Application Note

Biopharma

Agilent

Trusted Answers

Rapid Analysis of mRNA 5' Capping with High Resolution LC/MS

Author

Brian Liau
Agilent Technologies, Inc.

Introduction

Industrial-scale production of mRNA has received growing attention due to the ongoing SARS-CoV-2 pandemic. mRNA vaccines have proven to be one of the most effective tools against the virus and depend on high-quality mRNA synthesized by *in vitro* transcription.[1] Efficient translation of mRNA to protein depends critically on a 5'-terminal dinucleotide modification called capping which is appended to mRNA during[2] or after[3] the transcription process. Importantly, the percentage of successfully capped material and the type of capping structure appended is influenced by factors such as the quality of input materials, the reaction conditions, and the mRNA sequence. mRNA 5' capping is therefore a critical quality attribute that should be thoroughly characterized and monitored.

This work uses an Agilent 6545XT AdvanceBio LC/Q-TOF for rapid mRNA capping analysis with a combined sample preparation and analysis time of only 75 minutes. Sample preparation is accelerated using a thermostable enzyme to liberate 5' terminal oligonucleotides containing capping structures at an elevated reaction temperature (Figure 1). Capped oligonucleotides are then separated from the sample matrix using an AdvanceBio Oligonucleotide column using only simple cleanup procedures.

To demonstrate the utility of this method, we optimize a Vaccinia enzyme capping reaction for a challenging mRNA sequence. Our results show the value of high-resolution LC/MS as a sensitive and efficient method for process optimization and quality control of nucleic acid therapies.

Abbreviations used in this work

- DNA: Deoxyribose nucleic acid
- PCR: Polymerase chain reaction
- mRNA: Messenger ribonucleic acid
- ARCA: Anti-reverse cap analogue
- UTR: Untranslated region
- GTP: Guanosine triphosphate
- NTP: Nucleoside triphosphate
- SAM: S-adenosylmethionine

Experimental

All PCR primers and synthetic RNA/DNA chimeric probes were synthesized by Integrated DNA Technologies.

In vitro transcription and mRNA capping

A plasmid encoding a model ~3,800 nt sequence downstream of a T7 promoter was purchased from Sino Biological. The sequence included 5' and 3' UTRs as well as a coding region. The T7 promoter and sequence of interest were PCR-amplified using a Herculase II Fusion DNA Polymerase kit (Agilent part number 600677) and cleaned up using a StrataPrep PCR purification kit (Agilent part number 400771). The specificity of the PCR reaction and concentration of the product were measured on an Agilent 2100 Bioanalyzer Instrument with DNA 7500 kit (part number 5067-1506).

The PCR product was then transcribed to mRNA using T7 RNA polymerase (New England Biolabs M0251) using the manufacturer's recommended protocol. 100 µL *in vitro* transcription reactions were diluted with 100 µL nuclease-free water prior to DNase-I (New England Biolabs M0303) digestion to eliminate DNA template and residual PCR primers. mRNA was then precipitated by adding 70 µL of 8 M LiCl, chilling overnight at –20 °C, then centrifuging at 12,000x g for 15 minutes at 4 °C. mRNA pellets were washed twice in 70% ethanol, air-dried, then dissolved in nuclease-free water and quantitated using an Agilent 2100 Bioanalyzer Instrument with RNA 6,000 Nano kit (part number 5064-1511). The mRNA was then made into 5 µg aliquots and frozen at –80 °C.

Two methods of mRNA capping were used:

1) Co-transcriptional capping with ARCA

2) Enzymatic capping with Vaccinia capping enzyme

For ARCA capping, mRNA samples were transcribed with an NTP mix containing a 4:1 ratio of ARCA to GTP. For enzymatic capping, 5 to 10 µg of purified uncapped mRNA was capped using a Vaccinia Capping System kit (New England Biolabs M2080), then purified using a Monarch RNA cleanup kit (New England Biolabs T2040) and eluted in nuclease-free water prior to analysis.

Site-directed RNase-H cleavage

In-solution RNase-H site-directed cleavage was implemented as described by Lapham, J. *et al.*[4] 15 nt chimeric 2'-O-methyl RNA/DNA probes were designed complementary to the mRNA sample (Figure 1) and used to direct RNase-H cleavage under two conditions: (1) 1-hour thermal annealing program followed by 30-minute cleavage with RNase-H at 37 °C, or (2) 30-minute cleavage with thermostable RNase-H at 50 °C. After cleavage, oligonucleotides of length 40 nt or 50 nt containing the 5' cap were liberated and digested samples were purified using a Monarch RNA cleanup kit (New England Biolabs T2040) prior to visualization on an Agilent 2100 Bioanalyzer Instrument with Small RNA kit (part number 5067-1548).

LC-DAD/MS of 5' capped oligonucleotides

Instrumentation

- 1290 Infinity II LC with diode array detector (part number G7117B)
- 6545XT AdvanceBio LC/Q-TOF

Care was taken to eliminate glass from the flow path to reduce alkaline metal adduction. Agilent Nalgene bottles (part number 9301-6460) were used as mobile phase containers, and each solvent line was equipped with a steel frit. Agilent polypropylene sample vials were used (part number 5190-2242). Before the start of each day's experiments, the LC system and column were flushed with a 50% MeOH + 0.1% formic acid solution for 30 mins to further reduce alkaline metal adducts.

Digested mRNA samples were separated on an AdvanceBio Oligonucleotide column (2.1 × 50 mm, 2.7 µm, 120 Å, part number 659750-702). The mobile phase and LC gradients are shown in Table 1, mass spectrometer settings in Table 2, and data analysis was performed in MassHunter BioConfirm 10.0 with deconvolution settings shown in Table 3.

Table 1. Mobile phase and LC gradient.

