray, *IUBMB life* **2000**, *49*, 341-352.
[21] D. J. Williams, J. L. Boots, K. B. Hall, *RNA* **2001**, *7*, 44-53.
[22] K. E. Bohnsack, C. Höbartner, M. T. Bohnsack, *Genes* **2019**, *10*, 102.
[23] M. Penzo, L. Montanaro, *Biomolecules* **2018**, *8*, 38.
[24] X.-h. Liang, Q. Liu, M. J. Fournier, *Rna* **2009**, *15*, 1716-1728.
[25] K. E. Sloan, A. S. Warda, S. Sharma, K.-D. Entian, D. L. Lafontaine, M. T. Bohnsack, *RNA biology* **2017**, *14*, 1138-1152.
[26] G. Keith, *Biochimie* **1995**, *77*, 142-144.
[27] H. Grosjean, L. Droogmans, M. Roovers, G. Keith, *Methods in enzymology* **2007**, *425*, 55-101.
[28] J. Stanley, S. Vassilenko, *Nature* **1978**, *274*, 87-89.
[29] C. Burtis, *Journal of Chromatography A* **1970**, *51*, 183-194.
[30] C. W. Gehrke, K. C. Kuo, *Journal of Chromatography A* **1989**, *471*, 3-36.
[31] G. Nees, A. Kaufmann, S. Bauer, in *Innate DNA and RNA Recognition*, Springer, **2014**, pp. 3-14.
[32] S. C. Pomerantz, J. A. McCloskey, in *Methods in enzymology, Vol. 193*, Elsevier, **1990**, pp. 796-824.
[33] K. Thüring, K. Schmid, P. Keller, M. Helm, *Methods* **2016**, *107*, 48-56.
[34] Y. Yu, M. D. Shu, J. A. Steitz, *Rna* **1997**, *3*, 324.
[35] X. Zhao, Y.-T. Yu, *Rna* **2004**, *10*, 996-1002.
[36] D. P. Morse, B. L. Bass, *Biochemistry* **1997**, *36*, 8429-8434.
[37] N. Liu, M. Parisien, Q. Dai, G. Zheng, C. He, T. Pan, *Rna* **2013**, *19*, 1848-1856.
[38] M. Saikia, Q. Dai, W. A. Decatur, M. J. Fournier, J. A. Piccirilli, T. Pan, *Rna* **2006**, *12*, 2025-2033.
[39] Q. Dai, R. Fong, M. Saikia, D. Stephenson, Y.-t. Yu, T. Pan, J. A. Piccirilli, *Nucleic acids research* **2007**, *35*, 6322-6329.
[40] D. Dominissini, S. Moshitch-Moshkovitz, N. Amariglio, G. Rechavi, *Carcinogenesis* **2011**, *32*, 1569-1577.
[41] Z.-W. Dong, P. Shao, L.-T. Diao, H. Zhou, C.-H. Yu, L.-H. Qu, *Nucleic acids research* **2012**, *40*, e157-e157.
[42] J. Aschenbrenner, A. Marx, *Nucleic acids research* **2016**, *44*, 3495-3502.

[43] J. Aschenbrenner, S. Werner, V. Marchand, M. Adam, Y. Motorin, M. Helm, A. Marx, *Angewandte Chemie International Edition* **2018**, *57*, 417-421.
[44] M. Sakurai, T. Yano, H. Kawabata, H. Ueda, T. Suzuki, *Nature chemical biology* **2010**, *6*, 733-740.
[45] T. Suzuki, H. Ueda, S. Okada, M. Sakurai, *Nature protocols* **2015**, *10*, 715-732.
[46] aT. Rein, M. L. DePamphilis, H. Zorbas, *Nucleic acids research* **1998**, *26*, 2255-2264; bH. Thomassin, E. J. Oakeley, T. Grange, *Methods* **1999**, *19*, 465-475.
[47] M. Schaefer, T. Pollex, K. Hanna, F. Lyko, *Nucleic acids research* **2009**, *37*, e12-e12.
[48] A. Bakin, J. Ofengand, *Biochemistry* **1993**, *32*, 9754-9762.
[49] X. Li, P. Zhu, S. Ma, J. Song, J. Bai, F. Sun, C. Yi, *Nature chemical biology* **2015**, *11*, 592-597.
[50] aN. Krogh, U. Birkedal, H. Nielsen, in *RNA Methylation*, Springer, **2017**, pp. 189-209; bQ. Dai, S. Moshitch-Moshkovitz, D. Han, N. Kol, N. Amariglio, G. Rechavi, D. Dominissini, C. He, *Nature methods* **2017**, *14*, 695-698.
[51] K. D. Meyer, *Nature methods* **2019**, *16*, 1275-1280.
[52] D. Clyde, *Nature Reviews Genetics* **2020**, *21*, 387-387.
[53] aR. R. Pandey, R. S. Pillai, *Cell* **2019**, *178*, 515-517; bH.-X. Chen, Z. Zhang, D.-Z. Ma, L.-Q. Chen, G.-Z. Luo, *Methods* **2021**.
[54] S. D. Knutson, J. M. Heemstra, *Current protocols in chemical biology* **2020**, *12*, e82.
[55] K. D. Meyer, Y. Saletore, P. Zumbo, O. Elemento, C. E. Mason, S. R. Jaffrey, *Cell* **2012**, *149*, 1635-1646.
[56] B. Linder, A. V. Grozhik, A. O. Olarerin-George, C. Meydan, C. E. Mason, S. R. Jaffrey, *Nature methods* **2015**, *12*, 767-772.
[57] aK. Kruger, P. J. Grabowski, A. J. Zaug, J. Sands, D. E. Gottschling, T. R. Cech, *cell* **1982**, *31*, 147-157; bC. Guerrier-Takada, K. Gardiner, T. Marsh, N. Pace, S. Altman, *Cell* **1983**, *35*, 849-857.
[58] aA. D. Ellington, J. W. Szostak, *nature* **1990**, *346*, 818-822; bC. Tuerk, L. Gold, *science* **1990**, *249*, 505-510; cD. L. Robertson, G. F. Joyce, *Nature* **1990**, *344*, 467-468.
[59] G. F. Joyce, *Annual review of biochemistry* **2004**, *73*, 791-836.
[60] S. K. Silverman, *Wiley Encyclopedia of Chemical Biology: John Wiley & Sons, Inc* **2008**.
[61] R. R. Breaker, G. F. Joyce, *Chemistry & biology* **1994**, *1*, 223-229.
[62] aC. Höbartner, S. K. Silverman, *Biopolymers: Original Research on Biomolecules* **2007**, *87*, 279-292; bD. Morrison, M. Rothenbroker, Y. Li, *Small Methods* **2018**, *2*, 1700319; cC. Zhou, J. L. Avins, P. C. Klauser, B. M. Brandsen, Y. Lee, S. K. Silverman, *Journal of the American Chemical Society* **2016**, *138*, 2106-2109; dK. Liu, P. K. Lat, H.-Z. Yu, D. Sen, *Nucleic acids research* **2020**, *48*, 7356-7370.
[63] S. W. Santoro, G. F. Joyce, *Proceedings of the national academy of sciences* **1997**, *94*, 4262-4266.
[64] M. J. Cairns, A. King, L. Q. Sun, *Nucleic acids research* **2003**, *31*, 2883-2889.
[65] aK. Schlosser, J. Gu, L. Sule, Y. Li, *Nucleic acids research* **2008**, *36*, 1472-1481; bR. P. Cruz, J. B. Withers, Y. Li, *Chemistry & biology* **2004**, *11*, 57-67.
[66] aA. A. Fokina, D. A. Stetsenko, J.-C. François, *Expert Opinion on Biological Therapy* **2015**, *15*, 689-711; bS. Khan, B. Burciu, C. D. Filipe, Y. Li, K. Dellinger, T. F. Didar, *ACS nano* **2021**, *15*, 13943-13969.
[67] M. Buchhaupt, C. Peifer, K.-D. Entian, *Analytical biochemistry* **2007**, *361*, 102-108.

[68] M. Hengesbach, M. Meusburger, F. Lyko, M. Helm, *Rna* **2008**, *14*, 180-187.
[69] M. V. Sednev, V. Mykhailiuk, P. Choudhury, J. Halang, K. E. Sloan, M. T. Bohnsack, C. Höbartner, *Angewandte Chemie International Edition* **2018**, *57*, 15117-15121.
[70] aF. LuSTIG, T. Boren, C. Claesson, C. Simonsson, M. Barciszewska, U. Lagerkvist, *Proceedings of the National Academy of Sciences* **1993**, *90*, 3343-3347; bY. E. Chong, M. Guo, X.-L. Yang, B. Kuhle, M. Naganuma, S.-i. Sekine, S. Yokoyama, P. Schimmel, *Proceedings of the National Academy of Sciences* **2018**, *115*, 7527-7532.
[71] aD. Lambert, D. Leipply, D. E. Draper, *Journal of molecular biology* **2010**, *404*, 138-157; bA. Pyle, *JBIC Journal of Biological Inorganic Chemistry* **2002**, *7*, 679-690; cD. Kilburn, J. H. Roh, L. Guo, R. M. Briber, S. A. Woodson, *Journal of the American Chemical Society* **2010**, *132*, 8690-8696.
[72] Y. Chen, G. Varani, *eLS* **2010**.
[73] G. J. Quigley, A. Rich, *Science* **1976**, *194*, 796-806.
[74] A. Roth, R. R. Breaker, *Annual review of biochemistry* **2009**, *78*, 305.
[75] aR. T. Batey, S. D. Gilbert, R. K. Montange, *Nature* **2004**, *432*, 411-415; bM. Mandal, R. R. Breaker, *Nature structural & molecular biology* **2004**, *11*, 29-35; cK. F. Blount, J. X. Wang, J. Lim, N. Sudarsan, R. R. Breaker, *Nature chemical biology* **2007**, *3*, 44-49; dM. Mandal, M. Lee, J. E. Barrick, Z. Weinberg, G. M. Emilsson, W. L. Ruzzo, R. R. Breaker, *Science* **2004**, *306*, 275-279; eJ. E. Barrick, R. R. Breaker, *Genome biology* **2007**, *8*, 1-19; fY. Shi, G. Zhao, W. Kong, *Journal of Biological Chemistry* **2014**, *289*, 11353-11366; gC. E. Dann III, C. A. Wakeman, C. L. Sieling, S. C. Baker, I. Irnov, W. C. Winkler, *Cell* **2007**, *130*, 878-892.
[76] aW. C. Winkler, S. Cohen-Chalamish, R. R. Breaker, *Proceedings of the National Academy of Sciences* **2002**, *99*, 15908-15913; bS. Müller, *Chembiochem* **2003**, *4*, 817-819.
[77] aJ. M. Buzayan, W. L. Gerlach, G. Bruening, *Nature* **1986**, *323*, 349-353; bA. C. Forster, R. H. Symons, *Cell* **1987**, *49*, 211-220; cC. L. Peebles, P. Perlman, K. Mecklenburg, M. Petrillo, J. Tabor, K. Jarrell, H.-L. Cheng, *Cell* **1986**, *44*, 213-223.
[78] aH. F. Noller, V. Hoffarth, L. Zimniak, *Science* **1992**, *256*, 1416-1419; bP. Nissen, J. Hansen, N. Ban, P. B. Moore, T. A. Steitz, *Science* **2000**, *289*, 920-930.
[79] J. Kim, C. Cheong, P. B. Moore, *Nature* **1991**, *351*, 331-332.
[80] D. Bhattacharyya, G. Mirihana Arachchilage, S. Basu, *Frontiers in chemistry* **2016**, *4*, 38.
[81] A. Joachimi, A. Benz, J. S. Hartig, *Bioorganic & medicinal chemistry* **2009**, *17*, 6811-6815.
[82] aG. G. Jayaraj, S. Pandey, V. Scaria, S. Maiti, *RNA biology* **2012**, *9*, 81-89; bS. Kumari, A. Bugaut, J. L. Huppert, S. Balasubramanian, *Nature chemical biology* **2007**, *3*, 218-221; cJ. Christiansen, M. Kofod, F. C. Nielsen, *Nucleic acids research* **1994**, *22*, 5709-5716; dY. Xu, Y. Suzuki, K. Ito, M. Komiyama, *Proceedings of the National Academy of Sciences* **2010**, *107*, 14579-14584.
[83] aR. Shahid, A. Bugaut, S. Balasubramanian, *Biochemistry* **2010**, *49*, 8300-8306; bD. Gomez, A. Guedin, J.-L. Mergny, B. Salles, J.-F. Riou, M.-P. Teulade-Fichou, P. Calsou, *Nucleic acids research* **2010**, *38*, 7187-7198; cM. J. Morris, Y. Negishi, C. Pazsint, J. D. Schonhoft, S. Basu, *Journal of the American Chemical Society* **2010**, *132*, 17831-17839.
[84] aA. Decorsière, A. Cayrel, S. Vagner, S. Millevoi, *Genes & development* **2011**, *25*, 220-225; bM. Subramanian, F. Rage, R. Tabet, E. Flatter, J. L. Mandel, H. Moine, *EMBO reports* **2011**, *12*, 697-704; cH. Min, R. C. Chan, D. L. Black, *Genes & development* **1995**, *9*, 2659-2671.

[85] S. M. Lyons, D. Gudanis, S. M. Coyne, Z. Gdaniec, P. Ivanov, *Nature communications* **2017**, *8*, 1-11.

[86] aC. Wang, L. Zhao, S. Lu, *International journal of biological sciences* **2015**, *11*, 316; bG. Biffi, D. Tannahill, S. Balasubramanian, *Journal of the American Chemical Society* **2012**, *134*, 11974-11976.

[87] aJ.-L. Mergny, A.-T. Phan, L. Lacroix, *FEBS letters* **1998**, *435*, 74-78; bJ.-L. Mergny, L. Lacroix, M.-P. Teulade-Fichou, C. Hounsou, L. Guittat, M. Hoarau, P. B. Arimondo, J.-P. Vigneron, J.-M. Lehn, J.-F. Riou, *Proceedings of the National Academy of Sciences* **2001**, *98*, 3062-3067; cA. Y. Zhang, A. Bugaut, S. Balasubramanian, *Biochemistry* **2011**, *50*, 7251-7258.

[88] G. W. Collie, S. Sparapani, G. N. Parkinson, S. Neidle, *Journal of the American Chemical Society* **2011**, *133*, 2721-2728.

[89] M. Di Antonio, G. Biffi, A. Mariani, E. Raiber, R. Rodriguez, S. Balasubramanian, *Situ*, 11073-11078.

[90] aA. Renaud de la Faverie, A. Guedin, A. Bedrat, L. A. Yatsunyk, J.-L. Mergny, *Nucleic acids research* **2014**, *42*, e65-e65; bA. l. Laguerre, L. Stefan, M. Larrouy, D. Genest, J. Novotna, M. Pirrotta, D. Monchaud, *Journal of the American Chemical Society* **2014**, *136*, 12406-12414.

[91] S. Y. Yang, P. Lejault, S. Chevrier, R. Boidot, A. G. Robertson, J. M. Wong, D. Monchaud, *Nature communications* **2018**, *9*, 1-11.

[92] aE. J. Merino, K. A. Wilkinson, J. L. Coughlan, K. M. Weeks, *Journal of the American Chemical Society* **2005**, *127*, 4223-4231; bP. Tijerina, S. Mohr, R. Russell, *Nature protocols* **2007**, *2*, 2608-2623; cW. A. Ziehler, D. R. Engelke, *Current protocols in nucleic acid chemistry* **2000**, 6.1. 1-6.1. 21.

[93] aC. K. Kwok, G. Marsico, A. B. Sahakyan, V. S. Chambers, S. Balasubramanian, *Nature methods* **2016**, *13*, 841-844; bC. K. Kwok, A. B. Sahakyan, S. Balasubramanian, *Angewandte Chemie* **2016**, *128*, 9104-9107; cX. Weng, J. Gong, Y. Chen, T. Wu, F. Wang, S. Yang, Y. Yuan, G. Luo, K. Chen, L. Hu, *Nature chemical biology* **2020**, *16*, 489-492; dS. Y. Yang, D. Monchaud, J. M. Wong, *Nature Protocols* **2022**, 1-20.

[94] aR. W. Dirks, H. J. Tanke, *Biotechniques* **2006**, *40*, 489-496; bT. T. Weil, R. M. Parton, I. Davis, *Trends in cell biology* **2010**, *20*, 380-390; cB. A. Armitage, *Current opinion in chemical biology* **2011**, *15*, 806-812.

[95] S. Neubacher, S. Hennig, *Angewandte Chemie International Edition* **2019**, *58*, 1266-1279.

[96] aG. S. Filonov, J. D. Moon, N. Svensen, S. R. Jaffrey, *Journal of the American Chemical Society* **2014**, *136*, 16299-16308; bJ. S. Paige, K. Y. Wu, S. R. Jaffrey, *Science* **2011**, *333*, 642-646; cR. L. Strack, M. D. Disney, S. R. Jaffrey, *Nature methods* **2013**, *10*, 1219-1224; dC. Höbartner, C. Steinmetzger, N. Palanisamy, K. R. Gore, *Chemistry-A European Journal* **2018**.

[97] A. K. Sharma, J. J. Plant, A. E. Rangel, K. N. Meek, A. J. Anamisis, J. Hollien, J. M. Heemstra, *ACS chemical biology* **2014**, *9*, 1680-1684.

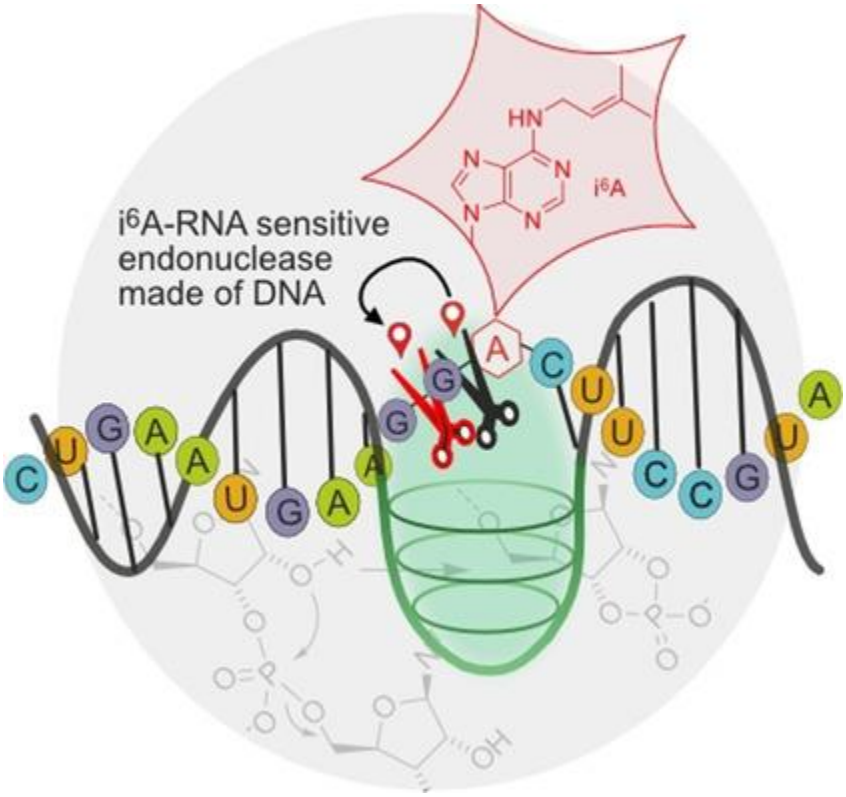
[98] R. I. McDonald, J. P. Guilinger, S. Mukherji, E. A. Curtis, W. I. Lee, D. R. Liu, *Nature chemical biology* **2014**, *10*, 1049-1054.

[99] R. Welz, K. Bossmann, C. Klug, C. Schmidt, H. J. Fritz, S. Müller, *Angewandte Chemie International Edition* **2003**, *42*, 2424-2427.

[100] M. Ghaem Maghami, C. P. Scheitl, C. Höbartner, *Journal of the American Chemical Society* **2019**, *141*, 19546-19549.
[101] M. Ghaem Maghami, S. Dey, A. K. Lenz, C. Höbartner, *Angewandte Chemie International Edition* **2020**, *59*, 9335-9339.
[102] aF. DUCONGÉ, J.-J. TOULMÉ, *Rna* **1999**, *5*, 1605-1614; bC. Boiziau, E. Dausse, L. Yurchenko, J.-J. Toulmé, *Journal of Biological Chemistry* **1999**, *274*, 12730-12737; cL. Aldaz-Carroll, B. Tallet, E. Dausse, L. Yurchenko, J.-J. Toulmé, *Biochemistry* **2002**, *41*, 5883-5893.
[103] F. Darfeuille, S. Reigadas, J. B. Hansen, H. Orum, C. Di Primo, J.-J. Toulmé, *Biochemistry* **2006**, *45*, 12076-12082.
[104] aC. P. Scheitl, M. Ghaem Maghami, A.-K. Lenz, C. Höbartner, *Nature* **2020**, *587*, 663-667; bH. Jiang, Y. Gao, L. Zhang, D. Chen, J. Gan, A. I. Murchie, *Nature Catalysis* **2021**, *4*, 872-881.

3. Paper I- N_6-Isopentenyladenosine in RNA Determines the Cleavage Site of Endonuclease Deoxyribozymes

Reprinted with the permission from "Liaqat A, Stiller C, Michel M, Sednev MV, Höbartner C. N^6 -Isopentenyladenosine in RNA Determines the Cleavage Site of Endonuclease Deoxyribozymes. *Angew Chem Int Ed Engl*. **2020**;59(42):18627-18631. doi:10.1002/anie.202006218"



DNAzymes Hot Paper

How to cite: *Angew. Chem. Int. Ed.* **2020**, *59*, 18627–18631
International Edition: doi.org/10.1002/anie.202006218
German Edition: doi.org/10.1002/ange.202006218

N^6-Isopentenyladenosine in RNA Determines the Cleavage Site of Endonuclease Deoxyribozymes

*Anam Liaqat, Carina Stiller, Manuela Michel, Maksim V. Sednev, and Claudia Höbartner**

Abstract: *RNA-cleaving deoxyribozymes can serve as selective sensors and catalysts to examine the modification state of RNA. However, site-specific endonuclease deoxyribozymes that selectively cleave post-transcriptionally modified RNA are extremely rare and their specificity over unmodified RNA is low. We report that the native tRNA modification N^6-isopentenyladenosine (i^6A) strongly enhances the specificity and has the power to reconfigure the active site of an RNA-cleaving deoxyribozyme. Using in vitro selection, we identified a DNA enzyme that cleaves i^6A-modified RNA at least 2500-fold faster than unmodified RNA. Another deoxyribozyme shows unique and unprecedented behaviour by shifting its cleavage site in the presence of the i^6A RNA modification. Together with deoxyribozymes that are strongly inhibited by i^6A, these results highlight that post-transcriptional RNA modifications modulate the catalytic activity of DNA in various intricate ways.*

Introduction

Cellular RNAs contain a variety of chemically diverse post-transcriptional modifications with important cellular functions.[1] Besides sophisticated analytical methods that generate transcriptome-wide modification maps, simple tools are required to validate predictions and to examine the modification state of distinct sites.[2] Focusing on N^6-methyladenosine (m^6A) as one of the most abundant mRNA modifications, we have recently reported m^6A-sensitive RNA-cleaving (endonuclease) deoxyribozymes for the site-specific interrogation of m^6A modification levels.[3] Deoxyribozymes with endonuclease activity bind to the target RNA via Watson–Crick base-pairing and catalyse site-specific RNA cleavage by promoting the attack of a specific 2′-hydroxy group onto the adjacent phosphodiester linkage, resulting in the formation of 2′,3′-cyclic phosphate and 5′-hydroxy termini in the fragments of the target RNA.[4] Inhibition of this mechanisms by 2′-*O*-methylated nucleotides was used to analyse ribose methylation in rRNA.[5] The m^6A-sensitive DNAzymes were generally applicable to analyse the presence of m^6A in DGACH sequence motifs, as shown for lncRNAs and a set of C/D box snoRNAs,[3] and these analyses were recently extended to other RNAs.[6] The detailed mechanism, how DNA enzymes recognize m^6A and modulate the cleavage response is currently unknown. Best discrimination was observed by the VMC10 DNA enzyme, which was inhibited by m^6A and cleaved unmodified RNA up to 150-fold faster. In contrast, DNA enzymes that preferred cleavage of m^6A-RNA were less specific and reached only 5–10-fold faster cleavage rates for modified versus unmodified RNA. From these observations, we concluded that a modified nucleotide like m^6A can easily interfere with formation of a catalytically competent DNA structure, but the opposite direction is more difficult: it is challenging to evolve catalytic DNA that is strictly dependent on the presence of a small chemical modification in the RNA target. Along these lines, one may expect that larger RNA modifications could establish energetically favourable interactions, and result in deoxyribozymes that more effectively discriminate modified from unmodified RNA. To explore this hypothesis, we chose to evolve deoxyribozymes that selectively cleave N^6-isopentenyladenosine (i^6A)-modified RNA and compared their activity to m^6A-sensitive endonucleases.

N^6-isopentenyladenosine (i^6A) is a structural analogue of m^6A that contains the bulkier dimethylallyl group as N^6-substituent. The i^6A modification is conserved in certain tRNAs in bacteria and eukaryotes.[7] Located at position 37 next to the anticodon, i^6A is suggested to enhance translation efficiency and fidelity by stabilizing cognate codon-anticodon interactions,[8] and may facilitate RNA localization by associating with membranes.[9] In some cases, i^6A is further modified, for example by thiomethylation to ms^2i^6A in mitochondrial tRNAs[7] and potentially in other (nuclear) RNAs,[10] or by oxidation/hydroxylation to ms^2io^6A, which is found for example, in salmonella tRNAs.[7,11] Interestingly, the i^6A modification is necessary for the expression of human selenoproteins,[12] and it has recently been shown as an essential determinant for the installation of additional tRNA anticodon modifications, including m^3C at position 32, and Um at position 34.[13] Given the importance of i^6A in many natural contexts, i^6A-sensitive deoxyribozymes could become diagnostically useful tools for the analysis of tRNA modification states.

Here we report the discovery and characterization of three classes of RNA-cleaving deoxyribozymes with distinct responses to the natural tRNA modification i^6A (Figure 1). First, we found that the activity of three AA deoxyribozymes is abolished by i^6A. Second, the AB08 deoxyribozyme is

[*] A. Liaqat, C. Stiller, M. Michel, Dr. M. V. Sednev, Prof. Dr. C. Höbartner
Institute of Organic Chemistry, University of Würzburg
Am Hubland, 97074 Würzburg (Germany)
E-mail: claudia.hoebartner@uni-wuerzburg.de
Supporting information and the ORCID identification number(s) for the author(s) of this article can be found under: https://doi.org/10.1002/anie.202006218.

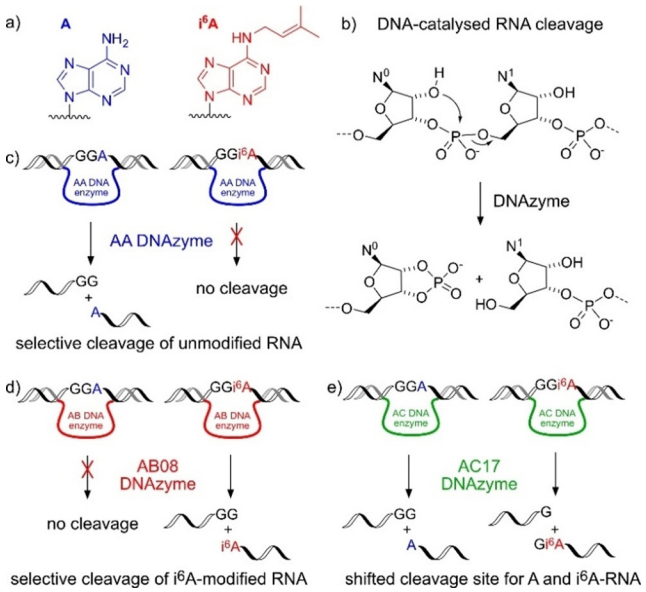


Figure 1. a) N^6-isopentenyladenosine (red) in comparison to unmodified adenosine (blue, only nucleobases shown). b) RNA-cleaving deoxyribozymes catalyse the cleavage of the phosphodiester backbone by transesterification. c–e) Depiction of three different classes of DNAzymes identified in this study; AA DNAzymes are inhibited by i^{6}A (c), the AB08 DNAzyme cleaves only i^{6}A-modified RNA (d). The cleavage site of the AC17 DNAzyme is influenced by the modification state of the RNA; that is, modified and unmodified RNA are cleaved at different positions (e).

strongly activated by i^{6}A and shows unprecedented selectivity in cleaving only i^{6}A-modified RNA next to the site of modification. Third, we found the deoxyribozyme AC17 for which the modified nucleotide i^{6}A causes a distinct shift of the cleavage site by one nucleotide.

Results and Discussion

In vitro selection of RNA-cleaving deoxyribozymes

A synthetic DNA library with 20 random nucleotides was used for the in vitro selection of RNA-cleaving deoxyribozymes, following established procedures (Supporting Information, Figure S1).[3] The 28-nt long RNA substrates were prepared by solid-phase synthesis (Supporting Information, Table S1), and covalently ligated to the DNA library (Table S2). The modified nucleotide i^{6}A was incorporated using a suitable i^{6}A phosphoramidite building block, which was obtained by regioselective N^6-alkylation of an N^6-acetylated intermediate (Supporting Information, Scheme S1).[14]

During 15 rounds of separation and amplification, catalytically active DNA sequences were trained to discriminate modified and unmodified RNA. To achieve the desired selectivity, two separate selection experiments were carried out, in which positive and negative selections were altered to favour unmodified RNA (AA selection) or i^{6}A-modified RNA (AB selection), respectively.

AA DNAzymes are inhibited by i^{6}A

After cloning of the enriched AA selection library, three DNA enzymes (AA07, AA14 and AA17) were identified that catalysed site-specific cleavage of unmodified RNA (R1) (Figure 2). RNA cleavage products were only observed with R1 as substrate (Figures 2b,c), but not with i^{6}A-containing RNA (R2). The cleavage sites were assigned by comparison to alkaline hydrolysis and RNase T1 digestion ladders and confirmed by high-resolution electrospray ionisation mass spectrometry (HR-ESI-MS) of both cleavage products (Figure 2d). AA17 cleaved R1 at position G16, while the other two DNA enzymes cleaved one nucleotide further upstream at G15. Interestingly, the i^{6}A at position 17 in the substrate RNA strongly inhibited the catalytic activity of all three DNA enzymes, independent of their cleavage site. Previously, m^{6}A was shown to inhibit the cleavage activity of DNAzymes when located directly next to the cleavage site (e.g. in VMC10),[3] but such a strong effect of a natural RNA modification at a more remote location, as is the case of i^{6}A on AA07 and AA14 activities, has not been observed before. The larger size of the isopentenyl group in i^{6}A compared to the methyl group in m^{6}A is likely responsible for this long-range effect. Indeed, when the AA DNAzymes were tested on m^{6}A-modified RNA

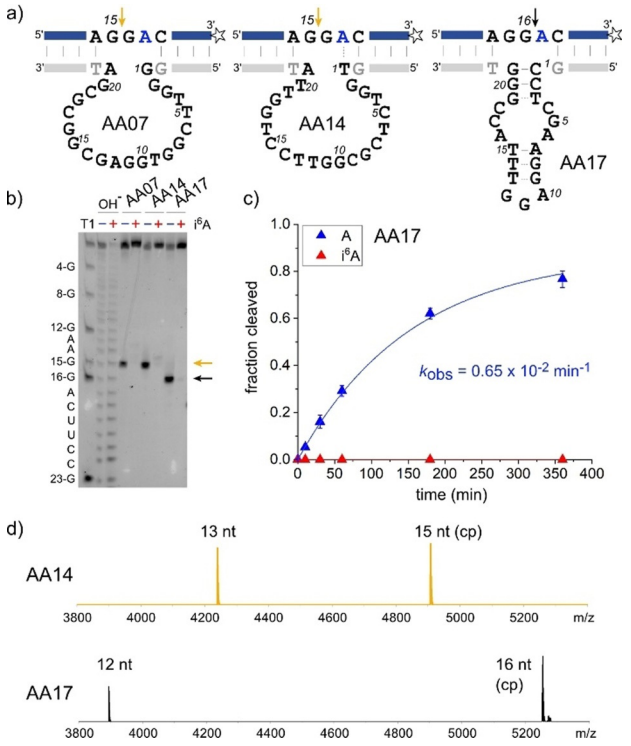


Figure 2. a) Sequences and predicted secondary structures of three AA DNAzymes, AA07, AA14, and AA17. b) PAGE analysis of the cleavage reaction with unmodified RNA (R1, labelled with blue −) and i^{6}A-modified RNA (R2, labelled with red +) shows selective cleavage of unmodified RNA, and identifies the cleavage site; AA07 and AA14 cleave at G15, and AA17 cleaves at G16. Experiments used 3′-fluorescein-labelled RNA. c) Single-turnover kinetics for AA17. Graphs for others and gel images are shown in Figure S3. d) HR-ESI-MS analyses of cleaved products. The longer 15/16-nt products contain a 2′,3′-cyclic phosphate, and the shorter 12/13-nt fragments contain 5′-OH and 3′-C_6-NH_2-termini.

(R3, same sequence as R2, with i^{6}A replaced by m^{6}A), AA17 was strongly inhibited by m^{6}A, while AA07 and AA14 retained substantial activity (Figure S3).

ABo8 DNAzyme cleaves only i^{6}A RNA

The AB selection revealed the deoxyribozyme AB08, which showed high activity for cleaving i^{6}A-RNA, but it did not cut unmodified RNA to any significant extent (Figure 3). This notable finding reports the first DNA enzyme that specifically cleaves a post-transcriptionally modified RNA. The AB08 DNA enzyme strictly required the modified nucleotide in the substrate RNA and yielded 80% cleaved product after 6 h at pH 7.5 with 20 mM $MgCl_2$. The cleavage site of AB08 was located directly next to the modified nucleotide i^{6}A, as confirmed by high-resolution mass spectrometry of the 12-nt 5′-OH-terminated i^{6}A-containing fragment (Figure 3c). These findings raised the question if AB08 could be activated by m^{6}A rather than i^{6}A in the RNA substrate. Using the analogous m^{6}A-modified RNA R3, no cleavage product was observed upon incubation with AB08 (Figure S4). Thus, the AB08 DNA enzyme is indeed specific for i^{6}A-modified RNA.

To learn more about the abundance of AB08 and potential related sequences, the enriched selection library was analysed by next generation sequencing (NGS). The round 15 library was separated into active and inactive fractions for cleavage of i^{6}A-RNA; this corresponded to one additional round of selection (round 16). Round 15 and both fractions of round 16 were amplified with the corresponding Illumina NGS primers (Figure S2), and 1.3 million reads were analysed (Table S3).[15] AB08 and its relatives with Levenshtein distance ≤ 4 represented 6.4% of reads in round 15. This fraction increased to 10.9% in round 16. Interestingly, the number of variants decreased from 73 to 54.[16] Comparison of the most enriched AB08 variants (Figure S5) and the sequence logo[17] plotted for the alignment of 54 relatives of AB08 shown in Figure 3d identified the nucleotides with highest variability at position 4 and at the 3′-end. The catalytic activities of several synthetic AB08 variants confirmed the conserved motif (Figure S6, Table S4), and may suggest a distant relationship of AB08 to the 8–17 DNA enzyme family, which has been identified in several in vitro selection experiments.[4c] The AGY triloop (position 10–12 in Figure 3d) and a conserved CG dinucleotide in the 3′-bulge (position 17–18 in the sequence logo) may be involved in a pseudoknot motif related to the one observed in the crystal structure of an 8–17 variant.[18] The highly invariable guanosine triple in the 5′-bulge is a distinctive feature of AB08 and is likely responsible for the astonishing selectivity for cleaving only i^{6}A RNA. In the absence of any structural insights into the organization of the catalytic core, the detailed mechanism of i^{6}A recognition remains unknown.

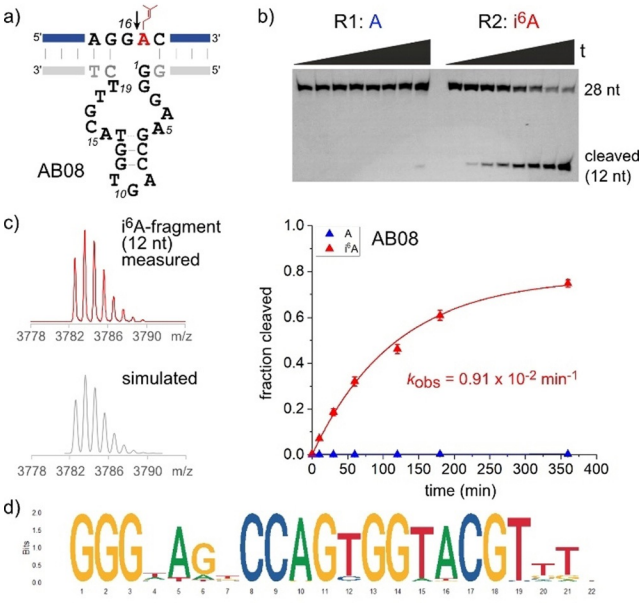


Figure 3. a) Secondary structure prediction of AB08 and indicated cleavage site of i^{6}A RNA. b) Gel image and kinetic graph demonstrating selectivity for i^{6}A-modified RNA. Time points are: 0 and 10 min; 0.5, 1, 2, 3, 6, and 22 h. c) HR-ESI-MS confirming the cleavage site of R2. Top: measured spectrum of 12-nt i^{6}A-modified fragment (i^{6}ACUUCCGUAACU). Bottom: simulated spectrum for $C_{117}H_{149}N_{40}O_{83}P_{11}$ with an exact mass of 3782.58 amu. d) Sequence logo for alignment of 54 sequences (within a Levenshtein distance of 4 from AB08), represented in the active fraction of the enriched DNA library.

i^{6}A shifts the cleavage site of AC17 DNAzyme

The AB selection surfaced an additional highly interesting DNA enzyme, which was named AC17.[19] Upon screening the catalytic activity, it was observed that AC17 yielded cleavage products with both i^{6}A-modified and unmodified RNAs (R1 and R2), and that these products exhibited different migration on the analysis gel (Figure 4). Comparison of the migration behaviour with the already assigned cleavage products of AA14 and AB08 revealed that AC17 cuts R1 at G16, but R2 at G15 (Figure 4b). Inspection of the sequence and secondary structure prediction of AC17 revealed that it is highly related to the previously identified VMC10 DNA enzyme, which showed strong inhibition by m^{6}A.[3] A single point mutation (C12T) converted AC17 into VMC10. Therefore, a more detailed characterization of the AC17 DNA enzyme in comparison to VMC10 was undertaken. First, AC17 was tested on m^{6}A-RNA and was found to yield only 10% cleaved RNA, and this cleavage occurred at G16 (Figure 4c, Figure S7). Thus, m^{6}A inhibited the ability of AC17 to cut at G16, but at the same time did not activate it to cut at G15 either. This confirmed that the unrivalled shift of the cleavage site is specific to i^{6}A. On the other hand, we also checked the activity of VMC10 to cut i^{6}A RNA and found about 25% cleavage at G15 (Figure 4d, Figure S7). Thus, both AC17 and VMC10 cut i^{6}A RNA at G15, but a cytidine at position 12 (C12) of the catalytic core enabled higher rate and yield compared to thymidine (T12).

Intrigued by the activity of AC17 to cut both RNAs, we analysed the enriched DNA library for AC17 abundance and potential relatives. The round 15 AB pool was separated into active and inactive fractions, now using unmodified RNA.

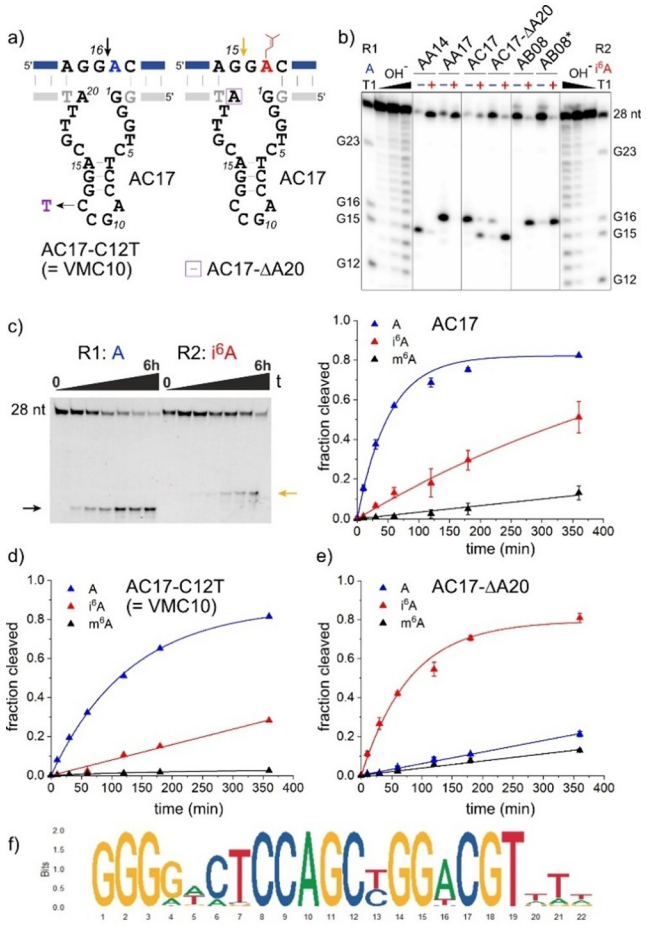


Figure 4. a) Predicted secondary structure of AC17 shown to cut unmodified RNA at G16 and i^{6}A-RNA at G15. Sites of important mutations are indicated. b) PAGE analysis of cleavage sites of AC17 in comparison to AA14, AA17 and AB08. Note that this analysis was performed with 5′-^{32}P-labelled RNA (in contrast to the 3′-fluorescein-labelled RNAs used in Figures 2b and 3b). c) PAGE analysis with 3′-fluorescein-labelled RNAs and kinetic graphs for cleavage of R1 (A), R2 (i^{6}A) and R3 (m^{6}A) RNA). d,e) kinetic graphs for cleavage of R1 (A), R2 (i^{6}A) and R3 (m^{6}A) RNA with VMC10 (d) and AC17-ΔA20 (e). f) Sequence logo for alignment of 45 AC17 variants with >50 reads and maximal 4 mutations.

This corresponded to a "negative" selection round 16. Both fractions were amplified with the corresponding NGS primers, and 1.4 million reads were analysed (Table S3). In accordance with the faster cleavage rate of R1, a larger number of AC17 variants (with Levenshtein distance ≤ 4) were found in the cleaved fraction (1177 variants versus 426 variants in the uncleaved fraction; the numbers were significantly reduced when filtered for more than 50 reads: 45 and 7 variants, respectively). The sequence logo for the alignment of 45 AC17 variants is shown in Figure 4f. The 3′-terminal nucleotides of the catalytic core showed highest variability. The importance of this region was investigated using synthetic variants of AC17. Deleting T19 resulted in a DNA enzyme that cleaved both R1 and R2 with comparable rates, and the ratio of the cleavage products can be directly translated into modification levels (Figure S8). Mutation at position 20 resulted in a series of DNA enzymes with very different behaviour: the activity of the A20G mutant was highly comparable to the parent deoxyribozyme, while the A20C variant showed strongly preferred cleavage of unmodified RNA with a 35-fold faster rate (Table S4, Figure S8). On the other hand, the deletion of A20 completely changed the activity of the deoxyribozyme.

Surprisingly, the AC17-ΔA20 variant showed highly accelerated cleavage of i^{6}A-RNA, while unmodified RNA R1 was cleaved only to a minor extent (Figure 4e, Figure S7). In other words, AC17-ΔA20 is the second i^{6}A-RNA-selective DNA enzyme discovered in this study, with comparable kinetic activity to AB08, but cleaving one nucleotide further upstream. Consistent with this result, the number of reads for AC17-ΔA20 (and its relatives with up to maximal four mutations) was almost twice as high in the active fraction of the i^{6}A NGS library compared to the uncleaved fraction (Table S2). Interestingly, AC17 contains conserved features reminiscent of the AGY/CG motif and the guanosine triple at the 5′-bulge, which were also found in AB08 (compare sequence logos in Figures 3d and 4f, and alignments in the Supporting Information). The relative location of these motifs seems preserved, but the position of the intervening GG dinucleotide involved in the formation of the base-paired region is shifted by one nucleotide. In other words, the AGY motif is located in a predicted triloop in AB08 and a tetraloop in AC17. It remains to be seen if these motifs are directly involved in the catalytic mechanism, and how the DNA nucleotides interact with the RNA bases near the cleavage site to distinctly modulate the activation of a specific 2′-OH group in response to the modification state of the adenosine under investigation. Future crystal structure analyses may also reveal the nucleotides contacting the cleavage site, enabling rational engineering in analogy to the structure-based design of RNA-ligating deoxyribozymes.[20]

Conclusion

In summary, this study reported three new classes of RNA-cleaving DNA enzymes that respond in distinct ways to the presence of the natural modified nucleotide i^{6}A in the target RNA. The AB08 DNA enzyme absolutely requires the presence of i^{6}A in the target RNA for catalytic activity. The finding that i^{6}A can finetune the active site of the DNA enzyme AC17 for cleavage at distinct positions is exceptional and may be further explored for analytical applications; for example, as a direct readout of chemically modulated adenosine isopentenylation states.[21] In the future, it will be interesting to investigate how other tRNA modifications, including hydroxylated or thiomethylated i^{6}A analogues are recognized and how they influence the catalytic activities of AB08, AC17 and future evolved variants for the study of native tRNAs.[22]

The presented data demonstrate the surprising plasticity of DNA's catalytic ability. The reported modes of action are distinct from previously observed responses of DNA enzymes to m^{6}A in RNA, and no other DNA catalysts sensitive to nucleobase modifications have been reported. The direct comparison of how i^{6}A and m^{6}A affect the catalytic activity

supports the conclusion that the larger and more hydrophobic isopentenyl group permits better discrimination due to larger energetic and kinetic differences. This observation correlates with the larger thermodynamically stabilizing effect of i^6A compared to m^6A in RNA hairpin structures.[23] Structural and mechanistic investigations will follow to allow deeper insights into the architecture of these new DNA catalysts and their modes of i^6A recognition.

Acknowledgements

This work was supported by the Deutsche Forschungsgemeinschaft (DFG, SPP1784 Chemical Biology of Native Nucleic Acid modifications) and by the European Research Council (ERC No. 682586). A.L. acknowledges funding by a PhD scholarship from the German Academic Exchange Service (DAAD). M.V.S. thanks the Graduate School for Life Sciences at the University of Würzburg for a Postdoc Plus fellowship. We thank Michaela Schraut and Tim Rieseberg for contributions to i^6A synthesis, and Sebastian Mayer for mass spectrometry. Illumina sequencing was performed at the Core Unit Systems Medicine at the University of Würzburg. Open access funding enabled and organized by Projekt DEAL.

Conflict of interest

The authors declare no conflict of interest.

Keywords: deoxyribozymes · epitranscriptomics · in vitro selection · RNA modification · site-specific RNA cleavage

[1] a) P. Boccaletto, M. A. Machnicka, E. Purta, P. Piatkowski, B. Baginski, T. K. Wirecki, V. de Crecy-Lagard, R. Ross, P. A. Limbach, A. Kotter, M. Helm, J. M. Bujnicki, *Nucleic Acids Res.* **2018**, *46*, D303–D307; b) P. J. McCown, A. Ruszkowska, C. N. Kunkler, K. Breger, J. P. Hulewicz, M. C. Wang, N. A. Springer, J. A. Brown, *Wiley Interdiscip. Rev. RNA* **2020**, https://doi.org/10.1002/wrna.1595.

[2] a) M. Helm, Y. Motorin, *Nat. Rev. Genet.* **2017**, *18*, 275; b) B. Linder, S. R. Jaffrey, *Cold Spring Harbor Perspect. Biol.* **2019**, *11*, a032201; c) K. Hartstock, A. Rentmeister, *Chem. Eur. J.* **2019**, *25*, 3455–3464.

[3] M. V. Sednev, V. Mykhailiuk, P. Choudhury, J. Halang, K. E. Sloan, M. T. Bohnsack, C. Höbartner, *Angew. Chem. Int. Ed.* **2018**, *57*, 15117–15121; *Angew. Chem.* **2018**, *130*, 15337–15341.

[4] a) R. R. Breaker, G. F. Joyce, *Chem. Biol.* **1994**, *1*, 223–229; b) S. K. Silverman, *Nucleic Acids Res.* **2005**, *33*, 6151–6163; c) K. Schlosser, Y. Li, *ChemBioChem* **2010**, *11*, 866–879.

[5] M. Buchhaupt, C. Peifer, K.-D. Entian, *Anal. Biochem.* **2007**, *361*, 102–108.

[6] M. Bujnowska, J. Zhang, Q. Dai, E. M. Heideman, J. Fei, *J. Biol. Chem.* **2020**, *295*, 6992–7000.

[7] U. Schweizer, S. Bohleber, N. Fradejas-Villar, *RNA Biol.* **2017**, *14*, 1197–1208.

[8] a) R. J. Maraia, J. R. Iben, *RNA* **2014**, *20*, 977–984; b) T. N. Lamichhane, A. G. Arimbasseri, K. Rijal, J. R. Iben, F. Y. Wei, K. Tomizawa, R. J. Maraia, *RNA* **2016**, *22*, 583–596; c) J. W. Yarham, T. N. Lamichhane, A. Pyle, S. Mattijssen, E. Baruffini, F. Bruni, C. Donnini, A. Vassilev, L. He, E. L. Blakely, *PLoS Genet.* **2014**, *10*, e1004424.

[9] T. Janas, T. Janas, M. Yarus, *RNA* **2012**, *18*, 2260–2268.

[10] a) V. Reiter, D. M. Matschkal, M. Wagner, D. Globisch, A. C. Kneuttinger, M. Müller, T. Carell, *Nucleic Acids Res.* **2012**, *40*, 6235–6240; b) M. Fakruddin, F. Y. Wei, S. Emura, S. Matsuda, T. Yasukawa, D. Kang, K. Tomizawa, *Nucleic Acids Res.* **2017**, *45*, 11954–11961.

[11] C. Mathevon, F. Pierrel, J. L. Oddou, R. Garcia-Serres, G. Blondin, J. M. Latour, S. Menage, S. Gambarelli, M. Fontecave, M. Atta, *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 13295–13300.

[12] G. J. Warner, M. J. Berry, M. E. Moustafa, B. A. Carlson, D. L. Hatfield, J. R. Faust, *J. Biol. Chem.* **2000**, *275*, 28110–28119.

[13] a) M. Zhou, T. Long, Z. P. Fang, X. L. Zhou, R. J. Liu, E. D. Wang, *RNA Biol.* **2015**, *12*, 900–911; b) A. G. Arimbasseri, J. Iben, F. Y. Wei, K. Rijal, K. Tomizawa, M. Hafner, R. J. Maraia, *RNA* **2016**, *22*, 1400–1410; c) L. K. Kim, T. Matsufuji, S. Matsufuji, B. A. Carlson, S. S. Kim, D. L. Hatfield, B. J. Lee, *RNA* **2000**, *6*, 1306–1315.

[14] V. I. Tararov, S. V. Kolyachkina, C. S. Alexeev, S. N. Mikhailov, *Synthesis* **2011**, 2483–2489.

[15] Analysis of the sequencing data was performed using fastaptamer perl scripts: K. K. Alam, J. L. Chang, D. H. Burke, *Mol. Ther. Nucleic Acids* **2015**, *4*, e230.

[16] Only variants with more than 50 reads were counted.

[17] O. Wagih, *Bioinformatics* **2017**, *33*, 3645–3647.

[18] a) H. Liu, X. Yu, Y. Chen, J. Zhang, B. Wu, L. Zheng, P. Haruehanroengra, R. Wang, S. Li, J. Lin, J. Li, J. Sheng, Z. Huang, J. Ma, J. Gan, *Nat. Commun.* **2017**, *8*, 2006; b) S. W. Santoro, G. F. Joyce, *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 4262–4266.

[19] This sequence was identified from clone number 17 of the AB selection (AB17), but because of its distinct properties and the sequence similarity to VMC10, it was renamed to AC17.

[20] A. Ponce-Salvatierra, K. Wawrzyniak-Turek, U. Steuerwald, C. Höbartner, V. Pena, *Nature* **2016**, *529*, 231–234.

[21] H. P. Cheng, X. H. Yang, L. Lan, L. J. Xie, C. Chen, C. Liu, J. Chu, Z. Y. Li, L. Liu, T. Q. Zhang, D. Q. Luo, L. Cheng, *Angew. Chem. Int. Ed.* **2020**, *59*, 10645–10650; *Angew. Chem.* **2020**, *132*, 10732–10737.

[22] AB08 and AC17 are not directly applicable for the examination of natural tRNAs, because of their preferred i^6A sequence context. Natural isopentenyl transferase enzymes preferably prenylate the central adenosine of a triple AAA motif,[24] while the DNA enzymes reported in this study require a guanine next to i^6A. The RNA substrate containing the Gi^6A motif was chosen to allow for direct comparison with m^6A-sensitive DNA enzymes. Analogous experiments with Ai^6AA-containing RNAs are expected to yield DNA enzymes in the future that can be used on native tRNA sequences.

[23] a) E. Kierzek, R. Kierzek, *Nucleic Acids Res.* **2003**, *31*, 4472–4480; b) A. P. Denmon, J. Wang, E. P. Nikonowicz, *J. Mol. Biol.* **2011**, *412*, 285–303.

[24] T. Soderberg, C. D. Poulter, *Biochemistry* **2000**, *39*, 6546–6553.

Manuscript received: April 29, 2020
Revised manuscript received: June 29, 2020
Accepted manuscript online: July 18, 2020
Version of record online: August 20, 2020

Angewandte Chemie
Eine Zeitschrift der Gesellschaft Deutscher Chemiker

Supporting Information

N^6-Isopentenyladenosine in RNA Determines the Cleavage Site of Endonuclease Deoxyribozymes

*Anam Liaqat, Carina Stiller, Manuela Michel, Maksim V. Sednev, and Claudia Höbartner**

anie_202006218_sm_miscellaneous_information.pdf

SUPPORTING INFORMATION

Table of Contents

Experimental Procedures

1. Materials and Methods

Unmodified DNA oligonucleotides were purchased from Microsynth and purified by denaturing PAGE (10-20% polyacrylamide) prior to use. Modified DNA (including primers) and RNA oligonucleotides were prepared in house by solid phase synthesis by using standard phosphoramidite chemistry. Solid supports for solid phase synthesis and unmodified DNA and 2'-*O*-TOM-protected RNA phosphoramidites were purchased from Sigma Proligo, ChemGenes, or Glen Research. Fluorescein-thiosemicarbazide and 6-carboxyfluorescein succinimidyl ester (NHS-fluorescein) were purchased from Sigma. 6-FAM-azide and deoxyribonucleotide triphosphate (dNTPs) were from Jena Bioscience. T4 Polynucleotide Kinase (PNK), T4 DNA Ligase and DreamTaq polymerase were purchased from Thermo Fisher Scientific.

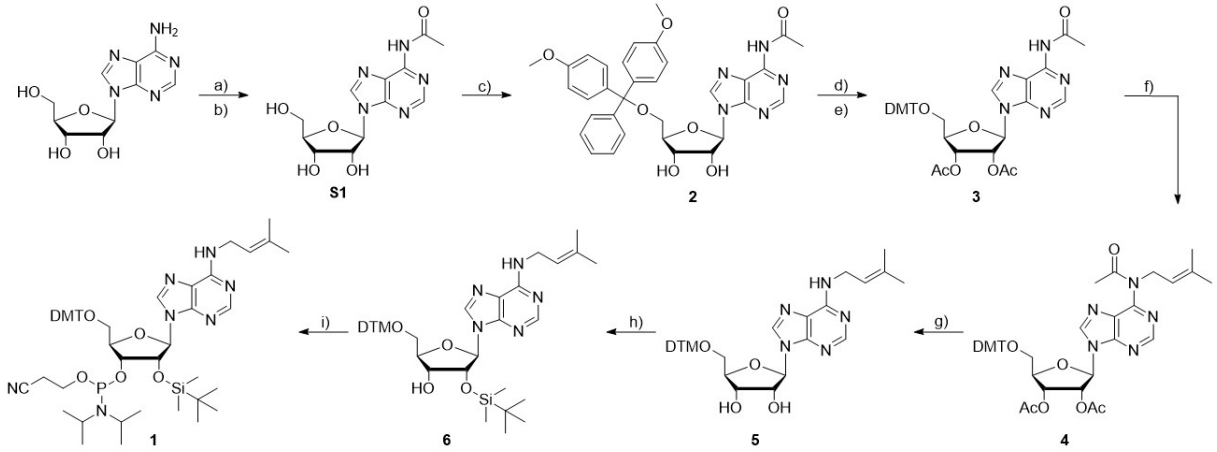
2. Synthetic procedures

General Information

All reaction were performed under nitrogen atmosphere. The obtained products were stored at 4 °C. The chemicals used for the reactions were purchased 'Pro analysis'- or 'For synthesis' quality and used without additional purification. Solvents for column chromatography were used in technical quality and distilled prior to use. Dry solvents (dichloromethane, acetonitrile, THF, DMF) were obtained from solvent purification system (SPS). Methanol was pre-dried over molecular sieves and further dried with magnesium turnings. Column chromatography was performed on silica gel (Kieselgel 60, Merck, 0.063 – 0.200 mm). TLC was performed on Alugram DIL G/UV254 (Machery-Nagel, UV visualization, 254 nm). NMR spectra were recorded on a Bruker Avance III HD 400 spectrometer. Chemical shifts were measured relative to the residual solvent signals as internal standards (in ppm; $CDCl_3$: ^1H = 7.26, ^{13}C = 77.16, DMSO: ^1H = 2.50, ^{13}C = 39.52). ESI-MS of the products and the oligonucleotides were recorded on a Brucker micrOTOF-Q III spectrometer.

SUPPORTING INFORMATION

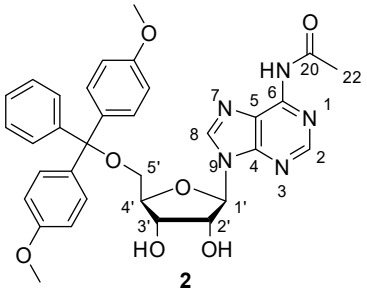
Synthesis of i^{6}A phosphoramidite



a) TMS-Cl, pyridine, 18 h, 0 °C - r.t. b) MeCN, AcCl, 1 h 0 °C - r.t. c) DMT-Cl, pyridine, 2 h, r.t. d) Ac_2O, pyridine, 1.5 h, r.t. e) imidazole, MeOH, 2 h, r.t. f) K_2CO_3, isopentenyl bromide, DMF, 20 h, 50 °C g) $MeNH_2$ in EtOH, 16 h, r.t. h) TBDMS-Cl, $AgNO_3$, THF/pyridine, 19 h, r.t. i) Me_2NEt, CEP-Cl, 2.5 h, r.t.

The synthesis of i^{6}A-phosphoramidite building block **1** started with acetyl protection of the *N*^{6} amino group of adenosine, followed by protection of the 5'-OH group with 4,4'-dimethoxytrityl chloride to produce compound **2**, following published procedures.[1] Compound **2** was treated with Ac_2O to give *N*^{6}-acetyl-2',3'-*O*-diacetyl-5'-*O*-(4,4'-dimethoxytrityl)adenosine (**3**).[2] The isopentenyl group was regioselectively introduced at *N*^{6} using isopentenyl bromide and K_2CO_3 to produce key intermediate **4**, following a procedure previously reported for *N*^{6}-alkylation of tetraacetyladenosine.[2] The acetyl groups where then removed under basic conditions to give **5**,[3] followed by treatment with TBDMS-Cl in the presence of silver nitrate.[4] The 2'-silylated compound **6** was then converted to the phosphoramidite **1**.

N^6-Acetyl-5'-*O*-(4,4'-dimethoxytrityl)adenosine (**2**)



Starting from adenosine, *N*^{6}-acetyl-5'-*O*-DMT-adenosine (**2**) was produced following the published procedure.[1] The analytical parameters coincided with the ones reported by Pitsch et al., but the assignments of H2 and H8 need to be revised, as described here.

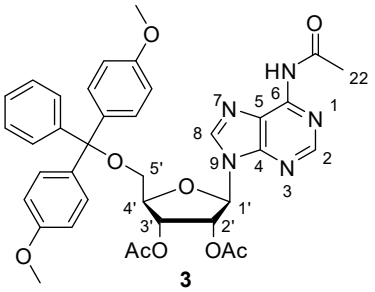
HR-MS (ESI$^+$): Exact mass calculated for $C_{33}H_{33}N_5NaO_7$ [M+Na]$^+$: 634.22722, found: 634.22888.

^{1}H NMR (400 MHz, $CDCl_3$) δ (ppm) = 8.63 (s, 1H, **2-H**), 8.22 (s, 1H, **8-H**), 7.30 – 7.21 (m, 2H, DMT), 7.20 – 7.14 (m, 7H, DMT), 6.77 – 6.69 (m, 4H, DMT), 6.03 (d, *J* = 5.7 Hz, 1H, 1'-H), 4.86 (dd, *J* = 5.7, 5.1 Hz, 1H, 2'-H), 4.46 (dd, *J* = 5.1, 2.3 Hz, 1H, 3'-H), 4.44 – 4.40 (m, 1H, 4'-H), 3.76 (d, *J* = 0.8 Hz, 6H, 2x CH_3), 3.44 (dd, *J* = 10.6, 3.5 Hz, 1H, 5'-H), 3.30 (dd, *J* = 10.6, 3.4 Hz, 1H, 5'-H), 2.61 (s, 3H, Acetyl-CH_3).

^{13}C NMR (101 MHz, $CDCl_3$) δ (ppm) = 170.38 (C_q, Acetyl), 158.68, 158.65 (C_q, C-OCH_3, DMT), 152.11 (C-2), 150.60 (C-4), 149.45 (C-6), 144.41 (C_q, DMT), 141.40 (C-8), 135.54, 135.53 (C_q, DMT), 130.11 (DMT), 130.06 (DMT), 128.10 (DMT), 127.99 (DMT), 127.08 (DMT), 122.14 (C-5), 113.27 (DMT), 90.73 (C-1'), 86.76 (C_q, C-O, DMT), 86.01 (C-4'), 75.88 (C-2'), 72.41 (C-3'), 63.55 (C-5'), 55.35 (2xCH_3, DMT), 25.83 (C-22, Acetyl-CH_3).

SUPPORTING INFORMATION

N^6-Acetyl-2',3'-*O*-diacetyl-5'-*O*-(4,4'-dimethoxytrityl)adenosin (**3**)



3

N^6-acetyl-5'-*O*-(4,4'-dimethoxytrityl)adenosine (**2**, 230 mg, 376 µmol, 1 eq) was dissolved in dry pyridine (1.5 mL), treated with Ac_2O (190 µL, 1.90 mmol, 5.05 eq) and stirred for 1.5 h at room temperature. After addition of EtOH (2.5 mL) the solvent was removed under reduced pressure. The reaction mixture was diluted with ethylacetate (20 mL) and washed with brine and the organic phase was dried over Na_2SO_4. After evaporation of the solvent under reduced pressure, the product (**3**, 208 mg, 299 µmol, 79%) was obtained as a colorless foam.

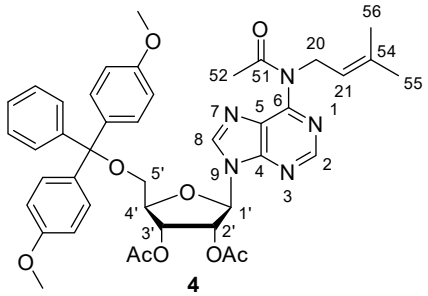
Chemical formula: $C_{37}H_{37}N_5O_9$, Molecular weight: 695.73 g·mol^{-1}

HR-MS (ESI$^+$): Exact mass calculated for $C_{37}H_{37}N_5NaO_9$ [M+Na]$^+$: 718.24835, found: 718.24876.

^{1}H NMR (400 MHz, $CDCl_3$) *δ* (ppm) = 8.64 (s, 1H, 2-H), 8.13 (s, 1H, 8-H), 7.43 – 7.39 (m, 2H, DMT), 7.34 – 7.23 (m, 6H, DMT), 7.24 – 7.19 (m, 1H, DMT), 6.84 – 6.75 (m, 4H, DMT), 6.32 (d, *J* = 6.6 Hz, 1H, 1'-H), 6.11 (dd, *J* = 6.6, 5.3 Hz, 1H, 2'-H), 5.69 (dd, *J* = 5.3, 3.0 Hz, 1H, 3'-H), 4.36 (q, *J* = 3.2 Hz, 1H, 4'-H), 3.78 (t, *J* = 0.7 Hz, 6H, 2 x CH_3), 3.51 (dd, *J* = 10.7, 3.1 Hz, 1H, 5'-H), 3.46 (dd, *J* = 10.7, 3.5 Hz, 1H, 5'-H), 2.59 (d, *J* = 1.5 Hz, 3H, Acetyl-CH_3), 2.13 (s, 3H, Acetly-CH_3), 2.05 (s, 3H, Acetyl-CH_3).

^{13}C NMR (101 MHz, $CDCl_3$) *δ* (ppm) = 170.28 (C_q, Acetyl), 169.87 (C_q, Acetyl, 169.54 (C_q, Acetyl), 158.81 (C_q, C-OCH_3, DMT), 152.83 (C-2), 151.43 (C-4), 149.36 (C-6), 144.19 (C_q, DMT), 141.08 (C-8), 135.34, 135.30 (C_q, DMT), 130.29, 130.27 (DMT), 128.34 (DMT), 128.16 (DMT), 127.27 (DMT), 121.97 (C-5), 113.42 (DMT), 87.27 (C_q, C-O, DMT), 85.16 (C-1'), 82.79 (C-4'), 73.41 (C-2'), 71.83 (C-3'), 63.12 (C-5'), 55.39 (2x CH_3, DMT), 25.81 (C-22, CH_3, Acetyl), 20.82 (CH_3, Acetyl), 20.56 (CH_3, Acetyl).

N^6-Acetyl-2',3'-*O*-diacetyl-5'-*O*-(4,4'-dimethoxytrityl)-*N*^{6}-isopentenyladenosine (**4**)



4

N^6-Acetyl-2',3'-*O*-diacetyl-5'-*O*-(4,4'-dimethoxytrityl)adenosine (**3**, 1.0 g, 1.44 mmol, 1 eq) in dry DMF (15 mL) was treated with K_2CO_3 (1.39 g, 10.1 mmol, 7 eq) and isopentenyl bromide (1.29 g, 996 µL, 8.62 mmol, 6 eq) and stirred for 20 h at 50 °C. The solution was allowed to cool down to room temperature, diluted with ethyl acetate (42 mL) and washed with brine. The organic phase was dried over Na_2SO_4 and the solvent was removed under reduced pressure. The residual was purified by column chromatography (DCM + 1% NEt_3 → DCM:EtOH = 99:1 + 1% NEt_3) to yield the product (**4**, 590 mg, 54%) as a slightly yellow foam. A significant fraction of unreacted starting material was recovered (**3**, 225 mg, 22%).

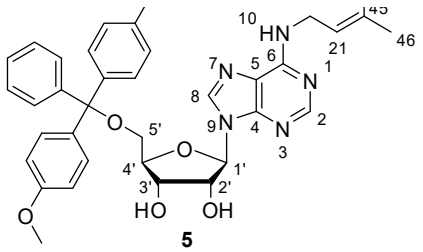
Chemical formula: $C_{42}H_{45}N_5O_9$, Molecular weight: 763.9 g·mol^{-1}

HR-MS (ESI$^+$): Exact mass calculated for $C_{42}H_{45}N_5NaO_9$ [M+Na]$^+$: 786.31095, found: 786.31438.

^{1}H NMR (400 MHz, $CDCl_3$) *δ* (ppm) = 8.72 (s, 1H, 2-H), 8.20 (s, 1H, 8-H), 7.43 – 7.39 (m, 2H, DMT), 7.33 – 7.29 (m, 4H, DMT), 7.26 (s, 2H, DMT), 7.24 – 7.18 (m, 1H, DMT), 6.84 – 6.78 (m, 4H, DMT), 6.35 (d, *J* = 6.5 Hz, 1H, 1'-H), 6.11 (dd, *J* = 6.4, 5.3 Hz, 1H, 2'-H), 5.71 (dd, *J* = 5.3, 3.2 Hz, 1H, 3'-H), 5.23 (ddt, *J* = 6.6, 5.3, 1.4 Hz, 1H, 21-H), 4.84 (d, *J* = 6.7 Hz, 2H, CH_2, 20-H), 4.37 (q, *J* = 3.3 Hz, 1H, 4'-H), 3.78 (d, *J* = 1.6 Hz, 6H, 2xCH_3, DMT), 3.52 (dd, *J* = 10.7, 3.1 Hz, 1H, 5'-H), 3.46 (dd, *J* = 10.8, 3.6 Hz, 1H, 5'-H), 2.27 (s, 3H, CH_3, H-52), 2.14 (s, 3H, CH_3, Acetyl), 2.07 (s, 3H, CH_3, Acetyl), 1.60 (d, *J* = 1.3 Hz, 3H), 1.59 (d, *J* = 1.3 Hz, 3H, H-55, H-56).

^{13}C NMR (101 MHz, $CDCl_3$) *δ* (ppm) = 171.41 (C_q C-51), 169.84 (C_q C-46), 169.52 (C_q C-45), 158.80 (C_q, C-OCH_3, DMT), 153.85 (C-6), 153.08 (C-4), 152.35 (C-2), 144.20 (C_q, DMT), 141.89 (C-8), 135.96 (C_q, C-54), 135.37, 135.33 (C_q, DMT), 130.26 (DMT), 128.30 (DMT), 128.13 (DMT), 127.40 (C-5), 127.23 (DMT), 120.28 (C-21), 113.41 (DMT), 87.23 (C-22), 85.37 (C-1'), 82.65 (C-4'), 73.30 (C-2'), 71.68 (C-3'), 63.03 (C-5'), 55.37 (2xCH_3, DMT), 45.79 (C-20), 25.78 (C-55, CH_3), 24.46 (C-52, CH_3), 20.81, 20.59 (C-48, C-49, 2xCH_3, Acetyl), 18.07 (C-56, CH_3).

N^6-Isopentenyl-5'-*O*-(4,4'-dimethoxytrityl)adenosine (**5**)



5

N^6-Acetyl-2',3'-*O*-diacetyl-5'-*O*-(4,4'-dimethoxytrityl)-*N*^{6}-isopentenyladenosine (**4**, 500 mg, 655 µmol, 1 eq) in MeOH (6 mL) was treated with methylamine in EtOH (33%, 5.7 mL, 45.8 mmol, 70 eq) and stirred at room temperature overnight. After evaporation of the solvent, the residue was purified by column chromatography (DCM + 1% NEt_3 → DCM:EtOH = 99:1 + 1% NEt_3 → DCM:EtOH = 97:3 + 1% NEt_3) to yield the product (293 mg, 459 µmol, 70%) as a colorless foam.

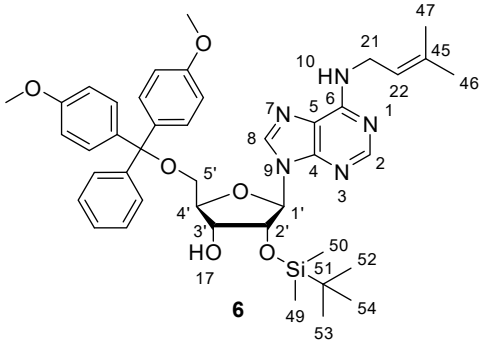
Chemical formula: $C_{36}H_{39}N_5O_6$, Molecular weight: 637.7 g·mol^{-1}

HR-MS (ESI$^+$): Exact mass calculated for $C_{36}H_{40}N_5O_6$ $[M+H]^+$: 638.29731, found: 638.29719

^{1}H NMR (400 MHz, $CDCl_3$) *δ* (ppm) = 8.34 (s, 1H, 2-H), 8.02 (s, 1H, 8-H), 7.28 – 7.24 (m, 2H, DMT), 7.20 – 7.14 (m, 7H, DMT), 6.76 – 6.70 (m, 4H, DMT), 5.94 (d, *J* = 6.0 Hz, 1H, 1'-H), 5.79 (s, 1H, NH), 5.41 – 5.34 (m, 1H, 21-H), 4.76 (dd, *J* = 6.1, 5.1 Hz, 1H, 2'-H), 4.42 (td, *J* = 3.3, 1.8 Hz, 1H, 4'-H), 4.37 (dd, *J* = 5.1, 1.9 Hz, 1H, 3'-H), 4.22 (s, 2H, CH_2, 20-H), 3.76 (d, *J* = 1.4 Hz, 6H, 2xCH_3, DMT), 3.43 (dd, *J* = 10.5, 3.5 Hz, 1H, 5'-H), 3.22 (dd, *J* = 10.5, 3.4 Hz, 1H, 5'-H), 1.74 (dd, *J* = 4.0, 1.3 Hz, 6H, 2xCH3,46-H, 47-H).

