Application Note

Pharma & Biopharma

Agilent

Trusted Answers

Analysis of mRNA Poly-A Sequence Variants by High-Resolution LC/MS

Author

Brian Liau
Agilent Technologies, Inc.

Introduction

The urgency engendered by the SARS-CoV-2 pandemic of 2020 has prompted policy makers and pharmaceutical firms alike to develop and deploy mRNA vaccines with unprecedented speed. mRNA vaccines have shown impressive safety and efficacy in clinical trials[1-4], outperforming vaccines based on alternative technologies. As mRNA vaccines are considered gene therapies[5], FDA guidance requires extensive characterization of product-related impurities. These may include populations of mRNA molecules with slight errors in their sequence, known as sequence variants. In addition, mRNA vaccines require lengthy, repetitive sections of A nucleotides (poly-A) at the 3' terminus for optimal stability and biological activity.[6] Both the length and the composition of poly-A sequences are therefore critical quality attributes.

This work uses an Agilent AdvanceBio 6545XT LC/Q-TOF to analyze poly-A tail sequences formed by *E. coli* Poly-A Polymerase (PAP), which is a common component of *in vitro* transcription systems. Findings show that PAP is not fully selective for ATP, and can act on both CTP and UTP precursors to incorporate significant quantities of undesirable C and U nucleotides under standard *in vitro* transcription conditions. As these sequence variants can be regarded as product-related impurities, the results caution against the use of PAP and show the value of LC/MS as a sensitive and efficient method for process optimization and quality control of nucleic acid therapies.

Abbreviations used in this work:

- ATP – adenosine triphosphate
- CTP – cytidine triphosphate
- UTP – uridine triphosphate
- GTP – guanosine triphosphate
- A, C, U, and G nucleotides – adenosine, cytidine, uridine, and guanosine monophosphate
- Poly-A – polyadenosine
- PAP – *E. coli* Poly-A Polymerase
- RNA-seq – RNA sequencing

Experimental

In-vitro transcription of mRNA

A pCMV3 plasmid encoding a 3822 nt gene flanked by an upstream T7 promoter and a downstream BGH terminator sequence was purchased from Sino Biological. The DNA sequence was PCR amplified for 35 cycles using T7 and BGH terminator primers (Agilent Herculase II Fusion DNA Polymerase, part number 600677). After cleanup (Agilent StrataPrep PCR Purification kit, part number 400771), the amplified dsDNA was analyzed on an Agilent 2100 Bioanalyzer with a DNA 7500 kit (part number 5067-1506) to measure its concentration and to assess the uniformity of amplification. The amplified dsDNA (~13 nM) was then transcribed *in vitro* using a HiScribe T7 ARCA mRNA Kit (New England Biolabs, part number E2060S) and tailed with the included PAP enzyme using the manufacturer's recommended protocol, then precipitated with LiCl. Aliquots of transcribed mRNA before and after PAP tailing were analyzed on a 2100 Bioanalyzer with an RNA 6000 Nano kit (part number 5064-1511) to monitor the reaction.

For PAP selectivity studies, a synthetic 10-mer poly-A sequence with 5' and 3'-OH (Integrated DNA Technologies) was extended with PAP enzyme using only one precursor nucleoside triphosphate per reaction (1 mM of either ATP, CTP, UTP, or GTP) for 30 minutes at 37 °C, as illustrated in Figure 1A.

Sample preparation

Approximately twenty picomoles of poly-A tailed mRNA was digested with 1,000 U of RNase T1 for 3 hours at 37 °C to liberate poly-A sequences. Each sample was subjected to five rounds of purification using 200 µL of oligo-dT magnetic beads to pull down poly-A sequences.[7] Each pull-down was eluted in 50 µL of 1x IDTE buffer (Integrated DNA Technologies, part number 11-05-01-05) and pooled into a final volume of 250 µL. Prior to LC/MS analysis, the pooled eluate was desalted into 60 µL of deionized water using Vivaspin 500 cartridges with 10 kDa MWCO (Sartorius, part number VS0102).

Poly-A tailing experiments

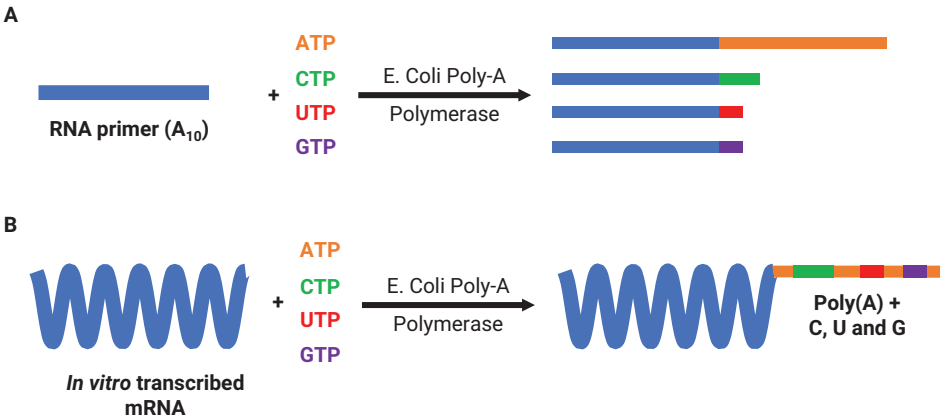


Figure 1. Schematic of tailing reactions performed in this application note. (A) Reactions on RNA primer carried out with only one precursor per reaction. (B) Reactions carried out on *in vitro* transcribed mRNA under standard conditions (all precursors).

LC-DAD/MS of poly-A sequences

Instrumentation consisted of:

- 1290 Infinity II LC with diode array detector (P/N 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 first use, the LC system and column were flushed with a 50% MeOH + 0.1% formic acid solution overnight to further reduce alkaline metal adducts. If required, a 30-minute flush with 50% MeOH + 0.1% formic acid was usually enough to clean the system between experiments.[8]

Poly-A sequences were separated on a PLRP-S column (2.1 × 50 mm, 5 µm, 1,000 Å, part number PL1912-1502). To achieve higher chromatographic resolution, an Infinity Poroshell 120 HPH-C18 column (2.1 × 50 mm, 1.9µm, 120 Å, part number 699675-702) was used in PAP selectivity experiments. The mobile phase and LC gradients are shown in Table 1. The mass spectrometer was operated in negative ion mode with settings in Table 2, and data analysis was performed in BioConfirm 10.0 with deconvolution settings in Table 3.