Agilent 1290 Infinity II LC System	
Column	AdvanceBio Oligonucleotide, 2.7 µm, 2.1 × 50 mm,120 Å
Solvent A	15 mM dibutylamine + 25 mM HFIP in DI water
Solvent B	15 mM dibutylamine + 25 mM HFIP in methanol
Gradient	**Time (min) % B** 0 10 4 30 25 51 25 to 28 90
Column Temperature	50 °C
Flow Rate	0.4 mL/min
Injection Volume	20 µL
mRNA Per Injection	2 µg

Table 2. Mass spectrometer settings.

Agilent 6545XT AdvanceBio LC/Q-TOF	
Acquisition Mode	Negative, standard (3,200 *m/z*) mass range, High Sensitivity (2 GHz)
Gas Tempurature	350 °C
Gas Flow	12 L/min
Nebulizer	25 psig
Sheath Gas Tempurature	275 °C
Sheath Gas Flow	10 L/min
V_{cap}	4,500 V
Nozzle Voltage	2,000 V
Fragmentor	250 V
Skimmer	65 V
MS1 Range	400 to 3,200 *m/z*
MS1 Scan Rate	2 Hz
Reference Mass	1,033.9881

Table 3. Deconvolution settings.

MassHunter BioConfirm 10.0 Settings	
Extract Chromatogram (MS)	Diode Array Detector
MS1 Delay Time	0.06 seconds
Decon. Algorithm	Maximum Entropy
Subtract Baseline	1
Adduct	Proton Loss
Mass Range	10,000 to 30,000 Da
Mass Step	0.5 Da
Use Limited *m/z* Range	1,040 to 3,200

Results and discussion

As full-length mRNA is typically too large for standard LC/MS analysis, we first optimized RNase-H site-directed cleavage. Two RNA/DNA chimeric probes were designed complementary to different sequences within the 5' UTR, directing RNase-H cleavage[4] to either nucleotide 40 or 50 (Figure 1A) so that oligonucleotides of 40 nt or 50 nt in length, bearing the 5' capping structure, would be liberated upon successful cleavage. These smaller 5' capped oligonucleotides are amenable to LC/MS analysis.

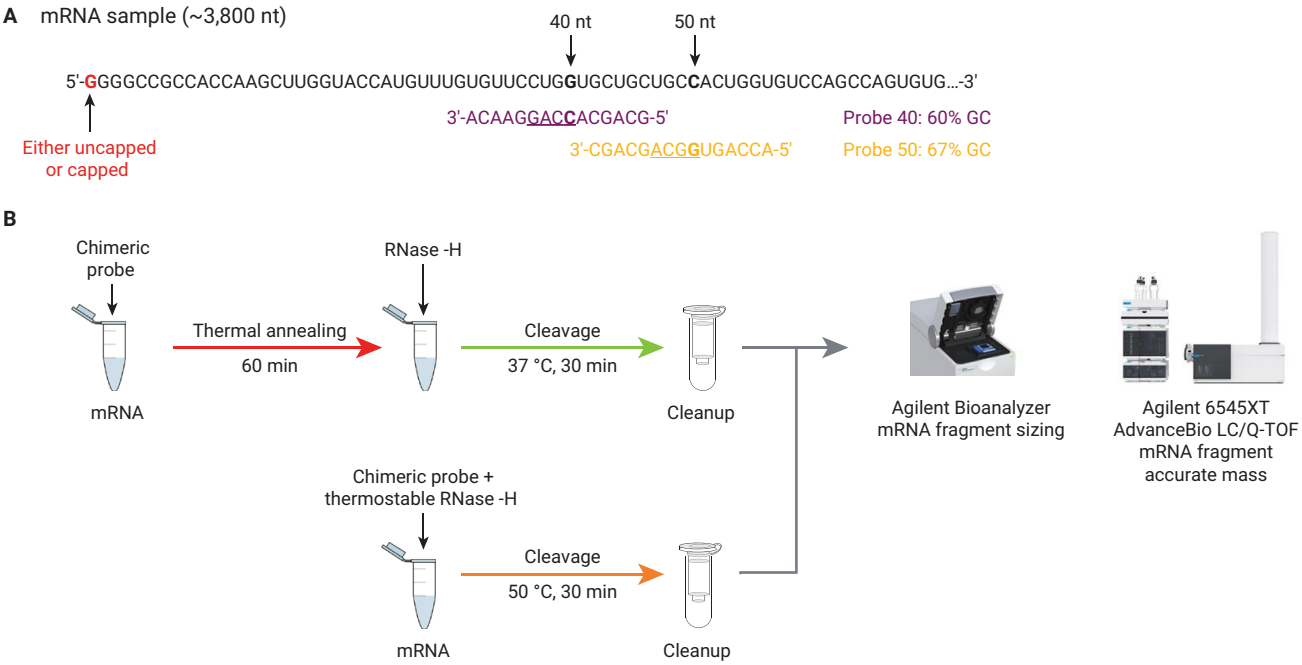

Figure 1. mRNA sample preparation and analysis. (A) Chimeric probes complementary to the mRNA sample were designed to direct RNase-H cleavage at 40 nt (purple) and 50 nt (gold). Underlined probe nucleotides were composed of DNA and were otherwise composed of 2'-O-methyl RNA. (B) Sample preparation scheme using non-thermostable (top scheme) or thermostable RNase-H (bottom scheme). Total analysis time using bottom scheme: cleavage (30 minutes) + cleanup (10 minutes) + LC/MS (25 minutes) + data analysis (10 minutes) = 75 minutes.

As published elsewhere[5], the typical workflow for such an experiment entails a separate one-hour thermal annealing step prior to the addition of non-thermostable RNase-H enzyme, followed by sample cleavage at 37 °C (Figure 1B, top scheme). Separate steps are necessary because the high temperature (~95 °C) used during thermal annealing would otherwise: (1) denature non-thermostable RNase-H, and (2) promote mRNA hydrolysis due to divalent cations present in the reaction buffer.[6]

We hypothesized that the cleavage reaction could be expedited by performing it at a moderately raised temperature of 50 °C with thermostable RNase-H (Figure 1B, bottom scheme) provided chimeric probes of sufficiently high GC content were used. In principle, the elevated temperature would be expected to unwind mRNA secondary structure, permitting high GC content chimeric probes to bind to their complementary mRNA regions and direct cleavage by thermostable RNase-H.