^{13}C NMR (101 MHz, $CDCl_3$) *δ* (ppm) = 158.62 (C_q, C-OCH_3), 154.87 (C-6), 152.70 (C-2), 147.89 (C-4), 144.40 (C_q, DMT), 137.92 (C-8), 137.36 (C-45), 135.70, 135.50 (C_q, DMT), 130.09 (DMT), 130.04 (DMT), 128.11 (DMT), 127.96 (DMT), 127.00 (DMT), 120.29 (C-5), 119.96 (C-22), 113.24 (DMT), 91.23 (C-1'), 86.71, 86.61 (C-4', C_q, C-O), 76.48 (C-2'), 73.15 (C-3'), 63.73 (C-5'), 55.32 (CH_3, DMT), 38.77 (C-21), 25.83 (CH_3, C-46), 18.17 (CH_3, C-47).

N^6-Isopentenyl-5'-*O*-(4,4'-dimethoxytrityl)-2'-*O*-(tert-butyldimethylsilyl)adenosine (**6**)



6

To *N*^{6}-isopentenyl-5'-*O*-(4,4'-dimethoxytrityl)adenosine (**5**, 335 mg, 525 µmol, 1 eq) in dry THF/pyridine (1:1, 4 mL) $AgNO_3$ (134 mg, 788 µmol, 1.5 eq) and TBDMS-Cl (87.1 mg, 578 µmol, 1.1 eq) were added and stirred for 19 h at room temperature. The reaction was stopped with MeOH (4 mL) and filtered over celite. The filtrate was evaporated to dryness under reduced pressure. The mixture of 2' and 3' silylated isomers was repeatedly separated by column chromatography (hexane:ethylacetate = 4:1 + 1% NEt_3 → 3:1 → 2:1) to yield the pure isomer compound **6** (182 mg, 242 µmol, 46%) as a colorless foam.

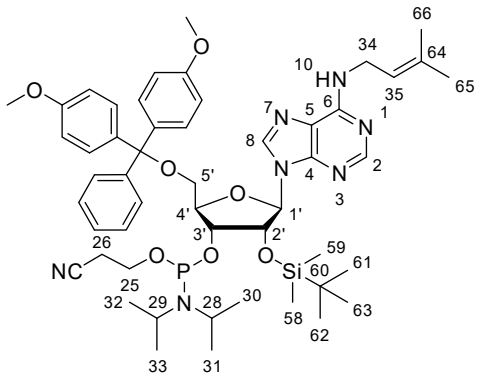
Chemical formula: $C_{42}H_{53}N_5O_6Si$, Molecular weight: 752.0 g·mol^{-1}

HR-MS (ESI$^+$): Exact mass calculated for $C_{42}H_{54}N_5O_6Si$ $[M+H]^+$: 752.38379, found: 752.38394

^{1}H NMR (400 MHz, $CDCl_3$) *δ* (ppm) = 8.31 (s, 1H, 2-H), 7.94 (s, 1H, 8-H), 7.47 – 7.43 (m, 2H, DMT), 7.36 – 7.31 (m, 4H, DMT), 7.27 (s, 2H, DMT), 7.23 – 7.18 (m, 1H, DMT), 6.84 – 6.77 (m, 4H, DMT), 6.01 (d, *J* = 5.5 Hz, 1H, 1'-H), 5.61 (s, 1H, NH), 5.38 (tdt, *J* = 7.1, 2.9, 1.4 Hz, 1H, 22-H), 4.99 (t, *J* = 5.3 Hz, 1H, 2'-H), 4.33 (dt, *J* = 5.2, 3.7 Hz, 1H, 3'-H), 4.24 (q, *J* = 3.5 Hz, 1H, 4'-H), 4.22 (s, 2H, CH_2, 21-H), 3.78 (d, *J* = 1.2 Hz, 6H, 2xCH_3, DMT), 3.51 (dd, *J* = 10.6, 3.3 Hz, 1H, 5'-H), 3.37 (dd, *J* = 10.6, 4.0 Hz, 1H, 5'-H), 2.72 (d, *J* = 4.0 Hz, 1H, OH), 1.78 – 1.73 (m, 6H, 2xCH3, 46-H, 47-H), 1.26 (t, *J* = 7.1 Hz, 1H), 0.84 (s, 9H, 3xCH_3, 52-H, 53-H, 54-H), -0.02 (s, 3H, CH_3, H-49), -0.14 (s, 3H, CH_3, 50-H).

^{13}C NMR (101 MHz, $CDCl_3$) *δ* (ppm) = 158.66 (C_q, C-OCH_3), 154.73 (C-6), 153.43 (C-2), 149.28 (C-4), 144.75 (C_q, DMT), 138.56 (C-8), 137.13 (C_q, C-45), 135.83 (C_q, DMT), 130.23 (DMT), 128.30 (DMT), 128.03 (DMT), 127.05 (DMT), 120.32 (C-5), 120.21 (C-22), 113.32 (DMT), 88.23 (C-1'), 86.71 (C_q, DMT), 84.18 (C-4'), 75.73 (C-2'), 71.72 (C-3'), 63.58 (C-5'), 55.36 (2xCH_3, DMT), 38.85 (C-21), 25.85 (C-46), 25.73 (C-52, C-53, C-54), 18.17 (C-47), -4.82, -5.11 (C-49, C-50).

N^6-Isopentenyl-5'-*O*-(4,4'-dimethoxytrityl)-2'-*O*-(*tert*-butyldimethylsilyl)adenosine 3' cyanoethyl *N*,*N*-diisopropyl phosphoramidite (**1**)



1

N^6-Isopentenyl-5'-*O*-DMT-2'-*O*-TBDMS-adenosine (**6**, 70 mg, 93.1 µmol, 1 eq) was dissolved in dry DCM (1 mL) and Me_2NEt (101 µL, 931 µmol, 10 eq) and CEP-Cl (33.0 mg, 140 µmol, 1.5 eq) were added. After stirring for 2.5 h at room temperature, the solvent was removed under reduced pressure. The residue was purified by column chromatography (ethylacetate:hexane = 2:1 + 1% NEt_3) to yield compound **1** (79 mg, 83.0 µmol, 89%).

Chemical formula: $C_{51}H_{70}N_7O_7PSi$, Molecular weight: 952.22 g·mol^{-1}

HR-MS (ESI$^+$): Exact mass calculated for $C_{51}H_{71}N_7O_7PSi$ $[M+H]^+$: 952.49164, found: 952.49298

^{1}H NMR (400 MHz, $CDCl_3$) δ (ppm) = 8.29 (s, 1H, 2-H), 8.27 (s, 1H, 2-H diast.), 7.97 (s, 1H, 8-H), 7.94 (s, 1H, 8-H diast.), 7.49 – 7.42 (m, 4H, DMT), 7.38 – 7.31 (m, 7H, DMT), 7.31 – 7.21 (m, 6H, DMT), 7.24 – 7.17 (m, 2H, DMT), 6.84 – 6.77 (m, 8H, DMT), 6.00 (d, *J* = 6.5 Hz, 1H, 1'-H), 5.96 (d, *J* = 6.1 Hz, 1H, 1'-H diast.), 5.58 (s, 2H, NH), 5.38 (dddd, *J* = 6.8, 5.4, 3.2, 1.5 Hz, 2H, 35-H), 5.07 (ddd, *J* = 8.6, 6.3, 4.6 Hz, 2H, 2'-H), 4.39 (m, 2H, 3'-H, 4'-H), 4.32 (q, *J* = 4.0 Hz, 1H, 4'-H), 4.21 (s, 4H, CH_2, 34-H), 4.01 – 3.92 (m, 1H, CH_2, 25-H), 3.88 (ddt, *J* = 10.3, 8.3, 6.7 Hz, 1H, CH_2, 25-H), 3.78 (d, *J* = 1.3 Hz, 7H, 2xCH_3, DMT), 3.78 (d, *J* = 0.6 Hz, 6H, 2xCH_3, DMT diast.), 3.71 – 3.62 (m, 1H, CH_2, 25-H diast.), 3.59 (dq, *J* = 10.2, 3.2 Hz, 4H, 25-H, 5'-H, 28-H, 29-H), 3.54 (dd, *J* = 10.7, 4.0 Hz, 1H, 5'-H diast.), 3.32 (dd, *J* = 9.8, 3.4 Hz, 1H, 5'-H), 3.28 (dd, *J* = 9.9, 3.4 Hz, 1H, 5'-H), 2.65 (td, *J* = 6.5, 4.4 Hz, 2H, CH_2, 26-H), 2.30 (td, *J* = 6.7, 1.9 Hz, 1H, CH_2, 26-H diast.), 1.76 (d, *J* = 1.3 Hz, 6H, 2xCH_3, 65-H, 66-H), 1.74 (d, *J* = 1.3 Hz, 5H, 2xCH_3, 65-H, 66-H diast.), 1.20 – 1.02 (m, 21H, 4xCH_3, 30-H, 31-H, 32-H, 33-H), 0.77 (s, 8H, 3xCH_3, 61-H, 62-H, 63-H), 0.76 (s, 8H, 3xCH_3, 61-H, 62-H, 63-H diast.), -0.03 (s, 2H, CH_3, 58-H), -0.06 (s, 3H, CH_3, 58-H diast.), -0.19 (s, 2H, CH_3, 59-H), -0.21 (s, 3H, CH_3, 59-H diast.).

^{13}C NMR (101 MHz, $CDCl_3$) δ (ppm) = 158.64 (C_q-OMe, DMT), 154.72 (C-6), 153.30 (C-2), 149.20 (C-4), 144.79, 144.69 (C_q, DMT), 139.02 (C-8), 137.10 (C.64), 135.97, 135.93, 135.80, 135.76 (C_q, DMT), 130.32, 130.27, 130.23 (C-DMT), 128.41, 128.29 (C-DMT), 128.02, 128.00 (C-DMT), 127.03 (C-DMT), 120.47 (C-5), 120.22 (C-35), 117.80, 117.46 (CN), 113.30, 113.26 (C-DMT), 88.40, 88.13 (C-1'), 86.77, 86.63 (C_q-O, DMT), 84.06, 83.81 (C-4'), 75.13, 75.11, 74.59, 74.54 (C-2'), 73.56, 73.46, 73.01, 72.86 (C-3'), 63.55, 63.38 (C-5'), 59.08, 58.91, 57.93, 57.72 (C-25), 55.37, 55.35 (O-CH_3, DMT), 43.58, 43.45, 43.12, 43.00 (C-28, C-29), 38.76 (C-34), 25.84, 25.80, 25.76 (C-61, C-62, C-63), 24.92, 24.85, 24.78, 24.67 (C-30, C-31, C-32, C-33), 20.62, 20.56, 20.22, 20.15 (C-26), 18.16, 18.10, 18.04 (C-65, C-66), -4.58, -4.60, -5.00 (C58, C-59).

^{31}P NMR (162 MHz, $CDCl_3$) δ (ppm) = 150.81, 148.84.

3. Synthesis and Labeling of RNA

The following solutions were used for automated solid-phase synthesis: 100 mM solutions in dry acetonitrile of the phosphoramidites, 0.25 M ethylthiotetrazole (ETT, activator) in dry acetonitrile, 3% dichloro acetic acid in 1,2-dichloroethane (detritylation), 0.5 M DMAP in acetonitrile (Cap A), acetic anhydride/sym-collidine/acetonitrile = 20/30/50 (Cap B), 10 mM I_2 in acetonitrile/sym-collidine/H_2O = 10/1/5 (oxidation). Oligonucleotide synthesis was carried out at 0.6 µmol scale. Forward primer for PCR was coupled with hexynyl group at 5' end and reverse primer was linked with a 12-nt tail via a non-extendable hexaethyleneglycol spacer. The coupling efficiencies were determined using absorbance of released dimethoxytrityl (DMT) cation and exceeded 98% for all syntheses. DNA oligonucleotides were deprotected using NH_4OH and $MeNH_2$ at 55°C for 5 hours. For RNA oligonucleotides deprotection was performed in two steps using $MeNH_2$ in aqueous ethanol, followed by 1 M TBAF in THF. 3'-Amino modified RNA was synthesized on amino-on-solid support and deprotected using NH_4OH and $MeNH_2$ in aqueous solution for 6 hours at 55°C, followed by TBAF in THF. Deprotected oligonucleotide samples were first desalted by size exclusion chromatography on 3×5 ml HiTrap columns (GE Healthcare), then purified by denaturing PAGE and analyzed via anion exchange HPLC and ESI-MS.

Anion exchange chromatography was performed on Dionex DNAPAc PA200 column, 2x250 mm, with 25 mM Tris.HCl pH 8.0, 6 M urea (buffer A) and 0.5 M $NaClO_4$ in 25 mM Tris.HCl pH 8.0, 6 M urea (buffer B), gradient: 0-48% B in 12 column volumes, flow rate 0.5 mL/min, 60°C. UV-detection at 260 nm.

PAGE purification was performed on 320x160x0.7mm denaturing polyacrylamide gels (10-20% acrylamide/ bisacrylamide 19:1, 7 M urea) with 1× TBE (89 mM Tris, 89 mM boric acid, 2 mM EDTA, pH=8.3) and run at 35 W power for 2 hours. Oligonucleotides were

visualized using UV shadowing on TLC plate, excised from the gel, extracted by crush and soak into TEN buffer (10 mM Tris-HCl, pH =8.0, 1 mM EDTA, 300 mM NaCl) and recovered by precipitation with ethanol.

Labeling of 5'-alkyne functionalized oligonucleotides

Oligonucleotides with 5'-alkyne group were labeled with 6-FAM-azide using copper(I)-catalyzed alkyne-azide cycloaddition (CuAAC). Oligonucleotide (5 nmol in 5 µL of H_2O) was mixed with 3 µL of DMSO/tBuOH (3:1 v/v) and 0.5 µL of 6-FAM-azide (50 mM in DMSO/tBuOH 3:1). The solution of CuBr (100 mM) was freshly prepared in DMSO/tBuOH 3:1 and mixed with a solution of TBTA (100 mM) in a 1:2 ratio. From this solution, 1.5 µL was added to the reaction mixture and incubated at 37°C for 3 hours in the dark. Afterwards, the sample was ethanol precipitated to remove excess fluorophore and labeled oligonucleotides were purified by PAGE.

Labeling of 3'-amino functionalized oligonucleotides

Oligonucleotides with 3'-amino group were labeled with 6-carboxyfluorescein succinimidyl ester (NHS-fluorescein). Amino-modified RNA (2.5 nmol in 2.5 µL of H_2O) was dissolved in carbonate buffer (20 µL, 100 mM, pH = 9.0) and solution of NHS-fluorescein in DMF (2.5 µL, 30 mM) was added. The reaction mixture was incubated at 37°C for 2 hours in the dark. Labeled RNA was purified by denaturing PAGE.

Labeling of native RNA 3'-ends

Unfunctionalized RNA was labeled at 3'-end with fluorescein-5-thiosemicarbazide by periodate oxidation method. RNA (5 nmol in 10 µL of H_2O) was mixed with sodium phosphate buffer (10 µL, 100 mM, pH = 7.4) and freshly prepared solution of $NaIO_4$ (2.5 µL, 100 mM). The reaction mixture was incubated at 37°C for 10 minutes. The reaction was quenched by Na_2SO_3 (5 µL, 100 mM) and incubated at 37°C for 5 minutes. To this reaction mixture, a solution of fluorescein -5-thiosemicarbazide (5 µL, 10 mM) in DMF was added and the reaction was incubated at 37°c for 3 hours in the dark. Labeled product was purified by denaturing PAGE.

5'-Phosphorylation of RNA substrates

Phosphorylation at 5'end of RNA substrates was carried with T4 PNK enzyme. RNA (2-5 nmol) incubated with 10× PNK buffer A (5 µL, 500 mM Tris-HCl, 10 mM $MgCl_2$, 50 mM DTT, 1 mM spermidine), ATP (5 µL, 10 mM) and T4 PNK enzyme (5 µL, 10 U/µL) at 37°C for 5 hours. The total volume of the reaction mixture was 50 µL. Afterwards, RNA was extracted with phenol/ chloroform/isoamylalcohol mixture followed by ethanol precipitation.

4. In vitro selection

Splint ligation of RNA substrate to deoxyribozyme selection pool

The RNA substrate was ligated with ssDNA pool using T4 DNA ligase in presence of complementary DNA splint. Deoxyribozyme pool (1.6 nmol), 5'-phosphorylated RNA and complementary splint were used in a molar proportion of 1: 1.5: 1.25 respectively in total volume of 12 µL. To this mixture, 10× annealing buffer (2 µL, 40 mM Tris-HCl, 150 mM NaCl, 1 mM EDTA, pH = 8.0) was added and heated at 95°C for 4 minutes. The mixture was allowed to cool down to 25°C for 10 minutes. Afterwards, 10× ligase buffer (2 µL, 400 mM Tris-HCl, 100 mM $MgCl_2$, 100 mM DTT, 5 mM ATP, pH = 7.8) and T4 DNA ligase (2 µL, 5 U/µL) was added to the solution and incubated at 37°C for 2-3 hours. Ligated product (DNA-RNA hybrid) was resolved and purified by denaturing PAGE.

Selection Step

The DNA-RNA hybrid (250 pmol in the first round and 10-30 pmol in further rounds) was incubated with 10× selection buffer (1 µL, 500 mM Tris-HCl, 1.5 M NaCl), pH = 7.5) in a total volume of 9.5 µL. To initiate the reaction, $MgCl_2$ (0.5 µL, 100 mM) with a final concentration of 5 mM was added and reaction mixture was incubated at 37°C for 16 hours. The active fraction of DNA enzymes was separated by denaturing PAGE and the areas corresponding to cleaved products (determined by comparison to size marker) were cut and extracted. Extracted products were ethanol precipitated and subjected to PCR amplification.

PCR amplification of active DNA library fraction

The active fraction of the DNA library recovered in the selection step was amplified through two subsequent asymmetrical PCRs. For the first PCR, the extracted product from previous step was used as template. Template (dissolved in 30 µL ultra pure water), forward primer (D2, 200 pmol), reverse primer (D3, 50 pmol), dNTP mixture (0.63 µL, 20 mM), 10× DreamTaq buffer (5 µL) and DreamTaq DNA polymerase (0.25 µL, 5 U/µL) were mixed and volume was adjusted to 50 µL with water. The reaction mixture was subjected to PCR reaction with the following conditions: [95°C for 4 minutes, 10× (95°C for 30 sec, 60°C for 30 sec, 72°C for 1 min) and 72°C for 5 minutes]. PCR-I product was stored at -20°C as backup sample (in case to repeat the selection round). For the second PCR, an aliquot (2 µL) from PCR-I was used as template and mixed with fluorescently labeled forward primer (D2-flu, 200 pmol), reverse primer (D3, 50 pmol), dNTPs mixture (0.63 µL, 20 mM), 10× Dream *Taq* buffer (5 µL) and Dream*Taq* DNA polymerase (0.5 µL, 5 U/µL) and water to a final volume of 50 µL. PCR conditions were: [95°C for 4 minutes, 30× (95°C for 30 sec, 60°C for 30 sec, 72°C for 1 min) and 72°C for 5 minutes]. The PCR product was resolved on denaturing PAGE and the fluorescently labeled shorter strand was isolated. The purified PCR product then ligated to the RNA substrate to initiate the next round of selection.

SUPPORTING INFORMATION

5. Deep sequencing analysis of enriched selection libraries

A schematic summary of sample preparation and sample nomenclature is shown in Figure S2a. In total, five Illumina libraries were prepared. The enriched DNA library from round 15 (AB_R15) was subjected to NGS and named ab1. It was also used for additional 16th round (AB_R16) with unmodified RNA (active: ab2, inactive: ab3), and with modified RNA (active: ab4, inactive: ab5). The isolated active and inactive fractions were amplified in a 2-step PCR protocol, in which Illumina adaptors and index sequences were added (Figure S2b). 1st PCR: forward primer (D6, 30 pmol), reverse primer (D7, 30 pmol), dNTPs mixture (0.63 µL, 20 mM), 10× DreamTaq buffer (5 µL) and DreamTaq DNA polymerase (0.25 µL, 5 U/µL) were mixed and volume was adjusted to 50 µL with water. PCR conditions were: [95°C for 4 minutes, 25× (95°C for 30 sec, 60°C for 30 sec, 72°C for 1 min) and 72°C for 5 minutes]. 2nd PCR was done with primers having index sequences. For each sample there was a unique index sequence in the reverse primer (D9-D13). An aliquot (2 µL) from 1st PCR was used as template and mixed with forward primer_D8 (100pmol), reverse primers_D9-D13 (100 pmol), dNTPs mixture (0.63 µL, 20 mM), 10× Dream *Taq* buffer (5 µL), Dream*Taq* DNA polymerase (0.25 µL, 5 U/µL) and water to a final volume of 200 µL. PCR conditions were: [95°C for 4 minutes, 30× (95°C for 30 sec, 60°C for 30 sec, 72°C for 1 min) and 72°C for 5 minutes]. The PCR products were purified on a 2% agarose gel and submitted to the Core Unit Systems Medicine (University of Würzburg) for amplicon sequencing. After quality control on a Bioanalyzer, the libraries were pooled, and sequenced as a fraction of a NextSeq-500 HighOutput 75nt single end run. After trimming and demultiplexing, the fastq and fastqc files were obtained from the Core unit and further processed. The raw fastq files were deduplicated of PCR duplicates with PRINSEQ[5] (prinseq-lite.pl -fastq INPUT_FILE -out_format 3 -derep 1 -out_good OUTPUT_FILE). Afterwards, the reads were trimmed of UMIs, and constant regions and filtered to the length of 18–21 nt with cutadapt[6] (cutadapt -g "CGACTAGTTACGGAAG;e=0.2" -a "CTTCATTCAGTTGGCC;e=0.3" -m 18 -M 21 --overlap 10 --times 2 -o OUTPUT_FILE INPUT_FILE). The processed reads were further analyzed using FASTAptamer[7] and custom R scripts. Multiple sequence alignment was carried out with msa[8] and sequence logos were generated using ggseqlogo.[9]

6. Single turnover kinetics characterization of deoxyribozymes

In a typical experiment for determining the trans-activity of individual deoxyribozymes, the synthetic deoxyribozyme (100 pmol) was mixed with 3' fluorescently labeled RNA substrate (10 pmol) in a final volume of 8.5 µL. Mixture was heated to 95°C for 4 minutes followed by cooling at 25°C for 10 minutes. To initiate the reaction, 10× kinetic assay buffer (1 µL, 500 mM Tris-HCl, 1.5 M NaCl), pH = 7.5) and $MgCl_2$ (0.5 µL, 400 mM) were added and reaction mixture was incubated at 37°C. Aliquots (1 µL) were taken at different time points (0, 10, 30, 60, 120, 180, and 360 min, sometimes 22h), quenched with loading buffer and analyzed by denaturing PAGE. The gels were imaged on a Chemidoc device and the cleavage yield was determined from band intensities quantified by ImageLab software. Values of k_{obs} (observed cleavage rate) and Y_{max} (maximum yield) were obtained by fitting cleavage yield *versus* time (min) with first order kinetics equation; $Y = Y_{max} * (1-e^{-kobs*t})$. Three independent replicates were performed for each experiment.

7. Supporting Tables

Table S1. Sequences and ESI-MS of RNA oligonucleotides

No	Description	5'-Sequence-3'	mol.wt. calc. amu	mol.wt. found m/z
R1	Unmodified RNA	AUAGACUGAAUGAAGGACUUCCGUAACU	8966.24	8966.31
R1a	Unmod. with 3'-C_6-NH_2	AUAGACUGAAUGAAGGACUUCCGUAACU-NH_2	9145.31	9145.49
R2	i^6A RNA	AUAGACUGAAUGAAGGi^{6}ACUUCCGUAACU	9034.30	9034.35
R2a	i^6A with 3'-C_6-NH_2	AUAGACUGAAUGAAGGi^{6}ACUUCCGUAACU-NH_2	9213.37	9213.48
R3	m^6A RNA	AUAGACUGAAUGAAGGm^{6}ACUUCCGUAACU	8980.34	8980.45
R3a	m^6A with 3'-C_6-NH_2	AUAGACUGAAUGAAGGm^{6}ACUUCCGUAACU-NH_2	9159.41	9159.33

Table S2. Sequences of DNA oligonucleotides

No	Description	5'-Sequence-3'
D1	Selection pool	GTGACGCGACTAGTTACGGAAGN_{20}CTTCATTCAGTTGGCGCCTCC
D2	Selection forward primer	Hexynyl-GTGACGCGACTAGTTAC
D3	Selection reverse primer (tailed)	$(CAA)_4E_3$GGAGGCGCCAACTGAATGAA
D4	Splint for selection pool (round 1)	TTCATTCAGTCTATGGAGGCGCCAACTG
D5	Splint for other selection rounds	TTCATTCAGTCTATTGGAGGCGCCAACTG
D6	1st_PCR_forward _NGS_UMI	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGNNNNNNNNCGACTAGTTACGGAAG
D7	1st_PCR_reverse _NGS	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGCCAACTGAATGAAG
D8	2nd _PCR_forward_NGS_i5_S502	AATGATACGGCGACCACCGAGATCTACACCTCTCTATTCGTCGGCAGCGTC
D9	2nd _PCR_reverse_NGS_i7_N706	CAAGCAGAAGACGGCATACGAGATCATGCCTAGTCTCGTGGGCTCGG
D10	2nd _PCR_reverse_NGS_i7_N707	CAAGCAGAAGACGGCATACGAGATGTAGAGAGGTCTCGTGGGCTCGG
D11	2nd _PCR_reverse_NGS_i7_N710	CAAGCAGAAGACGGCATACGAGATCAGCCTCGGTCTCGTGGGCTCGG
D12	2nd _PCR_reverse_NGS_i7_N711	CAAGCAGAAGACGGCATACGAGATTGCCTCTTGTCTCGTGGGCTCGG
D13	2nd _PCR_reverse_NGS_i7_N712	CAAGCAGAAGACGGCATACGAGATTCCTCTACGTCTCGTGGGCTCGG
		DNA enzymes
D14	AA07	TAGTTACGGAAGGGGTTCGGTGGAGCGGCGCGATTCATTCAGT
D15	AA14	TAGTTACGGAAGTGGTCTCGCGGTTCCTGGTTATTCATTCAGT
D16	AA17	TAGTTACGGAAGCCTGCAAGGAGGTTTACCGGGTTCATTCAGT
D17	AB08	TAGTTACGGAGGGGGAAGCCAGTGGTACGTTCTTCATTCAGT
D18	AB08_A4G	TAGTTACGGAGGGGGGAGCCAGTGGTACGTTCTTCATTCAGT
D19	AB08_A4C	TAGTTACGGAGGGGGCAGCCAGTGGTACGTTCTTCATTCAGT
D20	AB08_A4T	TAGTTACGGAGGGGGTAGCCAGTGGTACGTTCTTCATTCAGT
D21	AB08_ΔA4	TAGTTACGGAGGGGG_AGCCAGTGGTACGTTCTTCATTCAGT
D22	AB08 _T19A	TAGTTACGGAAGGGGAAGCCAGTGGTACGTACTTCATTCAGT
D23	AB08 _T19C	TAGTTACGGAAGGGGAAGCCAGTGGTACGTCCTTCATTCAGT
D24	AB08 _T19G	TAGTTACGGAGGGGGAAGCCAGTGGTACGTGCTTCATTCAGT
D25	AB08* (AB08_□T14,T19G)	TAGTTACGGAAGGGGAAGCCAGTGG_ACGTGCTTCATTCAGT
D26	AB08_restored binding arm	TAGTTACGGAAGGGGAAGCCAGTGGTACGTTCTTCATTCAGT
D27	AC17	TAGTTACGGAAGGGGTCTCCAGCCGGACGTTATTCATTCAGT
D28	AC17_ΔA20	TAGTTACGGAAGGGGTCTCCAGCCGGACGTT_TTCATTCAGT
D29	AC17_ΔT19	TAGTTACGGAAGGGGTCTCCAGCCGGACGT_ATTCATTCAGT
D30	AC17_A20C	TAGTTACGGAAGGGGTCTCCAGCCGGACGTTCTTCATTCAGT
D31	AC17_A20G	TAGTTACGGAAGGGGTCTCCAGCCGGACGTTGTTCATTCAGT
D32	AC17_C12T (VMC10)	TAGTTACGGAAGGGGTCTCCAGCTGGACGTTATTCATTCAGT

Table S3. Summary of NGS analysis data

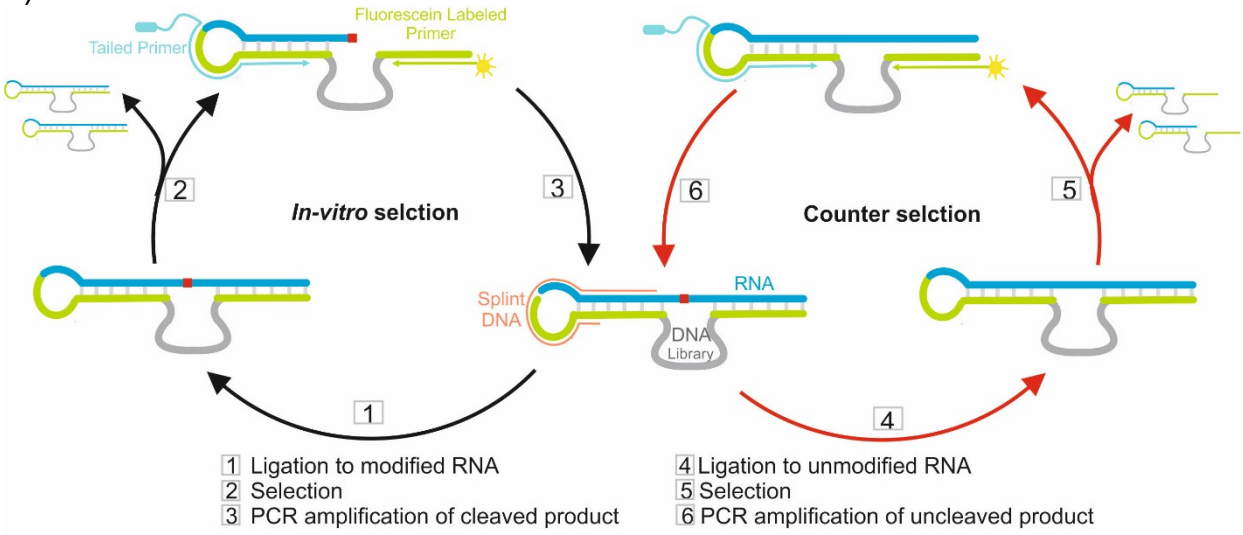
library		ab1	ab4	ab5	sum I (ab1+4+5)	ab2	ab3	sum II (ab2+3)
Round selection		15 reference	16 positive	16 positive		16 negative	16 negative	
RNA activity		- reference	i^6A (R2) cleaved active	i^6A (R2) uncleaved inactive		A (R1) cleaved active	A (R1) uncleaved inactive	
total reads (deduplicated)		750.106	345.138	252.618	1.347.862	773.935	616.504	1.390.439
unique sequences		115.993	16.248	43.496	175.737	62.876	104.136	167.012
seqs (> 50 reads)		796	340	275	1411	909	663	1572
AB08 (-d4 -f50)	cluster size	73	54	27	154	48	60	108
	reads	47.794	37.702	11.663	97.159	21.968	36.872	58.840
	% of total	6.4	10.9	4.6	21.9	2.8	6.0	8.8
AC17	# variants	538	125	225	888	1.177	426	1.603
AC17_ΔA20	# variants	1.277	420	524	2.221	2.172	1.025	3.197
	reads	15.930	6.491	3.626	26.047	26.979	10.047	37.026
	% of total	2.1	1.9	1.4	5.4	3.5	1.6	5.1

Table S4. Catalytic activities of DNA enzymes reported in this study

DNA	Unmodified RNA (R1)			i^6A-modified RNA (R2)		
	site	k_{obs} [$*10^{-2}min^{-1}$]	yield% @ 6h	site	k_{obs} [$*10^{-2}min^{-1}$]	yield% @ 6h
AA07	G15	0.11	42	-	no cleavage	-
AA14	G15	0.66	74	-	no cleavage	-
AA17	G16	0.65	77	-	no cleavage	-
AB08	-	no cleavage	-	G16	0.86	75
AB08-A4G	-	no cleavage	-	G16	0.27	65
AB08-A4C	-	no cleavage	-	G16	0.64	77
AB08-A4T	-	no cleavage	-	G16	0.30	67
AB08-ΔA4	-	no cleavage	-	-	no cleavage	-
AB08-T19G	-	no cleavage	-	G16	0.28	58
AB08-T19A	-	no cleavage	-	G16	0.26	54
AB08-T19C	-	no cleavage	-	-	no cleavage	-
AB08*Δ14T19G	-	no cleavage	-	G16	0.97	75
AC17	G16	1.8	82	G15	0.16	51
AC17-ΔT19	G16	0.41	60	G15	0.70	73
AC17-A20G	G16	1.1	84	G15	0.12	48
AC17-A20C	G16	1.7	87	G15	0.05	18
AC17-ΔA20	G16	0.04	21	G15	1.3	81
AC17-C12T	G16	0.77	81	G15	0.08	25

8. Supporting Figures

a)



b)

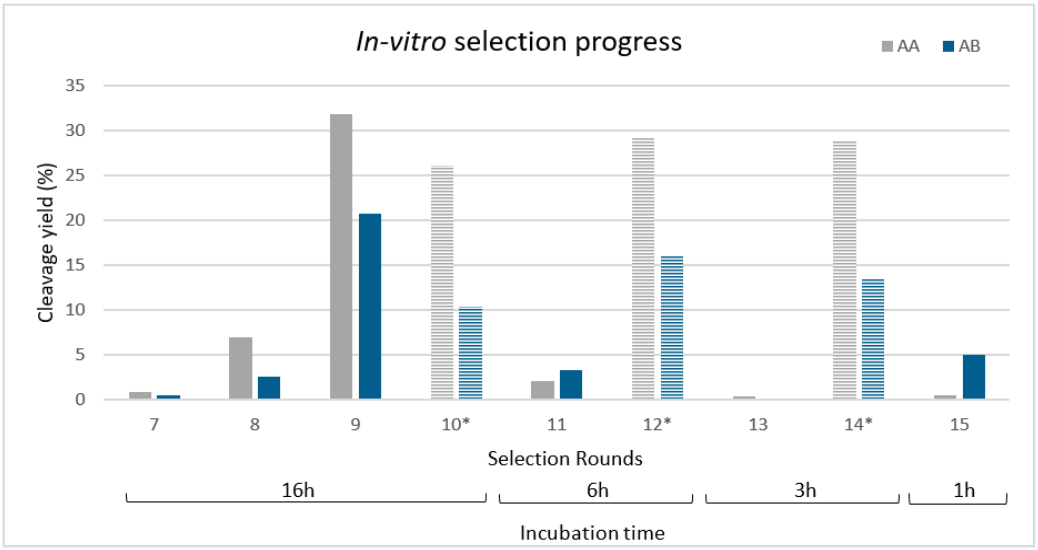
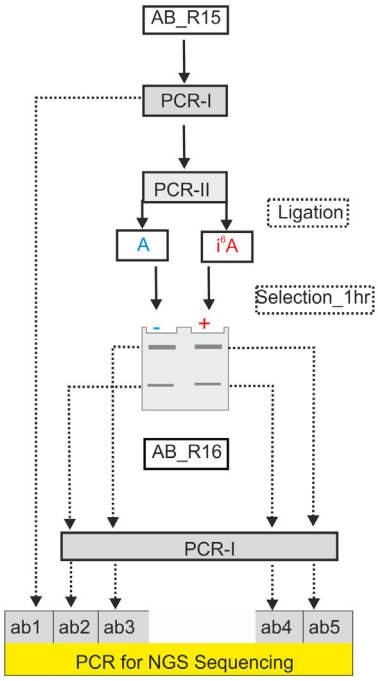


Figure S1. a) Schematic presentation of in vitro selection experiment to evolve i^{6}A-sensitive RNA cleaving deoxyribozymes (AB selection). In-Vitro selection: 1. Ligation of DNA library with modified RNA (R2). 2. Selection step in presence of $MgCl_2$ (5 mM, pH 7.5, 37°C). Cleaved fraction is isolated by PAGE. 3. PCR amplification of active fraction. Counter selection to enhance selectivity of the DNA enzymes: 4. Ligation of DNA library to unmodified RNA (R1). 5. Selection in presence of $MgCl_2$ (5 mM, pH 7.5, 37°C). The uncleaved fraction is isolated from PAGE. 6. PCR amplification of uncleaved fraction. For the AA selection, the procedure started with ligation of unmodified RNA, and the counter selection rounds were performed with i^{6}A-RNA. b) Progress of in-vitro selection plotted as % cleaved in each selection round. From round 1-6 fluorescence was below the detection limit. From round 7, the DNA pool was enriched enough to determine cleavage yield by measuring fluorescent band intensity during selection step by imaging the gel on a ChemiDoc imager. In round 11, the incubation time for selection was decreased to 6 h, and then further to 3 h in round 13 and finally to 1 h in round 15. The counter selection rounds are marked with asterisks and incubation time was 16 hour for each counter selection step.

a)



b)

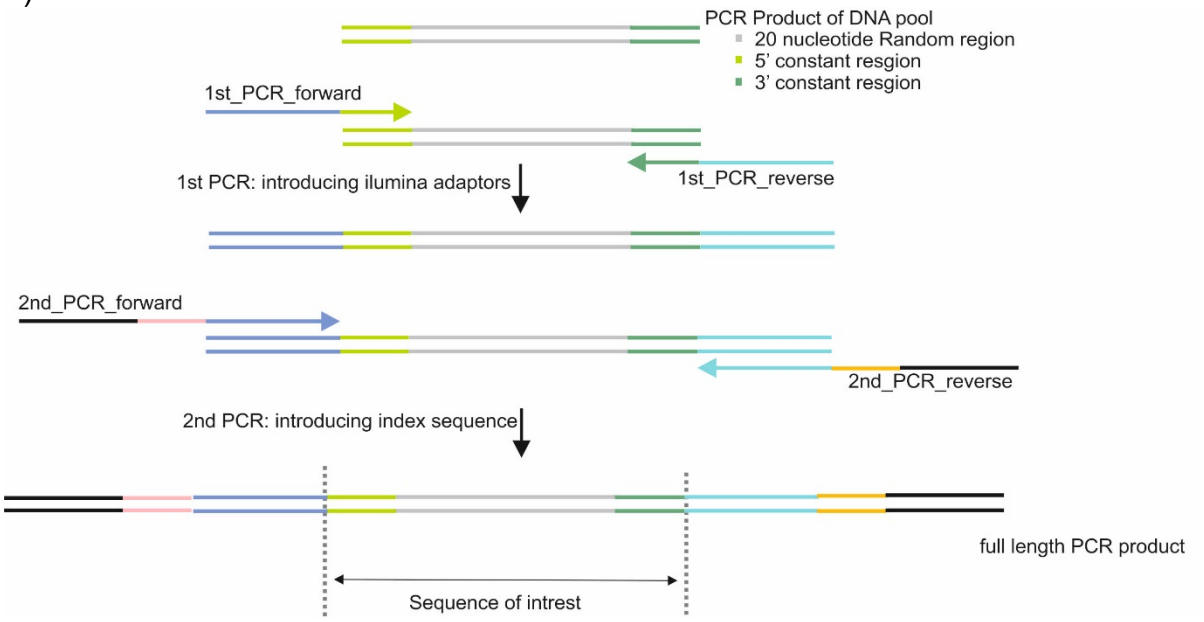
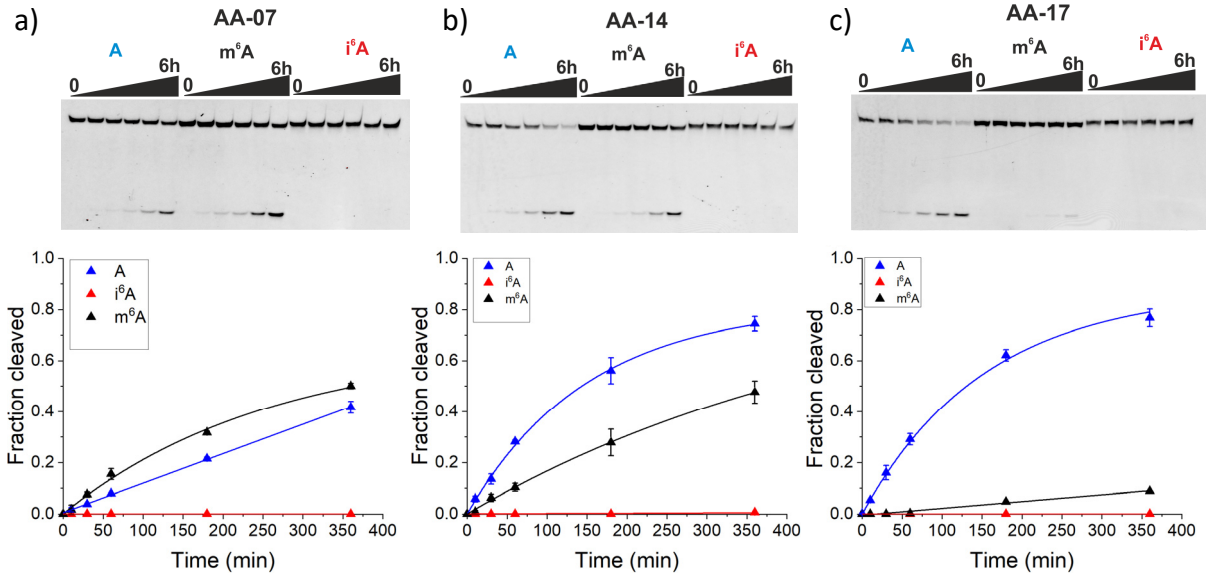


Figure S2. a) Schematic presentation of NGS libraries from round 15 and round 16. b) Schematic presentation of two-step-PCR for NGS library preparation to include Illumina adaptors and barcodes. The catalytic core (from 20 nt random region) is shown in grey, the 5' primer binding site in light green and 3' primer binding site in dark green. The 1st PCR introduces adaptor sequences (presented in blue and cyan, with primers D6 and D7 in Table S1), and the 2nd PCR introduces index sequences (pink and orange) from primers D8-D13. A unique index primer (2nd_PCR_reverse i7_primer, D9-D13) is used for each sample ab1-ab5.



RNA	AA07		AA14		AA17	
	k_{obs} [$*10^{-2}$ min^{-1}]	yield @ 6h	k_{obs} [$*10^{-2}$ min^{-1}]	yield @ 6h	k_{obs} [$*10^{-2}$ min^{-1}]	yield @ 6h
R1 A	0.12	42	0.66	74	0.65	77
R3 m^6A	0.30	49	0.19	47	0.02	8

Figure S3. Gel images, kinetic plots, rates and yields for AA DNA enzymes tested with unmodified RNA (R1), i^6A-RNA (R2) and m^6A-RNA (R3). Conditions: 1 µM RNA incubated with 10 µM DNA enzyme, 50 mM Tris-HCl, pH 7.5, 150 mM NaCl, 20 mM $MgCl_2$, 37°C. time points: 0, 10, 30, 60, 180, 360 min.

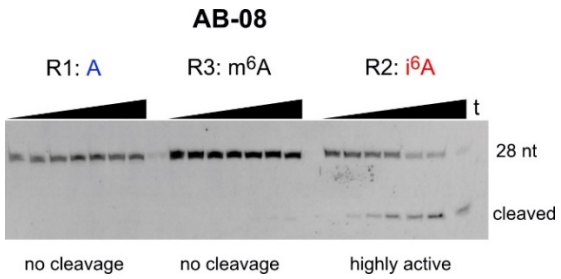
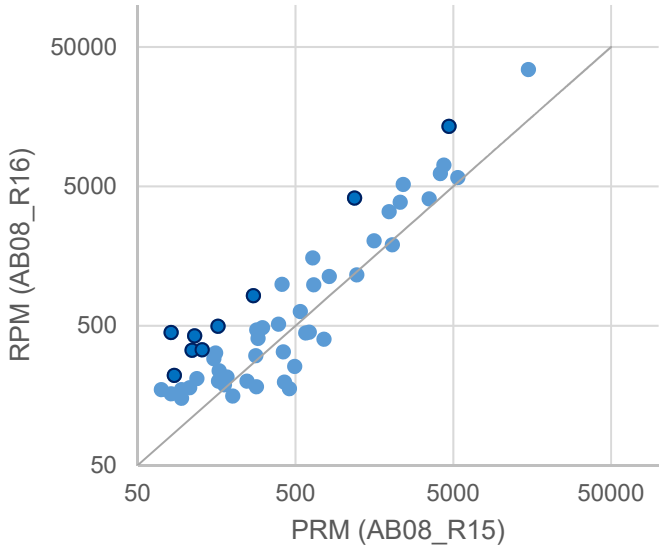


Figure S4. Gel image of AB08 tested on m^6A-RNA (R3), in comparison to R1 and R2. No cleavage was observed for m^6A-modified RNA, indicating that AB08 is specific for i^6A RNA. Conditions: 1 µM RNA incubated with 10 µM DNA enzyme in 50 mM Tris-HCl, pH 7.5, 150 mM NaCl, 20 mM $MgCl_2$, 37°C. Time points: 0, 10, 30, 60, 120, 180, 360 min.



	Sequence	RPM (x) R15(ab1)	RPM (y) R16(ab4)	enrichment (y/x)
AB08	GGGAAGCCAGTGGTACGTT	14981.9	34351.5	2.29
	GGGTAGCCAGTGG_ACGTG	81.3	446.2	5.49
	GGGTAGCCAGTGG_ACGTT	114.7	423.0	3.69
AB08*	GGGAAGCCAGTGG_ACGTG	1185.2	4111.4	3.47
	GGGAAGCCAGTGG_ACGTT	161.3	495.5	3.07
	GGGTAGCCAGTGG_ACGTA	270.6	820.0	3.03
	GGGCAGCCAGTGG_ACGTT	110.7	333.2	3.01
AB08_A4C	GGGCAGCCAGTGGTACGTT	4695.3	13461.3	2.87
	GGAAAGCCAGTGGTACGTT	128.0	336.1	2.63
	GGGAAGCCGGTGGTACGTT	85.3	220.2	2.58

Figure S5. Abundance of AB08 variants in NGS data from round 15 and round 16. Only Sequences with >50 RPM in round 16 are plotted. The most enriched variants are highlighted in dark blue, their sequences, RPM and enrichment factors are given, and mutations in comparison to AB08 are highlighted in red.

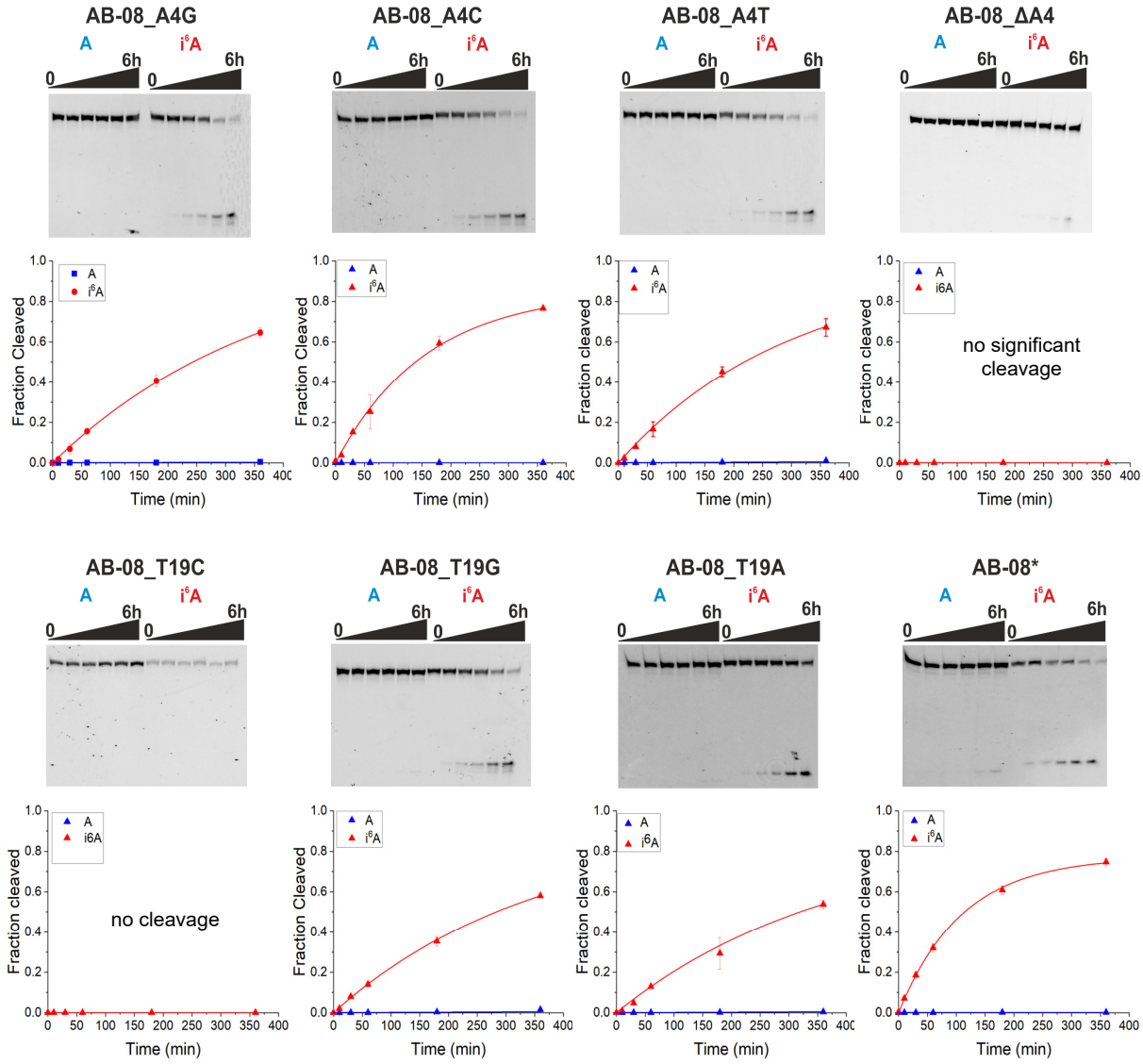


Figure S6. Gel images and kinetic plots for AB08 variants given in Table 2 in the manuscript. Conditions: 1 μM RNA incubated with 10 μM DNA enzyme in 50 mM Tris-HCl, pH 7.5, 150 mM NaCl, 20 mM $MgCl_2$, 37°C. Time points: 0, 10, 30, 60, 180, 360 min.

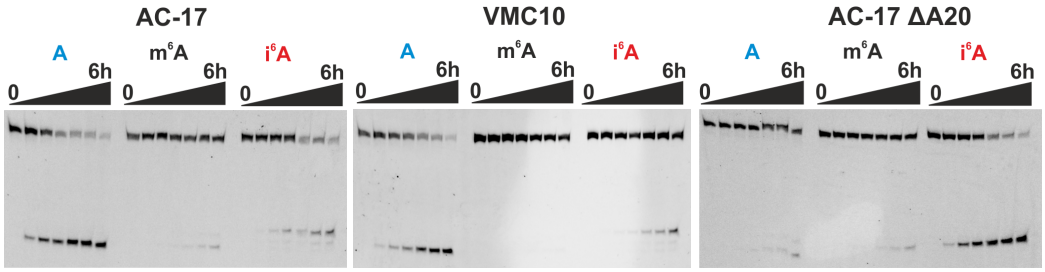


Figure S7. Gel images for plots in Figure 4. Conditions: 1 μM RNA incubated with 10 μM DNA enzyme in 50 mM Tris-HCl, pH 7.5, 150 mm NaCl, 20 mM $MgCl_2$, 37°C. Timepoints: 0, 10, 30, 60, 120, 180, 360 min.

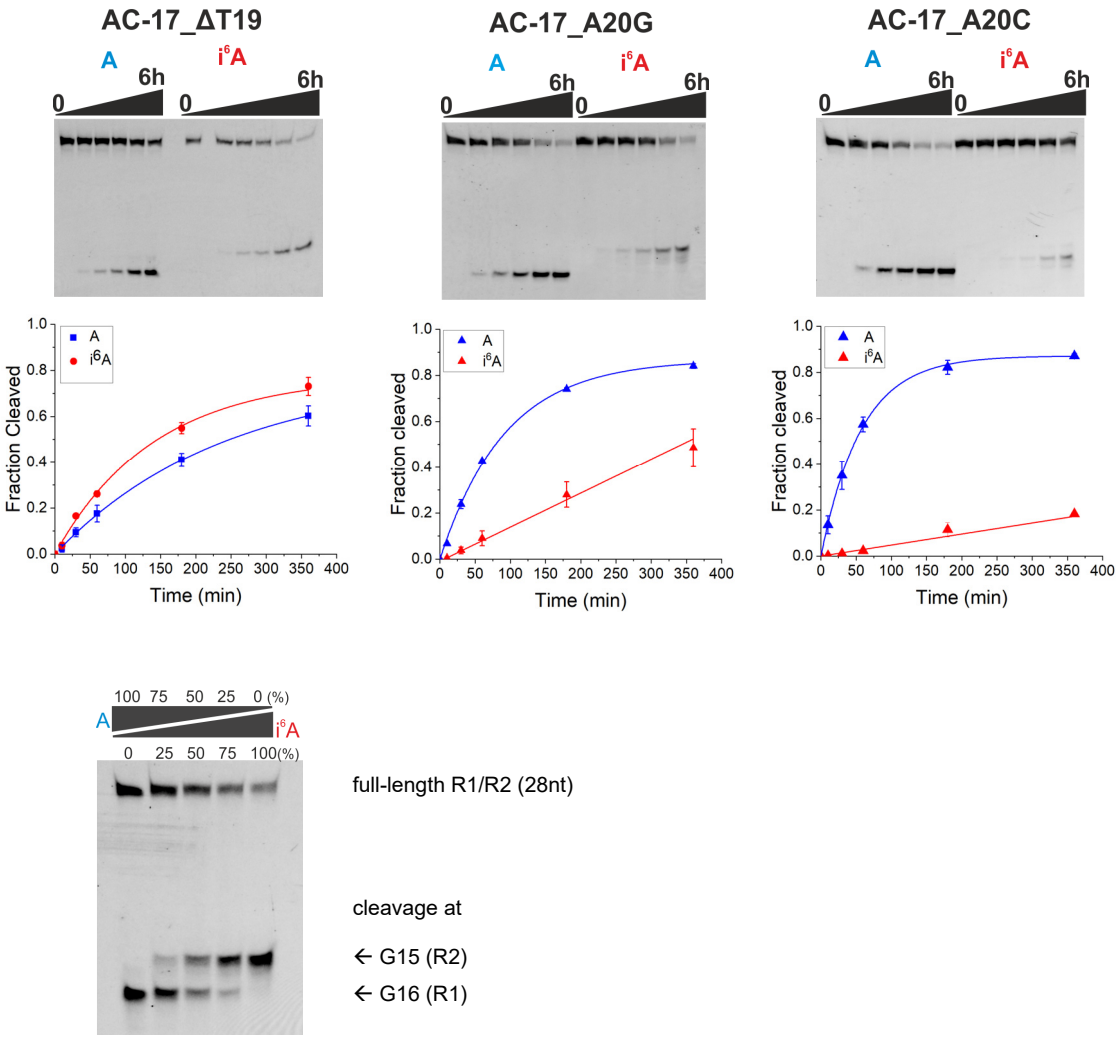


Figure S8. Top: Gel images and kinetic plots of AC17 variants given in Table 2 in the manuscript. Conditions: 1 µM RNA incubated with 10 µM DNA enzyme in 50 mM Tris-HCl, pH 7.5, 150 mm NaCl, 20 mM $MgCl_2$, 37°C. Timepoints: 0, 10, 30, 60, 180, 360 min. Bottom: Samples containing both R1 and R2 - with increasing amount of i^6A-RNA from left to right - were cleaved with AC17-ΔT19 for 6 h. The ratio of both cleavage products correlates to the i^6A content.

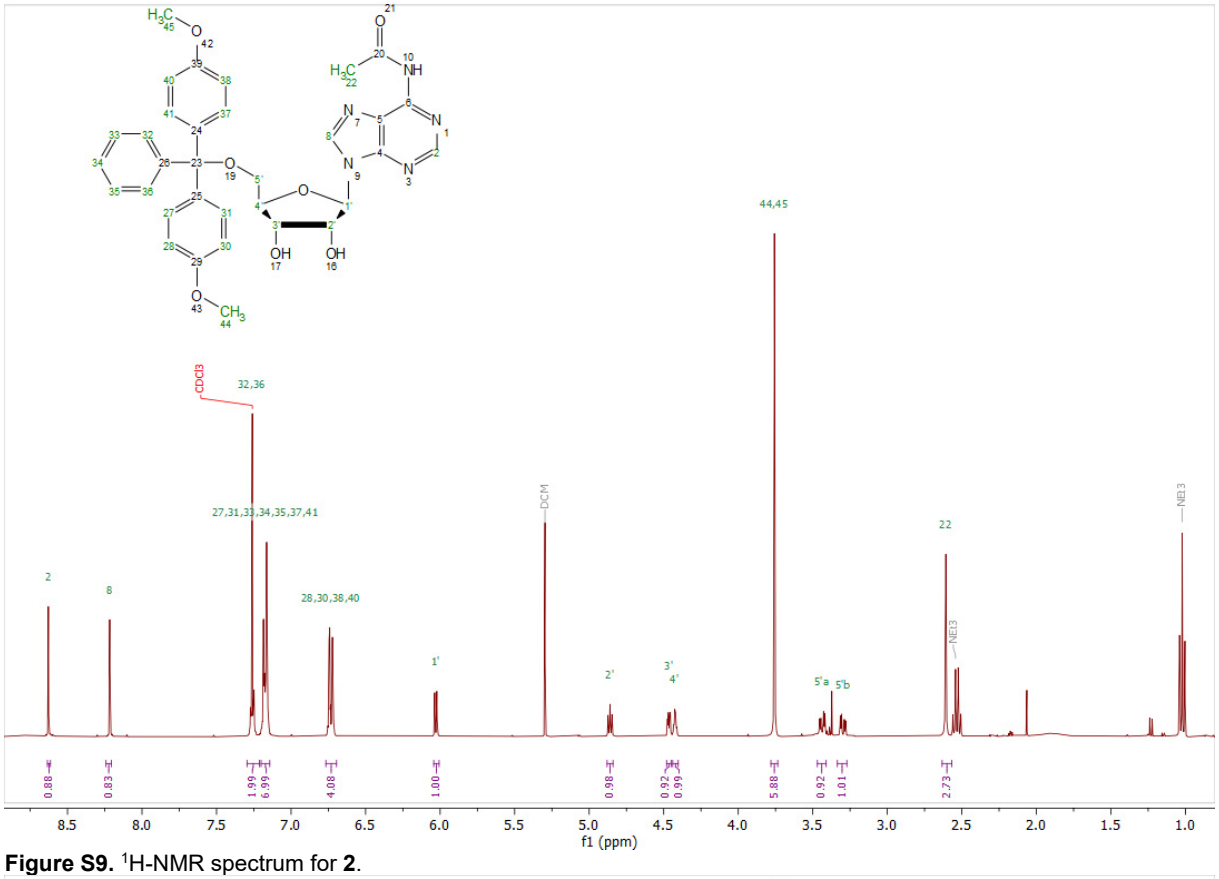


Figure S9. ^{1}H-NMR spectrum for **2**.

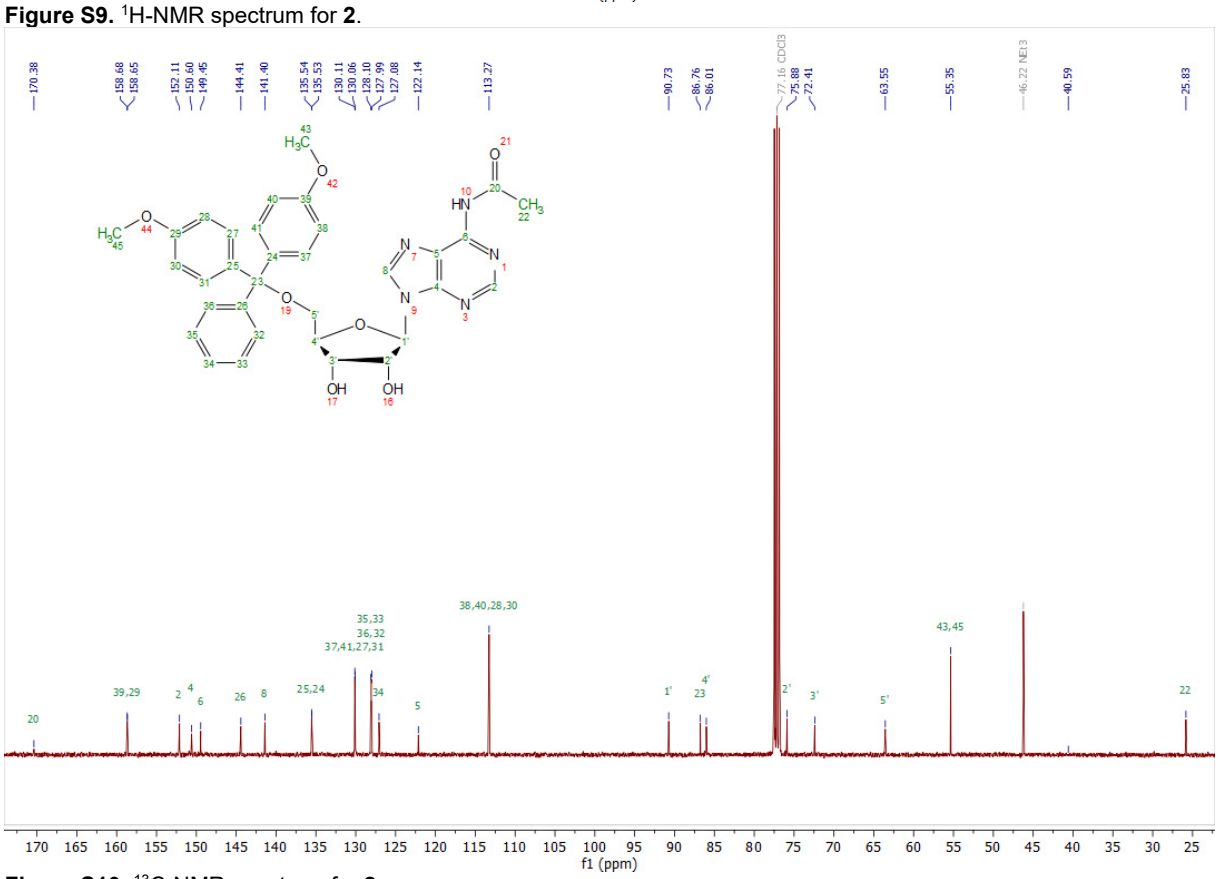


Figure S10. ^{13}C-NMR spectrum for **2**.

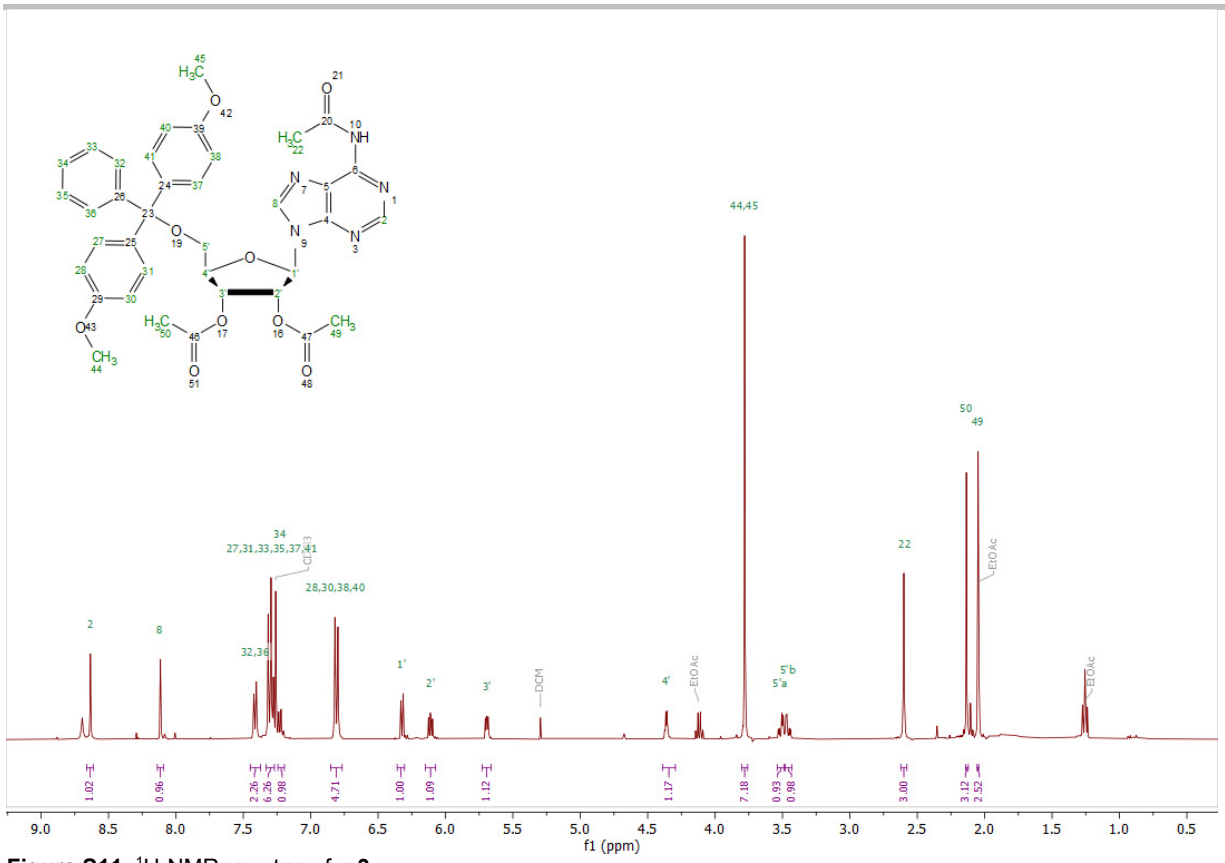


Figure S11. ^{1}H-NMR spectrum for **3**.

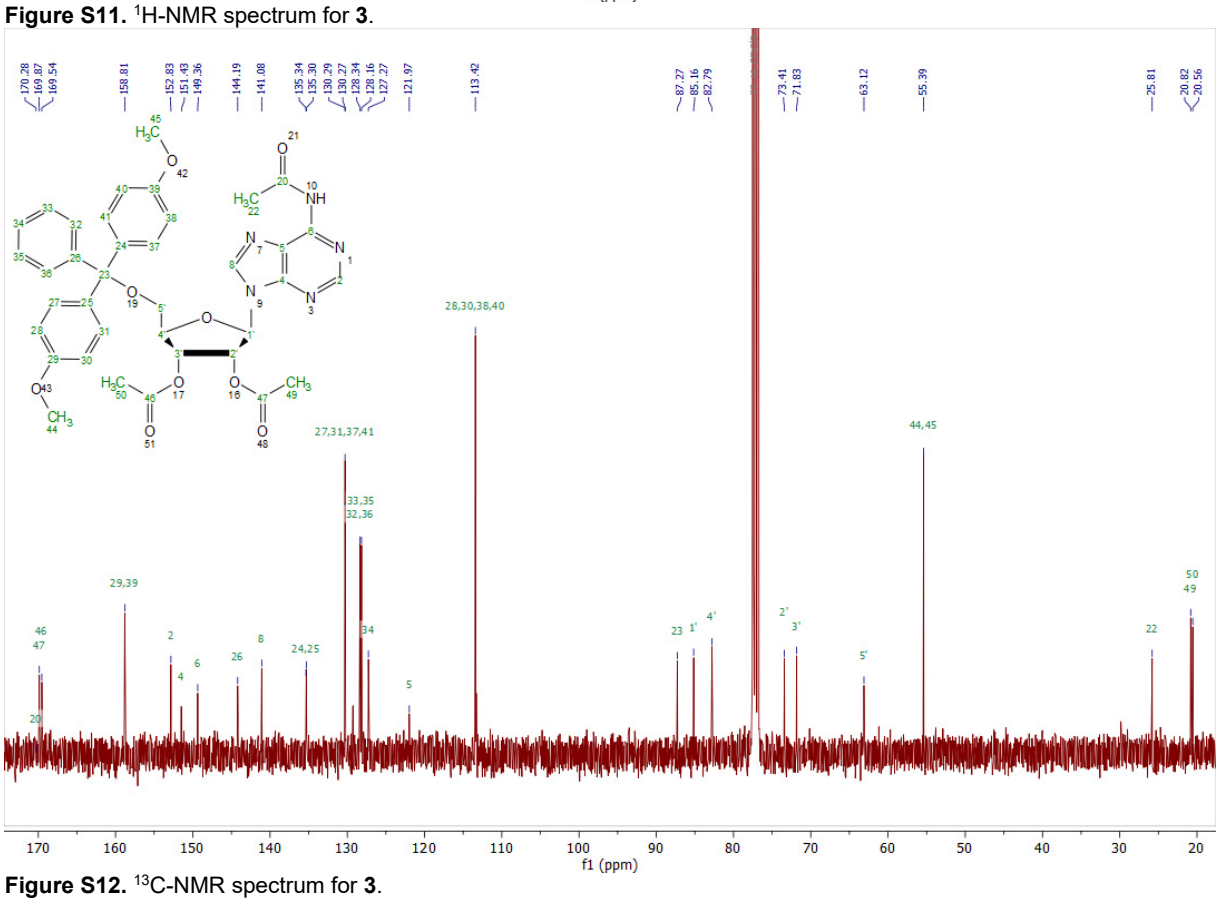


Figure S12. ^{13}C-NMR spectrum for **3**.

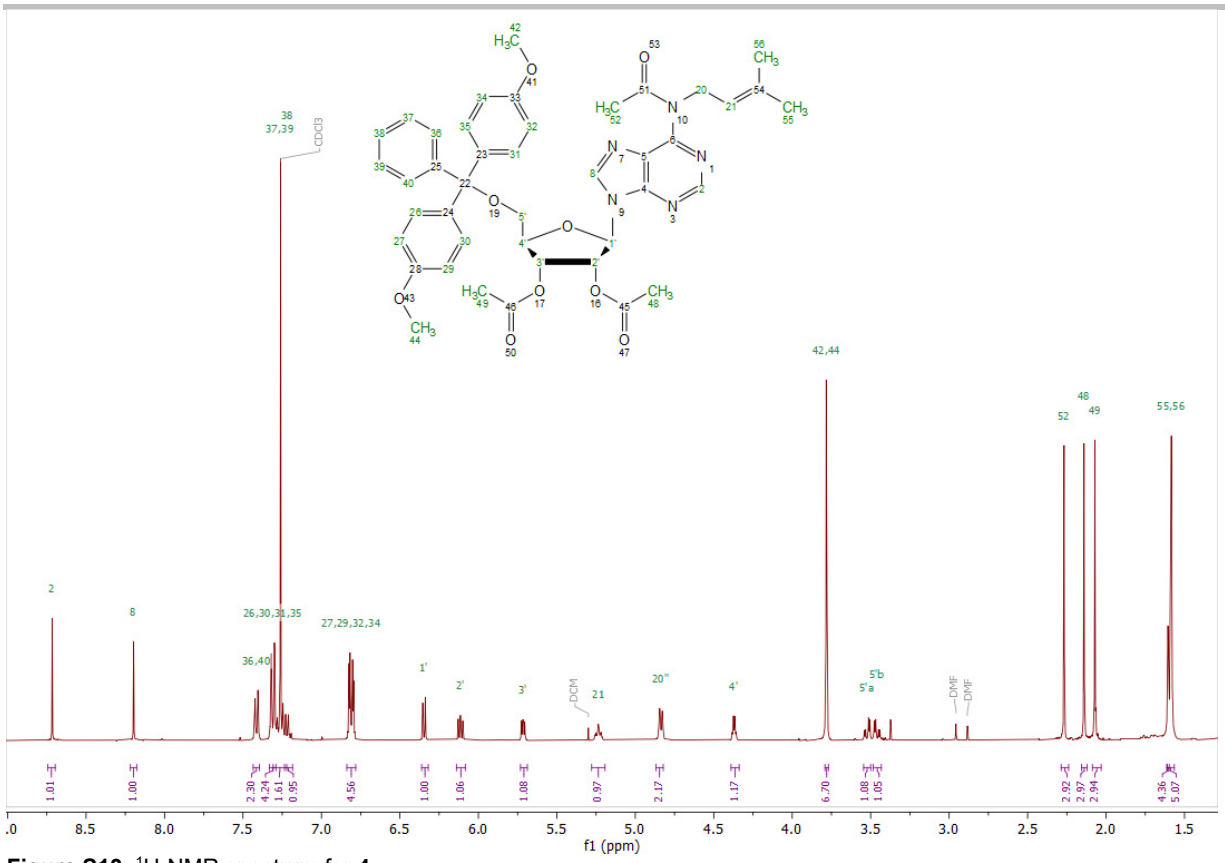


Figure S13. ^{1}H-NMR spectrum for **4**.