Table 1. Mobile phase and LC gradients.

Agilent 1290 Infinity II LC System		
Column	InfinityLab Poroshell 120 HPH-C18, 1.9 µm, 2.1 × 50 mm,120 Å	Agilent PLRP-S, 5 µm, 2.1 × 50 mm, 1,000 Å
Solvent A	15 mM dibutylamine + 25 mM HFIP in DI water	
Solvent B	15 mM dibutylamine + 25 mM HFIP in methanol	
Gradient	0 to 2 min, 15% B 12 min, 30% B 12.1 to 13 min, 90% B	0 to 1 min, 15% B 10.5 min, 45% B 10.6 to 11.5 min, 90% B
Column Temperature	50 °C	80 °C
Flow Rate	0.4 mL/min	
Injection Volume	10 to 20 µL	

Table 2. Mass spectrometer settings.

Agilent 6545XT AdvanceBio LC/Q-TOF		
	LC/MS	LC/MS/MS
Acquisition Mode	Negative, standard (3,200 *m/z*) mass range, high sensitivity (2 Ghz)	
Gas Temperature	350 °C	
Gas Flow	12 L/min	
Nebulizer	55 psig	
Sheath Gas Temperature	275 °C	
Sheath Gas Flow	10 L/min	
Vcap	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	5 Hz
MS2 Range	N/A	50 to 3,200 *m/z*
MS2 Scan Rate		3 Hz
MS2 Isolation Width		Medium (~4 amu)
Collision Energy		0, 40, 60 V
Threshold for MS2		On; 3 repeat then exclude for 0.2 min
Precursor Abundance Based Scan Speed		Yes
Target (Counts/Spectrum)		25,000
Use MS2 Accumulation Time Limit		Yes
Purity		100% stringency, 30% cutoff
Sort Precursors		By abundance only; +3, +2, +1
Reference Mass		1,033.9881

Table 3. Deconvolution settings.

Agilent MassHunter BioConfirm B10.0 Settings		
Oligonucleotide Length	≤30 nt	≥90 nt
Deconvolution Algorithm	Maximum Entropy	
Subtract Baseline	1	
Adduct	Proton loss	
Mass Range	3,000 to 10,000 Da	30,000 to 60,000 Da
Mass Step	0.05 Da	0.05 Da
Use Limited *m/z* Range	1,040 to 3,200	800 to 2,500

Results and discussion

This first test analyzed poly-A sequences extended by PAP on a synthetic RNA primer consisting of 10 repeated A nucleotides (A_{10}) in the presence of 1 mM ATP. As shown in Figure 2 this resulted in a bimodal distribution of poly-A sequences, with one population consisting of shorter oligonucleotides eluting from 2.5 to 6 minutes, and another consisting of longer nucleotides eluting in a broad peak at ~10.6 minutes. Mass spectrometric analysis indicated the shorter population ranged in size from 11 to 22 nt (Figure 3D), whereas the longer population ranged from 108 to 149 nt in length (35,492.89 to 48,990.17 Da, Figure 4C).

Extracted and deconvoluted mass spectra from three selected peaks from the shorter oligonucleotide population are shown in Figure 3. Mass spectra consisted primarily of doubly and triply charged ions generated through proton loss, as well as minor populations of sodium adducts. Isotopically resolved deconvoluted mass spectra were assigned identities of A_{20}, A_{21}, and A_{22} (Figures 3B to 3D) with <5 ppm error based on the respective monoisotopic peaks.

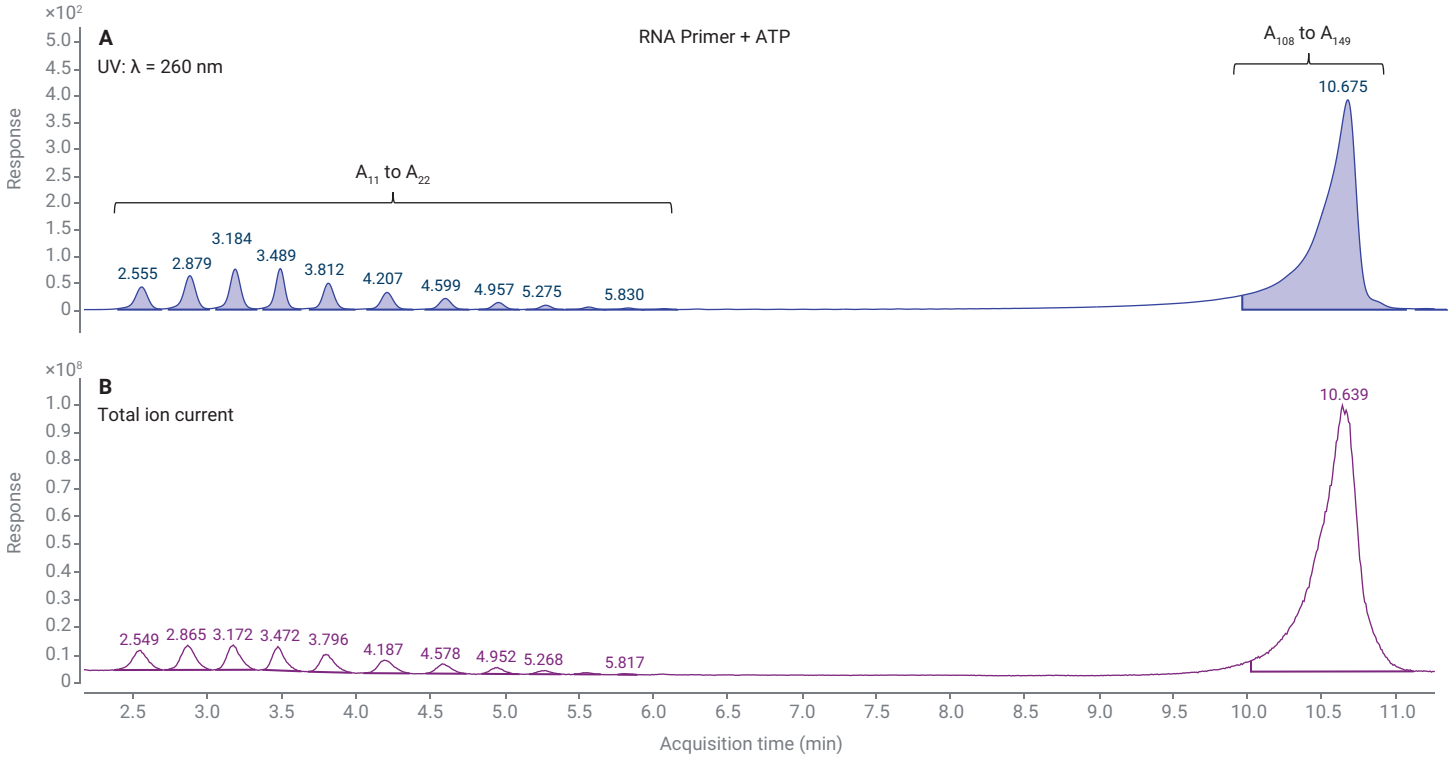


Figure 2. UV absorbance at 260 nm (A: reference = 360 nm) and total ion chromatogram (B) of RNA primers extended with PAP in the presence of only ATP. Separation was carried out on a PLRP-S column.

