RNase H-based analysis of synthetic mRNA 5' cap incorporation

Siu-Hong Chan*, Joseph M. Whipple, Nan Dai, Theresa M. Kelley, Kathryn Withers, George Tzertzinis, Ivan R Corrêa Jr, G. Brett Robb

New England Biolabs, 240 County Road, Ipswich, MA, 01938, USA

*Corresponding author:

Siu-Hong Chan

E. mail: chan@neb.com

Running title:

RNA 5' cap analysis

Keywords:

Synthetic mRNA, 5' cap, RNA capping, RNA cap analysis, RNase H, LC-MS

Abstract

Advances in mRNA synthesis and lipid nanoparticles technologies have helped make mRNA therapeutics and vaccines a reality. The 5' cap structure is a crucial modification required to functionalize synthetic mRNA for efficient protein translation *in vivo* and evasion of cellular innate immune responses. The extent of 5' capping is one of the critical quality attributes in mRNA manufacturing. RNA cap analysis involves multiple steps: generation of pre-defined short fragments from the 5' end of the kilobase-long synthetic mRNA molecules using RNase H or a DNAzyme, 5' cleavage products enrichment, and LC-MS intact mass analysis. In this communication, we describe 1) our results on the sequence preference of RNase H cleavage and a framework to design site-specific RNA cleavage; 2) a method to fluorescently label the RNase H cleavage fragments for readouts such as easily accessible urea-PAGE or high throughput capillary electrophoresis; 3) a simplified method for post-RNase H purification using desthiobiotinylated oligonucleotides and streptavidin magnetic beads followed by elution using water. By providing a design framework for RNase H-based RNA 5' cap analysis using less resource-intensive analytical methods, we hope to make RNA cap analysis more accessible to the scientific community.

Introduction

mRNA synthesis and nanoparticle technologies have advanced to a stage where mRNA therapeutics and vaccines have become a reality with global public health impact. The relatively short development and manufacturing time have been the main attributes propelling the rapid deployment of effective mRNA vaccines in a global health emergency such as COVID-19.

In eukaryotic cells, the 5' end of messenger RNA (mRNA) is characterized by a cap structure that plays key roles in many biological functions (1). Most relevant to mRNA as therapeutics or vaccines is the cap requirement for efficient protein translation (1–3), evading innate immune surveillance (4–6) and regulation of 5'-mediated decay (7, 8). Hence, a 5' cap structure is crucial for efficacy and safety of mRNA therapeutics (9) and vaccines (10, 11).

The 5' cap consists of a *N*7-methylated guanosine linked to the 5' end of RNA through a 5' to 5' triphosphate group. In metazoans, the m^7GpppN cap (Cap-0) is further modified by a cap-specific 2'-*O*-methyltransferase (MTase) to the m^7GpppNm structure (Cap-1) (Figure 1).

Currently, the cap structure can be incorporated into an *in vitro* transcript by two means: 1) incorporation of a chemically synthesized cap structure such as an anti-reverse cap analog (ARCA) or capped dinucleotides (12) during transcription, or 2) post-transcriptional capping using a mRNA capping enzyme. These approaches have well-established methods that are readily adaptable and scalable (13, 14). It is, however, less trivial to verify the identity of the cap and evaluate the extent of capping on kilobase-long synthetic mRNA transcripts. One of the most common methods for evaluating the efficiency of capping is a cell-based functional assay that measures the level of protein expression from the synthetic mRNA *in vivo* through immunometric or enzymatic reactions. These readouts can provide information about the efficacy of the synthetic mRNA translation, but do not directly reflect the extent of capping of a synthetic mRNA preparation.

The complexity of RNA cap analysis varies according to the method of cap incorporation. When a cap analog or a capped dinucleotide is used in co-transcriptional capping, the relevant RNA species in the process are Cap-1 or Cap-0 (depending on the choice of the cap analogue reagent), unreacted RNA 5'-triphosphate, and small amounts of 5'-diphosphate byproduct. For enzymatic capping, the nature of the catalytic pathway (15) requires the resolution and quantitation of Cap-1 or Cap-0 (depending on the use or not of a 2'-*O*-MTase) from capping intermediates, such as unmethylated G-cap and 5'-diphosphate RNA, and unreacted 5'-triphosphate RNAs (Figure 1).

The addition of a cap structure to a mRNA molecule amounts to a less-than 600 Da mass change over potentially hundreds of kilodaltons in mass. Analytical methods such as gel-electrophoresis and mass spectrometry may lack the resolution or detection range to confidently distinguish between Cap-1 or Cap-0 capped RNA and unreacted or intermediate capping products. To overcome this problem, methods that cleave a pre-defined fragment from the 5' end of the mRNA molecule using custom designed DNAzyme (16) or DNA-RNA chimera-guided RNase H have been developed (17–19). The 5' cleavage fragment, ranging from 5 to 30 nucleotides long, can be readily analyzed by denaturing gel electrophoresis or tandem liquid chromatography-mass spectrometry (LC-MS)(20).

In the cell, RNase H (RNase H1) is an endonuclease that removes the RNA primers from Okazaki fragments of the replicating DNA and processes R-Loops to modulate R-Loop-mediated biological processes, such as gene expression, DNA replication and DNA and histone modifications (21–23). It has been shown that *E. coli* RNase H can be constrained to cleaving ssRNA at specific sites *in vitro* using a DNA-RNA chimera. However, *E. coli* RNase H may also cleave one or more nucleotides away from the 5' or 3' of the target site, giving rise to multiple cleavage products different by 1 nt (17–19). In addition, it is not clear if the sequence of cleavage site plays a role in determining the RNase H cleavage specificity. For the analysis of RNA 5' capping, which is essentially the addition of one nucleotide at the 5' triphosphate group, multiple cleavage products of a few nucleotides difference in length can make it impossible to analyze by mobility-based methods such as gel- or capillary electrophoresis. As with other complementarity-based approaches,

efficient RNase H cleavage depends on factors such as overcoming the secondary structures of the analyte RNA molecules and annealing of DNA-RNA chimera to the pre-defined cleavage site under given reaction conditions. Although these factors are often not well-defined, a framework for designing precise RNase H cleavage is desired.

Previously it has been shown that using a biotinylated DNA-RNA chimera oligo, the 5' RNase H cleavage fragments could be readily isolated for qualitative and quantitative analyses using LC-MS (20). However, there has been insufficient guidance on the DNA-RNA chimera design to achieve uniform RNase H cleavage, and the readout method is limited to resource-intensive LC-MS. In this communication, by investigating the sequence context and cleavage position relative to the DNA-RNA chimera, we showed that RNase H can generate relatively uniform cleavage at pre-defined sites. We further simplified the post-RNase H enrichment method and showed that the 3' end of the RNase H cleavage fragments can be fluorescently labeled by the fill-in activity of Klenow fragment. Because the fill-in activity is dependent on the complementarity of the incoming fluorescently labeled dNTPs to the DNA-RNA chimera, the labeling step further constraints the size of the fluorescently labeled RNase H cleavage product to effectively eliminate the non-target cleavage products from the analysis. Therefore, the Klenow fill-in and fluorescently labeled RNase H cleavage products can be done using readily accessible methods such as urea-PAGE or high throughput methods such as capillary electrophoresis. By providing a framework for RNase H-based cap analysis design, we hope to make RNA cap analysis more accessible to the scientific community.

Methods

In vitro transcription

In vitro transcription of firefly luciferase (*FLuc*) and Cypridina Luciferase (C*Luc*) transcripts containing artificial 5' UTRs were performed using PCR-amplified templates and the HiScribe T7 High Yield RNA

Synthesis Kit (New England Biolabs) according to manufacturer's instructions. The RNA transcripts were purified using the Monarch RNA Cleanup Kit (New England Biolabs). RNA concentrations were determined using Qubit RNA BR Assay Kits (ThermoFisher).

RNA capping reaction

Purified *in vitro* transcripts or synthetic RNA oligonucleotides were capped using Vaccinia RNA capping enzyme (Vaccinia Capping System, New England Biolabs) according to manufacturer's instructions. In some cases, a lower concentration of Vaccinia RNA capping enzyme was used to generate partially capped RNA to demonstrate the resolution of reaction intermediates by the analytical methods.

RNase H reactions

DNA-RNA chimeras (targeting oligo or TO) composed of 5' deoxynucleotides and 3' ribonucleotides with or without a 3' TEG-desthiobiotin group were designed as described in the main text and synthesized chemically (Integrated DNA Technologies or Bio-Synthesis, Inc). For the investigation of RNase H cleavage specificity, 0.5 μM of synthetic RNA oligonucleotides containing a 5' FAM-label were combined with 2.5 μM of the corresponding TOs in a 10 μL reaction containing 1x RNase H reaction buffer (50 mM Tris-HCl, pH 8.3, 75 mM KCl, 3 mM $MgCl_2$, 10 mM DTT). The mixtures were heated at 80°C for 30 s then cooled down to 25°C at a rate of 0.1°C/s. The mixtures were then incubated with *E. coli* RNase H (New England Biolabs) or *T. thermophilus* RNase H (Thermostable RNase H, New England Biolabs) at a final concentration of 5 U/μL at 37°C for 1 h. Reactions were quenched by addition of EDTA to a final concentration of 10 mM. The quenched RNase H reactions were then subjected to 15% urea-PAGE followed by fluorescent imaging using the Typhoon RGB laser scanner (Cytiva). Alternatively, quenched reactions were analyzed by LC-MS directly.

For cap analysis of *in vitro* transcripts, RNA samples (0.5 μM) were heated at 80°C for 30 s in the presence of 2.5 μM of appropriate targeting oligos in a 10 μL reaction containing 1x RNase H reaction buffer or 1x RNA capping buffer (50 mM Tris-HCl, pH 8.0, 5 mM KCl, 1 mM $MgCl_2$, 1 mM DTT). After cooling down to 25°C at a rate of 0.1°C/s, the reactions were subjected to RNase H cleavage by incubation with Thermostable RNase H (New England Biolabs) at a final concentration of 0.5 U/μL at 37°C for 1 h.

Klenow fill-in and urea-PAGE analysis

The targeting oligos were designed to remain hybridized to the 5' cleavage fragment of the *FLuc* transcript with a 3' 1-base recessed end after RNase H cleavage. Because RNase H cleavage results in a hydroxyl group at the 3' end (24), the recessed 3' end was filled in using a fluorescently labeled deoxynucleotide complementary to the 5' most deoxynucleotide of the TO using the DNA Polymerase I Large (Klenow) Fragment (New England Biolabs). Briefly, 5 μL of the RNase H reaction was retrieved and added to 5 μL of 2x Klenow reaction mix that contained 2x NEBuffer 2, 0.1 mM FAM-12-dCTP (Perkin Elmer) and 0.5 U/μL DNA Polymerase I Large (Klenow) Fragment. The reactions were incubated at 37°C for 1 hour. In cases where a significant fraction of a 2-base recessive end was generated, an unlabeled complementary deoxynucleoside triphosphate was included to improve the yield of labeled cleavage products.

