# RNase 4 Enables Targeted Isolation and Quantitative Analysis of mRNA 5′ Cap Structures

Jeremy C. Henderson, Ph.D., Eric J. Wolf, Ph.D., Nan Dai, Ph.D., Gavin Kempskie, Joe Whipple, Ph.D., New England Biolabs Inc.

## INTRODUCTION

With the growth of synthetic messenger RNA (mRNA)-based therapeutics and vaccines, the development of analytical tools to characterize complex mRNA sequences and their key chemical features is essential. Characterization of lengthy mRNAs by liquid chromatography tandem mass spectrometry (LC-MS) requires enzymatic digestion into shorter, analyzable oligonucleotide sequences. In this application note, we demonstrate how RNase 4, an RNA endonuclease that specifically cleaves 3′ of uridine in uridine-purine sequences (U/A or U/G), is used to generate mRNA 5′ cap oligonucleotide sequences for LC-MS. As RNase 4 is specific for single-stranded RNA, this protocol requires the design and use of a protective biotin-DNA probe complementary to the first 16–45 nucleotides of the target mRNA 5′ end sequence. Enrichment of post-RNase 4-digested mRNA 5′ oligonucleotides is achieved via streptavidin pull-down. We also detail recommendations for streptavidin magnetic bead enrichment that minimize biotin-DNA probe carryover during elution of RNA analyte, to provide superior mRNA 5′ end signal-to-noise in LC-MS workflows. In summary, this application note demonstrates how RNase 4 can be used for precise chemical description of mRNA 5′ ends. Additional commentary and practical guidelines are included to highlight experimental design choices that contribute most to protocol success.

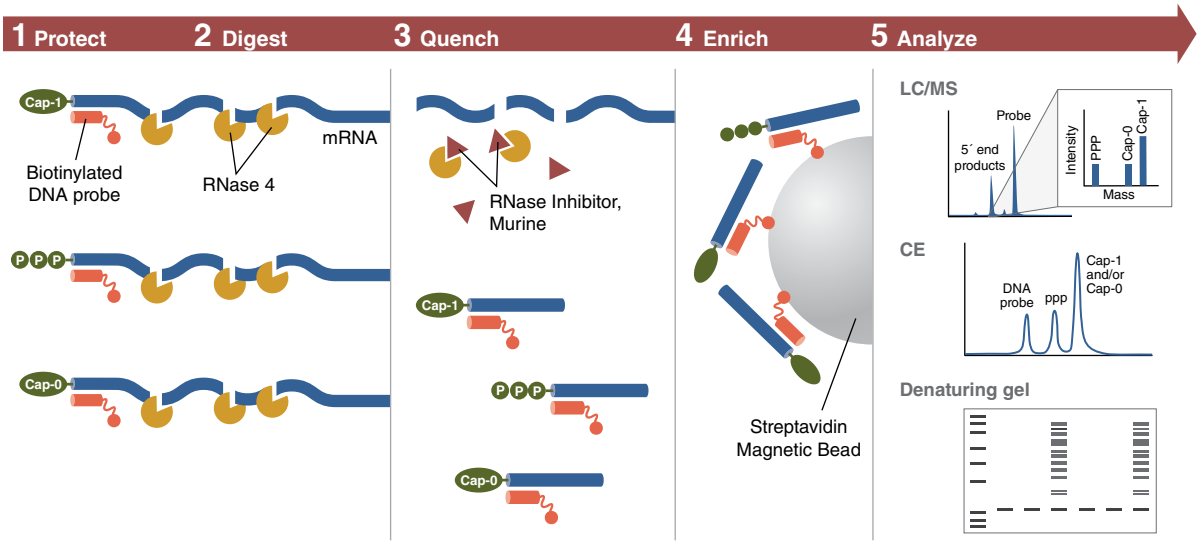
## MATERIALS

- RNase 4 (NEB #M1284)
- RNase Inhibitor, Murine (NEB #M0314)
- Streptavidin Magnetic Beads (NEB #S1420)
- NEBNext Magnetic Separation Rack (NEB #S1515)
- Biotin-DNA Probe