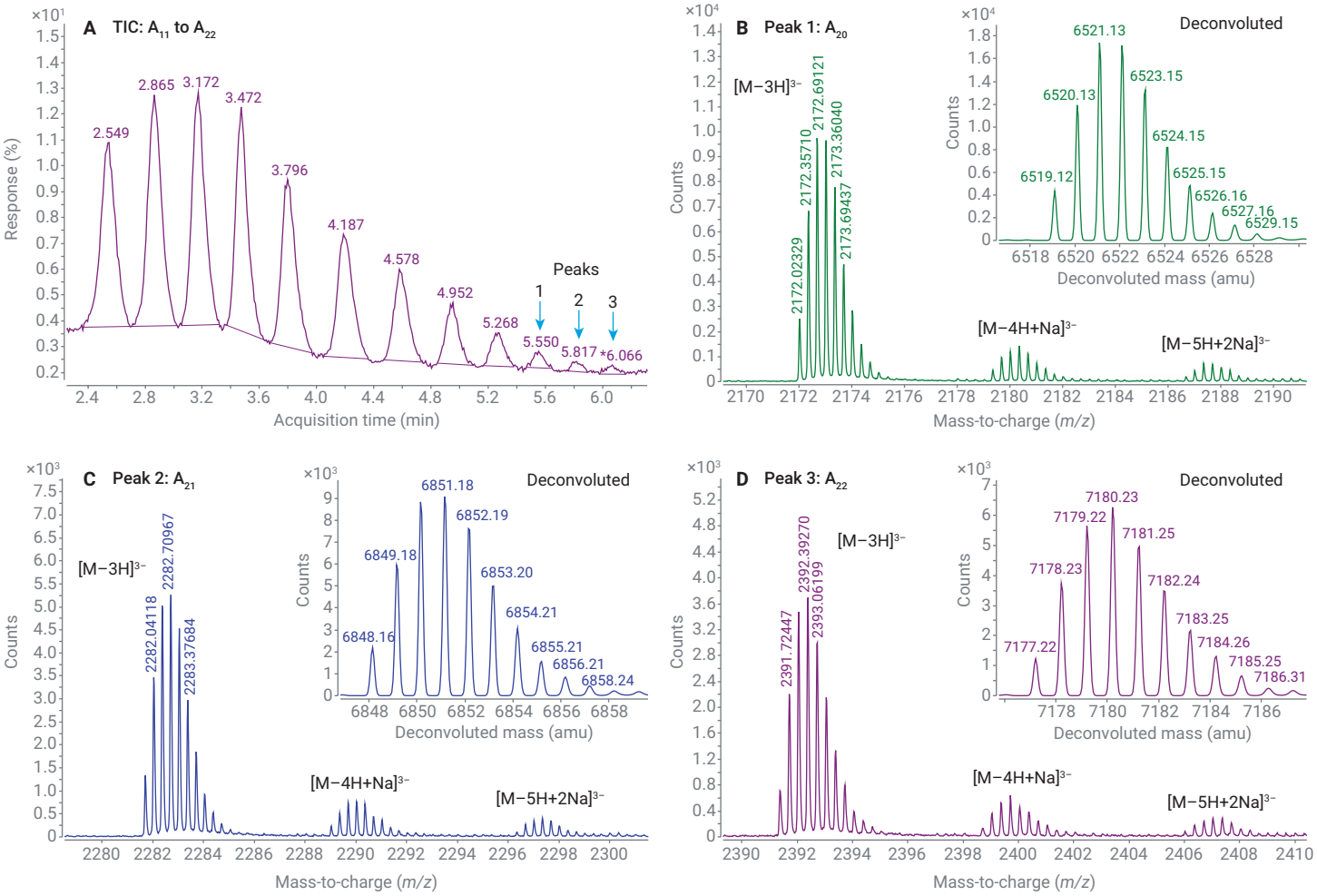


Figure 3. A_{11} to A_{22} oligonucleotides formed by PAP. (A) Total ion current chromatogram showing the three selected peaks with extracted mass spectra shown in (B to D). Deconvoluted mass spectra of A_{20} (M_{obs} = 6,519.12 Da, M_{theo} = 6,519.09 Da), A_{21} (M_{obs} = 6,848.16 Da, M_{theo} = 6,848.15 Da), and A_{22} (M_{obs} = 7,177.22 Da, M_{theo} = 7,177.20 Da) are shown as insets. M_{obs}: observed monoisotopic mass; M_{theo}: theoretical monoisotopic mass.

A portion of the longer oligonucleotides was sampled for deconvolution (Figure 4A). The charge envelope in the extracted mass spectra from 10 to 10.3 minutes primarily fell between 800 to 2,500 *m/z* (Figure 4B) and was deconvoluted to a destination mass range of 30 to 60 kDa. The deconvoluted mass spectra (Figure 4C) clearly showed a heterogenous population of sample peaks from 34 to 50 kDa, which were evenly separated by 329.2 ±1 Da (Figure 4D). These mass increments were consistent with single additions of A nucleotides, increasing the theoretical average mass by 329.209 Da. Table 4 shows that mass peaks in Figure 4D were confidently annotated as A_{121} to A_{138} with differences between theoretical and observed masses ≤1.16 Da.