For urea-PAGE analysis, 2 μL of the Klenow reactions were retrieved and added to 8 μL of 2x RNA Loading Dye (New England Biolabs). The mixtures were then analyzed by electrophoresis through a urea-15% polyacrylamide gel (Novex TBE-Urea gel 15%, Thermo Fisher Scientific). To acquire the fluorescent signals, the urea gels were scanned directly using a Amersham Typhoon RGB scanner (Cytiva) with the Cy2 channel. To acquire total RNA stain, the gels were stained using SYBR Gold (Thermo Fisher Scientific) and scanned using the Typhoon RGB scanner using the same settings. Data was analyzed using ImageQuant (Cytiva).

Capillary electrophoresis

Capillary electrophoresis of FAM-labeled RNA was carried out using an Applied Biosystems 3730xl Genetic Analyzer (96 capillary array) using the GeneScan® 120 LIZ dye Size Standard (Applied Biosystems) (25, 26). Peak detection and quantification were performed using the Peak Scanner software v.1.0 (Thermo Fisher Scientific) and an in-house data analysis suite. Details of method validation and sample electropherograms can be found in Supplemental Figure 2.

Purification of RNase H cleavage products

After RNase H cleavage and the optional Klenow fill-in, the 5' cleavage fragment/TO duplex was purified by size selection followed by affinity-purification (Figure 2). Briefly, 45 µL of nuclease-free water was added to 5 µL of RNase H cleavage reactions. The mixture was then added to 100 µL of NEBNext® Sample Purification Magnetic Beads (New England Biolabs) and incubated at room temperature for 5 min. The beads were then placed on a magnet for 2 min at room temperature. The clarified supernatant was retrieved and added to pre-cleared beads derived from 100 µL of NEBNext® Sample Purification Magnetic Beads. The beads were again placed on a magnet for 2 min at room temperature. The clarified supernatant was retrieved and added to pre-cleared beads derived from 50 µL of Dynabeads® MyOne Streptavidin C1 (Thermo Fisher Scientific). For the evaluation of biotin and water elution fractions, after washing with 200 µL of standard streptavidin beads wash buffer (5 mM Tris-HCl, pH 7.5, 1 M NaCl, 0.5 mM EDTA), the slurry was divided into two equal fractions. The first fraction was washed three more times with 100 µL of the standard streptavidin beads wash buffer, and the bound RNA was eluted by incubating the clarified beads in 10 µL of 0.1 M biotin solution at 37°C for 1 h. The other slurry fraction was washed three more times with 100 µL of a low salt wash buffer (5 mM Tris, pH 7.5, 0.5 mM EDTA, 60 mM NaCl). The bound RNA was eluted by incubating the clarified beads in 10 µL of nuclease-free water at 65°C for 5 min (27). Although higher molecular weight RNA was present when the size-selection step was omitted, the data quality of the LC-MS analysis was not affected (data not shown). Hence, the size-selection step was

considered optional. For cap analysis, the streptavidin magnetic beads with bound 5' cleavage fragment/TO duplexes were washed 4 times with 100 µL low salt wash buffer. Elution was carried out by incubating the washed beads in 10 µL of nuclease-free water at 65°C for 5 min.

LC-MS intact mass analysis

Intact oligonucleotide mass analyses were done using in-house facilities or by external contractor Novatia LLC. For in-house analyses, intact oligonucleotide mass analyses were performed by liquid chromatography-tandem mass spectrometry (LC-MS) on a Vanquish Horizon UHPLC equipped with a diode array detector and a Thermo Scientific Q-Exactive Plus orbitrap mass spectrometer operating under negative electrospray ionization mode (–ESI). UHPLC was performed using a Thermo DNAPac™ RP Column (2.1 × 50 mm, 4 µm) at 70 °C and 0.3 mL/min flow rate with a gradient mobile phase consisting of a hexafluoroisopropanol (HFIP) and *N,N*-diisopropylethylamine (DIEA) aqueous buffer and methanol. UV detection was performed at 260 nm. Intact mass analysis was performed under full scan mode at a resolution of 70,000 (FWHM) at *m/z* 200. ESI-MS raw data was deconvoluted using Promass HR (Novatia, LLC). The relative abundance of each deconvoluted mass peak was used to calculate the percentage of the substrate (ppp-), intermediates (pp-, Gppp-) and final products (m^7Gppp- or m^7GpppNm-) of enzymatic RNA capping. Methods and results for validation of deconvoluted mass peak intensities for RNA quantitation can be found in Supplemental Figure 3.

Results and Discussions

DNA-RNA chimera-guided RNase H cleavage at pre-selected site

By hybridizing the target RNA with a DNA-RNA chimera and generating an RNA/DNA duplex like its physiological substrate, *E. coli* RNase H has been shown to cleave ssRNA at specific sites *in vitro*. Two major cleavage sites, the adjacent phosphodiester bonds 5' and 3' to the ribonucleotide hybridized to the 5'

deoxynucleotide of the DNA-RNA chimera have been reported (Figure 3A) (17–19). Understanding the sequence and positional preference of RNase H cleavage could help guide the selection of cleavage sites and the design of the DNA-RNA chimera.

We compared two commercially available RNase H enzymes (from *E. coli* and from *Thermus thermophilus*) for their efficiency and specificity in cleavage site preference on an synthetic RNA oligonucleotide containing an artificial sequence (synfluc-AC). To maximize the availability of cleavage sites, we designed the DNA-RNA chimera (Targeting Oligos; TOs) to direct RNase H to cleavage within a loop region of the predicted secondary structure (Suppl. Figure 1). Four synthetic RNA oligonucleotides were designed with a variable nucleotide at the RNase H cleavage site. The substrate RNA oligonucleotides had a 5' FAM group such that the cleavage products can be analyzed using urea-PAGE electrophoresis and quantified using a flatbed laser scanner (Figure 3B).

As shown in Figure 3C, using the same quantity of enzyme under identical reaction conditions (37°C for 1 h), *T. thermophilus* (*Tth*) RNase H achieved higher levels of cleavage (80-90%) than the *E. coli* counterpart (10-40%) on all four RNA samples (lower left panel, Figure 3C). With respect to cleavage site preference, both RNase H enzymes generated >80% of the predicted cleavage product (24 nt) on cleavage sites containing A|C and A|U and 50-60% of the predicted cleavage product on A|A and A|G sites (lower right panel, Figure 3C).

To further investigate the effect of the nucleotide sequence on the position of *Tth* RNase H cleavage, multiple TOs (*FLuc*TOs) were designed and tested against the corresponding RNA substrates (syn*FLuc*) with a variable nucleotide downstream of the A|C site (Figure 4A). The RNase H cleavage events were evaluated using LC-MS intact mass analysis. Figure 4 shows the summary of the cleavage analysis. About 87% of the fragments generated from the pair syn*FLuc*-AC/*FLuc*TO-AC were products of cleavage at the designated site (A|C). About 11% of the cleavage events took place at the -1 position (A|A), and approximately 2% at the +1 position (C|U). This is consistent with the results derived from urea-PAGE (Figure 3C).

When the TO was shortened by 1 deoxynucleotide (flucTO-AC_-1), instead of at the -1 (A|A) site, the majority of the cleavage events on synfluc-AC still took place at the A|C site, albeit with a lower frequency (67%) than that seen with flucTO-AC. The remaining cleavage events took place at the -1 (A|A) site (27%) and at the -2 (U|A) site (6%). Consistently, when the TO was extended by 1 nt (flucTO-AC_+1), most cleavage events took place at the A|C site. The A|C site, now at -2 position, was cleaved 70% of the time, with the -1 site, now C|U, only cleaved at a 1% rate. Upon extending the TO to 2 nt (flucTO-AC_+2), most cleavage events took place at the C|U site (98%) at the -2 position. No cleavage event could be detected at the -1 (U|U) site. With the TO extended by 3 nt (flucTO-AC_+3), the major cleavage site remained at the -2 position but in context of a U|U sequence (98%). The -1 (U|U) site was cleaved at a low frequency (2%). By sampling the substrate loop region through TO designs, relatively uniform cleavage can be achieved by RNase H.

To further understand the influence of nucleotide sequence on RNase H cleavage specificity, we changed the ribonucleotide immediately 3' of the A|C site from uridine to cytosine (synfluc-ACC). Consistent with the sequence preference for A|C, 84% of the cleavage events took place at the -2 (A|C) site. Only 6% took place at the -1 (C|C) site. Extending the TO by 1 nt (flucTO-ACC_+1) resulted in a 60% cleavage at the -2 (C|C) site and 23% cleavage at the -3 (A|C) site. The -1 site (C|U) was only cleaved at 13%. Extending the TO further by 2 nt from the A|C site (flucTO-ACC_+2) resulted in 63% cleavage at -1 (U|U) site and 37% at the -2 (C|U) site.

When the ribonucleotide immediately 3' to the A|C site is changed to adenosine, no cleavage was detected at the -1 (C|A) site or the +1 (A|U) site. Instead, 93% of *Tth* RNase H cleavage took place at the -2 (A|C) site. Extending the TO by 1 or 2 nt (flucTO-ACA_+1 and *FLuc*TO-ACA_+2, respectively) resulted in major cleavage at the A|U site, irrespective of which TO was used. When the ribonucleotide immediately 3' to the A|C site was changed to guanosine (synfluc-ACG), no cleavage could be detected at the -1 (C|G) site or at the +1 (G|U) site. The A|C site, now at the -2 position remained the major cleavage site. Interestingly, extending the TO by 1 nt (flucTO-ACG_+1) resulted in redistribution of the cleavage events

to 47% at the -1 (G|U) site and 37% at the +1 (U|U) site. Further extending the TO by 2 nt (flucTO-ACG_+2) resulted in 90% cleavage at the -1 (U|U) site. Therefore, A|C and A|U and U|U sites are the preferred cleavage sites within the sequence context and secondary structure of our synthetic substrate RNAs.

To verify the applicability of the above findings to long RNA transcripts, a 1.7 kb firefly luciferase (*FLuc*) *in vitro* transcript containing the synfluc-AC sequence at the 5' end was cleaved by *Tth* RNase H in conjunction with TO-1 (Figure 5A). The 5' cleavage product was purified using the water elution method described below and analyzed by LC-MS intact mass analysis. Consistent with the findings with the short RNA oligonucleotides, the median cleavage frequency at the -1 (A|C) site was 91%, with 8% at the -2 (A|A) site, and 2% at the +1 (C|U) site (Fig. 5B), from eight replicated experiments.