## ANALYSIS OF CAP-1 mRNA NIST RESEARCH-GRADE TEST MATERIAL

An overview of the RNase 4 protocol for isolating mRNA 5′ cap structures for detailed structural analysis is provided in Figure 1. RNase 4 is a single-stranded RNA endonuclease that cleaves 3′ of uridine in uridine-purine sequences (cut sites: U/A or U/G). U-ending RNA oligonucleotides generated after RNase 4 digestion contain a 2′, 3′-cyclic phosphate or linear 3′-phosphate terminus. RNase 4 activity is specific for single-stranded RNA; thus, a biotin-DNA probe complementary to a desired RNA sequence of interest can be used to protect an RNA region from digestion. This enables targeted cleavage at the next available cut site downstream of the DNA:RNA duplex region. RNase 4 cut sites 2–4 nucleotides downstream of the DNA:RNA

FIGURE 1: **Visualized protocol of site-directed RNase 4 cleavage of mRNA 5′ ends & enrichment with protective biotin-DNA probe**




A scientist-designed biotinylated DNA oligomer is hybridized to the 5′ end of a complementary RNA sequence 2–4 nucleotides upstream of a chosen U/A or U/G RNase 4 cut site. As RNase 4 is specific for single-stranded RNA, any potential cut site within the hybridized DNA:RNA duplex region is protected from cleavage. RNase 4 digestions can be quenched with RNase Inhibitor (NEB #M0314), and probe-hybridized mRNA 5′ end products can be captured via Streptavidin Magnetic Beads (NEB #S1420). Enriched DNA:RNA duplexes are eluted and analyzed by HPLC-MS/MS, capillary electrophoresis, or gel based separation under denaturing conditions.

duplex allow for efficient cleavage (Figure 2). Post-digestion enrichment of cleaved mRNA 5′ oligonucleotides is achieved via magnetic streptavidin bead pull-down. The streptavidin enrichment protocol included in this application note is optimized to minimize carryover of the biotin-DNA probe in recovered RNA material, thus improving overall LC-MS data quality.

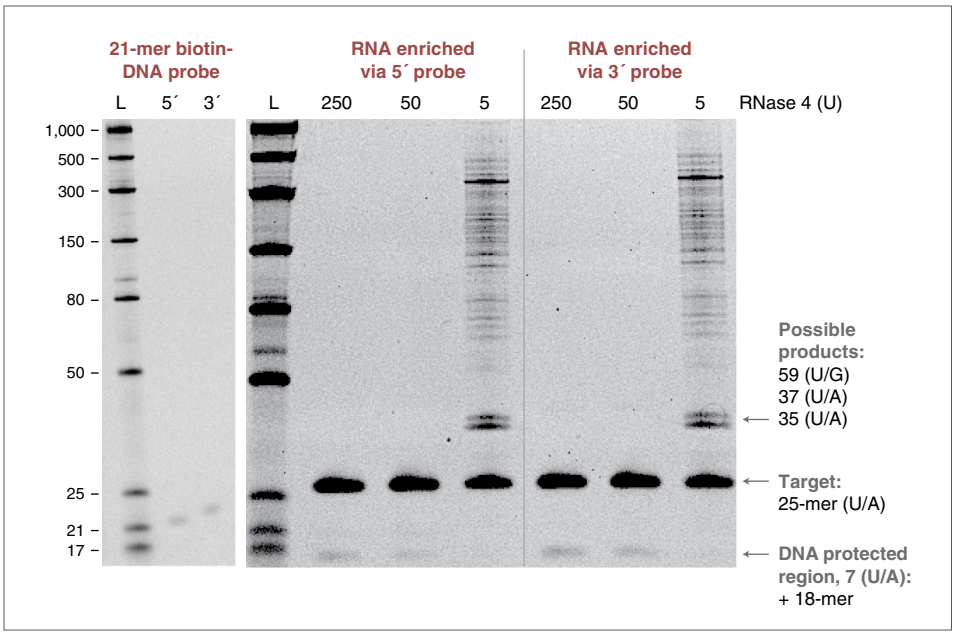
National Institute of Standards and Technology Research-Grade Test Material 10202 (NIST RGTM 10202) was obtained as part of an interlaboratory study to support assessment of mRNA Critical Quality Attributes (CQAs). These CQAs include sequence identity, concentration, intactness, 5′ capping efficiency, and 3′ poly(A) tail length. These attributes are measured for mRNA therapeutic products and to support future development of the RGTM 10202 material into a Standard Reference Material. A unit of RGTM 10202 consists of one 0.5 ml-cryovial containing approximately 50 µl of *in vitro* transcribed (IVT), purified FLuc mRNA (approximately 25 µg) in nuclease-free water, pH 6.4. The mRNA product has an enzymatically added Cap-1 structure on the 5′ end. The 3′ poly(A) tail was added co-transcriptionally during IVT. The primary nucleotide sequence consists of canonical (non-modified) bases linked through a phosphodiester backbone.