To assess the selectivity of PAP for ATP, duplicate experiments were conducted where PAP was added to the RNA primer in the presence of only 1mM CTP, UTP, or GTP. Although the extension of long polymeric chains were not observed, chromatographically resolved additions of up to two monomers of C nucleotides (Figure 5A) or one U nucleotide (Figure 5B) to the RNA primer, indicating that PAP was not wholly selective for ATP. The addition of guanosine monophosphate was not observed in this experiment (Figure 5C) but could not rule out the possibility that appreciable quantities might be added with longer reaction times or higher GTP concentrations. Overall, PAP showed the highest activity with ATP, followed by CTP, UTP, and GTP in descending order.

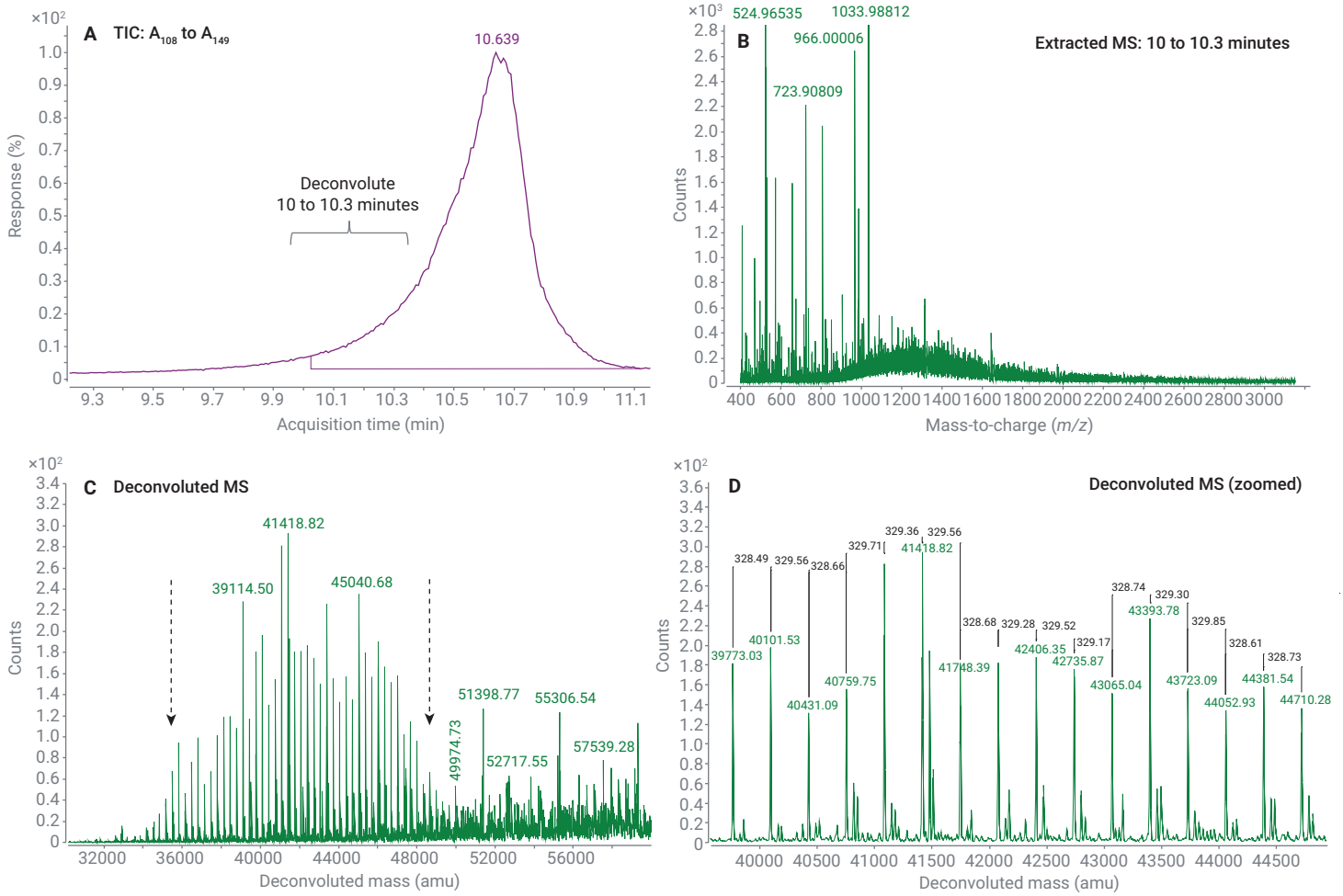
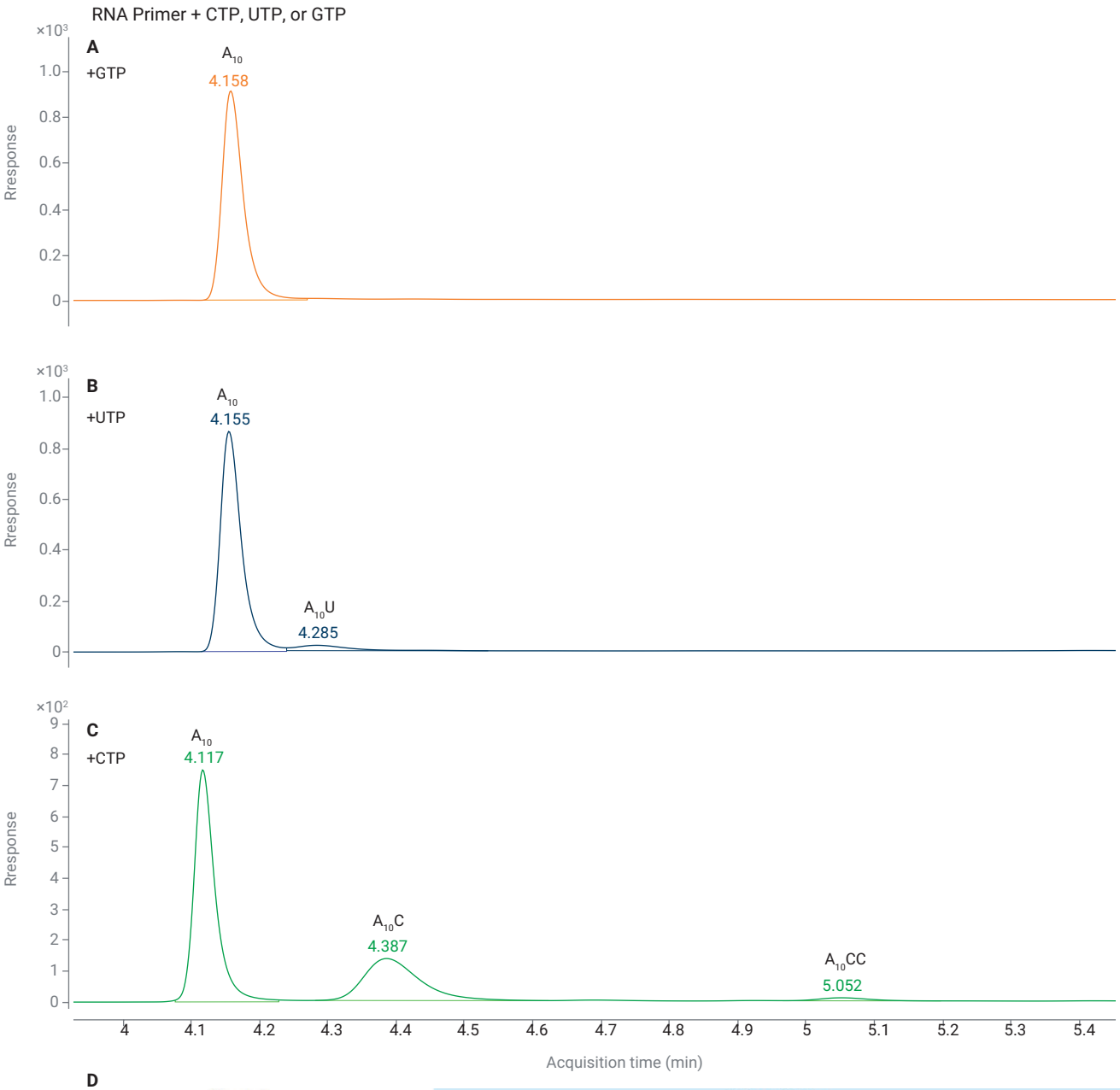