Therefore, a careful targeting oligo design coupled with *Tth* RNase H can result in highly uniform digestion fragments that can help simplify downstream analysis.

Efficient enrichment of RNase H cleavage products

To enrich the RNase H 5' cleavage fragment for downstream analysis, it has been described that the cleavage fragment could be released from a biotinylated targeting oligo by heating in a methanol solution (20). We sought to develop a simplified method using desthiobiotinylated targeting oligos and a milder elution method. To this end, we used a firefly luciferase (*FLuc*) *in vitro* transcript as a model to test our enrichment method (Figure 6). First, the long 3' cleavage product and the uncleaved full-length transcript were separated from the smaller 5' cleavage products and TO (TO-1; see Figure 6) by a size selection. Briefly, the RNase H treated sample (or the corresponding Klenow fill-in products; see below) were added to the NEBNext® Sample Purification Magnetic Beads. The unbound fraction containing smaller RNAs (ub1, Figure 6) was added to fresh beads for further enrichment. The subsequent unbound fraction (ub2, Figure 6), which contained only a small fraction of the larger RNA molecules, was then subjected to affinity enrichment using streptavidin magnetic beads. Two elution methods were compared: biotin in aqueous

solution versus water (27). The RNA-bound streptavidin magnetic beads were divided into two equal fractions. The first fraction was washed with a standard streptavidin-bead washing buffer (1 M NaCl) and elution was carried out by incubating the clarified beads with a 0.1 M biotin solution at 37°C for 1 h. The second fraction of RNA-bound streptavidin magnetic beads was washed with a low salt buffer (60 mM NaCl) and elution was carried out by incubating the clarified beads with nuclease-free water at 65°C for 5 min. Elution with either water or biotin solution recovered similar amounts of the TO-1 and the RNase H 5' cleavage products (Figure 6 and Supplemental Figure 4). For simplicity, a low salt wash buffer and elution with water was chosen to enrich desthiobiotin-tagged TO-RNase H cleavage product duplexes in subsequent experiments. A sample deconvoluted mass spectrum of the 5' cleavage fragment of a partially capped Cap-1 *CLuc* transcript purified by the above method is shown in Supplemental Figure. 5.

Upon further investigation, we found that although higher molecular weight RNA was present when the size-selection step was not performed, the data quality of the LC-MS analysis was not affected (data not shown). Hence, the size-selection step can be considered optional.

Fluorescent labeling of RNase H 5' cleavage products using Klenow fragment

Our cleavage site analysis showed that TO can be designed to generate a pre-defined 3' recessive end on the RNase H 5' cleavage product. Because RNase H cleavage results in a hydroxyl group at the 3' end (24), this 3' recessive end is amenable to labeling with FAM-modified deoxyribonucleotides using the Klenow fragment of *E. coli* DNA polymerase I (28, 29). Because the fill-in activity is dependent on the complementarity of the incoming fluorescently labeled dNTPs to the DNA-RNA chimera, the labeling step further constraints the size of the fluorescently labeled RNase H cleavage product to effectively eliminate the non-target cleavage products from the analysis. The filled-in 5' fragments can then be analyzed using urea-PAGE or capillary electrophoresis (25, 26), in addition to LC-MS.

As an example, a *FLuc* transcript capped using VCE was analyzed by treatment with RNase H followed by Klenow fill-in using FAM-labeled dCTP, as described in Materials and Methods (Figure 7A). Figure 7B showed that Klenow fill-in enabled visualization and quantification of the extent of 5' capping using urea PAGE. Compared to total RNA staining (SYBR Gold), the fluorescent FAM signal offers a higher signal-to-noise ratio and eliminates the interference from the TO in the data analysis. A small quantity of +1 nt cleavage product is visible in the FAM-signal image (Figure 7B, right panel). LC-MS analysis indicated that the +1 nt is probably derived from the +1 G transcript resulted from the slippage at the GGG trinucleotide at the 5' end of the Fluc transcript (data not shown) (30).

Urea-PAGE is a quick and easily implementable readout for 5' RNA capping, but it does not resolve 5' triphosphate (ppp-) from 5' diphosphate (pp-), or m^7G cap (Cap-0; m^7Gppp-) from unmethylated G cap (Gppp-). Rather, 5' triphosphate and 5' diphosphate co-migrate as uncapped RNA, while m^7G cap and unmethylated G cap co-migrate as capped RNA. An affinity-based denaturing PAGE method that can resolve m7G cap (Cap-0) from unmethylated G cap (Gppp-) and uncapped (ppp- and pp-) RNA has been described (31). Capillary electrophoresis, on the other hand, resolves nucleic acids by size and charge. We showed that capillary electrophoresis can be adapted to resolve and quantify the four 5' end structures on short RNA oligonucleotides (26) (Supplemental Figure 2). To show that the capillary electrophoresis platform is applicable to cap analysis of long IVT RNA molecules, a *FLuc* transcript partially capped by VCE was processed by the RNase H/Klenow fill-in method and enriched as described above (to remove most of the free FAM-dCTP). Capillary electrophoresis of the sample allowed quantification of 5' ppp-, pp-, Gppp- and m^7Gppp- (Figure 7C). The capability to resolve capping intermediates is of great interest for method development and quality control in platforms that utilize enzymatic capping for mRNA production.

To further validate the Klenow fill-in method, a partially capped *FLuc* preparation was subjected to the RNase H/Klenow fill-in processing and analyzed by urea-PAGE, capillary electrophoresis, and LC-MS (Figure 7D). We found that cap analysis using capillary electrophoresis and LC-MS intact mass analysis

prodyced highly comparable results. Although urea-PAGE cannot individually resolve m^7Gppp- from Gppp-, or pp- from ppp-, the percentage of capped (m^7Gppp- and Gppp-) and uncapped (pp- and ppp-) agreed well with the capillary electrophoresis and LC-MS results.

The effect of uridine modifications on RNase H cleavage

Uridine modifications (e.g., pseudouridine, 1-methylpseudouridine and 5-methoxyuridine) have been shown to help reduce cellular innate immune response and to improve protein translation of synthetic mRNA (32–34). Modified uridine has been widely used in mRNA vaccines, such as the FDA-approved SARS-CoV-2 vaccines manufactured by Pfizer/BioNTech and Moderna (35). Pseudouridine is known to exhibit different base-pairing properties compared to uridine (36, 37). Such differences can affect annealing of targeting oligos and/or RNase H cleavage.

To investigate the effect of pseudouridine (Ψ) on RNase H cleavage site preferences, a Cypridina luciferase (*CLuc*) transcript (1.8 kb) was transcribed *in vitro* in the presence of either UTP or ΨTP. A targeting oligo *CLuc*TO-26 was designed such that the RNase H cleavage site was restricted to a uridine residue. Upon *Tth* RNase H cleavage in conjunction with CLucTO-26, the 5' cleavage product of the uridine and the Ψ-substituted *CLuc* transcript were subjected to the enrichment and analysis as described above (Figure 8). Two major cleavage events were observed. 73% of the products resulted from cleavage 5' of the uridine (C|UC) and had a length of 25 nt, while 27% resulted from cleavage 3' of the uridine (CU|C) and had a length of 26 nt (Figure 8). When uridine residues were substituted with pseudouridine, cleavage events 5' of the pseudouridine (C|ΨC) dropped to a median frequency of 34%, with 66% of cleavage taking place 3' of the pseudouridine (CΨ|C). Hence, pseudouridine substitution can alter the cleavage specificity of RNase H through mechanisms possibly related to base-pairing with the TO and/or RNase H substrate binding.

Messenger RNA, as a single-stranded nucleic acid molecule, can adapt multiple stable conformations depending on formulation and reaction conditions. These stable conformers can affect RNase H or DNAzyme cleavage efficiency and specificity through factors such as substrate-TO/DNAzyme annealing and substrate-enzyme interactions. While these factors are still poorly understood, we have shown that RNase H cleavage is dependent on (1) the sequence and (2) the position of the cleavage site relative to the 5' end of the DNA-RNA chimera, and on (3) uridine substitution. By providing a framework to investigate RNase H cleavage specificity on a given *in vitro* transcript and a fluorescent labeling strategy for cleavage products, we hope to make RNA cap analysis more accessible to the scientific community.

Data Availability

Additional data is available upon request.

Funding

This work is supported by New England Biolabs, Inc.

Acknowledgements

We would like to thank Daniel Kneller for critical review of the manuscript. We are also grateful to the NEB sequencing core for the analysis of capillary electrophoresis samples as well as all the facilities personnel for daily support of operation.

Conflicts of interest

S.-H. Chan, J.M. Whipple, N. Dai, G. Tzertzinis, I.R. Corrêa Jr. and G.B. Robb are current employees of New England Biolabs that markets molecular biology reagents. The affiliation does not affect the authors' impartiality, adherence to journal standards and policies, or availability of data.

References

1. Gonatopoulos-Pournatzis,T. and Cowling,V.H.H. (2014) Cap-binding complex (CBC). *Biochem. J.*, **457**, 231–42.

2. Maquat,L.E., Tarn,W.-Y. and Isken,O. (2010) The Pioneer Round of Translation: Features and Functions. *Cell*, **142**, 368–374.

3. Goodfellow,I.G. and Roberts,L.O. (2008) Eukaryotic initiation factor 4E. *Int. J. Biochem. Cell Biol.*, **40**, 2675–80.

4. Hyde,J.L. and Diamond,M.S. (2015) Innate immune restriction and antagonism of viral RNA lacking 2′-O methylation. *Virology*, **479**–**480**, 66–74.

5. Devarkar,S.C., Wang,C., Miller,M.T., Ramanathan,A., Jiang,F., Khan,A.G., Patel,S.S. and Marcotrigiano,J. (2016) Structural basis for m7G recognition and 2'-O-methyl discrimination in capped RNAs by the innate immune receptor RIG-I. *Proc. Natl. Acad. Sci. U. S. A.*, **113**, 596–601.

6. Kumar,P., Sweeney,T.R., Skabkin,M.A., Skabkina,O. V, Hellen,C.U.T. and Pestova,T. V (2014) Inhibition of translation by IFIT family members is determined by their ability to interact selectively with the 5'-terminal regions of cap0-, cap1- and 5'ppp- mRNAs. *Nucleic Acids Res.*, **42**, 3228–45.