As shown in Figure 2, site-directed cleavage was performed using two different high GC content probes (Probe 40 and Probe 50, each >60% GC) with either non-thermostable or thermostable RNase-H. The resulting digests were separated using an Agilent 2100 Bioanalyzer automated electrophoresis tool with Small RNA kit (part number 5067-1548) and qualitatively assessed using a virtual gel image representation.

Both Probe 40 and Probe 50 successfully directed cleavage of the mRNA sample under all conditions. However, Probe 50 exhibited off-target binding when using non-thermostable RNase-H, resulting in undesirable products (Figure 2, red arrows). When thermostable RNase-H was used, Probe 50 directed specific cleavage, allowing a cleaner sample to be obtained in a shorter time frame.

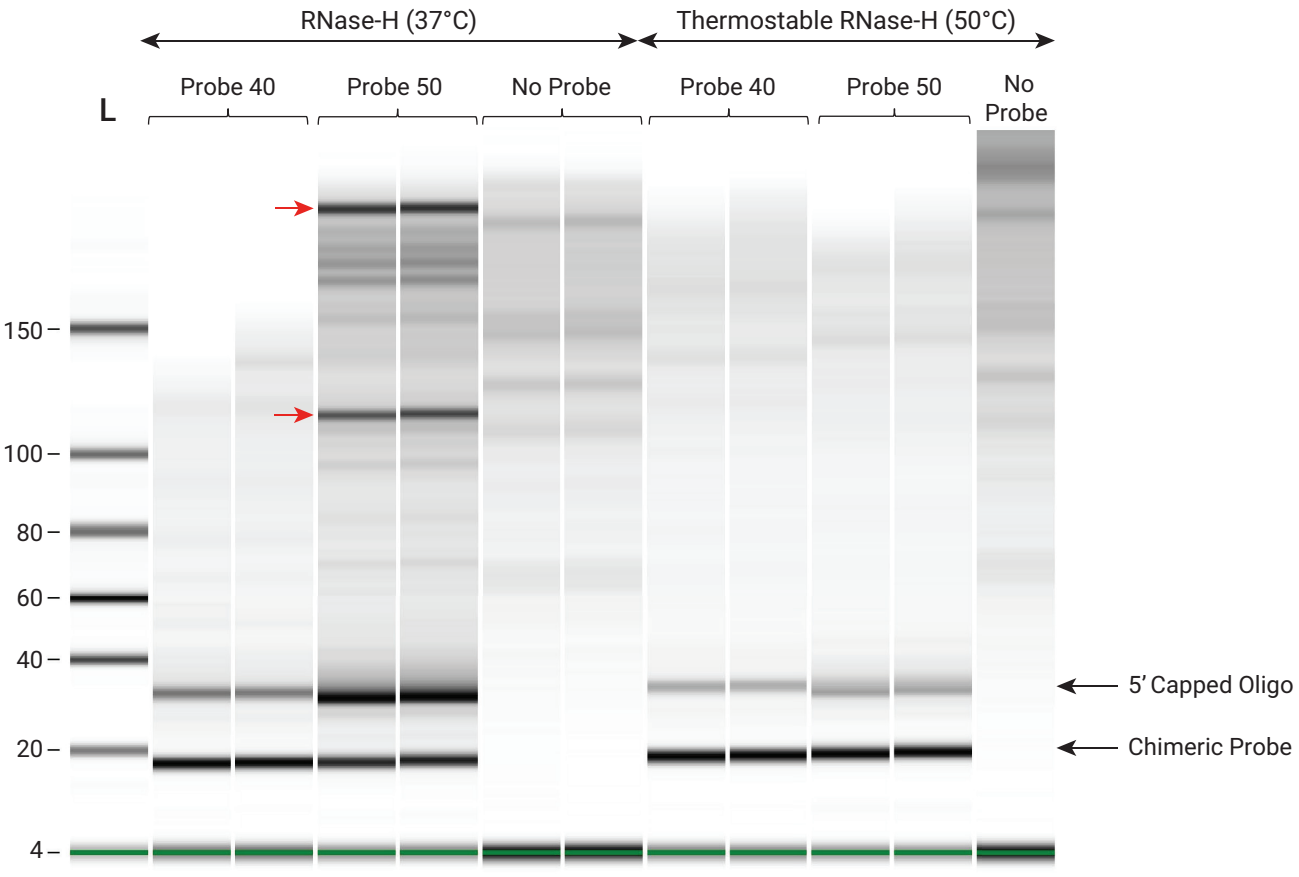

Figure 2. Assessment of site-directed cleavage using thermostable and non-thermostable RNase-H. Probe 40 and Probe 50 direct enzymatic cleavage at nucleotides 40 and 50 respectively. Red arrows indicate off-target cleavage products. L: ladder.

Next, uncapped mRNA digested using Probe 50 + thermostable RNase-H was analyzed by IP-RP LC/MS. Figure 3 shows the separation of uncapped 5' oligonucleotides from Probe 50 and the remaining mRNA on an Agilent AdvanceBio Oligonucleotide column (2.1 × 50 mm, 2.7 µm, 120 Å, part number 659750-702). Using automated data analysis workflows built into MassHunter BioConfirm 10.0, mass spectra were extracted from peaks identified in the UV chromatogram, then automatically deconvoluted using the settings shown in Table 3. Table 4 shows the expected uncapped, capping intermediate and fully capped species that may be present in each reaction.

As uncapped mRNA consists of a labile 5' terminal triphosphate group, some amount of hydrolyzed di- or monophosphate species3 might be observed depending on the reaction conditions. Figure 4 shows the extracted mass spectra and deconvoluted masses from triphosphate and diphosphate peaks. The optimized thermostable RNase-H cleavage protocol yielded only a small quantity (4.2%) of diphosphate and no detectable monophosphate species, suggesting that reaction conditions were suitably mild.