Figure 4. A_{108} to A_{149} oligonucleotides formed by PAP. (A) Total ion current chromatogram showing the region 10 to 10.3 minutes sampled for deconvolution. (B) Charge envelope and (C) Deconvoluted mass spectrum of sampled region. Dashed arrows (left = 35,492.89 Da, right = 48,990.17 Da) indicate the range of mass peaks that could be confidently assigned identities A_{108} to A_{149}. (D) Enlarged deconvoluted mass spectrum showing regular intervals between peaks from 39,773.03 to 44,710.28 Da.

Table 4. Annotated mass peaks from Figure 4D.

Oligonucleotide	Observed Mass (Da)	Theoretical Mass (Da)	Mass Difference (Da)
A_{121}	39,773.03	39,772.08	0.95
A_{122}	40,101.53	40,101.28	0.25
A_{123}	40,431.09	40,430.49	0.6
A_{124}	40,759.75	40,759.70	0.05
A_{125}	41,089.46	41,088.90	0.56
A_{126}	41,418.82	41,418.11	0.71
A_{127}	41,748.39	41,747.32	1.07
A_{128}	42,077.06	42,076.52	0.54
A_{129}	42,406.35	42,405.73	0.62
A_{130}	42,735.87	42,734.94	0.93
A_{131}	43,065.04	43,064.15	0.89
A_{132}	43,393.78	43,393.35	0.43
A_{133}	43,723.09	43,722.56	0.53
A_{134}	44,052.93	44,051.77	1.16
A_{135}	44,381.54	44,380.97	0.57
A_{136}	44,710.28	44,710.18	0.1



D

General				% Quantitation				
Mass	RT	Height	Area	Use for %Quant	Area (MS)	%Quant (Area)	Height (MS)	%Quant (Height)
3228.59365	3.998	1575689	6629732	☑	3250164	61.36	1575689	65.19
3533.6394	4.281	760517	5139482	☑	1896979	35.82	760517	31.47
3838.681	4.938	80709	3274129	☑	149355	2.82	80709	3.34

Figure 5. UV absorbance (Abs = 260 nm, Ref = 360 nm) chromatograms showing promiscuous activity of PAP towards (A) GTP, (B) UTP, and (C) CTP. No addition of guanosine monophosphate was detected. (D) Relative quantitation of A_{10}, $A_{10}C$, and $A_{10}CC$ as shown in panel (C). Separation was carried out on an Agilent Poroshell 120 HPH-C18 column.

The deconvoluted mass spectra of the unmodified RNA primer and those extended with C or U nucleotides are shown in Figure 6. Isotopically resolved deconvoluted mass spectra were assigned identities of A_{10}, $A_{10}C$, $A_{10}CC$ and $A_{10}U$ with <13 ppm error based on the respective monoisotopic peaks. MS/MS experiments showed that C and U nucleotides were indeed added to the 3' terminus of the RNA primer (Figure 7), resulting in the formation of characteristic doubly charged y-ions 1601.271 *m/z* and 1601.758 *m/z*. In contrast, the unmodified RNA primer was terminated with a 3' A nucleotide, yielding a doubly charged y-ion 1448.749 *m/z* upon fragmentation.

Next, full-length, *in vitro* transcribed mRNA were analyzed on a Bioanalyzer equipped with RNA 6000 Nano kit and by LC/MS. Before tailing, transcribed mRNA showed the expected length of ~3,800 nt which increased to ~4,200 nt after reaction with PAP (Figure 8A), indicating that successful poly-A tailing had been achieved.