7. Mugridge,J.S., Tibble,R.W., Ziemniak,M., Jemielity,J. and Gross,J.D. Structure of the activated Edc1-Dcp1-Dcp2-Edc3 mRNA decapping complex with substrate analog poised for catalysis. 10.1038/s41467-018-03536-x.

8. Floor,S.N., Borja,M.S. and Gross,J.D. (2012) Interdomain dynamics and coactivation of the mRNA decapping enzyme Dcp2 are mediated by a gatekeeper tryptophan. *Proc Natl Acad Sci U S A*, **109**, 2872–2877.

9. Ramaswamy,S., Tonnu,N., Tachikawa,K., Limphong,P., Vega,J.B., Karmali,P.P., Chivukula,P. and Verma,I.M. (2017) Systemic delivery of factor IX messenger RNA for protein replacement therapy.

Proc. Natl. Acad. Sci. U. S. A., **114**, E1941–E1950.

10. Pardi,N., Hogan,M.J., Porter,F.W. and Weissman,D. (2018) mRNA vaccines — a new era in vaccinology. *Nat. Rev. Drug Discov.*, **17**, 261–279.

11. VanBlargan,L.A., Himansu,S., Foreman,B.M., Ebel,G.D., Pierson,T.C. and Diamond,M.S. (2018) An mRNA Vaccine Protects Mice against Multiple Tick-Transmitted Flavivirus Infections. *Cell Rep.*, **25**, 3382-3392.e3.

12. Henderson,J.M., Ujita,A., Hill,E., Yousif-Rosales,S., Smith,C., Ko,N., McReynolds,T., Cabral,C.R., Escamilla-Powers,J.R. and Houston,M.E. (2021) Cap 1 Messenger RNA Synthesis with Co-transcriptional CleanCap® Analog by In Vitro Transcription. *Curr. Protoc.*, **1**, 1–17.

13. Fuchs,A.-L., Neu,A. and Sprangers,R. (2016) A general method for rapid and cost-efficient large-scale production of 5' capped RNA. *RNA*, **22**, 1454–1466.

14. Vaidyanathan,S., Azizian,K.T., Haque,A.K.M.K.M.A., Henderson,J.M., Hendel,A., Shore,S., Antony,J.S., Hogrefe,R.I., Kormann,M.S.D., Porteus,M.H., *et al.* (2018) Uridine Depletion and Chemical Modification Increase Cas9 mRNA Activity and Reduce Immunogenicity without HPLC Purification. *Mol. Ther. - Nucleic Acids*, **12**, 530–542.

15. Ramanathan,A., Robb,G.B. and Chan,S.H. (2016) mRNA capping: Biological functions and applications. *Nucleic Acids Res.*, **44**, 7511–7526.

16. Cairns,M.J., King,A. and Sun,L.Q. (2003) Optimisation of the 10-23 DNAzyme-substrate pairing interactions enhanced RNA cleavage activity at purine-cytosine target sites. *Nucleic Acids Res.*, **31**, 2883–2889.

17. Lapham,J. and Crothers,D.M. (1996) RNase H cleavage for processing of in vitro transcribed RNA for NMR studies and RNA ligation. *RNA*, **2**, 289–296.

18. Yu,Y.T., Shu,M. Di and Steitz,J.A. (1997) A new method for detecting sites of 2′-O-methylation in

RNA molecules. *RNA*, **3**, 324–331.

19. Lapham,J., Yu,Y.T., Shu,M. Di, Steitz,J.A. and Crothers,D.M. (1997) The position of site-directed cleavage of RNA using RNase H and 2′-O-methyl oligonucleotides is dependent on the enzyme source. *RNA*, **3**, 950.

20. Beverly,M., Dell,A., Parmar,P. and Houghton,L. (2016) Label-free analysis of mRNA capping efficiency using RNase H probes and LC-MS. *Anal. Bioanal. Chem.*, 10.1007/s00216-016-9605-x.

21. Huang,L., Kim,Y., Turchi,J.J. and Bambara,R.A. (1994) Structure-specific cleavage of the RNA primer from Okazaki fragments by calf thymus RNase HI. *J. Biol. Chem.*, **269**, 25922–25927.

22. Broccoli,S., Rallu,F., Sanscartler,P., Cerritelli,S.M., Crouch,R.J. and Drolet,M. (2004) Effects of RNA polymerase modifications on transcription-induced negative supercoiling and associated R-loop formation. *Mol. Microbiol.*, **52**, 1769–1779.

23. Parajuli,S., Teasley,D.C., Murali,B., Jackson,J., Vindigni,A. and Stewart,S.A. (2017) Human ribonuclease H1 resolves R-loops and thereby enables progression of the DNA replication fork. *J. Biol. Chem.*, **292**, 15216–15224.

24. Lima,W.F. and Crooke,S.T. (1997) Cleavage of single strand RNA adjacent to RNA-DNA duplex regions by Escherichia coli RNase H1. *J. Biol. Chem.*, **272**, 27513–27516.

25. Greenough,L., Schermerhorn,K.M., Mazzola,L., Bybee,J., Rivizzigno,D., Cantin,E., Slatko,B.E. and Gardner,A.F. (2016) Adapting capillary gel electrophoresis as a sensitive, high-throughput method to accelerate characterization of nucleic acid metabolic enzymes. *Nucleic Acids Res.*, **44**, e15–e15.

26. Wulf,M.G., Buswell,J., Chan,S.H., Dai,N., Marks,K., Martin,E.R., Tzertzinis,G., Whipple,J.M., Corrêa,I.R. and Schildkraut,I. (2019) The yeast scavenger decapping enzyme DcpS and its application for in vitro RNA recapping. *Sci. Rep.*, **9**, 8594.

27. Holmberg,A., Blomstergren,A., Nord,O., Lukacs,M., Lundeberg,J. and Uhlén,M. (2005) The biotin-

streptavidin interaction can be reversibly broken using water at elevated temperatures. *Electrophoresis*, **26**, 501–10.

28. Huang,Z. and Szostak,J.W. (1996) A simple method for 3'-labeling of RNA. *Nucleic Acids Res.*, **24**, 4360–4361.

29. Sandin,P., Stengel,G., Ljungdahl,T., Börjesson,K., Macao,B. and Wilhelmsson,L.M. (2009) Highly efficient incorporation of the fluorescent nucleotide analogs tC and tCO by Klenow fragment. *Nucleic Acids Res.*, **37**, 3924–3933.

30. Pleiss,J.A., Derrick,M.L. and UHLENBECK,O.C. (1998) T7 RNA polymerase produces 5′ end heterogeneity during in vitro transcription from certain templates. *Rna*, **4**, 1313–1317.

31. Matts,J.A., Sytnikova,Y., Chirn,G.W., Igloi,G.L. and Lau,N.C. (2014) Small RNA library construction from minute biological samples. *Methods Mol. Biol.*, **1093**, 123–136.

32. Karikó,K., Muramatsu,H., Welsh,F.A., Ludwig,J., Kato,H., Akira,S. and Weissman,D. (2008) Incorporation of pseudouridine into mRNA yields superior nonimmunogenic vector with increased translational capacity and biological stability. *Mol. Ther.*, **16**, 1833–40.

33. Karikó,K., Muramatsu,H., Keller,J.M. and Weissman,D. (2012) Increased erythropoiesis in mice injected with submicrogram quantities of pseudouridine-containing mRNA encoding erythropoietin. *Mol. Ther.*, **20**, 948–53.

34. Svitkin,Y. V., Cheng,Y.M., Chakraborty,T., Presnyak,V., John,M. and Sonenberg,N. (2017) N1-methyl-pseudouridine in mRNA enhances translation through eIF2α-dependent and independent mechanisms by increasing ribosome density. *Nucleic Acids Res.*, **45**, 6023–6036.

35. Nance,K.D. and Meier,J.L. (2021) Modifications in an Emergency: The Role of N1-Methylpseudouridine in COVID-19 Vaccines. *ACS Cent. Sci.*, **7**, 748–756.

36. Kierzek,E., Malgowska,M., Lisowiec,J., Turner,D.H., Gdaniec,Z. and Kierzek,R. (2014) The

contribution of pseudouridine to stabilities and structure of RNAs. *Nucleic Acids Res.*, **42**, 3492–501.

37. Deb,I., Popenda,Ł., Sarzyńska,J., Małgowska,M., Lahiri,A., Gdaniec,Z. and Kierzek,R. (2019) Computational and NMR studies of RNA duplexes with an internal pseudouridine-adenosine base pair. *Sci. Rep.*, **9**, 1–13.

Figure legends

FIGURE 1. Enzymatic reactions involved in RNA 5' capping. RNA transcripts generated by *in vitro* transcription (usually done using T7 RNA polymerase or its variants) contains a triphosphate group at the 5' end (pppR-). The 5' triphosphate group can be converted to the Cap-1 structure (m^7GpppRm-) through four enzymatic reactions carried out by RNA capping enzymes such as vaccinia RNA capping enzyme (reactions 1 through 3) in conjunction with Vaccinia cap 2'-*O*-methyltransferase (reaction 4).

FIGURE 2. A general scheme of RNase H-based RNA cap analysis. A DNA-RNA or DNA-2'-*O*-methyl-ribonucleotide chimera is designed to be complementary to part of the 5' end of the target RNA molecules, such that the chimera stays annealed to the cleavage fragment after RNase H cleavage. The chimera (called Targeting oligo or TO in this communication) contains a 3'-desthiobiotin group. After denaturation and annealing, RNase H cleaves at a pre-defined site within the RNA-TO duplex and generate a 1-base recessive end at the 3' end of the cleaved RNA. Because RNase H cleavage results in a 3' hydroxyl group (24), this recessive 3' end can be filled in with a fluorescently-labeled deoxynucleotide using the Klenow fragment of DNA polymerase I. The fluorescently-labeled 5' cleavage fragment can be analyzed by denaturing PAGE directly without enrichment. The 5' duplex cleavage fragment can be enriched using

streptavidin magnetic beads. The enriched RNase H cleavage products can be analyzed by LC/MS or capillary electrophoresis (if filled in with a fluorescent deoxynucleotide).

FIGURE 3. Directed RNA cleavage with RNase H (A) Previously reported RNase H cleavage sites (17-19). **(B)** Synthetic RNA oligonucleotides with artificial sequences and corresponding Targeting Oligos (TOs) designed based on (A). RNase H cleavage sites are indicated by inverted triangles. The variable nucleotides of the RNA substrates are colored in red. Deoxynucleotides in TOs are colored in blue; Ribonucleotides are colored in green. The predicted loop region of the synthetic RNA oligonucleotides based on an RNA folding algorithm (RNAFold, University of Vienna) is indicated. **(C)** RNA cleavage using *E. coli* or *Thermus thermophilus* RNase H at 37°C. Both RNase H enzymes prefer A|C and A|U over A|G or A|A; the *Tth* enzyme is more efficient in cleavage in general.