About 15% of triphosphate oligonucleotides in the uncapped sample were found to be sequence variants containing an additional non-templated +G nucleotide (Figure 4D). This presumably occurs due to T7 transcriptional slippage, which is commonly seen when repeating G nucleotides are present at the start of a transcribed sequence.[7] As seen in Figure 5, slippage sequence variants were also evident in mRNA samples co-transcriptionally capped with ARCA, both before and after a subsequent methylation reaction to Cap 1. Slippage sequence variants could be identified as distinct chromatographic peaks but were not baseline separated. All capped oligonucleotides were well-separated from uncapped oligonucleotides.

Table 4. Uncapped, capping intermediate, and capped species. Red letters indicate the 5' terminal nucleotide with phosphates marked by 'p', which may be elaborated or substituted by enzymatic reactions into capping structures.

Uncapped		Capping Intermediates		Capped Species	
Name	**Sequence**	**Name**	**Sequence**	**Name**	**Sequence**
Triphosphate	pppGGGGCC…	Diphosphate	ppGGGGCC…	ARCA Cap 0	3'-O-Me-m7GpppGGGGCC…
		G-Cap	GpppGGGGCC…	ARCA Cap 1	3'-O-Me-m7GpppmGGGGCC…
				Cap 0	m7GpppGGGGCC…

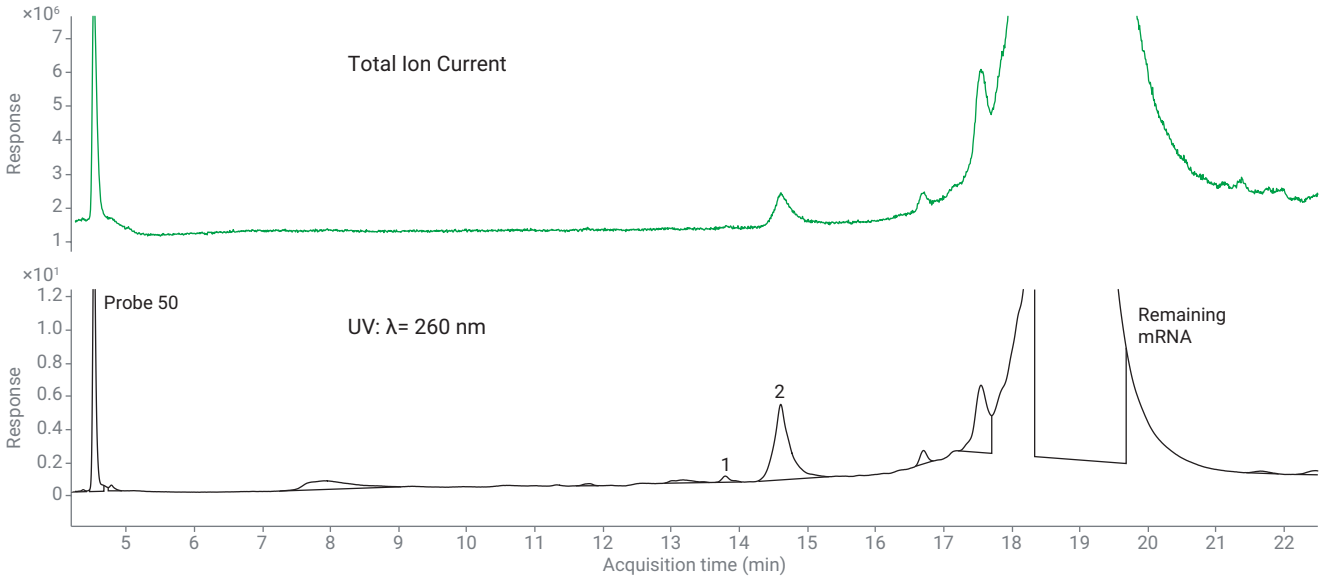

Figure 3. Chromatographic separation of 5' uncapped oligonucleotides from chimeric probe and sample matrix. Peak 1: 5' diphosphate, Peak 2: 5' triphosphate oligonucleotides.