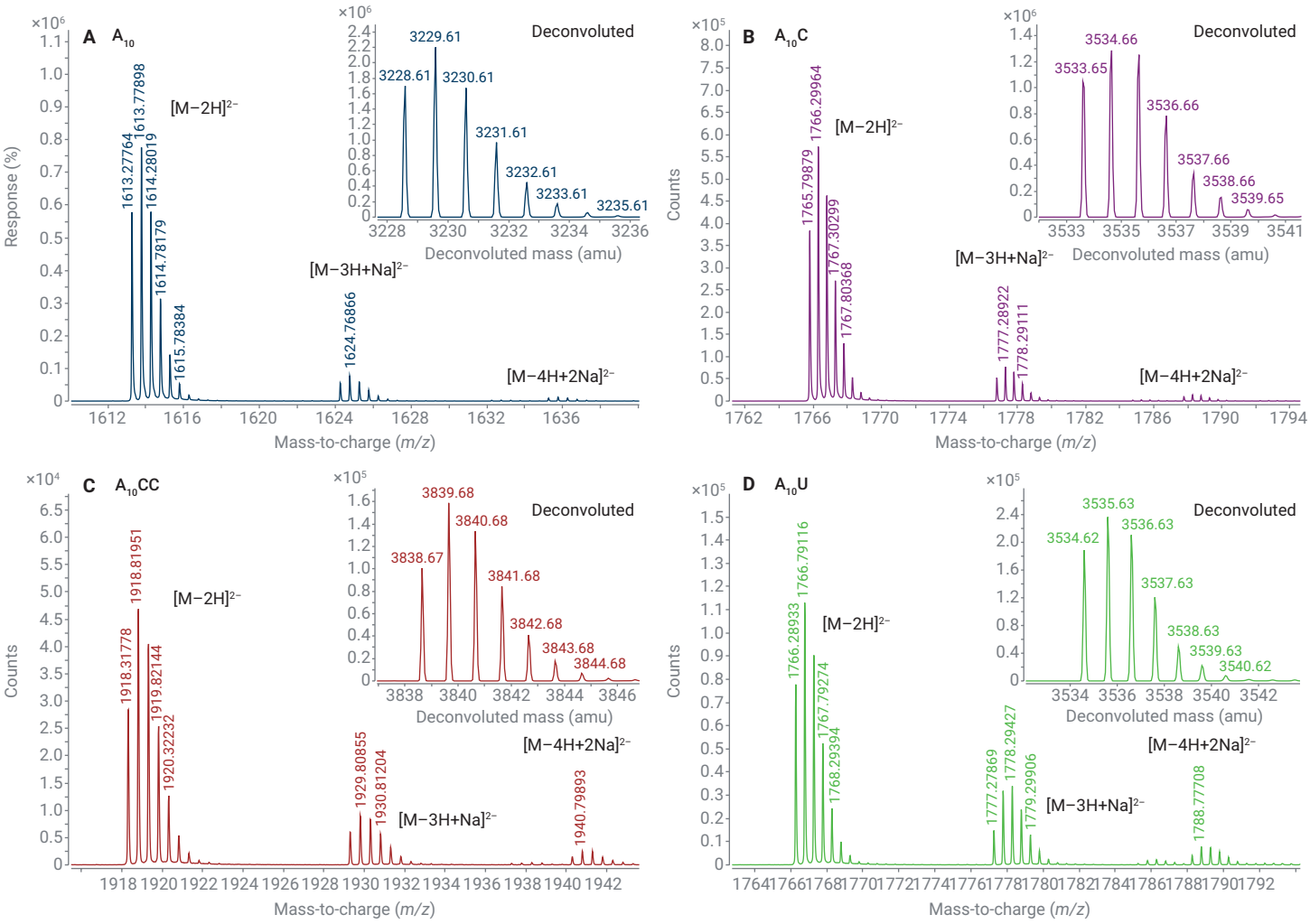


Figure 6. Extracted and deconvoluted mass spectra of (A) Unmodified A_{10} RNA primer (M_{obs} = 3,228.61 Da, M_{theo} = 3,228.57 Da), (B) Extended with one C nucleotide (M_{obs} = 3,533.65 Da, M_{theo} = 3,533.61 Da), (C) Extended with two C nucleotides (M_{obs} = 3,838.67 Da, M_{theo} = 3,838.65 Da), (D) Extended with one U nucleotide (M_{obs} = 3,534.62 Da, M_{theo} = 3,534.59 Da). M_{obs}: observed monoisotopic mass; M_{theo}: theoretical monoisotopic mass.

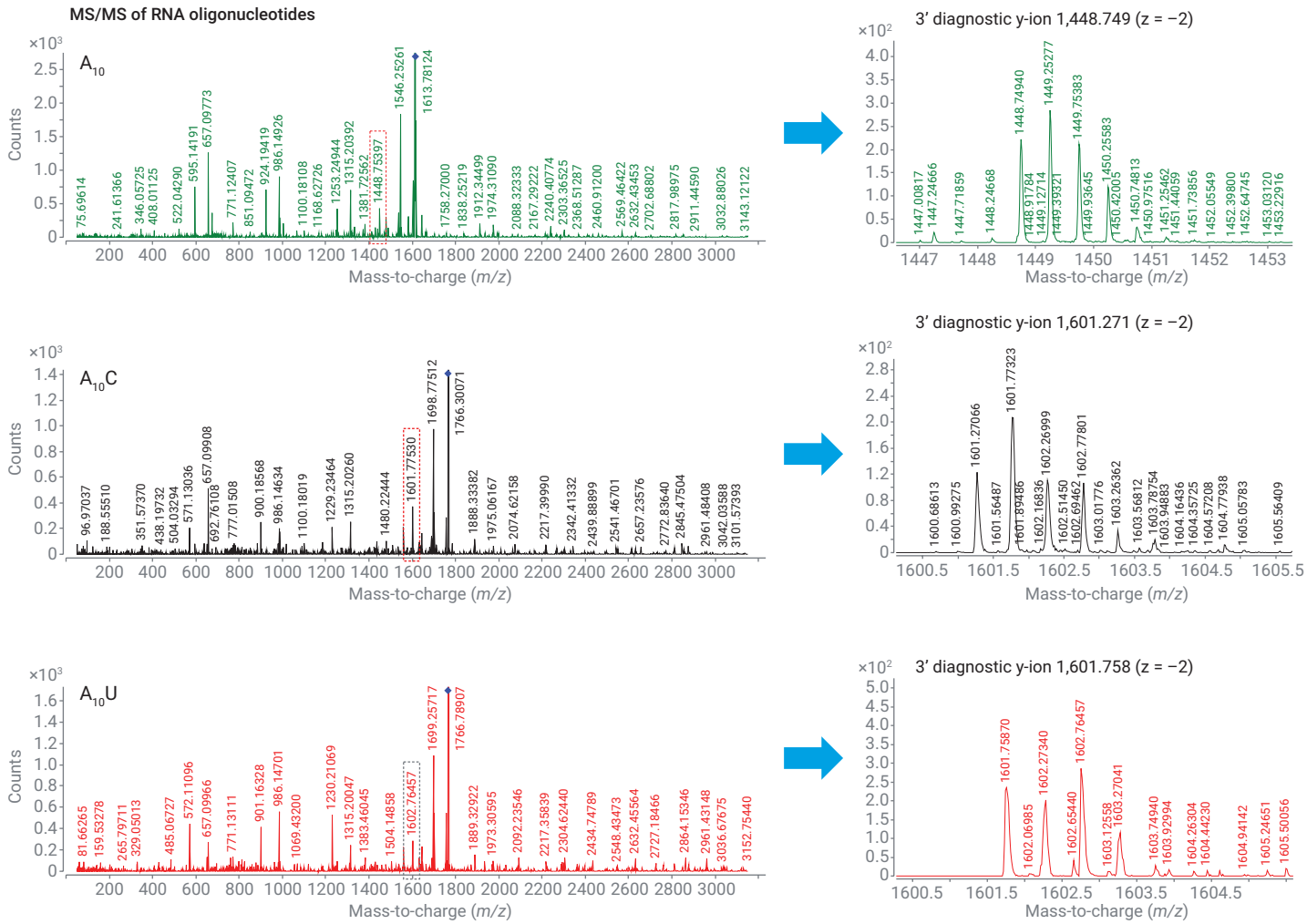


Figure 7. MS/MS of selected oligonucleotides showing diagnostic ions characteristic of their different 3' termini.