We first determined the optimal amount of RNase 4 units to digest NIST RGTM 10202 mRNA for 5′ cap analysis. Digestions contained 40 pmol biotin-DNA probe heat-hybridized to 40 pmol mRNA 5′ ends in a 10 mM Tris pH 7.0 buffered solution. We examined a set of biotin-DNA probes for their ability to yield the simplest population of mRNA 5′ ends after enrichment. These probes differed in 5′- or 3′- conjugation of biotin. Streptavidin-enriched RNase 4 digestion products were resolved by TBU-15% polyacrylamide gel electrophoresis and visualized with fluorescent nucleic acid dye (Figure 3). Consistent with the known length and mass quantity of NIST RGTM 10202 mRNA (1,934 nucleotides, 25 µg), 50 units of RNase 4 was sufficient to generate a mostly singular population of mRNA 5′ ends. This observed optimum was independent of 5′- or 3′-biotin conjugation. A single, protected RNase 4 U/A cut site is present at nucleotides 7/8. Minor levels of cleavage at this RNA site were observed, with a diagnostic ~18-mer oligonucleotide species visible on a gel (Figure 3). The intensity of the band representing the 18-mer correlates with the increasing units of RNase 4 included in digestion reactions.

FIGURE 2: **Design of protective biotin-DNA Probes for 5′ Cap analysis of NIST RGTM 10202 mRNA**


A reverse-complementary biotin-DNA probe (5′ shown or 3′ shown) is designed for the first 21 nucleotides of the NIST RGTM 10202 mRNA sequence. This positions a target RNase 4 site 4 nucleotides downstream of the RNA:biotin-DNA duplex region, which is expected to yield a pool of 25-mer mRNA 5′ species for characterization.

FIGURE 3: **15–50 units of RNase 4 optimal for digestion of 40 pmol NIST RGTM 10202 mRNA protected with 40 pmol Biotin-DNA**




Streptavidin enrichment of probe protected RNase 4 digestion products visualized by fluorescent nucleic acid stain after tris-borate urea (TBU) 15% polyacrylamide gel electrophoresis. Digestion of 40 pmol NIST RGTM 10202 mRNA sequence (~1934 nucleotides, 25 µg), was optimal for production of the target 25-mer RNA oligonucleotide in digestions that contain 50 units of RNase 4. This is consistent with recommended digestion conditions for general nucleotide sequence mapping for transcripts 1,000–10,000 nucleotides in length (NEB #M1284). A combination of two NEB ladders was mixed and loaded into a single reference lane: Low Range ssRNA Ladder (NEB #N0364S) and microRNA marker (NEB #N2102S). Probe only controls are included for reference, with either 5′ or 3′ biotin-DNA 21mer comigrating with length 21 nucleotide reference band.