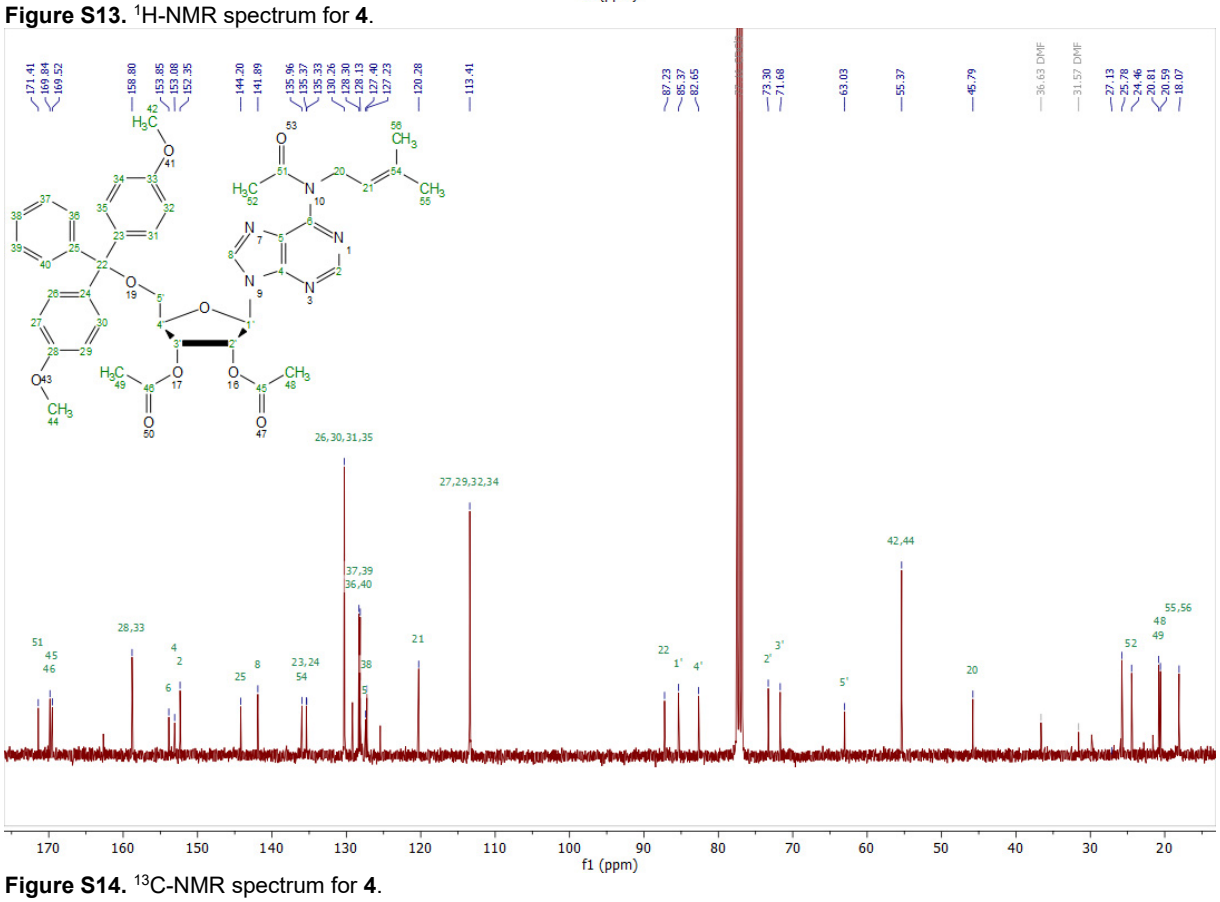


Figure S14. ^{13}C-NMR spectrum for **4**.

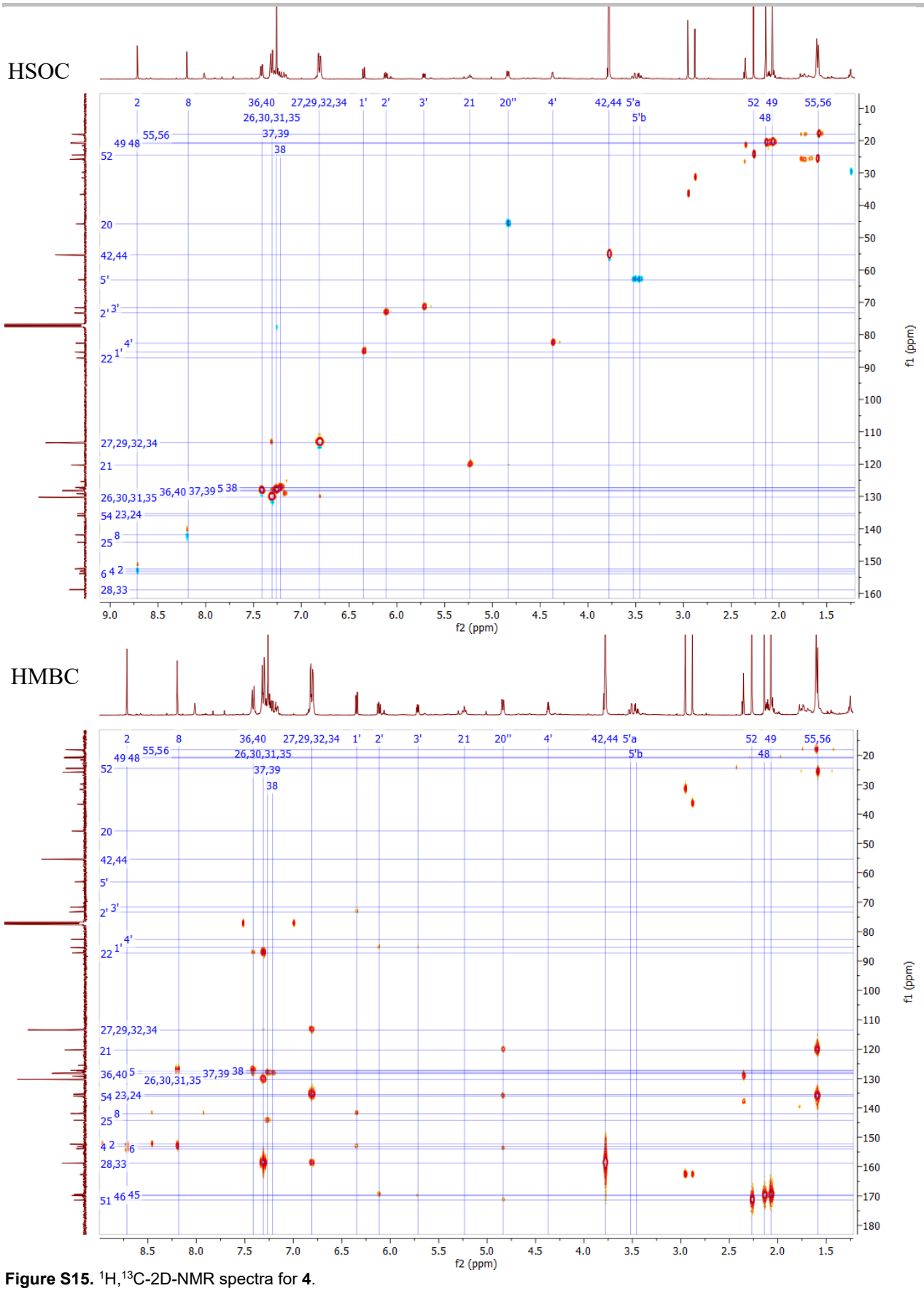


Figure S15. $^{1}H,^{13}C$-2D-NMR spectra for **4**.

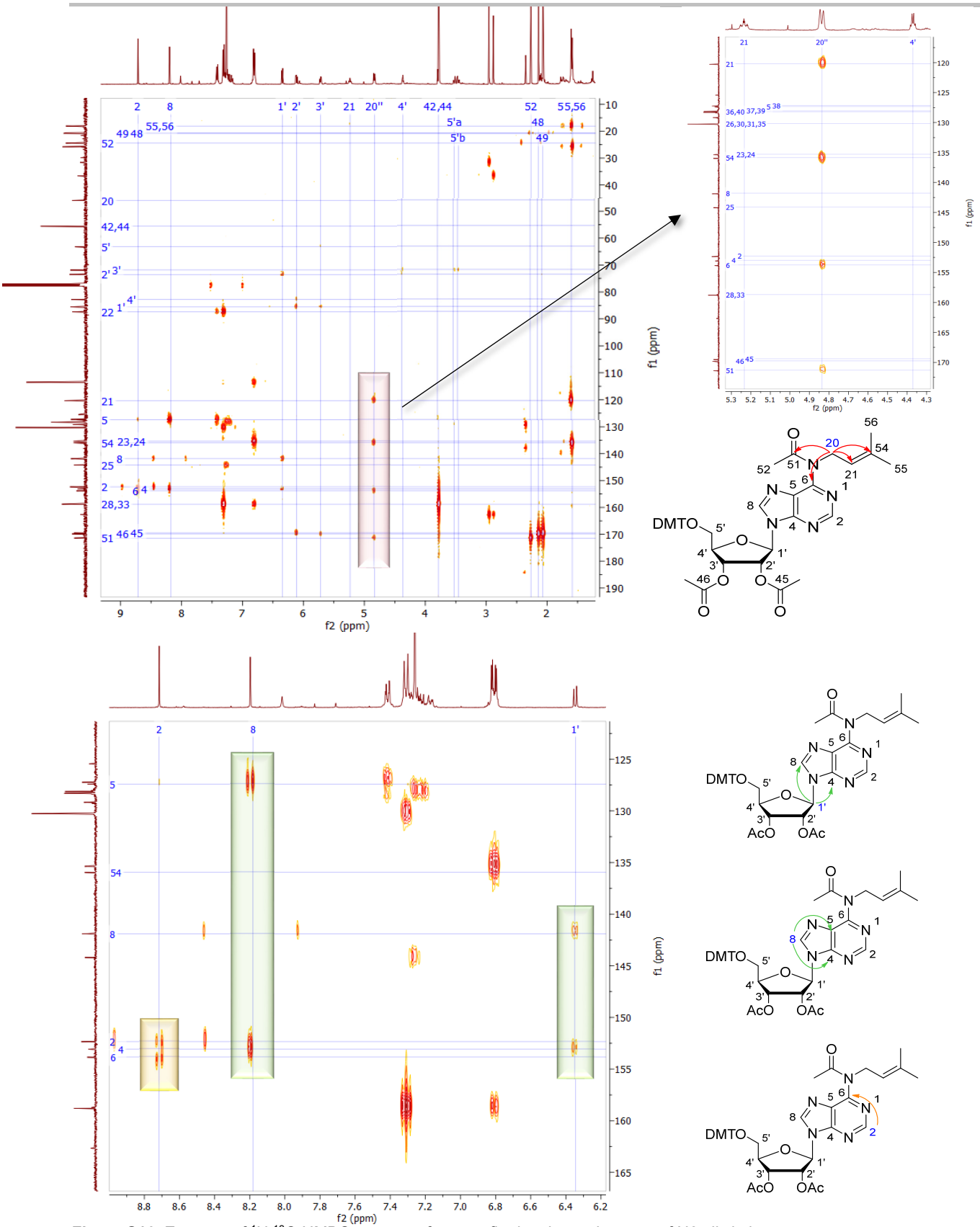


Figure S16. Excerpts of $^{1}H,^{13}C$-HMBC spectrum for **4** confirming the assignment of N6 alkylation.

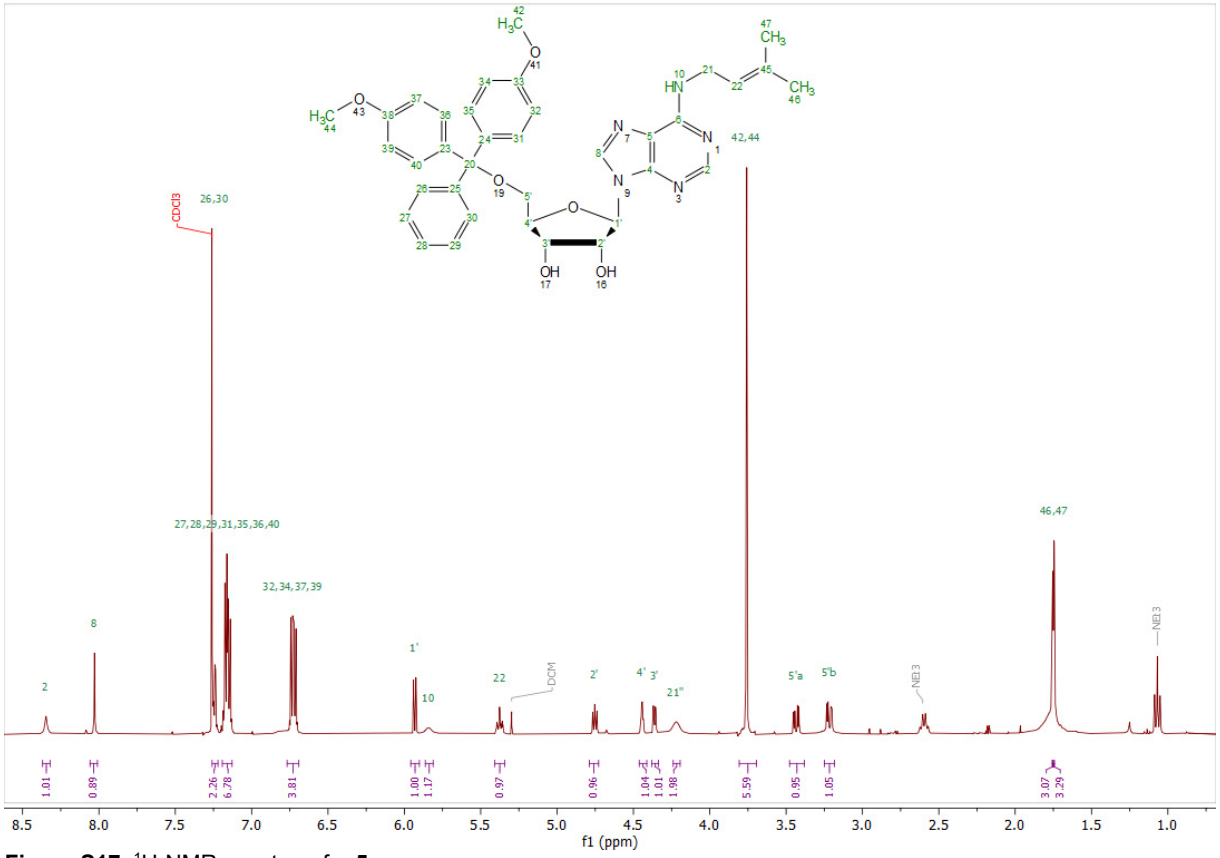


Figure S17. ^{1}H-NMR spectrum for **5**.

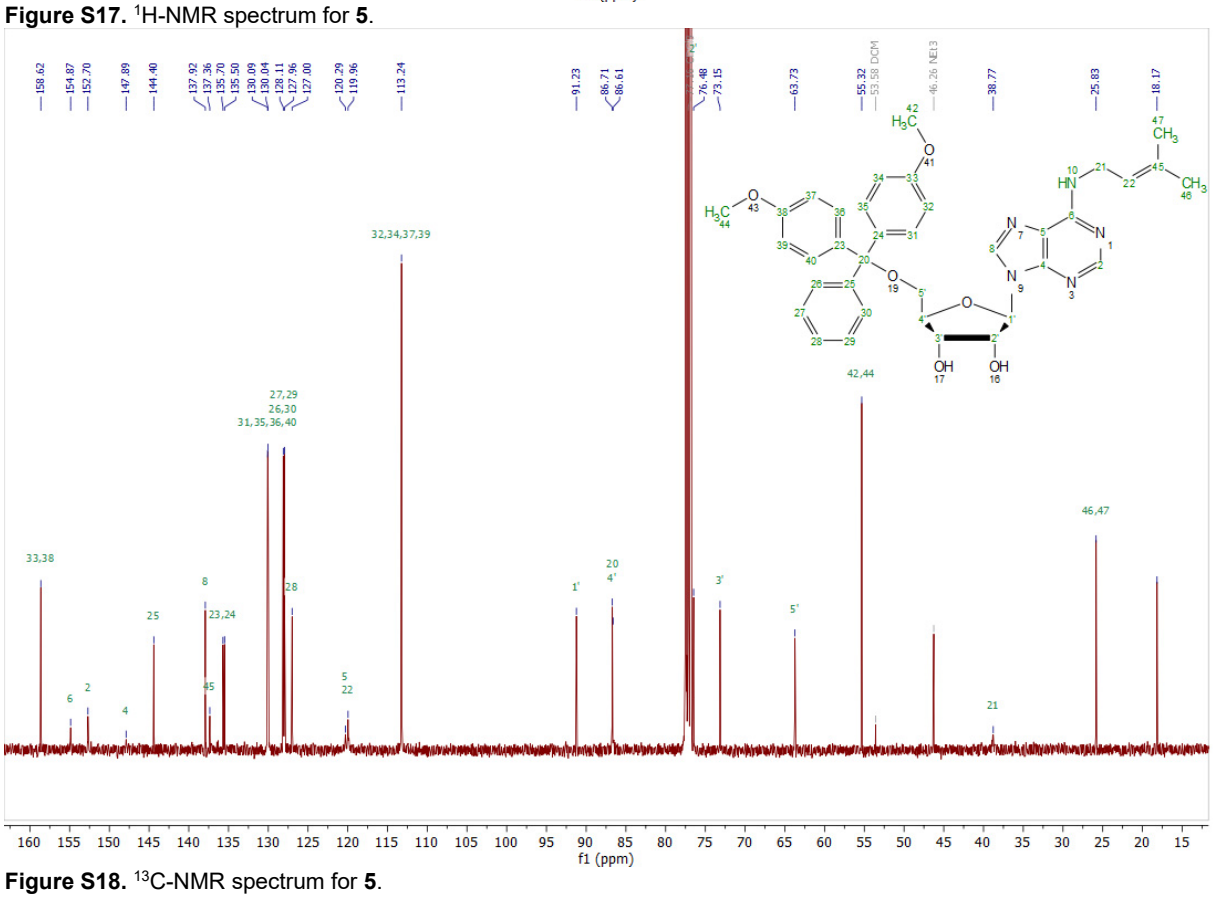


Figure S18. ^{13}C-NMR spectrum for **5**.

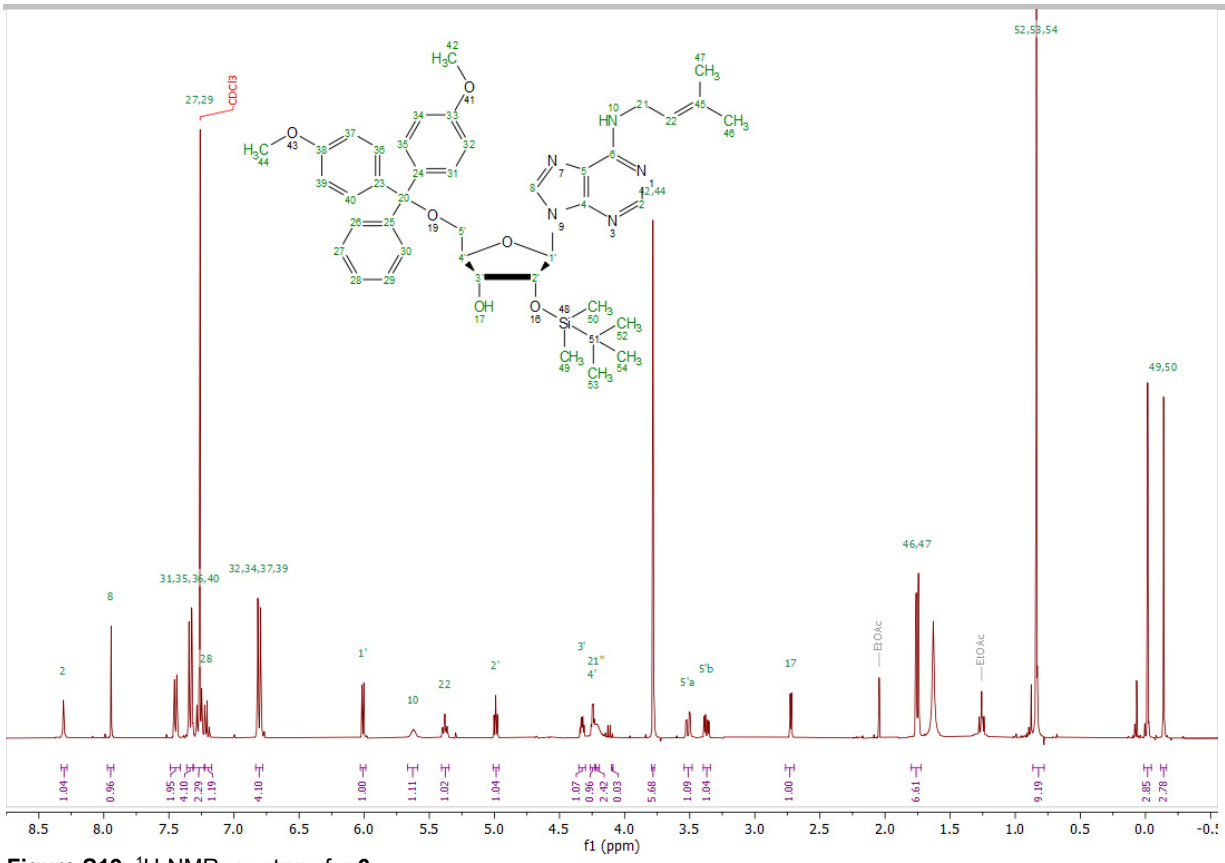


Figure S19. ^{1}H-NMR spectrum for **6**.

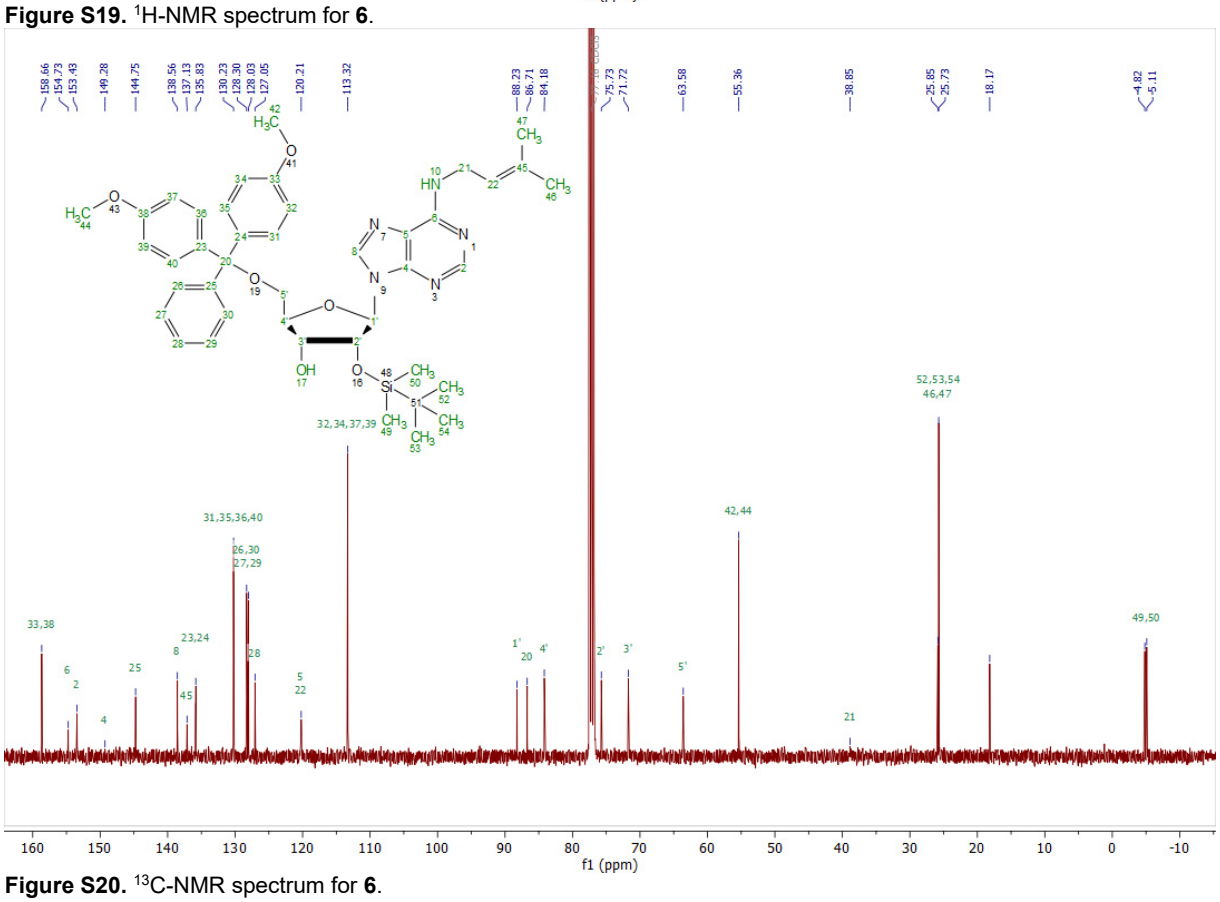


Figure S20. ^{13}C-NMR spectrum for **6**.

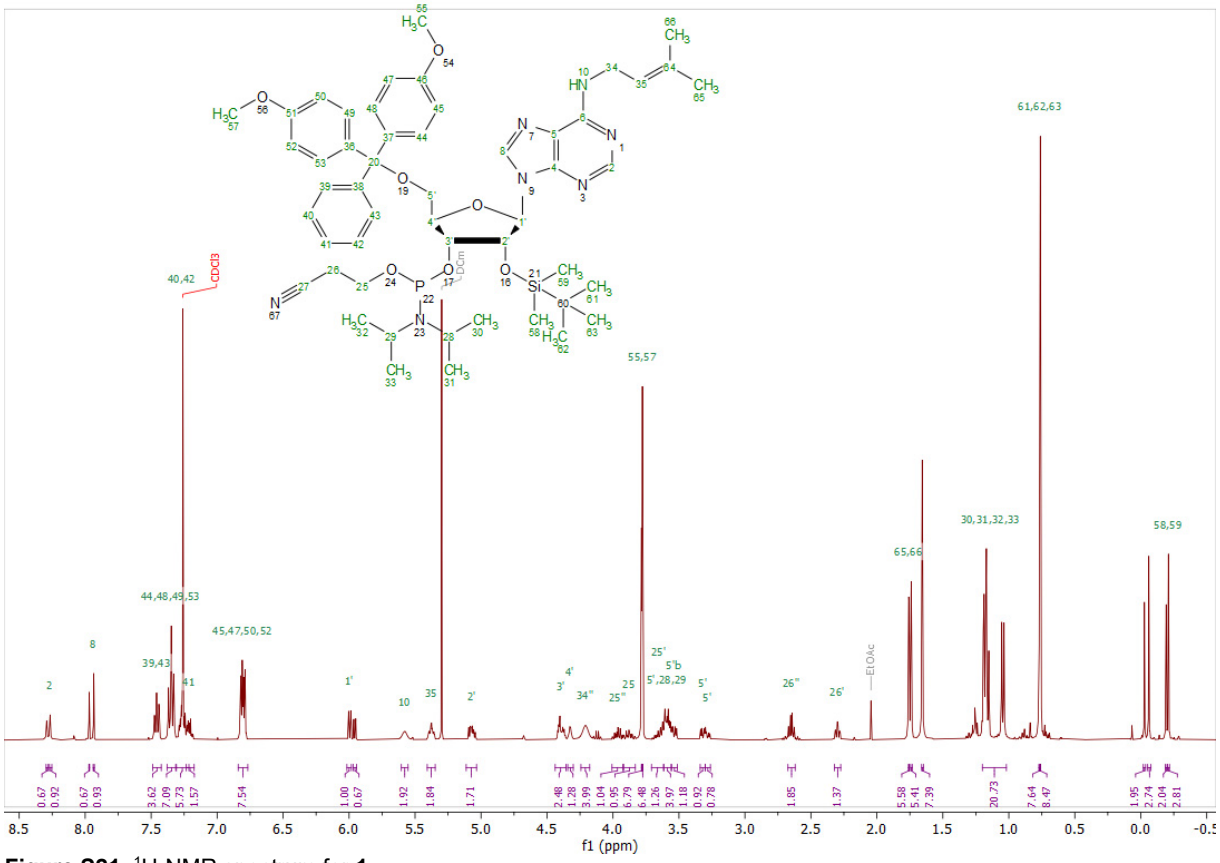


Figure S21. ^{1}H-NMR spectrum for **1**.

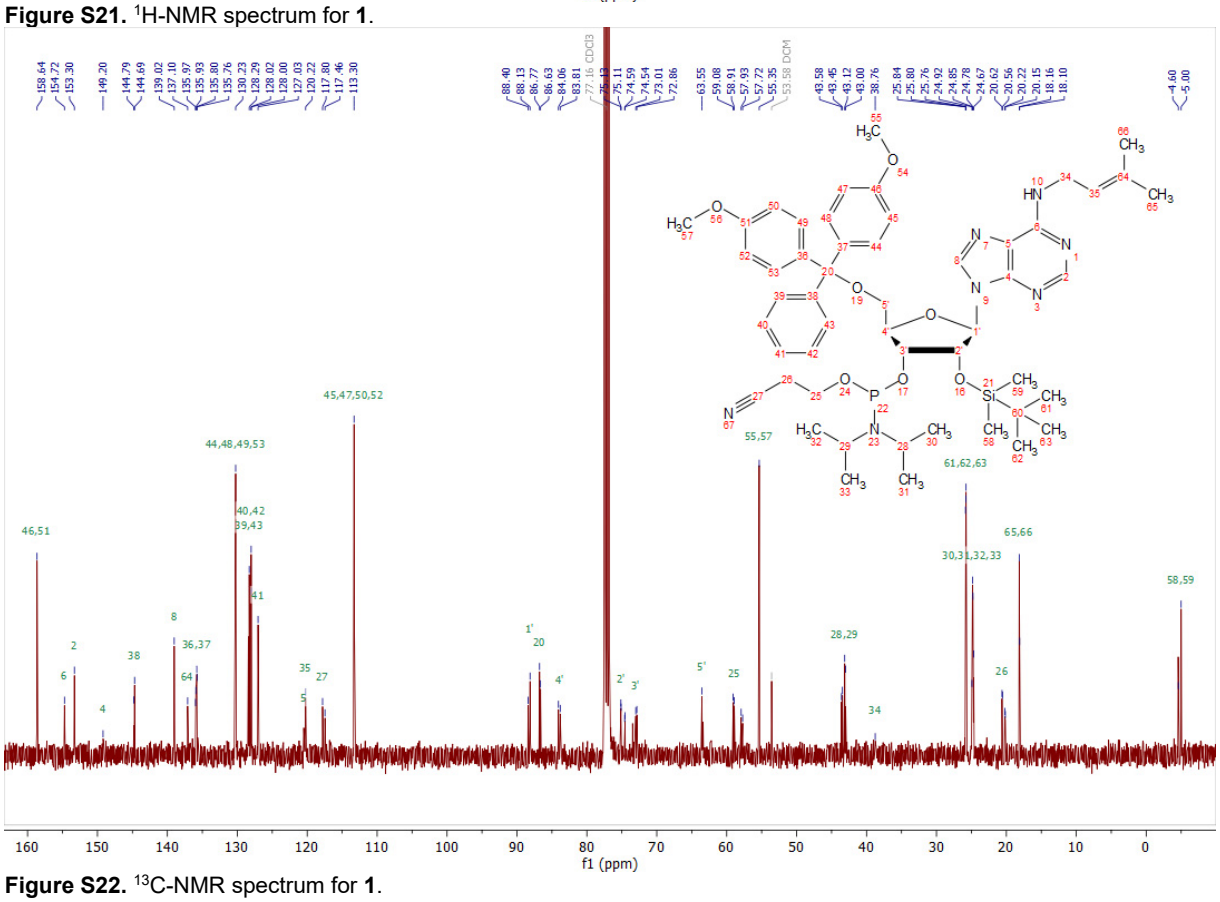


Figure S22. ^{13}C-NMR spectrum for **1**.

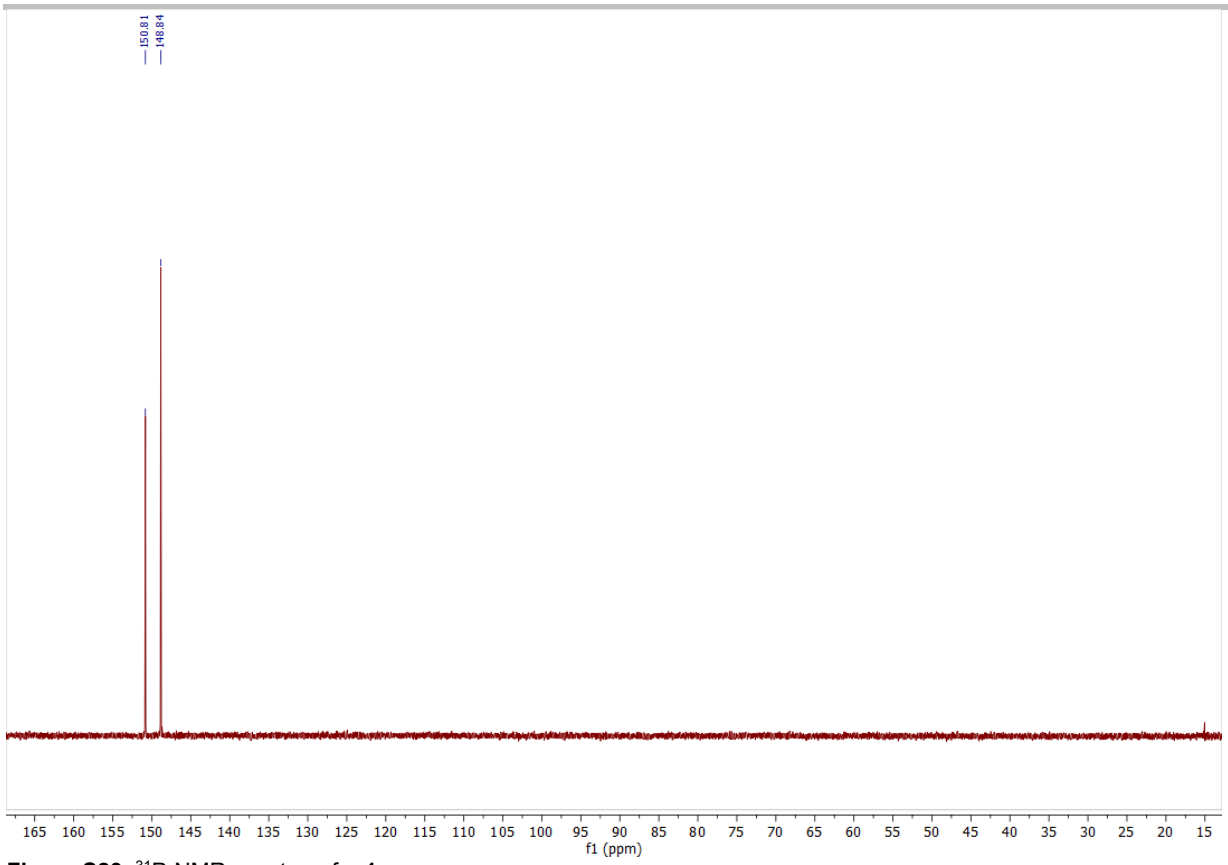


Figure S23. ^{31}P-NMR spectrum for **1**.

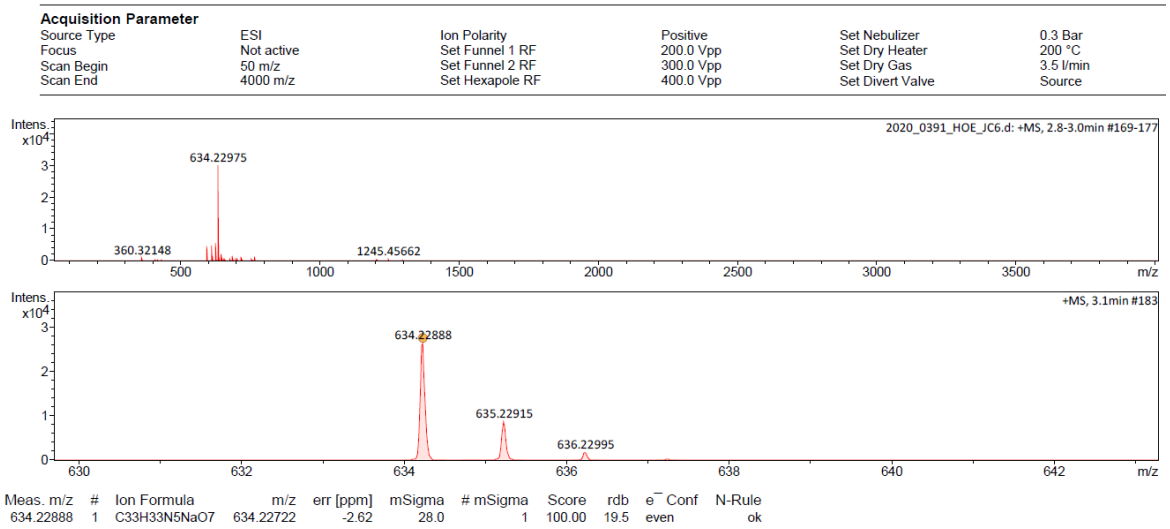
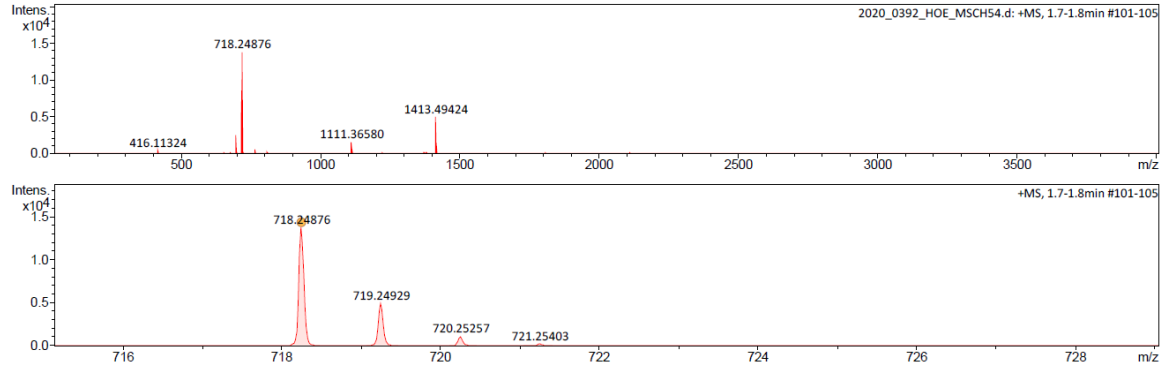


Figure S24. ESI(+)-MS spectrum for **2**.

Acquisition Parameter

Source Type	ESI	Ion Polarity	Positive	Set Nebulizer	0.3 Bar
Focus	Not active	Set Funnel 1 RF	200.0 Vpp	Set Dry Heater	200 °C
Scan Begin	50 m/z	Set Funnel 2 RF	300.0 Vpp	Set Dry Gas	3.5 l/min
Scan End	4000 m/z	Set Hexapole RF	400.0 Vpp	Set Divert Valve	Source

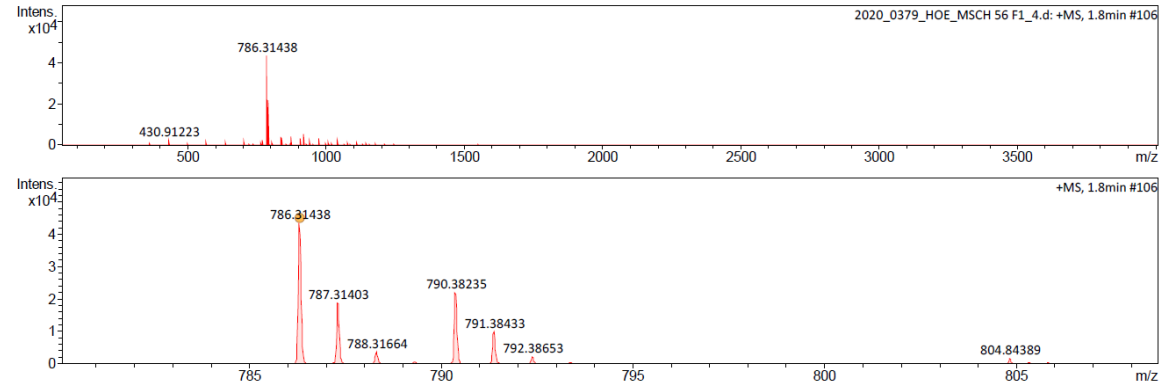


Meas. m/z	#	Ion Formula	m/z	err [ppm]	mSigma	# mSigma	Score	rdb	e¯ Conf	N-Rule
718.24876	1	C37H37N5NaO9	718.24835	-0.57	36.8	1	100.00	21.5	even	ok

Figure S25. ESI(+)-MS spectrum for **3**.

Acquisition Parameter

Source Type	ESI	Ion Polarity	Positive	Set Nebulizer	0.3 Bar
Focus	Not active	Set Funnel 1 RF	200.0 Vpp	Set Dry Heater	200 °C
Scan Begin	50 m/z	Set Funnel 2 RF	300.0 Vpp	Set Dry Gas	3.5 l/min
Scan End	4000 m/z	Set Hexapole RF	400.0 Vpp	Set Divert Valve	Source

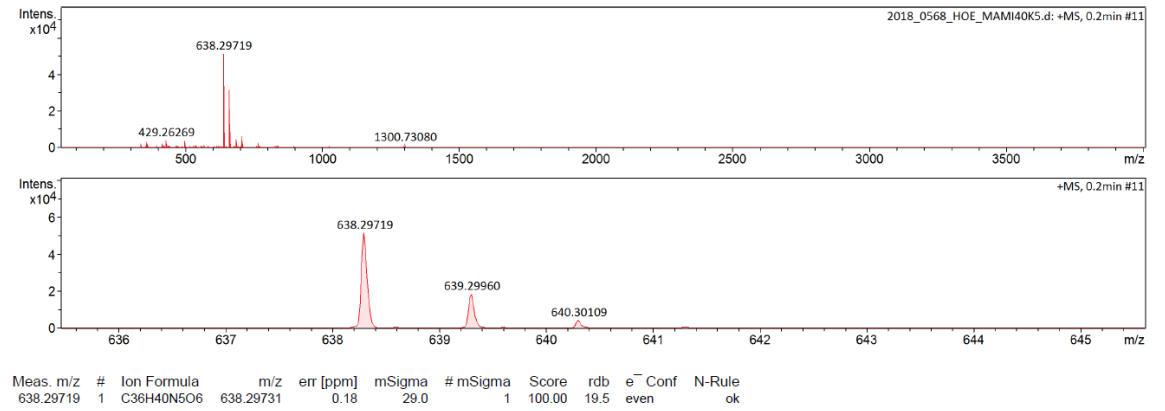


Meas. m/z	#	Ion Formula	m/z	err [ppm]	mSigma	# mSigma	Score	rdb	e¯ Conf	N-Rule
786.31438	1	C42H45N5NaO9	786.31095	-4.36	35.3	1	100.00	22.5	even	ok

Figure S26. ESI(+)-MS spectrum for **4**.

Acquisition Parameter

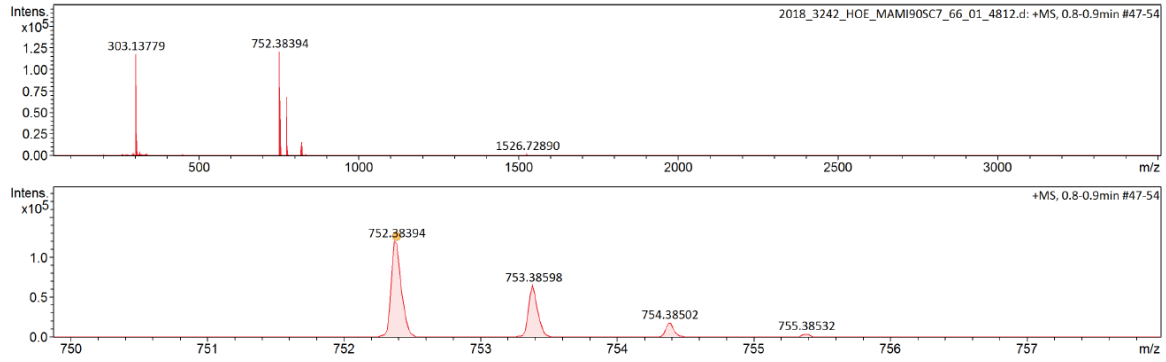
Source Type	ESI	Ion Polarity	Positive	Set Nebulizer	0.4 Bar
Focus	Not active	Set Funnel 1 RF	200.0 Vpp	Set Dry Heater	200 °C
Scan Begin	50 m/z	Set Funnel 2 RF	300.0 Vpp	Set Dry Gas	4.0 l/min
Scan End	4000 m/z	Set Hexapole RF	400.0 Vpp	Set Divert Valve	Source



Meas. m/z	#	Ion Formula	m/z	err [ppm]	mSigma	# mSigma	Score	rdb	e¯ Conf	N-Rule
638.29719	1	C36H40N5O6	638.29731	0.18	29.0	1	100.00	19.5	even	ok

Figure S27. ESI(+)-MS spectrum for **5**.

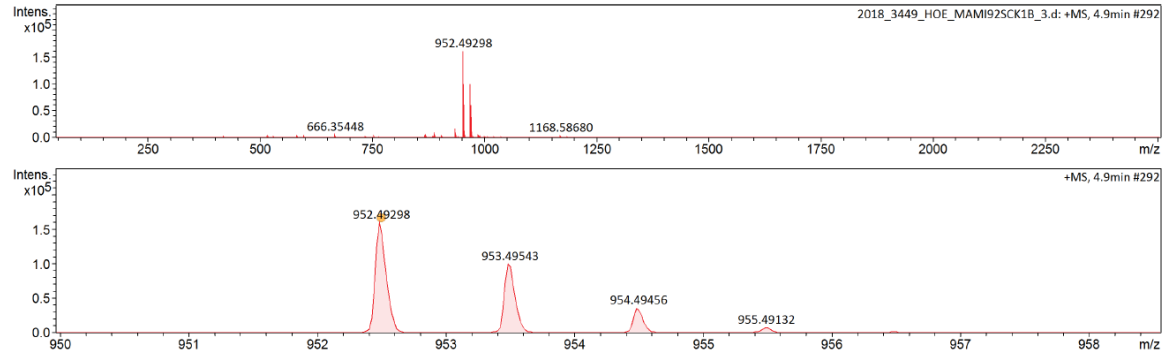
Acquisition Parameter					
Source Type	ESI	Ion Polarity	Positive	Set Nebulizer	0.3 Bar
Focus	Not active	Set Funnel 1 RF	100.0 Vpp	Set Dry Heater	200 °C
Scan Begin	50 m/z	Set Funnel 2 RF	200.0 Vpp	Set Dry Gas	5.0 l/min
Scan End	3500 m/z	Set Hexapole RF	300.0 Vpp	Set Divert Valve	Source



Meas. m/z	#	Ion Formula	m/z	err [ppm]	mSigma	# mSigma	Score	rdb	e¯ Conf	N-Rule
752.38394	1	C42H54N5O6Si	752.38379	-0.20	18.9	1	100.00	19.5	even	ok

Figure S28. ESI(+)-MS spectrum for **6**.

Acquisition Parameter					
Source Type	ESI	Ion Polarity	Positive	Set Nebulizer	0.3 Bar
Focus	Not active	Set Funnel 1 RF	300.0 Vpp	Set Dry Heater	200 °C
Scan Begin	50 m/z	Set Funnel 2 RF	400.0 Vpp	Set Dry Gas	4.0 l/min
Scan End	2500 m/z	Set Hexapole RF	400.0 Vpp	Set Divert Valve	Source



Meas. m/z	#	Ion Formula	m/z	err [ppm]	mSigma	# mSigma	Score	rdb	e¯ Conf	N-Rule
952.49298	1	C51H71N7O7PSi	952.49164	-1.41	19.7	1	100.00	21.5	even	ok

Figure S29. ESI(+)-MS spectrum for **1**.

9. Alignment of NGS reads for sequence logos

Alignment of 54 AB08 sequence variants used for the sequence logo in Figure 3f

using msa (Bioconductor R package) for multiple sequence alignment[8]

```
GGGCAG-CCA GTGGTACGTT --
GGGCAG-CCA GTGG-ACGTT --
GGGCAG-CCA GTGGTACGTT T-
GGGCAG-CCA GTGGTACGTG --
GGGCAG-CCA GTGGTACGTA --
GGGTAG-CCA GTGG-ACGTC TT
GGGTAG-CCA GTGG-ACGTC T-
GGGTAG-CCA GTGG-ACGTC TG
GGGTAG-CCA GTGG-ACGTC TA
GGGTAG-CCA GTGG-ACGTG --
GGGTAG-CCA GTGG-ACGTG T-
GGGTATTCCA GTGG-ACGTT --
GGGTATTCCA GTGGTACGTT --
GGGTATACCA GTGG-ACGTT --
GGGTAATCCA GTGG-ACGTT --
GGGTAATCCA GTGG-ACGTG --
GGGTAG-CCA GTGG-ACGTA --
GGGTAG-CCA GTGGTACGTG --
GGGTAG-CCA GTGGATCGTT --
GGGTAG-CCA GTGGATCGTG --
GGGTAG-CCA GTGG-ACGTT T-
GGGTAG-CCA GTGG-ACGTT TT
GGGTAG-CCA GTGG-ACGTT TG
GGGTAG-CCA GTGG-ACGTT TA
GGGTAG-CCA GTGG-ACGTT A-
GGGTAG-CCA GTGG-ACGTT --
GGGTAG-CCA GTGG-ACGTT G-
GGGTAG-CCA GTGGTACGTT --
GGGGAG-CCA GTGGTACGTT --
GGGAAG-CCA GTGGTACGTG --
GGGAAG-CCA GTGG-ACGTG --
GGGAAG-CCA GTGG-ACGTA --
GGGAAG-CCA GTGG-ACGTA T-
GGGAAG-CCA GTGGTACGTA --
GGGAAG-CCA GTGGTACGTC --
GGGAAG-CCA GTGG-ACGTT --
GGGAAG-CCA GTGG-ACGTT G-
GGGAAG-CCA GTGG-ACGTT T-
GGGAAG-CCA GTGGTACGTT T-
-GGGAATCCA GTGGTACGTT --
-GGAAAGCCA GTGGTACGTT --
GGGGAAGCCA GTGGTACGTT --
--GGAAGCCA GTGGTACGTT --
-GGGAAGCCA GTGGTACGTT --
GGGAAG-CCG GTGGTACGTT --
-GGGAGGCCA GTGGTACGTT --
GGGAAG-CCA GAGGTACGTT --
GGGAAG-CCA GTGGTACGCT --
GGGAAG-CCA GTGGTGCGTT --
GGGAAG-CCA GCGGTACGTT --
GGGTTGTCCA GCGGTACGTT --
GGGTTGCCCA GCGGTACGTT --
GGGATGCCCA GCGGTACGTT --
GGGATGTCCA GCGGTACGTT --
```

SUPPORTING INFORMATION

Alignment of 45 AC17 sequence variants used for the sequence logo in Figure 4f

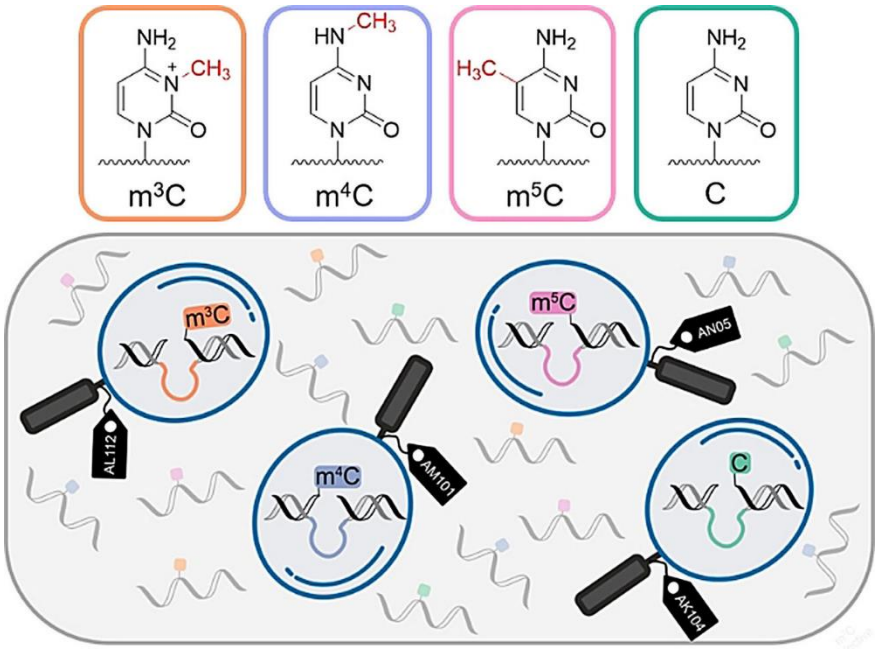
```
-GGGAC-CCA GCTGGACGTT TG
-GGGAC-CCA GCTGGACGTT T-
-GGGAC-CCA GCTGGACGTT TA
-GGGAC-CCA GCTGGACGTT AT
-GGGAC-CCA GCTGGACGTT --
-GGGAC-CCA GCTGGACGTT TT
-GGGAC-CCA GCCGGACGTT TG
-GGGAC-CCA GCCGGACGTT T-
-GGGAC-CCA GCCGGACGTT TA
-GGGAC-CCA GCCGGACGTT TT
-GGGAC-CCA GCCGGACGTT G-
-GGGAC-CCA GCCGGACGTT --
-GGGAC-CCA GCCGGACGTT A-
-GGGACTCCA GCCGGACGTT --
-GGGACTCCA GCCGGTCGTT --
-GGGACTCCA GCCGGACGTG --
-GGGAC-CCA GCCGGACGTG --
-GGGAC-CCA GCCGGACGTC TT
-GGGAC-CCA GCCGGACGTA --
-GGGAC-CCA GCTGGACGTA --
--GGGCTCCA GCTGGTCGTC TA
--GGGCTCCA GCTGGACGTC TT
--GGGCCCCA GCTGGACGTC TA
-GGGACTCCA GCTGGTCGTT --
GGGAT-TCCA GCTGGACGTA --
GGGATATCCA GCTGGACGTA --
GGGAT-TCCA GCTGGACGTT --
GGGAT-TCCA GCCGGACGTA --
GGGATATCCA GCCGGACGTA --
GGGAT-TCCA GCCGGACGTG --
GGGAT-TCCA GCCGGACGTT --
-GGGTCTCCA GCTGGTCGTA --
-GGGTCTCCA GCCGGTCGTA --
-GGGTCTCCA GCTGGTCGTT --
-GGGTCTCCA GCTGGCCGTT --
-GGGTCTCCA GCTGGACGTT --
-GGGTCTCCA GCTGGACGTG --
-GGGTCTCCA GCCGGACGTA --
-GGGTCTCCA GCCGGACGTG --
-GGGTCTCCA GCCGGACGTT --
-GGGCCTCCA GCTGGACGTA --
-GGGTAGCCA G-TGGACGTT A-
-GGGTATCCA G-TGGACGTT --
-GGTTATCCA GCTGGACGTT --
-GGGTATCCA G-TGGACGTA --
```

10. References

[1] S. Pitsch, P. A. Weiss, L. Jenny, A. Stutz, X. Wu, *Helv. Chim. Acta* **2001**, *84*, 3773-3795.
[2] V. I. Tararov, S. V. Kolyachkina, C. S. Alexeev, S. N. Mikhailov, *Synthesis* **2011**, *2011*, 2483-2489.
[3] V. I. Tararov, A. Tijsma, S. V. Kolyachkina, V. E. Oslovsky, J. Neyts, M. S. Drenichev, P. Leyssen, S. N. Mikhailov, *Eur. J. Med. Chem.* **2015**, *90*, 406-413.
[4] G. H. Hakimelahi, Z. A. Proba, K. K. Ogilvie, *Tet. Lett.* **1981**, *22*, 4775-4778.
[5] R. Schmieder, R. Edwards, *Bioinformatics* **2011**, *27*, 863-864.
[6] M. Martin, *EMBnet.journal; Vol 17, No 1: Next Generation Sequencing Data Analysis* **2011**. DOI: 10.14806/ej.17.1.200.
[7] K. K. Alam, J. L. Chang, D. H. Burke, *Molecular Therapy - Nucleic Acids* **2015**, *4*.
[8] U. Bodenhofer, E. Bonatesta, C. Horejš-Kainrath, S. Hochreiter, *Bioinformatics* **2015**, *31*, 3997-3999.
[9] O. Wagih, *Bioinformatics* **2017**, *33*, 3645-3647.

4. Paper-II RNA-Cleaving Deoxyribozymes Differentiate Methylated Cytidine Isomers in RNA

Reprinted with the permission from "Liaqat A, Sednev MV, Stiller C, Höbartner C. RNA-Cleaving Deoxyribozymes Differentiate Methylated Cytidine Isomers in RNA. *Angew Chem Int Ed Engl*. **2021**;60(35):19058-19062. doi:10.1002/anie.202106517"



RNA Modification Hot Paper

How to cite: *Angew. Chem. Int. Ed.* **2021**, *60*, 19058–19062
International Edition: doi.org/10.1002/anie.202106517
German Edition: doi.org/10.1002/ange.202106517

RNA-Cleaving Deoxyribozymes Differentiate Methylated Cytidine Isomers in RNA

*Anam Liaqat[+], Maksim V. Sednev[+], Carina Stiller, and Claudia Höbartner**

In memory of Professor Siegfried Hünig

Abstract: *Deoxyribozymes are emerging as modification-specific endonucleases for the analysis of epigenetic RNA modifications. Here, we report RNA-cleaving deoxyribozymes that differentially respond to the presence of natural methylated cytidines, 3-methylcytidine (m^3C), N^4-methylcytidine (m^4C), and 5-methylcytidine (m^5C), respectively. Using in vitro selection, we found several DNA catalysts, which are selectively activated by only one of the three cytidine isomers, and display 10- to 30-fold accelerated cleavage of their target m^3C-, m^4C- or m^5C-modified RNA. An additional deoxyribozyme is strongly inhibited by any of the three methylcytidines, but effectively cleaves unmodified RNA. The m^XC-detecting deoxyribozymes are programmable for the interrogation of natural RNAs of interest, as demonstrated for human mitochondrial tRNAs containing known m^3C and m^5C sites. The results underline the potential of synthetic functional DNA to shape highly selective active sites.*

Posttranscriptional nucleotide modifications are indispensable for the functional diversity of cellular RNA. More than 170 types of modified nucleotides are known,[1] but detection and quantification remain challenging. Transcriptome-wide maps are available for some modifications,[2] but the predicted sites and derivatization levels must be further validated by alternative methods.[3] Recently, we reported RNA-cleaving deoxyribozymes that reliably report the presence of N^6-alkyladenosines.[4] Deoxyribozymes responsive to N^6-methyladenosine (m^6A) have been used to interrogate m^6A levels in lncRNA, snoRNA and mRNA in DGACH sequence motifs,[4a,5] and the results were in agreement with those obtained by the current benchmark in the field (the SCARLET assay).[6] We found that DNA enzymes activated by m^6A reached 5–10-fold faster cleavage rates for modified vs. unmodified substrate.[4a] The bulkier and more hydrophobic N^6-isopentenyladenosine (i^6A) in the RNA had a much stronger activating effect on deoxyribozyme-catalyzed RNA cleavage, leading up to 2500-fold faster cleavage of the i^6A-modified versus unmodified RNA substrate.[4b]

The finding of m^6A- and i^6A-sensitive deoxyribozymes inspired the search for DNA catalysts that target other RNA modifications. Deoxyribozymes that are able to differentiate the three natural monomethylated cytidine isomers 3-methylcytidine (m^3C), N^4-methylcytidine (m^4C) and 5-methylcytidine (m^5C), and distinguish them from unmodified cytidine, are of particular interest. These methylated cytidines exhibit different structural properties depending on the location of the methyl group (Figure 1a). The methyl group in m^3C disrupts Watson–Crick base pairing and confers a positive charge to the nucleobase.[7] As a result, m^3C may promote the formation of alternative secondary structures. Similarly to m^6A, the methyl group in m^4C is attached to the exocyclic amino group, and depending on its conformation, m^4C can either retain (in *anti*-conformation) or disrupt Watson–Crick base pairing (in *syn*-conformation).[8] The *anti*-conformation is enforced in a base-paired structure but destabilized by steric repulsion (Figure 1a). In contrast to m^4C, the methyl group in m^5C does not disturb base pairing directly but may

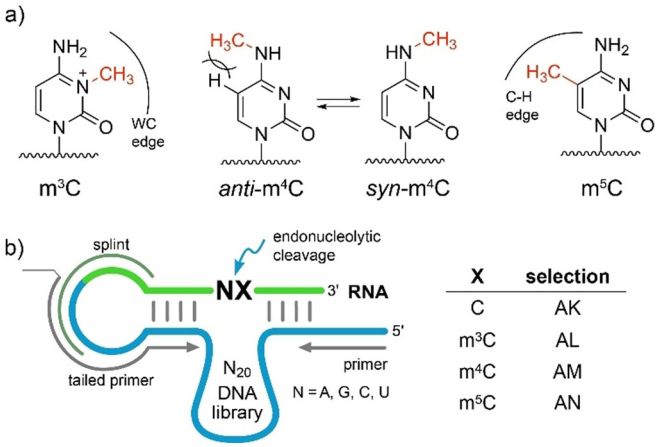


Figure 1. a) Structures of three natural methylated cytidines and their structural consequences. m^3C blocks the Watson–Crick (WC) edge and confers a positive charge to the nucleobase. m^4C retains WC base pairing in *anti*-conformation which is destabilized by steric clash between 5-H and N^4-Me. In *syn*-conformation, m^4C hinders WC base pairing. m^5C enhances hydrophobicity in the major groove and improves base stacking. b) Design of in vitro selection libraries. Splint for ligation and primer binding sites are indicated. Detailed selection is shown in Figure S1.

[*] A. Liaqat,[+] Dr. M. V. Sednev,[+] C. Stiller, Prof. Dr. C. Höbartner
Institute of Organic Chemistry, University of Würzburg
Am Hubland, 97074 Würzburg (Germany)
E-mail: claudia.hoebartner@uni-wuerzburg.de

[+] These authors contributed equally to this work.

Supporting information and the ORCID identification number(s) for the author(s) of this article can be found under: https://doi.org/10.1002/anie.202106517.

interact with neighboring nucleotides via the C–H-edge. The m^5C modification is known to enhance base stacking.[9] Targeting these methylated cytidine isomers also addresses fundamental questions on scope and plasticity of DNA catalysis.

To identify modification-specific catalysts, we performed gel-based in vitro selection experiments[4] (arbitrarily named AL, AM and AN, Figure 1b) using an N_{20} DNA library and three synthetic RNA substrates, each containing one of the three monomethylated cytidines (R2-R4, Table S1). Negative selection rounds were carried out with unmodified RNA (R1) to enhance the selectivity for the modified RNA. In this way, DNA enzymes that can cleave both unmodified and modified RNA were eliminated (Figure S1). In parallel, we sought DNA enzymes that preferentially cleave unmodified RNA and are inhibited by m^xC (AK selection). In this case, the positive selection was performed with R1 and the negative selection used an equimolar mixture of R2-R4. To evolve potentially universal DNA catalysts that respond to the cytidine modifications within all four possible dinucleotide sequence motifs NC, the RNA substrates contained a degenerate nucleotide upstream of the target cytidine.

After 18 rounds of selection, the enriched DNA pools were cloned, and 10 clones of each selection were sequenced. We identified 36 unique sequences (D6-D41, Table S2), suggesting a high diversity in the selection libraries. To obtain further insights into the library composition and to potentially identify further DNAzyme candidates, we applied next-generation sequencing (NGS) for a deeper analysis of AK, AL, AM and AN selection pools. In total, eight libraries were sequenced, including rounds 7 and 18 from each of the four in vitro selections. Deoxyribozymes specific for m^3C, m^4C or m^5C should be enriched in selection libraries AL, AM or AN, respectively, and depleted in selection AK. Conversely, the deoxyribozymes inhibited by a modification would be highly enriched in selection AK and depleted in selections AL, AM or AN. In this way, using $\log_2$ fold change in abundance ($\log_2 f_{CA}$) between selection rounds 7 and 18 we identified nine additional candidate DNA sequences (D42-D50, Table S2).[10]

Overall, 45 DNA sequences were then tested individually for their ability to cleave sixteen 3′-fluorescently labeled RNA substrates R5-R20 in independent cleavage reactions. The degenerate position was substituted with one of the four canonical ribonucleotides. Together R5-R20 represented every possible combination of four different junctions and four cytidine variants. After removing the 5′-overhang and 3′-loop region, 28 deoxyribozymes retained trans-cleaving activity (Table S3).

Using RNase T1 digestion and alkaline hydrolysis ladders, we determined the cleavage site for every active DNA enzyme. In accordance with the initial design, all tested deoxyribozymes cleaved the RNA substrate near the target cytidine (Table S3). The activity data partially supported our hypothesis that differences in enrichment in the selection libraries correlate with the specificity of deoxyribozymes. The correlation was good for DNAzymes found in the AK, AL and AM selections. However, poor correlation was observed for the selection AN, as the majority of the AN DNA enzyme candidates cleaved both m^5C-modified and unmodified RNA nearly to the same extent despite large differences in fold change abundance values (Table S3).

Interestingly, the inspection of predicted secondary structures revealed that none of the active deoxyribozymes contained the 8–17 motif, which had been repeatedly found in independent in vitro selections of RNA-cleaving deoxyribozymes.[4a,11] Its absence is especially surprising for AK selection, which used unmodified RNA as substrate. Probably, the 8–17 variants, that were initially present in the selection pools, were not particularly selective for either modified or unmodified RNA, and therefore were depleted during negative selection rounds. None of previously reported 8–17 variants was particularly efficient at cleaving N|C junctions.[11]

With respect to the degenerate position in the RNA substrate that was introduced to evolve for sequence generality, only deoxyribozyme AL112 distinguished a cytidine modification almost equally well in all four NC sequence contexts. On the other hand, many DNA catalysts were good at differentiating methylated from unmethylated RNA in one or two sequence contexts. In fact, we identified a corresponding deoxyribozyme for almost every combination of modification and junction (Table S3). For the UC junction in particular, we found that AL112, AM101, AN05 and AK104 (Figure 2a) deoxyribozymes together achieved efficient differentiation for all three monomethylated cytidines. Among these, AL112, AM101 and AN05 preferentially cleaved Um^3C-, Um^4C- and Um^5C-containing RNAs, respectively (fraction cleaved (FC) for the corresponding modified RNA $FC_{UmXC} \geq 70\,\%$, and unmodified RNA $FC_{UC} \leq 30\,\%$ after 6 h, see Figure 2b and Table S3), while AK104 efficiently cleaved only the UC substrate and was inhibited by all the modified nucleotides ($FC_{UC} = 85\,\%$, $FC_{UmXC} \approx 0$ after 6 h). Therefore, this set of deoxyribozymes was further characterized.

First, we confirmed the cleavage sites determined by the gel-based assay (Figure S3) by mass spectrometry. Consistent with the general mechanism proposed for DNA-catalyzed RNA cleavage,[12] the cleavage products for all four selected catalysts contained 2′,3′-cyclic phosphate and 5′-OH termini (Figure S4).

Deoxyribozyme AL112 preferentially cleaved m^3C-modified RNA in all four NC sequence contexts with the highest cleavage yield for the UC junction ($FC_{Um3C} = 70\,\%$, Figure S5). In the UC context, AL112 showed a ca. 30-fold higher k_{obs} value for m^3C-RNA compared to unmodified RNA (Figure 2b). The cleavage site was situated directly upstream of the modified nucleotide m^3C. Analysis of the predicted secondary structure showed that presence of m^3C might be required for disruption of the dG1-rC16 base pair that otherwise inhibited RNA cleavage for the NC-, Nm^4C- and Nm^5C-RNA substrates (Figure 2b, Table S3).

Deoxyribozyme AM101 was specific for m^4C and cleaved Um^4C-RNA ca. 10 times faster than UC-RNA. The cleavage site of AM101 was one nucleotide downstream of the modification site. Analysis of the secondary structure revealed the presence of a dG18-rU15 wobble base pair, which might be responsible for selectivity of AM101 for the UC junction (Figure 2a). However, in the CC sequence

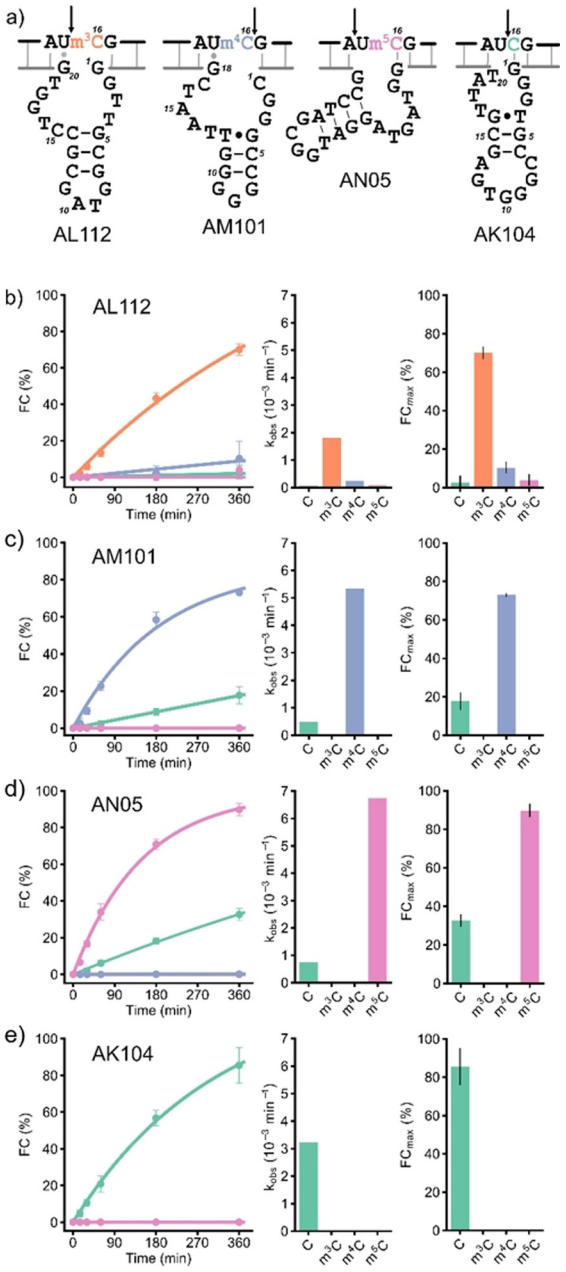


Figure 2. a) Sequences and predicted secondary structures of selected deoxyribozymes found in this study. The arrow indicates the cleavage site. Kinetic characterization of b) AL112, c) AM101, d) AN05, e) AK104 with four RNA substrates R8 (UC), R12 (Um^{3}C), R16 (Um^{4}C) and R20 (Um^{5}C). b)—e): Single-turnover kinetics plots (left). Observed rate constants k_{obs} (middle). Cleavage yields FC_{max} after 6 h incubation (right: mean of three independent experiments with error bars showing SD); representative gel images are in the supporting information). Reaction conditions: 1 μM RNA, 10 μM deoxyribozyme, 20 mM $MgCl_2$, 50 mM Tris-HCl, 150 mM NaCl, pH 7.5, 37 °C.

context, where formation of the stronger Watson–Crick base pair dG18-rC15 is expected, a significant decrease in cleavage activity was observed (Figure S6, Table S3). Interestingly, AM101 also discriminated against N^4,N^4-dimethylcytidine (m^{4,4}C), which was cleaved at nearly the same rate as unmodified RNA (Figure S7).

The deoxyribozyme AN05 for cleavage of m^{5}C-modified RNA was found in Sanger sequencing data of the AN selection. Analysis of NGS data showed that it was also present in the AK selection, where it was strongly depleted from round 7 to round 18 (Table S3). In the UC context, AN05 cleaved only m^{5}C-modified and unmodified substrates ($FC_{Um^5C} = 90\,\%$, $FC_{UC} = 33\,\%$), and was completely inhibited by m^{3}C and m^{4}C. The kinetic discrimination ability of AN05 was comparable to AM101: AN05 cleaved Um^{5}C-RNA ca. 10 times faster than the corresponding unmodified substrate (Figure 2b, S8, and Table S3). The cleavage site of AN05 was located one nucleotide upstream of the target cytosine. The predicted secondary structure contained the dG1-rC16 base pair that might stabilize an active site which is responsible for preference of AN05 to m^{5}C and C (Figure 2a). In case of m^{3}C and m^{4}C, this base pair might be destabilized or disrupted and thus lead to inhibition of cleavage.