FIGURE 4. *Tth* RNase H cleavage frequency with respect to sequence context and TO design as analyzed by LC-MS. Within the sequence context and predicted structure of the artificial sequences in this study, high cleavage frequency was achieved at A|C, A|U, C|U or U|U sites. Variable nucleotides in the RNA substrate are colored in red; phosphodiester bonds are shown as "-". Deoxynucleotides in the TOs are shown as grey bars; Ribonucleotides are shown as green dots. Numbers represent frequency of cleavage detected at corresponding phosphodiester bonds.

FIGURE 5. Uniform RNase H cleavage with designed targeting oligos (A) 5' sequence of a 1.7 kb *in vitro FLuc* transcript containing an artificial 5' UTR and the corresponding targeting oligo TO-1. The targeting oligo contains six deoxynucleotides (blue) at the 5' end followed by 19 ribonucleotides (green) and a desthiobiotin (DTB) group at the 3' end. The size of the expected RNase H cleavage products are

shown. **(B)** Frequency of cleavage events expressed as percentage of detected cleavage product using LC-MS. Median of cleavage frequencies was 8%, 91% and 1% at (A|A), (A|C) and (C|U), respectively.

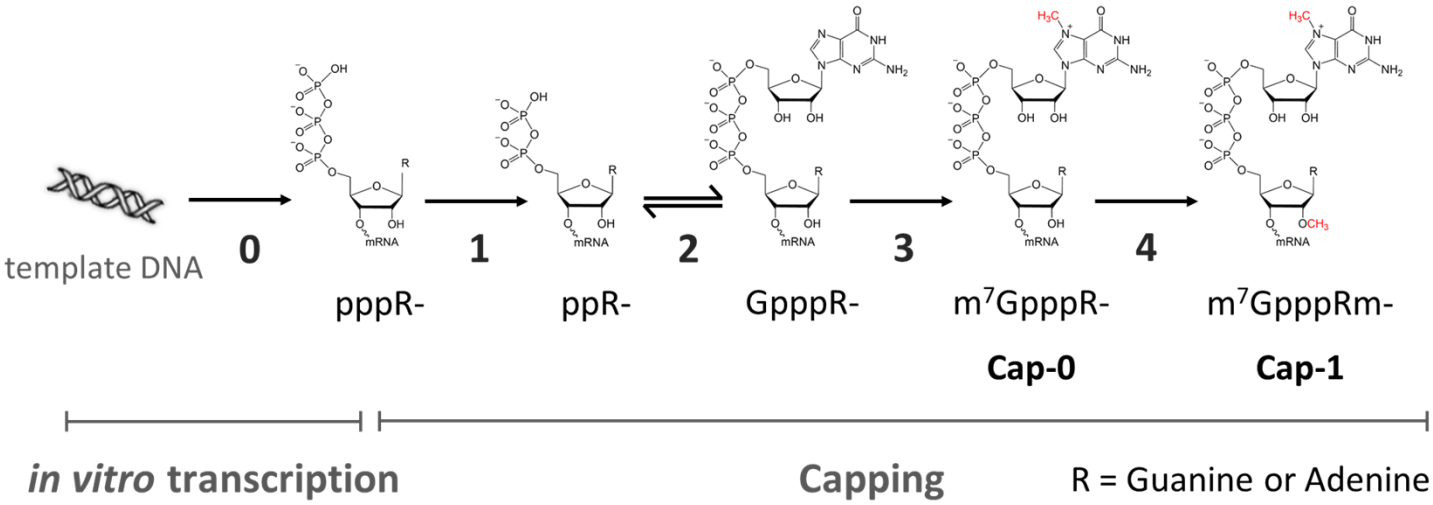
FIGURE 6. Enrichment of RNase H cleavage products by size and affinity selection. The RNase H cleavage products were first selected by size by two rounds of NEBNext magnetic beads. The clarified unbound fraction of the second round of size selection (ub2) was then added directly to streptavidin magnetic beads for affinity selection. After the first wash using a standard wash solution containing 1 M NaCl (W1), the resuspended bead slurry was divided into two fractions. One fraction was washed three more time using the standard wash buffer (W4_HS) and eluted using biotin (Eluate_biotin). The other fraction of the slurry was washed three more time using a low NaCl wash solution (W4_LS) followed by elution using the same volume of water (Eluate_water). Similar amount of RNase H cleavage product and TO were eluted (also see suppl. Figure. 4).

FIGURE 7. Fluorescent labeling of RNase H 5' cleavage products and analyses. (A) Schematic representation of targeting, cleavage, and labeling. The targeting oligo was designed to guide RNase H to generate a 1-nt 3' recessive end, which can be filled in by a fluorescently labeled deoxynucleotide using the Klenow fragment. In this example, a FAM-labeled dCTP, was incorporated to the 3' end of the *FLuc* RNase H cleavage fragment, which can be directly analyzed by denaturing PAGE (B) or capillary electrophoresis (C) after enrichment. **(B)** Polyacrylamide gel analysis of cleaved RNA fragments. The 1.7 kb *FLuc* transcript capped using Vaccinia RNA capping enzyme (VCE) was subjected to the RNase H/Klenow fill-in treatments. Reactions were analyzed directly by urea PAGE followed by laser scanning of total RNA stain using SYBR Gold (left panel) or fluorescent signal (right panel). The targeting oligo TO-1 is invisible when the gel was scanned using the FAM channel and did not interfere with quantitation of the 5' cleavage products. **(C)** Resolution and quantification of capping and capping intermediates using

capillary electrophoresis. The *FLuc* transcript was capped using a low concentration of VCE (10 nM) and subjected to the RNase H/Klenow fill-in reactions. After enrichment, the RNA was analyzed using capillary electrophoresis. In addition to substrate 5' triphosphate (ppp-) and the product m7Gppp-capped forms, enzymatic intermediate products 5' diphosphate (pp-) and the unmethyl-cap (Gppp-) can be resolved and quantified. **(D)** Synthetic mRNA cap analysis using gel electrophoresis, capillary electrophoresis, and LC-MS intact mass analysis produce comparable results. After RNase H/Klenow fill-in reactions and enrichment, an uncapped or partially capped *FLuc* transcript was analyzed using all three available methods. Capillary electrophoresis and LC-MS yielded comparable results in quantification of substrate (ppp-), product (m7Gppp-) and intermediate products (pp- and Gppp-). Urea-PAGE does not resolve pp- from ppp- or Gppp- from m7Gppp-. Considering ppp- and pp- as uncapped and Gppp- and m7Gppp- as capped species, quantitation of fluorescently labeled RNase H cleavage products using urea-PAGE generate results comparable to CE or LC-MS albeit the lack of resolution for intermediate products.

FIGURE 8. The effect of pseudouridine on RNase H cleavage. A *Cypridina* luciferase transcript (*CLuc*; 1.8 kb) was cleaved using *Tth* RNase H in conjunction with clucTO-26, which was designed to include a uridine residue in the RNase H cleavage site. The expected size of the cleavage products and cleavage sites are indicated. When unsubstituted, median cleavage frequency at the C|UC site was 73% (25 nt) and at the CU|C site was 27% (26 nt). When all uridines were substituted with pseudouridine, cleavage at the C|ΨC site decreased to a median frequency of 34%, with 66% cleavage at the CΨ|C site.

Fig. 1



Reactions	Enyme activity	Shorthand	Example enzymes
0	*in vitro* transcription	IVT	T7 RNA polymerase
1	RNA triphosphatase	TPase	Vaccinia RNA capping enzyme
2	RNA guanylyltransferase	GTase	
3	mRNA cap guanine-*N*7 methyltransferase	N7MTase	
4	Cap 2'-*O*-Methyltransferase	2'OMTase	Vaccinia cap 2'-*O*-methyltransferase

FIGURE 1. Enzymatic reactions involved in RNA 5' capping. RNA transcripts generated by *in vitro* transcription (usually done using T7 RNA polymerase or its variants) contains a triphosphate group at the 5' end (pppR-). The 5' triphosphate group can be converted to the Cap-1 structure (m7GpppRm-) through four enzymatic reactions carried out by RNA capping enzymes such as vaccinia RNA capping enzyme (reactions 1 through 3) in conjunction with Vaccinia cap 2'-*O*-methyltransferase (reaction 4).

Fig. 2

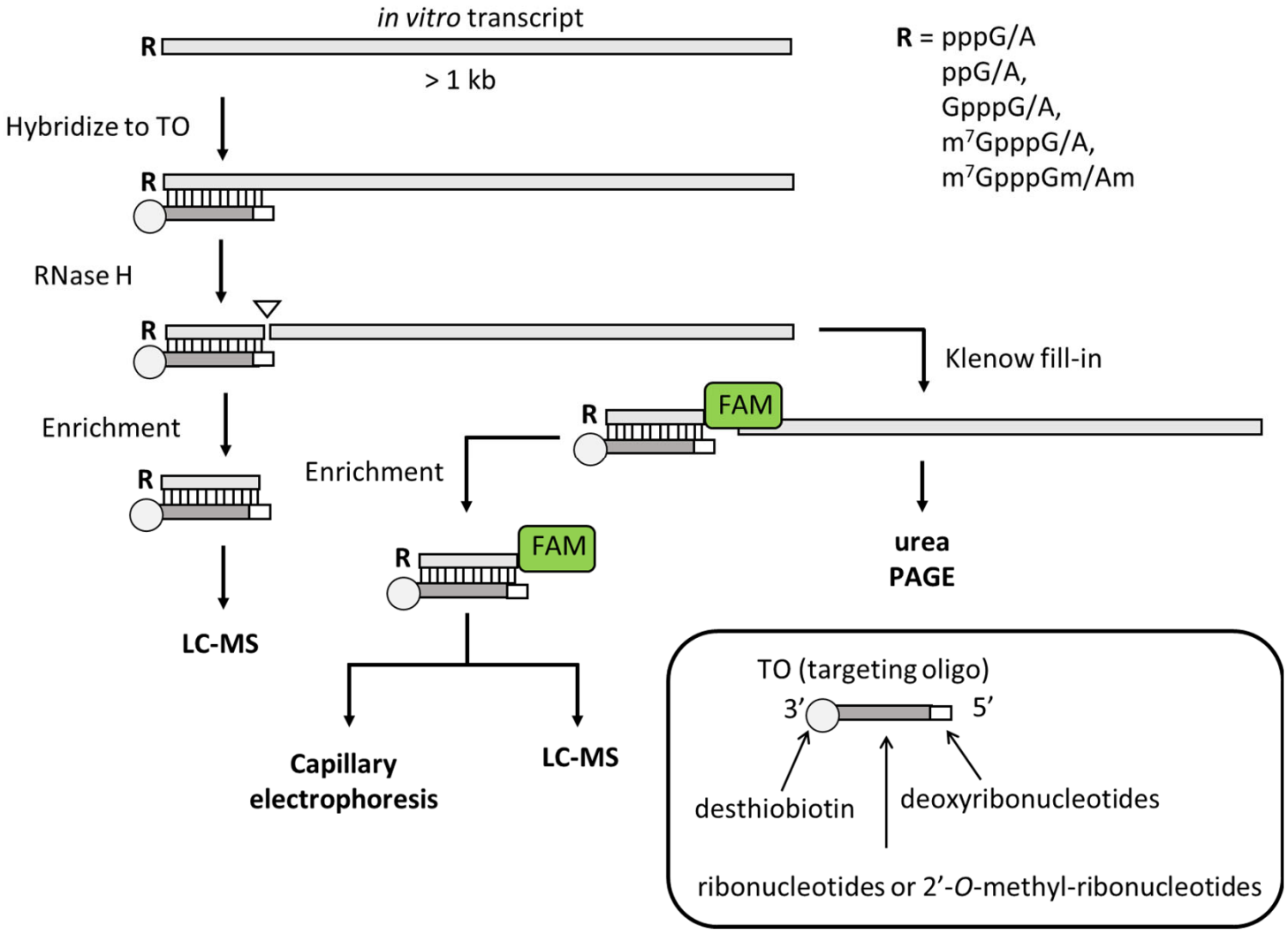
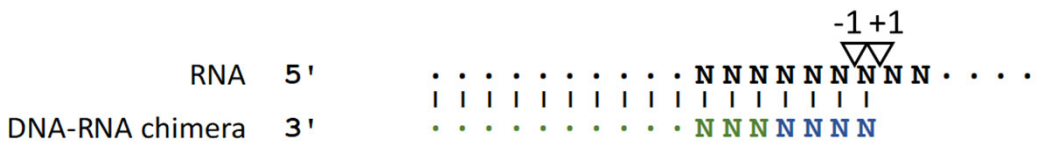