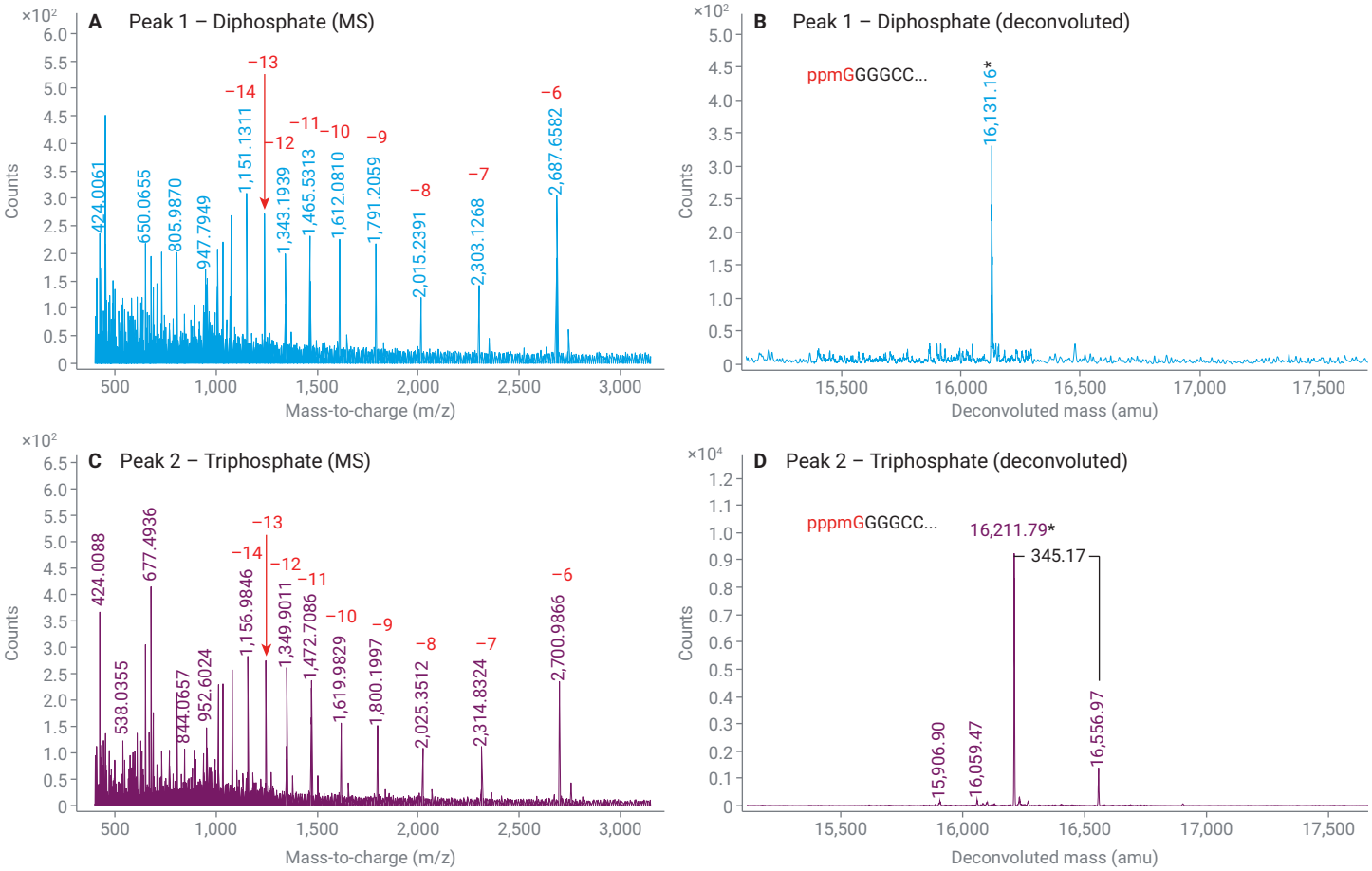

Figure 4. Mass spectra of 5' uncapped mRNA oligonucleotides. (A) Extracted and (B) deconvoluted mass spectra from Peak 1 of Figure 3. (C) Extracted and (D) deconvoluted mass spectra from Peak 2 of Figure 3. Numbers in red indicate the charge state of each peak in the extracted mass spectra arising from the diphosphate and triphosphate oligonucleotides. Mass peaks marked with asterisks (16131.16 Da and 16211.79 Da) were matched to their putative diphosphate and triphosphate identities (sequences inset) with <10 ppm mass error (Table 5). In panel D, triphosphate + G sequence variant (+345.17 Da) was observed co-eluting with triphosphate uncapped oligonucleotides.

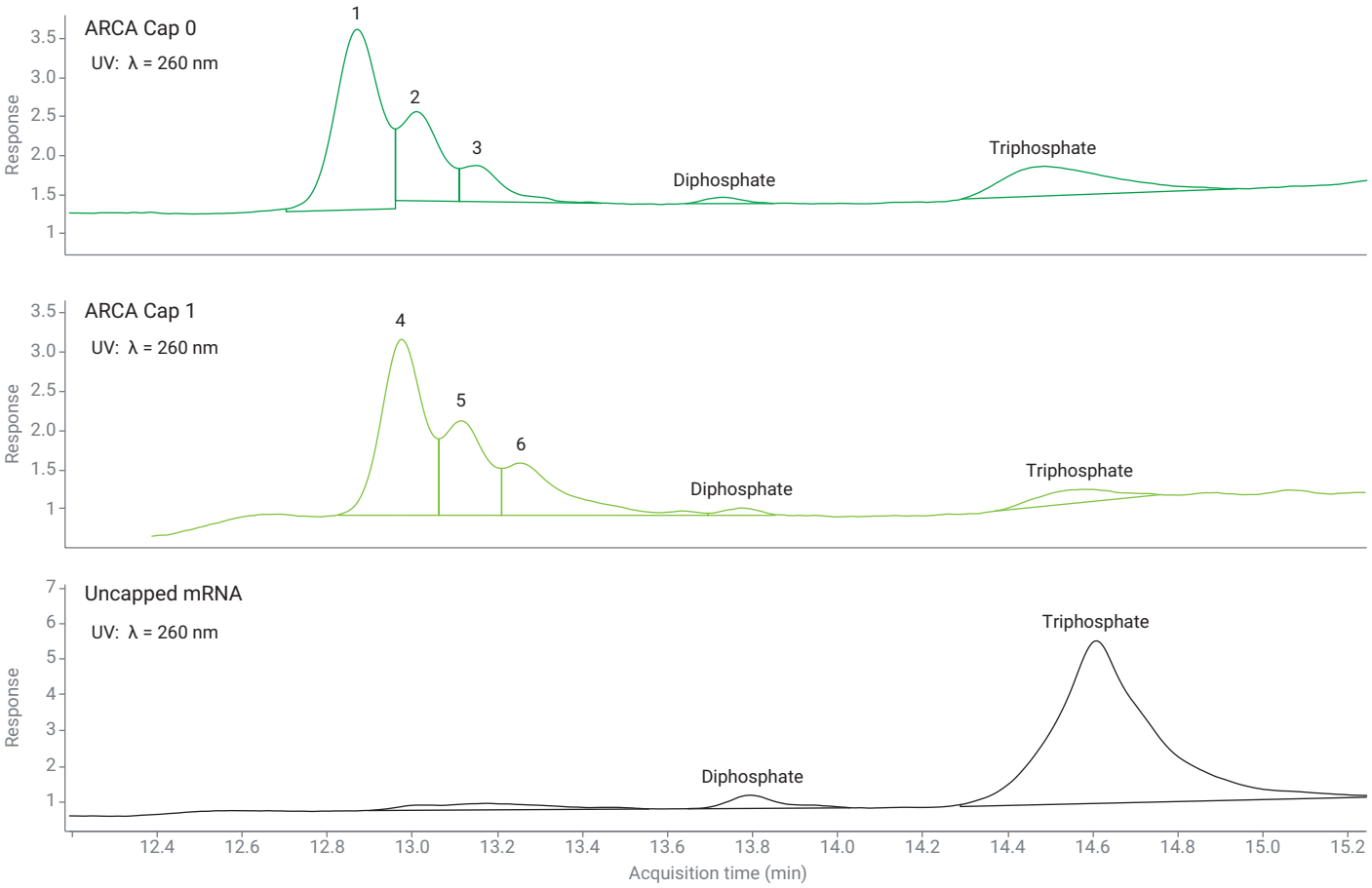

Figure 5. Separation of capped from uncapped oligonucleotides. Peaks 1 to 3: ARCA Cap 0, Peaks 4 to 6: ARCA Cap 1. Capped species contain 0 to 2 non-templated G nucleotides as slippage sequence variants.