Full length mRNA samples were digested with RNase T1, followed by repeated pull-downs with oligo dT magnetic beads to yield purified tail sequences. As with PAP-extended RNA primers, tail sequences derived from *in vitro* transcribed mRNA consisted of both a shorter population of oligonucleotides eluting between 3.7 to 7.5 minutes and a longer population eluting ~10.6 minutes (Figure 8B). Extracted and deconvoluted mass spectra from selected peaks in the shorter population revealed poly-A sequences ranging in length from 16 to 27 nt, with each containing a single misincorporated U nucleotide (Figure 9). Although not seen in this dataset, misincorporated C nucleotides were also observed in other experiments.

As noted by M. Beverly *et al.*[7], tail sequences formed by PAP are considerably more heterogenous in length as compared to genetically templated poly-A sequences. The results indicate that this heterogeneity is compounded by the misincorporation of differing numbers of C and U nucleotides when the tailing reaction takes place under standard conditions with all four precursor nucleoside triphosphates present, making the mass spectra of longer tail sequences very challenging to deconvolute.

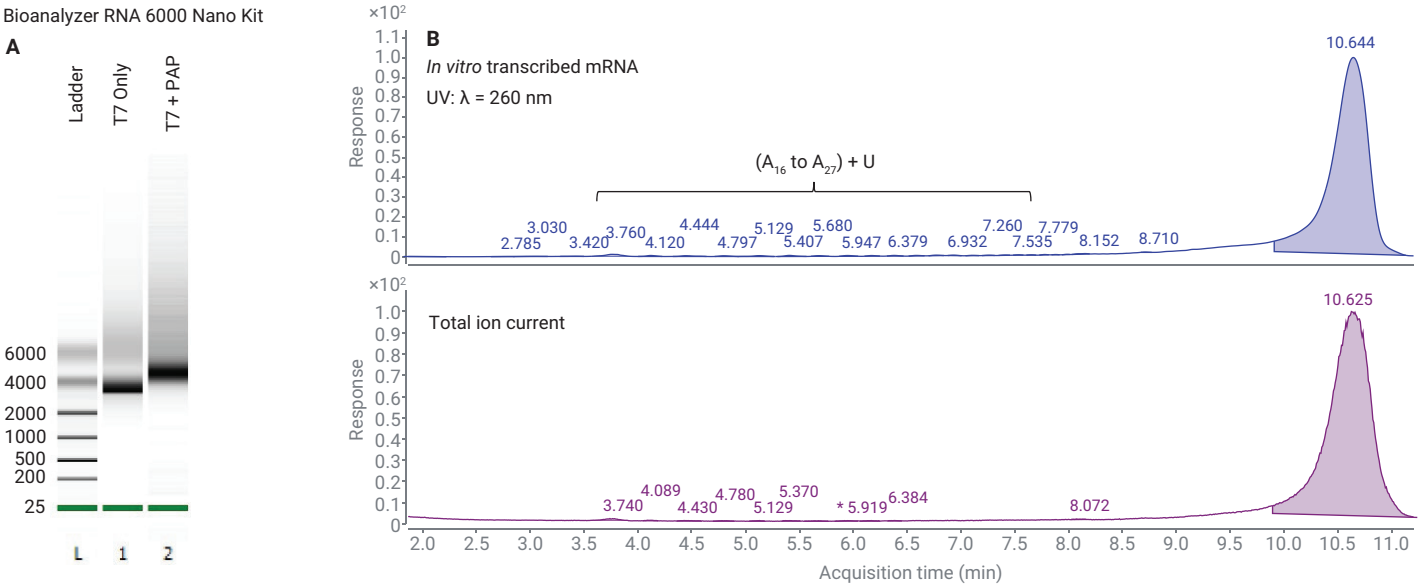


Figure 8. (A) Bioanalyzer analysis of *in vitro* transcribed mRNA before (lane 1) and after (lane 2) tailing with PAP. (B) UV absorbance at 260 nm (top panel) and total ion chromatogram (bottom panel) of poly-A sequences appended to *in vitro* transcribed mRNA in the presence of all four nucleoside phosphate precursors. Separation was carried out on a PLRP-S column.