The recognition of m^{5}C by AN05 raised the question about the oxidized variants 5-hydroxymethylcytosin (hm^{5}C) and 5-formylcytosin (f^{5}C), previously reported in RNA as products of TET2 and ALKBH1 enzymes.[13] AN05 cleaved hm^{5}C-containing RNA slightly slower but to a similar extent as m^{5}C RNA, while f^{5}C was comparable to unmodified RNA (Figure S9). Surprisingly, this result is in line with bisulfite sequencing, which is unable to distinguish m^{5}C from hm^{5}C, and f^{5}C from C.[14] Given the functional importance of hm^{5}C and f^{5}C in natural RNA,[13] future in vitro selection experiments may be directed to evolve deoxyribozymes specific for the individual oxidized m^{5}C variants.

Deoxyribozyme AK104 efficiently cleaved unmodified RNA (FC ≥ 85 % after 6 h) and was inhibited by all cytosine modifications in AC, GC and UC sequence contexts. In this case, the cleavage site was situated directly upstream of the target cytosine. However, when the RNA substrates with the CC junction were used, AK104 cleaved the m^{3}C-containing RNA with a moderate yield (FC_{Cm^3C} ca. 30 %) and was inactive with the unmodified substrate. The cleavage site for the Cm^{3}C-RNA was shifted one nucleotide upstream (Figure S10, Table S3). This odd behavior can be explained by disruption of the dG1-rC16 base pair by m^{3}C in Am^{3}C-, Gm^{3}C- and Um^{3}C-RNAs and formation of an alternative GC base pair between dG1 and rC15 in case of Cm^{3}C-RNA. The new base pair leads to repositioning of the AK104 active site and partial rescue of the cleavage activity. On the other hand, both rC15 and rC16 in the CC-substrate can form base pairs with dG2 and dG1, respectively, thereby disrupting the functional active site and abolish the catalytic activity of AK104. Thus, AK104 is a RNA-cleaving DNA enzyme for DC junctions (D = A,G,U), with high selectivity for unmodified RNA even under accelerated cleavage conditions at elevated Mg^{2+} concentrations. None of the modified RNAs was cleaved, and RNAs with rC16 changed to U, G, or A were also not cleaved (Figures S11, S12, S13).

Next, we examined the deoxyribozymes in the more challenging situation when all three RNA isomers (Um^{3}C-, Um^{4}C- and Um^{5}C-RNAs) are present at the same time. For this experiment, we labeled the three methylated UC-RNA substrates (R12, R16 and R20) at the 2′-OH of rA9 with fluorescent N^6-(6-aminohexyl)-ATP derivatives, carrying

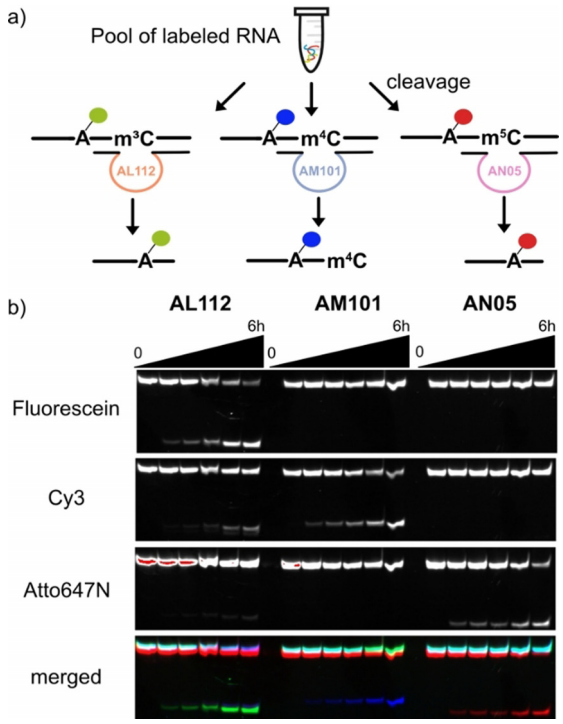


Figure 3. a) Schematic representation of selective RNA cleavage with AL112, AM101 and AN05. RNA substrates R12, R16 and R20 were separately labeled with 6FAM, Cy3 and Atto647N, respectively,[15] pooled and the mixture was subjected to cleavage with each of the three modification-specific DNAzymes. b) PAGE analysis of the experiment. Individual imaging channels for the three fluorophores are shown in black and white, and the merged image in false-color representation. Reaction conditions: 1 μM RNA mixture, 10 μM deoxyribozyme, 20 mM $MgCl_2$, 50 mM Tris-HCl, 150 mM NaCl, 37 °C.

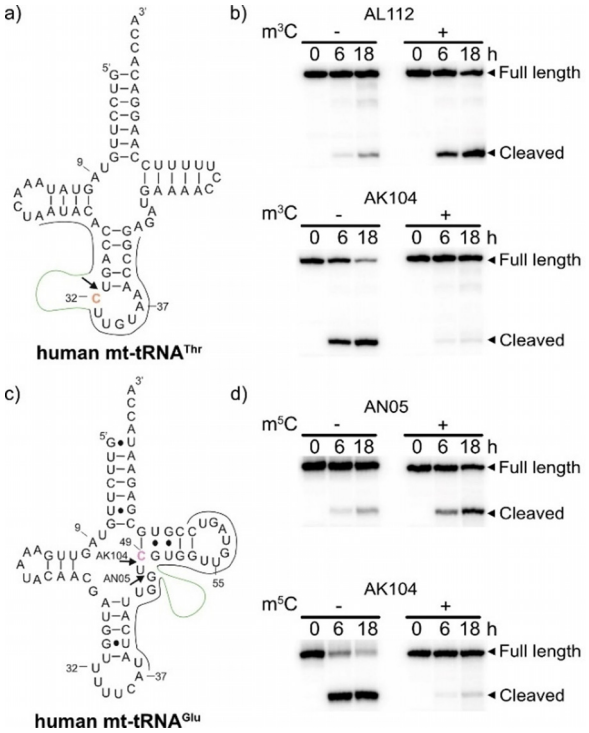


Figure 4. a) Secondary structure of mt-tRNA^{Thr} with target C32 in orange. b) PAGE analysis of mt-tRNA^{Thr} cleaved with AL112 and AK104. c) Secondary structure of mt-tRNA^{Glu} with target C49 in violet. d) PAGE analysis of mt-tRNA^{Glu} cleaved with AN05 and AK104. The cleavage reactions were performed at 37 °C. The arrows indicate the cleavage sites. Binding sites for DNAzymes are shown (binding arms in gray, catalytic core in green).

6FAM, Cy3 and ATTO-647N, respectively.[15] An equimolar mixture of the three labeled RNAs was then incubated with 10-fold excess of the AL112, AM101 or AN05 deoxyribozyme, respectively (Figure 3a). The cleavage reactions were analyzed by denaturing PAGE and imaged using a dedicated channel for each fluorescent label (Figure 3b). As expected, AL112 showed preferential cleavage of Um^{3}C-RNA with negligible signals in the m^{4}C- and m^{5}C-channels. At the same time, AM101 and AN05 cleaved only their cognate substrates with no detectable cleavage products in other channels. The results demonstrated that the specificity of each DNAzyme for its target modification was retained, and this collection of deoxyribozymes can be used to examine the identity of methylated cytidines in UC motifs. In this regard, m^{4}C-specific RNA cleavage is particularly important, as in contrast to m^{3}C and m^{5}C, there are no specific chemical reactions reported for the detection of m^{4}C in RNA.[16]

Next, we tested deoxyribozymes AL112, AN05 and AK104 on longer RNA substrates with complex secondary structures. To this end, we chose human mt-tRNA^{Thr}, which contains a known m^{3}C in the anticodon loop (C32, Figure 4a),[17] and mt-tRNA^{Glu} with a known m^{5}C at C49 (Figure 4c).[17] The single-modified tRNAs were assembled from synthetic RNA fragments by splinted ligation with T4 RNA ligase, and the binding arms of deoxyribozymes were designed to cleave the tRNAs upstream of the corresponding modification sites.[18] The results of the cleavage reactions for unmodified and modified tRNAs were compared (Figure 4b,d). AK104 cleaved both unmodified tRNAs with a high yield (>90% after 18 h) and was nearly completely inhibited by m^{3}C and by m^{5}C. AL112 and AN05 cleaved m^{3}C-modified mt-tRNA^{Thr} and m^{5}C-modified mt-tRNA^{Glu}, respectively, both somewhat less efficiently than their cognate substrates R12 and R20 (58% and 42% after 18 h), but the unmodified tRNAs did not undergo any significant cleavage. Therefore, AL112 and AN05 can faithfully identify m^{3}C- and m^{5}C-modified tRNA, respectively.

In summary, we have identified and characterized a series of RNA-cleaving deoxyribozymes that differentially respond to the presence of methylated cytidines in RNA. On the one hand, the methyl groups in m^{3}C, m^{4}C and m^{5}C can dramatically inhibit deoxyribozyme-catalyzed RNA cleavage (as in case of AK104). On the other hand, the presence of methylcytidines can lead to accelerated cleavage of modified RNA (as in case of AL112, AM101 and AN05). Owing to these kinetic differences, new deoxyribozymes can be used to examine the identity of methylcytidines and the degree of their methylation at selected target sites. The effect of the methyl group on catalytic activity might be related to steric, electrostatic (m^{3}C) or hydrophobic interaction of the modified nucleotide with its neighbors and Mg^{2+} in the active site.

The detailed structural basis for the selective recognition of methylcytidines remains to be identified.

As the enriched libraries for selection of methylcytidine-sensitive deoxyribozymes remained highly diverse even after 18 rounds of selection, we expect that many DNA enzymes with potentially interesting properties remained undiscovered. To this end, massively parallel activity assays that are able to screen cleavage reactions of thousands of candidate deoxyribozymes are required. Given the recent success in using deep sequencing for large-scale characterization of (deoxy)ribozyme kinetics,[19] new catalytic DNA motifs can likely be revealed in the future. The related experiments are currently ongoing in our laboratory and expected to uncover hidden treasures in in vitro selection libraries.

Acknowledgements

This work was supported by the DFG (SPP1784) and by the ERC (grant no. 682586). A.L. acknowledges funding by a PhD scholarship from the German Academic Exchange Service (DAAD). M.V.S. thanks the Graduate School of Life Sciences at the University of Würzburg for a Postdoc Plus fellowship. Illumina sequencing was performed at the Core Unit Systems Medicine at the University of Würzburg. Open access funding was enabled and organized by Projekt DEAL.

Conflict of Interest

The authors declare no conflict of interest.

Keywords: deoxyribozymes · epitranscriptomics · in vitro selection · RNA modification · site-specific RNA cleavage

[1] a) P. Boccaletto, M. A. Machnicka, E. Purta, P. Piątkowski, B. Bagiński, T. K. Wirecki, V. de Crécy-Lagard, R. Ross, P. A. Limbach, A. Kotter, M. Helm, J. M. Bujnicki, *Nucleic Acids Res.* **2018**, *46*, D303 – D307; b) P. J. McCown, A. Ruszkowska, C. N. Kunkler, K. Breger, J. P. Hulewicz, M. C. Wang, N. A. Springer, J. A. Brown, *Wiley Interdiscip. Rev. RNA* **2020**, *11*, e1595.

[2] J.-J. Xuan, W.-J. Sun, P.-H. Lin, K.-R. Zhou, S. Liu, L.-L. Zheng, L.-H. Qu, J.-H. Yang, *Nucleic Acids Res.* **2018**, *46*, D327 – D334.

[3] a) M. Helm, Y. Motorin, *Nat. Rev. Genet.* **2017**, *18*, 275 – 291; b) B. Linder, S. R. Jaffrey, *Cold Spring Harbor Perspect. Biol.* **2019**, *11*, a032201.

[4] a) M. V. Sednev, V. Mykhailiuk, P. Choudhury, J. Halang, K. E. Sloan, M. T. Bohnsack, C. Höbartner, *Angew. Chem. Int. Ed.* **2018**, *57*, 15117 – 15121; *Angew. Chem.* **2018**, *130*, 15337 – 15341; b) A. Liaqat, C. Stiller, M. Michel, M. V. Sednev, C. Höbartner, *Angew. Chem. Int. Ed.* **2020**, *59*, 18627 – 18631; *Angew. Chem.* **2020**, *132*, 18786 – 18790; c) R. Micura, C. Höbartner, *Chem. Soc. Rev.* **2020**, *49*, 7331 – 7353.

[5] M. Bujnowska, J. Zhang, Q. Dai, E. M. Heideman, J. Fei, *J. Biol. Chem.* **2020**, *295*, 6992 – 7000.

[6] N. Liu, M. Parisien, Q. Dai, G. Zheng, C. He, T. Pan, *RNA* **2013**, *19*, 1848 – 1856.

[7] a) L. Flemmich, S. Heel, S. Moreno, K. Breuker, R. Micura, *Nat. Commun.* **2021**, *12*, 3877; b) S. Mao, P. Haruehanroengra, S. V. Ranganathan, F. Shen, T. J. Begley, J. Sheng, *ACS Chem. Biol.* **2021**, *16*, 76 – 85.

[8] a) R. Micura, W. Pils, C. Höbartner, K. Grubmayr, M. O. Ebert, B. Jaun, *Nucleic Acids Res.* **2001**, *29*, 3997 – 4005; b) S. Mao, B. Sekula, M. Ruszkowski, S. V. Ranganathan, P. Haruehanroengra, Y. Wu, F. Shen, J. Sheng, *Nucleic Acids Res.* **2020**, *48*, 10087 – 10100.

[9] a) L. Trixl, A. Lusser, *Wiley Interdiscip. Rev. RNA* **2019**, *10*, e1510; b) E. M. Harcourt, A. M. Kietrys, E. T. Kool, *Nature* **2017**, *541*, 339 – 346; c) S. Wang, E. T. Kool, *Biochemistry* **1995**, *34*, 4125 – 4132.

[10] We picked the candidates with the largest differences between $\log_2 f_{CA}$ and $\log_2 f_{CA,neg}$ (see Table S3).

[11] a) S. W. Santoro, G. F. Joyce, *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 4262 – 4266; b) J. Li, W. Zheng, A. H. Kwon, Y. Lu, *Nucleic Acids Res.* **2000**, *28*, 481 – 488; c) R. P. G. Cruz, J. B. Withers, Y. Li, *Chem. Biol.* **2004**, *11*, 57 – 67; d) K. Schlosser, Y. Li, *Biochemistry* **2004**, *43*, 9695 – 9707; e) J. C. F. Lam, J. B. Withers, Y. Li, *J. Mol. Biol.* **2010**, *400*, 689 – 701.

[12] a) R. R. Breaker, G. F. Joyce, *Chem. Biol.* **1994**, *1*, 223 – 229; b) S. K. Silverman, *Nucleic Acids Res.* **2005**, *33*, 6151 – 6163; c) K. Schlosser, Y. Li, *ChemBioChem* **2010**, *11*, 866 – 879.

[13] a) C. He, J. Bozler, K. A. Janssen, J. E. Wilusz, B. A. Garcia, A. J. Schorn, R. Bonasio, *Nat. Struct. Mol. Biol.* **2021**, *28*, 62 – 70; b) S. Haag, K. E. Sloan, N. Ranjan, A. S. Warda, J. Kretschmer, C. Blessing, B. Hubner, J. Seikowski, S. Dennerlein, P. Rehling, M. V. Rodnina, C. Höbartner, M. T. Bohnsack, *EMBO J.* **2016**, *35*, 2104 – 2119.

[14] a) C.-X. Song, C. Yi, C. He, *Nat. Biotechnol.* **2012**, *30*, 1107 – 1116; b) Y. Dai, B.-F. Yuan, Y.-Q. Feng, *RSC Chem. Biol.* **2021**, https://doi.org/10.1039/D1CB00022E.

[15] RNA labeling was performed with FH14 ribozyme: M. Ghaem Maghami, C. P. M. Scheitl, C. Höbartner, *J. Am. Chem. Soc.* **2019**, *141*, 19546 – 19549.

[16] a) I. Behm-Ansmant, M. Helm, Y. Motorin, *J. Nucleic Acids* **2011**, 408053; b) M. Heiss, S. Kellner, *RNA Biol.* **2017**, *14*, 1166 – 1174.

[17] T. Suzuki, Y. Yashiro, I. Kikuchi, Y. Ishigami, H. Saito, I. Matsuzawa, S. Okada, M. Mito, S. Iwasaki, D. Ma, X. Zhao, K. Asano, H. Lin, Y. Kirino, Y. Sakaguchi, T. Suzuki, *Nat. Commun.* **2020**, *11*, 4269.

[18] To facilitate access of the deoxyribozymes, disruptor oligonucleotides were included complementary to the 5′ or 3′ end of the acceptor stem, respectively (see D56 and D59 in the Supporting Information, Table S2).

[19] a) S. Kobori, Y. Yokobayashi, *Angew. Chem. Int. Ed.* **2016**, *55*, 10354 – 10357; *Angew. Chem.* **2016**, *128*, 10510 – 10513; b) V. Dhamodharan, S. Kobori, Y. Yokobayashi, *ACS Chem. Biol.* **2017**, *12*, 2940 – 2945; c) A. D. Pressman, Z. Liu, E. Janzen, C. Blanco, U. F. Müller, G. F. Joyce, R. Pascal, I. A. Chen, *J. Am. Chem. Soc.* **2019**, *141*, 6213 – 6223; d) J. O. L. Andreasson, A. Savinov, S. M. Block, W. J. Greenleaf, *Nat. Commun.* **2020**, *11*, 1663; e) Y. Yokobayashi, *Acc. Chem. Res.* **2020**, *53*, 2903 – 2912; f) Y. Shen, A. Pressman, E. Janzen, I. A. Chen, *Nucleic Acids Res.* **2021**, https://doi.org/10.1093/nar/gkab199.

Manuscript received: May 14, 2021
Revised manuscript received: June 20, 2021
Accepted manuscript online: June 29, 2021
Version of record online: July 26, 2021

SUPPORTING INFORMATION

Supporting Information

69451 Weinheim, Germany

RNA-cleaving deoxyribozymes differentiate methylated cytidine isomers in RNA

Anam Liaqat, Maksim V. Sednev, Carina Stiller and Claudia Höbartner*

Abstract: Deoxyribozymes are emerging as modification-specific endonucleases for the analysis of epigenetic RNA modifications. Here, we report RNA-cleaving deoxyribozymes that differentially respond to the presence of natural methylated cytidines, 3-methylcytidine (m^3C), N^4-methylcytidine (m^4C), and 5-methylcytidine (m^5C), respectively. Using *in vitro* selection, we found several DNA catalysts, which are selectively activated by only one of the three cytidine isomers, and display 10- to 30-fold accelerated cleavage of their target m^3C-, m^4C- or m^5C-modified RNA. An additional deoxyribozyme is strongly inhibited by any of the three methylcytidines, but effectively cleaves unmodified RNA. The m^XC-detecting deoxyribozymes are programable for the interrogation of natural RNAs of interest, as demonstrated for human mitochondrial tRNAs containing known m^3C and m^5C sites. The results underline the potential of synthetic functional DNA to shape highly selective active sites.

DOI: 10.1002/anie.2021XXXXX

SUPPORTING INFORMATION

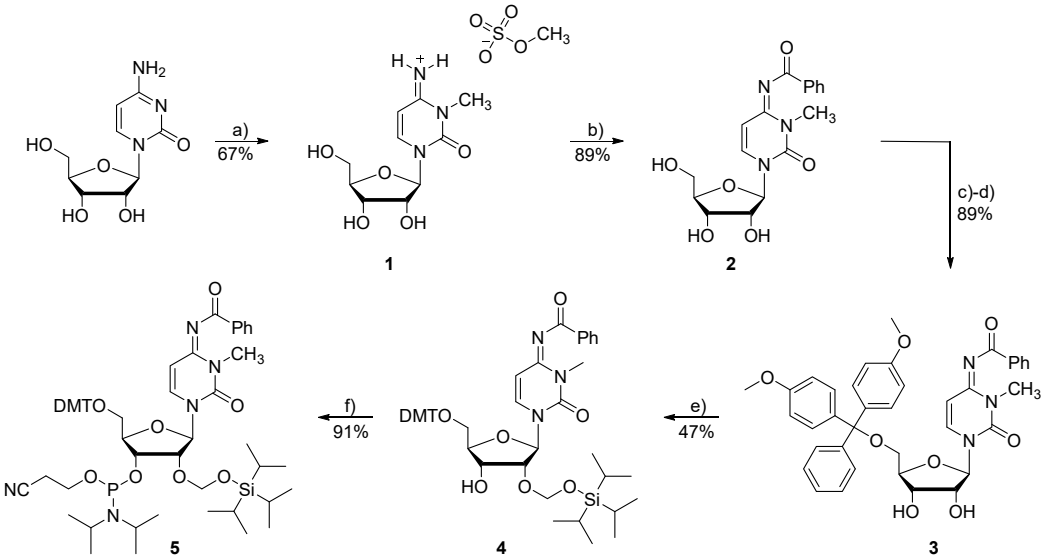
Table of Contents

General information

DNA oligonucleotides were purchased from Microsynth. RNA oligonucleotides R1−R26 (Table S1) were prepared by solid-phase synthesis using 5'-*O*-DMT-2'-*O*-TOM-protected 3'-β-cyanoethyl phosphoramidites of N^6-acetyladenosine, N^4-acetylcytidine, N^2-acetylguanosine and uridine (purchased from Chemgenes), N^4-methylcytidine (synthesized according to the known procedure[1]), N^3-methylcytosine (prepared using the procedure described below), 5'-*O*-DMT-2'-*O*-TBDMS-N^4-benzoyl-5-methylcytidine 3'-β-cyanoethyl phosphoramidite (from Chemgenes), and (2-cyanoethoxy)-5-hexyn-1-yl-*N*,*N*-diisopropylaminophosphine (prepared according to the known procedure[2]). N^4,N^4-Dimethylcytidine-modified RNA oligonucleotide R27 was prepared by solid-phase synthesis using 2'-*O*-TOM-protected ribonucleoside phosphoramidites for unmodified positions and "convertible" 5'-*O*-(4,4'-Dimethoxytrityl)-2'-*O*-(triisopropylsilyl)oxymethyl-O^4-(4-chlorophenyl)uridine 3'-(2-cyanoethyl-*N*,*N*-diisopropylphosphoramidite) (prepared according to the known procedure[3]) for the modified position followed by conversion of the resulting O^4-chlorophenyluridine-modified RNA with *N,N*-dimethylamine (see below for the conditions). RNA oligonucleotides R28−R34 were prepared by in vitro transcription from the corresponding DNA templates. Fluorescein thiosemicarbazide was purchased from Sigma-Aldrich. 6-FAM-azide, N^6-(6-aminohexyl)-ATP-6-FAM, N^6-(6-aminohexyl)-ATP-Cy3, N^6-(6-aminohexyl)-ATP-ATTO-647N, nucleotide triphosphates (NTPs) and deoxyribonucleotide triphosphates (dNTPs) were purchased from Jena Bioscience. m^5CTP, hm^5CTP and f^5CTP were purchased from TriLink Biotechnologies. T4 PNK, T4 DNA Ligase, and DreamTaq polymerase were purchased from Thermo Fisher Scientific. γ-^{32}P-Adenosine 5'-triphosphate (γ-^{32}P-ATP) was purchased from Hartmann Analytic. All DNA and RNA oligonucleotides were purified by denaturing PAGE (10–20% acrylamide/bis-acrylamide 19:1, 7 M urea) with running buffer 1x TBE (89 mM Tris, 89 mM boric acid and 2 mM EDTA, pH 8.3), extracted by crush and soak into TEN buffer (10 mM Tris-HCl, pH 8.0, 1 mM EDTA, 300 mM NaCl) and recovered by precipitation with ethanol. For preparative gels, oligonucleotides were visualized by UV shadowing on a TLC plate. Fluorescence gel images were recorded with a Biorad ChemiDoc MP using epi illumination. Analytical anion-exchange HPLC was carried out on a GE Healthcare Life Sciences ÄKTA™ micro system using DNAPac™ PA200 columns (4 x 250 mm, Thermo Scientific) at 60 °C. Solvent A: 25 mM Tris-HCl (pH 8.0), 6 M Urea. Solvent B: 25 mM Tris-HCl (pH 8.0), 6 M Urea, 0.5 M $NaClO_4$. Gradient: linear, 0–54% solvent B, 4% solvent B per 1 CV. The solvents were filtered through 0.2 µm cellulose acetate filters (Sartorius, Germany) prior to usage. HR-ESI-MS spectra of the synthetic products and oligonucleotides were recorded on a Bruker micrOTOF-Q III spectrometer.

Synthesis of m^3C phosphoramidite

All synthetic procedures were performed under nitrogen atmosphere. The chemicals were purchased 'Pro analysis'- or 'For synthesis' grade and used without additional purification. Dry solvents (DCM, MeCN, DMF, THF) were obtained in a solvent purification system (SPS). DCE was dried over molecular sieves. Solvents for column chromatography (technical quality) were distilled prior to use. The obtained compounds were stored at 4 °C (-20 °C in case of TOM-protected compounds and the final phosphoramidite). Column chromatography was performed on silica gel (Kieselgel 60, Merck, 0.040-0.063 mm). TLC was performed on Alugram DIL G/UV254 (Machery-Nagel, UV visualization, 254 nm). NMR spectra were recorded on a Bruker Avance III HD 400 spectrometer. Chemical shifts are reported relative to the residual solvent signals. $CDCl_3$: 7.26 ppm (^{1}H), 77.16 ppm (^{13}C). DMSO: 2.50 ppm (^{1}H), 39.52 ppm (^{13}C).



Scheme S1. Synthesis of m^3C phosphoramidite. a) DMF, dimethyl sulfate, 37 °C, 30 min. b) TMSCl, pyridine, 1 h, r.t., then NEt_3, BzCl, 4 h, r.t. c) DMF-DMA, pyridine, o.n., r.t. d) DMT-Cl, pyridine, r.t., 2 h. e) iPr$_2$NEt, tBu$_2$SnCl$_2$, DCE, r.t., 1 h, then TOM-Cl, 80 °C, 20 min. f) Me_2NEt, CEP-Cl, DCM, r.t.

SUPPORTING INFORMATION

The synthesis of m^{3}C-phosphoramidite building block **5** was performed in five steps. The methyl group was introduced at position N3 of cytosine with dimethyl sulfate,[4] resulting in formation of salt **1**. The transient protection of the ribose hydroxy groups with TMSCl, allowed selective protection of the amino group with benzoyl chloride to obtain compound **2**.[5] In the next step, the 2'- and 3'-hydroxy groups of compound **2** were transiently protected as acetal with DMF-DMA, and the 5'-hydroxy group was selectively protected upon reaction with DMT-Cl. Compound **3** was then activated with *t*Bu_2SnCl_2 and treated with TOM-Cl to yield compound **4**. Finally, compound **4** was treated with CEP-Cl to afford m^{3}C phosphoramidite **5**.

N^3-Methylcytidinum methyl sulfate (**1**)

As described earlier,[4] cytidine (500 mg, 2.06 mmol, 1.00 eq.) was suspended in anhydrous dimethylformamide (5 mL) and warmed to 37 °C under nitrogen atmosphere. Afterwards, dimethyl sulfate (2.00 mL, 21.1 mmol, 10.2 eq.) was added. After 30 min stirring, the resulting clear solution was diluted to 20 mL with methanol, and ethyl acetate was added until a precipitate formed. The title compound was obtained by filtration as a colorless solid (510 mg, 1.38 mmol, 67%).
^{1}H NMR (MeOD, 400 MHz): δ (ppm) = 3.50 (s, 3H, NCH_3), 3.68 (s, 3H, $^-OS(O)_2OCH_3$), 3.78 (dd, J = 12.4, 2.5 Hz; 1H, H^1-5'), 3.94 (dd, J = 12.4, 2.5 Hz; 1H, H^2-5'), 4.08 (dt, J = 6.1, 2.5 Hz; 1H, H-4'), 4.15 (dd, J = 6.1, 4.9 Hz; 1H, H-3'), 4.18 (dd, J = 4.9, 2.7 Hz; 1H, H-2'), 5.87 (d, J = 2.7 Hz; 1H, H-1'), 6.16 (d, J =7.9 Hz; 1H, H-5), 8.50 (d, J = 7.9 Hz; 1H, H-6).
^{13}C NMR (MeOD): δ (ppm) = 30.93 (N-CH_3), 55.11 ($^-OS(O)_2OCH_3$), 61.15 (C5'), 70.09 (C3'), 76.28 (C2'), 86.22 (C4'), 93.05 (C1'), 94.75 (C5), 143.57 (C6), 149.13 (C=O), 161.29 (C=NH_2^+).
HR-MS (ESI$^+$): Exact mass calculated for $C_{10}H_{16}N_3O_5$ [M+H]$^+$: 258.1084, found: 258.1086.

N^4-Benzoyl-*N*^{3}-methylcytidine (**2**)

N^3-Methylcytidinium methyl sulfate **1** (395 mg, 1.07 mmol, 1.00 eq.) was treated with trimethylsilyl chloride (TMSCl, 528 µL, 4.17 mmol, 3.90 eq.) in anhydrous pyridine (7 mL) under nitrogen atmosphere. After stirring for 1 h, triethylamine (297 µL, 2.14 mmol, 2.00 eq.) was added, and the reaction mixture was stirred for further 5 min at room temperature. Afterwards, benzoyl chloride (149 µL, 1.28 mmol, 1.20 eq.) was added dropwise. After stirring for 4 h at room temperature, the reaction mixture was quenched with water (1.5 mL). The resulting mixture was stirred for 5 min at room temperature followed by addition of aqueous ammonia (2.6 mL). After further 15 min stirring, the volatiles were removed under reduced pressure. The crude residue was purified by column chromatography (DCM/MeOH, 20:1) to give the title compound as a colorless foam (344 mg, 952 µmol, 89%).
^{1}H NMR (MeOD, 400 MHz): δ (ppm) = 3.54 (s, 3H, N-CH_3), 3.74 (dd, J = 12.3, 2.9 Hz; 1H, H^1-5'), 3.86 (dd, J = 12.3, 2.9 Hz; H^2-5'), 4.02 (dt, J = 5.2, 2.9 Hz; 1H, H-4'), 4.15 (t, J = 5.2 Hz; 1H, H-3'), 4.19 (dd, J = 5.2, 4.0 Hz; 1H, H-2'), 5.94 (d, J = 4.0 Hz; 1H, H-1'), 6.41 (d, J = 8.2 Hz; 1H, H-5), 7.42–7.48 (m, 2H, bz), 7.53–7.58 (m, 1H, bz), 7.93 (d, J = 8.2 Hz; 1H, H-6), 8.09-8.13 (m, 2H, bz).
^{13}C NMR (MeOD, 100 MHz): δ (ppm) = 30.48 (N-CH_3), 61.99 (C5'), 70.96 (C3'), 75.94 (C2'), 86.19 (C4'), 91.79 (C1'), 98.54 (C5), 129.34 (bz), 130.72 (bz), 133.71 (bz), 137.22 (C_q-bz), 138.80 (C6), 151.94 (C2), 158.18 (C4), 179.35 (C=O).
HR-MS (ESI$^+$): Exact mass calculated for $C_{17}H_{19}N_3NaO_6$ [M+Na]$^+$: 384.1166, found: 384.1171.

N^4-Benzoyl-5'-*O*-(4,4'-dimethoxytrityl)-*N*^{3}-methylcytidine (**3**)

Compound **2** (265 mg, 734 µmol, 1.00 eq.) was co-evaporated with pyridine (2 mL) and re-dissolved in anhydrous pyridine (2.5 mL) under nitrogen atmosphere. *N*,*N*-dimethylformamide dimethyl acetal (DMF-DMA, 175 mg, 196 µL, 1.47 mmol, 2.00 eq.) was added. The resulting mixture was stirred for 18 h at room temperature and then evaporated to dryness. The residue was re-dissolved in dry pyridine (2.5 mL) and 4,4'-dimethoxytrityl chloride (298 mg, 881 µmol, 1.20 eq.) was added in one portion. After stirring for 2 h at room temperature, the reaction mixture was quenched with methanol (3 mL). The reaction mixture was evaporated to dryness and then co-evaporated with toluene (3 mL). The residue was diluted with dichloromethane (30 mL) and then washed with aqueous citric acid (5%), water and saturated aqueous $NaHCO_3$. The organic layer was dried with Na_2SO_4 and concentrated under reduced pressure. The crude product was purified by column chromatography (DCM/MeOH, 80:1 with 2% v/v NEt_3) to give the title compound as a colorless foam (433 mg, 653 µmol, 89%).
^{1}H NMR ($CDCl_3$, 400 MHz): δ (ppm) = 3.38 (dd, J = 11.0, 3.0 Hz; 1H, H^1-5'), 3.47 (dd, J = 11.0, 3.0 Hz; 1H, H^2-5'), 3.56 (s, 3H, N-CH_3), 3.77 (d, J = 2.2 Hz; 6H, OCH_3), 4.23 (dt, J = 5.0, 3.0 Hz; 1H, H-4'), 4.28 (dd, J = 5.0, 3.6 Hz; 1H, H-2'), 4.35 (t, J = 5.0 Hz; 1H, H-3'), 5.85 (d, J = 3.6 Hz; 1H, H-1'), 6.21 (d, J = 8.2 Hz; H-5), 6.78-6.83 (m, 4H, trityl), 7.16-7.21 (m, 1H, trityl), 7.21-7.28 (m, 6H, trityl), 7.32-7.35 (m, 2H, trityl), 7.41-7.46 (m, 2H, bz), 7.50-7.55 (m, 1H, bz), 7.66 (d, J = 8.2 Hz; H-6), 8.11-8.14 (m, 2H, bz).
^{13}C NMR ($CDCl_3$, 100 MHz): δ (ppm) = 30.17 (N-CH_3), 55.37 (OCH_3), 62.48 (C5'), 70.86 (C3'), 76.41 (C2'), 84.73 (C4'), 87.14 (C_q-trityl), 91.85 (C1'), 98.24 (C5), 113.40, 113.51, 113.53 (trityl), 127.24 (trityl), 128.11 (trityl), 128.22 (trityl), 128.32 (bz), 129.83 (bz), 130.12, 130.16 (trityl), 132.55 (bz), 135.30 (C_q-trityl), 135.74 (C6), 136.02 (C_q-bz), 144.18 (C_q-trityl), 151.27 (C2), 155.86 (C4), 158.77 (C_q-OCH_3, trityl), 177.43 (*C*=O, bz).
HR-MS (ESI$^+$): Exact mass calculated for $C_{38}H_{37}N_3NaO_8$ [M+Na]$^+$: 686.2473, found: 686.2467.

N^4-Benzoyl-5'-*O*-(4,4'-dimethoxytrityl)-N^3-methyl-2'-*O*-{[(triisopropylsilyl)oxy]methyl}cytidine (**4**)

To a solution of compound **3** (345 mg, 520 µmol, 1.00 eq.) in anhydrous 1,2-dichloroethane (4 mL), *N,N*-diisopropylethylamine (311 µL, 1.82 mmol, 3.50 eq.) and di-*tert*-butyltin dichloride (174 mg, 572 µmol, 1.1 eq.) were added. After stirring at room temperature for 1 h, the reaction mixture was heated to 80 °C, and TOM-Cl (150 mg, 676 µmol, 1.30 eq.) was added. After 20 min, the reaction mixture was cooled down to room temperature and diluted with dichloromethane (40 mL). The organic layer was separated, washed with saturated aqueous $NaHCO_3$ (3 x 20 mL) and dried over Na_2SO_4. The crude compound was purified by column chromatography (hexane/EtOAc, 5:1 with 2% v/v NEt_3), to give the title compound as a colorless foam (207 mg, 243 µmol, 47%).
^{1}H NMR ($CDCl_3$, 400 MHz): δ (ppm) = 1.05-1.13 (m, 21H, $Si(CH(CH_3)_2)_3$), 3.20 (d, J = 6.4 Hz; 1H, C3'-O*H*), 3.48-3.50 (m; 2H, H-5'), 3.54 (s, 3H, N-CH_3), 3.78 (d, J = 1.0 Hz; 6H, OCH_3), 4.09 (dt, J = 6.6, 2.5 Hz; 1H, H-4'), 4.21 (dd, J = 5.0, 2.9 Hz; 1H, H-2'), 4.39-4.48 (m, 1H, H-3'), 5.06 (d, J = 4.7 Hz; 1H, OCH_2O), 5.25 (d, J = 4.7 Hz; 1H, OCH_2O), 6.04 (d, J = 2.9 Hz; 1 H, H-1'), 6.06 (d, J = 8.2 Hz; H-5), 6.80-6.84 (m, 4H, trityl), 7.17-7.22 (m, 1H, trityl), 7.23-7.29 (m, 6H, trityl, bz), 7.34-7.38 (trityl), 7.41-7.46 (bz), 7.50-7.55 (m, 1H, bz), 7.79 (d, J = 8.2 Hz; H-6), 8.11-8.14 (m, 1H, bz).
^{13}C NMR ($CDCl_3$, 100 MHz): δ (ppm) = 11.99, 17.94 ($Si(CH(CH_3)_2)_3$), 30.02 (N-CH_3), 55.36 (OCH_3), 61.98 (C5'), 69.10 (C3'), 83.27 (C2'), 83.66 (C4'), 87.15 (C_q-OCH_2), 88.72 (C1'), 90.80 (OCH_2O), 98.29 (C5), 113.40 (trityl), 127.24 (trityl), 128.13 (trityl), 128.30, 128.35 (trityl, bz), 129.82 (bz), 130.15, 130.24 (trityl), 132.50 (bz), 135.34, 135.64 (trityl, C_q), 135.88 (C6), 136.07 (bz, C_q), 144.18 (trityl, C_q), 150.40 (NC(=O)N), 155.83 (C=N), 158.77 (trityl, C_q-OCH_3), 177.44 (C=O, bz).
HR-MS (ESI$^+$): Exact mass calculated for $C_{48}H_{59}N_3NaO_9Si$ $[M+Na]^+$: 872.3913, found: 872.3895.

N^4-Benzoyl-5'-*O*-(4,4'-dimethoxytrityl)-N^3-methyl-2'-*O*-{[(triisopropylsilyl)oxy]methyl}cytidine 3'-*O*-(2-cyanoethyl diisopropylphosphoramidite) (**5**)

To a solution of compound **4** (188 mg, 221 µmol, 1.00 eq.) in anhydrous dichloromethane (4 mL), *N,N*-dimethylethylamine (240 µL, 2.21 mmol, 10.0 eq.) and 2-cyanoethyl *N,N*-diisopropylchlorophosphoramidite (78.6 mg, 332 µmol, 1.50 eq.) were added, and the reaction mixture was stirred for 2.5 h at room temperature. Afterwards, the reaction mixture was evaporated to dryness under reduced pressure. The crude product was purified by column chromatography (hexane/EtOAc, 3:1 with 2% v/v NEt_3) to yield the title compound as a colorless foam (211 mg, 201 µmol, 91%).
^{1}H NMR ($CDCl_3$, 400 MHz): δ (ppm) = 0.99-1.08 (m, 42H, CH_3 (isopropyl-N); $Si(CH(CH_3)_2)_3$, diast.), 1.14-1.19 (m, 18H, CH_3 (isopropyl-N), diast.) 2.37 (td, 2H, J = 6.5, 2.0 Hz; $POCH_2CH_2$, diast.), 2.64 (td, 2H, J = 6.1, 4.0 Hz; $POCH_2CH_2$, diast.), 3.38 (td, 2H, J = 10.6, 2.90 Hz; H-5', diast.), 3.53 (s, 3H, diast., N-CH_3), 3.53 (s, 3H, N-CH_3), 3.53-3.60 (m, 4H, diast., H-5', N($CH(CH_3)_2)_2$ and H^2, $POCH_2$), 3.61-3.68 (m, 1H, H^2, $POCH_2$), 3.78 (d, 5H, diast., OCH_3), 3.79 (d, 6H, J = 0.8 Hz; OCH_3), 3.81-3.88 (m, 1H, H^1, $POCH_2$), 3.94 (ddt, 1H, J = 10.5, 7.5, 6.1 Hz; diast., H^1, $POCH_2$), 4.20 (dt, 1H, J = 4.7, 2.7 Hz; diast., H-4'), 4.27 (dd, 1H, J = 5.3, 2.5 Hz; H-4'), 4.39-4.42 (m, diast. H-3', H-2'), 4.35-4.38 (m, 1H, H-2'), 4.45 (dt, 1H, J = 10.3, 5.4 Hz; H-3'), 5.03 (d, 1H, J = 4.8 Hz; H^2, OCH_2O), 5.05 (d, 1H, diast. H^2, OCH_2O), 5.09 (d, 2H, J = 4.8 Hz, H^1, OCH_2O, diast.), 6.08 (d, 1H, diast., H-5), 6.13 (d, 1H, J = 8.2 Hz, H-5), 6.16 (d, 1H, J = 4.2 Hz; H-1'), 6.18 (d, 1H, diast., H-1'), 6.79-6.85 (m, 8H, trityl, diast.), 7.17-7.23 (m, 2H, trityl, diast.), 7.24-7-30 (m, 7H, trityl), 7.33-7.40 (m, 4H, trityl, diast.), 7.41-7.46 (m, 4H, bz, diast.), 7.50-7.55 (m, 2H, bz, diast.), 7.66 (d, 1H, diast., J = 8.2 Hz; H-6), 7.74 (d, 1H, J = 8.2 Hz; H-6), 8.11-8.15 (m, 4H, bz, diast.).
^{13}C NMR ($CDCl_3$, 100 MHz): δ (ppm) = 12.1, 18.0 ($Si(CH(CH_3)_2)_3$), 20.3, 20.5 ($POCH_2CH_2$), 24.7, 24.8 (CH_3, isopropyl-N), 30.1 (N-CH_3), 43.2, 43.4 (N($CH(CH_3)_2)_2$), 55.4 (OCH_3), 57.4, 58.1, 58.9, 59.1 ($POCH_2$), 61.9 (C5'), 70.2, 70.7 (C3'), 77.9, 78.5 (C2'), 83.2, 83.3 (C4'), 87.1, 87.3 (trityl, C_qOCH_2), 88.1, 88.3 (C1'), 89.2, 89.5 (OCH_2O), 113.4 (trityl), 117.5, 117.8 (CN), 127.3 (trityl), 128.1, 128.3, 128.4 (bz, trityl), 128.5 (trityl), 129.8 (bz), 130.2, 130.3 (trityl), 132.4 (bz), 135.3, 135.4, 135.6 (trityl), 136.1, 136.2 (bz, C_q), 144.1, 144.2 (trityl), 150.4 (C2), 155.8, 155.9 (C4), 158.8 (C_q-OCH_3), 177.4 (C=O, bz).
^{31}P-NMR ($CDCl_3$, 162 MHz): 150.12. 150.50.
HR-MS (ESI$^+$): Exact mass calculated for $C_{57}H_{76}N_5NaO_{10}PSi$ $[M+Na]^+$: 1072.4991, found: 1072.4968.

Solid-phase synthesis of oligoribonucleotides

RNA oligonucleotide syntheses were carried out at 0.7–1 µmol scale on ABI 394 DNA/RNA synthesizer using standard β-cyanoethyl phosphoramidite chemistry with the building blocks mentioned above. Deprotection of m^3C-modified oligoribonucleotides was performed with a mixture of concentrated aqueous NH_3 and EtOH (3:1) at 55 °C for 6 h followed by incubation with 1 M TBAF in THF under standard conditions. Conversion of the O^4-chlorophenyluridine-modified oligoribonucleotide into the $m^{4,4}$C-modified RNA (Table S1, R27) was performed by incubation of the solid support with 0.7 ml of 40% aqueous Me_2NH at 37°C for 5 h followed by addition of 0.5 ml of aqueous NH_3 and further incubation at 55 °C for 2 h to ensure complete deprotection of cyanoethyl and acetyl groups. Deprotection of TOM groups was achieved with TBAF in THF under standard conditions. Deprotection of other oligonucleotides was performed under standard conditions with $MeNH_2$ in aqueous ethanol and TBAF in THF. The crude oligonucleotides were desalted by size exclusion chromatography on GE Healthcare Life Sciences ÄKTA™ start system (3x5 ml HiTrap columns), purified by PAGE, and analyzed by anion exchange HPLC and HR-ESI-MS.

SUPPORTING INFORMATION

Labeling of oligonucleotides

Labeling of 5'-alkyne functionalized oligonucleotides

Oligonucleotides with 5'-alkynes were labeled with 6-FAM-azide using copper(I)-catalyzed alkyne-azide cycloaddition (CuAAC). Oligonucleotide (5 nmol in 5 µl of H_2O) was mixed with 3 µl of DMSO/tBuOH (3:1 v/v) and 0.5 µl of 6-FAM-azide (50 mM in DMSO/tBuOH 3:1). The solution of CuBr (100 mM) was freshly prepared in DMSO/tBuOH 3:1 and mixed with a solution of TBTA (100 mM) in 1:2 ratio. From this solution, 1.5 µl was added to the reaction mixture and incubated at 37°C for 3 hours in the dark. Afterwards, the reaction mixture was precipitated with ethanol to remove excess fluorophore, and the labeled oligonucleotide was purified using PAGE.

Labeling of RNA at 3'-end

Unfunctionalized oligoribonucleotides were labeled by periodate oxidation followed by a reaction with fluorescein-5-thiosemicarbazide. In a typical procedure, RNA (0.3–1 nmol) was dissolved in 7.5 µl water, followed by addition of 5x sodium phosphate buffer (2 µl, 250 mM, pH 7.4) and a freshly prepared aqueous solution of $NaIO_4$ (0.5 µl, 400 mM). The resulting mixture was incubated for 10–15 min at 37 °C. The excess of $NaIO_4$ was quenched by treatment with aqueous Na_2SO_3 (1 µl, 1 M) for 5–10 min at 37 °C. Afterwards, a solution of fluorescein-5-thiosemicarbazide (1 µl, 10 mM) in DMF was added, and the reaction mixture was incubated in the dark for additional 1 h at 37 °C. The labeled product was then purified by PAGE (10–20% polyacrylamide) and recovered by extraction and ethanol precipitation.

Radioactive labeling of RNA at 5'-end

Radioactive labeling of RNA was performed by phosphorylation using γ-^{32}P-ATP. In a typical procedure, RNA (200–500 pmol) was incubated with 10x PNK buffer A (5 µl, 500 mM Tris-HCl, 10 mM $MgCl_2$, 50 mM DTT, 1 mM spermidine), γ-^{32}P-ATP (0.5 µl, 10 µCi/µl) and T4 PNK enzyme (0.5 µl, 10 U/µl) at 37 °C for 1 hours. The labeled RNA was recovered by ethanol precipitation.

Ribozyme-catalyzed fluorescent labeling of RNA

Oligoribonucleotides R12, R16 and R20 were labeled at 2'-OH of internal adenosine (A9) with N^6-(6-aminohexyl)-ATP-6-FAM, N^6-(6-aminohexyl)-ATP-Cy3, N^6-(6-aminohexyl)-ATP-ATTO-647N, respectively, according to the known procedure.[6] In a typical procedure, RNA (500 pmol) and ribozyme FH14 (700 pmol) were incubated in total volume of 20 µl of the reaction buffer (5 mM NaCl, 120 mM KCl, 50 mM HEPES, pH 7.5) including 40 mM $MgCl_2$ and 200 µM of the N^6-modified ATP. The annealing step (95 ºC for 3 min, then r.t. for 10 min) was performed prior to addition of $MgCl_2$ and ATP. The reaction mixture was incubated overnight at 37 ºC and then quenched with 20 µl of stop solution. The labeled RNA was purified by PAGE (20%).

In-vitro selection

Phosphorylation of RNA substrates for selection

Phosphorylation at 5'-end of RNA substrates was carried out with T4 PNK as follows: RNA (2–5 nmol) was incubated with 10x PNK buffer A (5 µl, 500 mM Tris-HCl, 10 mM $MgCl_2$, 50 mM DTT, 1 mM spermidine), ATP (5 µl, 10 mM) and T4 PNK (5 µl, 10 U/µl) at 37 °C for 5 hours. Total volume of the reaction mixture was 50 µl. The phosphorylated RNA was extracted with phenol/ chloroform/isoamylalcohol mixture and ethanol precipitated.

Splinted ligation of RNA substrate and deoxyribozyme library

Splinted ligation was carried out with T4 DNA ligase as described earlier.[7] In a typical procedure, the deoxyribozyme pool, 5'-phosphorylated RNA and complementary DNA splint (1.8 nmol, 2.4 nmol and 2 nmol, respectively) were dissolved in water (9 µl), and 10x annealing buffer (1 µl; 40 mM Tris-HCl, 150 mM NaCl, 1 mM EDTA, pH 8.0) was added. After 2 min incubation at 95 °C, the solution was allowed to cool down to r.t. for 15 min and then further cooled down to 0 °C for 10 min. 10x DNA Ligase buffer (2 µl; 400 mM Tris-HCl, 100 mM $MgCl_2$, 100 mM DTT, 5 mM ATP, pH 7.8), T4 DNA Ligase (2 µl; 5 U/µl) and water to a total volume of 20 µl were added, and the resulting mixture incubated for 2 h at 37 °C to yield a DNA-RNA hybrid product which was purified by PAGE.

Intramolecular cleavage of RNA-DNA hybrids (key selection step)

The DNA-RNA hybrid (250 pmol in first round and 10–30 pmol in further rounds) was incubated with 10x selection buffer (1 µl, 500 mM Tris-HCl, 1.5 M NaCl, pH 7.5) in a total volume of 9 µl. To initiate the reaction, $MgCl_2$ (1 µl, 100 mM) was added to the final concentration of 10 mM and reaction mixture was incubated at 37 °C for 16 hours. Active fraction of DNA enzymes was separated using denaturing PAGE (10%) and the gel areas corresponding to the cleaved products (determined by comparison with a size marker) were cut and extracted. Extracted products were ethanol precipitated and amplified with PCR.

PCR amplification of the enriched DNA library

The enriched DNA library was amplified via two subsequent asymmetric PCR reactions.[7] In the first PCR, the extracted product from previous step was used as a template. Template (30 µl), forward primer D2 (200 pmol), reverse primer D3 (50 pmol), dNTP mixture (0.625 µl, 20 mM each dNTP), 10x DreamTaq buffer (5 µl) and DreamTaq DNA polymerase (0.25 µl, 5 U/µl) were mixed, and the total volume was adjusted to 50 µl with water. The resulting mixture was subjected to a 10-cycle PCR reaction (95 °C for 4 minutes, 10× [95 °C for 30 s, 60 °C for 30 s, 72 °C for 1 min] and 72 °C for 5 minutes) using a Bio-Rad thermal cycler. For the second PCR reaction, an aliquot (1–2 µl) from PCR-I was used as template. This aliquot was mixed with fluorescently labeled forward primer D2.1 (200 pmol), reverse primer D3 (50 pmol), dNTP mixture (0.6 µl, 20 mM), 10x DreamTaq buffer (5 µl), DreamTaq DNA polymerase (0.5 µl, 5 U/µl) and water to a final volume of 50 µl. The resulting solution was subjected to a 30-cycle PCR reaction (95 °C for 4 minutes, 30x [95 °C for 30 s, 60 °C for 30 s, 72 °C for 1 min], and 72 °C for 5 min). The PCR product was subjected to denaturing PAGE and the short strand with the fluorescent label was isolated. The resulting ssDNA was used for ligation to RNA substrate to initiate the next round of selection.

Cloning and sequencing

After 18 rounds of selection, the enriched libraries AK, AL, AM and AN were cloned using TOPO-TA cloning and transformed into *E.Coli.* The plasmids of 40 randomly picked colonies were isolated and sequenced using Sanger sequencing. Enriched libraries from rounds 7 and 18 were also subjected to deep sequencing. For this purpose, total eight selection libraries were amplified in a 2-step PCR protocol, in which Illumina adaptors and unique indices were introduced. The PCR products were purified on PAGE (10%) and with ZYMO Oligo Clean & Concentrator kit. After quality control on a Bioanalyzer, the libraries were pooled, and sequenced as a fraction of a NextSeq 500 High Output 75 nt single-end run in the Core Unit Systems Medicine (University of Würzburg). After demultiplexing, the resulting fastq files were further processed using custom bash, R and Python scripts.

Kinetics characterization of deoxyribozyme-catalyzed RNA cleavage

Kinetics characterization was carried out as described previously[8] to analyze trans-activity of individual deoxyribozymes or active deoxyribozyme pools under single turnover conditions. In a typical procedure, a deoxyribozyme (100 pmol) was mixed with a fluorescently labeled RNA substrate (10 pmol) in a final volume of 8 µl. The resulting mixture was heated to 95 °C for 4 min and then cooled to 25 °C for 10 min. To initiate the reaction, 10x kinetic assay buffer (1 µl, 500 mM Tris-HCl, 1.5 M NaCl), pH 7.5) and $MgCl_2$ (1 µl, 200 mM) were added and the resulting mixture was incubated at 37 °C. Aliquots (1 µl) were taken at different time points (e.g. 0, 15 min, 30 min, 60 min, 180 min, and 360 min), quenched with loading buffer and analyzed by denaturing PAGE. Cleavage yields were determined by measuring fluorescent intensities of the corresponding bands. Values of k_{obs} (observed cleavage rate) and Y_{max} (maximum fraction cleaved) were obtained by fitting fraction cleaved versus time with the first-order kinetics equation $Y = Y_{max} * (1 - e^{-kobs*t})$. Each experiment was repeated at least trice, and values are reported as mean of three individual experiments.

Synthesis and analysis of mt-tRNAs

Unmodified mt-tRNA^{Glu} was prepared by *in vitro* transcription from a synthetic DNA template using T7 RNA polymerase. Reactions were performed in aqueous solution containing 40 mM Tris-HCl, pH 8.0, 10mM DTT, 4 mM each NTP, 30 mM $MgCl_2$ and 2 mM spermidine and 20 µg of T7 RNA polymerase (prepared in house) at 37 °C for 5 h. The transcription products were purified by denaturing PAGE (10% polyacrylamide) and dephosphorylated by calf intestinal alkaline phosphatase. The m^{5}C-modified mt-tRNA^{Glu} was assembled by splint ligation of two RNA fragments (R25 and R26). The m^{3}C-modified mt-tRNA^{Thr} and unmodified mt-tRNA^{Thr} were constructed by splint ligation of three synthetic RNA fragments (R21, R22, R23) and (R21, R22, R24) respectively. In a typical procedure, RNA fragments and complementary splint (D52/D53) in equimolar ratio were annealed in 10x annealing buffer (40 mM Tris-HCl, 150 mM NaCl, 1 mM EDTA, pH 8.0) and ligation reactions were then performed with T4 DNA ligase for 16 h at 37 °C. Full length products were separated by PAGE (10% polyacrylamide), and labeled with γ-^{32}P-ATP at their 5'end.Aliquots of the full-length tRNA samples were then subjected to DNAzyme-mediated cleavage using the corresponding DNAzymes (D54/D55/D57/D58) as described above.

SUPPORTING INFORMATION

To disrupt the secondary structure of tRNAs close to DNAzyme binding site, disruptor oligonucleotides (D56/D59) were added in the cleavage reactions. Labeled tRNAs (approx. 10 pmol) were annealed with corresponding DNAzyme (100 pmol) and disruptor oligonucleotides (100 pmol). The reactions were incubated in a final volume of 10 µl of 1x selection buffer (50 mM Tris-HCl, 150 mM NaCl, pH 7.5) including 40 mM $MgCl_2$ at 37°C for 18 h. Aliquots (2 µl) were taken at 0, 6 h and 18 h, quenched with loading buffer and analyzed by denaturing PAGE. The gels were dried under vacuum for 30 min at 80 °C, exposed overnight to a phosphor storage screen and imaged on a Typhoon phosphorimager.

SUPPORTING INFORMATION

Supporting Tables

Table S1. Sequences of RNA oligonucleotides

No	Description/motif	5'-sequence-3'	mol. wt. calc., amu	mol. wt. found. m/z
R1	unmodified/NC	AUAGACUGAAUGAANCGUAACUAGUC	-	-
R2	Nm^{3}C	AUAGACUGAAUGAANm^{3}CGUAACUAGUC	-	-
R3	Nm^{4}C	AUAGACUGAAUGAANm^{4}CGUAACUAGUC	-	-
R4	Nm^{5}C	AUAGACUGAAUGAANm^{5}CGUAACUAGUC	-	-
R5	unmodified/AC	AUAGACUGAAUGAACGUAACUAGUC	8339.17	8339.34
R6	unmodified/GC	AUAGACUGAAUGAAGCGUAACUAGUC	8355.17	8355.25
R7	unmodified/CC	AUAGACUGAAUGAACCGUAACUAGUC	8315.16	8315.19
R8	unmodified/UC	AUAGACUGAAUGAAUCGUAACUAGUC	8316.14	8316.21
R9	Am^{3}C	AUAGACUGAAUGAAAm^{3}CGUAACUAGUC	8353.19	8353.00
R10	Gm^{3}C	AUAGACUGAAUGAAGm^{3}CGUAACUAGUC	8369.18	8369.27
R11	Cm^{3}C	AUAGACUGAAUGAACm^{3}CGUAACUAGUC	8329.18	8329.31
R12	Um^{3}C	AUAGACUGAAUGAAUm^{3}CGUAACUAGUC	8330.16	8330.27
R13	Am^{4}C	AUAGACUGAAUGAAAm^{4}CGUAACUAGUC	8353.19	8253.20
R14	Gm^{4}C	AUAGACUGAAUGAAGm^{4}CGUAACUAGUC	8369.18	8369.19
R15	Cm^{4}C	AUAGACUGAAUGAACm^{4}CGUAACUAGUC	8329.18	8329.20
R16	Um^{4}C	AUAGACUGAAUGAAUm^{4}CGUAACUAGUC	8330.16	8330.18
R17	Am^{5}C	AUAGACUGAAUGAAAm^{5}CGUAACUAGUC	8353.19	8353.25
R18	Gm^{5}C	AUAGACUGAAUGAAGm^{5}CGUAACUAGUC	8369.18	8369.24
R19	Cm^{5}C	AUAGACUGAAUGAACm^{5}CGUAACUAGUC	8329.18	8329.25
R20	Um^{5}C	AUAGACUGAAUGAAUm^{5}CGUAACUAGUC	8330.16	8330.20
R21	mt-tRNA^{Thr} 5' fragment	GUCCUUGUAGUAUAAACUAAUACA		
R22	mt-tRNA^{Thr} 3' fragment	pGAUGAAAACCUUUUUCCAAGGACAp		
R23	mt-tRNA^{Thr} m^{3}C ASL	CCAGUm^{3}CUUGUAAACCGGA		
R24	mt-tRNA^{Thr} unmod. ASL	CCAGUCUUGUAAACCGGA		
R25	mt-tRNA^{Glu} 5'-fragment	GUUCUUGUAGUUGAAAUACAACGAUGGUUUUUCAUA		
R26	mt-tRNA^{Glu} m^{5}C 3'-fragment	UCAUUGGUm^{5}CGUGGUUGUAGUCCGUGCGAGAAUA		
R27	Um^{4,4}C	AUAGACUGAAUGAAUm^{4,4}CGUAACUAGUC	8344.14	8344.24
R28	unmodified/UC transcript	GGAUAGACUGAAUGAAUCGUAACUAGUCA[a]	9575.14	9575.18
R29	unmodified/UU transcript	GGAUAGACUGAAUGAAUUGUAACUAGUCAA[a]	9905.14	9905.28
R30	unmodified/UA transcript	GGAUAGACUGAAUGAAUAGUAACUAGUCAA[a]	9928.15	9928.30
R31	unmodified/UA transcript	GGAUAGACUGAAUGAAUGGUAACUAGUCAA[a]	9944.14	9944.30
R32	Um^{5}C transcript	GGAUAGAm^{5}CUGAAUGAAUm^{5}CGUAAm^{5}CUAGUm^{5}CA[a]	9630.19	9630.31
R33	Uhm^{5}C transcript	GGAUAGAhm^{5}CUGAAUGAAUhm^{5}CGUAAhm^{5}CUAGUhm^{5}CA[a]	9694.17	9694.27
R34	Uf^{5}C transcript	GGAUAGAf^{5}CUGAAUGAAUf^{5}CGUAAf^{5}CUAGUf^{5}CA[a]	9686.11	9686.19

[a] non-templated A added by T7 RNA polymerase in vitro transcription

Table S2. Sequences of DNA oligonucleotides

No	Description	5'-Sequence-3'[a]
D1	Selection pool	GTGACGCGACTAGTTACN$_{20}$TTCATTCAGTTGGCGACTCC
D2	Selection forward primer	GTGACGCGACTAGTTAC
D2.1	Selection forward primer (for 5'-labeling)	Alk-GTGACGCGACTAGTTAC
D3	Selection reverse primer (tailed)	(CAA)$_4$E$_3$GGAGGCGCCAACTGAATGAA
D4	Splint for selection pool (round1)	TTCATTCAGTCTATGGAGTCGCCAACTG
D5	Splint for other selection rounds	TTACGTTTCATTCAGTCTATTGGAGTCGCCAACT
D6	AK02	GACTAGTTAC CAAGGAGCACGGAACTTTC TTCATTCAGT
D7	AK03	GACTAGTTAC GGCAGGAGCAGTCACATCG TTCATTCAGT
D8	AK04	GACTAGTTAC CCAACCAGGCCGGCACATA TTCATTCAGT
D9	AK11	GACTAGTTAC CAACCAGGATGCGGAACCA TTCATTCAGT
D10	AK12	GACTAGTTAC GGGGTATACAGTGGGGCGA TTCATTCAGT

D11	AK13	GACTAGTTAC GGGGTATACAGTGGCGCGAA TTCATTCAGT
D12	AK24	GACTAGTTAC CCAACGGTTGAGTCACCCC TTCATTCAGT
D13	AL01	GACTAGTTAC CGCAGGTGTTTGTGGGGTAA TTCATTCAGT
D14	AL03	GACTAGTTAC GGTTCGGGACGGCCTGGTAA TTCATTCAGT
D15	AL04	GACTAGTTAC GGGGTAGCCAGTGGACGTT TTCATTCAGT
D16	AL07	GACTAGTTAC GGGGCAGGTGGTGGGAAGGG TTCATTCAGT
D17	AL10	GACTAGTTAC CGAGGCGCGGAACCCCCAC TTCATTCAGT
D18	AL12	GACTAGTTAC CGTTACGGAGTGCGTGGTAA TTCATTCAGT
D19	AL13	GACTAGTTAC GGGGGTAAGACTGGGGTAG TTCATTCAGT
D20	AL15	GACTAGTTAC GGTTACGTAGTGCCTGGTTG TTCATTCAGT
D21	AL18	GACTAGTTAC CCGATCGTGGACTCCGGAGA TTCATTCAGT
D22	AL21	GACTAGTTAC GGCACCAACTGGAGTCTGGG TTCATTCAGT
D23	AM05	GACTAGTTAC GGGGTACCGGTGGAGACTA TTCATTCAGT
D24	AM06	GACTAGTTAC CCAACAAGGAGGGTCACCC TTCATTCAGT
D25	AM11	GACTAGTTAC CAAACGATCGCCGGCACCG TTCATTCAGT
D26	AM12	GACTAGTTAC CACCACAATCGCGGAACCG TTCATTCAGT
D27	AM13	GACTAGTTAC GGGAGCGACACAACCATCA TTCATTCAGT
D28	AM17	GACTAGTTAC CACGTGGTGAGCACTCTAAC TTCATTCAGT
D29	AM18	GACTAGTTAC CACGTGGTGAGCACTTCTAC TTCATTCAGT
D30	AM20	GACTAGTTAC GGCACCCACTGGAGTCTGGT TTCATTCAGT
D31	AM21	GACTAGTTAC CGCTGGGTTGGGTTGCGCGG TTCATTCAGT
D32	AN01	GACTAGTTAC CCAACAGACAGCGGCACTC TTCATTCAGT
D33	AN02	GACTAGTTAC GGGGCGGAACGTAATCTTA TTCATTCAGT
D34	AN03	GACTAGTTAC CCACAGGCACGAGGATGTAA TTCATTCAGT
D35	AN04	GACTAGTTAC GGGGTATACAGTGGCGCGA TTCATTCAGT
D36	AN05	GACTAGTTAC GGTAGTAGGATGGCGATCCG TTCATTCAGT
D37	AN07	GACTAGTTAC GGACACAGCTGGTACCGGT TTCATTCAGT
D38	AN12	GACTAGTTAC CACCAACTAGGAGCACCCG TTCATTCAGT
D39	AN14	GACTAGTTAC CCCATACTAGCGTCGTCCG TTCATTCAGT
D40	AN19	GACTAGTTAC CGGCTGGACGTTGACGCGGT TTCATTCAGT
D41	AN24	GACTAGTTAC CCAGTAACGTGTCCGCCAT TTCATTCAGT
D42	AK104	GACTAGTTAC GGGTGCCGGGTGAGCGTTAT TTCATTCAGT
D43	AK118	GACTAGTTAC GGGGTAGCCAGTGGGACGTG TTCATTCAGT
D44	AL107	GACTAGTTAC GGGGTAGGTGGTGGGAAGGG TTCATTCAGT
D45	AL112	GACTAGTTAC GGTTGCGGTAGCGCCTGGTG TTCATTCAGT
D46	AM101	GACTAGTTAC CGGGCCGGGGGTTAATCG TTCATTCAGT
D47	AM102	GACTAGTTAC TCAACAGCCGAAACGACCA TTCATTCAGT
D48	AM106	GACTAGTTAC GGGGTATACAGTGGTGCGA TTCATTCAGT
D49	AM108	GACTAGTTAC CGGCTGGACGTTGACGCGA TTCATTCAGT
D50	AN109	GACTAGTTAC TCAACAGCCGAAACGACAC TTCATTCAGT
D51	AL112_G20A	GACTAGTTAC GGTTGCGGTAGCGCCTGGTA TTCATTCAGT
D52	Splint mt-tRNA^{Thr}	AAGGTTTTCATCTCCGGTTTACAAGACTGGTGTATTAGTTT
D53	Splint mt-tRNA^{Glu}	GACCAATGATATGAAAAACCATCG
D54	AK104 for mt-tRNA^{Thr}	TCCGGTTTACAA GGGTGCCGGGTGAGCGTTAT CTGGTGTATT
D55	AL112 for mt-tRNA^{Thr}	TCCGGTTTACAA GGTTGCGGTAGCGCCTGGTG CTGGTGTATT
D56	Disruptor for mt-tRNA^{Thr}	TTTATACTACAAGGAC
D57	AK104 for mt-tRNA^{Glu}	GACTACAACCAC GGGTGCCGGGTGAGCGTTAT CCAATGATAT
D58	AN05 for mt-tRNA^{Glu}	GACTACAACCAC GGTAGTAGGATGGCGATCCG CCAATGATAT
D59	Disruptor for mt-tRNA^{Glu}	TGGTATTCTCGCACGGAC

[a] Alk = 5-hexynyl, E_3 = triethylene glycol

SUPPORTING INFORMATION

Table S3. Summary of enrichment data and cleavage activity for deoxyribozymes reported in this study.

Deoxyribozyme	log_2f_{CA}[a]	$log_2f_{CA,neg}$[b]	Cleavage site	N[c]	Activity, FC[d](k_{obs}* 10^{-3} min^{-1})[e]			
					C	m^3C	m^4C	m^5C
AK12	4.76	-2.94 (AL)	AN | CG	A	12	0	2	2
				G	89	0	69	49
				C	0	0	0	0
				U	6	0	0	0
AK13	1.32	-4.36 (AL)	AN | CG	A	12	0	0	0
				G	77	0	48	51
				C	0	0	0	0
				U	3	0	0	0
AK104[f]	10.64	-5.15 (AM) -12.91 (AN)	AN | CG A | NCG AN | CG	A	95	0	0	0
				G	94	0	0	0
				C	0	29	0	0
				U	85(3.2)	0	0	0
AK118	4.73	-8.94 (AN) -6.28 (AL)	AN | CG	A	0	0	0	0
				G	62	0	30	53
				C	0	0	0	0
				U	23	10	3	12
AL01	-0.49	-1.45	AN | CG	A	12	27	14	9
				G	12	29	14	8
				C	0	0	0	0
				U	0	0	0	0
AL04	13.04	2.39	AN | CG	A	0	0	0	0
				G	13	58	1	4
				C	0	24	0	0
				U	0	0	0	0
AL07	3.00	-1.30	AN | CG	A	0	15	0	0
				G	0	11	1	0
				C	2	60	7	0
				U	27	87	53	21
AL12	-3.18	-4.17	AN | CG	A	0	13	6	0
				G	18	28	23	13
				C	0	0	0	0
				U	0	0	0	0
AL13	11.32	4.07	ANC | G	A	<1	68	0	0
				G	6	73	14	8
				C	0	3	0	0
				U	<1	12	0	0
AL15	-0.86	-3.19	AN | CG	A	0	31	3	0
				G	25	63	47	17
				C	0	7	1	0
				U	1	11	4	2
AL18	1.41	0.53	AN | CG	A	2	3	3	2
				G	40	26	45	26
				C	0	0	0	0
				U	0	0	0	0
AL21	7.90	4.17	AN | CG	A	0	5	0	0
				G	0	4	0	0
				C	<1	30	2	0
				U	6	55	15	5
AL107	2.93	-1.37	AN | CG	A	0	14	0	0
				G	0	9	0	0
				C	0	54	2	0
				U	21	93	45	18
AL112[f]	-2.94	-	AN | CG	A	0	45	0	0
				G	0	40	0	0
				C	0	57	0	0
				U	3(0.06)	70(1.8)	10(0.2)	4(0.08)
AM05	12.82	1.64	ANC | G	A	0	0	0	0
				G	0	0	3	0
				C	0	0	0	0
				U	66	8	81	56
AM17	5.48	2.21	AN | CG	A	0	0	0	0
				G	47	49	18	39
				C	0	0	0	0
				U	0	0	0	0

SUPPORTING INFORMATION

Deoxyribozyme	log_2f_{CA}[a]	$log_2f_{CA,neg}$[b]	Cleavage site	N[c]	Activity, FC[d] (k_{obs}* 10^{-3} min^{-1})[e]			
					C	m^3C	m^4C	m^5C
AM18	4.28	1.56	AN | CG	A	0	0	0	0
				G	21	24	17	12
				C	0	0	0	0
				U	0	0	0	0
AM21	0.59	-6.63	AN | CG	A	29	71	72	35
				G	6	14	3	4
				C	0	0	0	0
				U	9	8	9	8
AM101[f]	10.02	1.41	ANC | G	A	0	0	0	0
				G	0	0	0	0
				C	0	0	22	0
				U	18(0.5)	0	73(5.3)	0
AM102	6.38	-2.90	ANC | G	A	27	44	52	25
				G	71	74	82	64
				C	49	47	84	57
				U	88	71	91	88
AM106	7.78	4.61	AN | CG	A	9	0	0	0
				G	81	0	55	31
				C	0	0	0	0
				U	0	0	0	0
AM108	5.86	0.21	AN | CG	A	10	19	15	7
				G	42	41	50	29
				C	0	0	0	0
				U	0	0	0	0
AN03	2.07	-0.01	AN | CG	A	0	10	5	6
				G	17	13	12	13
				C	0	1	0	0
				U	0	0	0	0
AN04	3.15	2.28	AN | CG	A	16	0	6	4
				G	78	0	66	73
				C	0	0	0	0
				U	5	0	2	12
AN05[f]	12.35	-4.68	A | NCG	A	0	0	0	0
				G	0	0	0	0
				C	0	0	0	0
				U	33(0.7)	0	0	90(6.7)
AN07	9.29	3.01	AN | CG	A	0	0	0	0
				G	0	0	0	0
				C	0	0	0	0
				U	47	0	28	39
AN19	3.81	-3.05	AN | CG	A	0	22	0	0
				G	0	9	2	0
				C	0	2	0	0
				U	67	60	82	62
AN109	8.96	-4.81	ANC | G	A	25	42	49	28
				G	86	71	87	67
				C	43	47	83	62
				U	88	79	93	93

Color code	Activity	FC[d] range
	Low	0 ≤ FC < 10
	Moderate	10 ≤ FC < 50
	High	50 ≤ FC ≤ 100

[a] log_2 fold change in abundance between rounds 18 and 7 of the modification-specific selection (AK for C, AL for m^3C, AM for m^4C and AN for m^5C) log_2f_{CA} is defined as $log_2(CPM_{R18} / CPM_{R7})$, where CPM_{R18} is the read counts per million for the corresponding deoxyribozyme in round 18, and CPM_{R7} is read counts per million in round 7. [b] log_2 fold abundance change between rounds 18 and 7 of the negative selection (AK for m^3C, m^4C, m^5C). [c] Nucleotide preceding the target cytosine site in the RNA substrate. [d] Fraction cleaved (%) after 6 h incubation (100 pmol deoxyribozyme, 10 pmol RNA substrate, 20 mM $MgCl_2$, 50 mM Tris-HCl, 150 mM NaCl, pH 7.5). [e] Observed kinetic constants for single-turnover cleavage. [f] The activity assays were performed at least twice to check for reproducibility. Mean FC values are presented.