Figure 6 shows the extracted mass spectra and deconvoluted masses from ARCA Cap 0 and ARCA Cap 1 peaks. Consistent with published literature[2], we found the ARCA Cap 0 co-transcriptional capping efficiency to be high (91.3 ±1.76%), but not complete. In samples subsequently methylated to ARCA Cap 1, the capping efficiency was not significantly different (90.6 ±1.53%) from ARCA Cap 0, indicating an essentially complete conversion of all Cap 0 structures to Cap 1.

Finally, the LC/MS method was used to assess the efficiency of Vaccinia enzymatic capping. Although this reaction has been claimed[3] to be more efficient than ARCA co-transcriptional capping and relatively independent of sample sequence, in our hands it proved to be otherwise. Using the manufacturer's recommended reaction conditions, Vaccinia enzymatic capping resulted in a mixture of uncapped material, capping intermediates, and fully capped sample (Figure 7 and Table 6). As shown in Figure 8, most of the sample (90.9 ±0.6%) remained uncapped with 5' terminal di- or triphosphate, and 6.5 ±0.7% was capped but unmethylated i.e. G-Cap. Only 2.6 ±0.5% of the sample was capped and methylated to Cap 0.

Table 5. Mass accuracy of ARCA-capped and capping intermediate oligonucleotides.

ARCA Co-Transcriptional Capping			
Identity	**Theoretical Mass**	**Observed Mass**	**Accuracy (ppm)**
ARCA Cap 0	16504.59	16504.62	2.01
ARCA Cap 0 + G	16849.80	16849.69	6.37
ARCA Cap 0 + 2G	17195.01	17195.03	1.50
ARCA Cap 1	16518.59	16518.85	15.77
ARCA Cap 1 + G	16863.80	16864.11	18.47
ARCA Cap 1 + 2G	17209.01	17208.73	16.13
Di-phosphate	16131.52	16131.37	9.21
Tri-phosphate	16211.51	16211.50	0.38
Tri-phosphate + G	16556.55	-	-

Table 6. Mass accuracy of enzymatically capped and capping intermediate oligonucleotides.

Enzymatic Capping			
Identity	**Theoretical Mass**	**Observed Mass**	**Accuracy (ppm)**
Cap 0	16490.59	16490.65	3.55
Cap 0 + G	16835.80	16836.00	11.97
G-Cap	16476.57	16476.65	4.85
Di-phosphate	16131.52	16131.49	1.95
Tri-phosphate	16211.51	16211.61	6.24
Tri-phosphate + G	16556.72	16556.97	15.33