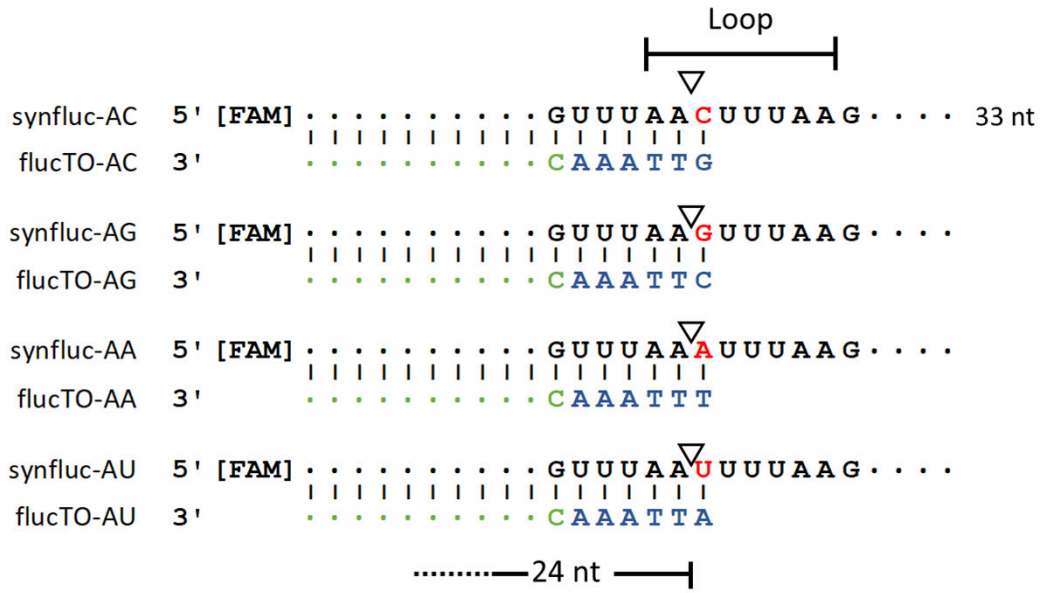
FIGURE 2. A general scheme of RNase H-based RNA cap analysis. A DNA-RNA or DNA-2'-O-methyl-ribonucleotide chimera is designed to be complementary to part of the 5' end of the target RNA molecules, such that the chimera stays annealed to the cleavage fragment after RNase H cleavage. The chimera (called Targeting oligo or TO in this communication) contains a 3'-desthiobiotin group. After denaturation and annealing, RNase H cleaves at a pre-defined site within the RNA-TO duplex and generate a 1-base recessive end at the 3' end of the cleaved RNA. Because RNase H cleavage results in a 3' hydroxyl group (24), this recessive 3' end can be filled in with a fluorescently-labeled deoxynucleotide using the Klenow fragment of DNA polymerase I. The fluorescently-labeled 5' cleavage fragment can be analyzed by denaturing PAGE directly without enrichment. The 5' duplex cleavage fragment can be enriched using streptavidin magnetic beads. The enriched RNase H cleavage products can be analyzed by LC/MS or capillary electrophoresis (if filled in with a fluorescent deoxynucleotide).

Fig. 3

A



B



C

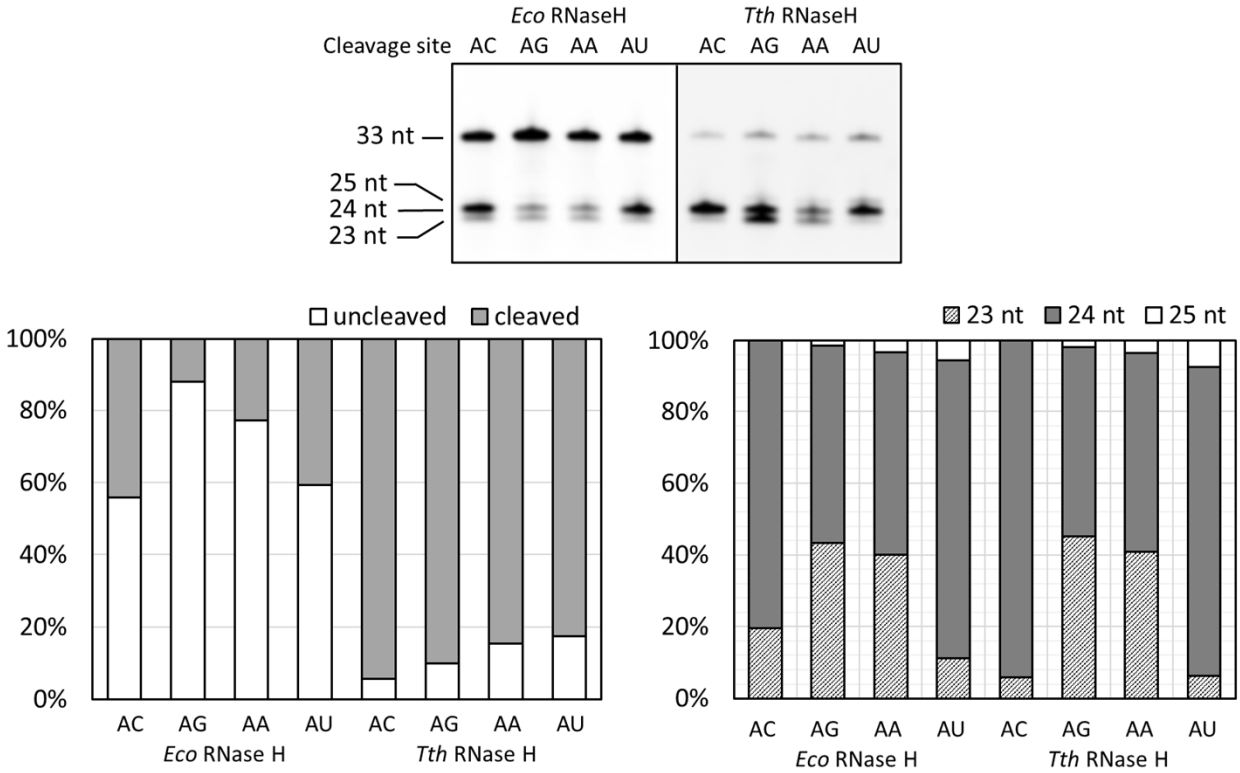


FIGURE 3. Directed RNA cleavage with RNase H (A) Previously reported RNase H cleavage sites (17-19). **(B)** Synthetic RNA oligonucleotides with artificial sequences and corresponding Targeting Oligos (TOs) designed based on (A). RNase H cleavage sites are indicated by inverted triangles. The variable nucleotides of the RNA substrates are colored in red. Deoxynucleotides in TOs are colored in blue; Ribonucleotides are colored in green. The predicted loop region of the synthetic RNA oligonucleotides based on an RNA folding algorithm (RNAFold, University of Vienna) is indicated. **(C)** RNA cleavage using E. coli or Thermus thermophilus RNase H at 37°C. Both RNase H enzymes prefer A|C and A|U over A|G or A|A; the Tth enzyme is more efficient in cleavage in general.