SUPPORTING INFORMATION

Supporting figures

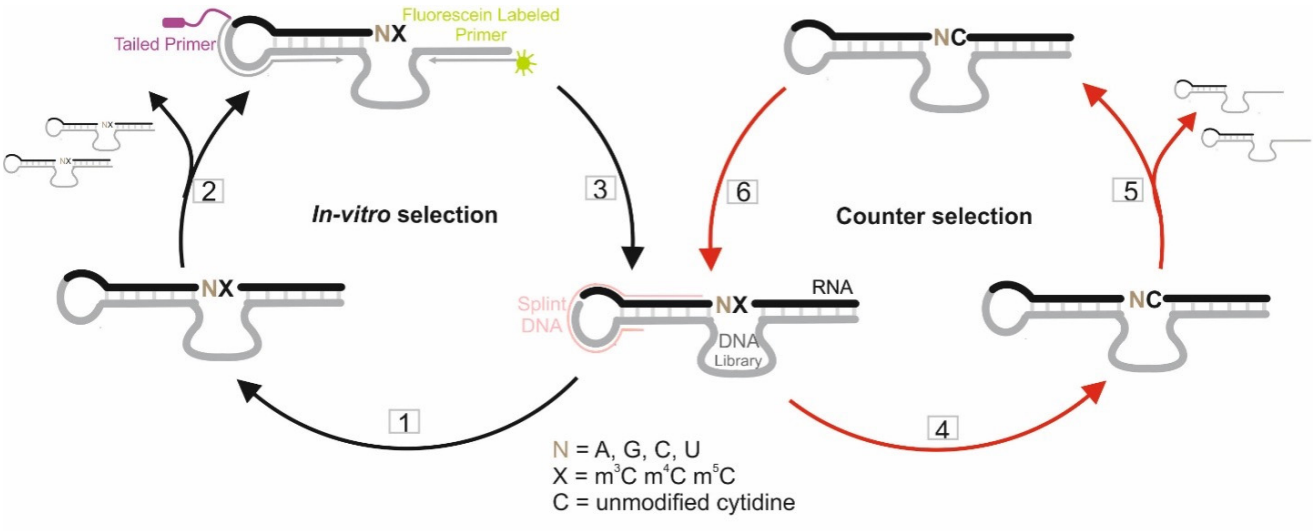


Figure S1. Schematic of *in vitro* selection of RNA-cleaving deoxyribozymes sensitive to m^3C, m^4C or m^5C modifications (AL, AM and AN selections, respectively): 1) Ligation of DNA library to modified RNA (R2, R3, R4). 2) Incubation in presence of $MgCl_2$. The cleaved fraction is isolated by PAGE. 3) PCR amplification of active fraction. 4) Ligation of DNA library to unmodified RNA (R1). 5) Incubation in presence of $MgCl_2$. The uncleaved fraction is isolated by PAGE. 6) PCR amplification of uncleaved fraction.

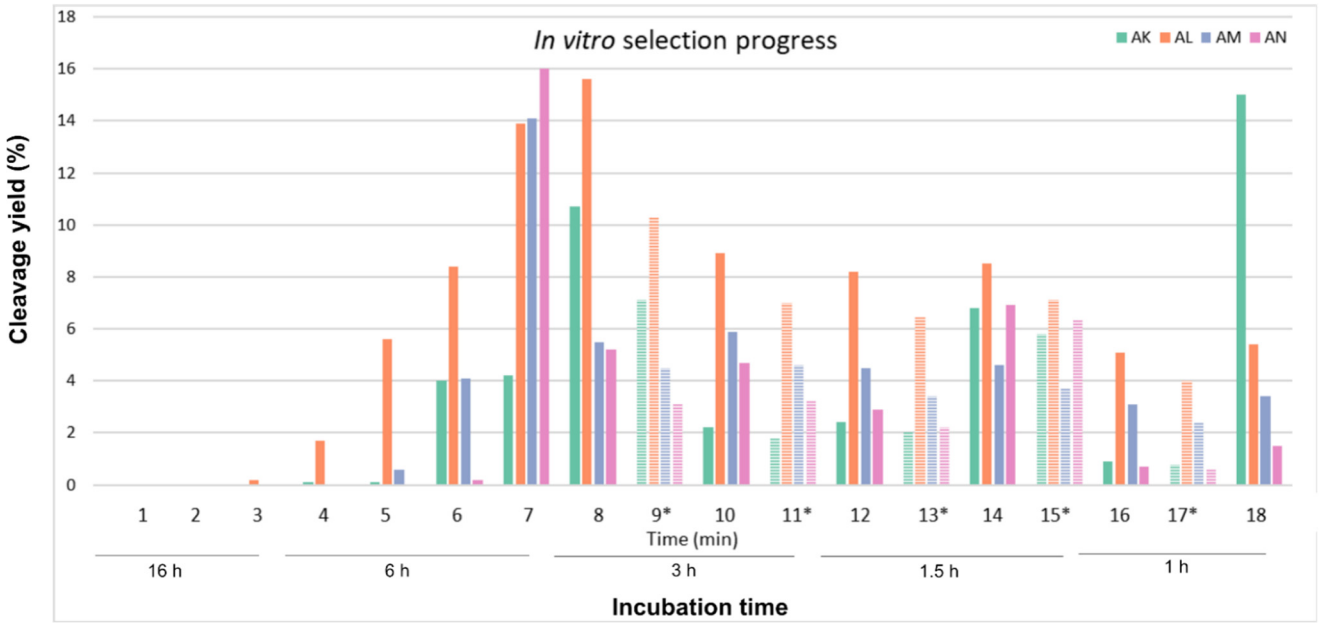


Figure S2. Progress of in-vitro selection plotted as cleavage yield in every selection round. From round 4, cleavage activity was visible and cleavage yield was determined by quantification of fluorescent bands in the selection step. The counter selection rounds are marked with asterisks and incubation time was 16 h for every counter selection round.

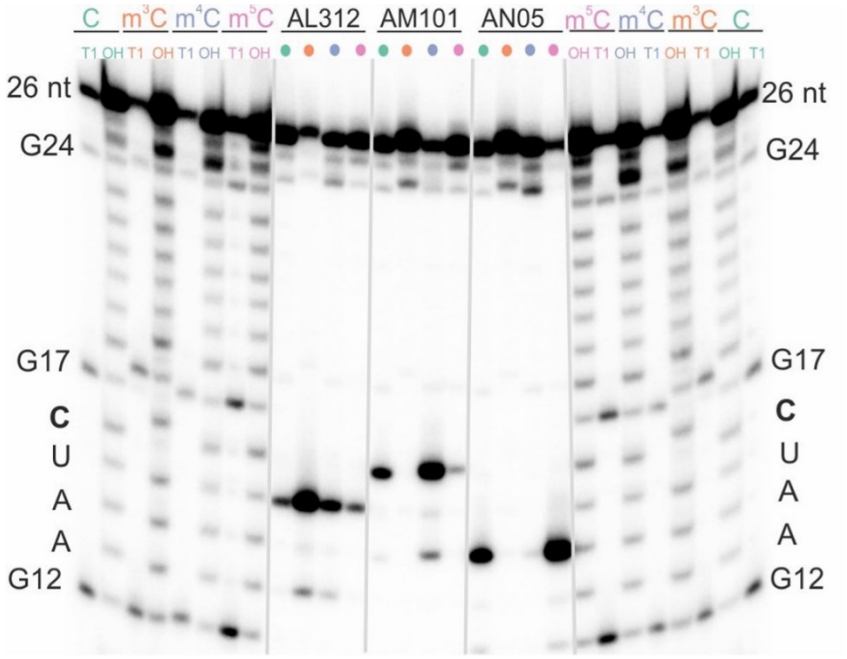


Figure S3. PAGE analysis of RNA cleavage mediated by deoxyribozymes AL112, AM101 and AN05 and comparison of cleavage activity for unmodified (green), m^3C- (orange), m^4C- (blue), and m^5C-modified (magenta) RNA. 5'-Radioactively labeled RNA was incubated with DNA enzymes in presence of 20 mM $MgCl_2$ at 37 °C for 6 h. Lanes OH: alkaline hydrolysis of RNA in presence of 25 mM NaOH at 95 °C for 3 min. Lanes T1: RNase T1 ladder showing positions of Gs in the RNA substrate.

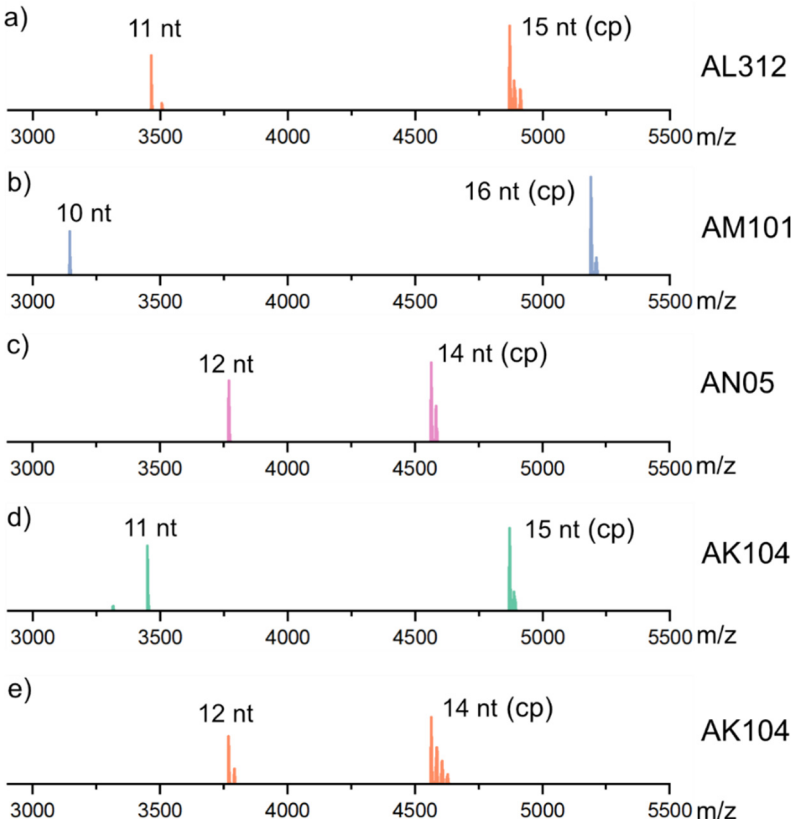


Figure S4. High-resolution electrospray ionization mass spectrometry (HR-ESI-MS) analysis of cleaved products for AL112, AM101, AN05 and AK104. a) RNA fragments formed by cleavage of R12 (Um^3C) with AL312. b) RNA fragments formed by cleavage of R16 (Um^4C) with AM101. c) RNA fragments formed by cleavage of R20 (Um^5C). d) AK104 cleaved R8 (UC motif) with formation of 15-nt fragment which contained 2',3'-cyclic phosphate (cp) and 11-nt fragment which contained 5'-OH. e) AK104 shifted its cleavage site one nucleotide upstream in reaction with R11 (Cm^3C) as confirmed by formation of 14-nt and 12-nt fragments.

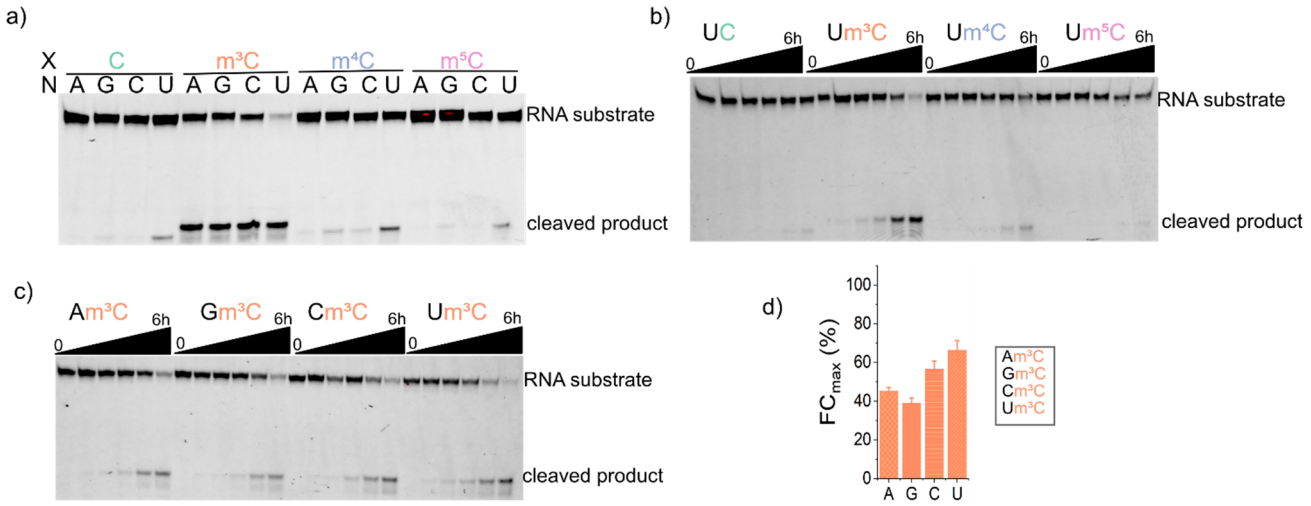


Figure S5. Activity assays for deoxyribozyme AL112. Cleavage yields (FC) and k_{obs} values are presented in Table S3. a) Gel image for cleavage of R5–R20. X – modification at the target cytosine site. N – nucleotide preceding the target cytosine site. Conditions: 100 pmol deoxyribozyme, 10 pmol RNA substrate, 20 mM $MgCl_2$, 50 mM Tris-HCl, 150 mM NaCl, pH 7.5, 37 °C, 6 h. b) Gel-based kinetics assays for cleavage of R8 (UC), R12 (Um^3C), R16 (Um^4C) and R20 (Um^5C). Time points: 0, 10, 30, 60, 180, 360 min. Kinetics plots are presented in Figure 2b. c) Gel-based kinetics assay for cleavage of R9 (Am^3C), R10 (Gm^3C), R11 (Cm^3C) and R12 (Um^3C). d) Comparison of FC values obtained after 6 h cleavage of R9–R12. Each bar represents the mean from three independent experiments with error bars showing the ± SD.

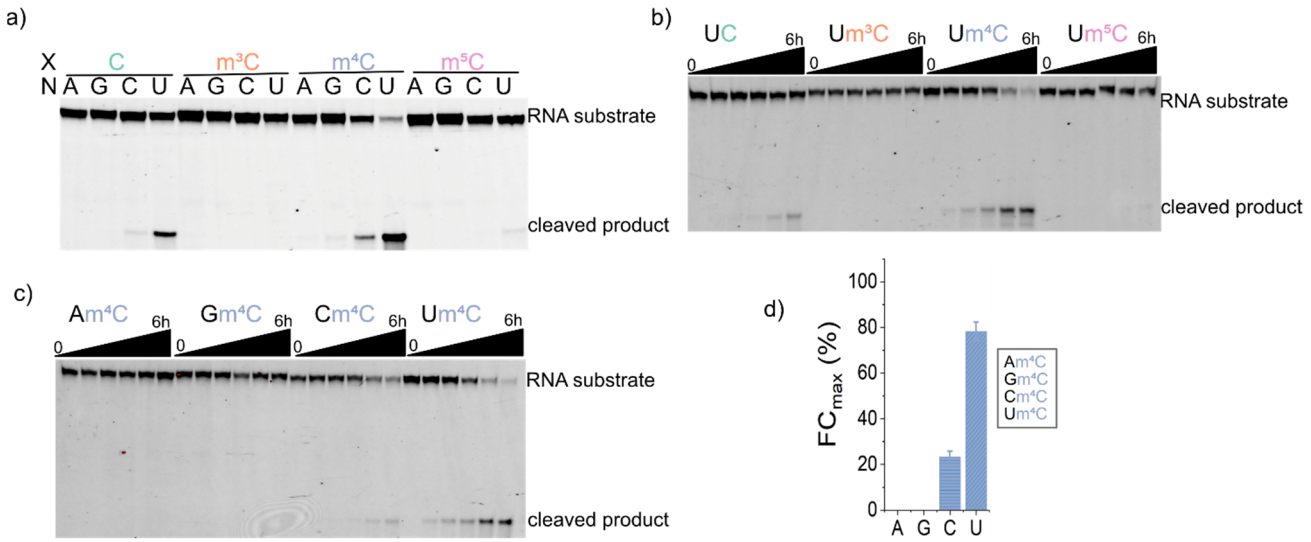


Figure S6. Activity assays for deoxyribozyme AM101. Cleavage yields (FC) and k_{obs} values are presented in Table S3. a) Gel image for cleavage of R5–R20. X – modification at the target cytosine site. N – nucleotide preceding the target cytosine site. Conditions: 100 pmol deoxyribozyme, 10 pmol RNA substrate, 20 mM $MgCl_2$, 50 mM Tris-HCl, 150 mM NaCl, pH 7.5, 37 °C, 6 h. b) Gel-based kinetics assays for cleavage of R8 (UC), R12 (Um^3C), R16 (Um^4C) and R20 (Um^5C). Time points: 0, 10, 30, 60, 180, 360 min. Kinetics plots are presented in Figure 2b. c) Gel-based kinetics assay for cleavage of R13 (Am^4C), R14 (Gm^4C), R15 (Cm^4C) and R16 (Um^4C). d) Comparison of FC values obtained after 6 h cleavage of R13–R16. Each bar represents the mean from three independent experiments with error bars showing the ± SD.

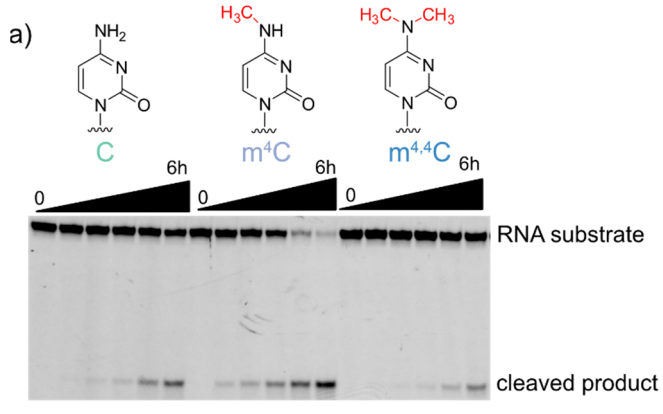


Figure S7. Single-turnover cleavage of unmodified, m^4C-modified and $m^{4,4}C$-modified RNA substrates with m^4C-specific deoxyribozyme AM101. a) Gel images for cleavage of R8, R20 and R27. Conditions: 100 pmol deoxyribozyme, 10 pmol RNA substrate, 20 mM $MgCl_2$, 50 mM Tris-HCl, 150 mM NaCl, pH 7.5, 37 °C, 6 h. b) Kinetic plots. d) Comparison of FC values obtained after 6 h cleavage.

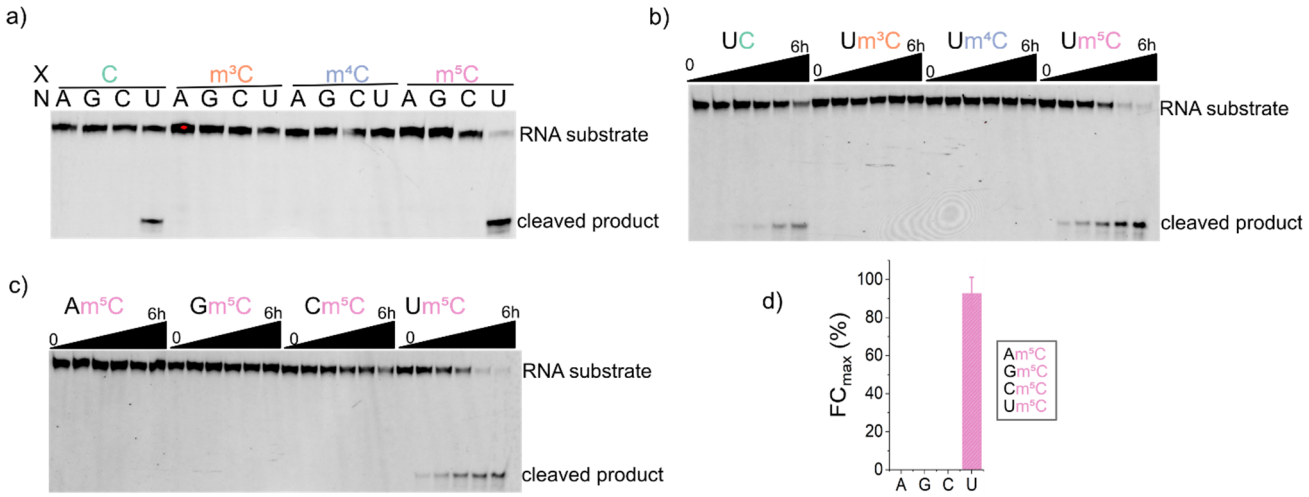


Figure S8. Activity assays for deoxyribozyme AN05. Cleavage yields (FC) and k_{obs} values are presented in Table S3. a) Gel image for cleavage of R5–R20. X – modification at the target cytosine site. N – nucleotide preceding the target cytosine site. Conditions: 100 pmol deoxyribozyme, 10 pmol RNA substrate, 20 mM $MgCl_2$, 50 mM Tris-HCl, 150 mM NaCl, pH 7.5, 37 °C, 6 h. b) Gel-based kinetics assays for cleavage of R8 (UC), R12 (Um^3C), R16 (Um^4C) and R20 (Um^5C). Time points: 0, 10, 30, 60, 180, 360 min. Kinetics plots are presented in Figure 2b. c) Gel-based kinetics assay for cleavage of R17 (Am^5C), R18 (Gm^5C), R19 (Cm^5C) and R20 (Um^5C). d) Comparison of FC values obtained after 6 h cleavage of R17–R20. Each bar represents the mean from three independent experiments with error bars showing the ± SD.

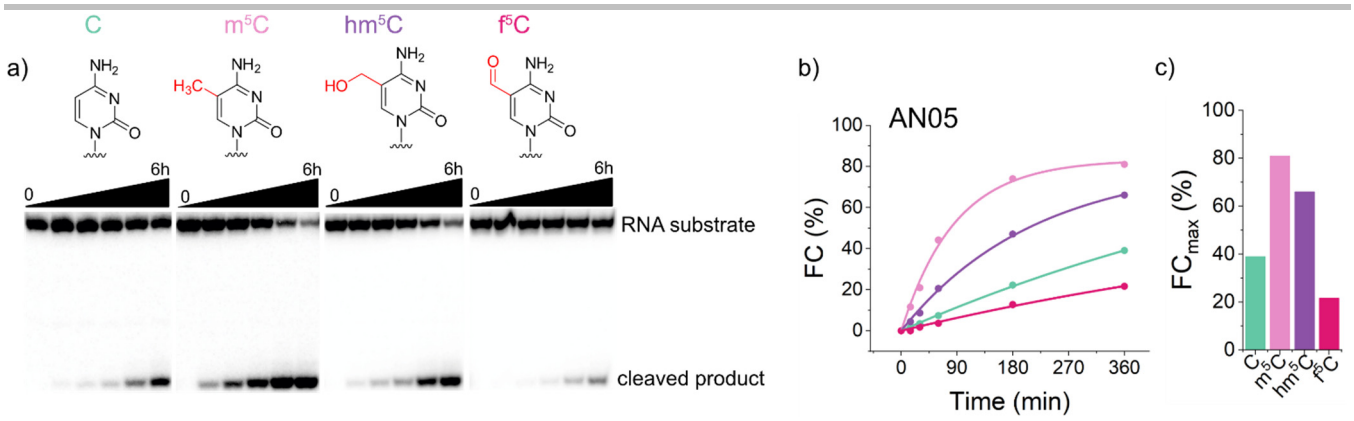


Figure S9. Single-turnover cleavage of unmodified, m^{5}C-, hm^{5}C- and f^{5}C-modified RNA substrates with m^{5}C-specific deoxyribozyme AN05. a) Gel images for cleavage of R28 and R32−R34. Conditions: 100 pmol deoxyribozyme, 10 pmol RNA substrate, 20 mM $MgCl_2$, 50 mM Tris-HCl, 150 mM NaCl, pH 7.5, 37 °C, 6 h. b) Kinetic plots. d) Comparison of FC values obtained after 6 h cleavage.

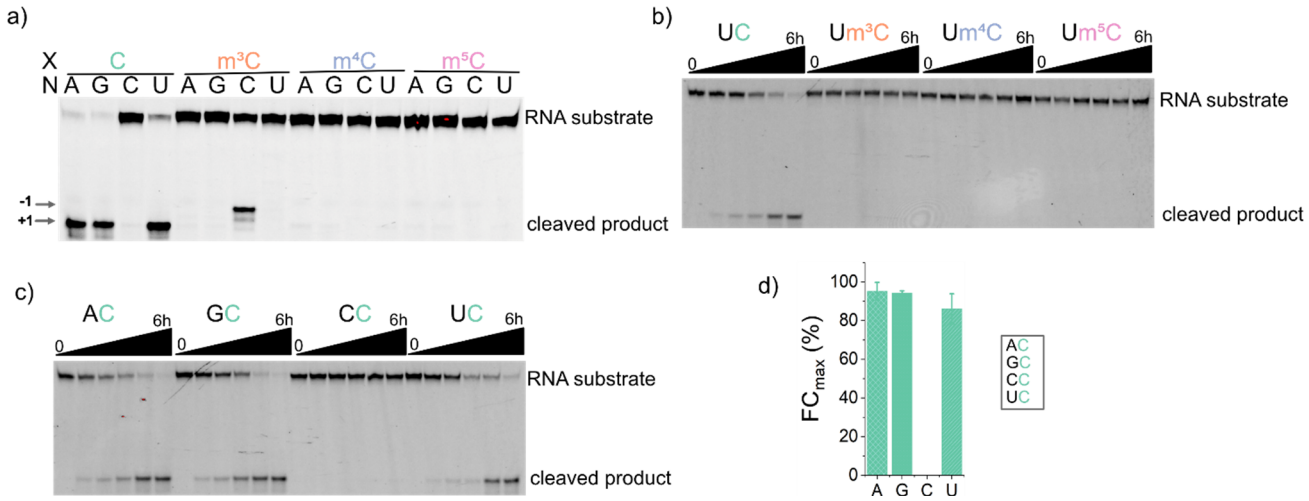


Figure S10. Activity assays for deoxyribozyme AK104. Cleavage yields (FC) and k_{obs} values are presented in Table S3. a) Gel image for cleavage of R5–R20. X – modification at the target cytosine site. N – nucleotide preceding the target cytosine site. Conditions: 100 pmol deoxyribozyme, 10 pmol RNA substrate, 20 mM $MgCl_2$, 50 mM Tris-HCl, 150 mM NaCl, pH 7.5, 37 °C, 6 h. b) Gel-based kinetics assays for cleavage of R8 (UC), R12 (Um^{3}C), R16 (Um^{4}C) and R20 (Um^{5}C). Time points: 0, 10, 30, 60, 180, 360 min. Kinetics plots are presented in Figure 2b. c) Gel-based kinetics assay for cleavage of R5 (AC), R6 (GC), R7 (CC) and R8 (UC). d) Comparison of FC values obtained after 6 h cleavage of R5–R8. Each bar represents the mean from three independent experiments with error bars showing the ± SD.

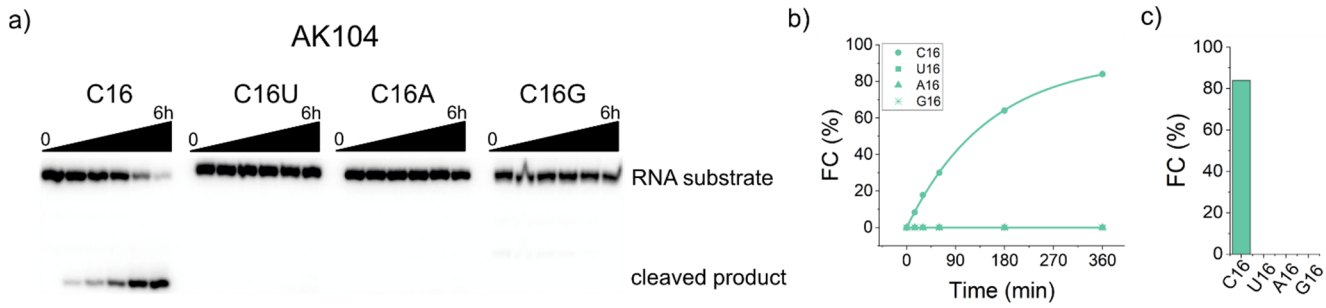


Figure S11. Single-turnover cleavage of the unmodified selection substrate and its C16U, C16A and C16G mutants with deoxyribozyme AK104. a) Gel images for cleavage of R28−R31. Conditions: 100 pmol deoxyribozyme, 10 pmol RNA substrate, 20 mM $MgCl_2$, 50 mM Tris-HCl, 150 mM NaCl, pH 7.5, 37 °C, 6 h. b) Kinetic plots. d) Comparison of FC values obtained after 6 h cleavage.

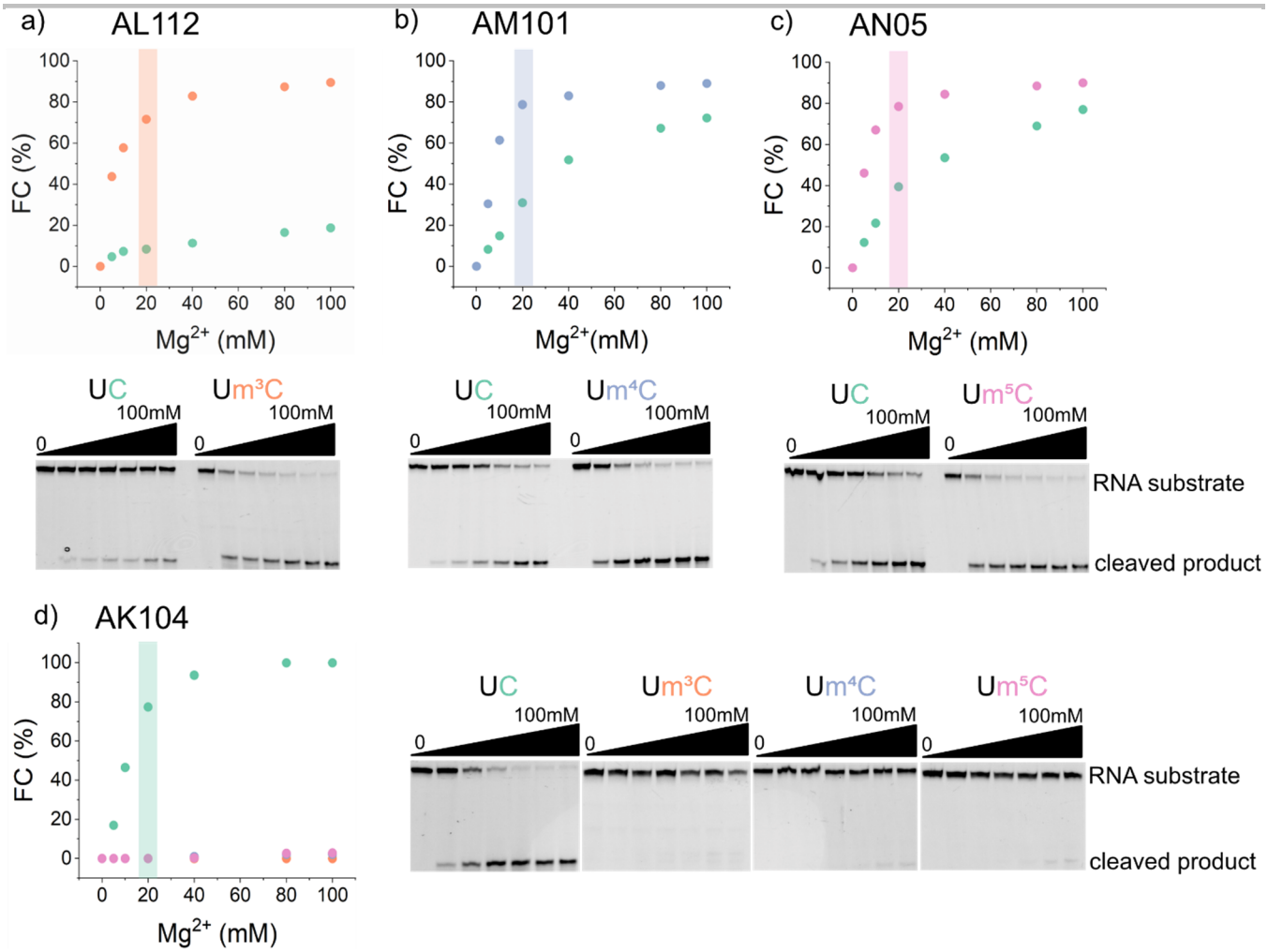


Figure S12. Dependence of cleavage yield on concentrations of Mg^{2+} for the reactions of the selection substrates with deoxyribozymes AL112, AM101, AN05 and AK104. a)−c) Top panel: cleavage yields plotted vs concentration of Mg^{2+} for deoxyribozymes AL112, AM101 and AN05, respectively. Bottom panel: gel images for cleavage of R8 and R12 with AL112 (a), R8 and R16 with AM101 (b), R8 and R20 with AN05 (c). d) Cleavage yields plotted vs concentration of Mg^{2+} for deoxyribozyme AK104 (left). Gel images for cleavage of R8, R12, R16 and R20 with AK104 (right). Conditions: 100 pmol deoxyribozyme, 10 pmol RNA substrate, indicated concentration of $MgCl_2$, 50 mM Tris-HCl, 150 mM NaCl, pH 7.5, 37 °C, 6 h.

SUPPORTING INFORMATION

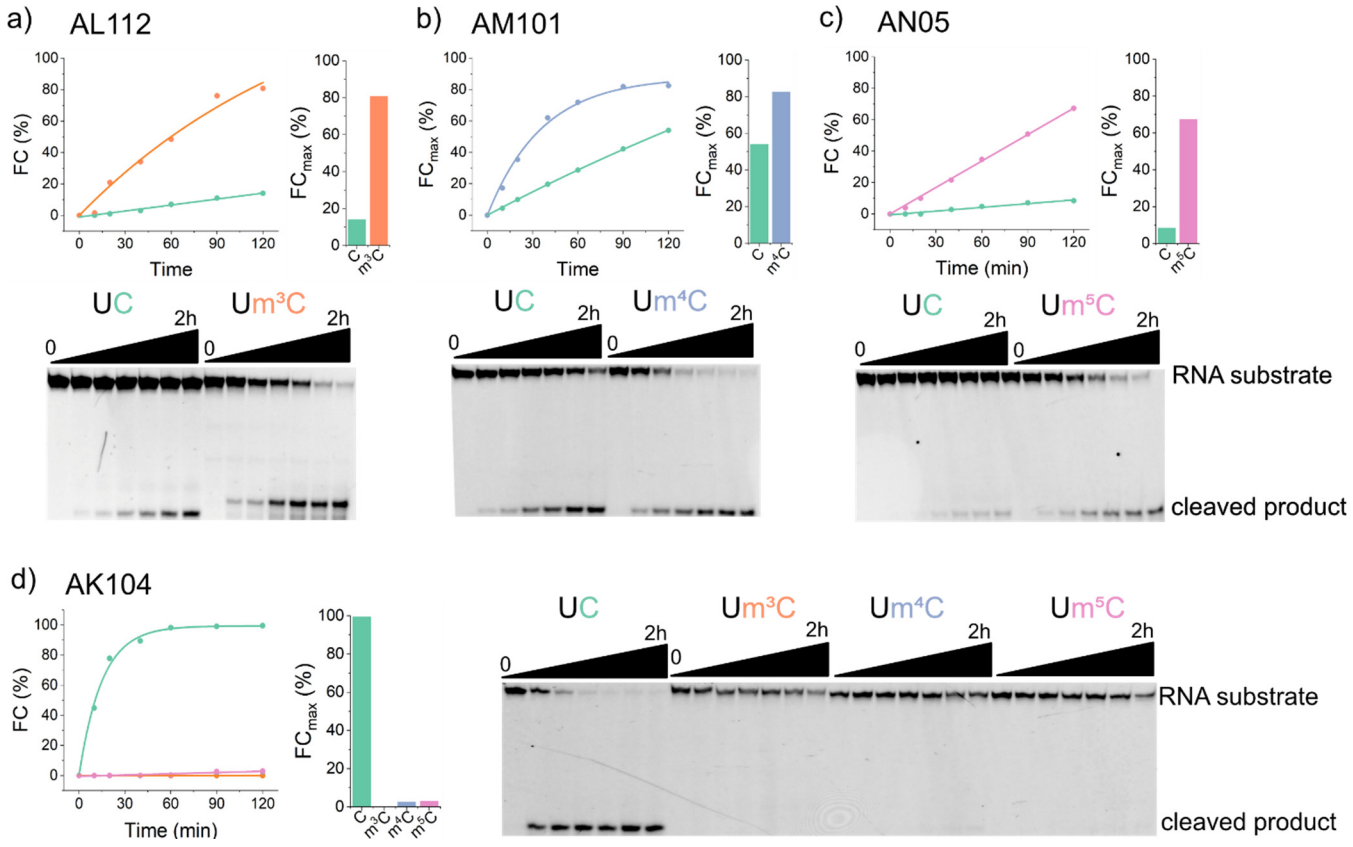
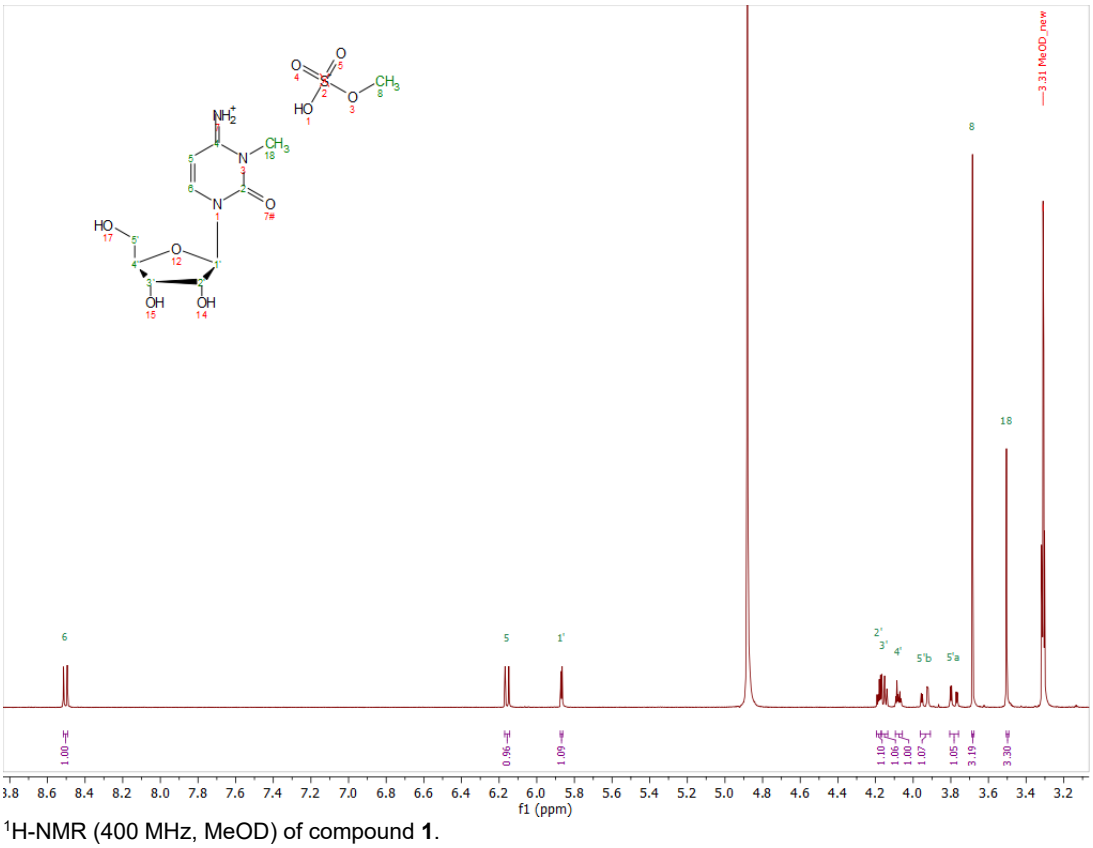
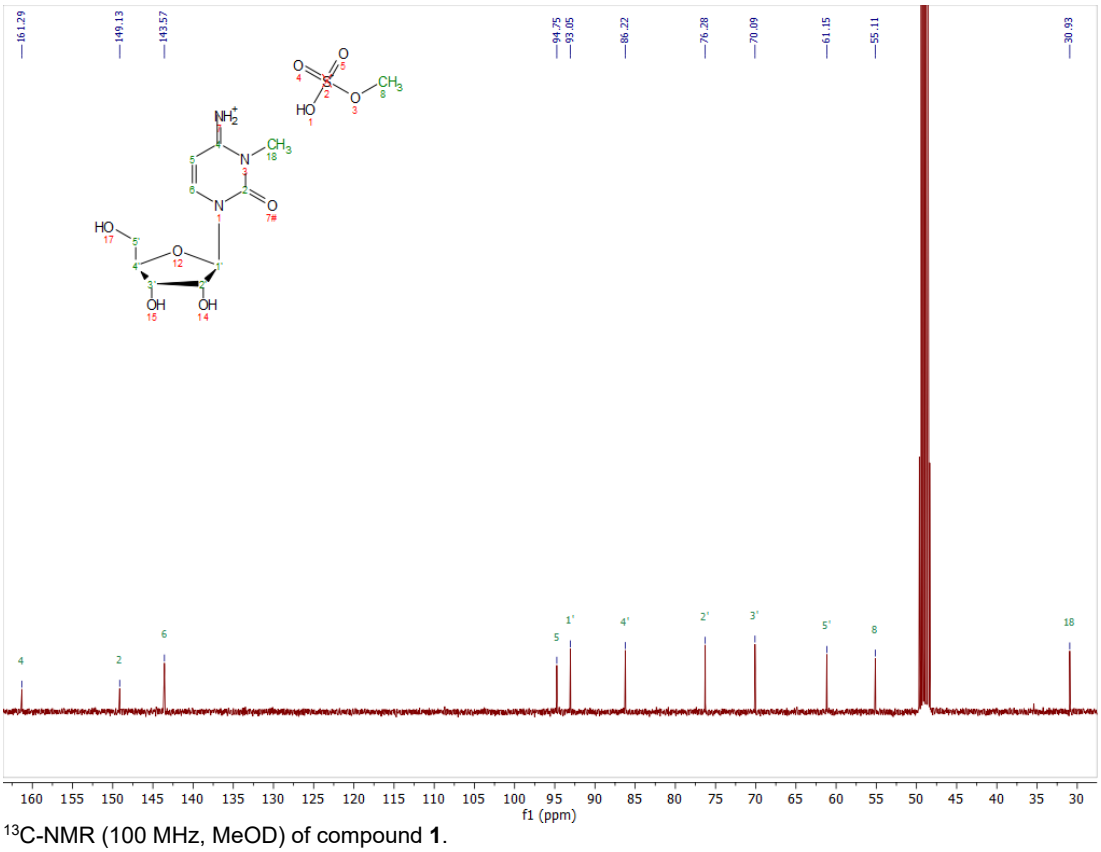


Figure S13. Cleavage activity of deoxyribozymes AL112, AM101, AN05 and AK104 in the presence of Mn^{2+}. a)−c) Top panel: kinetic plots and comparison of FC values obtained after 2 h cleavage for deoxyribozymes AL112, AM101 and AN05, respectively. Bottom panel: gel images for cleavage of R8 and R12 with AL112 (a), R8 and R16 with AM101 (b), R8 and R20 with AN05 (c). d) kinetic plots and comparison of FC values obtained after 2 h cleavage for deoxyribozyme AK104 (left). Gel image for cleavage of R8, R12, R16 and R20 with AK104 (right). Conditions: 100 pmol deoxyribozyme, 10 pmol RNA substrate, 20 mM $MnCl_2$, 50 mM Tris-HCl, 150 mM NaCl, pH 7.5, 37 °C, 2 h.

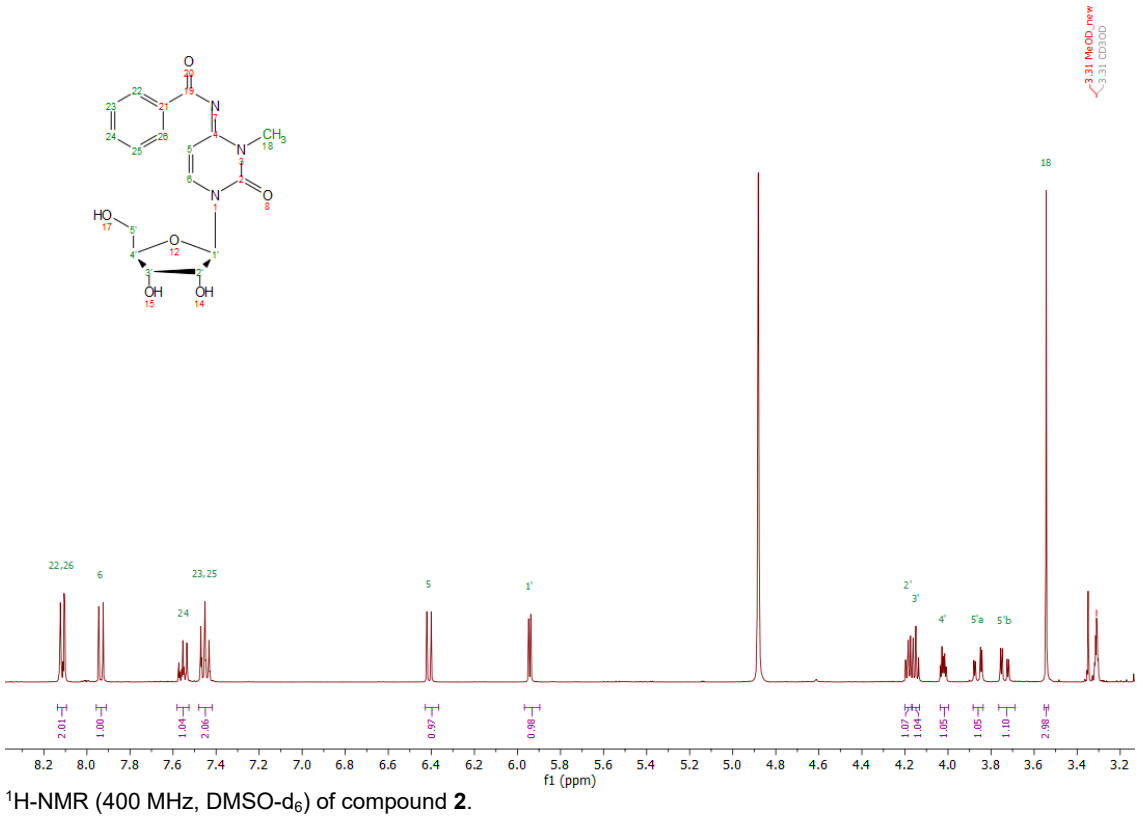
NMR Spectra



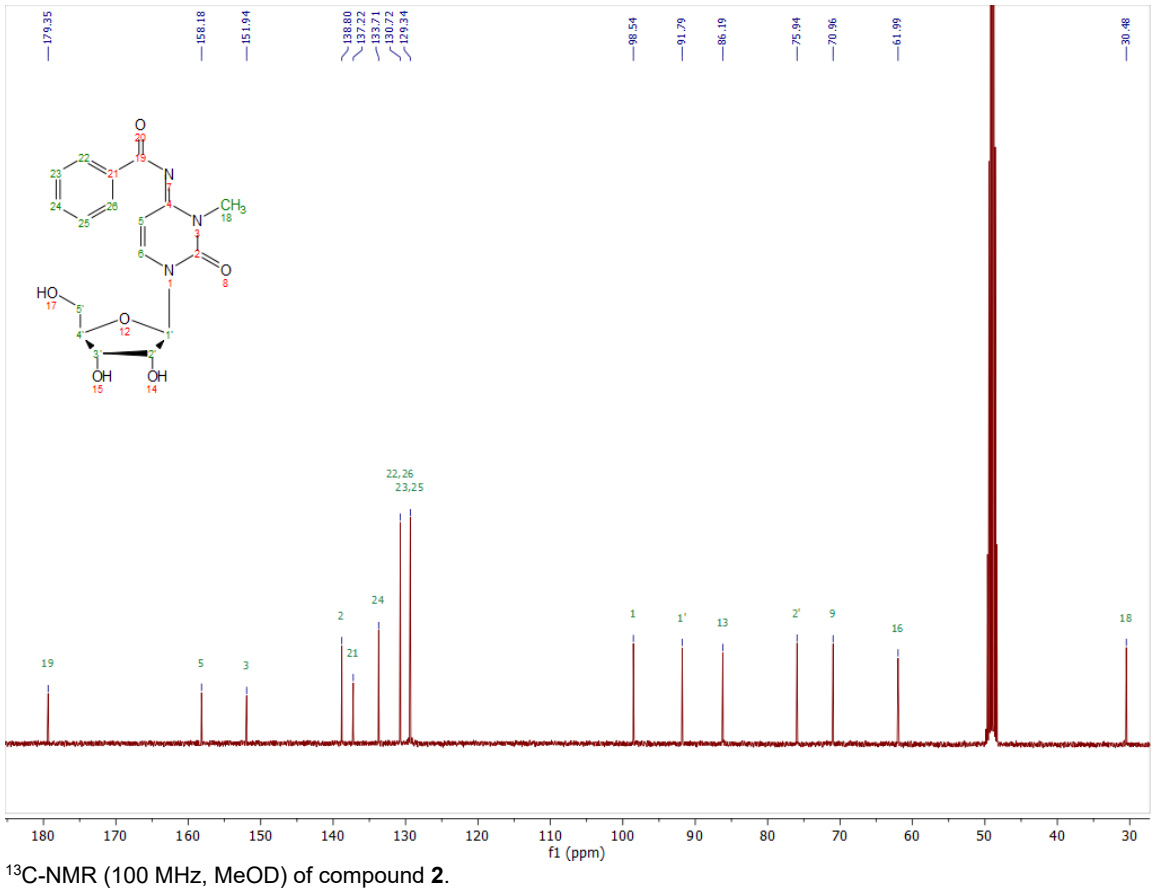
^{1}H-NMR (400 MHz, MeOD) of compound **1**.



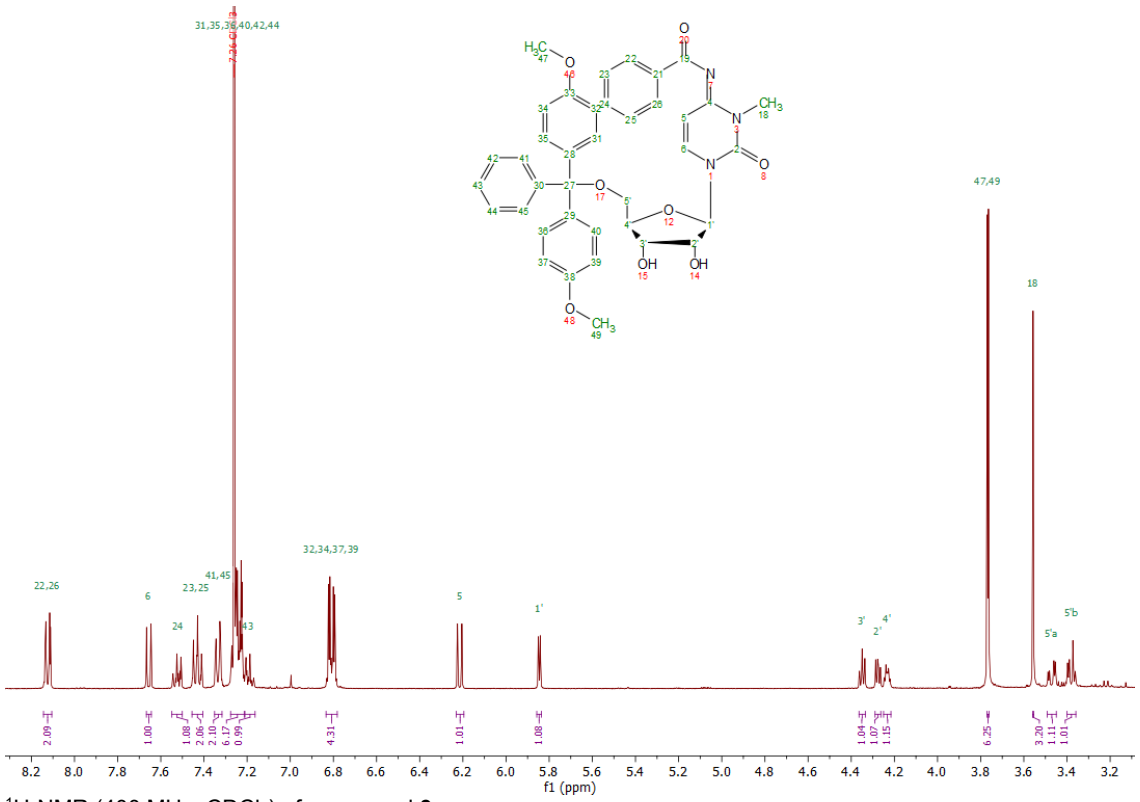
^{13}C-NMR (100 MHz, MeOD) of compound **1**.



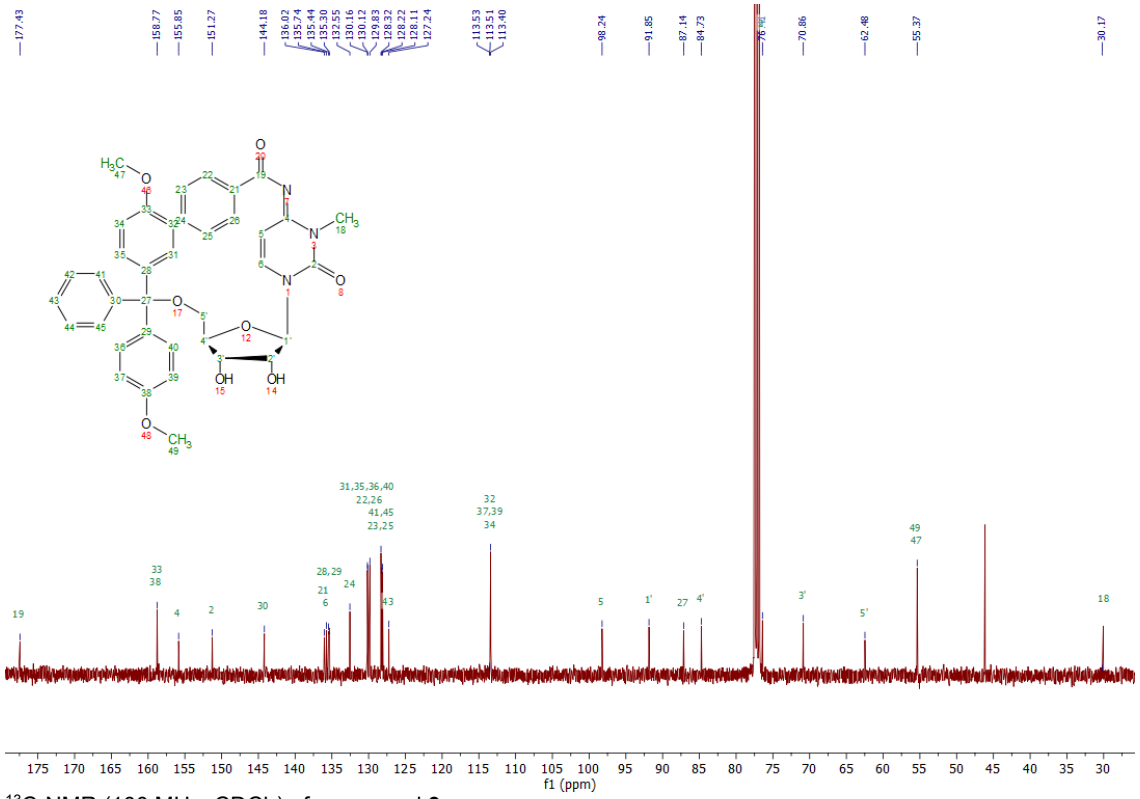
^{1}H-NMR (400 MHz, DMSO-d_6) of compound **2**.



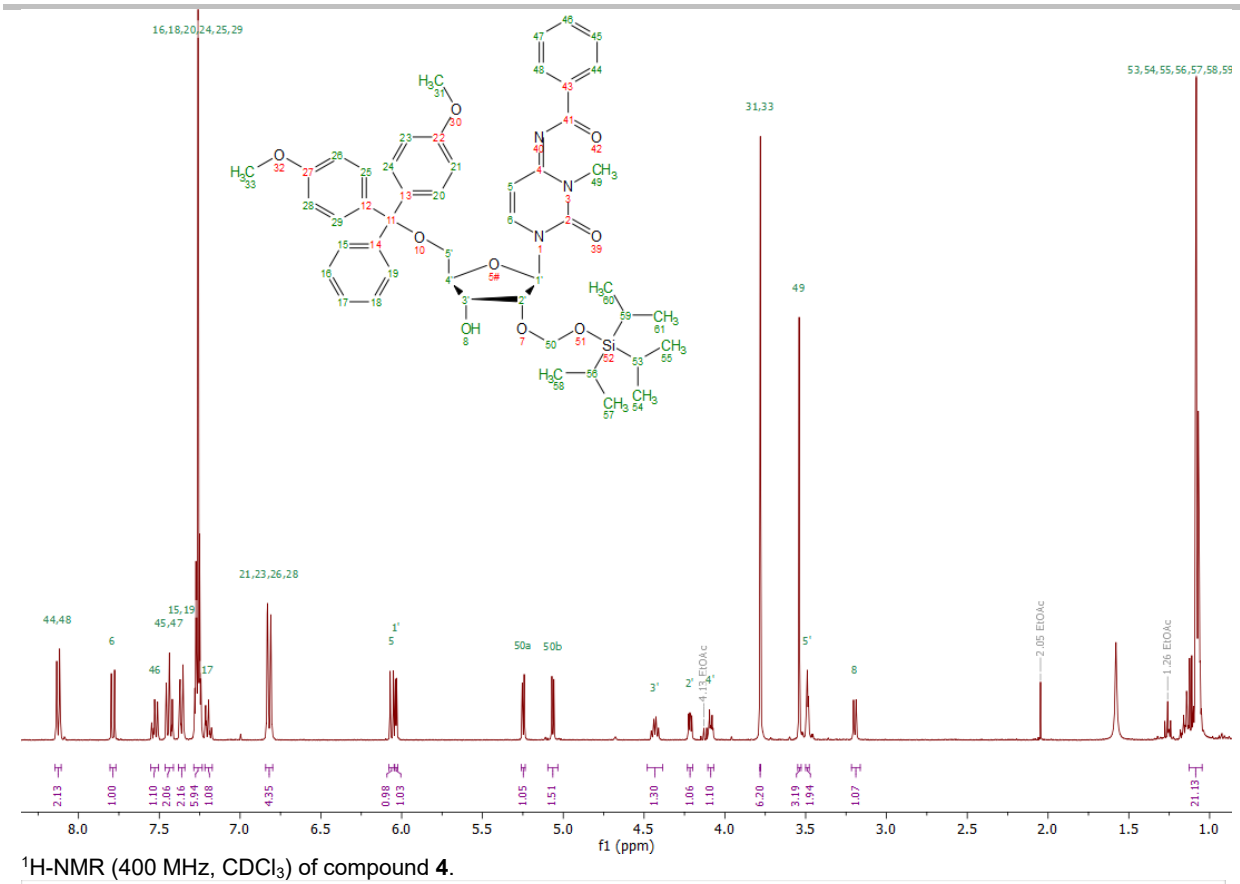
^{13}C-NMR (100 MHz, MeOD) of compound **2**.



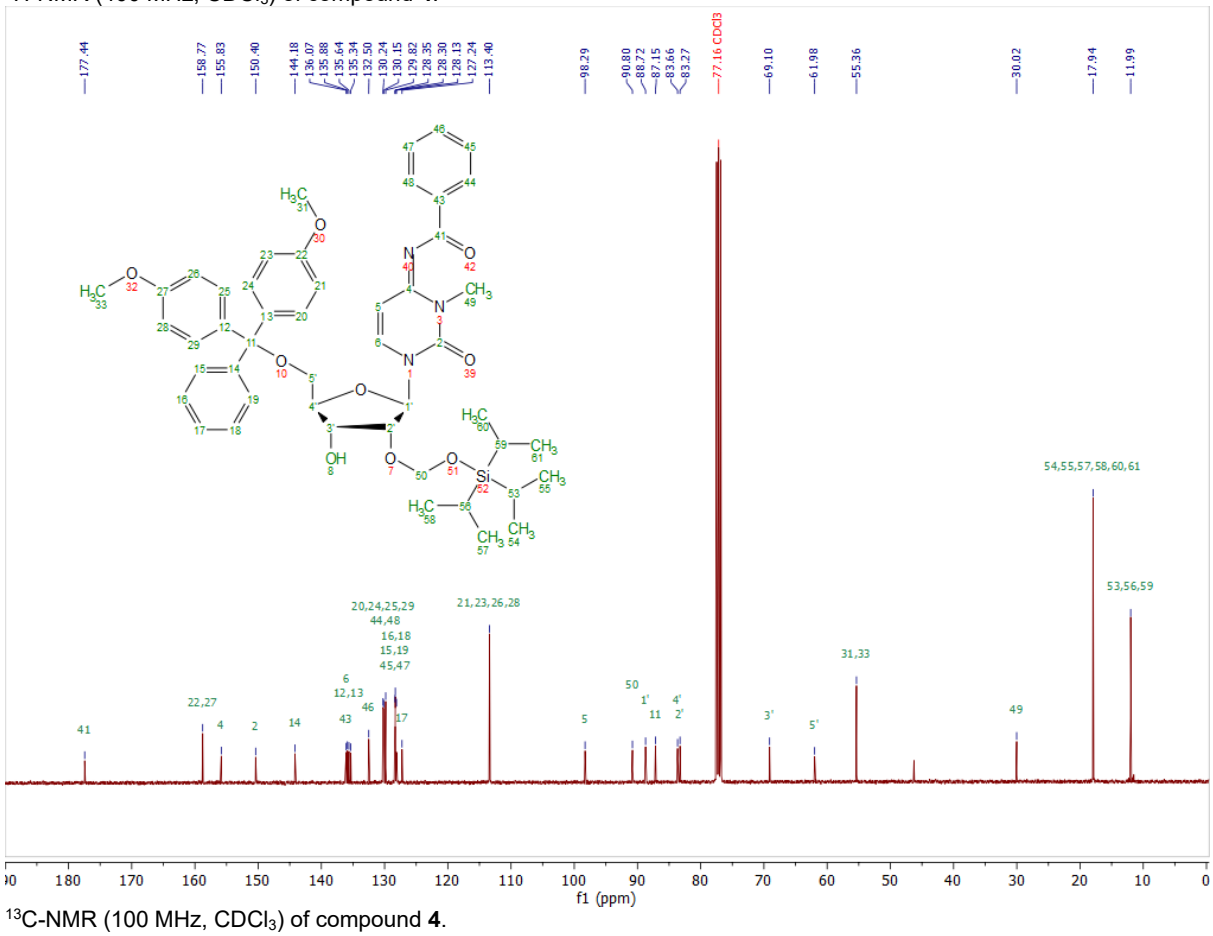
^{1}H-NMR (400 MHz, $CDCl_3$) of compound **3**.



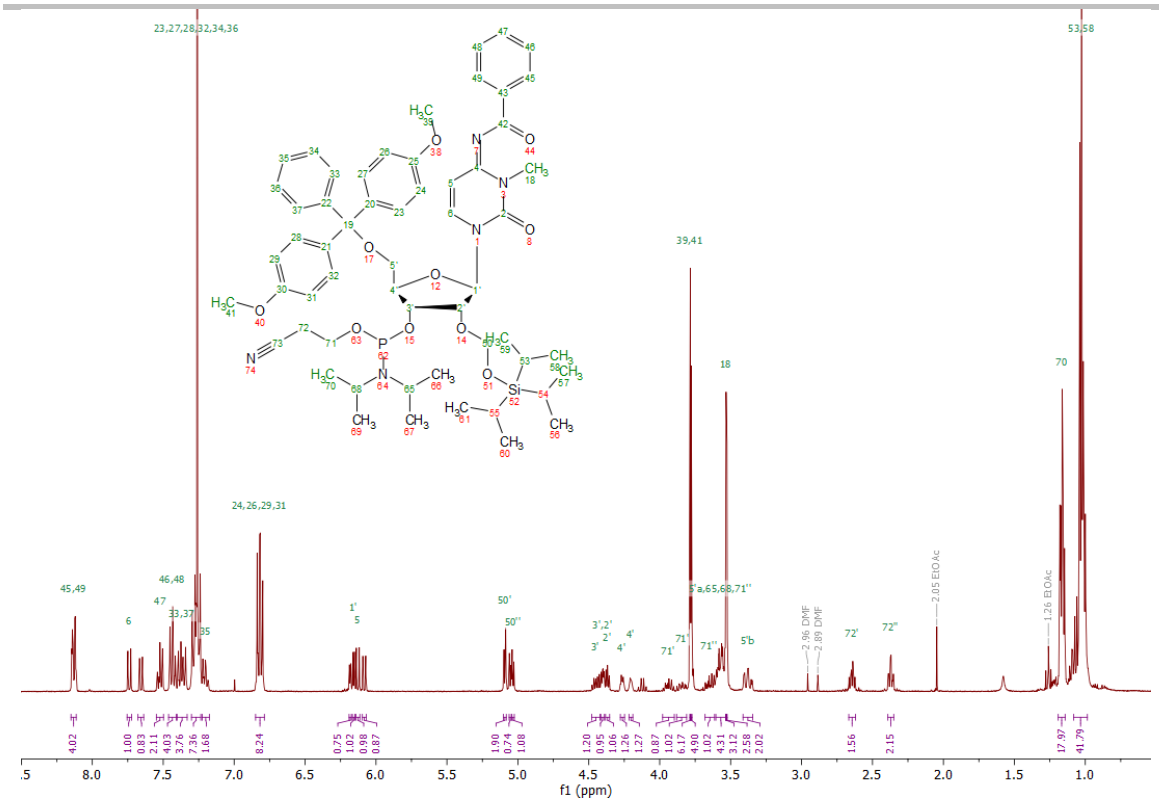
^{13}C-NMR (100 MHz, $CDCl_3$) of compound **3**.



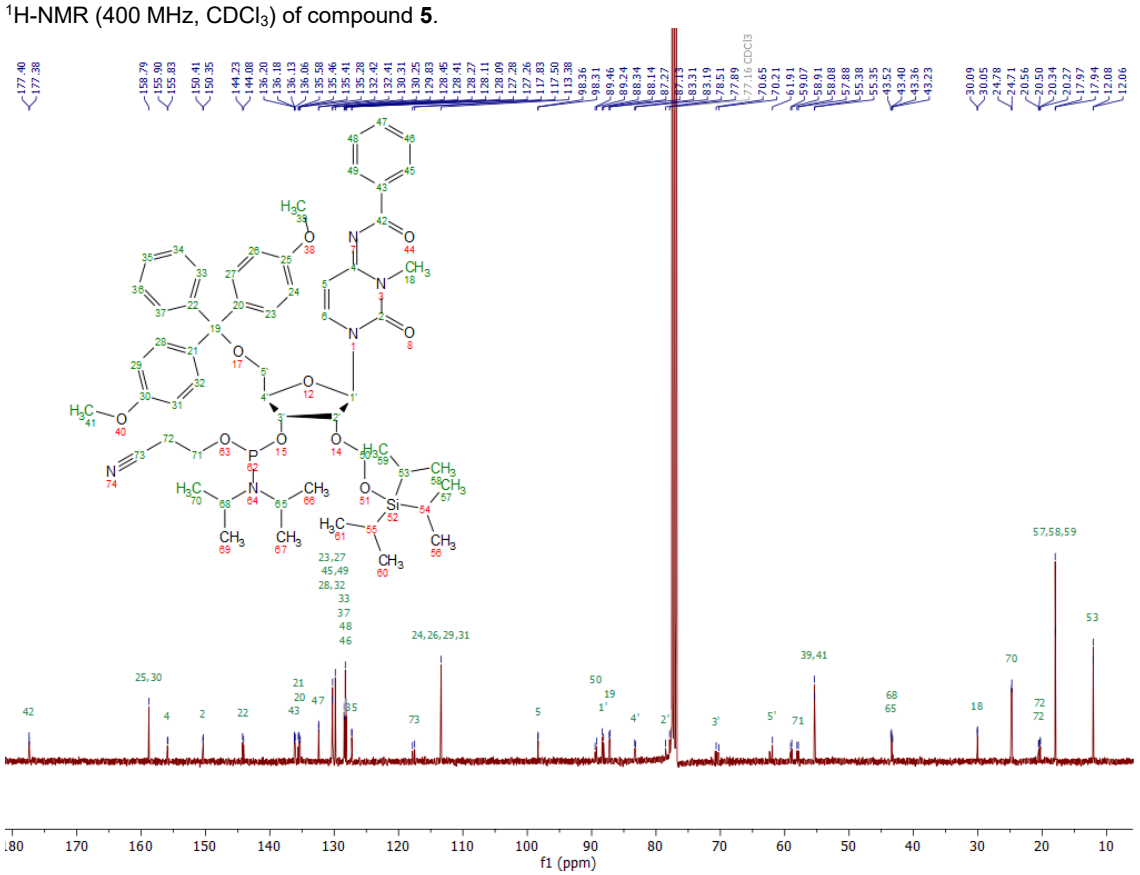
^{1}H-NMR (400 MHz, $CDCl_3$) of compound **4**.



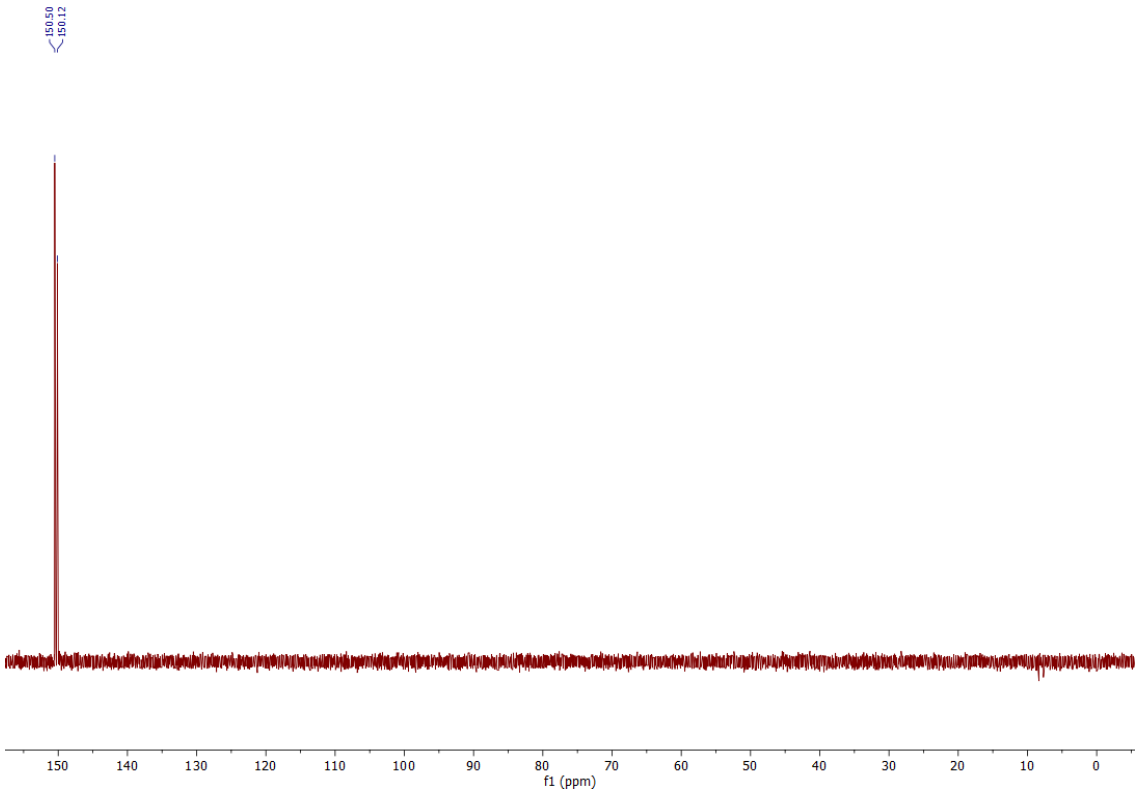
^{13}C-NMR (100 MHz, $CDCl_3$) of compound **4**.



^{1}H-NMR (400 MHz, $CDCl_3$) of compound **5**.



^{13}C-NMR (100 MHz, $CDCl_3$) of compound **5**.