Lower units of RNase 4 (5 units) yield more RNA species, especially those of length > 50 nucleotides (Figure 3, page 2). Overall, the optimal units of RNase 4 for complete digestion of 40 pmol of an unmodified RNA, approximately 2,000 nucleotides in length, to NIST RGTM 10202 mRNA, is between 15–50 units.

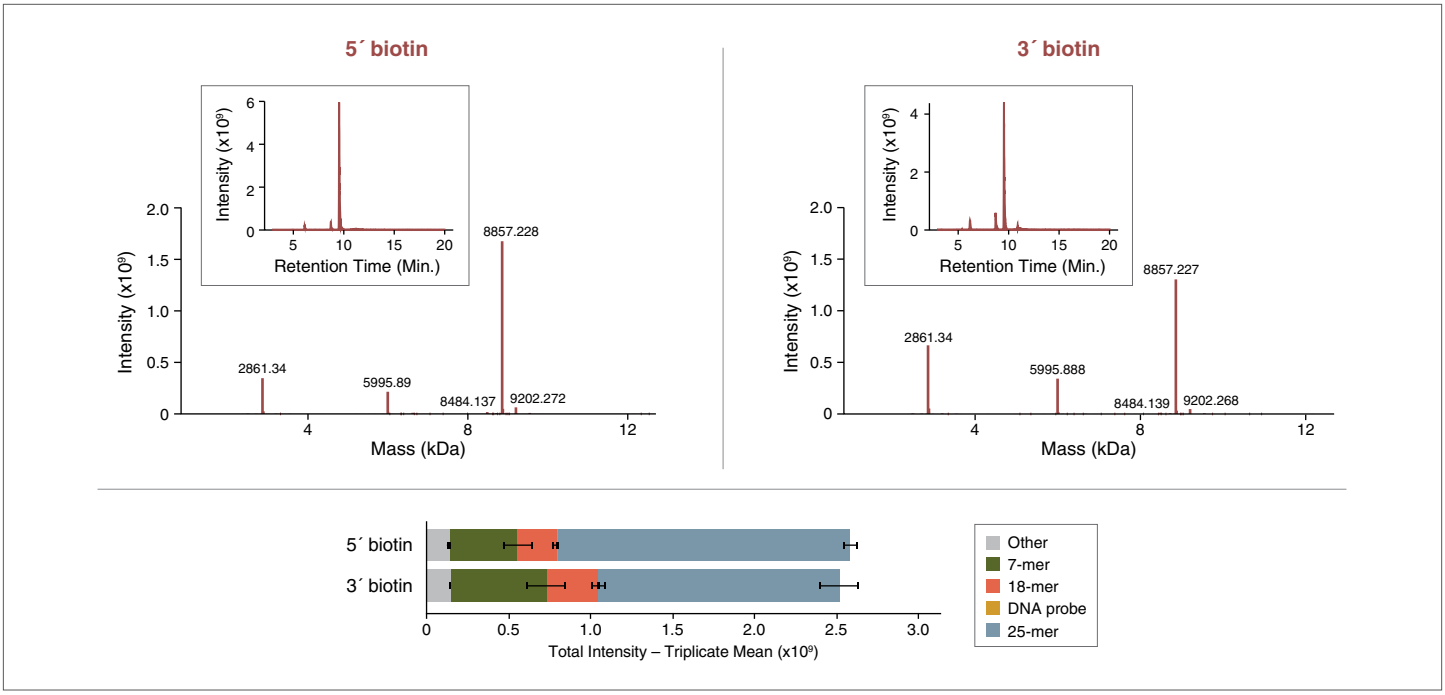
After optimization of RNase 4 digestion conditions, LC-MS data were collected on replicate digestions of NIST RGTM 10202 mRNA 5′ ends (Figure 4). We characterized 5′ ends obtained with either a 5′ or a 3′ biotin-DNA protective probe to determine whether any potential biases could be identified. By TBU-15% polyacrylamide gel, no apparent differences can be seen in the populations of 5′ ends obtained (Figure 3, page 2). In brief, our LC-MS workflow included separation of product RNA oligonucleotides on an ACQUITY® Premier Oligonucleotide C18 Column (Waters®) using a Vanquish™ HPLC, upstream to mass spectra acquisition via Orbitrap® Eclipse Tribrid mass spectrometer (Thermo Fisher®). Detected molecular ions were deconvoluted (Thermo Fisher BioPharma Finder), annotated into m/z peak table lists, and quantified as indicated (Figure 4).