Fig. 4

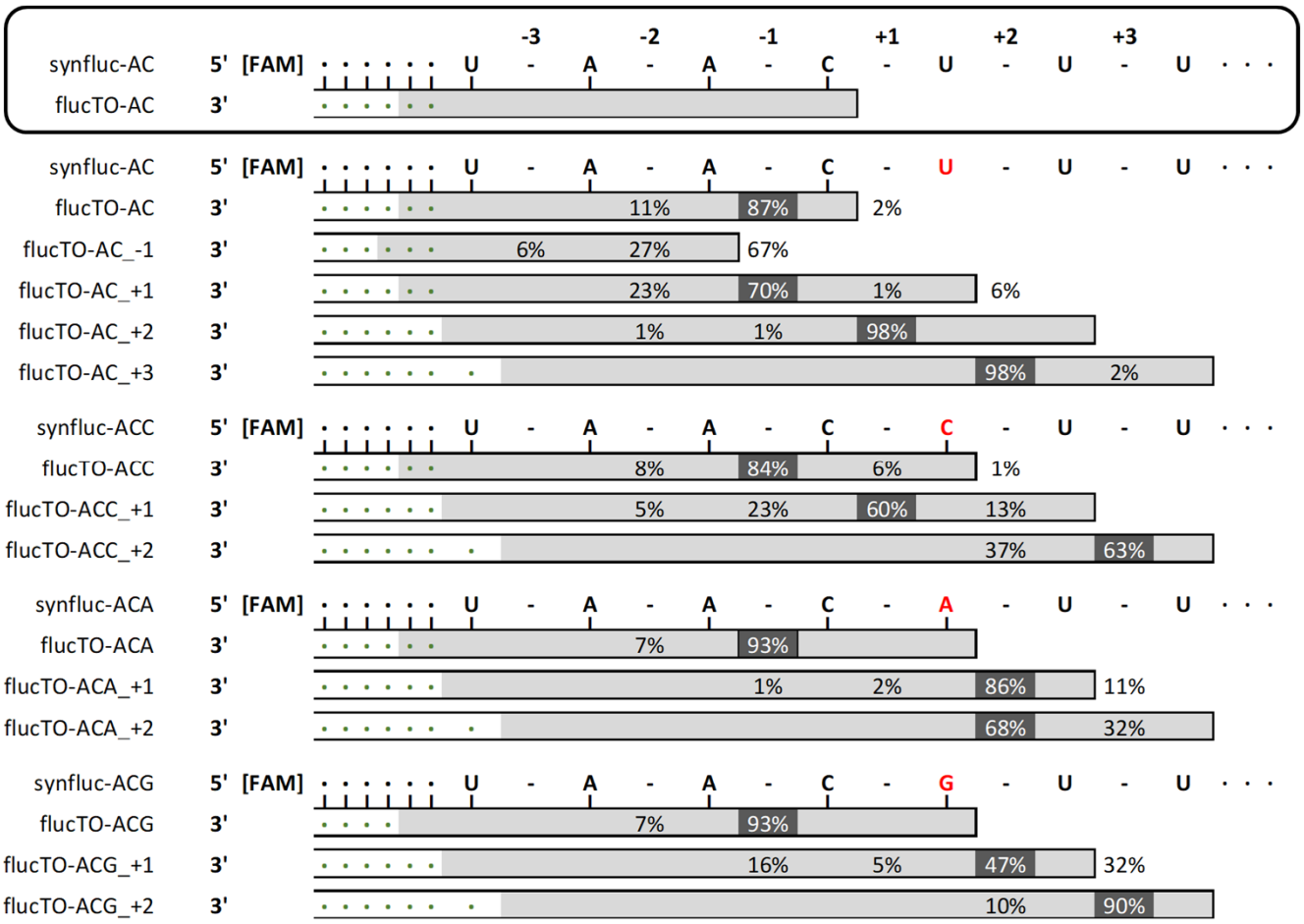


FIGURE 4. *Tth* RNase H cleavage frequency with respect to sequence context and TO design as analyzed by LC-MS. Within the sequence context and predicted structure of the artificial sequences in this study, high cleavage frequency was achieved at A|C, A|U, C|U or U|U sites. Variable nucleotides in the RNA substrate are colored in red; phosphodiester bonds are shown as "-". Deoxynucleotides in the TOs are shown as grey bars; Ribonucleotides are shown as green dots. Numbers represent frequency of cleavage detected at corresponding phosphodiester bonds.

Fig. 5

A

cleavage fragments (nt) 23 24 25

FLuc RNA (1.7 kb) 5′ GGGUCUAGAAAUAAUUUUGUUUAACUUUAAGAAGG...

TO-1 3′DTB–CCCAGAUCUUUAUUAAAACAAATTG 5′

B

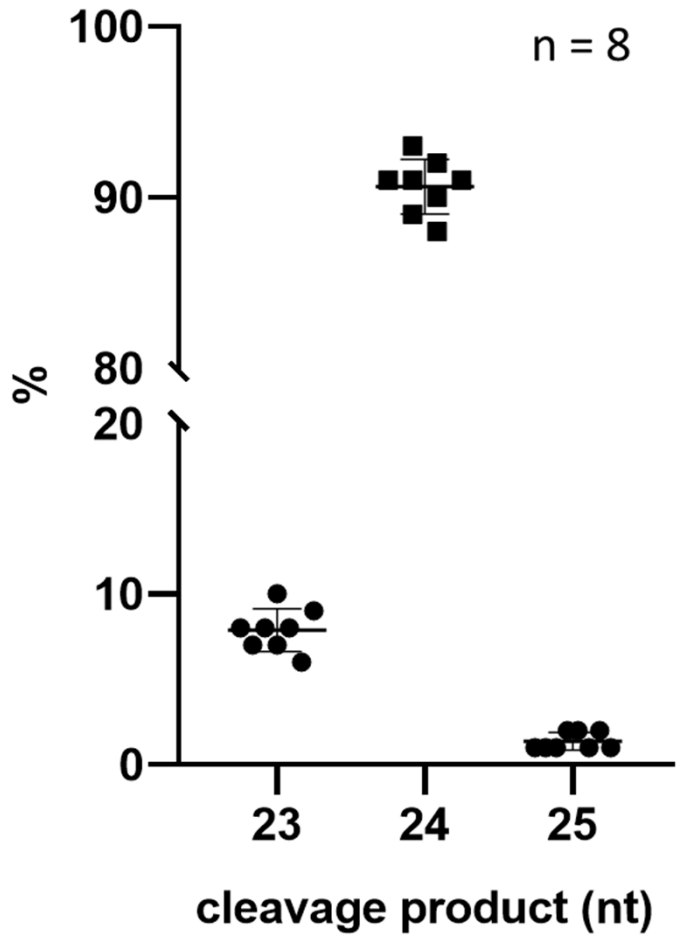


FIGURE 5. Uniform RNase H cleavage with designed targeting oligos (A) 5′ sequence of a 1.7 kb *in vitro FLuc* transcript containing an artificial 5′ UTR and the corresponding targeting oligo TO-1. The targeting oligo contains six deoxynucleotides (blue) at the 5′ end followed by 19 ribonucleotides (green) and a desthiobiotin (DTB) group at the 3′ end. The size of the expected RNase H cleavage products are shown. **(B)** Frequency of cleavage events expressed as percentage of detected cleavage product using LC-MS. Median of cleavage frequencies was 8%, 91% and 1% at (A|A), (A|C) and (C|U), respectively.

Fig. 6

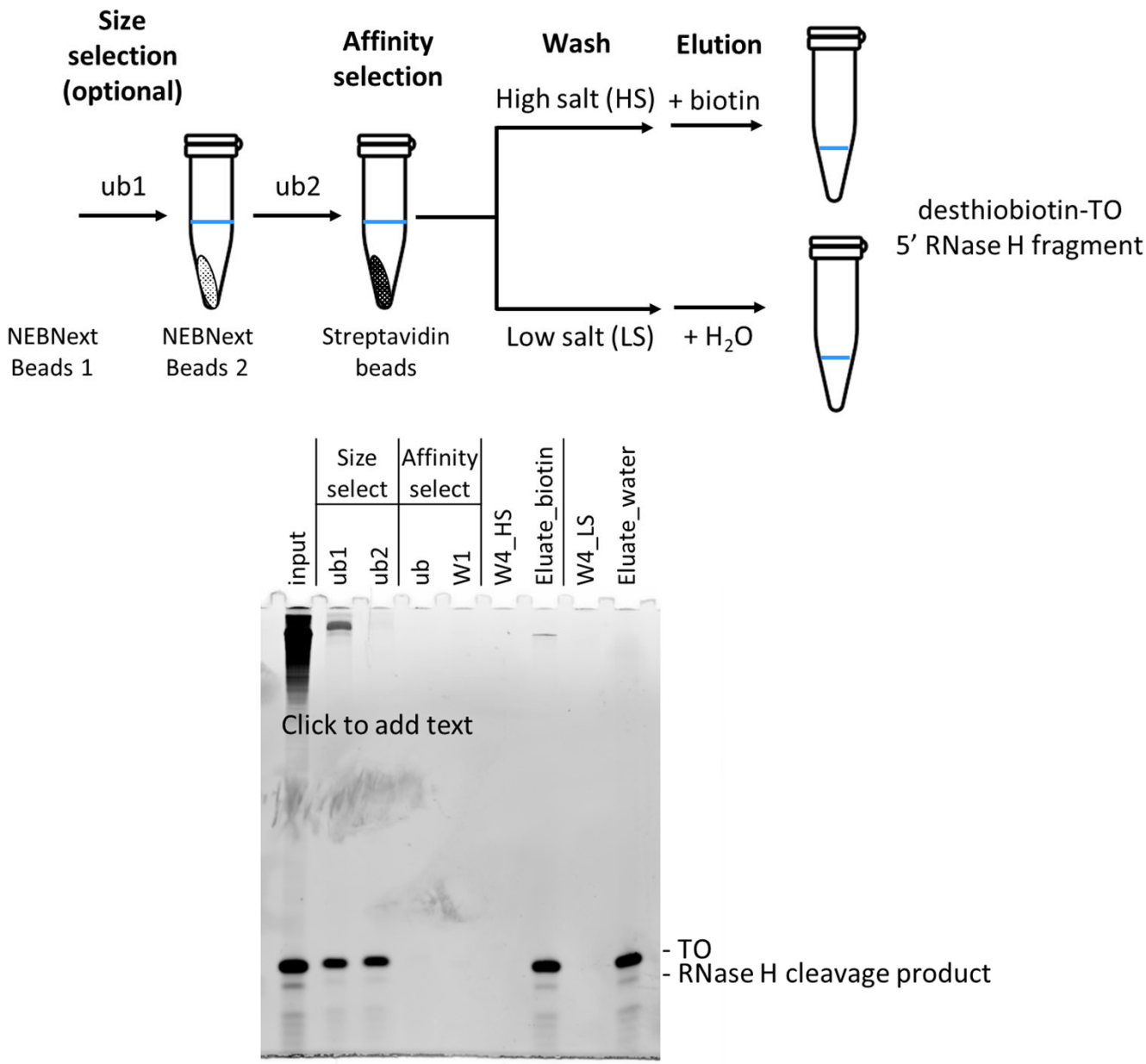
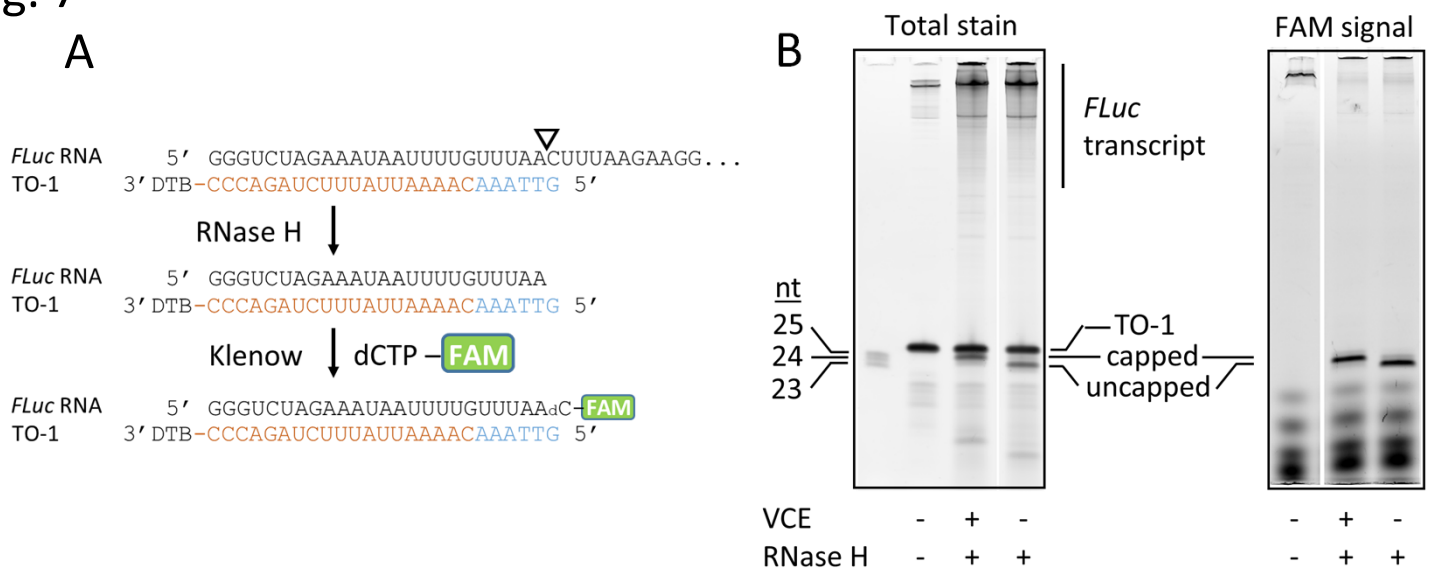
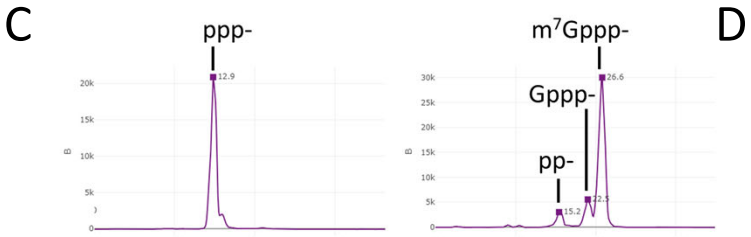