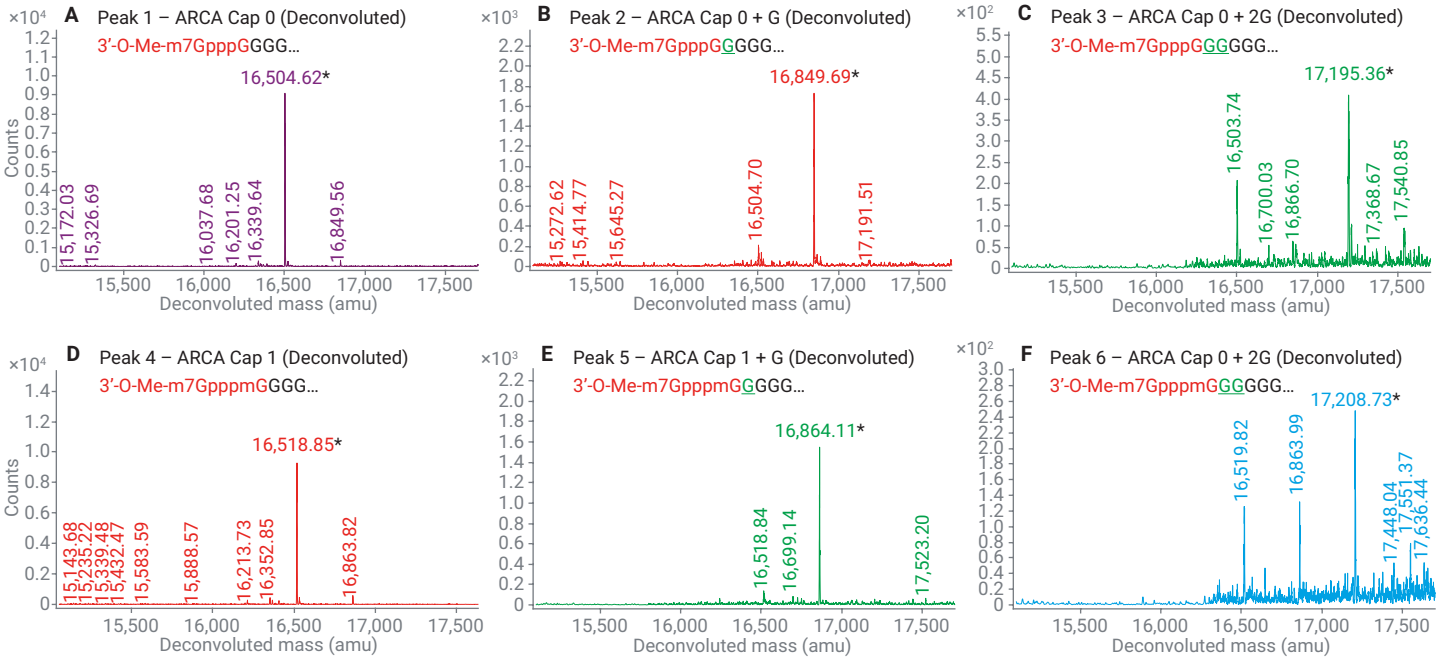

Figure 6. Mass spectra of ARCA Cap 0 mRNA oligonucleotides. (A – C) Deconvoluted mass spectra from Peaks 1 – 3 of Figure 5. (D – F) Deconvoluted mass spectra from Peaks 4 – 6 of Figure 5. Mass peaks marked with asterisks (16504.62 Da, 16849.69 Da, 17195.36 Da, 16518.85 Da, 16864.11 Da and 17208.73 Da) were matched to their putative identities (sequences inset) with < 20 ppm mass error (Table 5). Underlined letters in green indicate non-templated nucleotides likely due to T7 transcriptional slippage.

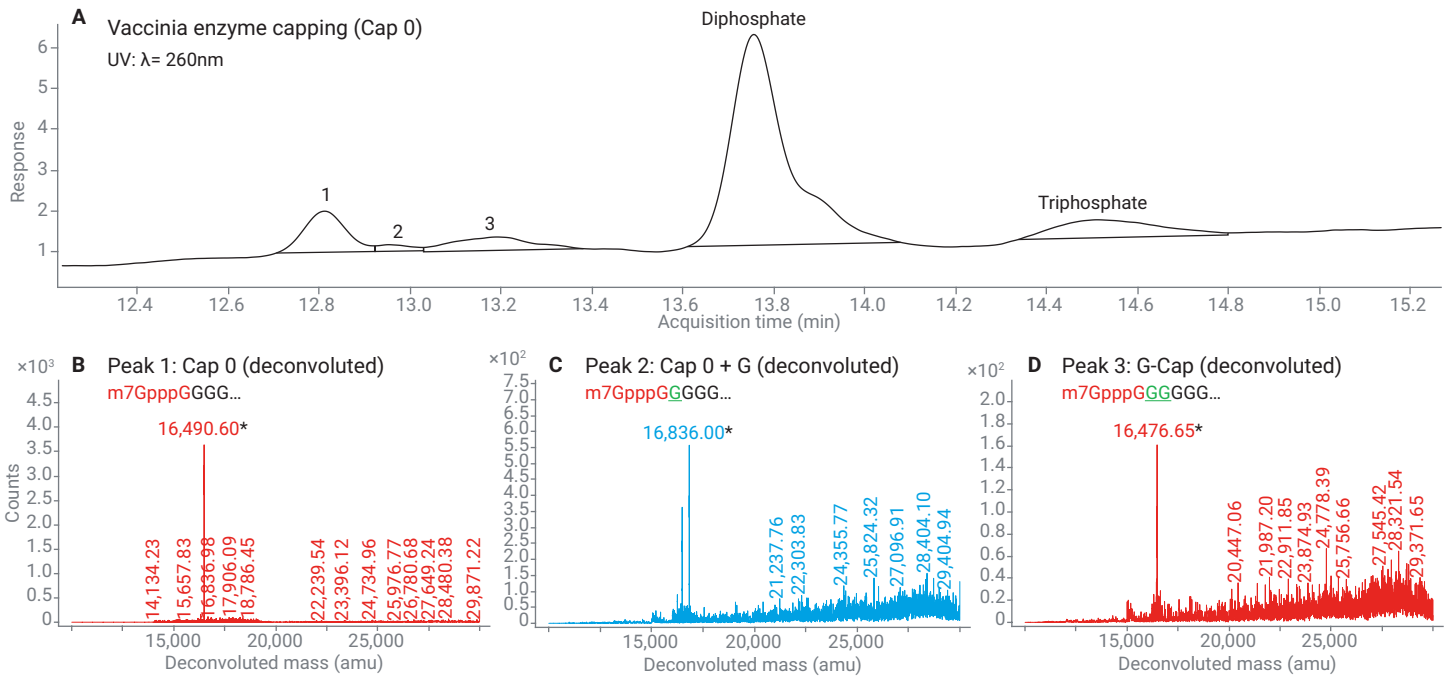

Figure 7. LC/MS of Vaccinia enzyme capped oligonucleotides and capping intermediates. (A) Peaks 1 to 2: Cap 0 and +G sequence variant, Peak 3: G-capped oligonucleotide, (B to D) Deconvoluted mass spectra of capped oligonucleotides. All mass peaks marked with asterisks (16,490.60, 16,836.00, and 16,476.65 Da) were matched to their putative identities (sequences inset) with <12 ppm (Table 6). Underlined letters in green indicate non-templated nucleotides likely due to T7 transcriptional slippage.