The majority of observed molecular ions correspond to target 25-mer species generated by RNase 4 cleavage at the first available U/A cut site downstream of the RNA:DNA duplex region (Figure 5, page 4). The 5′ biotin-DNA probe captured a slightly higher total signal intensity for this 25-mer RNA species, compared with the 3′ biotin-DNA probe. Regardless of the 5′ or 3′ biotin conjugation strategy, the relative % intensity of observed 5′ cap structures did not vary in a statistically significant manner. By our analysis of the 25-mer population, NIST RGTM 10202 is estimated to contain 98.6 ± 0.1% Cap-1 (Figure 5, page 4). We did observe minor levels of intermediate Cap-0 and Gppp cap structures (~0.3-0.4% relative population), as well as some signal for 5′ diphosphate cap species (~1% relative population).

Surveying the other molecular ion species detected by our LC-MS analysis, we identified some 7-mer and 18-mer mRNA 5′ ends. As discussed for the TBU-15% polyacrylamide gel, these products arise due to an RNase 4 U/A cut site positioned at nucleotide 7/8 of the NIST RGTM 10202 sequence. Since they are complementary to the biotin-DNA probe, they are expected to be enriched during the streptavidin pulldown. Quantification of the relative 5′ cap species population based on the captured 7-mers reveals a nearly identical distribution of cap species to the 25-mer population (Figure 6, page 4). Further optimization of digestion conditions can reduce protected region cutting (e.g., by decreasing RNase 4 units from 50 units to 15 units).

Other minor $m/z$ species of note include putative T7 RNA polymerase mis-initiation events that result in +1 G or +1 A appendage to the templated transcript start sequence. These proposed sequences comprise < 2% of the relative signal of observed molecular ion species, with the majority of signal arising from +1 G. Additionally, some of the molecular ions binned as "other" species, including HPLC solvent adducts, were omitted for simplicity of quantification. We note these contaminants as further demonstration of the remarkable sensitivity of our RNase 4 digestion and streptavidin enrichment LC-MS protocol.

FIGURE 4: **LC-MS spectra obtained with streptavidin-enriched NIST RGTM 10202 mRNA 5′ ends, generated by the RNase 4 workflow**



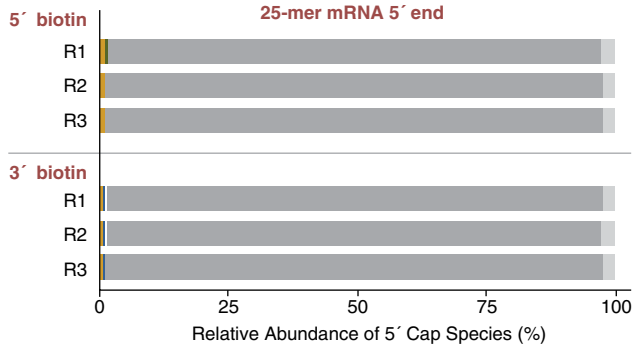

Relative intensity peak chromatograms of deconvoluted mass spectra, obtained for mRNA 5′ ends enriched via the 5′- or 3′- biotin-DNA probe. HPLC Total Ion chromatograms are also shown (inset). LC-MS summary graph visualizes total intensity of major species of interest. Error bars represent standard deviation obtained from triplicate sample analysis.