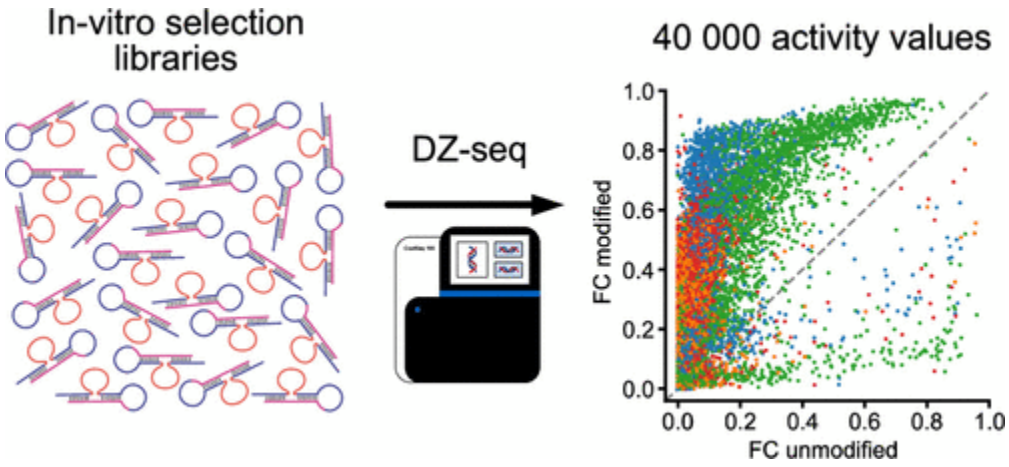


^{31}P-NMR (162 MHz, $CDCl_3$) of compound **5**.

References

[1] C. Höbartner, C. Kreutz, E. Flecker, E. Ottenschläger, W. Pils, K. Grubmayr, R. Micura, *Monatshefte für Chemie / Chemical Monthly* **2003**, *134*, 851-873.
[2] M. Alvira, R. Eritja, *Chemistry & Biodiversity* **2007**, *4*, 2798-2809.
[3] L. Büttner, J. Seikowski, K. Wawrzyniak, A. Ochmann, C. Höbartner, *Bioorganic & Medicinal Chemistry* **2013**, *21*, 6171-6180.
[4] P. Brookes, P. D. Lawley, *J. Chem. Soc.* **1962**, 1348-1351.
[5] X.-F. Zhu, H. J. Williams, A. I. Scott, *Synthetic Communications* **2003**, *33*, 1233-1243.
[6] M. Ghaem Maghami, C. P. M. Scheitl, C. Höbartner, *Journal of the American Chemical Society* **2019**, *141*, 19546-19549.
[7] a) C. Zhou, J. L. Avins, P. C. Klauser, B. M. Brandsen, Y. Lee, S. K. Silverman, *Journal of the American Chemical Society* **2016**, *138*, 2106-2109; b) J. Chandrasekar, S. K. Silverman, *Proceedings of the National Academy of Sciences* **2013**, *110*, 5315.
[8] a) M. V. Sednev, V. Mykhailiuk, P. Choudhury, J. Halang, K. E. Sloan, M. T. Bohnsack, C. Höbartner, *Angewandte Chemie International Edition* **2018**, *57*, 15117-15121; b) A. Liaqat, C. Stiller, M. Michel, M. V. Sednev, C. Höbartner, *Angewandte Chemie International Edition* **2020**, *59*, 18627-18631.

5. Paper-III High-Throughput Activity Profiling of RNA-Cleaving DNA Catalysts by Deoxyribozyme Sequencing (DZ-seq)



JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY
pubs.acs.org/JACS
Communication

High-Throughput Activity Profiling of RNA-Cleaving DNA Catalysts by Deoxyribozyme Sequencing (DZ-seq)

Maksim V. Sednev,*[,†] Anam Liaqat,[†] and Claudia Höbartner*

Cite This: J. Am. Chem. Soc. 2022, 144, 2090−2094
Read Online
ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: RNA-cleaving deoxyribozymes have found broad application as useful tools for RNA biochemistry. However, tedious in vitro selection procedures combined with laborious characterization of individual candidate catalysts hinder the discovery of novel catalytic motifs. Here, we present a new high-throughput sequencing method, DZ-seq, which directly measures activity and localizes cleavage sites of thousands of deoxyribozymes. DZ-seq exploits A-tailing followed by reverse transcription with an oligo-dT primer to capture the cleavage status and sequences of both deoxyribozyme and RNA substrate. We validated DZ-seq by conventional analytical methods and demonstrated its utility by discovery of novel deoxyribozymes that allow for cleaving challenging RNA targets or the analysis of RNA modification states.

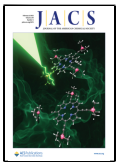
DNA enzymes or deoxyribozymes are single-stranded DNA sequences that form three-dimensional active sites for the catalysis of chemical reactions. Although deoxyribozymes have not been yet found in nature, in vitro selection procedures allowed the discovery of artificial deoxyribozymes with a vast repertoire of catalytic activities.[1−4] A typical in vitro selection starts from a combinatorial library of up to ∼10^{15} unique sequences and consists of repetitive rounds of activity screening, selection, and amplification.[5] Next-generation sequencing (NGS) techniques have been increasingly replacing the low-throughput Sanger method in the analysis of in vitro selection libraries of aptamers[6−11] and (deoxy)ribozymes.[12−16] The candidate catalysts are usually selected from the NGS data using the relative sequence abundance or sequence enrichment as proxies for the catalytic activity. However, due to PCR and other biases,[17,18] the candidate catalysts have to be validated by conventional low-throughput biochemical assays. Novel NGS-based strategies provided direct activity measurement for large mutant libraries of RNA-cleaving ribozymes,[19,20] DNA-cleaving deoxyribozymes[21] and in vitro selection libraries of aminoacylating ribozymes,[22] but RNA-cleaving deoxyribozymes still lack a comparable technique.

In our efforts to develop tools for site-specific detection of epigenetic modifications in RNA, we performed in vitro selections to identify RNA-cleaving deoxyribozymes that differentiate between modified and unmodified RNA.[23−25] Recently, we reported DNAzymes that distinguish between unmodified cytosine (C), 3-methylcytidine (m^3C), N^4-methylcytidine (m^4C), and 5-methylcytidine (m^5C).[25] Using Sanger sequencing and analysis of sequence enrichment in the NGS data with subsequent characterization of candidate catalysts by gel shift-based kinetic assays, we found few DNA enzymes with the desired properties. Screening of further deoxyribozymes was time-consuming and laborious as every candidate sequence had to be individually assayed for its cleavage activity.

To facilitate the discovery of further RNA-cleaving deoxyribozymes, we implemented deoxyribozyme sequencing (DZ-seq), an NGS-based strategy that allows massively parallel quantitation of RNA cleavage activity directly from sequencing data. In the first step of DZ-seq, the in vitro selection library with the attached 5′-sequencing adapter was subjected to self-cleavage (Figure 1). Next, the 2′,3′-cyclic phosphate residues resulting from the intramolecular attack of an RNA 2′-hydroxyl group on the adjacent phosphodiester linkage,[26] were removed by T4 PNK, and both cleaved and uncleaved species were extended with poly(A) tails using *E. coli* poly(A) polymerase. The resulting poly(A) tails were then used as primer binding sites for the anchored oligo(dT) primer (D4, Table S1), which introduced the second sequencing adapter to the resulting cDNA during reverse transcription. After this step, the cDNA contained information about sequences of deoxyribozyme and RNA substrate, cleavage status of RNA, and position of the cleavage site. Finally, the cDNA was used as a template in a PCR with two primers (D5 and D6) which introduced sequencing barcodes to produce sequencing-ready fragments.

We applied DZ-seq to analyze end point activity (10 mM Mg^{2+}, 6 h, 37 °C) of in vitro selection pools AK, AL, AM, and AN, which have been trained previously to cleave NC-, Nm^3C-, Nm^4C-, and Nm^5C-dinucleotide containing all-RNA substrates, respectively (N = A, C, G, U).[25] We prepared in total 16 sequencing libraries by combining eight deoxyribozyme libraries from rounds 7 and 18 of each in vitro selection with four RNA substrates (R1−R4, Table S2) in various combinations. Eight sequencing libraries contained DNA pools

Received: November 27, 2021
Published: January 26, 2022



https://doi.org/10.1021/jacs.1c12489
J. Am. Chem. Soc. 2022, 144, 2090−2094

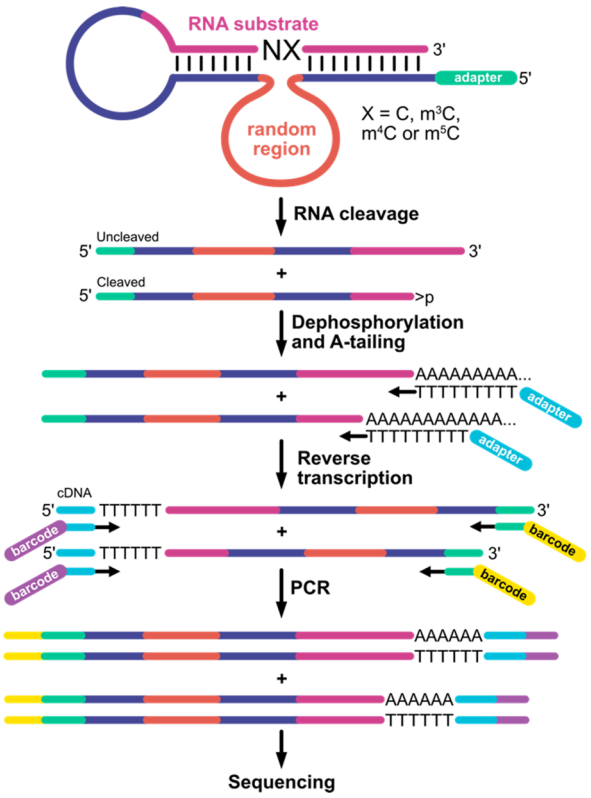


Figure 1. Outline of DZ-seq. ">p" denotes 2′,3′-cyclic phosphate residue.

ligated to the unmodified NC-substrate R1. Six libraries were prepared by combination of DNA pools from AL, AM, and AN selections with their corresponding cognate RNA substrates R2−R4, respectively. Further two sequencing libraries contained DNA pools from the selection AK and Nm^{5}C-RNA R4.

The sequencing-ready DNA fragments were subjected to NGS, and a total number of ∼220 million reads with ∼10−20 million reads per library were obtained (Table S3). The raw reads were trimmed of binding arm and connecting loop sequences and dereplicated to obtain unique sequences and their counts. For every unique combination of deoxyribozyme and RNA substrate we determined the total numbers of cleaved and intact reads and the position of the cleavage site. For about 40 000 deoxyribozymes at least one such combination had at least 100 reads. This data was then used for calculation of cleaving activity, which was expressed as fraction of cleaved reads (FC).

We validated the FC values obtained from DZ-seq by conventional PAGE-based kinetic assays. To this end, we chose ten deoxyribozymes AM301, AM319−AM327 (D4−D13) with varying predicted activities and tested them individually in cis-cleavage reactions, i.e., when the RNA substrate was covalently attached to the deoxyribozyme. DZ-seq correctly identified the position of the cleavage site for all ten cleavage reactions, but systematically overestimated FC values in comparison to PAGE (Figure 2A and Table S4). The overestimation can be attributed to additional cleavage events which occur during the dephosphorylation and A-tailing steps of DZ-seq as the corresponding reaction buffers contain Mg^{2+} ions and thus promote deoxyribozyme-catalyzed RNA cleavage. Nevertheless, the FC values obtained from DZ-seq displayed a good linear correlation with the PAGE assay (R^2 = 0.91, Figure 2A). Therefore, the obtained DZ-seq data set provides a quantitative insight into catalytic activity of tens of thousands of RNA-cleaving deoxyribozymes.

To illustrate the distribution of cleaving activities and specificities of DNA enzymes, we plotted FC values for the m^{3}C-, m^{4}C-, or m^{5}C-modified RNA substrates against those for the unmodified RNA (Figure 2B). The top left region of the plot for the m^{3}C modification is densely populated corresponding to numerous highly active DNA enzymes which are also highly specific for m^{3}C. In contrast, the majority of data points in the plots for m^{4}C and m^{5}C is concentrated along the diagonal line corresponding to low specificity of the DNAzymes in the AM and AN selection pools for the modified substrate. This observation is consistent with the stronger influence of the positively charged and base pair-disrupting m^{3}C modification on secondary structure of nucleic acids as compared to m^{4}C and m^{5}C.[27] The high number of very active m^{3}C-specific deoxyribozymes may indicate that m^{3}C-containing RNA is an "easier" target for DNA-catalyzed cleavage than are m^{4}C-, m^{5}C-modified, or unmodified RNA. Interestingly, few active deoxyribozymes that were inhibited by m^{3}C survived selection pressure in the AL selection, as several data points appear at the very right bottom of the corresponding plot.

DZ-seq interrogates the activity of RNA-cleaving deoxyribozymes in intramolecular reactions (in-cis) with the RNA substrates. For detection of RNA modifications an intermolecular arrangement of the DNA enzyme and the RNA substrate (in-trans) is more advantageous because it allows straightforward examination of RNA modifications without additional laborious experimental steps. Cis- and trans-activities do not necessarily correlate with each other, as the tertiary interactions between the connecting loop and the catalytic core may be critical for proper folding into a functional catalytic structure.[28] Surprisingly, some of the deoxyribozymes used for the validation of the DZ-seq results showed even higher FC values when they were tested with the corresponding RNA substrates in-trans under pseudo-first-order conditions (Table S4). However, most of the data points in the DZ-seq data set agreed well with the in-trans selectivity profiles of the previously characterized deoxyribozymes (exemplary data points are highlighted in Figure 2B for deoxyribozymes AL112, AM101 and AK104).[25] Hence, DZ-seq can be used for identification of deoxyribozymes that differentiate between modified and unmodified RNA substrates in intermolecular cleavage reactions.

To this end, we picked candidate sequences AM328−AM332 (D14−D18), which were predicted by DZ-seq to preferentially cleave Gm^{3}C-, Am^{3}C-, Cm^{3}C-, Gm^{4}C-, and Am^{5}C-containing RNAs, respectively (Figure 2C). Among them, the Gm^{4}C- and Am^{5}C-contexts were particularly challenging for specific cleavage with the previously reported DNA enzymes.[25] The new candidate DNA enzymes were individually tested in-trans in gel-based kinetic assays with their cognate and the corresponding unmodified RNA substrates (Table S5). As expected, all five selected deoxyribozymes preferentially cleaved the modified RNA substrates, although in the case of AM332 the FC value reported by PAGE (20%) was lower than that of DZ-seq (61%), which may be attributed to the influence of the connecting loop. Noteworthy, the DNA sequences here chosen for further analysis were low-abundant with no more than 5−300 counts per million (CPM) in the

https://doi.org/10.1021/jacs.1c12489
J. Am. Chem. Soc. 2022, 144, 2090−2094

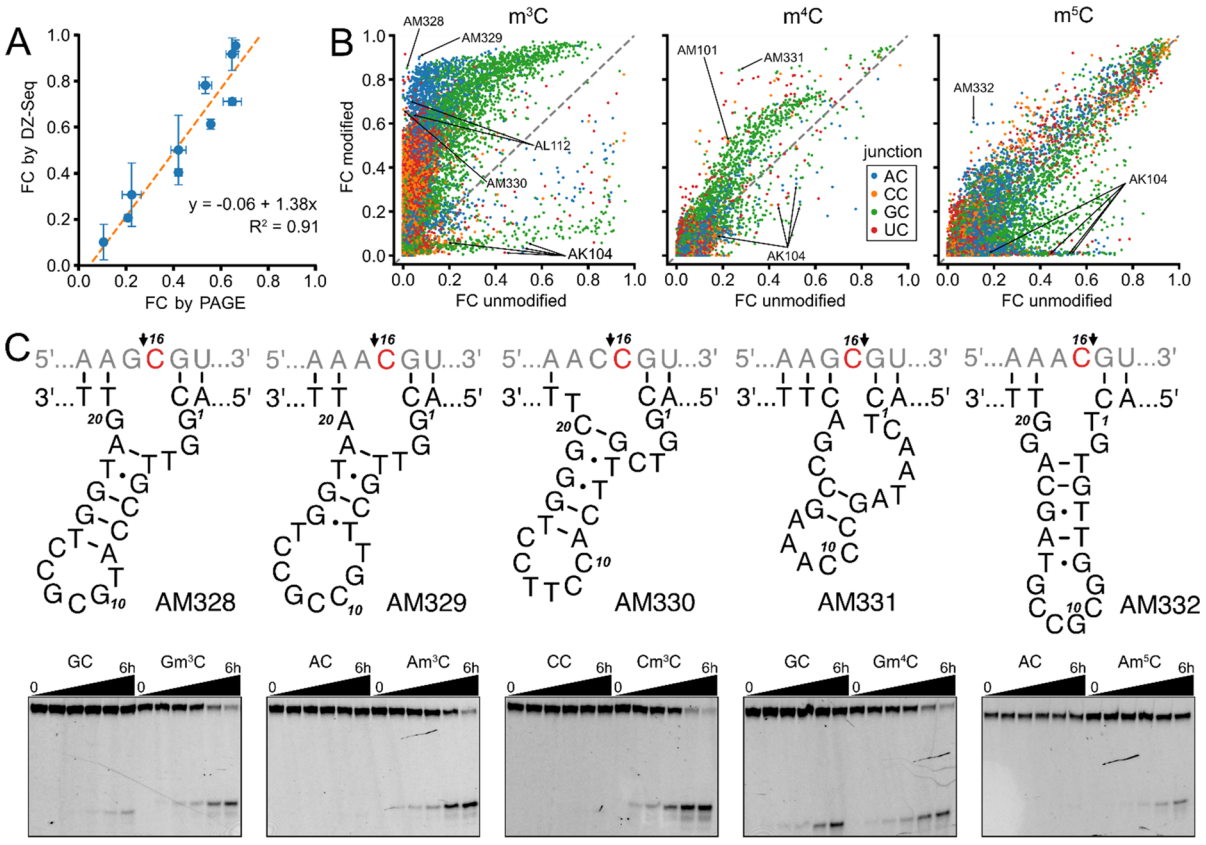


Figure 2. (A) Correlation of cis-cleavage yields for ten combinations of deoxyribozymes and RNA substrates measured by DZ-seq and PAGE. (B) DZ-seq data set visualized as cleavage yields for modified RNA plotted vs cleavage yields for unmodified RNA for each modification (m^3C, m^4C, or m^5C). (C) Predicted secondary structures of deoxyribozymes AM328−AM332 (top) and their trans-cleavage activity with 3′-fluorescently labeled modified and unmodified RNA substrates analyzed by gel-based kinetic assays (bottom). Modified cytosine site is highlighted in red. Cleavage sites reported by DZ-seq are indicated by arrows. Cleaved fragments containing m^3C migrate slower in PAGE, i.e., are shifted upward due to the positive charge at the nucleobase.

final selection rounds (Table S6), and thus most likely would have never been identified using conventional Sanger sequencing, which typically yielded sequences with much higher abundance (5000−30 000 CPM). Also, analysis of NGS data solely based on abundance and/or enrichment would have missed these candidates.

In fact, DZ-seq revealed a poor correlation between end point cleaving activity and abundance of deoxyribozymes in the final rounds of in vitro selections AK, AL, AM, and AN (Figure S2). A representative example is deoxyribozyme AM301 (Figure 3A), which displayed FC values that were among the highest found in the DZ-seq data set. Specifically, AM301 cleaved CG junction in Am^5C-, CC-, Cm^4C-, Cm^5C-, Um^4C-, and Um^5C-RNA substrates with very high yields (FCs of 80−98% after 6 h according to the PAGE assay, see Table S5). Despite these favorable properties, AM301 was depleted in most of the selection pools and its abundance in the final rounds was very low (~90 CPM). This indicates that cleaving activity is not the only decisive factor for the outcome of the in vitro selection, and that PCR amplification and/or ligation may distort the pool composition toward less active species.[29] In this respect, it would be compelling to analyze the similar in vitro selections of the pre-NGS era[30,31] with DZ-seq for the presence of depleted but nevertheless highly active (deoxy)-ribozyme motifs.

DZ-seq uncovered another low-abundant (<18 CPM) and moderately active (FCs up to 60% after 6 h, Table S5)

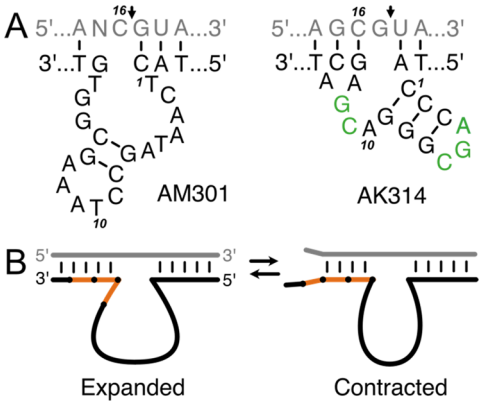


Figure 3. (A) Predicted secondary structures of deoxyribozymes AM301 and AK314. (B) Contraction of a deoxyribozyme's catalytic core due to strand slippage on tandem repeats. Repeating elements are shown as orange segments.

deoxyribozyme which had unusually short 13-nt catalytic core emerged from the originally 20-nt randomized region of the starting selection library (AK314, Figure 3A).[25] AK314 possesses the structural features that are characteristic for the 8−17 family: a 3 bp stem flanked by an AGC triple in the apical loop and a CG-dinucleotide in the 3′-direction of the stem (shown green in Figure 3A).[32] In fact, the sequence and the secondary structure of AK314 are identical up to a single

https://doi.org/10.1021/jacs.1c12489
J. Am. Chem. Soc. 2022, 144, 2090−2094

base pair swap to those of the 8−17 variant discovered by Santoro et al.[30] The unusually short length of AK314 raised the question about the mechanism that led to a such significant contraction of the in vitro selection library. Screening of the sequencing data set afforded a more abundant variant of AK314 with an additional TTCA fragment at the 3′-end of the catalytic core (AK314L, Table S1). Coincidentally, the 5′-end of the adjacent binding arm started with two further consecutive TTCA fragments. The resulting tandem repeat $(TTCA)_3$ is prone to strand slippage[33] and therefore might have induced contraction of AK314L to form AK314 during the PCR step of the in vitro selection. Moreover, the same repeat may provide possibility for the formation of an alternative fold so that the TTCA fragment of the core is displaced into the binding arm and the actual catalytic core of AK314L is identical with that of AK314 (Figure 3B).

In conclusion, in the present study we reported DZ-seq, a high-throughput NGS-based method suitable for profiling activity of RNA-cleaving deoxyribozymes in in vitro selection libraries. DZ-seq provided reliable information on cleavage activity and cleavage sites of thousands of deoxyribozymes. The novel method was successfully applied for discovery of rare DNA catalytic motifs that were inaccessible by conventional methods. We expect DZ-seq can be applied for comprehensive examination of in vitro selection or other combinatorial libraries and thus enable rapid discovery of catalysts with desired properties, e.g., deoxyribozymes specifically cleaving a modified RNA. Since DZ-seq should be indifferent toward the type of nucleic acid catalyst (i.e., deoxyribozyme or ribozyme), DZ-seq can be also applicable for discovery of novel RNA-cleaving motifs.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/jacs.1c12489.

Detailed methods, sequences, NGS data analysis, representative gel images, abundance plots, tables with abundance and activity data (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

Maksim V. Sednev − *Institute of Organic Chemistry, University of Würzburg, Am Hubland 97074 Würzburg, Germany*; Email: maksim.sednev@uni-wuerzburg.de

Claudia Höbartner − *Institute of Organic Chemistry, University of Würzburg, Am Hubland 97074 Würzburg, Germany*; orcid.org/0000-0002-4548-2299; Email: claudia.hoebartner@uni-wuerzburg.de

Author

Anam Liaqat − *Institute of Organic Chemistry, University of Würzburg, Am Hubland 97074 Würzburg, Germany*

Complete contact information is available at: https://pubs.acs.org/10.1021/jacs.1c12489

Author Contributions

†M.V.S. and A.L. contributed equally to this work.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was supported by the DFG (SPP1784) and by the ERC (Grant No. 682586). M.V.S. thanks the Graduate School of Life Sciences at the University of Würzburg for a Postdoc Plus fellowship. A.L. acknowledges funding by a PhD scholarship from the German Academic Exchange Service (DAAD). Illumina sequencing was performed at the Core Unit Systems Medicine and at the Biocenter of the University of Würzburg.

■ REFERENCES

(1) Silverman, S. K. Catalytic DNA: Scope, Applications, and Biochemistry of Deoxyribozymes. *Trends Biochem. Sci.* **2016**, *41* (7), 595−609.

(2) Hollenstein, M. DNA Catalysis: The Chemical Repertoire of DNAzymes. *Molecules* **2015**, *20* (11), 20777−20804.

(3) Schlosser, K.; Li, Y. Biologically Inspired Synthetic Enzymes Made from DNA. *Chem. Biol.* **2009**, *16* (3), 311−322.

(4) Micura, R.; Höbartner, C. Fundamental studies of functional nucleic acids: aptamers, riboswitches, ribozymes and DNAzymes. *Chem. Soc. Rev.* **2020**, *49* (20), 7331−7353.

(5) Jijakli, K.; Khraiwesh, B.; Fu, W.; Luo, L.; Alzahmi, A.; Koussa, J.; Chaiboonchoe, A.; Kirmizialtin, S.; Yen, L.; Salehi-Ashtiani, K. The in vitro selection world. *Methods* **2016**, *106*, 3−13.

(6) Rosenthal, M.; Pfeiffer, F.; Mayer, G. A Receptor-Guided Design Strategy for Ligand Identification. *Angew. Chem., Int. Ed.* **2019**, *58* (31), 10752−10755.

(7) Schmitz, A.; Weber, A.; Bayin, M.; Breuers, S.; Fieberg, V.; Famulok, M.; Mayer, G. A SARS-coV-2 spike binding DNA aptamer that inhibits pseudovirus infection by an RBD-independent mechanism. *Angew. Chem., Int. Ed.* **2021**, *60* (18), 10279−10285.

(8) Kacherovsky, N.; Yang, L. F.; Dang, H. V.; Cheng, E. L.; Cardle, I. I.; Walls, A. C.; McCallum, M.; Sellers, D. L.; DiMaio, F.; Salipante, S. J.; Corti, D.; Veesler, D.; Pun, S. H. Discovery and Characterization of Spike N-Terminal Domain-Binding Aptamers for Rapid SARS-CoV-2 Detection. *Angew. Chem., Int. Ed.* **2021**, *60* (39), 21211−21215.

(9) Schütze, T.; Wilhelm, B.; Greiner, N.; Braun, H.; Peter, F.; Mörl, M.; Erdmann, V. A.; Lehrach, H.; Konthur, Z.; Menger, M.; Arndt, P. F.; Glökler, J. Probing the SELEX Process with Next-Generation Sequencing. *PLoS One* **2011**, *6* (12), No. e29604.

(10) Blind, M.; Blank, M. Aptamer Selection Technology and Recent Advances. *Mol. Ther. Nucleic Acids* **2015**, *4*, No. e223.

(11) Liu, M.; Yin, Q.; Chang, Y.; Zhang, Q.; Brennan, J. D.; Li, Y. In Vitro Selection of Circular DNA Aptamers for Biosensing Applications. *Angew. Chem., Int. Ed.* **2019**, *58* (24), 8013−8017.

(12) Streckerová, T.; Kurfürst, J.; Curtis, E. A. Single-round deoxyribozyme discovery. *Nucleic Acids Res.* **2021**, *49* (12), 6971−6981.

(13) Zhang, C.; Li, Q.; Xu, T.; Li, W.; He, Y.; Gu, H. New DNA-hydrolyzing DNAs isolated from an ssDNA library carrying a terminal hybridization stem. *Nucleic Acids Res.* **2021**, *49* (11), 6364−6374.

(14) Popović, M.; Ellingson, A. Q.; Chu, T. P.; Wei, C.; Pohorille, A.; Ditzler, M. A. In vitro selections with RNAs of variable length converge on a robust catalytic core. *Nucleic Acids Res.* **2021**, *49* (2), 674−683.

(15) Inomata, R.; Zhao, J.; Miyagishi, M. Zn^{2+}-dependent DNAzymes that cleave all combinations of ribonucleotides. *Commun. Biol.* **2021**, *4* (1), 221.

(16) Ameta, S.; Winz, M.-L.; Previti, C.; Jäschke, A. Next-generation sequencing reveals how RNA catalysts evolve from random space. *Nucleic Acids Res.* **2014**, *42* (2), 1303−1310.

(17) Tolle, F.; Wilke, J.; Wengel, J.; Mayer, G. By-Product Formation in Repetitive PCR Amplification of DNA Libraries during SELEX. *PLoS One* **2014**, *9* (12), No. e114693.

(18) Komarova, N.; Kuznetsov, A. Inside the Black Box: What Makes SELEX Better? *Molecules* **2019**, *24* (19), 3598.

(19) Kobori, S.; Nomura, Y.; Miu, A.; Yokobayashi, Y. High-throughput assay and engineering of self-cleaving ribozymes by sequencing. *Nucleic Acids Res.* **2015**, *43* (13), No. e85.
(20) Kobori, S.; Yokobayashi, Y. High-Throughput Mutational Analysis of a Twister Ribozyme. *Angew. Chem., Int. Ed.* **2016**, *55* (35), 10354−10357.
(21) Dhamodharan, V.; Kobori, S.; Yokobayashi, Y. Large Scale Mutational and Kinetic Analysis of a Self-Hydrolyzing Deoxyribozyme. *ACS Chem. Biol.* **2017**, *12* (12), 2940−2945.
(22) Pressman, A. D.; Liu, Z.; Janzen, E.; Blanco, C.; Müller, U. F.; Joyce, G. F.; Pascal, R.; Chen, I. A. Mapping a Systematic Ribozyme Fitness Landscape Reveals a Frustrated Evolutionary Network for Self-Aminoacylating RNA. *J. Am. Chem. Soc.* **2019**, *141* (15), 6213−6223.
(23) Sednev, M. V.; Mykhailiuk, V.; Choudhury, P.; Halang, J.; Sloan, K. E.; Bohnsack, M. T.; Höbartner, C. N^6-Methyladenosine-Sensitive RNA-Cleaving Deoxyribozymes. *Angew. Chem., Int. Ed.* **2018**, *57* (46), 15117−15121.
(24) Liaqat, A.; Stiller, C.; Michel, M.; Sednev, M. V.; Höbartner, C. N^6-Isopentenyladenosine in RNA Determines the Cleavage Site of Endonuclease Deoxyribozymes. *Angew. Chem., Int. Ed.* **2020**, *59* (42), 18627−18631.
(25) Liaqat, A.; Sednev, M. V.; Stiller, C.; Höbartner, C. RNA-Cleaving Deoxyribozymes Differentiate Methylated Cytidine Isomers in RNA. *Angew. Chem., Int. Ed.* **2021**, *60* (35), 19058−19062.
(26) Silverman, S. K. In vitro selection, characterization, and application of deoxyribozymes that cleave RNA. *Nucleic Acids Res.* **2005**, *33* (19), 6151−6163.
(27) Mao, S.; Haruehanroengra, P.; Ranganathan, S. V.; Shen, F.; Begley, T. J.; Sheng, J. Base Pairing and Functional Insights into N^3-Methylcytidine (m^3C) in RNA. *ACS Chem. Biol.* **2021**, *16* (1), 76−85.
(28) De la Peña, M.; Gago, S.; Flores, R. Peripheral regions of natural hammerhead ribozymes greatly increase their self-cleavage activity. *EMBO J.* **2003**, *22* (20), 5561−5570.
(29) Takahashi, M.; Wu, X.; Ho, M.; Chomchan, P.; Rossi, J. J.; Burnett, J. C.; Zhou, J. High throughput sequencing analysis of RNA libraries reveals the influences of initial library and PCR methods on SELEX efficiency. *Sci. Rep.* **2016**, *6* (1), 33697.
(30) Santoro, S. W.; Joyce, G. F. A general purpose RNA-cleaving DNA enzyme. *Proc. Natl. Acad. Sci. U. S. A.* **1997**, *94* (9), 4262−4266.
(31) Pan, T.; Uhlenbeck, O. C. A small metalloribozyme with a two-step mechanism. *Nature* **1992**, *358* (6387), 560−563.
(32) Schlosser, K.; Li, Y. A Versatile Endoribonuclease Mimic Made of DNA: Characteristics and Applications of the 8−17 RNA-Cleaving DNAzyme. *ChemBioChem.* **2010**, *11* (7), 866−879.
(33) Shinde, D.; Lai, Y.; Sun, F.; Arnheim, N. Taq DNA polymerase slippage mutation rates measured by PCR and quasi-likelihood analysis: (CA/GT)n and (A/T)n microsatellites. *Nucleic Acids Res.* **2003**, *31* (3), 974−980.

https://doi.org/10.1021/jacs.1c12489
J. Am. Chem. Soc. 2022, 144, 2090−2094

Supporting information for

High-throughput activity profiling of RNA-cleaving DNA catalysts by deoxyribozyme sequencing (DZ-seq)

Maksim V. Sednev[‡*], Anam Liaqat[‡] and Claudia Höbartner[*]

Institute of Organic Chemistry, University of Würzburg, Am Hubland,

97074 Würzburg, Germany

Table of Contents

General information

DNA oligonucleotides were purchased from Microsynth. Synthesis of RNA oligonucleotides is described earlier.[1] Fluorescein thiosemicarbazide was purchased from Sigma-Aldrich. Nucleotide triphosphates (NTPs) and deoxyribonucleotide triphosphates (dNTPs) were purchased from Jena Bioscience. T4 PNK, T4 DNA Ligase, Superscript III, and DreamTaq polymerase were purchased from Thermo Fisher Scientific. *E. coli* Poly(A) Polymerase and RNase inhibitor (recombinant, human placenta) were purchased from New England Biolabs. γ-^{32}P-Adenosine 5'-triphosphate (γ-^{32}P-ATP) was purchased from Hartmann Analytic. All DNA and RNA oligonucleotides were purified by denaturing PAGE (10–20% acrylamide/bis-acrylamide 19:1, 7 M urea) with running buffer 1x TBE (89 mM Tris, 89 mM boric acid and 2 mM EDTA, pH 8.3), extracted by crush and soak into TEN buffer (10 mM Tris-HCl, pH 8.0, 1 mM EDTA, 300 mM NaCl) and recovered by precipitation with ethanol. For preparative gels, oligonucleotides were visualized by UV shadowing on a TLC plate. Fluorescence gel images were recorded with a Biorad ChemiDoc MP using epi illumination. Analytical anion-exchange HPLC was carried out on a GE Healthcare Life Sciences ÄKTA™ micro system using DNAPac™ PA200 columns (4 x 250 mm, Thermo Scientific) at 60 °C. Solvent A: 25 mM Tris-HCl (pH 8.0), 6 M Urea. Solvent B: 25 mM Tris-HCl (pH 8.0), 6 M Urea, 0.5 M $NaClO_4$. Gradient: linear, 0–54% solvent B, 4% solvent B per 1 CV. The solvents were filtered through 0.2 µm cellulose acetate filters (Sartorius, Germany) prior to usage. HR-ESI-MS spectra of the synthetic products and oligonucleotides were recorded on a Bruker micrOTOF-Q III spectrometer.

Preparation of sequencing libraries for DZ-seq

<u>Phosphorylation of RNA substrates</u>

Phosphorylation at 5'-end of RNA substrates was carried out with T4 PNK as follows: RNA (2–5 nmol) was incubated with 10x PNK buffer A (5 µl, 500 mM Tris-HCl, 10 mM $MgCl_2$, 50 mM

DTT, 1 mM spermidine), ATP (5 µl, 10 mM) and T4 PNK (5 µl, 10 U/µl) at 37 °C for 5 hours. Total volume of the reaction mixture was 50 µl. The phosphorylated RNA was extracted with phenol/chloroform/isoamylalcohol mixture and ethanol precipitated.

Introduction of 5'-sequencing adapter

5'-Adapter was introduced by an asymmetric PCR reaction. *In-vitro* selection back-up sample (obtained after initial PCR amplification of the enriched library, see Liaqat et al.[1]) was used as the template. The template (2 µl), forward primer D1 (200 pmol), reverse tailed primer D2 (50 pmol), dNTP mixture (2 µl, 10 mM each dNTP), 10x DreamTaq buffer (10 µl) and DreamTaq DNA polymerase (1 µl, 5 U/µl) were mixed, and the total volume was adjusted to 100 µl with water. The resulting mixture was subjected to a 30-cycle PCR reaction (95 °C for 4 minutes, 30 x [95 °C for 30 s, 52 °C for 30 s, 72 °C for 1 min] and 72 °C for 5 minutes) using a Bio-Rad thermal cycler. The PCR product was subjected to denaturing PAGE and the short strand was isolated.

Splinted ligation of RNA substrate and deoxyribozyme library

Splinted ligation was carried out with T4 DNA ligase as described earlier.[1-3] In a typical procedure, the deoxyribozyme library with 5'-adapter, 5'-phosphorylated RNA and complementary DNA splint D3 (1/2 of the amount obtained after the previous step, 75 pmol and 50 pmol, respectively) were dissolved in water (9 µl), and 10x annealing buffer (1 µl; 40 mM Tris-HCl, 150 mM NaCl, 1 mM EDTA, pH 8.0) was added. After 2 min incubation at 95 °C, the solution was allowed to cool down to r.t. for 15 min and then further cooled down to 0 °C for 10 min. 10x DNA Ligase buffer (2 µl, 400 mM Tris-HCl, 100 mM $MgCl_2$, 100 mM DTT, 5 mM ATP, pH 7.8), T4 DNA Ligase (1 µl, 5 U/µl), RNase inhibitor (0.5 µl, 40 U/µl) and water to a total volume of 20 µl were added, and the resulting mixture incubated overnight at 37 °C to yield a DNA-RNA hybrid product which was purified by denaturing PAGE.

Intramolecular cleavage of DNA-RNA hybrids

The DNA-RNA hybrid (10–30 pmol) was incubated with 10x in-vitro selection buffer (2 µl, 500 mM Tris-HCl, 1.5 M NaCl, pH 7.5) in a total volume of 18 µl. To initiate the reaction, $MgCl_2$ (2 µl, 100 mM) was added to the final concentration of 10 mM and the reaction mixture was incubated at 37 °C for 6 hours. Afterwards, the reaction mixture was quenched with TEN buffer (280 µl) and ethanol precipitated.

Removal of 2,3-cyclic phosphate residues

Dephosphorylation was carried out with T4 PNK as follows. The DNA-RNA hybrid library from the previous step was incubated with 10x PNK buffer A (5 µl, 500 mM Tris-HCl, 10 mM $MgCl_2$, 50 mM DTT, 1 mM spermidine), RNase inhibitor (0.5 µl, 40 U/µl) and T4 PNK (1 µl, 10 U/µl) in a total volume of 50 µl at 37 °C for 2 h. The reaction product was recovered using an Oligo Clean & Concentrator™ kit (Zymo Research).

A-tailing

The dephosphorylated DNA-RNA hybrid library (~180 ng) was incubated with ATP (0.5 µl, 1 mM), 10x reaction buffer (5 µl, 500 mM Tris-HCl, 2.5 M NaCl, 100 mM $MgCl_2$, pH 8.1) and *E. coli* Poly(A) Polymerase (0.5 µl, 5 U/µl) in total volume of 20 µl at 37 °C for 1 h. The reaction product was purified by an Oligo Clean & Concentrator™ kit (Zymo Research).

Introduction of 3'-adapter by reverse transcription

3'-Sequencing adapter was attached using reverse transcription with Superscript III as follows. The A-tailed DNA-RNA-hybrid library (1/2 of the amount obtained after A-tailing), anchored oligo(dT) primer D4 (10 pmol) and dNTP mixture (1 µl, 10 mM) were incubated at 95 °C for 3 min in a total volume of 17.5 µl and then cooled down to room temperature. 10x reaction buffer (2 µl, 500 mM Tris-HCl, 750 mM KCl, 30 mM $MgCl_2$, 100 mM DTT, pH 8.3) and Superscript III (0.5 µl, 200 U/µl) were added and the reaction mixture was incubated at 50 °C for 30 min. The resulting cDNA was isolated by PAGE.

Introduction of Illumina indices by PCR

The cDNA from the previous step, an indexed primer set (D5 and D6, 25 pmol each), dNTP mixture (1 µl, 10 mM), 10x DreamTaq buffer (5 µl) and DreamTaq polymerase (0.5 µl) were mixed, and the total reaction volume was adjusted to 50 µl with water. The resulting solution was subjected to a 5-cycle PCR reaction (95 °C for 4 minutes, 5 x [95 °C for 30 s, 60 °C for 30 s, 72 °C for 1 min] and 72 °C for 5 minutes) using a Bio-Rad thermal cycler. The reaction mixture was ethanol precipitated and the PCR product was isolated by PAGE followed by purification with an Oligo Clean & Concentrator™ kit (Zymo Research).

Sequencing and data analysis

The DZ-seq libraries were sequenced on an Illumina NextSeq 500 (75 cycles, single-end) with 10% PhiX. The demultiplexed reads were trimmed of constant sequences using cutadapt[4] and further analyzed using custom bash, R and python scripts (available on Github: https://github.com/msednev/DZ-seq). In short, the cleavage status of each read was identified by the presence of an A-tail at the 3'-end (Figure S1). The start position of the A-tail relative to the residual RNA sequence gave the position of the deoxyribozyme cleavage site. Fraction cleaved (FC) for every unique combination of deoxyribozyme and RNA substrate has been calculated from the numbers of corresponding intact and cleaved reads.

Labeling of oligonucleotides

Labeling of RNA by oxidation at 3'-end

Oligoribonucleotides were labeled by periodate oxidation followed by a reaction with fluorescein-5-thiosemicarbazide. In a typical procedure, RNA (0.3–1 nmol) was dissolved in 7.5 µl water, followed by addition of 5x sodium phosphate buffer (2 µl, 250 mM, pH 7.4) and a freshly prepared aqueous solution of $NaIO_4$ (0.5 µl, 400 mM). The resulting mixture was incubated for 10–15 min at 37 °C. The excess of $NaIO_4$ was quenched by treatment with aqueous Na_2SO_3 (1 µl, 1 M) for 5–10 min at 37 °C. Afterwards, a solution of fluorescein-5-thiosemicarbazide (1 µl, 10 mM) in DMF was added, and the reaction mixture was incubated in the dark for additional 1 h at 37 °C. The labeled product was then purified by PAGE (10–20% polyacrylamide) and recovered by extraction and ethanol precipitation.

Radioactive labeling of RNA at 5'-end

Radioactive labeling of RNA was performed by phosphorylation using γ-^{32}P-ATP. In a typical procedure, RNA (200–500 pmol) was incubated with 10x PNK buffer A (5 µl, 500 mM Tris-HCl, 10 mM $MgCl_2$, 50 mM DTT, 1 mM spermidine), γ-^{32}P-ATP (0.5 µl, 10 µCi/µl) and T4 PNK enzyme (0.5 µl, 10 U/µl) at 37 °C for 1 hours. The labeled RNA was purified by PAGE (20% polyacrylamide) and recovered by extraction and ethanol precipitation.

Kinetics characterization of deoxyribozyme-catalyzed RNA cleavage by gel-based assays

Intramolecular cleavage *(cis-activity)*

To determine intramolecular cleavage activity, deoxyribozymes were ligated to 5'-radioactively labeled RNA substrates using splinted ligation. 5'-Labeling of RNA was performed by using γ-^{32}P-ATP. To this end, RNA (200–500 pmol) was incubated with 10x PNK buffer A (1 µl, 500 mM Tris-HCl, 10 mM $MgCl_2$, 50 mM DTT, 1 mM spermidine), γ-^{32}P-ATP (0.5 µl, 10 µCi/µl) and T4 PNK enzyme (0.5 µl, 10 U/µl) at 37 °C for 1 hour. The phosphorylated RNA was purified by PAGE (20% polyacrylamide) and then covalently attached to DNAzymes by ligation. For this purpose, equimolar amounts of phosphorylated RNA, deoxyribozyme and complementary splint (D3) were dissolved in 9 µl water and annealed in 10x annealing buffer (1 µl, 40 mM Tris-HCl, 150 mM NaCl, 1 mM EDTA, pH 8.0). The ligation was initiated by addition of 10x DNA ligase buffer (2 µl, 400 mM Tris-HCl, 100 mM $MgCl_2$, 5 mM ATP, 100 mM DTT, pH 7.8) and T4 DNA ligase (1 µl, 5 U/µl) in a final volume of 20 µl. The resulting mixture was incubated for ~16 h at 37 °C and the products were isolated by denaturing PAGE (10% polyacrylamide). To initiate intramolecular cleavage, the RNA-DNA hybrid product (approx. 1 µCi) was incubated in a total volume of 20 µl in the presence of 10 mM $MgCl_2$, 50 mM Tris-HCl and 150 mM NaCl at pH 7.5 at 37 °C. After 6 h incubation, the reaction mixture was quenched with loading buffer and analyzed by denaturing PAGE. The gels were dried under vacuum for 30 min at 80 °C, exposed overnight to a phosphor storage screen and imaged on a Typhoon phosphorimager.

Intermolecular cleavage (trans-activity)

Kinetics characterization was carried out as described previously[5-6] to analyze trans-activity of individual deoxyribozymes or active deoxyribozyme pools under single turnover conditions. In a typical procedure, a deoxyribozyme (100 pmol) was mixed with a fluorescently labeled RNA substrate (10 pmol) in a final volume of 8 µl. The resulting mixture was heated to 95 °C for 4 min and then cooled to 25 °C for 10 min. To initiate the reaction, 10x kinetic assay buffer (1 µl, 500 mM Tris-HCl, 1.5 M NaCl), pH 7.5) and $MgCl_2$ (1 µl, 200 mM) were added, and the resulting mixture was incubated at 37 °C. Aliquots (1 µl) were taken at different time points (e.g. 0, 15 min, 30 min, 60 min, 180 min, and 360 min), quenched with loading buffer and analyzed by denaturing PAGE. Cleavage yields (FC) were determined by measuring fluorescent intensities of the corresponding bands. Each experiment was repeated 2-3 times, and FC values are reported as means of individual experiments.

Supporting tables

Table S1. Sequences of DNA oligonucleotides

No	Description	5'-Sequence-3'[a]
D1	Forward primer 1st PCR	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGTGACGCGACTAGTTAC
D2	Reverse primer 1nd PCR	(CAA)$_4$E$_3$GGAGTCGCCAACTGAATGAA
D3	Ligation splint	TTACGTTTCATTCAGTCTATTGGAGTCGCCAACT
D4	Anchored oligo(dT) primer	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTTTTTTTTTTTTTTTTTTTTTTTTTVN
D5	Forward primer 2nd PCR	AATGATACGGCGACCACCGAGATCTACAC[i5]TCGTCGGCAGCGTC
D6	Reverse primer 2nd PCR	CAAGCAGAAGACGGCATACGAGAT[i7]GTCTCGTGGGCTCGG
D4	AM301 cis	GTGACGCGACTAGTTAC TCAATAGCCTAAAGCGGTG TTCATTCAGTTGGCGACTCCA
D5	AM319	GTGACGCGACTAGTTAC TCAATAGCCCCAAAGCGGAT TTCATTCAGTTGGCGACTCCA
D6	AM320	GTGACGCGACTAGTTAC GGTTCGGAACGGCCTGGTAA TTCATTCAGTTGGCGACTCCA
D7	AM321	GTGACGCGACTAGTTAC CGTTGCATCGCGCGTGGTTA TTCATTCAGTTGGCGACTCCA
D8	AM322	GTGACGCGACTAGTTAC TCAATAGCCGAAAGCGGTA TTCATTCAGTTGGCGACTCCA
D9	AM323	GTGACGCGACTAGTTAC GGTTGCGCAGCGCCTGGTAA TTCATTCAGTTGGCGACTCCA
D10	AM324	GTGACGCGACTAGTTAC GGTTGTGTAACGCCTGGTAA TTCATTCAGTTGGCGACTCCA
D11	AM325	GTGACGCGACTAGTTAC CGGCTGGACGTCGACGCGGT TTCATTCAGTTGGCGACTCCA
D12	AM326	GTGACGCGACTAGTTAC TCAATAGCCCAAAGCGGTC TTCATTCAGTTGGCGACTCCA
D13	AM327	GTGACGCGACTAGTTAC GGTTGTGGAACGCCTGGCAA TTCATTCAGTTGGCGACTCCA
D14	AM328	GACTAGTTAC GGTTGCCATGCGCCTGGTAG TTCATTCAGT
D15	AM329	GACTAGTTAC GGTTGCTTGCCGCCTGGTAA TTCATTCAGT
D16	AM330	GACTAGTTAC GGTCGTTCACCTTCCTGGGC TTCATTCAGT
D17	AM331	GACTAGTTAC TCAATAGCCCAAAGCCGAC TTCATTCAGT
D18	AM332	GACTAGTTAC TGTGTTGGCGCCGTAGCAGG TTCATTCAGT
D19	AK314	GACTAGTTAC CCAGCGGGACGAGCA TTCATTCAGT
D20	AK314L	GACTAGTTAC CCAGCGGGACGAGCATTCA TTCATTCAGT
D21	AM301 trans	GACTAGTTAC TCAATAGCCTAAAGCGGTG TTCATTCAGT

[a] E$_3$ = triethylene glycol, i5 = CTCTCTAT or TATCCTCT, i7 = TCGCCTTA, CTAGTACG, TTCTGCCT, GCTCAGGA, AGGAGTCC, CATGCCTA, GTAGAGAG or CAGCCTCG.

Table S2. Sequences of RNA oligonucleotides

No	Description/motif	5'-Sequence-3'
R1	unmodified/NC	AUAGACUGAAUGAA**NC**GUAACUAGUC
R2	Nm^{3}C	AUAGACUGAAUGAA**Nm^{3}C**GUAACUAGUC
R3	Nm^{4}C	AUAGACUGAAUGAA**Nm^{4}C**GUAACUAGUC
R4	Nm^{5}C	AUAGACUGAAUGAA**Nm^{5}C**GUAACUAGUC
R5	unmodified/AC	AUAGACUGAAUGAA**AC**GUAACUAGUC
R6	unmodified/GC	AUAGACUGAAUGAA**GC**GUAACUAGUC
R7	unmodified/CC	AUAGACUGAAUGAA**CC**GUAACUAGUC
R8	unmodified/UC	AUAGACUGAAUGAA**UC**GUAACUAGUC
R9	Am^{3}C	AUAGACUGAAUGAA**Am^{3}C**GUAACUAGUC
R10	Gm^{3}C	AUAGACUGAAUGAA**Gm^{3}C**GUAACUAGUC
R11	Cm^{3}C	AUAGACUGAAUGAA**Cm^{3}C**GUAACUAGUC
R12	Um^{3}C	AUAGACUGAAUGAA**Um^{3}C**GUAACUAGUC
R13	Am^{4}C	AUAGACUGAAUGAA**Am^{4}C**GUAACUAGUC
R14	Gm^{4}C	AUAGACUGAAUGAA**Gm^{4}C**GUAACUAGUC
R15	Cm^{4}C	AUAGACUGAAUGAA**Cm^{4}C**GUAACUAGUC
R16	Um^{4}C	AUAGACUGAAUGAA**Um^{4}C**GUAACUAGUC
R17	Am^{5}C	AUAGACUGAAUGAA**Am^{5}C**GUAACUAGUC
R18	Gm^{5}C	AUAGACUGAAUGAA**Gm^{5}C**GUAACUAGUC
R19	Cm^{5}C	AUAGACUGAAUGAA**Cm^{5}C**GUAACUAGUC
R20	Um^{5}C	AUAGACUGAAUGAA**Um^{5}C**GUAACUAGUC

Table S3. Overview of sequencing libraries

Library	In-vitro selection	Round	RNA substrate	RNA modification	Number of reads	Number of unique deoxyribozymes[b]
L1	AK	7	R1	C	10 867 685	44 683
L2	AK	18	R1	C	16 660 578	187 132
L3	AK	7	R4	m^{5}C	9 460 402	37 640
L4	AK	18	R4	m^{5}C	20 765 478	231 246
L5	AL	7	R1	C	13 589 510	67 327
L6	AL	18	R1	C	15 496 535	49 429
L7	AL	7	R2	m^{3}C	13 134 905	64 920
L8[a]	AL	18	R2	m^{3}C	6 231 593 + 6 574 428	41 875
L9	AM	7	R1	C	12 069 675	124 241

Library	In-vitro selection	Round	RNA substrate	RNA modification	Number of reads	Number of unique deoxyribozymes[b]
L10	AM	18	R1	C	15 014 642	72 004
L11[a]	AM	7	R3	m^4C	6 516 617 + 4 914 923	120 364
L12	AM	18	R3	m^4C	12 283 917	60 696
L13	AN	7	R1	C	13 534 181	37 475
L14	AN	18	R1	C	13 413 247	98 602
L15	AN	7	R4	m^5C	18 575 228	46 669
L16	AN	18	R4	m^5C	14 606 354	110 872

[a] Sequenced in two different lanes to improve sequencing depth. [b] Deoxyribozymes with at least 10 counts are shown.

Table S4. Validation of sequencing results by PAGE-based activity assays. Representative gel images are shown in Figures S3 and S4

Deoxyribozyme	RNA substrate	DZ-seq				PAGE				
		cis-FC[a]	SD[b]	n[c]	Cleavage site[d]	*cis*-FC[a]	SD[b]	n[e]	Cleavage site[d]	*trans*-FC[f]
AM301	CC	0.95	0.02	4	ACC|GU	0.66	0.01	3	ACC|GU	0.81
AM319	Am^5C	0.78	0.04	2	AAm^5C|GU	0.53	0.03	3	AAm^5C|GU	0.74
AM320	Gm^4C	0.61	0.02	2	AG|m^4CGU	0.56	0.01	3	AG|m^4CGU	0.70
AM321	Gm^4C	0.40	0.02	2	AG|m^4CGU	0.42	0.01	3	AG|m^4CGU	0.40
AM322	Cm^5C	0.21	0.02	3	ACm^5C|GU	0.21	0.01	3	ACm^5C|GU	0.13
AM323	Gm^3C	0.92	0.07	2	AG|m^3CGU	0.65	0.03	3	AG|m^3CGU	0.86
AM324	Gm^4C	0.71	0.02	2	AG|m^4CGU	0.65	0.04	3	AG|m^4CGU	0.86
AM325	UC	0.50	0.15	5	AU|CGU	0.42	0.03	3	AU|CGU	0.25
AM326	CC	0.31	0.14	5	ACC|GU	0.22	0.04	2	ACC|GU	0.20
AM327	AC	0.10	0.08	5	AA|CGU	0.10	0.01	3	AA|CGU	0.03

[a] Mean fraction cleaved values for cis-cleavage after 6 h incubation (10 mM $MgCl_2$, 50 mM Tris-HCl, 150 mM NaCl, pH 7.5, 37 °C). [b] Standard deviation. [c] Number of occurrences in the DZ-seq dataset. [d] A fragment of the RNA substrate is shown. The cleavage site is indicated as the vertical bar. [e] Number of performed measurements. [f] Fraction cleaved for trans-cleavage after 6 h incubation (100 pmol deoxyribozyme, 10 pmol RNA substrate, 10 mM $MgCl_2$, 50 mM Tris-HCl, 150 mM NaCl, pH 7.5, 37 °C).

Table S5. Cleaving activity for selected deoxyribozymes. Representative gel images are shown in the main text and in Figures S5 and S6

Deoxyribozyme	RNA substrate	*cis*-FC by DZ-seq[a]	*trans*-FC by PAGE[b]
AM328	GC	0.02	0.08
	Gm^3C	0.85	0.53
AM329	AC	0.07	<0.01
	Am^3C	0.90	0.62
AM330	CC	0.00	<0.01
	Cm^3C	0.66	0.95
AM331	GC	0.27	0.25
	Gm^4C	0.84	0.61
AM332	AC	0.11	<0.01
	Am^5C	0.61	0.20
AK314	GC	0.62[c]	0.60
	Gm^5C	0.25[c]	0.48
AK314L	GC	0.31	0.47
	Gm^5C	0.31	0.38
AM301	AC	0.88	0.55
	Am^3C	0.44	0.05
	Am^4C	0.81[c]	0.53
	Am^5C	0.88	0.82
	GC	0.68	0.24
	Gm^3C	0.12	<0.01
	Gm^4C	0.34[c]	0.18
	Gm^5C	0.75	0.44
	CC	0.95	0.82
	Cm^3C	0.52	0.05

Deoxyribozyme	RNA substrate	*cis*-FC by DZ-seq[a]	*trans*-FC by PAGE[b]
	Cm^4C	0.89[c]	0.94
	Cm^5C	0.91	0.82
	UC	0.88	0.90
	Um^3C	0.34	<0.01
	Um^4C	0.89[c]	0.87
	Um^5C	0.82	0.89

[a] The mean value is given if the indicated combination of the DNA enzyme and the RNA substrate appeared several times in the DZ-seq dataset. [b] The mean of two measurements. [c] Less than 100 reads were available for calculation.

Table S6. Abundances of deoxyribozymes in pools AK, AL, AM and AN

Deoxyribozyme	Round	Abundance, CPM[a]			
		Pool AK	Pool AL	Pool AM	Pool AN
AM319	7	49.61	2.75	0.18	60.04
	18	0.25	n.d.	0.77	31.08
AM320	7	378.04	1194.30	111.58	731.02
	18	40.96	607.08	818.47	254.88
AM321	7	47.52	551.42	139.40	61.50
	18	0.54	136.33	112.42	6.88
AM322	7	182.66	14.99	0.57	152.88
	18	4.97	0.08	4.80	234.32
AM323	7	5443.21	2488.30	343.10	2958.97
	18	273.78	123.77	587.30	853.12
AM324	7	1112.58	1788.86	174.54	1599.55
	18	448.29	640.93	2052.04	2503.45
AM325	7	114.99	114.41	48.88	26.90
	18	4.27	92.76	90.20	137.58
AM326	7	916.06	126.06	4.94	358.02
	18	17.34	0.23	59.16	297.22
AM327	7	14.17	288.01	46.69	25.14
	18	2.13	414.91	37.47	8.90
AM328	7	1.12	22.06	26.30	2.34
	18	0.06	8.64	0.64	0.16
AM329	7	3.41	49.60	19.80	19.22
	18	0.06	5.54	0.85	0.16
AM330	7	0.55	4.64	1.69	0.08
	18	n.d	48.44	1.06	n.d
AM331	7	4.36	3.17	8.86	0.29
	18	0.20	8.88	286.15	5.60
AM332	7	73.04	0.23	n.d.	n.d.
	18	n.d.	n.d.	n.d.	n.d.
AK314	7	3.09	0.15	n.d.	1.02
	18	17.04	0.15	0.73	17.50
AK314L	7	895.54	2.45	2.54	740.63
	18	4408.19	51.52	838.85	6461.32
AM301	7	564.91	303.46	3.68	112.17
	18	6.47	0.11	5.51	82.06
AK104	7	5.01	27.40	194.13	38530.98
	18	7865.13	52.98	5.43	5.77
AL112	7	1.256620	30.935492	48.588913	0.74658
	18	n.d.	4.05	0.18	n.d.
AM101	7	8.11	804.50	18.24	1.48
	18	21.20	10.24	18339.24	309.88

[a] Mean counts per million.

Supporting figures

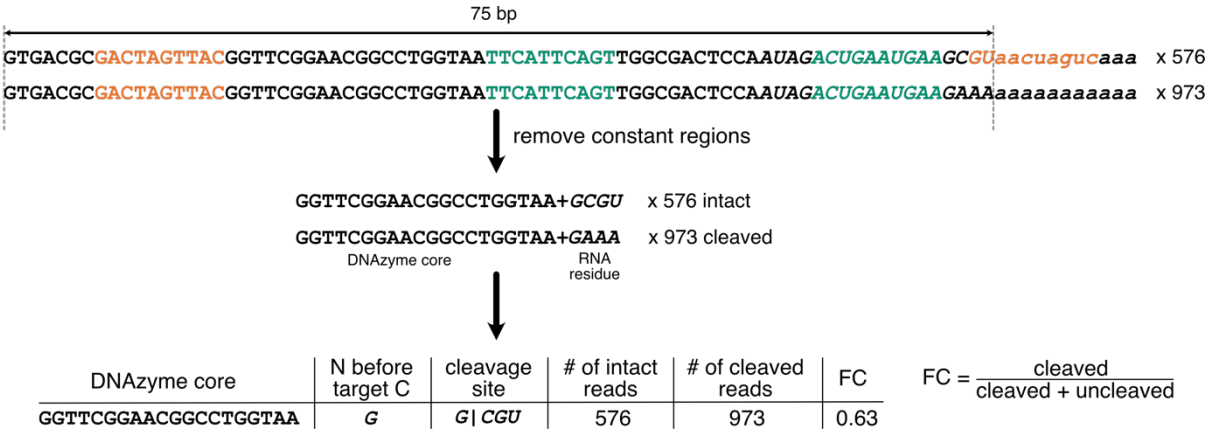


Figure S1. DZ-seq data analysis pipeline illustrated by processing of sequencing reads obtained from library L12 and related to the cleavage of Gm^{4}C-RNA (R14) with deoxyribozyme AM320 (D6). The sequence corresponding to the RNA substrate is shown in italic. The binding arms of the deoxyribozyme and the corresponding complementary regions in the RNA substrate are shown in orange and green. The regions of the DNA-RNA hybrid not covered by sequencing are shown in lowercase.

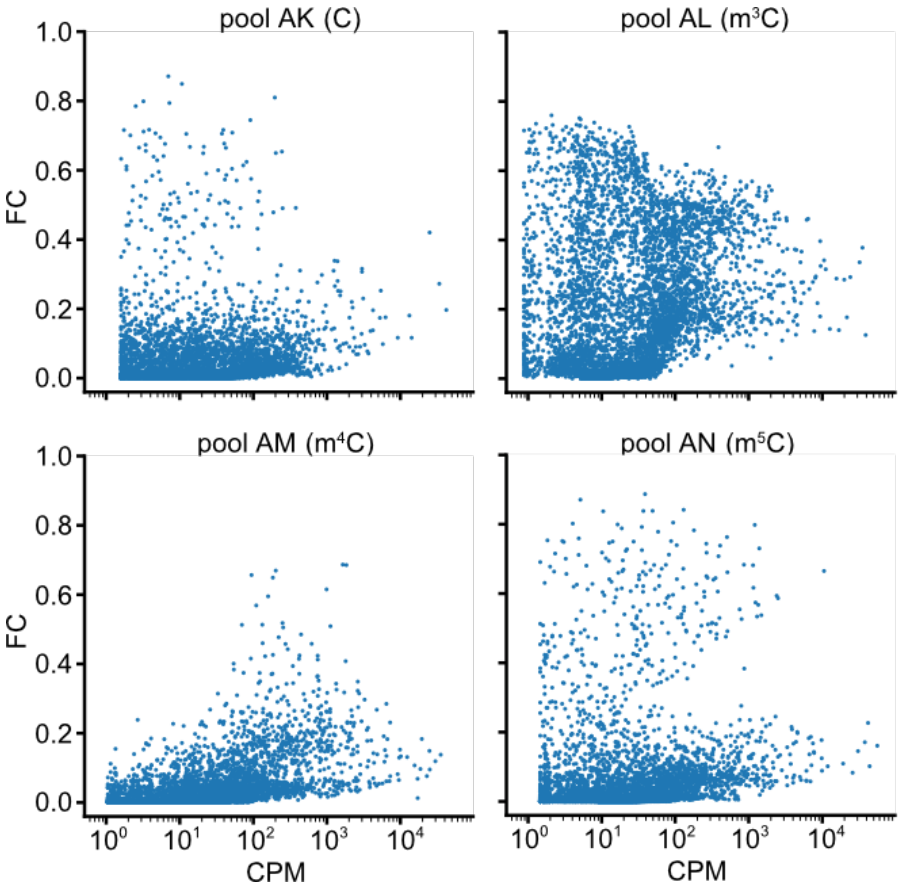


Figure S2. Average FC (FC = [FC_{AC} + FC_{CC} + FC_{GC} + FC_{UC}] / 4) plotted vs abundance for deoxyribozymes in in-vitro selection pools AK, AL, AM and AN which were trained to selectively cleave unmodified (C), m^{3}C-, m^{4}C or m^{5}C-modified RNA, respectively.

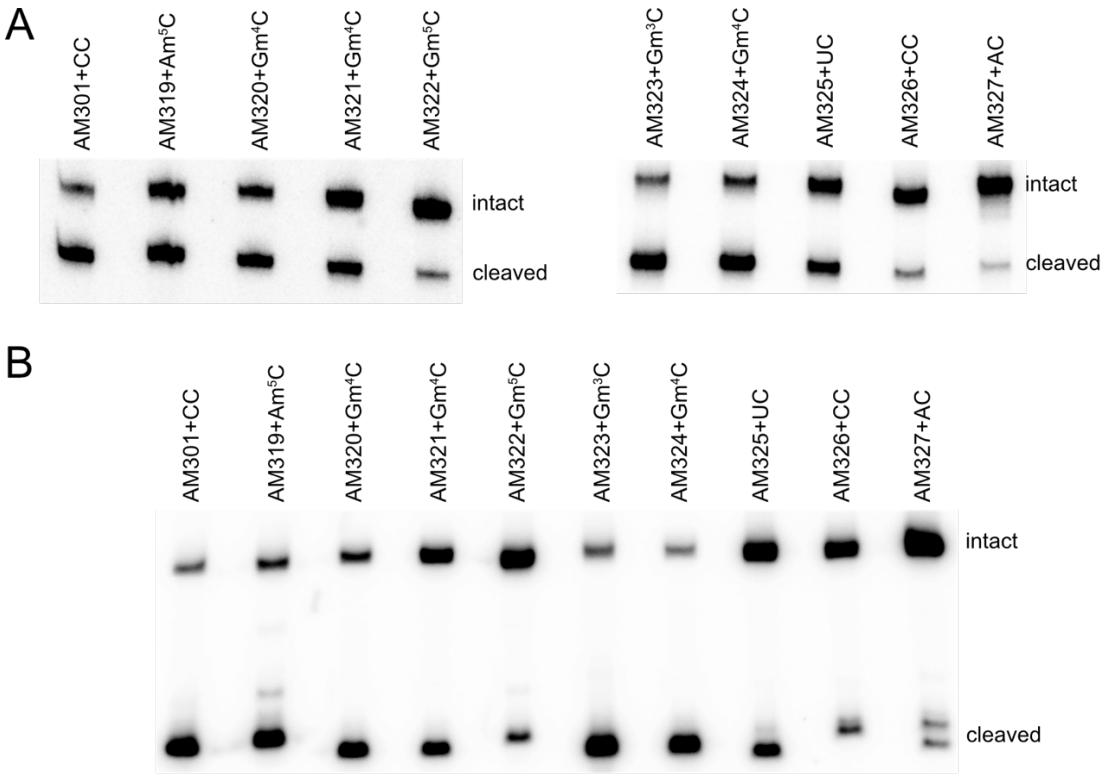


Figure S3. Representative gel-based assays for cis- (A) and trans- (B) endpoint activity (10 mM $MgCl_2$ at 37 °C for 6 h) of deoxyribozymes AM301 and AM319–AM327 with selected RNA substrates. FC values are presented in Table S4.

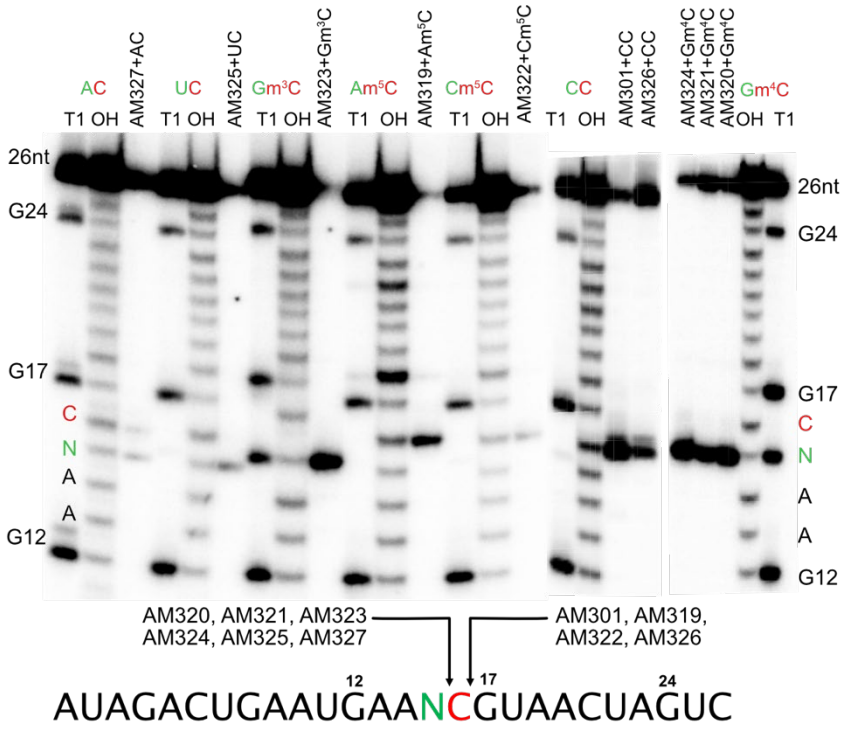


Figure S4. Determination of cleavage site position for deoxyribozymes AM301 and AM319–AM327. 5'-Radioactively labeled RNA was incubated with the DNA enzymes in presence of 10 mM $MgCl_2$ at 37 °C for 6 h. Lanes OH: alkaline hydrolysis of RNA in the presence of 25 mM NaOH at 95 °C for 3 min. Lanes T1: RNase T1 ladder showing positions of Gs in the RNA substrate.

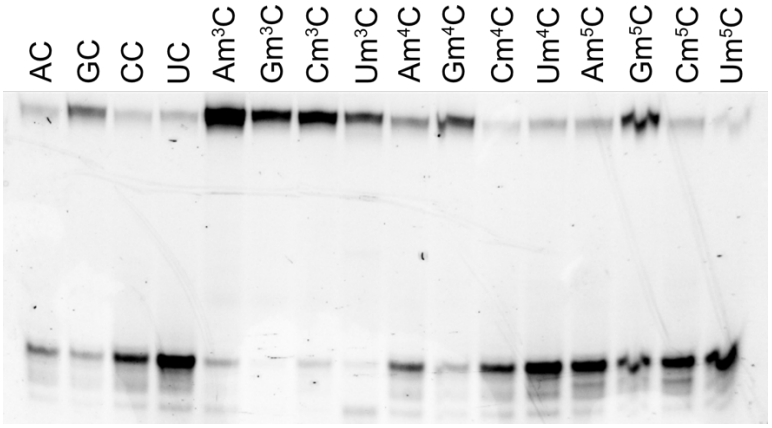


Figure S5. PAGE assay for single-turnover cleavage of RNA substrates R5–R20 with deoxyribozyme AM301. FC values are given in Table S6. Conditions: 100 pmol deoxyribozyme, 10 pmol 3'-fluorescently labeled RNA substrate, 10 mM $MgCl_2$, 50 mM Tris-HCl, 150 mM NaCl, pH 7.5, 37 °C, 6 h.

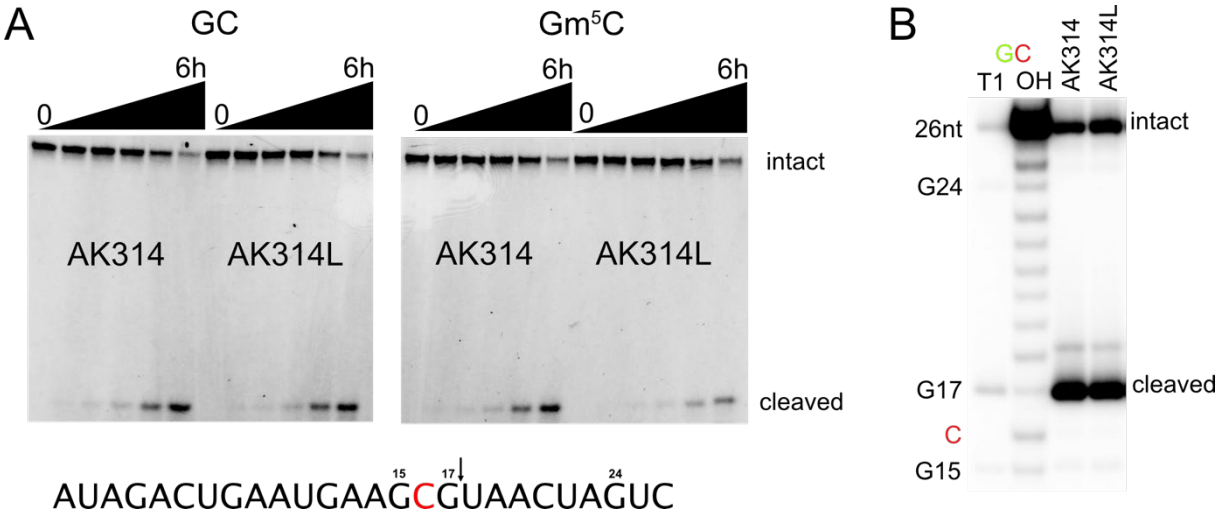


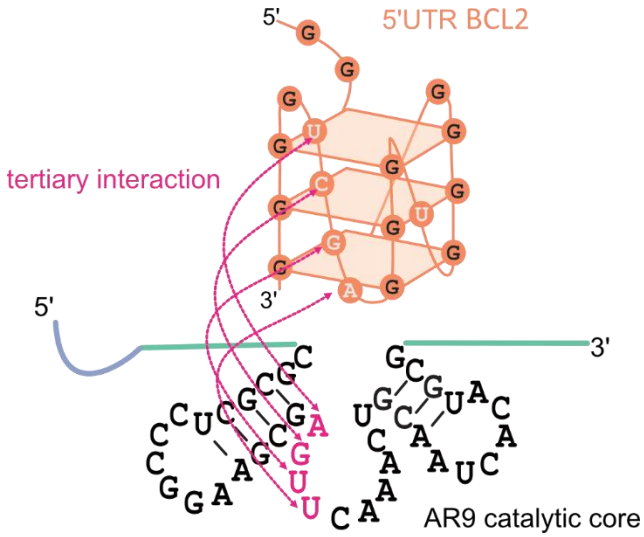
Figure S6. (A) Activity assays for deoxyribozymes AK314 and AK314L and RNA substrates R6 (GC) and R18 (Gm^5C). FC values are presented in Table S5. Conditions: 100 pmol deoxyribozyme, 10 pmol RNA substrate, 10 mM $MgCl_2$, 50 mM Tris-HCl, 150 mM NaCl, pH 7.5, 37 °C, 6 h. (B) Determination of cleavage site position for deoxyribozymes AK314 and AK314L. 5'-Radioactively labeled RNA was incubated with the DNA enzymes in presence of 10 mM $MgCl_2$ at 37 °C for 6 h. Lanes OH: alkaline hydrolysis of RNA in the presence of 25 mM NaOH at 95 °C for 3 min. Lanes T1: RNase T1 ladder showing positions of Gs in the RNA substrate.

