

mRNA Mapping with IP-RPLC-MS on a Low-Adsorption Flow Path

Suitable for Agilent
1290 Infinity III LC

Using an Agilent 1290 Infinity II Bio LC System

Authors

Helena Vanluchene,
Kris Morreel, Jelle De Vos,
Pat Sandra, and Koen Sandra
RIC group, President
Kennedypark 6, 8500 Kortrijk,
Belgium

Sonja Schneider and
Udo Huber
Agilent Technologies, Inc.

Abstract

As increasingly more messenger ribonucleic acid (mRNA)-based medicines are developed, new methods emerge to analyze the attributes of these biopharmaceuticals. One such method is mRNA mapping, where RNases are used to generate oligonucleotides that are amenable to LC/MS using IP-RPLC or HILIC. The latter methodology offers valuable insights into primary structural properties such as mRNA sequence, 3' poly A tail length and distribution, 5' capping structure/capping efficiency, posttranscriptional modifications, and sequence variants. Oligonucleotide measurement, however, is biased by the adverse effects of the stainless-steel (SS) surfaces in the LC/MS flow path. To mitigate these effects, this application note demonstrates IP-RPLC-MS analyses of mRNA digests on a low-adsorption flow path using an Agilent Altura Oligo HPH-C18 column with Ultra Inert technology and an Agilent 1290 Infinity II Bio LC System with an Agilent 6530 LC/Q-TOF.

Introduction

Vaccination technology based on mRNA (see Figure 1) became known worldwide due to the coronavirus disease 19 (COVID-19) pandemic, which was eventually confined when regulatory agencies such as the U.S. Food and Drug Administration (FDA) and the European Medical Agency (EMA) approved the global administration of the Comirnaty (BioNTech/Pfizer) and Spikevax (Moderna) vaccines.^{1,2} The mode of operation relies on the mRNA, encapsulated in lipid nanoparticles (LNPs), which upon entering the cell is translated into the spike protein of the severe acute respiratory syndrome coronavirus 2 (SARS-COV-2) virus, thereby triggering a targeted immune response.^{3,4} The acceptance of these landmark vaccines has subsequently paved the way for full exploration of the mRNA technology.⁵⁻⁷

As illustrated in Figure 1, mRNA is composed of phosphodiester-linked ribonucleosides, e.g., adenosine (A), guanosine (G), cytosine (C) and uridine (U), in which the different nucleobases are attached to a ribose moiety. mRNA is rather fragile because the ribosyl 2'-hydroxy function might readily hydrolyze the neighboring phosphodiester linkage, either purely chemically or catalyzed by a ribonuclease (RNase). Digestion by these RNases might occur at the termini (exoribonucleases) or at internal sites (endoribonucleases) of the RNA sequence. Therefore, the open reading frame (ORF, Figure 1) of mRNA, which encodes the protein of interest, is protected by upstream 5'-terminal and downstream 3'-terminal untranslated regions (5'-UTR and 3'-UTR, respectively). These 5'- and 3'-UTRs are additionally protected by bearing a cap and a poly A tail, respectively (Figure 1), which are also prerequisites to enable translation. The eukaryotic cap structure consists of a 7-methylguanosine which is 5'-5'-linked to the next nucleotide by a triphosphate moiety. The poly A tail, flanking the 3'-UTR, is a long chain of adenine nucleotides. Its creation typically yields a mRNA pool displaying a distribution of poly A tail lengths.^{8,9}