## PROBE DESIGN AND MAGNETIC STREPTAVIDIN BEAD RECOMMENDATIONS

Design of an effective biotin-DNA protective probe simplifies the population of mRNA 5′ ends generated by the RNase 4-based workflow. Whether the downstream chemical characterization method is capillary gel electrophoresis or a highly sensitive LC-MS/MS protocol, data quality and ease of interpretation rely on fundamentally sound probe design. We have obtained efficient mRNA 5′ end enrichment when the RNase 4 target site is 2–4 nucleotides downstream of the RNA:DNA duplex and the biotin-DNA probe ranges from 16–45 nucleotides in length. Heat elution at 80–90°C for 5 minutes in a 10 mM Tris pH 7.0 buffer solution has yielded better results than when elutions are performed in nuclease-free water (data not shown). In Figure 7 (page 5), the relative abundance of molecular species detected by LC-MS obtained with DNA probes targeting the 5′ end of a 5,000-nucleotide *in vitro* transcribed RNA is shown. The probes are either 5′ or 3′ chemically conjugated and contain either biotin or desthiobiotin. As a general trend, capture of RNA with biotin-DNA probes provided the highest signal of RNA 5′ ends compared with other detected molecular species (Figure 5). Due to the lower avidity of desthiobiotin for streptavidin compared with biotin, more desthiobiotin probe is released during elution. This increase in background probe signal limits the abundance of RNA species that can be detected during LC-MS characterization of mRNA 5′ ends. As observed with NIST RGTM 10202 Cap-1 mRNA, the 5′-biotin DNA probe enriched for a more abundant population of RNA 5′ ends when the 5k-nucleotide transcript was digested. The 5k-nucleotide transcript was not enzymatically capped and contained a 5′ triphosphate end.

Another trend evident from the LC-MS summary data (Figure 7, page 5), is how increased available binding capacity improves the relative intensity of captured mRNA 5′ end signal. A higher volume of beads used to capture RNA:DNA duplexes increased the relative proportion of RNA-specific signal by LC-MS. This effect was independent of biotin versus desthiobiotin, or of 5′- versus 3′-linkage. As the binding capacity of our magnetic streptavidin beads is functionally quantified using a single-stranded biotin-DNA oligomer, it makes spatial logic that an RNA:DNA duplex would require at least > 2X more equivalents of binding capacity for optimal capture performance. Of equal value, we consistently observe better performance when hydrophobic streptavidin beads are used to capture RNA:biotin-DNA duplexes compared

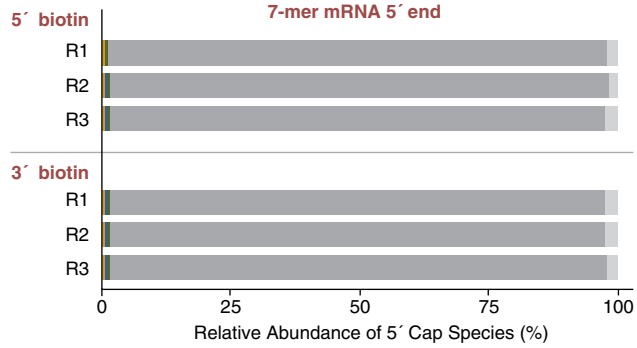
FIGURE 5: **Relative quantification of 5′ cap structures detected with target 25-mer species obtained by RNase 4 digestion of NIST RGTM 10202 mRNA**




| mRNA 5′ structure | 5′ biotin probe | 3′ biotin probe |
|---|---|---|
| **5′ pp-** GGGAAAUAAGAGAGAAAAGAAGAGU **>cp** | 1.03 ±0.07% | 0.93 ±0.06% |
| **5′ Gppp-** GGGAAAUAAGAGAGAAAAGAAGAGU **>cp** | 0.29 ±0.03% | 0.31 ±0.02% |
| **5′ Cap-0-** GGGAAAUAAGAGAGAAAAGAAGAGU **>cp** | 0.12 ±0.07% | 0.16 ±0.03% |
| **5′ Cap-1-** GGGAAAUAAGAGAGAAAAGAAGAGU **>cp** | 95.90 ±0.14% | 96.10 ±0.35% |
| **5′ Cap-1-** GGGAAAUAAGAGAGAAAAGAAGAGU **-p** | 2.67 ±0.20% | 2.50 ±0.27% |
| **Sum of detected Cap-1 species** | **98.57** ±0.24% | **98.60** ±0.44% |