References

(1) Liaqat, A.; Sednev, M. V.; Stiller, C.; Höbartner, C., RNA-Cleaving Deoxyribozymes Differentiate Methylated Cytidine Isomers in RNA. *Angew. Chem. Int. Ed.* **2021,** *60* (35), 19058–19062.
(2) Zhou, C.; Avins, J. L.; Klauser, P. C.; Brandsen, B. M.; Lee, Y.; Silverman, S. K., DNA-Catalyzed Amide Hydrolysis. *J. Am. Chem. Soc.* **2016,** *138* (7), 2106–2109.
(3) Chandrasekar, J.; Silverman, S. K., Catalytic DNA with phosphatase activity. *Proc. Natl. Acad. Sci.* **2013,** *110* (14), 5315.
(4) Martin, M., Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet j.* **2011,** *17* (1), 3.
(5) Sednev, M. V.; Mykhailiuk, V.; Choudhury, P.; Halang, J.; Sloan, K. E.; Bohnsack, M. T.; Höbartner, C., N^6-Methyladenosine-Sensitive RNA-Cleaving Deoxyribozymes. *Angew. Chem. Int. Ed.* **2018,** *57* (46), 15117–15121.
(6) Liaqat, A.; Stiller, C.; Michel, M.; Sednev, M. V.; Höbartner, C., N^6-Isopentenyladenosine in RNA Determines the Cleavage Site of Endonuclease Deoxyribozymes. *Angew. Chem. Int. Ed.* **2020,** *59* (42), 18627–18631.

6. Paper-IV In Vitro Evolution of Ribozymes for Recognition and Labeling of RNA G-quadruplex

Anam Liaqat and Claudia Höbartner, 2022, *Manuscriot in preparation*



In Vitro Evolution of Ribozymes for Recognition and Labeling of RNA G-quadruplex

Abstract

Ribozymes are efficient tools for site-specific labeling of RNA. Previously known ribozymes catalyzed the labeling of specific target sites in RNAs based on recognition via Watson-Crick base pairing. Herein, we performed the in vitro selection for RNA labeling ribozymes for structure-based recognition of BCL2 RNA G-quadruplex. The ribozymes evolved from two subsequent selections attached fluorescently labeled adenosine to the RNA of interest and accept a variety of N^6-modified-ATP. AO10.2 ribozyme was identified from the initial selection and was found to label uridine in the binding arm instead of the BCL2 region. However, it could serve as a potential tool for site-specific labeling of RNA. AR9 is a trans-active, self-labeling ribozyme that catalyzes nucleotide transferase reaction in the presence of the BCL2 sequence. Our preliminary findings already indicate that AR9 has the potential to serve as a reporter RNA by self-labeling upon structure-based recognition of RNA G-quadruplexes.

Introduction

The RNA molecules perform diverse functions in living cells by modulating their structure, which directly affects the regulation mechanism or enables them to bind with other RNAs, proteins, or ligands leading to biochemical events that impact cellular metabolism. The RNA structures are dynamic and have motifs as simple as stems, bulges, and hairpin loops, called RNA secondary structures, and as complex as pseudoknots or kink turns, called RNA tertiary structures. The secondary structures of RNA consist of short double-stranded regions separated by a stretch of single-stranded nucleotides; tertiary structures are assembled through the formation of diverse interactions between secondary structure elements [1]. RNA secondary structures are mostly dominated by Watson-Crick base pairing while tertiary structures are formed by weaker interactions such as coaxial stacking and long-range tertiary interactions, including non-canonical base pairs and hydrogen bonds. Unlike proteins, RNA secondary structures are inherently stable

and provide a rigid subunit for tertiary folding facilitated by magnesium, other cations, and proteins. These interacting elements often stabilize the RNA assembly by overcoming the electrostatic barrier, becoming an integral structural component, or remodeling the RNA structure[2]

Tertiary RNA structures are rival in their complexity to that of proteins, and are observed in ribozymes, riboswitches, long non-coding RNAs, and ribonucleoproteins like ribosomal RNA considered as complex RNA machinery to perform diverse functions. RNA being a single stranded molecule has the liberty to remodel its structure during molecular crowding and various ionic conditions in the cytoplasm and act as a tunable device depending on cellular conditions[3]. The role of RNA structure modulation has been exemplified by evolutionarily conserved mRNA stem-loop structures that stall ribosome movement on mRNA leading to the accumulation of truncated polypeptides and no-go mRNA decay[4].

Among RNA structures, G-quadruplexes are non-canonical secondary structures formed by repeated stretches of guanine-rich sequences. Four guanines form a planer quartet (G-quartet), which stack on each other to assemble into a four-stranded structure with the central cavity stabilized by cations (i.e. K+, Na+)[5]. RNA G-quadruplex (rG4s) are more compact and thermodynamically more stable than their DNA counterpart (dG4s) due to i) the 2'-hydroxyl group in ribose sugar increases the intramolecular interactions in rG4s, leading to enhanced stability ii) rG4s topology is constrained to parallel conformation by the 2' hydroxyl group as it limits the orientation of the guanosine to anti-conformation and imparts 3' endo puckering to the structure iii) The 2'-hydroxyl group also affects ligand and metal ion binding. For example, K+ stabilizes both G4s but Na+ has a less stabilizing effect on rG4s than dG4s[6].

The rG4s in mRNA (5’UTR, 3’UTR, or ORF) have been found to play a role in translation regulation (as enhancer or repressor), 3'end processing, alternative splicing, mRNA localization, polyadenylation, and transcription termination. The 5'UTR BCL2 is a prime example of posttranscriptional regulatory pathways involving structural elements. The sequence is highly conserved among several species and reported to form an RNA G-quadruplex in 5' UTR, thereby modulating translation efficiency[7]. Moreover, rG4s has been identified in non-coding RNAs (ncRNAs). For example, i) telomere-associated RNAs (TERRA), which are involved in regulating

telomere length and telomerase activity, and ii) tRNA-derived stress-induced RNAs (tiRNAs), which inhibit translation in response to stress[8]

A strong interest in mapping the distribution of RNA G-quadruplexes in the transcriptome has been sparked by the emerging role of rG4s in the cell. Initially, biophysical assays were used to determine rG4s based on the changes in biophysical parameters such as melting temperature, UV absorbance, intrinsic fluorescence, and migration on polyacrylamide gel[9]. Nevertheless, these assays might not reflect the actual folding of rG4s in the native environment, as they are usually applied to homogenous and short RNA oligonucleotides. Reverse transcriptase (RT) footprinting assays are the techniques that exploit the intrinsic properties of reverse transcriptase in which cDNA synthesis can be blocked by covalently modified RNA or stable RNA structures on the template RNA. The aborted cDNA products are then identified by denaturing polyacrylamide gel electrophoresis. These methods involve the development of chemical reagents which selectively modify unstructured regions of RNA. Well-established techniques in this area include SHAPE (2'-hydroxyl acylation), DMS probing (N7-methylation of guanine), and kethoxal footprinting (N1 and N-NH2 alkylation of guanine). As a result, the structure and unstructured regions of RNA can be identified according to the rate of chemical modification indicated by the frequency of RT stops[10].

Similarly, ligand binding assays depend on chemical ligands that specifically bind to G4s by stacking interactions. The rG4-ligand interaction can be detected by fluorescence measurement induced by ligand and attachment of molecular tags as a fluorophore for imaging or biotin for pull-down [11]. In addition, specific antibodies were developed for detection of RNA G-quadruplex in fixed human cells [12]. With the advent of RNA-seq technology, high throughput methods were developed based on RT-footprinting and ligand-binding assays. Few of these methods such as rG4 seq, SHALiPE-seq, Keth -seq, RT-stop profiling, and G4RP-seq, are well-known rG4 detection approaches[13]. The methods may be ambiguous when the targeted footprint is not unique to rG4s or ligands that bind to rG4s to stabilize the structure that does not occur under physiological conditions.

Furthermore, the intrinsic ability of RNA to form RNA-RNA interactions has been exploited to develop aptamers that identify target hairpin structures using loop-loop interactions[14]. An RNA

aptamer was evolved by in vitro selection that forms a kissing-loop complex with HIV-1 TAR element and has shown to be a more specific binder than the antisense oligonucleotide[15]. These studies take advantage of peculiarities (loops) for choosing the target sequence and maintaining the functional folding of the target RNA. Aptaswitches were then designed based on these facts to detect specific ligands and precursors of microRNAs[16]. Recently, L-RNA aptamers have been developed to target TERRA-rG4s and provide a new approach in the selective targeting of rG4s.[17]

Our work was inspired by these examples of tertiary interactions, and we aimed to develop ribozymes that could recognize RNA-quadruplexes through tertiary interactions, followed by fluorescent labeling using small cofactors. Ribozymes are catalytic RNA molecules that are limited to performing peptidyl transferase, RNA-cleavage, and ligation reactions in nature. However, the advent of the in vitro selection methods facilitated the discovery of artificially evolved ribozymes with a broad range of activities. RNA catalysts recently added to the repository include nucleotidyltransferase and methyltransferase that catalyze the attachment of chemical moieties at the internal position of the target RNA[18]. Among these, trans-acting ribozymes provide efficient tools for sequence-based recognition followed by site-specific labeling of the target RNA. The aim of this project is to develop RNA labeling ribozyme for structure-based recognition of RNA quadruplex. The 5'UTR BCL2 was selected as a target structure as it is known to regulates the expression of proteins involved in apoptosis[19].

Two new structured RNA pools were designed to target the BCL2 RNA G-quadruplex for covalent labeling using an in vitro selection strategy previously established in our lab for the development of RNA-labeling ribozymes[18a]. We envisioned that the bulged-out rG4s directly facing the random region of the library would be recognized by the potentially active candidates through tertiary interactions. In the first selection experiment, we identified a ribozyme that catalyzes a new RNA labeling reaction, independent of an intact quadruplex structure. The second approach yielded a quadruplex-dependent ribozyme, which upon further characterization and optimization may become a new tool for quadruplex research.

Results and Discussion

Initial library design and assembly

The pool design is depicted in **Figure 1** in analogy to a previously used strategy[18a], the selection pool was designed to have an N_{40} random region (black) flanked with constant primer binding arms (green) complementary to the binding arms in substrate sequence (pink) where BCL2 (orange) was bulged out as the target sequence and connected through a loop region (grey) (**Figure 1a**). The substrate sequence was designed in a way that a known RNA G-quadruplex (BCL2) sequence was flanked with binding arms that later form base pairing with primer binding sites. As shown schematically, this bulging-out region of BCL2 was directly facing the random region and intended modification site. The 14-nucleotide connecting loop was introduced to ensure the physical connection between active ribozyme species and modified substrate sequence.

The DNA template for the selection library was constructed using two synthetic DNA fragments (D1, D2) to avoid the difficulty of synthesizing a 146nt long DNA template (including T7 promoter) using the standard solid-phase synthesis method. The primer D1 included T7 promoter, BCL2, binding arms, the connecting loop sequences, and a 5’ phosphate group. The primer D2 covered primer binding sites, N_{40} random region, and connecting loop. Two primers were complementary at their 3'-end in the connecting loop segment and used to synthesize a full-length dsDNA template using Klenow fragment (-exo). It turned out that the in vitro transcription of the full-length pool RNA including the quadruplex region was extremely inefficient when standard transcription conditions were employed. Through careful tests and optimization, it was found that transcription of a quadruplex-containing sequence is more efficient when a single-stranded DNA template is used, and only the T7 promoter region is double-stranded. For this reason, the sense-strand comprising the G-rich sequence was removed by digestion with lamda-exonucleases, which specifically degrades the 5’-phosphorylated strand. The purified ssDNA was then annealed to a short promoter oligonucleotide and transcribed by T7 RNA polymerase to give ~2.5 nmol of the full-length RNA pool (~3 x 10^{14} unique variants with an average copy number of 6) (**Figure S1**). A fraction of the pool (~300pmol) was fluorescently labeled with Lucifer yellow at the 3'end, which was later used to determine the activity of the enriched pool.

Selection process

Biotinylated cofactors (N^6-(6-Aminohexyl)-ATP-Biotin and Biotinylated O6-benzylguanine) were used as the selection substrates in AO and AP selections, respectively (**Figure S2**). The schematic representation of the selection process is shown in (**Figure 1b**). The RNA pool was incubated in the selection conditions along with the respective biotinylated selection substrate. The selection buffer contained monovalent ions (120 mM K^+, 5 mM Na^+) to stabilize the RNA secondary and tertiary interactions particularly important in the stabilization of RNA G-quadruplex (BCL2). In addition, divalent ions have been found to be directly involved in catalysis along with providing structural stability [20]. Therefore, in AO and AP selections, Mg^{+2} (40mM) was used as a divalent metal ion.

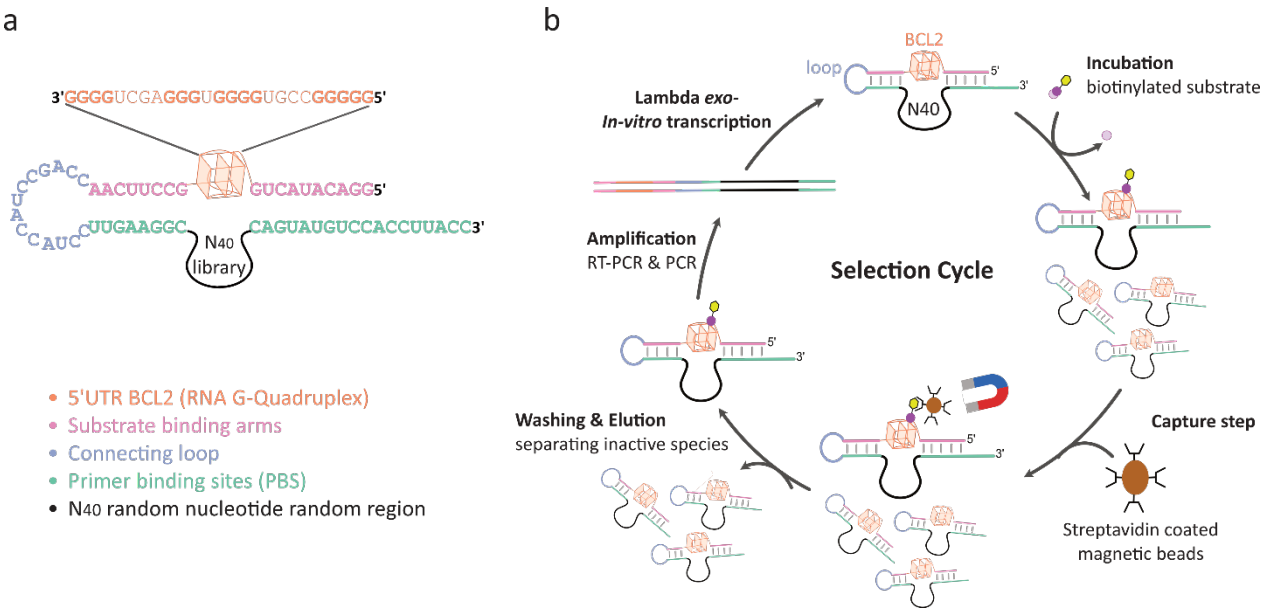


Figure 1 Schematic presentation of pool design and in vitro selection procedure. **a)** Schematic representation of the selection pool design. **b)** Schematic in vitro selection cycle. During the incubation step the active species react with the biotinylated substrate. The "self-biotinylated" species are separated by streptavidin/neutravidin magnetic beads during the capture step. The inactive species are washed out and the active species are separately eluted from the beads. The eluted RNA is reverse transcribed and amplified through PCR in the amplification step. The 5'phosphorylated sense strand is removed using lambda exonuclease and single-stranded DNA is then transcribed to generate the enriched pool for the next selection round.

The initial incubation rounds were performed at 37°C for overnight to provide enough time for all the active species in the pool to be modified. During the reaction, active species got self-biotinylated and were separated from the rest of the pool by a bead-based capture step. Streptavidin/neutravidin-coated beads were used alternatively to avoid non-specific enrichment of streptavidin/neutravidin binders during the selection. Before incubation with pool species, the beads were also treated with *E.coli* tRNA to block all the potential sites for nonspecific interactions. After washing off excess tRNA, beads were incubated with a pool mixture (a fraction of the selection sample was stored for later to determine pool activity). The samples were then extensively washed using the denaturing wash buffer containing urea as a denaturing agent. After the beads were washed in the denaturing wash buffer, streptavidin/neutravidin binders were removed from the pool and biotinylated molecules remained. Active species were eluted under harsh conditions (incubation with elution buffer containing 95% formamide followed by heating at 95°C for 8-10 minutes) to disrupt the strong interaction between biotin and streptavidin/neutravidin. Eluted RNA was recovered through ethanol precipitation. The percentage of active species in the pool was determined by comparing the fluorescence intensity of the eluted sample to the total intensity of the selection sample (determined from the sample stored before the capture step). After fluorescent measurement, the active fraction was amplified with two subsequent PCR reactions. The first reaction, called the RT-PCR reaction, was perfomed to reverse transcribe the recovered RNA to cDNA, and then it was amplified by 5-10 PCR cycles. The number of PCR cycles were decreased as selection progressed to avoid amplification biases. The forward primer (D4) used in RT-PCR was bound to the loop region and ensured to amplify the pool excluding the proposed target sequence i.e., BCL2 and 5' binding arm. A fraction (~5-10µl) of the RT-PCR product was used as a template for the second PCR reaction. The forward pool primer (D1) was used to introduce substrate sequence including the T7 promotor at its 5' end. The schematic presentation of primer binding and two-step amplification is illustrated in **Figure S3**.

A total of 15 rounds were conducted for AO selection and 8 rounds for AP selection. For AO selection, enrichment was observed at round 7 with 0.8% of the RNA eluted from the capture step. In rounds 8 and 9, the eluted RNA amount decreased to ~0.4%, although reaction conditions

and incubation time remained constant. The maximum amount of RNA was eluted at round 10 and amounted to 2.5%. In the following rounds, selection stringency was increased by reducing the concentration of N^6-ATP-Biotin and $MgCl_2$. The selection pressure dropped the activity to 0.5% in round 11, and it increased to 0.9% in round 12 under similar conditions. To encourage the survival of fast ribozymes, the incubation time was reduced in round 13 from overnight to 3h. This led to a dramatic reduction of the eluted fraction but increased again in the subsequent round. In the last round, the incubation time was further reduced to 1 hour which slightly decrease the amount of eluted RNA. In the case of AP selection, enrichment was not observed until round 8, and it was decided not to continue further. The summary of the selection progress and selection conditions is shown in **Figure S4**.

As the fluorescence measurement after each selection round gave us an approximate value of the activity, an active fraction from different selection rounds was subjected to streptavidin gel-shift for estimation of the true cis-activity of the enriched pool. There were four rounds selected as round 6 presumably had the most diverse active species, round 10 was the round with the highest percentage of eluted fraction, and the round 14, and 15 were supposed to have the fastest and most active species. RNAs were incubated overnight with the biotinylated substrate under selection conditions at 37°C. Afterward, samples were incubated with a large excess of streptavidin so that biotinylated species can bind to the protein and show a gel shift when resolved on 10% native PAGE as shown in **Figure S5**. Gel analyses revealed that only a small percentage of the pool species were self-biotinylated, and they appeared similar in all selection rounds. Nevertheless, round 6 and 10 were selected for cloning as they were more likely to contain a more diverse set of active ribozymes compared to later rounds 14 and 15. This assumption was due to the fact that as selection progressed, the pool is expected to converge to similar families of ribozymes and the most efficiently amplified sequences may dominate the active species. Therefore, the most abundant species might not necessarily be the most active. The pool was cloned using the TOPO-TA system and transformed *E.coli* were selected via blue-white screening. Five plasmids from each selection round were isolated and submitted for sanger sequencing.

Analysis of sequencing results revealed two families of ribozymes named AO10.2 and AO6.16. (**Table S3**) Representative members were tested using a streptavidin gel shift assay. The individual clones were PCR amplified from the plasmid using pool forward primer (D1) and reverse primer (D3) to obtain the dsDNA template. The PCR product was purified, and the sense strand was removed using lambda exonuclease. ssDNA template was in vitro transcribed and full-length ribozyme covalently attached with substrate sequence was subjected to a self-biotinylation reaction. The reaction mixture included 40 mM $MgCl_2$ and 200 μM N^6-ATP-Biotin. After overnight incubation, activity was determined using a streptavidin gel shift assay. The results indicated that the AO 6.16 family was inactive while AO10.2 and AO6.2 were self-biotinylated (**Figure S6**). AO10.2 and AO6.4 had identical catalytic cores and differed only in the binding arm by a single nucleotide substitution. Only the left binding arm of the constant region was not included in any primer and thus had chances to get mutated during the PCR reactions. The AO10.2 family contains six more members that are single mutants of the parent ribozyme.

Probing the modification site

Secondary structures of the ribozymes were predicted using the Vienna RNAfold server[21]. AO10.2 and AO6.4 showed similar predicted secondary structures (**Figure 2a**). Based on predicted secondary structure, it is important to probe the modification site before testing trans-activity, in which the connecting loop would be removed, which may affect secondary interactions and the activity of enzymes.

DNAzyme assisted cleavage reaction

To confirm the modification site to be located in the substrate sequence, 8-17N|G deoxyribozyme was used. The full-length RNA including the ribozyme and the quadruplex (AO10.2-BCL2) was fluorescently labeled at 3'-end using fluorescein. N^6-(6-Aminohexyl)-ATP-ATTO550 was used as fluorophore-labeled ribozyme substrate. The modified RNA was site-specifically cleaved with 8-17 enzyme (D9) in the connecting loop between A and G as shown in the **Figure 2b-c**. After the cleavage reaction, RNA fragments were resolved on denaturing PAGE. Depending on the location of the modification site, we either observe a single band with dual labels (**Figure 2b**) or two bands with green and red fluorescence (**Figure 2c**). The gel will display a single band with dual labels

(green and red) if the modification is inserted in the connecting loop, primer binding site, or even the catalytic core. In contrast to this, if the BCL2 sequence including binding arms is modified with N^6-ATP-ATTO550, two bands will appear on the gel. The green-fluorescent band contains the catalytic core, primer binding sites, and part of the connecting loop while a red band contains BCL2 with binding arms. The analyses of the gel image demonstrated that AO10.2 was labeling the BCL2 sequence (including binding arms) which appeared as an additional red band on the PAGE as shown in **Figure 2d**.

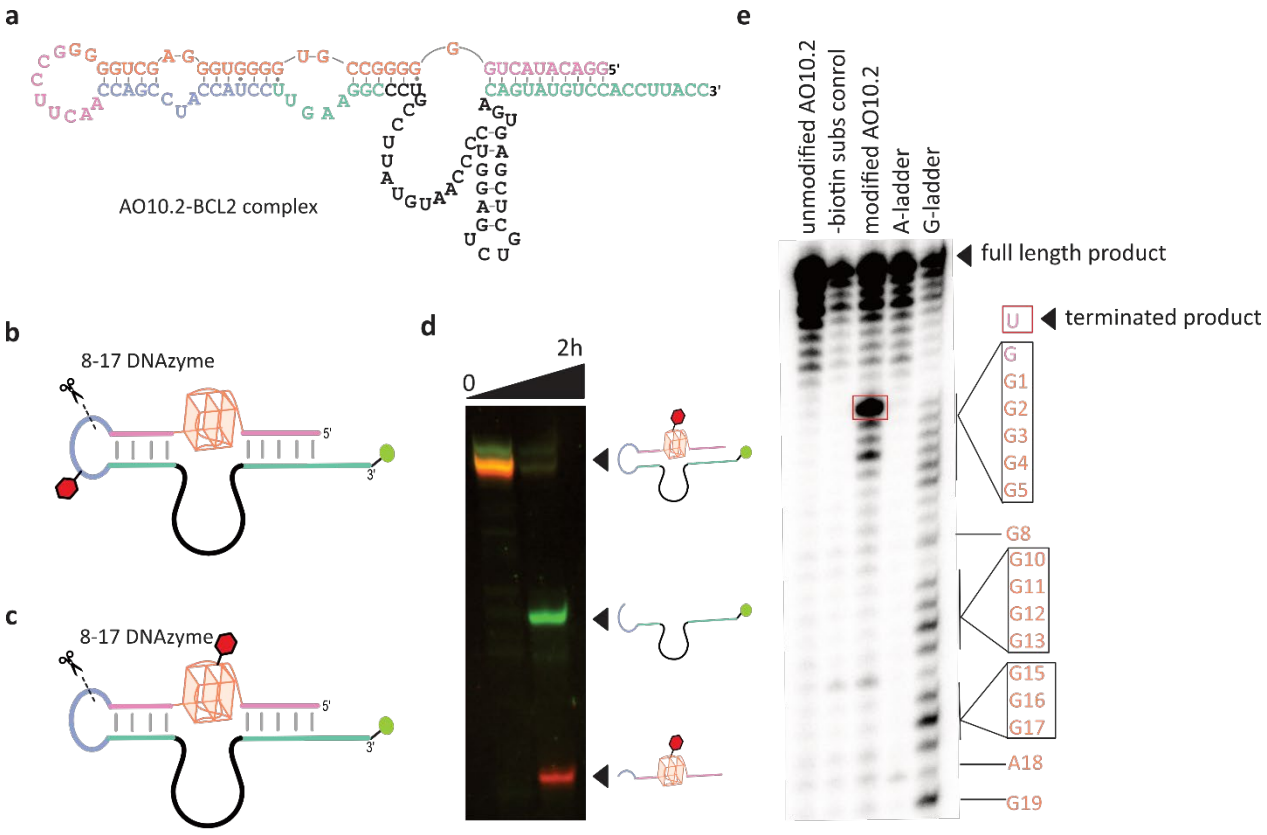


Figure 2 Probing of the modification site of AO10.2 ribozyme. a) Predicted secondary structure of AO10.2-BCL2 RNA b) Schematic presentation of AO10.2-BCL2 RNA with possible modification site outside the substrate region or c) within the substrate region. The attached fluorescent adenosyl-moiety is represented in red, the cleavage site of the 8-17 DNAzyme is indicated with scissor cutting in the loop region, and 3'-fluorescein-end labeling is shown in green. d) DNAzyme mediated cleavage products were resolved on analytical denaturing PAGE. Time points were taken at 0 and 2 hours from the start of the reaction. A clear red and green band were visible after 2 h of reaction. e) RT primer extension stop assay confirms the uridine as a modification site in the right binding arm of the AO10.2-BCl2 RNA

RT primer extension stop

The exact modified position was determined by reverse transcriptase-based primer extension assay. The modification in the template strand blocks the reverse transcriptase and results in the termination of the cDNA synthesis. The aborted product can be identified using 5’ end-labeled primers followed by resolution of the product on denaturing analytical PAGE. A 5'- P^{32} labeled primer (D8) was used complementary to the connecting loop region. The primer was annealed to the modified RNA and subjected to primer extension using superscript-III reverse transcriptase. Unmodified RNA and a sample that was incubated without N^{6-}ATP-Biotin were used as controls. Full-length cDNA products were observed in the case of controls while modified RNA lane showed abortive bands. To read the exact termination position, sequencing ladders were generated based on dideoxynucleotide chain termination. As the BCL2 sequence is rich in G-sequences, we included G and A ladders using ddCTP and ddTTP respectively. The most pronounced abortive band appeared one nucleotide 3' to the modification site, which indicates the presence of a steric block, at the modified positions. Since the most pronounced abortive band appeared at G in the binding arm, the next U nucleotide is most likely the modified position (**Figure 2e and S7**).

Activity analysis

In-trans activity assay of AO10.2

To determine if the AO10.2 ribozyme is also active in trans, i.e. in an intermolecular setup, in which the target quadruplex RNA substrate is not covalently attached to the ribozyme, the full-length RNA was cleaved in the connecting loop region, and at two different positions within the BCL2 substrate using 8-17 and 10-23 DNAzymes (D9-11) (**Figure 3a and S8**). The ribozyme-BCL2 RNA was fluorescently labeled at the 3’ end, followed by cleavage with the respective DNAzyme. Afterwards, DNase-I was used to remove the deoxyribozyme. The two resulting RNA fragments were annealed (95°C for 3 min followed by cooling at RT) and then subjected to nucleotidyltransferase reaction using N^{6-}ATP-ATTO550 as a cofactor. Samples were resolved on analytical denaturing PAGE and multichannel imaging showed that indeed the ribozyme modified the target uridine in the in-trans setup as well. Surprisingly, this was the case for all three tested positions.

G-quadruplex (BCL2) dependent activity analysis

After it was demonstrated that AO10.2 ribozyme is trans-active and modified near the target structure, the recognition of BCL2 sequence by ribozyme was investigated further. For this purpose, a set of target sequences with desired mutations were designed to disrupt the G-quadruplex (**Figure S9a**). In vitro transcription was performed using mutated forward primers (D12-15) to generate the AO10.2-BCL2 RNAs with desired mutations. Nucleotidyltransferase reaction was performed using N^6-ATP-Biotin as a cofactor. Streptavidin gel shift assay was used to determine the activity of the mutated AO10.2-BCL2 RNAs, followed by RT primer extension stop experiments to confirm the modification site. According to the gel shift assay, the mutation of G to A next to the modification site had a detrimental effect on ribozyme activity. Also, G1A, G2A, and G3A mutations at 5'-end of BCL2 negatively impacted the enzyme activity, supporting the hypothesis that the ribozyme recognizes the G-quadruplex sequence and should not function when the structure is disrupted. However, mutations G23A, G24A, G25A at 3'-end of BCL2 and the variant with a truncated BCL2 (removed 15-nt at 3'-end of BCL2) still showed activity based on the results of the the gel shift assay. Surprisingly, clear shifts in bands were observed in both mutated and truncated sequences, which means that the AO10.2 riboyzme retains activity even if the absence of an intact quadruplex, which contradicts the hypothesis (**Figure S9b**).

Possibly, the ribozyme activity is only adversely affected by mutations close to the modification site due to proximity rather than because it recognizes the sequence of BCL2. Therefore, when the sequence was mutated at the 3' end or even deleted 15 nt from 3'-end, these chnages had no influence on enzyme activity. RT-stop primer extension further confirmed the results, showing the abortive bands exactly one nucleotide 3' to the modification site, i.e. U in the binding arm (**Figure S9c**).

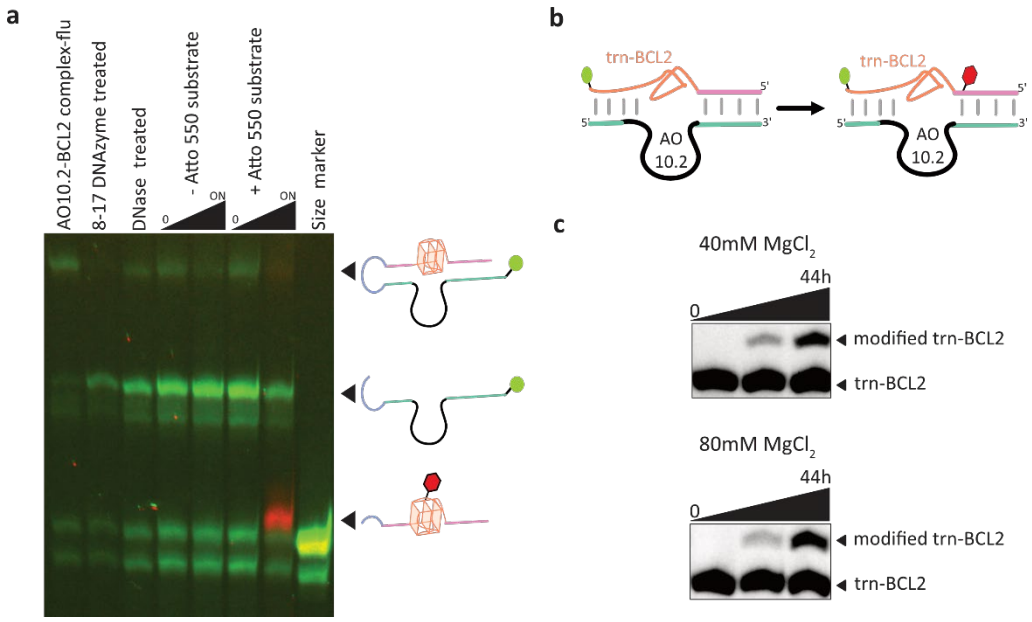


Figure 3 In-trans activity analysis of AO10.2 ribozyme. **a)** 3-flu-end labeled AO10.2-BCL2 cleaved with 8-17 DNAZyme in the loop region followed by DNAase treatment to remove DNAZyme. Two RNA fragments were incubated under nucleotidyltransferase reaction conditions using N^6-ATP-ATTO550 overnight at 37°C. Samples were resolved on 10% denaturing PAGE and labeled RNA fragments appeared as a red band. **b)** Schematic presentation of trans-activity which means ribozyme is not covalently attached to RNA substrate**. c)** Gel image of the single-turnover kinetic assay performed on trnBCL2. Time points were taken at 0, 24h, and 44h. Aliquots of samples analytical denaturing PAGE.

In order to confirm the activity in the absence of an intact quadruplex, and to investigate the trans activity further, a truncated BCL2(trn BCL2) was synthesized by solid-phase synthesis and fluorescently labeled at 3'-end. A DNA template for AO10.2 was ordered in which the loop region was removed, and this was used for generating a truncated AO10.2 ribozyme by in vitro transcription. A single turnover kinetic experiment was performed, in which the ratio between substrate and ribozyme was 1:10. The reaction conditions were identical to the selection process except for the $MgCl_2$ concentration which was increased from 40 mM to 80 mM. The samples were resolved on gel, which showed that the yield was quite low even with a prolonged incubation time (44h) and a higher concentration of $MgCl_2$. Nevertheless, this analysis confirmed that AO10.2 was able to modify the trn-BCL2 substrate and did not require intact G-quadruplex for its active conformation.

Optimized selection design:

As outlined above, the first design of the quadruplex-containing RNA library resulted in the identification of ribozymes that do not depend on an intact quadruplex structure, but rather labeled a nucleotide in the binding arm. We recognized that the fixed binding arms in the previous design, which were intended to hold bulged-out rG4s directly opposite to the random region, actually forced the random region to open the BCL2 structure. We proposed that such a probability could be ruled out by directly attaching the rG4 to the library through the small connecting loop. Therefore, the redesign of the starting RNA library resulted in a free-hanging BCL2 region, more relaxed to fold into the quadruplex (**Figure-4a**). Additionally, we used pyridostatin (PDS) as a ligand to stabilize the conformation of RNA G-quadruplex during selection[22]. The selection of a new library was performed analogous to that of previous SELEX. Two selections were conducted: a) with rG4s binding ligand (AR-selection) b) without rG4s binding ligand (AQ-selection). N^6-ATP-Biotin was used as a cofactor in both selections.

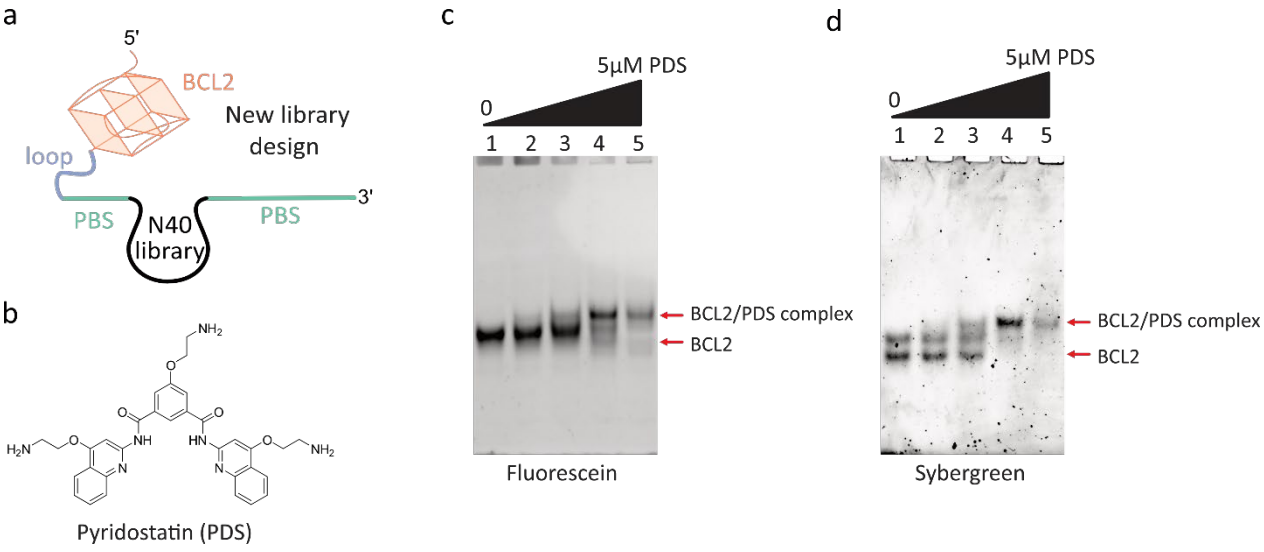


Figure 4 Schematic presentation of optimized library design a) A new library was designed by removing fixed binding arms and directly attaching rG4s to the random library through connecting loop. b) Molecular structure of pyridostatin (PDS) used as a ligand to bind RNA G-quadruplex. c) The ligand (PDS) binding assay with 3-flu-end labeled 5'UTR BCL2 RNA G-quadruplex and d) with unlabeled 5'UTR BCL2 RNA G-quadruplex (stained with Sybergreen). Lane-1 is 1µM RNA and no PDS, lane-2 is 1µM RNA and 0.5µM PDS, lane-3 is 1µM RNA and 1µM PDS, lane-4 is 1µM RNA and 2µM PDS and lane-5 is 1µM RNA and 2.5µM PDS.

The first sign of enrichment appeared in round 6 in the case of AQ selection (selection without PDS). Activity levels continued to increase from 0.2% to 3.5% over the next four rounds (**Figure 5a**). No enrichment was observed in AR selection (selection with PDS) until round 8. We speculated that PDS might directly interfere with the N^6-ATP-Biotin or change the reaction environment that led to the degradation of the cofactor into N^6-AMP-Biotin. In order to test this, we incubated PDS with ATP under similar conditions to those used during the selection reaction (i.e.,1x selection buffer, 40mM $MgCl_2$ overnight at 37°C). The anion exchange HPLC analysis showed that ATP was not degraded to AMP during selection conditions (including incubation with PDS) (**Figure S10**). The other explanation for stalling the enrichment is that the PDS may bind to the rG4s too tightly, making it less likely for rG4s to interact with the random library. To rule out this possibility, we continued the next round of AR selection without PDS. Interestingly, 0.3% of the eluted fraction was recovered from streptavidin beads and the amount was increased to 2% in the next round (**Figure 5a**).

Streptavidin gel-shift assay was used to determine the true cis-activity of the enriched pools from round 10. RNAs were incubated overnight with the biotinylated substrate under selection conditions at 37°C followed by incubation with a large excess of streptavidin protein. The gel shift analyses indicated that a very low percentage of the pool species were self-biotinylated in both selections (**Figure 5b**). However, we continued to investigate the scope of NTP substrates, and pools were tested with ATP analogs in which biotin was replaced by fluorophores. For this purpose, N^6-ATP-ATTO647N and N^6-ATP-Cy3 were used as modified substrates. Interestingly, both substrates were successfully accepted by AQ and AR pools. This initial assay already suggested that the potential ribozymes evolved from these selections could be used as a tool for rG4s labeling using fluorescent dyes (**Figure 5c-d**)

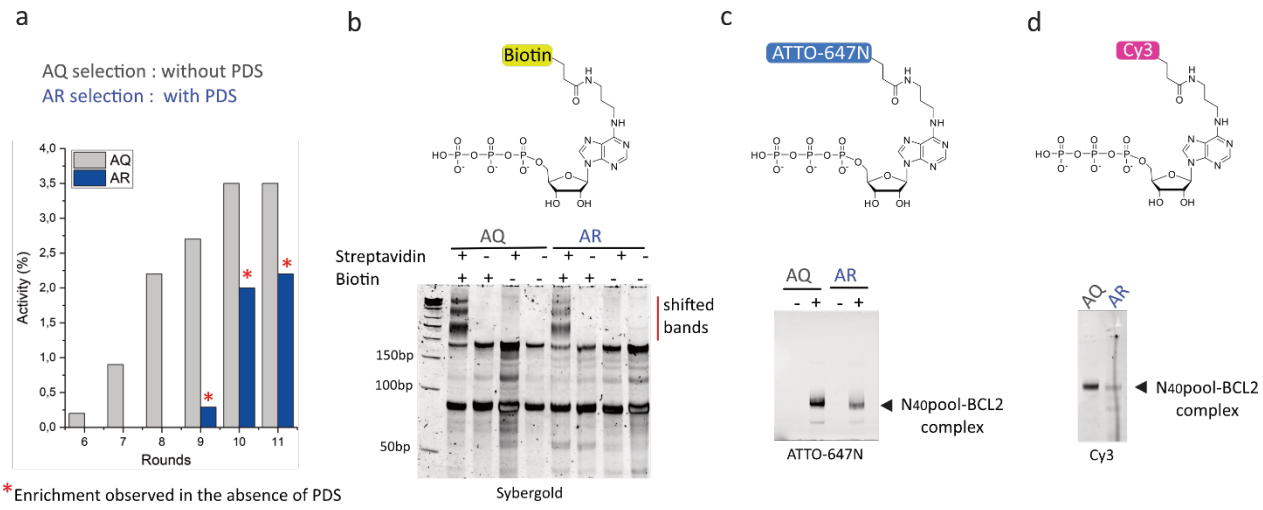


Figure 5 Selection progress and cis activity of the enriched pools with ATP analogs. **a)** The enrichment level at each selection round was determined by fluorescence measurement of the eluted RNA. In case of AR selection enrichment was observed in the absence of PDS. **b)** Streptavidin gel-shift assay of the round 10 pools. **c)** Labeling of pool using N^6-ATP-ATTO647N **d)** Labeling of pool using N^6-ATP-Cy3. Reaction conditions: ~10 µM of RNA Pool, 200 µM ATP analogs and 40 mM $MgCl_2$ incubated overnight at 37°C. A fraction of samples was resolved on 10% PAGE.

To probe the modification site, the RNA-BCL2 RNAs were modified using N^6-aminohexyl-ATTO647N under similar reaction conditions. Afterward, modified RNAs were site-specifically cleaved with 8-17NG deoxyribozyme in the connecting loop region. The cleaved products were resolved on 10% denaturing PAGE. The analyses of the gel image showed that active sequences in the pool labeled themselves rather than rG4s (**Figure S11**). It raised the question of whether the active pool was even able to recognize rG4s. To address this question, we performed in-trans (the BCL2 sequence is not covalently attached to the RNA pool) activity analyses. The RNA pool was annealed with the BCL2 substrate and then subjected to nucleotidyltransferase reaction using N^6-ATP-ATTO647N as a cofactor. In addition, a negative control reaction was performed without the BCL2 substrate. The samples were resolved on denaturing PAGE, and the analysis revealed that the active species in the pool labeled themselves only in the presence of rG4s. These results demonstrated that active ribozyme sequences identified BCL2 structure through tertiary interaction and reported its presence by self-labeling reaction (**Figure 6a**).

After these encouraging results from the activity analysis of the pool, we submitted 6 clones from AQ selection and 14 clones from AR selection for Sanger sequencing. Analysis of sequencing

results uncovered that both selections were converged to a single sequence (named AR9) after 10 rounds of selection (**Table S4**). Therefore, it was expected that AR9 showed similar activity as the selection pool. However, the actual yield and exact labeling site within the ribozyme core remain to be determined.

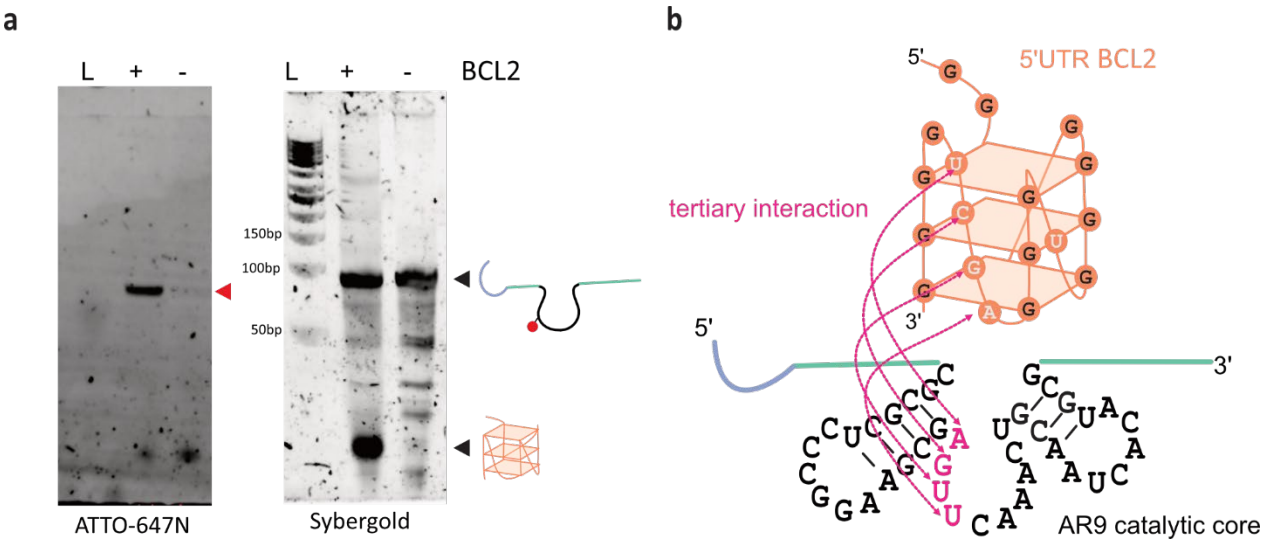


Figure 6 Trans-activity of the enriched AQ-pool and predicted secondary structure of AR9 **a)** Labeling of the pool using N^6-ATP-ATTO647N in the presence or absence of BCL2 visualized in ATTO647N channel followed by staining in sybergold. Reaction conditions: ~10 µM of RNA Pool, 200 µM N^6-ATP-ATTO647N, and 40 mM $MgCl_2$ incubated overnight at 37°C. A fraction of samples was resolved on 10% denaturing PAGE. **b)** Predicted secondary structure of AR9. The proposed tertiary interaction between AR9 and BCL2 RNA is highlighted in pink.

Interestingly, predicted AR9 secondary structure[21] showed structural similarities with FH14 ribozyme, which may indicate similar binding pockets as they both use N^6-ATP-Biotin as cofactors (**Figure S12**). Additionally, the 4-nt thread from the catalytic core of AR9 appears to form a tertiary interaction with the loop region of BCL2 that might help the ribozyme to recognize the BCL2 quadruplex. These are the preliminary observations and further experiments are needed to reveal the detailed interactions between AR9 and BCL2. Nevertheless, AR9 is the first quadruplex-dependent self-labeling ribozyme reported to date.

Conclusion

In summary, we identified AO10.2 and AR9 ribozymes from two subsequent selections. In our initial selection design, we observed enrichment of sequences that label uridine in the binding arm instead of the target rG4s region. The fixed binding arm presumably forced the random region to open the BCL2 structure, which is why active molecules were unable to recognize rG4s and catalyze labeling reactions even with a truncated BCL2 sequence. Even though AO10.2 was unable to recognize rG4s, it still represents the first ribozyme that has been identified to modify uridine. As it is trans-active, it could serve as a site-specific labeling tool for target RNA.

Subsequently, we modified the design of the RNA pool by removing the binding arms and attaching the rG4 directly to the library via a small connecting loop. We proposed that the free-hanging BCL2 region is more relaxed to fold in a quadruplex conformation. Moreover, we used rG4s binding ligand (pyridostatin) to ensure that the quadruplex retained its functional folding during the selection process. We identified a promising candidate AR9 after 10 rounds of selection. This AR9 is a trans-active, self-labeling ribozyme that catalyzes nucleotide transferase reaction in the presence of the BCL2 sequence. We hypothesized that AR9-BCL2 tertiary interactions could promote the folding of ribozymes into active conformation, resulting in the self-labeling of AR9 that reports the detection of RNA G-quadruplex. The AR9 ribozyme is slow when compared to the nucleotidyltransferase ribozyme that has already been reported[18a]. The exact modification site remains to be investigated, but our preliminary findings already indicate that AR9 has the potential to serve as a reporter RNA by self-labeling in the presence of an RNA G-quadruplex. In the future, it will be interesting to investigate how AR9 recognizes the rG4 and how this interaction modulates the ribozyme structure for self-labeling.

References

[1] P. Klaff, D. Riesner, G. Steger, *Post-Transcriptional Control of Gene Expression in Plants* **1996**, 89-106.

[2] aY. Xin, C. Laing, N. B. Leontis, T. Schlick, *Rna* **2008**, *14*, 2465-2477; bS. E. Butcher, A. M. Pyle, *Accounts of chemical research* **2011**, *44*, 1302-1311.

[3] X. Ji, H. Sun, H. Zhou, J. Xiang, Y. Tang, C. Zhao, *nucleic acid therapeutics* **2011**, *21*, 185-200.

[4] C. Bao, S. Loerch, C. Ling, A. A. Korostelev, N. Grigorieff, D. N. Ermolenko, *Elife* **2020**, *9*, e55799.

[5] D. Bhattacharyya, G. Mirihana Arachchilage, S. Basu, *Frontiers in chemistry* **2016**, *4*, 38.

[6] A. Joachimi, A. Benz, J. S. Hartig, *Bioorganic & medicinal chemistry* **2009**, *17*, 6811-6815.

[7] P. Kharel, S. Balaratnam, N. Beals, S. Basu, *Wiley Interdisciplinary Reviews: RNA* **2020**, *11*, e1568.

[8] aY. Xu, Y. Suzuki, K. Ito, M. Komiyama, *Proceedings of the National Academy of Sciences* **2010**, *107*, 14579-14584; bC. Wang, L. Zhao, S. Lu, *International journal of biological sciences* **2015**, *11*, 316; cP. Ivanov, E. O'Day, M. M. Emara, G. Wagner, J. Lieberman, P. Anderson, *Proceedings of the National Academy of Sciences* **2014**, *111*, 18201-18206.

[9] aJ.-L. Mergny, A.-T. Phan, L. Lacroix, *FEBS letters* **1998**, *435*, 74-78; bJ.-L. Mergny, L. Lacroix, M.-P. Teulade-Fichou, C. Hounsou, L. Guittat, M. Hoarau, P. B. Arimondo, J.-P. Vigneron, J.-M. Lehn, J.-F. Riou, *Proceedings of the National Academy of Sciences* **2001**, *98*, 3062-3067; cA. Y. Zhang, A. Bugaut, S. Balasubramanian, *Biochemistry* **2011**, *50*, 7251-7258.

[10] aE. J. Merino, K. A. Wilkinson, J. L. Coughlan, K. M. Weeks, *Journal of the American Chemical Society* **2005**, *127*, 4223-4231; bP. Tijerina, S. Mohr, R. Russell, *Nature protocols* **2007**, *2*, 2608-2623; cW. A. Ziehler, D. R. Engelke, *Current protocols in nucleic acid chemistry* **2000**, 6.1. 1-6.1. 21.

[11] aR. Haudecoeur, L. Stefan, F. Denat, D. Monchaud, *Journal of the American Chemical Society* **2013**, *135*, 550-553; bP. Chilka, N. Desai, B. Datta, *Molecules* **2019**, *24*, 752; cA. Renaud de la Faverie, A. Guedin, A. Bedrat, L. A. Yatsunyk, J.-L. Mergny, *Nucleic acids research* **2014**, *42*, e65-e65; dA. l. Laguerre, L. Stefan, M. Larrouy, D. Genest, J. Novotna, M. Pirrotta, D. Monchaud, *Journal of the American Chemical Society* **2014**, *136*, 12406-12414.

[12] G. Biffi, M. Di Antonio, D. Tannahill, S. Balasubramanian, *Nature chemistry* **2014**, *6*, 75-80.

[13] aC. K. Kwok, G. Marsico, A. B. Sahakyan, V. S. Chambers, S. Balasubramanian, *Nature methods* **2016**, *13*, 841-844; bC. K. Kwok, A. B. Sahakyan, S. Balasubramanian, *Angewandte Chemie* **2016**, *128*, 9104-9107; cX. Weng, J. Gong, Y. Chen, T. Wu, F. Wang, S. Yang, Y. Yuan, G. Luo, K. Chen, L. Hu, *Nature chemical biology* **2020**, *16*, 489-492; dS. Y. Yang, D. Monchaud, J. M. Wong, *Nature Protocols* **2022**, 1-20.

[14] aF. DUCONGÉ, J.-J. TOULMÉ, *Rna* **1999**, *5*, 1605-1614; bC. Boiziau, E. Dausse, L. Yurchenko, J.-J. Toulmé, *Journal of Biological Chemistry* **1999**, *274*, 12730-12737; cL. Aldaz-Carroll, B. Tallet, E. Dausse, L. Yurchenko, J.-J. Toulmé, *Biochemistry* **2002**, *41*, 5883-5893.

[15] F. Darfeuille, S. Reigadas, J. B. Hansen, H. Orum, C. Di Primo, J.-J. Toulmé, *Biochemistry* **2006**, *45*, 12076-12082.

[16] aA. Sett, L. Zara, E. Dausse, J.-J. Toulme, *Analytical Chemistry* **2020**, *92*, 9113-9117; bG. Durand, S. Lisi, C. Ravelet, E. Dausse, E. Peyrin, J. J. Toulmé, *Angewandte Chemie* **2014**, *126*, 7062-7065.
[17] C. Y. Chan, C. K. Kwok, *Angewandte Chemie International Edition* **2020**, *59*, 5293-5297.
[18] aM. Ghaem Maghami, C. P. Scheitl, C. Höbartner, *Journal of the American Chemical Society* **2019**, *141*, 19546-19549; bC. P. Scheitl, M. Ghaem Maghami, A.-K. Lenz, C. Höbartner, *Nature* **2020**, *587*, 663-667; cH. Jiang, Y. Gao, L. Zhang, D. Chen, J. Gan, A. I. Murchie, *Nature Catalysis* **2021**, *4*, 872-881.
[19] aK. W. Sherrill, M. P. Byrd, M. E. Van Eden, R. E. Lloyd, *Journal of Biological Chemistry* **2004**, *279*, 29066-29074; bR. Shahid, A. Bugaut, S. Balasubramanian, *Biochemistry* **2010**, *49*, 8300-8306.
[20] S. A. Woodson, *Current opinion in chemical biology* **2005**, *9*, 104-109.
[21] I. L. Hofacker, *Nucleic acids research* **2003**, *31*, 3429-3431.
[22] M. Di Antonio, G. Biffi, A. Mariani, E. A. Raiber, R. Rodriguez, S. Balasubramanian, *Angewandte Chemie* **2012**, *124*, 11235-11240.

Supporting Information

Table of Contents

General Information

DNA oligonucleotides were purchased from Microsynth. RNA oligonucleotides were synthesized by solid-phase synthesis using phosphoramidite chemistry (2'-O-TOM protected) on controlled-pore glass solid support as described earlier[1]. The NTPs, dNTPs and N^6-modified ATP analogues were purchased from Jena biosciences. SNAP-biotin was purchased from New England Biolabs. Lucifer-yellow carbohydrazide and Fluorescein thiosemicarbazide was purchased from Sigma-Aldrich. Klenow exo- fragment, T4 PNK, Superscript III, and DreamTaq polymerase were purchased from Thermo Fisher Scientific. Lambda exonuclease exo- was purchased from New England Biolabs. γ32P-Adenosine 5'-triphosphate (γ32P-ATP) was purchased from Hartmann Analytic. The TOPO-TA cloning kit from Invitrogen. Illustra Ready-to-Go RT-PCR beads were purchased from GE-healthcare. Streptaviding-coated magnetic beads (DynaBeads MyOne streptavidin T1) were purchased from Thermofisher scientific. Sera-Mag SpeedBead Neutravidin-Coated Magnetic Particles were purchased from GE-healthcare. All DNA and RNA oligonucleotides were purified by denaturing PAGE (10–20% acrylamide/bisacrylamide 19:1, 7 M urea) with running buffer 1x TBE (89 mM Tris, 89 mM boric acid and 2 mM EDTA, pH 8.3), extracted by crush and soak into TEN buffer (10 mM Tris-HCl, pH 8.0, 1 mM EDTA, 300 mM NaCl) and recovered by precipitation with ethanol. For preparative gels, oligonucleotides were visualized by UV shadowing on a TLC plate. Fluorescence gel images were recorded with a Biorad ChemiDoc MP using epi illumination. Analytical anion-exchange HPLC was carried out on a GE Healthcare Life Sciences ÄKTA™ micro system using DNAPac™ PA200 columns (4 x 250 mm, Thermo Scientific) at 60 °C. Solvent A: 25 mM TrisHCl (pH 8.0), 6 M Urea. Solvent B: 25 mM Tris-HCl (pH 8.0), 6 M Urea, 0.5 M NaClO4. Gradient: linear, 0–54% solvent B, 4% solvent B per 1 CV. The solvents were filtered through 0.2 µm cellulose acetate filters (Sartorius, Germany) prior to usage. HR-ESI-MS spectra of the synthetic products and oligonucleotides were recorded on a Bruker micrOTOF-Q III spectrometer. Fluorescent measurement of the selection samples was performed using Jasco FP-8300 spectrofluorometer.

Labeling of oligonucleotides

Labeling of RNA at 3'-end using periodate glycol oxidation

Unfunctionalized oligoribonucleotides were labeled by periodate oxidation followed by a reaction with fluorescein-5-thiosemicarbazide or Lucifer-yellow carbohydrazide. In a typical procedure, RNA (0.3–1 nmol) was dissolved in 7.5 µL water, followed by the addition of 5x sodium phosphate buffer (2 µL, 250 mM, pH 7.4) and a freshly prepared aqueous solution of NaIO4 (0.5 µL, 400 mM). The resulting mixture was incubated for 10–15 min at 37 °C. The excess of NaIO4 was quenched by treatment with aqueous Na_2SO_3 (1 µL, 1 M) for 5–10 min at 37 °C. Afterward, a solution of fluorescein-5-thiosemicarbazide (1 µL, 10 mM in DMF) or Lucifer-yellow carbohydrazide (1 µL, 10 mM in H_2O) was added, and the reaction mixture was incubated in the dark for additional 1 h at 37 °C. The labeled product was then purified by denaturing PAGE (10–20% polyacrylamide) and recovered by extraction and ethanol precipitation.

Labeling of 3'-amino functionalized oligonucleotides

Oligonucleotides with 3'-amino group were labeled with 6-carboxyfluorescein succinimidyl ester (NHS-fluorescein). Amino-modified RNA (2.5 nmol in 2.5 µL of H2O) was dissolved in carbonate buffer (20 µL, 100 mM, pH = 9.0). The solution of NHS-fluorescein in DMF (2.5 µL, 30 mM) was added and the reaction mixture was incubated at 37°C for 2 hours in the dark. Labeled RNA was purified by denaturing PAGE.

Radioactive labeling of RNA/DNA at 5'-end

Radioactive labeling of RNA/DNA was performed by phosphorylation using γ32P-ATP. In a typical procedure, RNA (200–500 pmol) was incubated with 10x PNK buffer A (5 µL, 500 mM Tris-HCl, 10 mM MgCl2, 50 mM DTT, 1 mM spermidine), γ32P-ATP (0.5 µL, 10 µCi/µL) and T4 PNK enzyme (0.5 µL, 10U/µL) at 37 °C for 1 hours. The labeled oligonucleotide was purified with PAGE and recovered by ethanol precipitation.

In vitro selection

Assembly of the selection pool

The DNA template for in vitro transcription of the starting RNA library was assembled by overlap extension using the Klenow fragment. Two DNA strands (D1+D2) with connecting loop sequence acting as the overlap region were mixed in the equimolar ratio (~250 pmol each) and annealed by heating at 95°C for 4 minutes and cooled down to room temperature for 10 minutes. The remaining reagents were added to the reaction mixture in a final concentration of 1x Klenow buffer, 200 µM of each dNTP, and 0.2 U/µL of Klenow exo-. The reaction was then incubated at 37°C for 1 hour. The sample was subjected to P/CI extraction and ethanol precipitation.

Lambda exonuclease treatment

A double-stranded DNA template having a 5' phosphate group at sense strand (non-template strand) was subjected to degradation by lambda exonuclease. In a typical reaction, ds DNA (2-3 µg) was incubated with 1x lambda exonuclease buffer and lambda exonuclease (1 U/µg of DNA) for 30 minutes at 37°C. ssDNA template was then purified using a PCR clean-up kit (Qiagen).

In vitro transcription

RNA was prepared by in vitro transcription from a synthetic DNA template using T7 RNA polymerase. Single-stranded DNA templates and T7 promoter in equimolar concentration were annealed by heating at 95°C for 3 min and then cooled down to RT for 10 min. The remaining reagents were added to the reaction in a final concentration of 40 mM Tris-HCl, pH 8.0, 10mM DTT, 4 mM each NTP, 30 mM MgCl2, and 2 mM spermidine and 20 µg of T7 RNA polymerase (prepared in house). The reaction was incubated at 37°C for 4-16 h. The reaction was quenched by 0.5 M EDTA pH 8.0 and added high-dye gel-loading solution (80% formamide, 89 mM Tris, 89

mM boric acid, 52 mM EDTA, 0.025% W/V bromophenol blue, 0.025% W/V xylene cyanol). The transcription product was purified by denaturing PAGE.

Selection reaction

In the first selection round, 2-3 nmol of starting RNA libraries (containing ~300 pmol of the fluorescently labeled RNA with Lucifer Yellow as mentioned above) were heated in presence of selection buffer (50 mM HEPES pH = 7.5, 120 mM KCl and 5 mM NaCl) at 95°C for 3 minutes followed by cooling at room temperature for 10 minutes to ensure proper folding of RNA pool and hybridization of RNA-substrate complex. The reaction was initiated with 40 mM MgCl2 and 200 µM N^6-ATP-Biotin or 100 µM of Biotinylated O^6-benzylguanine (SNAP-biotin). The reaction mixture was incubated overnight at 37°C.

The reaction volume and the amount of RNA used were continuously reduced in further selection rounds, but concentration was kept constant at ~ 40-50 µM. All reagent concentrations were kept the same as in the first selection round. After each incubation step, the reaction was stopped by the addition of TEN buffer, up to 300 µL. Afterward, the reaction mixture was ethanol precipitated and washed twice with 70% ethanol. Pellets were dried and dissolved in milli-Q water.

Capture step

Biotinylated RNA was captured by using streptavidin or neutravidin-coated magnetic beads (altered the beads in every second round to avoid the selection for neutravidin/streptavidin binders) using the established method[2]. The active fraction was then eluted with 100 µL of the elution buffer (95% formamide, 1 mM EDTA pH = 8) and heating at 95°C for 8 minutes. The eluted RNA was precipitated by ethanol and the pellet was dissolved in an appropriate amount of milli-Q water.

Quantifying the selection enrichment

After each selection reaction (biotinylation reaction) and subsequent ethanol precipitation, 10 µL of the sample was taken and diluted to a total volume of 50 µL. This sample was then subjected to fluorescent measurement using Jasco fluorescent spectrophotometer. The eluted RNA from the capture step was also dissolved in 50 µL of milli-Q water and subjected to fluorescent measurement using the same parameters. The percentage of RNA eluted was estimated after each round using the following formula

(Fe = integrated peak area of the fluorescent emission of the eluted sample, Fi = integrated peak area of the fraction before capture step, Df = dilution factor):

$$\%Eluted = \frac{Fi}{Fe * Df} * 100$$

Amplification step

Active RNA fraction was amplified by a two-step protocol. The first step involved an RT-PCR reaction to reverse transcribe the eluted RNA to cDNA followed by 10-5 cycles of PCR (also serve

as a backup sample for each selection round). In a typical reaction, 50 µL of eluted RNA sample, 30 pmol of forward primer (D4), and 50 pmol of reverse primer (D3) were transferred to the tube containing one Illustra Ready-to-Go RT-PCR bead. The resulting mixture was subjected to the reaction (42°C for 30 min, 95 °C for 4 minutes, 10× [95 °C for 30 s, 60 °C for 30 s, 72 °C for 1 min] and 72 °C for 5 minutes) using a Bio-Rad thermal cycler.

The second step involved further amplification using another forward primer (D1) that install the substrate sequence and T7 RNA promoter sequence to the 5' end of the pool. The reverse primer was the same in both amplification steps. An aliquot (5-10 µL) from RT-PCR was used as a template. This aliquot was mixed with forward primer D1 (100 pmol), reverse primer D3 (150 pmol), dNTP mixture (1 µL, 20 mM), 10x DreamTaq buffer (10 µL), DMSO (10 µL), DreamTaq DNA polymerase (0.5 µL, 5 U/µL) and water to a final volume of 100 µL. The resulting solution was subjected to a 30-cycle PCR reaction (95 °C for 4 minutes, 30x [95 °C for 30 s, 60 °C for 30 s, 72 °C for 1 min], and 72 °C for 5 min).

The ds PCR product was purified using a PCR clean-up kit (Qiagen) and subjected to lambda exonuclease treatment as described above. ssDNA template was then transcribed to RNA using the already described method of in vitro transcription.

Cloning and Sanger Sequencing

In the case of selection AO, cloning was performed in rounds 6 and 10. As we did not observe any activity in AP selection, it did not proceed for cloning. Selection AQ and AR were also cloned at round 10. From backup samples, PCR was performed using primers (D3 and D4) and following the above-stated protocol. The product was cleaned up using a PCR clean-up kit (Qiagen). TOPO TA cloning kit was used according to the manufacturer's protocol to clone the enriched pool. Individual clones were then submitted to Sanger sequencing after the confirmation of insert by colony PCR.

Ten plasmids from AO selection (5 from each round 6/10) were submitted to Sanger sequencing, and the results showed two sequence groups (**Table S3**). In the case of AQ and AR selection, ten clones were submitted from each selection, and the results were again classified into two sequence groups, as shown in table (**Table S4**).

Activity assay

Intramolecular (in-cis) activity assay

To determine the intramolecular activity of individual ribozymes or active pools, a streptavidin gel-shift assay was performed. 100 pmol of a ribozyme or active pool was incubated with the biotinylated substrate under selection condition in a total reaction volume of 10 µL. Then the excess of the unreacted biotinylated substrate was removed by ethanol precipitation. The pellet was dissolved in 20 µL water and an aliquot of 2 µL was incubated with 1x TBS buffer and 1 µL of streptavidin (1 mg/mL) at room temperature for 5 min. Samples were mixed with 1 µL of the 6x glycerol loading dye. The samples were loaded on a 10% PAGE (Biorad mini gel 10 x 8cm) and run

for 45 min at 200 V. The gel was stained with sybergreen (1x solution in TBE buffer) prior to imaging.

Intermolecular (in-trans) activity assay

Kinetics characterization was carried out to analyze the trans-activity of individual ribozymes under single turnover conditions. In a typical procedure, a ribozyme (100 pmol) was annealed (heated to 95 °C for 4 min and then cooled to RT for 10 min) with a fluorescently labeled RNA substrate (10 pmol) in 5x selection buffer (2 µL, 50 mM HEPES pH = 7.5, 120 mM KCl and 5 mM NaCl). To initiate the reaction, cofactor (0.4 µL, 2.5/5 mM), and $MgCl_2$ (1 µL, 400 mM) were added in a final volume of 10 µL. The resulting mixture was incubated at 37 °C. Aliquots (1 µL) were taken at different time points and quenched with loading buffer (4 µL) and analyzed by denaturing PAGE. Reaction yields were determined by measuring the fluorescent intensities of the corresponding bands.

Probing the modification site

Dual labeling reaction

In order to localize the modification site, ~50pmol of 3'end flu-labeled Ribozyme was annealed (heated to 95 °C for 4 min and then cooled to RT for 10 min) in 5x selection buffer (1 µL, 50 mM HEPES pH = 7.5, 120 mM KCl and 5 mM NaCl). Afterwards, N^6-ATP-ATTO550 (1 µL, 1 mM) and $MgCl_2$ (0.5 µL, 400 mM) were added in a final volume of 5 µL. The reaction was incubated overnight at 37°C and excess of cofactor was then removed by EtOH precipitation. Dried pellet was dissolved in appropriate amount (~10 µL) of milli-Q H_2O and used for subsequent reactions.

Cleavage of ribozyme using deoxyribozymes

To perform the DNAzyme mediated cleavage experiment, an aliquot (2 µL) of dual labeled ribozyme was annealed with corresponding 8-17NG deoxyribozyme (D9,D11) in 5x selection buffer (2 µL, 250 mM HEPES pH 7.5, 2 M KCl, 500 mM NaCl).To initiate the reaction, 10 mM $MnCl_2$ and 10 mM $MgCl_2$ were added to a final volume of 10 µL. In case of 10-23 deoxyribozyme (D10) a different 5x selection buffer (2 µL, 500 mM Tris-HCl pH 7.5 and 1.5 M NaCl) was used and 20 mM $MgCL_2$ was added to initiate the reaction. The reaction was then incubated at 37°C for 1h. An aliquot (1 µL) was taken at 0, and 1-hour timepoint and mixed with 4 µL of the stop solution. 2.5 µL of these samples were resolved on 15% PAGE and visualized by dual-channel fluorescent imaging.

To determine if ribozyme could work in-trans, 3'end flu-labeled ribozyme was first cleaved with deoxyribozymes and subsequently modified with N^6-ATP-ATTO550 or N^6-ATP-ATTO647N using above-mentioned protocol.

RT primer extension stop

RNA was self-modified as described earlier in section in-cis activity assay. Briefly, 100 pmol of RNA was annealed in a 5x selection buffer and a selection reaction was performed with N^6-aminohexyl-ATP-Biotin (200 µM final concentration) and 40 mM $MgCl_2$ in a total volume of 10 µL. The reaction was incubated at 37°C for ~18h and was subjected to ethanol precipitation. The pellet was dissolved in 10 µL of H_2O and used ~5 pmol in the primer extension reaction. γ32P-labeled primer (~5 pmol) was annealed with modified or non-modified RNA by heating at 95°C for 3 minutes, followed by incubation at room temperature for 10 minutes in 5x annealing buffer (1 µL, Tris 5 mM pH= 7.5, EDTA 0.1 mM). Afterward, 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 were added in a final volume of 10 µL. The reaction was incubated at 55°C for 1 hour followed by the addition of 1 µL of 2 N NaOH and incubation for 5 minutes at 95°C. The cDNA strand was then recovered by ethanol precipitation. The dried pellet was dissolved in 12 µL of H_2O. 3 µL of this sample was mixed with high loading dye resolved on 15% PAGE (40 cm long analytical gel, run at 45 W for 2 hours and 30 minutes). The gel was dried at 80°C under a vacuum for 30 minutes and cooled down to RT for 30 minutes. The gel was exposed to the phosphorus screen overnight and the screen was scanned on a Typhoon phosphorus imager.

The sequencing ladder was generated by following the established method[3]. In general, ~5 pmol of the P^{32}-labeled primer was annealed with ~5 pmol of the unmodified RNA in 1 µL of 5x annealing buffer. For the A sequencing ladder dTTP, was added to the reaction to a final concentration of 0.5 mM and dTTP at 0.05 mM. Other dNTP were added in a final concentration of 0.5 mM. For G sequencing ladder, reaction mixture included 0.5 mM ddCTP, 0.05 mM dCTP and 0.5 mM of the rest of the dNTPs. All the other reagents and reaction conditions were the same as the primer extension protocol.