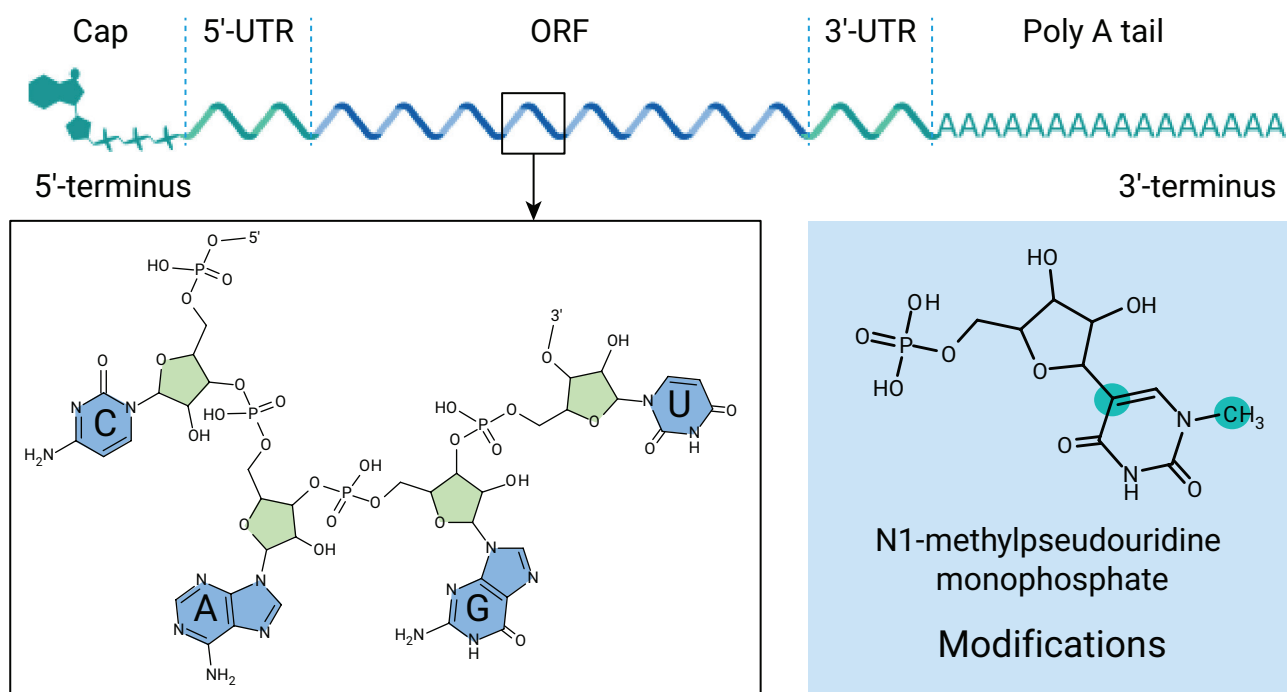


Figure 1. An mRNA construct and its constituting nucleotides. The RNA building blocks, called ribonucleotides, consist of a ribose moiety (green background) to which a nucleobase is attached (blue background; nucleobase/ribonucleoside: A adenine/adenosine, C cytosine/cytidine, G guanine/guanosine, and U uracil/uridine) and that are connected through phosphodiester linkages. Modified nucleosides such as N1-methylpseudouridine are commonly introduced during in vitro transcription to enhance mRNA stability, reduce immunogenicity, and improve protein translation. ORF = open reading frame; UTR = untranslated region.

The cellular uptake of foreign mRNA readily triggers the immune system, which can be mitigated by the introduction of modified nucleosides into the default mRNA structure (Figure 1).^{10,11} The most prevalent modification includes the replacement of uridine by pseudouridine or N1-methyl-pseudouridine, as is present in the Comirnaty and Spikevax vaccines. Furthermore, mRNA modifications might unintentionally occur during synthesis and storage resulting from oxidation or lipid adduct formation, thereby rendering the mRNA untranslatable.^{13,14} Noticeably, Katalin Kariko and Drew Weissman were awarded the Nobel prize in 2023 for their discovery of the increased stability and decreased anti-RNA immune response of pseudouridine-containing RNA.¹²

In mRNA-based drug development, it is of crucial importance to meticulously study structural features such as mRNA sequence, 3' poly A tail length and distribution, 5' capping structure and capping efficiency, posttranscriptional modifications, and sequence variants.¹⁵ Liquid chromatography (LC) with mass spectrometry (MS) is gaining attention for assessing these attributes.¹⁶⁻²⁶ However, successful LC/MS analysis requires RNase-mediated digestion of the mRNA to oligonucleotides.²³ The predominant LC method for oligonucleotide separation is ion pairing-reversed phase LC (IP-RPLC) using either hexylamine or triethylamine (TEA), and hexafluoroisopropanol (HFIP) as ion-pair reagent (IPR) and counterion, respectively.²⁷ This IP-RPLC method yields a higher MS sensitivity compared to hydrophilic interaction chromatography (HILIC), regarded as the alternative and more sustainable option for LC/MS analysis of oligonucleotides.²⁴