Relative intensities tabulated for target mRNA 25-mers enriched via the 5′- or 3′- biotin-DNA probe. A low level of 3′- linear phosphate (3′|-p|) was observed in the most abundant Cap-1 pool (>98% total 25mer pool). All other species contain 2′,3′-cyclic phosphate (3′|>cp|). The LC-MS summary graph visualizes the intensity of major 25-mer species of interest. Each row indicates an individual replicate obtained with either 5′- or 3′- biotin-DNA probe.

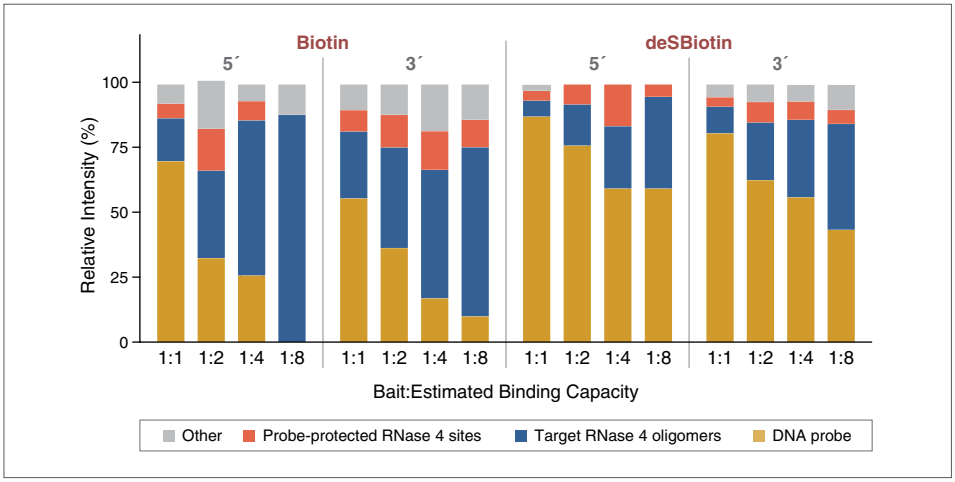
FIGURE 6: **Relative quantification of 5′ cap structures detected with 7-mer species obtained by RNase 4 digestion of NIST RGTM 10202 mRNA**




| mRNA 5′ structure | 5′ biotin probe | 3′ biotin probe |
|---|---|---|
| **5′ pp-** GGGAAAU **>cp** | 0.86 ±0.03% | 0.89 ±0.06% |
| **5′ Gppp-** GGGAAAU **>cp** | 0.22 ±0.06% | 0.30 ±0.05% |
| **5′ Cap-0-** GGGAAAU **>cp** | 0.30 ±0.00% | 0.32 ±0.02% |
| **5′ Cap-1-** GGGAAAU **>cp** | 96.44 ±0.51% | 96.11 ±0.26% |
| **5′ Cap-1-** GGGAAAU **-p** | 2.18 ±0.58% | 2.39 ±0.34% |
| **Sum of detected Cap-1 species** | **98.62** ±0.77% | **98.50** ±0.43% |

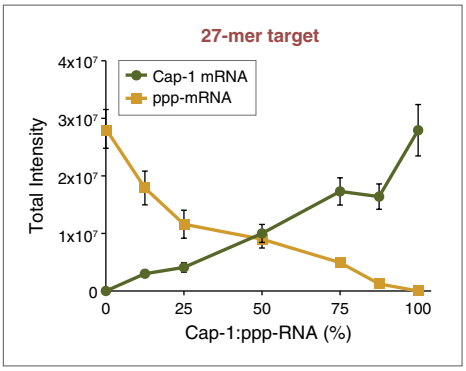
Relative intensities tabulated for mRNA 7-mers enriched via the 5′- or 3′- biotin-DNA probe. A low level of 3′- linear phosphate (3′|-p|) was observed in the most abundant Cap-1 pool (>98% total 7-mer pool). All other species contain 2′,3′-cyclic phosphate (3′|>cp|). The LC-MS summary graph visualizes the intensity of major 7-mer species of interest. Each row indicates an individual replicate obtained with either 5′- or 3′- biotin-DNA probe.