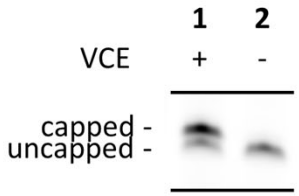
FIGURE 6. Enrichment of RNase H cleavage products by size and affinity selection. The RNase H cleavage products were first selected by size by two rounds of NEBNext magnetic beads. The clarified unbound fraction of the second round of size selection (ub2) was then added directly to streptavidin magnetic beads for affinity selection. After the first wash using a standard wash solution containing 1 M NaCl (W1), the resuspended bead slurry was divided into two fractions. One fraction was washed three more time using the standard wash buffer (W4_HS) and eluted using biotin (Eluate_biotin). The other fraction of the slurry was washed three more time using a low NaCl wash solution (W4_LS) followed by elution using the same volume of water (Eluate_water). Similar amount of RNase H cleavage product and TO were eluted (also see suppl. Figure. 4).

Fig. 7





	No enzyme	VCE
ppp-	100.0%	0.0%
pp-	0.0%	7.3%
Gppp-	0.0%	14.3%
m7Gppp-	0.0%	78.4%



Rxn 1	Urea-PAGE	CE	LC-MS
m7Gppp-	70%	62%	60%
Gppp-		1%	6%
pp-	30%	37%	33%
ppp-		0%	0%

FIGURE 7. Fluorescent labeling of RNase H 5' cleavage products and analyses. (A) Schematic representation of targeting, cleavage, and labeling. The targeting oligo was designed to guide RNase H to generate a 1-nt 3' recessive end, which can be filled in by a fluorescently labeled deoxynucleotide using the Klenow fragment. In this example, a FAM-labeled dCTP, was incorporated to the 3' end of the *FLuc* RNase H cleavage fragment, which can be directly analyzed by denaturing PAGE (B) or capillary electrophoresis (C) after enrichment. **(B)** Polyacrylamide gel analysis of cleaved RNA fragments. The 1.7 kb *FLuc* transcript capped using Vaccinia RNA capping enzyme (VCE) was subjected to the RNase H/Klenow fill-in treatments. Reactions were analyzed directly by urea PAGE followed by laser scanning of total RNA stain using SYBR Gold (left panel) or fluorescent signal (right panel). The targeting oligo TO-1 is invisible when the gel was scanned using the FAM channel and did not interfere with quantitation of the 5' cleavage products. **(C)** Resolution and quantification of capping and capping intermediates using capillary electrophoresis. The *FLuc* transcript was capped using a low concentration of VCE (10 nM) and subjected to the RNase H/Klenow fill-in reactions. After enrichment, the RNA was analyzed using capillary electrophoresis. In addition to substrate 5' triphosphate (ppp-) and the product m7Gppp-capped forms, enzymatic intermediate products 5' diphosphate (pp-) and the unmethyl-cap (Gppp-) can be resolved and quantified. **(D)** Synthetic mRNA cap analysis using gel electrophoresis, capillary electrophoresis, and LC-MS intact mass analysis produce comparable results. After RNase H/Klenow fill-in reactions and enrichment, an uncapped or partially capped *FLuc* transcript was analyzed using all three available methods. Capillary electrophoresis and LC-MS yielded comparable results in quantification of substrate (ppp-), product (m7Gppp-) and intermediate products (pp- and Gppp-). Urea-PAGE does not resolve pp- from ppp- or Gppp- from m7Gppp-. Considering ppp- and pp- as uncapped and Gppp- and m7Gppp- as capped species, quantitation of fluorescently labeled RNase H cleavage products using urea-PAGE generate results comparable to CE or LC-MS albeit the lack of resolution for intermediate products.

Fig. 8

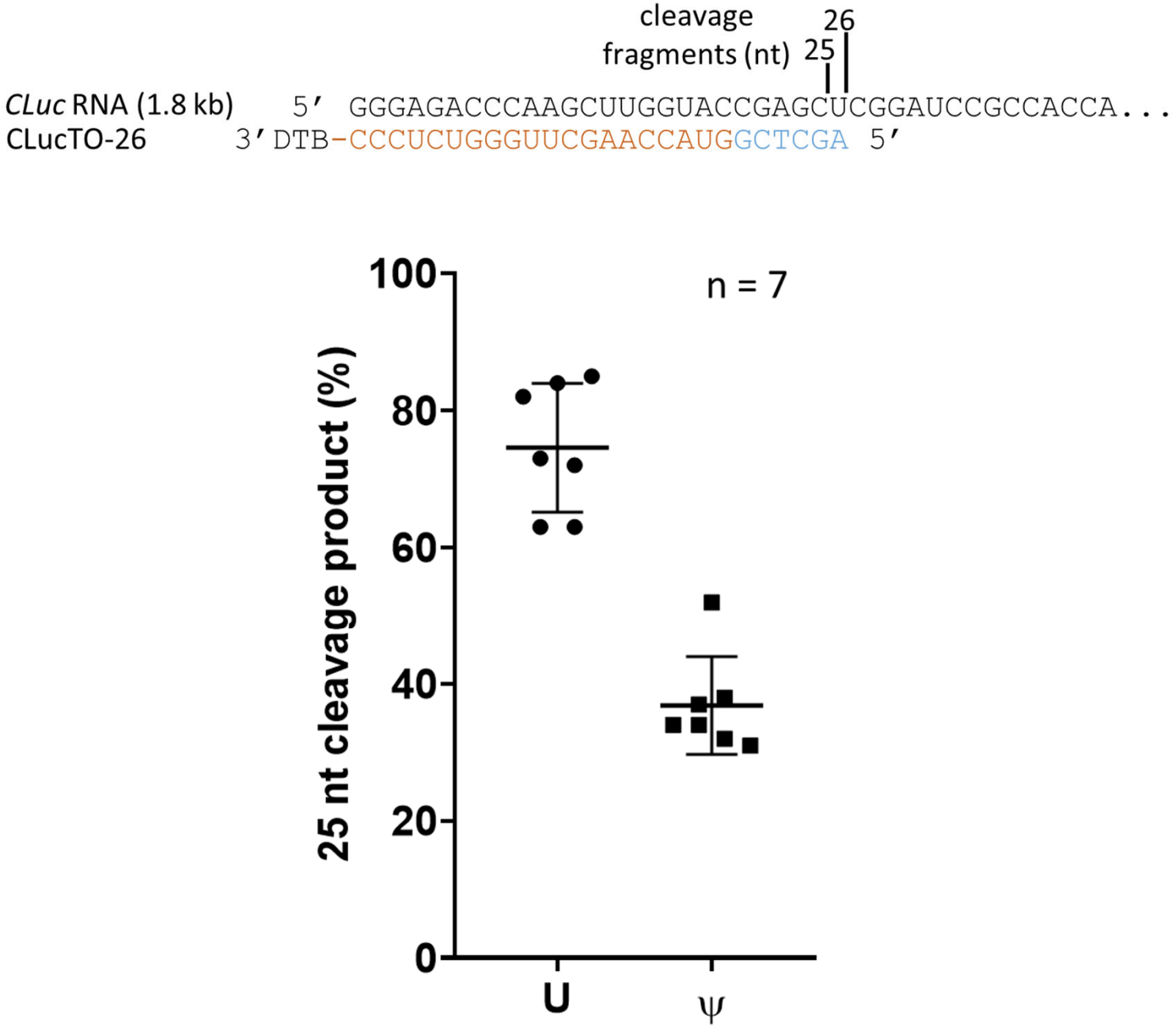


FIGURE 8. The effect of pseudouridine on RNase H cleavage. A *Cypridina* luciferase transcript (*CLuc*; 1.8 kb) was cleaved using *Tth* RNase H in conjunction with clucTO-26, which was designed to include a uridine residue in the RNase H cleavage site. The expected size of the cleavage products and cleavage sites are indicated. When unsubstituted, median cleavage frequency at the C|UC site was 73% (25 nt) and at the CU|C site was 27% (26 nt). When all uridines were substituted with pseudouridine, cleavage at the C|ΨC site decreased to a median frequency of 34%, with 66% cleavage at the CΨ|C site.