Supporting Tables

Table S1. Sequence of DNA oligonucleotides

No	Discription	5'-Sequence-3'
D1	Pool forward primer	CTGTAATACGACTCACTATAGGACATACTGGGGGGCCGTGGGGTG GGAGCTGGGGGCCTTCAACCAGCCTACCATCC
D2	Pool reverse primer	GGTAAGGTGGACATACTGN_{40}GCCTTCAAGGATGGTAGGCTGG
D3	Selection reverse primer	GGTAAGGTGGACATACTG
D4	Forward primer (RT-PCR)	GCCTTCAACCAGCCTACCATCC
D6	Forward primer excluding substrate sequence	CTGTAATACGACTCACTATAGGCCAGCCTACCATCC
D7	T7 RNA promotor	CTGTAATACGACTCACTATA
D8	Primer for RT stop	CAAGGATGGTAGGCTGGT
D9	8-17 NG DNAzyme cutting in loop region	AAGGATGGTAGGTGTCAGCGACTCGAAGGTTGAAGGC
D10	10-23 DNAzyme cutting at pos 8 in BCL2	CTCCCACCCCAGGCTAGCTACAACGAGGCCCCCCAG
D11	8-17 NG DNAzyme cutting at pos 18 in BCL2	GAAGGCCCCCAGTGTCAGCGACTCGAACCCACCCCAC
D12	Forward primer G to A in 5'binding arm	CTGTAATACGACTCACTATAGGACATACTAGGGGGCCGTGGGGTG GGAGCTGGGGGCCTTCAACCAGCCTACCATCC
D13	Forward primer G1A, G2A, G3A in BCL2	CTGTAATACGACTCACTATAGGACATACTGAAAGGCCGTGGGGTG GGAGCTGGGGGCCTTCAACCAGCCTACCATCC
D14	Forward primer G23A, G24A, G25A in BCL2	CTGTAATACGACTCACTATAGGACATACTGGGGGGCCGTGGGGTG GGAGCTGAAAGCCTTCAACCAGCCTACCATCC
D15	Forward primer truncated BCL2 (-15nt)	CTGTAATACGACTCACTATAGGACATACTGGGGGGCCGTGGCCTT CAACCAGCCTACCATCC
D16	Transcription template for AO6.4 ribozyme	GGACATACTGTCACTCGAGAGACCTCCAGGGGTTACATAAGGCAG GGCCTTCAACCTATAGTGAGTCGTATTACAG
D17	Transcription template for AO10.2 ribozyme	GGACATACTGTCACTCGAGAGACCTCCAGGGGTTACATAGGGTAG GGCCCTCAACCTATAGTGAGTCGTATTACAG
D19	Pool forward primer (no binding arms)	CTGTAATACGACTCACTATAGGGGGCCGUGGGGUGGGAGCUGGGG CCAGCCTACCATCC
D20	Forward primer (RT-PCR)	CCAGCCTACCATCCTTGA
D21	8-17 NG DNAzyme cutting in loop region	AAGGATGGTAGGTGTCAGCGACTCGAAGGCCCCAGCT

Table S2. Sequence of RNA oligonucleotides

No	Discription	5'-Sequence-3'
R1	BCL2 RNA G-Quadruplex	GGGGGCCGUGGGGUGGGAGCUGGGG
R2	Truncated BCL2	GGACAUACUGGGGGGCCGUG-NH2

Table S3. Ribozymes form AO and AP selection identified by sequencing

Name	Frequency	5'-Sequence-3'
AO6.4*	1	UUGAAGGC CCUGCCUUAUGUAACCCCUGGAGGUCUCUCGAGUGA
AO10.2*	1	UUGAGGGC CCUGCCUUAUGUAACCCCUGGAGGUCUCUCGAGUGA
AO6.7	3(AO6.12+AO6.14)	UUGAAGGC CCUACCUUAUGUAACCCCUGGAGGUCUCUCGAGUGA
AO6.9	1	UUGAAGGC CCUACCUUAUGUAACCC-UGGAGGUCUCUCGAGUGA
AO6.6	1	UUGA-GGC CCUACCUUAUGUAACCCCUGGAGGUCUCUCGAGUGA
AO10.14	1	UUGAAGGC CCUACCUUAUGUAUCCCCUGGAGGUCUCUCGAGUGA
AO10.13	1	UUGAGGGC CCUGCCCUAUGUAACCCCUGGAGGUCUCUCGAGUGA
AO6.11	1	UUGAAGGC CCUACCUUAUGUAACCCCUGGAGGUCUCUCGAGUGA
AO6.16*	1	UUGAGGGC CAUCUUGUGUCCUUGACAGCACCGUUCUCCUCCCGGAGGG
AO6.13	1	UCGAAGGC CAUCUUGUGUUCUGGACAGCACCGUCCUCCCCCCGGAGGG

AO = selection with N^6-ATP-Biotin as cofactor, 6.4 = round-6 clone no.4, 10.2= round-10 clone no.2. * Representative member of the selection and used for further analysis.

Table S4. Ribozymes from AQ and AR selection identified by sequencing

Name	Frequency	5'-Sequence-3'
AQ4	4(Q8, Q18, R5)	UUGAAGGC CGCGCUCUCGGAAGCGAGUUCAAACUGCAAUCACAUGCG
AR9*	1	UUGAAGGC CGCGCCCUCGGAAGCGAGUUCAAACUGCAAUCACAUGCG
AR6	5(R3, R18, R3, R18)	UUGACGGC CGCGCUCUCGGAAGCGAGUUCAAACUGCAAUCACAUGCG
AR10	4(R16, R10`, R19)	UUGACGGC CGCGCCCUCGGAAGCGAGUUCAAACUGCAAUCACAUGCG

* Representative member of the selection and used for further analysis.

Supporting Figures

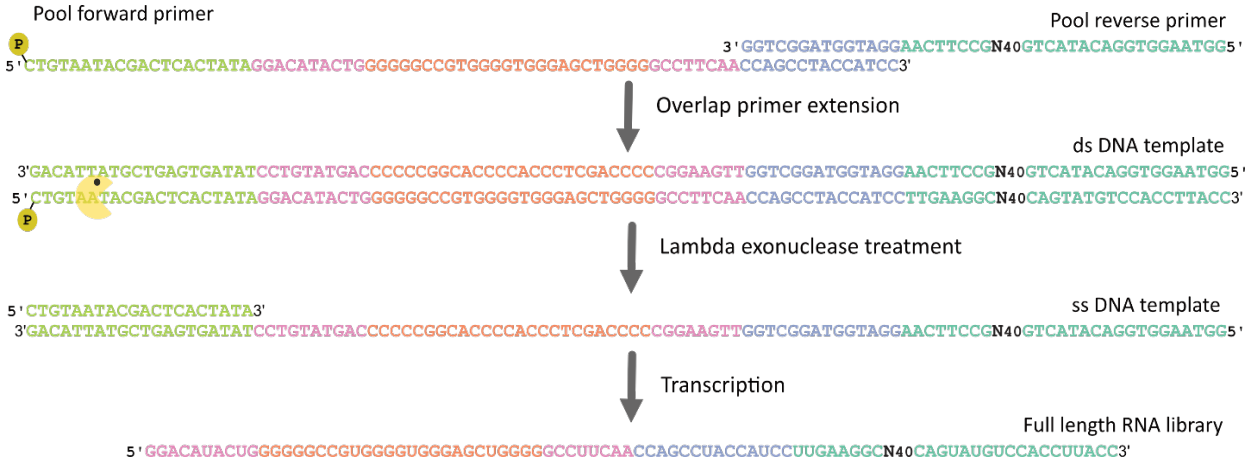
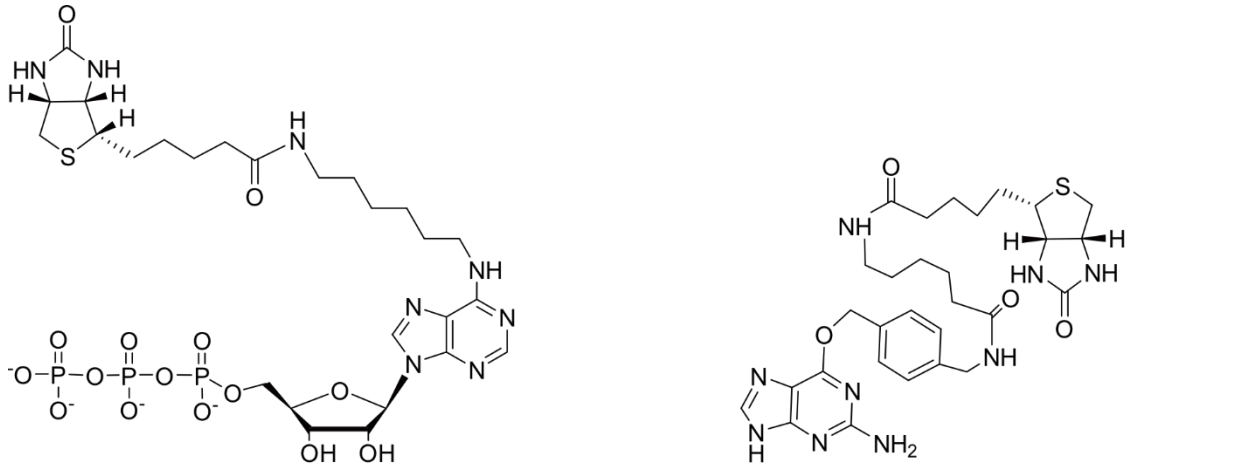


Figure S1 Schematic presentation of pool assembly. The pool is assembled via overlap extension using Klenow fragment. Two primers complementary to each other at 3'-end are extended to form ds DNA template. 5' phosphorylated sense strand is removed using lambda exonuclease and single stranded DNA is transcribed to full length RNA library.



N^6-(6-Aminohexyl)-ATP-Biotin

Biotinylated *O*^{6}-benzylguanine (SNAP-Biotin)

Figure S2 Molecular structure of N^6- (6-Aminohexyl)-ATP-Biotin and Biotinylated O6-benzylguanine (SNAP-Biotin) used as cofactors for AO and AP selection, respectively.

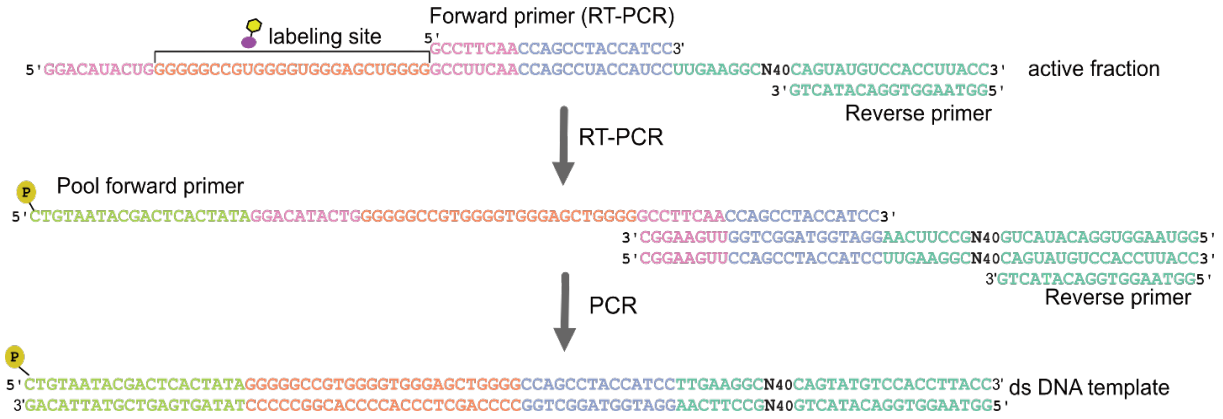


Figure S3 Schematic presentation of the amplification step. In the first PCR (RT-PCR) a forward primer was used complementary to binding arm and loop region. In second PCR amplification cycle, a long forward primer was used to restores the cDNA to its full-length and adds the T7 promoter.

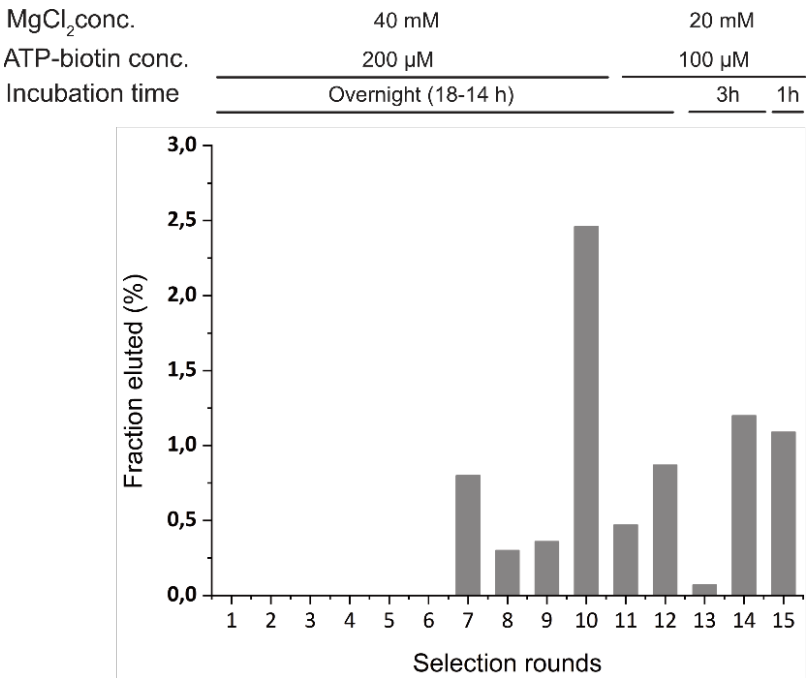


Figure S4 Progress of AO selection. The percentage of the eluted RNA after each round was estimate using fluorescence measurement. The selection stringency was increased by reducing the incubatio time, cofactor concentration and MgCl2 concentration.

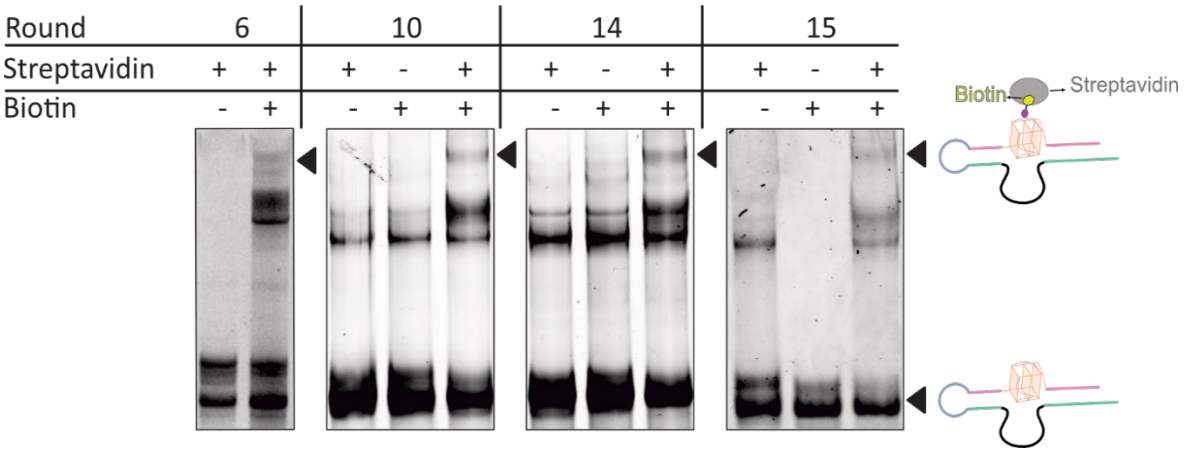


Figure S5 Cis-activity analysis of the enriched pool. Streptavidin gel-shift assay of the rounds 6, 10, 14 and 15 pool. Reaction conditions: 10 µM RNA Pool incubated in selection buffer including 200 µM biotinylated ATP and 40 mM MgCl2 overnight at 37°C followed by incubation with streptavidin protein. The samples were subjected to 10% PAGE. Shifted bands shows the self-biotinylation of active fraction in the pool.

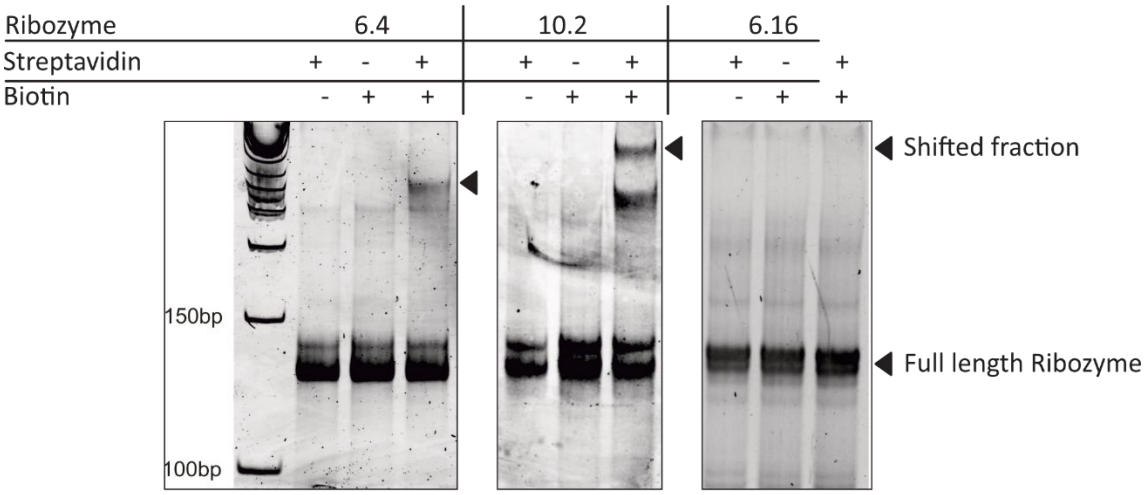


Figure S6 Cis-activity analysis of the individual ribozymes. Streptavidin gel-shift assay of AO6.4, AO10.2 and AO6.16 ribozymes. Reaction conditions: 10 µM RNA Pool incubated in selection buffer including 200 µM biotinylated ATP and 40 mM MgCl2 overnight at 37°C followed by incubation with streptavidin protein. The samples were subjected to 10% PAGE. Shifted bands shows the self-biotinylation of ribozymes.

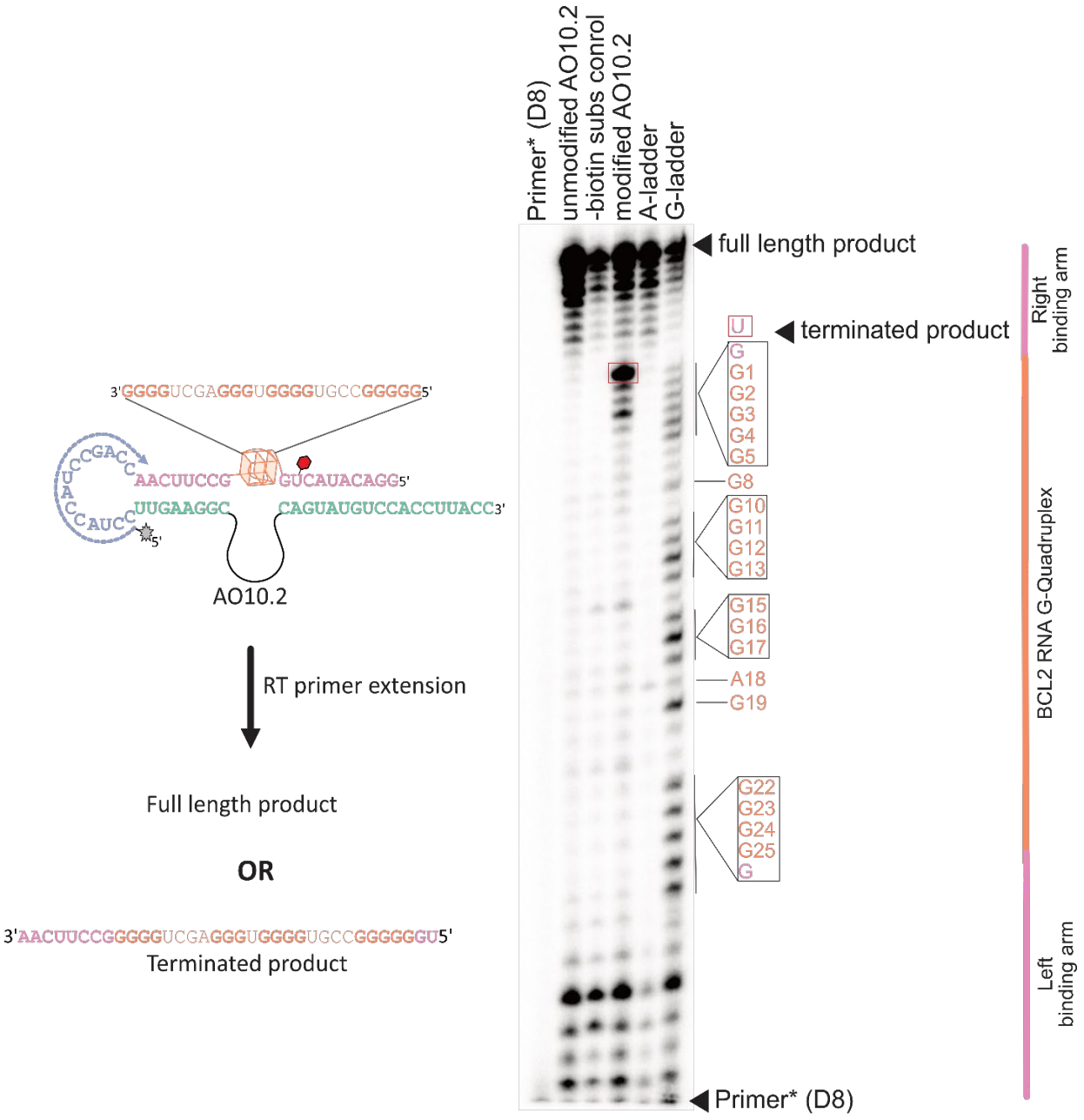


Figure S7 Probing the modification site using RT-primer extension stop assay. AO10.2 was self -biotinylated using N^6-ATP-Biotin. 32P-labeled primer complementary to the loop region was used for primer extension reaction. The reaction was performed on unmodified, RNA incubated without cofactor and modified RNA. A and G sequence ladders were made using primer extension in the presence of ddTTP and ddCTP, respectively.

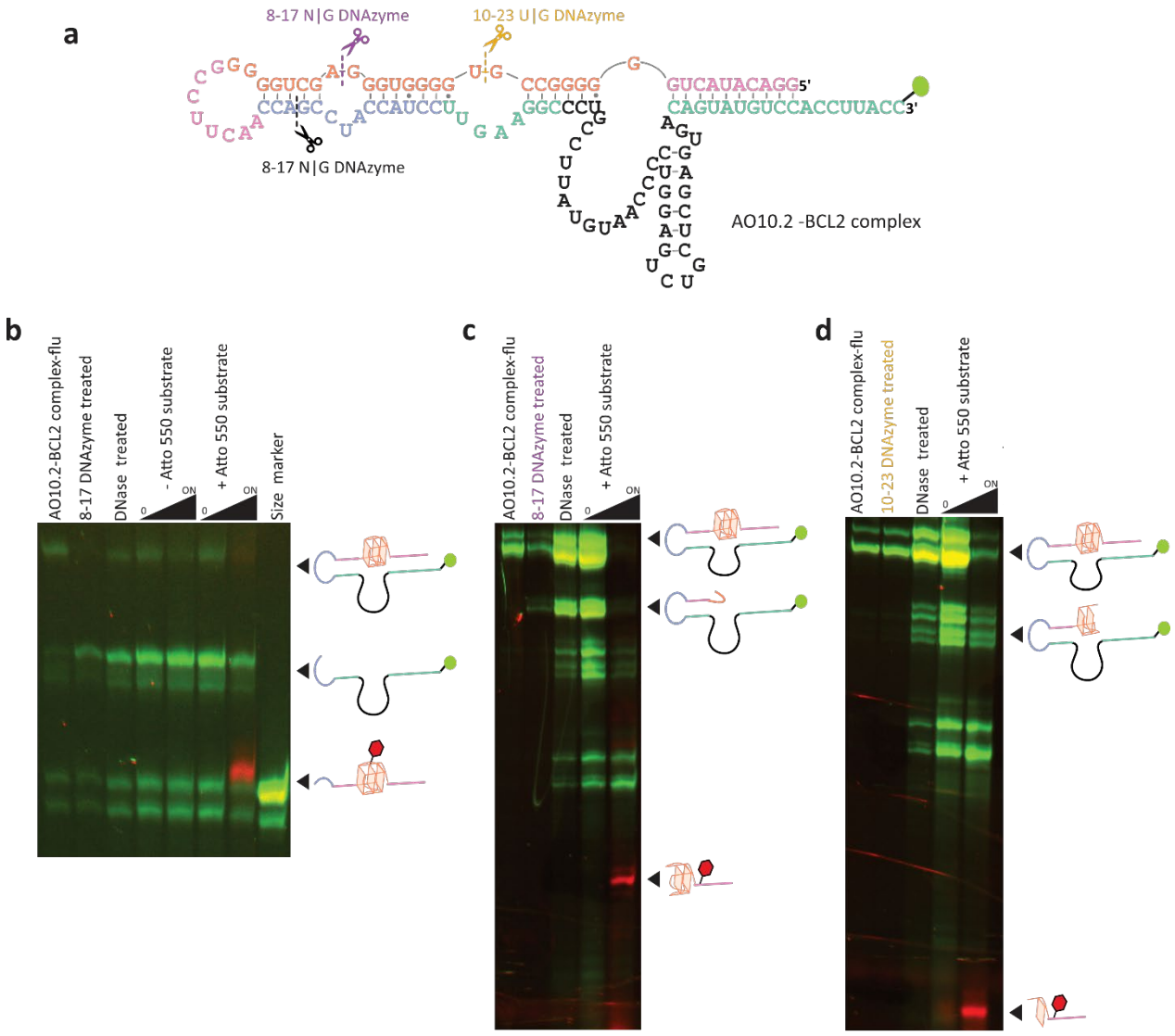


Figure S8 In-trans activity analysis of AO10.2 ribozyme. a) Schematic presentation of AO10.2-BCL2 complex illustrating three different cleavage sites for DNAzymes. Black: represent the 8-17 NG cleavage site in the loop region. Purple: represent the 8-17 cleavage site in the BCL2 region. Yellow: represent the 10-23 cleavage site in BCL2 region. b) 3-flu-end labeled AO10.2-BCL2 complex cleaved with 8-17 DNAZyme (black) c) 3-flu-end labeled AO10.2-BCL2 complex cleaved with 8-17 DNAZyme (purple). d) 3-flu-end labeled AO10.2-BCL2 complex cleaved with 10-23 DNAZyme (yellow). Reaction conditions: DNAzyme mediated cleavage reaction was performed with 10pmol of RNA complex and 100pmol of respective DNAzyme followed by DNAase treatment to remove DNAZyme. Two RNA fragments were incubated in selection buffer including 200 µM N^6-ATP-ATTO550 and 40 mM MgCl2 overnight at 37°C. Samples were resolved on 10% denaturing PAGE and labeled RNA fragment appeared as red band.

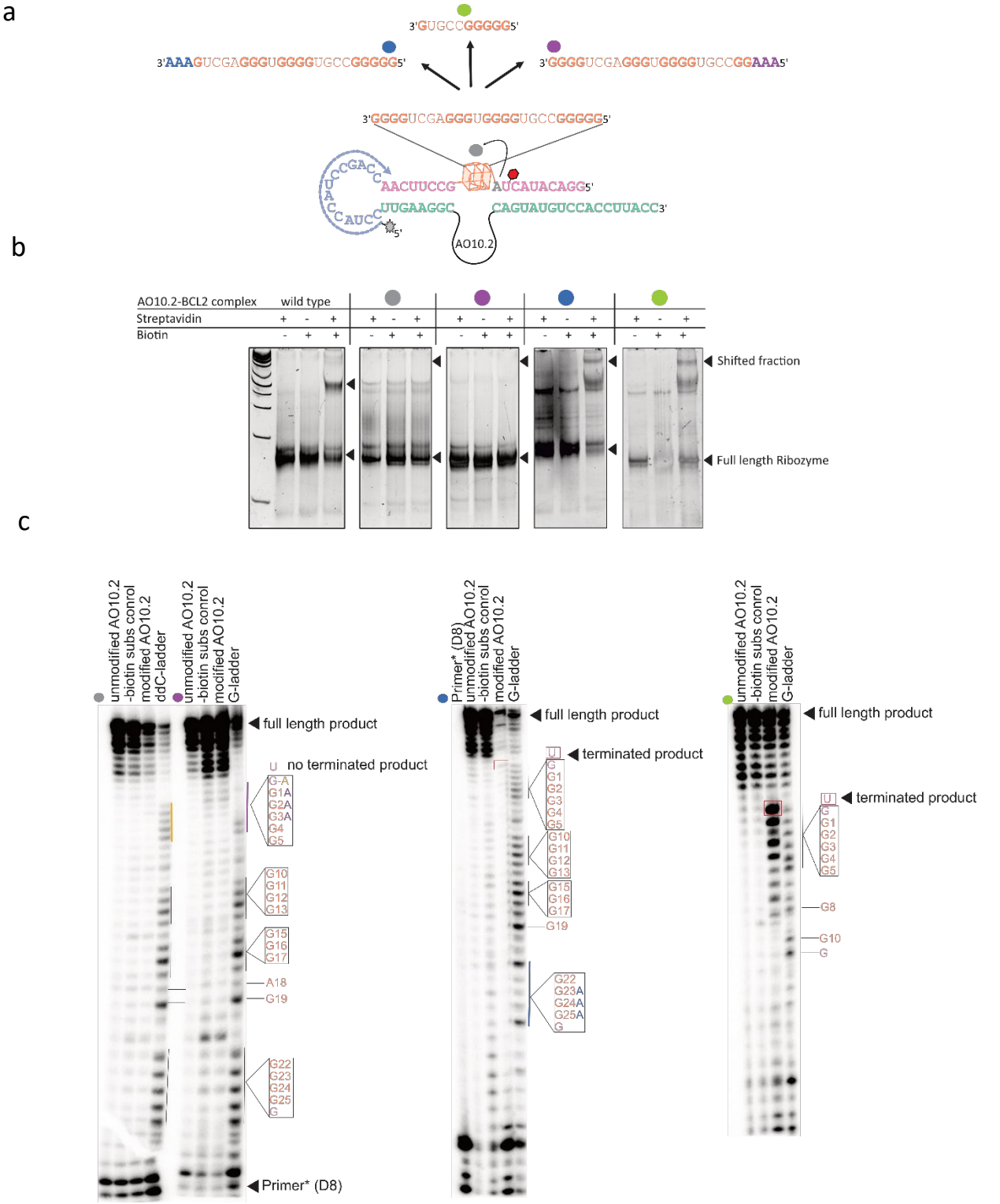


Figure S9 rG4s dependent activity analysis of AO10.2 ribozyme. **a)** Schematic presentation of AO10.2-BLC2 complex with four possible mutation sites. Grey: G next to the modification site mutated to A, Purple: G1A, G2A, G3A mutations at 5'-end of BCL2, Blue: G23A, G24A, G25A mutations at 3'-end of BCL2 and Green: 15 nucleotides were deleted form 3'-end of BCL2. **b)** Streptavidin gel-shift assay of self-biotinylated all four AO10.2-BCL2 complexes. Reaction conditions: 10 µM RNA complexes incubated in selection buffer including 200 µM biotinylated ATP and 40 mM $MgCl_2$ overnight at 37°C followed by incubation with streptavidin protein. The samples were subjected to 10% PAGE. Shifted bands shows the self-biotinylation of active complexes. **c)** P^{32}-labeled primer complementary to the loop region was used for primer extension reaction. The reaction was performed on unmodified, RNA incubated without cofactor and modified RNA complexes. G sequence ladders were prepared using primer extension in the presence of ddCTP.

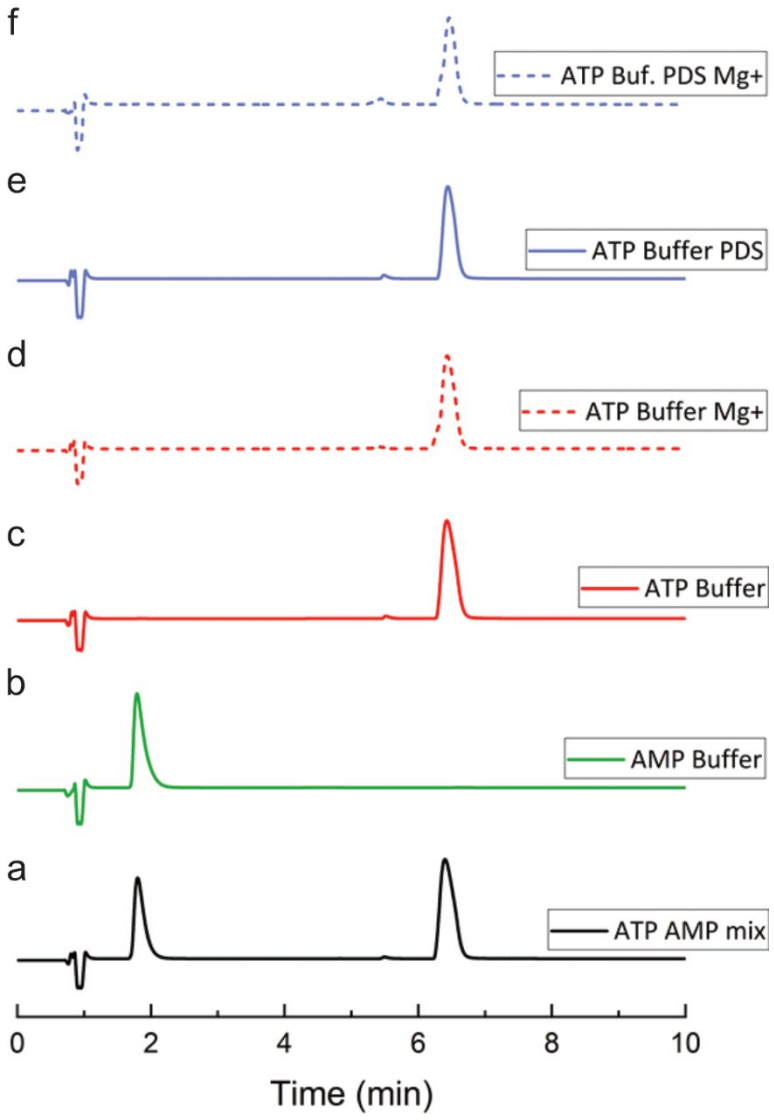


Figure S10 Anion exchange HPLC analysis of ATP during reaction conditions. **a)** ATP and AMP coinjected to HPLC. **b)** AMP after overnight incubation in 1x selection buffer. **c)** ATP after overnight incubation in 1x selection buffer. **d)** ATP after overnight incubation in 1x selection buffer and 40mM $MgCl_2$. **e)** ATP after overnight incubation in 1x selection buffer and 200 µM of PDS. **f)** ATP after overnight incubation in 1x selection buffer, 200 µM of PDS and 40mM $MgCl_2$.

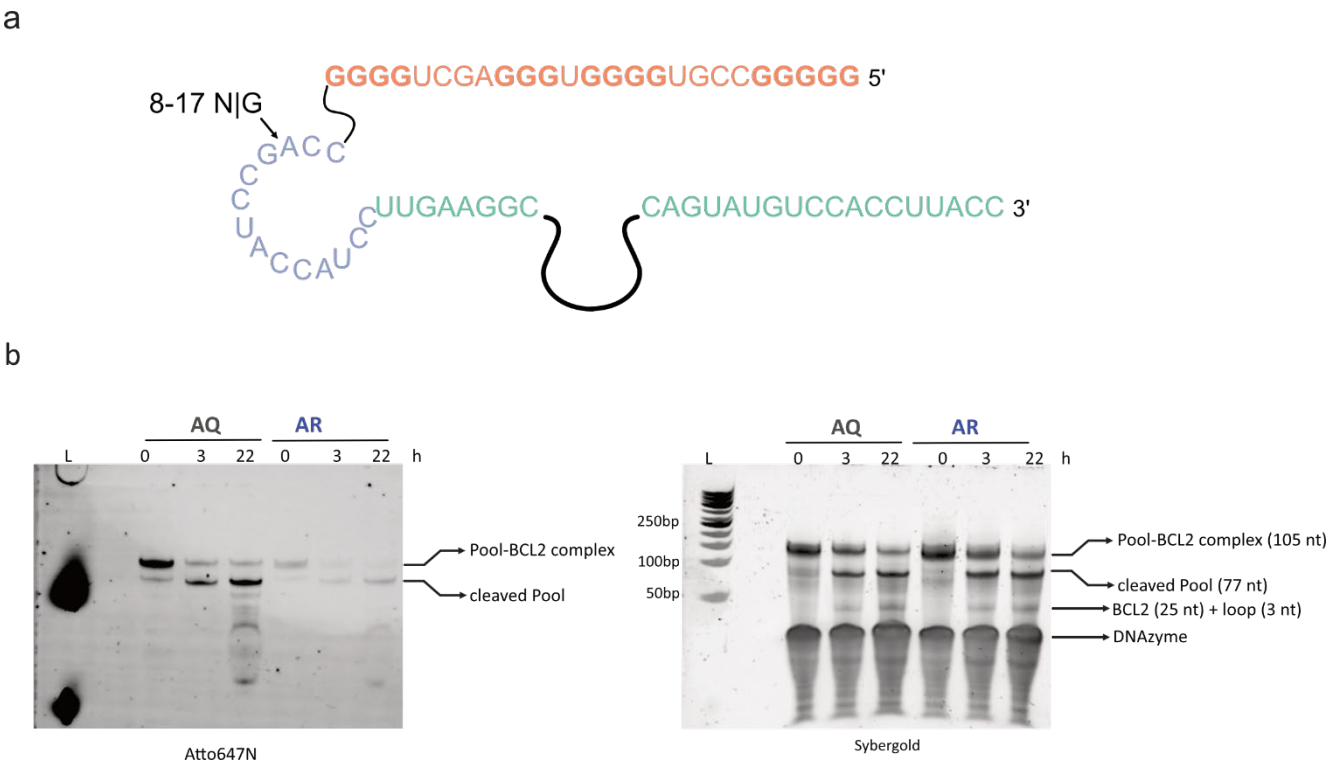


Figure S11 In-trans activity analysis of AQ and AR pool. a**)** Schematic presentation of Pool-BCL2 complex illustrating cleavage sites for DNAzymes. **b)** Unlabeled pool-BCL2 complexes were incubated under selection conditions including 200 µM N^6-ATP-ATTO647N and 40 mM $MgCl_2$ overnight at 37°C. Labeled pools were cleaved with 8-17 DNAZyme. Time points were taken at 0, 3h and 22h and aliquots of samples were resolved on 10% denaturing PAGE.

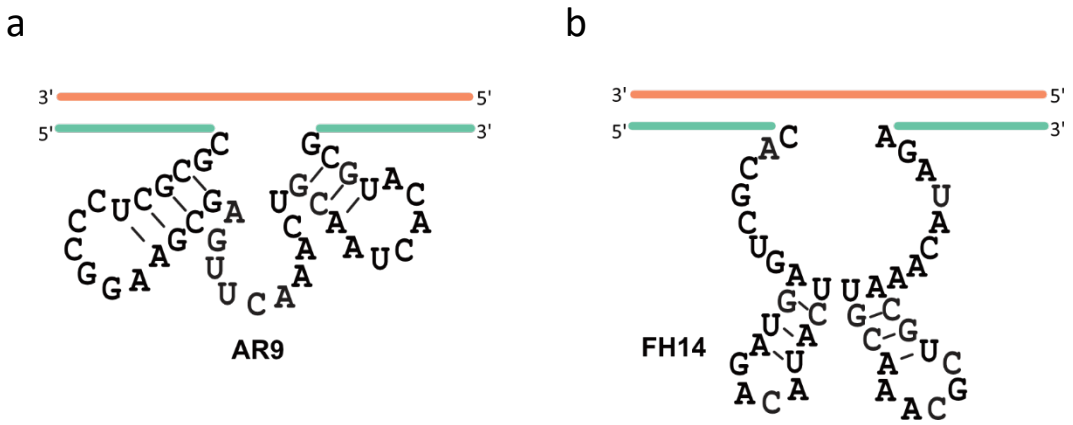


Figure S12 Predicted secondary structures of AR9 and FH14. The two stem-loops and single stranded core region is similar in both ribozymes.

References

[1] C. Höbartner, C. Kreutz, E. Flecker, E. Ottenschläger, W. Pils, K. Grubmayr, R. Micura, *Monatshefte für Chemie/Chemical Monthly* **2003**, *134*, 851-873.

[2] M. Ghaem Maghami, C. P. Scheitl, C. Höbartner, *Journal of the American Chemical Society* **2019**, *141*, 19546-19549.

[3] C. Shahn, E. G. Strauss, J. H. Strauss, in *Methods in enzymology, Vol. 180*, Elsevier, **1989**, pp. 121-130.

7. Discussion and Conclusion

Nucleic acids (DNA and RNA) are structurally diverse molecules primarily known to store, transfer and express the genetic information in nature. An additional functional perspective on RNA was exposed with the discovery of the catalytic activity of RNA, in the form of group-I introns and RNase P ribozymes[1]. In order to discover functional nucleic acids with new activities, a powerful method was developed called SELEX (or in vitro selection). The SELEX method was initially used to artificially evolve functional RNA molecules i.e., aptamers or ribozymes[2]. Afterwards, DNA as a catalyst was discovered using in vitro selection[3]. RNA is the only catalytic functional nucleic acid known to exist in nature, whereas catalytic DNA molecules are synthetic and evolved in the laboratory by vitro selection.

Single-stranded DNA and RNA perform diverse functions by forming specific structures through stacking interaction, hydrogen bonding, or metal ion coordination. These structural domains function as a catalytic unit during the enzymatic reaction, binding pockets for metabolites/metal ions or interacting motifs towards other RNAs and proteins. This structural diversity in single-stranded nucleic acids set the basis for their functional diversity.

Here we discuss nucleic acid catalysts as a potential tool for the detection of RNA modifications and RNA structures in two parts.

7.1 Artificial evolution of modification sensitive RNA-cleaving deoxyribozymes

Among all DNA enzymes, RNA-cleaving deoxyribozymes are the well-studied group and have been employed in various applications including therapeutics and biosensors[4]. RNA-cleaving deoxyribozymes also has promising application in the field of epitranscriptomics as a tool for the detection and validation of RNA modifications. Already existing 10-23 and 8-17 deoxyribozymes were employed for the detection of RNA modifications followed by radioactive labeling of cleaved fragments and thin-layer chromatography. The method is dependent on equal cleavage efficiency of DNAzymes for modified and unmodified RNA, thus limited to study the small modification usually present at non-Watson-Crick edges (e.g., m^5C, m^5U, Ψ)[5]. RNA-cleaving deoxyribozymes usually catalyze a transesterification reaction where the 2'-hydroxyl group attacks the adjacent phosphodiester linkage, forming 2'-3' cyclic phosphate and 5'-hydroxyl termini. Therefore,

modifications (such as 2'-O-Me) directly attached to the functional group involved in the cleavage reaction were identified using deoxyribozymes, where the presence of modification at the cleavage site prevents DNA- catalyzed cleavage[6].

In contrast, a direct approach was developed where chemical modification influences the catalytic performance of RNA cleaving DNAzymes when present in close proximity to the cleavage site. The library design and in vitro selection strategy was modified to evolve DNAzymes for site-specific identification of m^6A[7]. All-RNA substrate containing DRACH (GGACU in particular) motif was used during selection unlike an embedded single ribonucleotide in a DNA substrate used in previous selections for RNA-cleaving deoxyribozymes[4a]. The DNAzyme library was designed to have an N_{20} random region (advantageous to cover full sequence space in a single selection experiment) flanked by binding arms complementary to the RNA substrate and a loop region to facilitate the RNA substrate ligation via splint ligation. The binding arms of the DNA pool were designed to keep two nucleotides unpaired upstream from the modification site to evolve deoxyribozymes having enhanced variability in terms of sequence context at cleavage site. The selection was performed under mild conditions using as little as 5 mM Mg^{2+} ions. This avoids the high metal ion concentration or combination with transition metal ions (such as Mn^{2+}), which can lead to unspecific degradation of RNA substrate. The selection successfully identified DNAzymes sensitive to m^6A nucleotide; VMC10 was completely inhibited in the presence of the m^6A modification presumably due to methyl group interfering with the active conformation of catalytic core. On the contrary, VMA8 and VMA15 were two DNA enzymes that exhibited accelerated rate of reaction for m^6A modified RNA but it was not possible to completely abolish the activity in the absence of methyl group. However, the kinetic differences were significant to directly report the modification state of RNA, in particular with the help of a calibration curve generated with synthetic RNA substrates of known m^6A stoichiometry.

In addition to the fundamental interest in training deoxyribozymes for other post-transcriptional RNA modifications, DNA enzymes' response to different analogues and derivatives is also important to understand their catalytic potential and sensitivity as enzymes. Our study started with evolution of deoxyribozymes for identification of RNA modifications related to m^6A.

7.1.1 N^6-isopentenyladenosine sensitive DNAzymes

The discovery of m^6A-sensitive deoxyribozymes was a milestone in the development of DNAzymes for the detection of RNA modifications. However, the detailed interaction of the methyl group with the catalytic core of DNA and the role of magnesium ion in stabilizing those interactions is currently unknown. Without structure and mechanistic studies of deoxyribozymes, it is challenging to engineer existing enzymes to identify other RNA modifications.

Therefore, we performed a new in vitro selection to evolve deoxyribozymes that selectively cleave N^6-isopentenyladenosine (i^6A)-modified RNA. Because i^6A is a structural analogue of m^6A that contains a bulkier functional group (dimethylallyl) at the N^6-position of adenosine, we followed the same library design and selection strategy used to evolve m^6A-sensitive DNAzymes. The purpose was to identify deoxyribozymes that could discriminate structural analogues in addition to the site-specific detection of i^6A modification. The selection surfaced three classes of deoxyribozymes that have distinct responses to i^6A modification. We found AA deoxyribozymes which are similar in their activity to the VMC10 DNAzymes and catalyze the cleavage of unmodified RNA but are inhibited by i6A. In addition, we reported AB08 as the first DNAzyme that specifically and efficiently cleaved i^6A -modified RNA and was inhibited by unmodified RNA. The bulkier and more hydrophobic i^6A modification in the RNA presumably had a much stronger activating effect on deoxyribozyme-catalyzed RNA cleavage. Furthermore, it possesses the ability to discriminate the i^6A and i^6A analogues to a significant extent and supports DNAzyme as a sensitive tool for the detection of specific modifications. In this series of i^6A-sensitive enzymes, AC17 showed a distinct behavior by cleaving both i^6A-modified and unmodified RNA at different positions. AC17 cleaved unmodified RNA directly next to the target position and it shifted to one nucleotide upstream in the case of i^6A-modified RNA suggesting that i^6A can modulate the active site of the DNA enzyme.

A remarkable degree of similarity was observed between AC17 and VMC10 upon sequence alignment and secondary structure prediction. AC17 is a single point mutant (C12T) of VMC10 which showed strong inhibition by m^6A. After AB08, AC17 was reported as the second enzyme in this series that can distinguish i^6A from m^6A. Guanosine triplet at the 5'-bulge was a common motif observed in both AB08 and AC17, but the functional significance of this motif is still

unknown. AC17 and AB08 also had conserved motifs such as 5'-AGY-3' tri-loop and CG dinucleotide bulge identified in the 8-17 DNAzymes family[4a].

The Watson-Crick base-paired binding arms of DNAzymes are exchangeable to target various RNA substrates as has already been shown for many RNA-cleaving deoxyribozymes[4a]. In contrast, the critical part is the identity of the non-Watson-Crick base-paired nucleotides close to the cleavage site, which cannot always be freely chosen. Similarly, we have observed that mutation of nucleotide near the cleavage site affects the cleavage activity and reduces the fidelity of modification recognition by the deoxyribozymes. Since m^6A is mostly found in the DRACH motif of natural RNAs, it was not a problem for m^6A -sensitive enzymes. i^6A modification, however, is located at position 37 of certain tRNAs in different sequence contexts. Therefore, using a similar selection substrate as used for in vitro selection of m^6A -sensitive DNAzymes turns out not to be a good choice because it limits the utility of DNAzymes as a tool for detection of i^6A modification in natural RNA sequences.

7.1.2 DNAzymes sensitive to methylated cytidine isomers

The finding of i^6A-sensitive deoxyribozymes inspired the search for DNA catalysts that target other RNA modifications such as monomethylated cytidine isomers 3-methylcytidine (m^3C), N^4-methylcytidine (m^4C), and 5-methylcytidine (m^5C) and discriminate them from unmodified cytidine.

We performed four gel-based in vitro selections using three modified-RNA substrates, each containing one of the methylated cytidines and an unmodified RNA. The selection approach was similar to the previous experiments, but the sequence of selection substrate was altered to evolve versatile deoxyribozymes that can identify cytidine modifications within all four possible dinucleotide junctions (NC; N = A, G, C, or U). Therefore, the RNA selection substrate contained a degenerate nucleotide upstream of the target nucleotide. Unlike the previous design, binding arms of the DNA library were intended to keep one nucleotide unpaired upstream from the modification site (i.e., degenerate nucleotide position). Consequently, it will be possible to isolate DNA catalysts that can accept any nucleotide upstream of the modified nucleotide and can be designed to target any RNA substrate by the Watson-Crick base-pairing.

The evolution and catalytic behavior of methylcytidine isomers sensitive deoxyribozymes depend on the different structural properties of methylated cytidines. The selection evolved AL112 deoxyribozymes that discriminate m^3C-modified RNA from unmodified RNA in all four NC sequence contexts and cleaved upstream of the modification site. The predicted secondary structure showed that m^3C might disrupt the Watson-Crick base pairing[8] between m^3C and dG1 (dG1 = first nucleotide from the enzyme's catalytic core) therefore supporting the cleavage of m^3C modified RNA. On the other hand, m^4C, m^5C, and C exhibit canonical base pairing and significantly inhibited the cleavage of respective substrates.

DNAzyme AM101 deoxyribozyme was selective for m^4C-modified RNA in YC (Y= C or U) context and cleaved one nucleotide downstream of the modification site. Um^4C was the preferred cleavage junction for AM101 with a higher cleavage rate compared to Cm^4C, presumably due to the wobble base pairing between rU15-dG18 (rU15= position before modification in RNA substrate, dG18 = first nucleotide from the enzyme's catalytic core) whereas rC15-dG18 Watson-Crick base pairing significantly reduced the cleavage yield. The methyl group in m^4C is attached to the exocyclic amino group that can either retain (in anti-conformation) or disrupt Watson–Crick base pairing (in syn-conformation)[9]. These unique structural properties probably contributed to the evolution of m^4C-specific enzyme that was completely inhibited by other methylated cytidines (e.g. m^3C and m^5C). In addition, AM101 also discriminated the structural analogue N^4,N^4 -dimethylcytidine ($m^{4,4}C$) from m^4C, demonstrating that DNAzymes are a specific tool for the detection of RNA modification.

The deoxyribozyme AN05 cleaved m^5C-modified RNA with an accelerated rate in the UC context one nucleotide upstream of the modified nucleotide. Like AM101, showed complete selectivity and was inhibited by the other two methylated cytidines m^3C and m^4C. A plausible rationalization is based on the fact that m^5C does not disturb base-pairing directly[10] thus stabilizing m^5C-dG1 base pairing (dG1 = first nucleotide from the enzyme's catalytic core) responsible for m^5C specific cleavage. In the case of m^3C and m^4C, this base pair might be destabilized or disrupted and thus led to inhibition of cleavage.

Lastly, AK104 was identified that efficiently cleaved unmodified RNA and inhibited cytosine modifications in the DC context (D= A, G, U). In this case, the cleavage site was situated directly

upstream of the target cytosine. The behavior of AK104 showed resemblance to m^6A-sensitive VMC10 and i^6A-sensitive AA deoxyribozymes as they also report the modification status by inhibiting the cleavage of modified RNA.

The idea of using degenerate nucleotide upstream of modified nucleotide partially enabled the selection of deoxyribozymes with sequence generality as in the case of AL112, AM101, and AK104. In addition, many DNA catalysts were good at differentiating methylated from unmethylated RNA in more than one sequence context, or at least we identified one deoxyribozyme for almost every combination of modification and junction (shown in Chapter 3).

Interestingly, analysis of predicted secondary structures showed no similarity to the 8-17 DNAzymes family, which was repeatedly observed in various other independent in vitro selections of RNA cleaving deoxyribozymes[4a, 7]. This is probably because we used RNA substrates with NC at the cleavage junction, and 8-17 variants are known to have poor activity at this junction[4a, 7].

DNA catalysts as a tool should easily be designed to target various RNA sequences for site-specific detection or validation of post-transcriptional RNA modifications. As mentioned earlier, Watson-Crick base-paired binding arms are usually flexible to alter for desired RNA substrate, but in this selection, non-Watson-Crick base-paired nucleotides are also flexible due to the use of degenerate nucleotide in selection substrate at the cleavage junction. The potential of AL112, AK104, and AN05 deoxyribozymes for the detection of RNA modifications in heavily structured RNAs was tested by targeting two different sites in human mt-tRNAs. AL112 successfully cleaved human mt-$tRNA^{Thr}$, which contains m^3C in the anticodon loop (C32). AN05 designed for mt-$tRNA^{Glu}$ with m^5C at position 49 showed an accelerated cleavage rate compared to the unmodified substrate. AK104 cleaved both unmodified tRNAs with a high yield and was significantly inhibited by m^3C and m^5C.

7.1.3 High throughput activity profiling of DNAzymes

In the previous selections for methylcytidine-sensitive deoxyribozymes, we observed high diversity in libraries even after 18 rounds of selection. Sanger sequencing of 40 clones identified 36 unique sequences, suggesting a high diversity in the selection libraries. To potentially identify further DNAzymes, we performed next-generation sequencing (NGS) of selection pools. Overall,

we tested 45 DNA sequences for 16 RNA substrates, which means ~720 independent reactions to identify the cleavage activity. The process is time-consuming and laborious as every candidate sequence had to be individually assayed for its cleavage activity.

We implemented deoxyribozyme sequencing (DZ-seq), an NGS-based strategy that allows massively parallel quantitation of RNA cleavage activity directly from sequencing data. The key process was the construction of the sequencing library as we were dealing with a hybrid system in which an active DNA pool was ligated to RNA substrate. In contrast, already reported methods for self-cleaving deoxy(ribozymes) employed different techniques for library preparation[11]. In case of self-cleaving twister ribozyme, a high throughput strategy was employed to assay ribozymes mutants[11a]. A partially randomized ribozyme library was generated by in vitro transcription from a DNA template followed by self-cleavage reaction under desired conditions. The cleaved and uncleaved ribozymes were reverse transcribed into cDNA and processed to attach adapter and barcode sequences for deep sequencing. In a follow-up study, Zn^{2+}-dependent self-hydrolyzing deoxyribozyme I-R3 was subjected to high throughput mutational and kinetic analysis[11c]. A doped deoxyribozyme library was prepared by solid-phase synthesis in which the positions targeted for mutagenesis contained 79% of the original (wild-type) base and 7% of each of the remaining three bases. The resulting library was allowed to undergo self-cleavage under chosen conditions. Subsequently, an adapter oligonucleotide was ligated to the 5’ end of both cleaved and uncleaved fraction. The DNAs were circularized by CircLigase as 3’ ends of the DNAs were heterogeneous (due to cleavage and randomization). Circular DNA libraries were then PCR amplified for high throughput sequencing.

In Dz-seq, 5’-adapter sequence was introduced by an asymmetric PCR reaction using an in-vitro selection backup sample (obtained after initial PCR amplification of the enriched library) followed by ligation to RNA substrate. The in vitro selection library (DNA-RNA hybrid) was then subjected to self-cleavage. Next, the 2’,3’-cyclic phosphate residues resulting from the transesterification reaction were removed by T4-polynucleotide kinase (T4-PNK), and both cleaved and uncleaved species were polyadenylated using poly(A) polymerase to generate homogenous 3’ ends as it was more efficient compared to circularization in our case.

The data obtained from NGS was used to determine the total numbers of cleaved and intact reads and the position of the cleavage site for each unique sequence that had at least 100 reads. This data was then used for the calculation of cleaving activity, which was expressed as the fraction of cleaved reads (FC). The FC values and position of the cleavage site obtained from DZ-seq were validated by PAGE analysis for AM301 and AM319-AM327 deoxyribozyme sequences that were chosen to span the full activity range. DZ-seq showed a good linear correlation with the PAGE assay i.e. $R^2 = 0.91$ and correctly identified the position of the cleavage site. The slight deviation of PAGE assays from DZ-seq was attributed to the reaction conditions during library preparation (for example dephosphorylation and poly(A)-tailing reaction buffers contain Mg^{+2} ions that promote RNA cleavage).

Interestingly, DZ-seq revealed a poor correlation between enzyme activity and the abundance of deoxyribozymes in the final rounds of selection. Low abundant deoxyribozymes had very low chances of being identified from conventional Sanger sequencing, but DZ-seq discovered few interesting DNA sequences. For example, AM301 had a very low abundance in the final rounds (~90 CPM) but showed high yields (~ 80– 98% FC after 6h) in various cleavage junctions. AK314 was another low abundant (~18 CPM) and moderately active (~60% FC after 6h) DNAzyme. This enzyme has remarkable similarity with 8-17 DNAzymes (i.e., 5'-AGC-3' tri-loop and CG dinucleotide bulge) and a very short catalytic core of 13-nt evolved from 20-nt random region of the library[4a].

7.2 Artificial evolution of structure-specific RNA labeling ribozymes

Ribozymes evolved from in vitro selections are known to perform a variety of functions. Among them, ribozyme-based RNA labeling is a quite efficient approach to study RNA folding, structural dynamics, and localization. Self-alkylating ribozymes such as epoxide reactive ribozyme derived from *Aeropyrum prenix* genomic RNA and the fluorescein–iodoacetamide reactive ribozyme have been developed for RNA labeling[12]. However, self-alkylating ribozymes needed to be inserted into the target RNA. To overcome this limitation, a twin-ribozyme was engineered to cleave a section from target RNA and replaced it with a labeled RNA fragment[13]. But the low labeling efficiency of the multistep reaction is an obstacle to their application. Along these lines, FH14 and FH31 were reported as efficient trans-acting ribozymes for site-specific labeling of RNA[14]. These

ribozymes attached N^6-labeled adenosine to an internal 2'-OH group through a 2',5'-phosphodiester bond. They showed a high versatility to target various RNA sequences simply by designing the binding arms complementary to the target RNA and a good scope for a broad range of N^6-labeled ATP derivatives.

These existing approaches rely on the sequence-specific recognition of the target RNA by Watson-Crick base pairing. To overcome this limitation, our study aimed at the development of structure-specific ribozymes. The design of our selection library was a modified version of the FH14 ribozyme selection pool. Instead of targeting a single bulged-out nucleotide, we target an RNA G-quadruplex (5'UTR BCL2) for covalent labeling using N^6-ATP-Biotin after recognition by ribozyme through tertiary interaction. In parallel, we also performed a selection using SNAP-Biotin as a cofactor, which has already been used as a cofactor for in vitro selection of trans-acting MTR1 ribozymes[15]. In the selection with N^6-ATP-Biotin as a cofactor, we observed enrichment of sequences that label uridine in the binding arm instead of the target BCL2 region. The fixed binding arms presumably forced the random region to open the G-quadruplex structure, which is why active molecules were unable to recognize BCL2 and catalyzed labeling reaction even with a truncated BCL2 sequence. Unfortunately, we did not observe enrichment in the selection performed with SNAP-Biotin as a cofactor.

Subsequently, we modified the design of the RNA pool by removing the binding arms and attaching the G-quadruplex directly to the library via a small connecting loop. Free-hanging BCL2 region is relatively more relaxed and can easily fold into a native G-quadruplex. Moreover, we used rG4s binding ligand (pyridostatin) to ensure that the quadruplex retained its functional folding during the selection process. The adenylyl transferase ribozyme identified in this selection was represented as AR9. The ribozyme AR9 was a trans-active, self-labeling ribozyme that catalyzes nucleotide transferase reaction in the presence of the BCL2 sequence suggesting that AR9 required at least BCL2 sequence to perform its activity. However, it is not yet clear whether AR9 recognizes BCL2 based on its structure. Further research is needed to understand how AR9 interacts with the BCL2 RNA.

7.3 Conclusion and Outlook

During this thesis project, deoxyribozymes were established as tools for the site-specific detection and validation of a series of post-transcriptional RNA modifications. Our selection process evolved DNAzymes that discriminate different analogues of adenosine and single methylated isomers of cytidine. The dedicated design of the in vitro selection scheme led to the identification of versatile DNA catalysts that could be easily engineered for various target RNAs. We demonstrated that AL112, AN05, and AK104 had high target specificity and successfully validated the m^3C and m^5C modification sites in specific human mitochondrial tRNAs. Furthermore, to accelerate the DNAzymes discovery, we established an NGS-based strategy (DZ-seq) for profiling activity of thousands of RNA-cleaving deoxyribozymes from sequencing data. In the future, we envision that DZ-seq can be employed for a comprehensive analysis of in vitro selections and library composition. Also, DZ-seq has been shown quite an efficient and reliable approach, it can be applied for the identification of DNA enzymes for the plethora of RNA modifications.

In the second project, we discovered two ribozymes that catalyze RNA labeling reactions using fluorescent ATP analogous as substrates. The ribozymes AO10.2 and AR9 originated from two subsequent selections with alternative designs that aimed at labeling an RNA quadruplex structure. Even though AO10.2 was unable to recognize RNA G-quadruplex, it is a trans-active ribozyme that labeled uridine in target RNA. In the future, AO10.2 ribozyme could serve as a site-specific RNA labeling tool after further characterization. This requires mass spectrometry of reaction product, analysis of target sequence scope, and NTP substrate scope. Our preliminary characterization of the second ribozyme indicates that AR9 may have the potential to serve as a quadruplex reporter, as it catalyzes the RNA self-labeling reaction only in the presence of RNA G-quadruplex. Future work will investigate how AR9 recognizes the rG4 structure, and how the RNA-RNA interactions activate the ribozyme.

References

[1] aK. Kruger, P. J. Grabowski, A. J. Zaug, J. Sands, D. E. Gottschling, T. R. Cech, *cell* **1982**, *31*, 147-157; bC. Guerrier-Takada, K. Gardiner, T. Marsh, N. Pace, S. Altman, *Cell* **1983**, *35*, 849-857.

[2] aA. D. Ellington, J. W. Szostak, *nature* **1990**, *346*, 818-822; bC. Tuerk, L. Gold, *science* **1990**, *249*, 505-510; cD. L. Robertson, G. F. Joyce, *Nature* **1990**, *344*, 467-468.

[3] R. R. Breaker, G. F. Joyce, *Chemistry & biology* **1994**, *1*, 223-229.

[4] aS. W. Santoro, G. F. Joyce, *Proceedings of the national academy of sciences* **1997**, *94*, 4262-4266; bA. A. Fokina, D. A. Stetsenko, J.-C. François, *Expert Opinion on Biological Therapy* **2015**, *15*, 689-711; cJ. Li, Y. Lu, *Journal of the American Chemical Society* **2000**, *122*, 10466-10467.

[5] M. Hengesbach, M. Meusburger, F. Lyko, M. Helm, *Rna* **2008**, *14*, 180-187.

[6] M. Buchhaupt, C. Peifer, K.-D. Entian, *Analytical biochemistry* **2007**, *361*, 102-108.

[7] M. V. Sednev, V. Mykhailiuk, P. Choudhury, J. Halang, K. E. Sloan, M. T. Bohnsack, C. Höbartner, *Angewandte Chemie International Edition* **2018**, *57*, 15117-15121.

[8] S. Mao, P. Haruehanroengra, S. V. Ranganathan, F. Shen, T. J. Begley, J. Sheng, *ACS Chemical Biology* **2020**, *16*, 76-85.

[9] aR. Micura, W. Pils, C. Höbartner, K. Grubmayr, M.-O. Ebert, B. Jaun, *Nucleic acids research* **2001**, *29*, 3997-4005; bS. Mao, B. Sekula, M. Ruszkowski, S. V. Ranganathan, P. Haruehanroengra, Y. Wu, F. Shen, J. Sheng, *Nucleic acids research* **2020**, *48*, 10087-10100.

[10] aE. M. Harcourt, A. M. Kietrys, E. T. Kool, *Nature* **2017**, *541*, 339-346; bS. Wang, E. T. Kool, *Biochemistry* **1995**, *34*, 4125-4132.

[11] aS. Kobori, Y. Nomura, A. Miu, Y. Yokobayashi, *Nucleic acids research* **2015**, *43*, e85-e85; bS. Kobori, Y. Yokobayashi, *Angewandte Chemie International Edition* **2016**, *55*, 10354-10357; cV. Dhamodharan, S. Kobori, Y. Yokobayashi, *ACS chemical biology* **2017**, *12*, 2940-2945.

[12] aR. I. McDonald, J. P. Guilinger, S. Mukherji, E. A. Curtis, W. I. Lee, D. R. Liu, *Nature chemical biology* **2014**, *10*, 1049-1054; bA. K. Sharma, J. J. Plant, A. E. Rangel, K. N. Meek, A. J. Anamisis, J. Hollien, J. M. Heemstra, *ACS chemical biology* **2014**, *9*, 1680-1684.

[13] R. Welz, K. Bossmann, C. Klug, C. Schmidt, H. J. Fritz, S. Müller, *Angewandte Chemie International Edition* **2003**, *42*, 2424-2427.

[14] M. Ghaem Maghami, C. P. Scheitl, C. Höbartner, *Journal of the American Chemical Society* **2019**, *141*, 19546-19549.

[15] C. P. Scheitl, M. Ghaem Maghami, A.-K. Lenz, C. Höbartner, *Nature* **2020**, *587*, 663-667.

8. Annex

8.1 List of publications

I. Liaqat A, Stiller C, Michel M, Sednev MV, Höbartner C. N^6-Isopentenyladenosine in RNA Determines the Cleavage Site of Endonuclease Deoxyribozymes. *Angew Chem Int Ed Engl*. **2020**;59(42):18627-18631. doi:10.1002/anie.202006218

II. Liaqat A, Sednev MV, Stiller C, Höbartner C. RNA-Cleaving Deoxyribozymes Differentiate Methylated Cytidine Isomers in RNA. *Angew Chem Int Ed Engl*. **2021**;60(35):19058-19062. doi:10.1002/anie.202106517

III. Sednev MV, Liaqat A, Höbartner C. High-Throughput Activity Profiling of RNA-Cleaving DNA Catalysts by Deoxyribozyme Sequencing (DZ-seq). *J Am Chem Soc*. **2022**;144(5):2090-2094. doi:10.1021/jacs.1c12489

IV. Liaqat A, Höbartner C. In Vitro Evolution of Ribozymes for Recognition and Labeling of RNA G-quadruplex., **2022**, *manuscript in preparation*.

Publications not included in the thesis

I. Liaqat A, Sednev MV, Höbartner C. In Vitro Selection of Deoxyribozymes for the Detection of RNA Modifications. *Methods Mol Biol*. **2022**; 2533:167-179. doi:10.1007/978-1-0716-2501-9_10

8.2 Decalaration of authorship

Statement of individual author contributions and of legal second publication rights to

Manuscript 1: Liaqat A, Stiller C, Michel M, Sednev MV, Höbartner C. N^6 -Isopentenyladenosine in RNA Determines the Cleavage Site of Endonuclease Deoxyribozymes. *Angew Chem Int Ed Engl*. **2020**;59(42):18627-18631. doi:10.1002/anie.202006218					
Participated in	**Author Initials,** Responsibility decreasing from left to right				
Study Design Methods Development	CH, AL AL, MVS	CS, MVS CS, MM	 CH		
Data Collection	AL	CS	MVS	MM	
Data Analysis and Interpretation	AL, MVS	CH, CS			
Manuscript Writing Writing of Introduction Writing of Materials & Methods Writing of Discussion Writing of First Draft	 CH, AL AL, CS CH AL	 MVS AL, MVS			

Manuscript 2: Liaqat A, Sednev MV, Stiller C, Höbartner C. RNA-Cleaving Deoxyribozymes Differentiate Methylated Cytidine Isomers in RNA. *Angew Chem Int Ed Engl*. **2021**;60(35):19058-19062. doi:10.1002/anie.202106517					
Participated in	**Author Initials,** Responsibility decreasing from left to right				
Study Design Methods Development	CH, AL, MVS AL, MVS				
Data Collection	AL	MVS	CS		
Data Analysis and Interpretation	AL	MVS	CS	CH	
Manuscript Writing Writing of Introduction Writing of Materials & Methods Writing of Discussion Writing of First Draft	 MVS AL MVS MVS, AL	 CH CS CH	 AL MVS AL	 CH CS	

Manuscript 3: Sednev MV, Liaqat A, Höbartner C. High-Throughput Activity Profiling of RNA-Cleaving DNA Catalysts by Deoxyribozyme Sequencing (DZ-seq). *J Am Chem Soc*. **2022**;144(5):2090-2094. doi:10.1021/jacs.1c12489					
Participated in	**Author Initials,** Responsibility decreasing from left to right				
Study Design Methods Development	MVS, AL, CH MVS				
Data Collection	AL	MVS			
Data Analysis and Interpretation	MVS, AL				
Manuscript Writing Writing of Introduction Writing of Materials & Methods Writing of Discussion Writing of First Draft	 MVS MVS, AL MVS, CH, AL MVS, AL				

Manuscript 4: Liaqat A, Höbartner C. In Vitro Evolution of Ribozymes for Recognition and Labeling of RNA G-quadruplex, 2022, *manuscript in preparation*					
Participated in	**Author Initials,** Responsibility decreasing from left to right				
Study Design Methods Development	AL, CH AL				
Data Collection	AL				
Data Analysis and Interpretation	AL, CH				
Manuscript Writing Writing of Introduction Writing of Materials & Methods Writing of Discussion Writing of First Draft	 AL AL AL, CH AL				

Anam Liaqat

_____

Doctoral Researcher's Name Date Place Signature

Prof. Dr. Claudia Höbartner

_____

Primary Supervisor's Name Date Place Signature

Statement of individual author contributions to figures/tables of manuscripts included

Manuscript 1: Liaqat A, Stiller C, Michel M, Sednev MV, Höbartner C. N^6-Isopentenyladenosine in RNA Determines the Cleavage Site of Endonuclease Deoxyribozymes. *Angew Chem Int Ed Engl*. **2020**;59(42):18627-18631. doi:10.1002/anie.202006218					
Figure	**Author Initials,** Responsibility decreasing from left to right				
1	CH	AL			
2	AL				
3	AL	MVS	CS		
4	AL	MVS	CS		
5					
Table	**Author Initials,** Responsibility decreasing from left to right				
1					
2					

Manuscript 2: Liaqat A, Sednev MV, Stiller C, Höbartner C. RNA-Cleaving Deoxyribozymes Differentiate Methylated Cytidine Isomers in RNA. *Angew Chem Int Ed Engl*. **2021**;60(35):19058-19062. doi:10.1002/anie.202106517					
Figure	**Author Initials,** Responsibility decreasing from left to right				
1	MVS	AL	CH		
2	AL	MVS	CH		
3	AL	MVS	CH		
4	AL	MVS	CS		
5					
Table	**Author Initials,** Responsibility decreasing from left to right				
1					
2					

Manuscript 3 : Sednev MV, Liaqat A, Höbartner C. High-Throughput Activity Profiling of RNA-Cleaving DNA Catalysts by Deoxyribozyme Sequencing (DZ-seq). *J Am Chem Soc*. **2022**;144(5):2090-2094. doi:10.1021/jacs.1c12489					
Figure	**Author Initials,** Responsibility decreasing from left to right				
1	MVS				
2	MVS	AL	CH		
3	AL	MVS	CH		
4					
5					
Table	**Author Initials,** Responsibility decreasing from left to right				
1					
2					

Manuscript 4 : Liaqat A, Höbartner C. In Vitro Evolution of Ribozymes for Recognition and Labeling of RNA G-quadruplex, 2022, *manuscript in preparation*					
Figure	**Author Initials,** Responsibility decreasing from left to right				
1	AL				
2	AL				
3	AL				
4	AL				
5	AL				
6	AL				
Table	**Author Initials,** Responsibility decreasing from left to right				
1					
2					

Anam Liaqat

_____

Prof. Dr. Claudia Höbartner

_____

Date Place Signature

8.3 Affidavit

I hereby confirm that my thesis entitled **Artificial evolution of nucleic acid catalysts and their use for studying RNA** is the result of my own work. I did not receive any help or support from commercial consultants. All sources and/or materials applied are listed and specified in the thesis.

Furthermore, I confirm that this thesis has not yet been submitted as part of another examination process neither in identical nor in similar form.

Würzburg, 15.07.2022

Place, Date Signature

Hiermit erkläre ich an Eides statt, Artifizielle **Evolution von katalytischen Nucleinsäuren und deren Anwendung für die Untersuchung von RNA** eigenständig, d.h. insbesondere selbstständig und ohne Hilfe eines kommerziellen Promotionsberaters, angefertigt und keine anderen als die von mir angegebenen Quellen und Hilfsmittel verwendet zu haben.

Ich erkläre außerdem, dass die Dissertation weder in gleicher noch in ähnlicher Form bereits in einem anderen Prüfungsverfahren vorgelegen hat.

Würzburg, 15.07.2022

Ort, Datum Unterschrift