FIGURE 7: **Trends in target RNA enrichment based on chemical conjugation strategy & use of increased binding capacity with magnetic streptavidin beads**




A 5000 nucleotide RNA transcript was protected with a 20-mer DNA probe and digested with RNase 4. 5´- or 3´- conjugation of biotin- or desthiobiotin (deSBiotin) were compared. The ratio of biotin-DNA probe to excess streptavidinmagnetic beads was also compared, varying the molar ratio of biotin bait to known molar binding capacity (1:1, 1:2, 1:4, or 1:8 as indicated). Optimization of target RNase 4 oligomer capture (Blue) was observed with the biotin-DNA interaction. Higher ratios of excess binding capacity to bait biotin similarly improved capture of target RNA species (e.g., ~50 µl NEB #S1420 to 40 pmol RNA:biotin-DNA duplex).

FIGURE 8: **LC-MS total intensity of target 27-mer 5´ cap species derived from RNase 4 digestion of known molar mixtures of *in vitro* transcribed 5´ Cap-1 or triphosphate-mRNA**




Known mixtures of *in vitro* transcribed mRNA (2110 nucleotides) were protected with a 5´ biotin-DNA 22-mer and digested with RNase 4. Prepared molar mixtures were formulated to a final concentration of 40 pmol mRNA with digestions performed in triplicate. Streptavidin magnetic beads were used to capture target 27-mer RNA species. Total intensity of molecular ions corresponding to either Cap-1 or triphosphate 5´-end species were quantified. Values plotted are based on % molar Cap-1 mRNA of total 40 pmol mixture. Error bars indicate standard deviation in total intensity across triplicate digestions.

with hydrophilic beads (data not shown). We hypothesize that the hydrophobic beads promote higher avidity of the biotin and streptavidin interaction, and limit non-specific binding of nucleic acids to bead surface matrix; factors that enhance the effectiveness of wash and elution steps. Our current protocols ideally capture 40 pmol of RNA:DNA duplex with 50 µl of magnetic streptavidin beads (NEB #S1420).

## NO APPARENT BIAS IN CAPTURE OF CAP-1 RELATIVE TO TRIPHOSPHATE mRNA 5´ SPECIES

The effectiveness of mRNA therapeutics, in part, depends on the chemical identity of 5´ cap structures that promote desirable molecular phenotypes, such as *in vivo* stability and translational efficiency. By extension, in a manufacturing process, it is important to quantify 5´ capping efficiency of synthetically produced transcripts to ensure the overall quality of mRNA drug substances. To better understand whether our streptavidin-based enrichment method, or biotin-DNA probe strategy, biases the specific capture of mRNA 5´ cap structures, we deployed our RNase 4 protocol against known molar mixtures of Cap-1 and/or triphosphate RNA. Varying molar ratios of a 2,000-nucleotide transcript were prepared to a final concentration of 40 pmol. Absolute LC-MS intensities were obtained for a target 27-mer 5´ sequence and plotted relative to % molar Cap-1 mRNA (Figure 8). Within our specific LC-MS analysis workflow, the total intensity of 40 pmol of Cap-1 mRNA did not vary significantly from 40 pmol of triphosphate mRNA. This indicates comparable ionization efficiency of either target 27-mer RNA sequences. Similarly, the overall proportion of detected molecular ions correlates strongly with the known input ratio of Cap-1 to triphosphate mRNA. Given these observations, our RNase 4-based workflow can be used to effectively quantify variable mixtures of 5´ cap mRNA material, enabling assessment of mRNA drug substance CQAs such as 5´ capping efficiency.