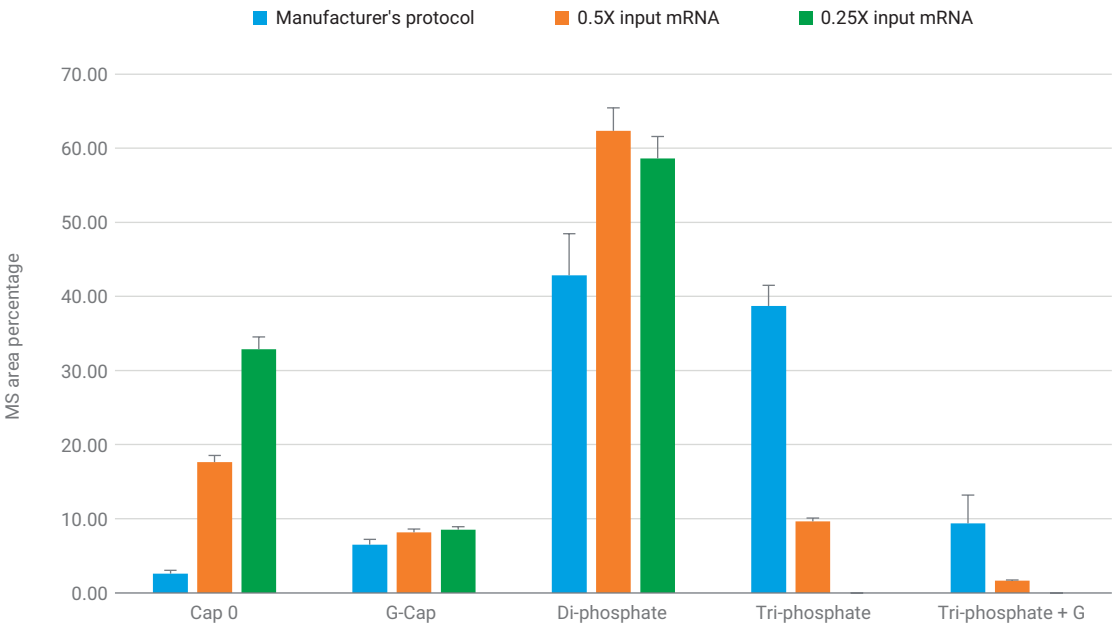

Figure 8. Optimization of Vaccinia enzymatic capping guided by LC/MS.

To increase capping efficiency, we reduced the quantity of input mRNA by 0.5x and 0.25x, thus raising the molar ratio of SAM, GTP and Vaccinia capping enzyme relative to mRNA. This strategy was moderately successful, resulting in a decrease in uncapped material (0.5x = 73.6 ±1.3% uncapped, 0.25x = 58.6 ±0.8% uncapped) and a corresponding increase in capped material (0.5x = 17.7 ±1.8% Cap 0, 0.25x = 32.9 ±0.4% Cap 0). Increasing the concentration of individual reactants in isolation did not increase capping efficiency (data not shown), suggesting that no individual component was faulty.

Conclusion

In this study we have developed a faster LC/MS method for quantifying mRNA 5' capping. One of the key enhancements in this workflow has been the use of thermostable RNase-H in conjunction with chimeric RNA/DNA oligonucleotide probes of sufficiently high GC content to expedite mRNA cleavage without a separate thermal annealing step. Probes with excessively high GC content may be unsuitable due to the increased likelihood of off-target binding or the formation of secondary structures. Indeed, we observed off-target cleavage with Probe 50 (67% GC content, Figure 2) when performing cleavage at 37 °C. As off-target binding is favored at lower reaction temperatures, performing cleavage at elevated temperatures resulted in cleaner samples in addition to hastening the reaction. In separate experiments, we found that probes with low GC content (40%) were unsuccessful at directing thermostable RNase-H cutting (data not shown).

We have demonstrated this workflow using two different capping methodologies: (1) co-transcriptional capping with ARCA, (2) enzymatic capping with Vaccinia capping enzyme. Although ARCA capping results agreed well with the literature, our experience with Vaccinia enzymatic capping was disappointing. Analysis of the sample sequence highlights a possible cause: the first 12 nucleotides have a very high GC content of over 90%, increasing the likelihood of secondary structure formation which may have interfered with Vaccinia enzymatic capping. As homopolymeric G repeats can also lead to inclusion of non-templated G nucleotides as sequence variants (Figure 5), these results indicate that redesigning the 5' UTR of this sequence could improve ease of processing and product quality.

www.agilent.com/chem

Printed in the USA, September 21, 2021
5994-3984EN
PR7000-3098

We note that a similar analytical approach was published recently by M.Beverly *et al.*[5],where chimeric probes were covalently conjugated to magnetic beads to enable simultaneous site-directed cutting by RNase-H and affinity enrichment of cleaved oligonucleotides. In our hands, this approach failed to direct specific cleavage of our sample despite numerous optimizations (data not shown). This could have been because the chimeric probes complementary to the sample's 5' terminus possessed very high GC content, thereby causing widespread nonspecific binding. We found our workflow to be more flexible as chimeric probes could be designed to avoid challenging regions in the sample. Moreover, sample cleanup using silica-based spin columns is likely to be less sequence-dependent than affinity-based approaches, and will probably lead to greater sample recovery.

In conclusion, this application note highlights the utility of the Agilent 6545XT AdvanceBio LC/Q-TOF for high-resolution, rapid analysis of mRNA 5' capping. We anticipate this workflow to be useful in both drug development and quality control laboratories.

References

1. Hassett, K. J. *et al.* Optimization of Lipid Nanoparticles for Intramuscular Administration of mRNA Vaccines. *Mol. Thera. Nucleic Acids* **2019**, *15*, 1–11.
2. Stepinski, J. *et al.* Synthesis and Properties of mRNAs Containing the Novel "Anti-Reverse" Cap Analogues 7-methyl(3'-O-methyl)GpppG and 7-methyl(3'-deoxy) GpppG. *RNA* **2001**, *7*, 1486–1495.
3. Fuchs, A. L. *et al.* A General Method for Rapid and Cost-Efficient Large-Scale Production of 5' Capped RNA. *RNA* **2016**, *22*, 1454–1466.
4. Lapham, J. *et al.* Site-Specific Cleavage of Transcript RNA. *Methods in Enzymology* **2000**, *317*, 132–139.
5. Beverly, M. *et al.* Label-Free Analysis of mRNA Capping Efficiency using RNase H Probes and LC-MS. *Anal. Bioanal. Chem.* **2016**, *408*, 5021–5030.
6. AbouHaidar, M. G. *et al.* Non-Enzymatic RNA Hydrolysis Promoted by the Combined Catalytic Activity of Buffers and Magnesium Ions. *Z Naturforsch C J Biosci* **1999**, *54*, 542–548.
7. T.Conrad *et al.* Maximizing Transcription of Nucleic Acids with Efficient T7 Promoters. *Nat. Comm. Biol.* **2020**, *3*, 1–8.

Agilent
Trusted Answers