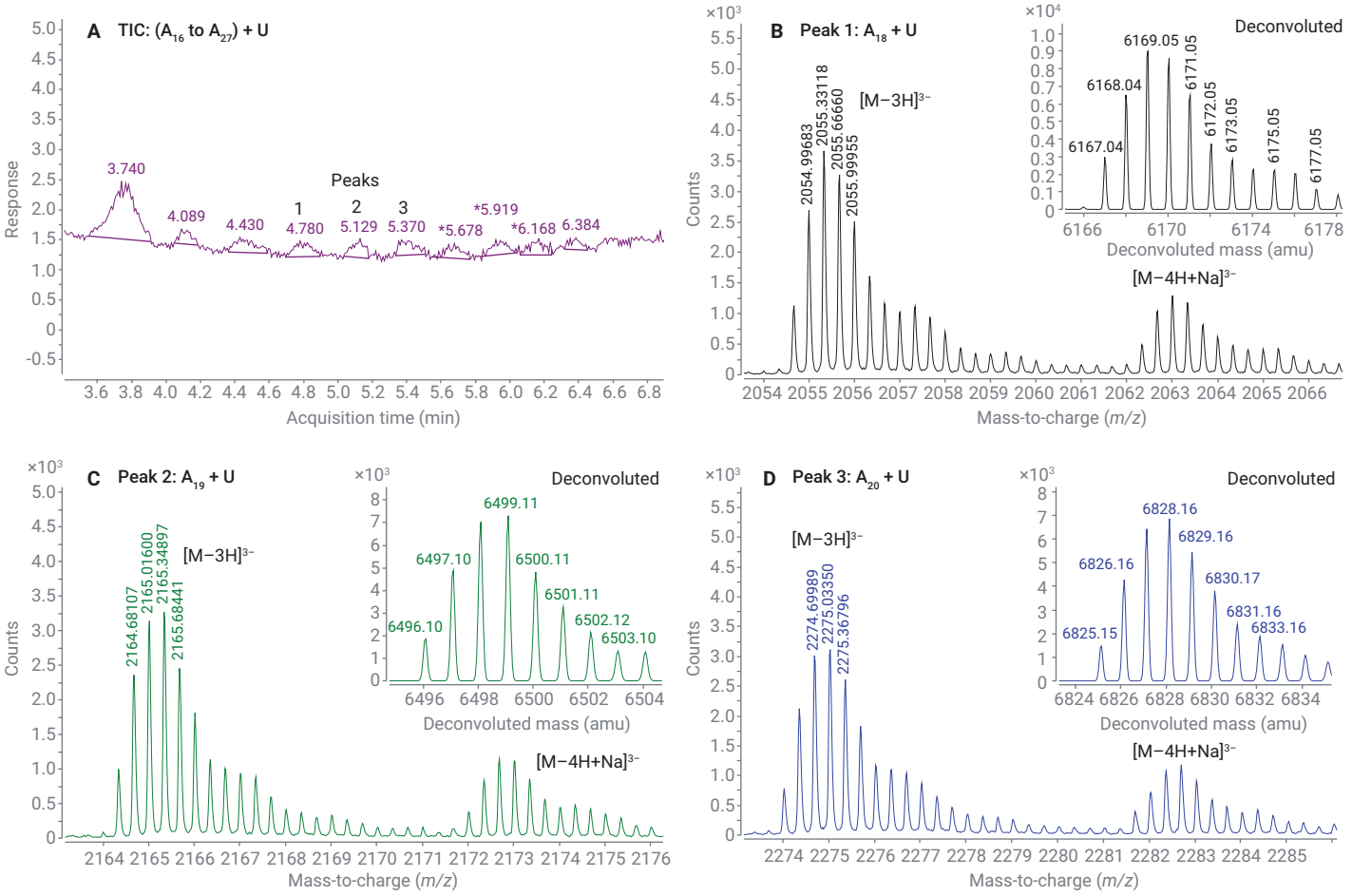


Figure 9. A_{16} to A_{27} oligonucleotides with misincorporated U nucleotide. (A) Total ion current chromatogram showing three selected peaks with extracted mass spectra shown in (B to D). Deconvoluted mass spectra of A_{18} + U (M_{obs} = 6,167.04 Da, M_{theo} = 6,167.01 Da), A_{19} + U (M_{obs} = 6,496.10 Da, M_{theo} = 6,496.07 Da), and A_{20} + U (M_{obs} = 6,825.15 Da, M_{theo} = 6,825.12 Da) are shown as insets. M_{obs}: observed monoisotopic mass; M_{theo}: theoretical monoisotopic mass.

Conclusion

This study shows that: (1) the intact masses of long (121 to 136 nt), heterogenous poly-A sequences can be accurately measured by deconvolution of their ensemble mass spectra, and (2) PAP is not fully selective for ATP under standard *in vitro* transcription conditions, causing both C and U nucleotides to be added to poly-A tail sequences.

Although these sequence variants may be inconsequential for *in vitro* studies, they are highly significant from a regulatory standpoint. Notably, other *in vitro* transcription enzymes such as T7 polymerase may also produce sequence variants through mechanisms such as slippage or transcriptional arrest[9], underscoring the need for highly sensitive and selective methods for detecting these impurities.

To achieve such sensitivity and selectivity, one prior study demonstrated PAP's off-target activity by using radiolabeled nucleotides.[10] Such techniques can be hazardous and are ill-suited for production environments. LC/MS can achieve single-nucleotide selectivity without the need for such reagents. Moreover, LC/MS can detect and quantify sequence variants without the lengthy reverse transcription, ligation and amplification steps characteristic of RNA-seq, which are known to introduce biases and artifacts.

References

1. Jackson, L. A. *et al.* An MRNA Vaccine against SARS-CoV-2 — Preliminary Report. *N. Engl. J. Med.* **2020**, *383*, 1920–1931.
2. Mulligan, M. J. *et al.* Phase I/II Study of COVID-19 RNA Vaccine BNT162b1 in Adults. *Nature* **2020**, *586*, 589–593.
3. Pfizer and BioNTech Announce Vaccine Candidate Against COVID-19 Achieved Success in First Interim Analysis from Phase 3 Study.
4. Promising Interim Results from Clinical Trial of NIH-Moderna COVID-19 Vaccine.
5. Wadhwa, A. *et al.* Opportunities and Challenges in the Delivery of MRNA-Based Vaccines. *Pharmaceutics* **2020**, *12*, 102.
6. Kahvejian, A. *et al.* Mammalian Poly (A)-Binding Protein Is a Eukaryotic Translation Initiation Factor, Which Acts via Multiple Mechanisms. *Genes Dev.* **2005**, *19*, 104–113.
7. Beverly, M.; Hagen, C.; Slack, O. Poly A Tail Length Analysis of *In Vitro* Transcribed MRNA by LC/MS. *Anal. Bioanal. Chem.* **2018**, *410*, 1667–1677.
8. Birdsall, R. E. *et al.* Reduction of Metal Adducts in Oligonucleotide Mass Spectra in Ion-Pair Reversed-Phase Chromatography/Mass Spectrometry Analysis. *Rapid Commun. Mass Spectrom.* **2016**, *30*, 1667–1679.
9. Tateishi-Karimata, H.; Isono, N.; Sugimoto, N. New Insights into Transcription Fidelity: Thermal Stability of Non-Canonical Structures in Template DNA Regulates Transcriptional Arrest, Pause, and Slippage. *PLOS ONE* **2014**, *9*.
10. Yehudai-Resheff, S.; Schuster, G. Characterization of the E.Coli Poly(A) Polymerase: Nucleotide Specificity, RNA-Binding Affinities and RNA Structure Dependence. *Nucleic Acids Res.* **2000**, *28*, 1139–1144.

www.agilent.com/chem

RA44231.5456365741

Printed in the USA, March 8, 2021
5994-3005EN

Agilent
Trusted Answers