To operate at high pressures, the typical LC hardware (instrument, column, and tubing) consists of stainless-steel (SS), which may pose significant challenges when working with highly negatively charged oligonucleotides. Indeed, the phosphate moieties of oligonucleotides adsorb to the positively charged metal oxide surfaces and interact with leached metal-ions retained by the stationary phase, manifesting a low analyte recovery and peak tailing.²⁸ Leaching metals might be dealt with by adding metal chelators to the mobile phase such as ethylenediaminetetraacetic acid (EDTA) or medronic acid. To avoid adsorption at the SS surface, passivation of the SS surface is routinely performed, either by using mobile phase additives such as citrate or phosphate, or by using the oligonucleotides themselves, i.e., by conditioning the column with repeated injections of the sample.²⁹ Even though passivation improves the peak shape and recovery

of oligonucleotides, it might neither fully cover all active sites at the SS surface nor provide a sufficiently long-term passivation. Furthermore, the use of mobile phase additives might result in a lower MS response due to ion suppression. A superior alternative is the implementation of low adsorption and biocompatible LC flow paths from which SS is eliminated (e.g. MP35N) or metal surfaces deactivated or covered with, for example, polyether ether ketone (PEEK).

This application note describes the IP-RPLC-MS-based mRNA mapping of RNase T1- and RNase 4-digested IVT-mRNA on a low adsorption flow path using an Altura Oligo HPH-C18 column with Ultra Inert technology and biocompatible 1290 Infinity II Bio LC System in combination with a 6530 Q-TOF MS. The Altura column was compared with an AdvanceBio Oligonucleotide column, a conventional stainless-steel HPLC column packed with the same high-pH-tolerant C18 stationary phase.

Experimental

Materials

Water (HPLC-grade; shipped in polyethylene water bottles) and methanol (ULC/MS – CC/SFC) were supplied by Biosolve. TEA and HFIP were sourced from Sigma-Aldrich and Thermo Fisher Scientific. RNase T1 and RNase 4 were purchased from New England Biolabs. Tris(hydroxymethyl)aminomethane (Tris) pH 7.5 buffer and ethylenediaminetetraacetic acid (EDTA) were obtained from Sigma-Aldrich. RNA resolution standard (part number 5190-9028) was received from Agilent Technologies, and CleanCap Firefly Luciferase (FLuc) mRNA from TriLink BioTechnologies. The length of FLuc mRNA and its ORF are 1,922 and 1,653 nucleotides (nt), respectively.

Sample preparation

The RNA resolution standard was dissolved in 1 mL of water. Prior to RNA mapping, FLuc mRNA was enzymatically digested by either RNase T1 or RNase 4. To create the RNase T1 digestion mixture, 100 µg of FLuc mRNA was digested with 5,000 U of RNase T1 in a 100 mM Tris pH 7.5 buffer with 40 mM EDTA for 30 minutes at 37 °C with shaking at 300 rpm. To prepare the RNase 4 digestion mixture, 100 µg of FLuc mRNA was first denatured by mixing with 3 M urea, followed by incubation at 90 °C for 10 minutes. After the sample was cooled to room temperature, digestion proceeded with 500 U RNase 4 in 1x NEBuffer r1.1 for 1 hour at 37 °C with shaking at 300 rpm. Finally, both resultant digests were concentrated up to 3 µg/µL using a refrigerated Centrивap Concentrator (Labconco).

Instrumentation

Samples were run on a 1290 Infinity II Bio LC System consisting of an Agilent 1290 Infinity II Bio High-Speed Pump (G7132A), an Agilent 1290 Infinity II Bio Multisampler (G7137A) with integrated sample thermostat, an Agilent 1290 Infinity II Multicolumn Thermostat (G7116B) with Agilent InfinityLab Quick Connect Bio heat exchanger, standard flow (G7116-60071), and an Agilent 1290 Infinity II DAD (G7117B) with Agilent InfinityLab Max-Light Cartridge Cell, LSS, 10 mm (G7117-60020). LC/MS was performed by hyphenating the 1290 Infinity II Bio LC to a 6530 LC/Q-TOF (G6530A). Method parameters are summarized in Tables 1 and 2.

Table 1. LC method parameters.

Parameter	Value												
Columns	Agilent AdvanceBio Oligonucleotide, 2.1 × 150 mm, 2.7 µm (p/n 653750-702) Agilent Altura Oligo HPH-C18 with Ultra Inert technology, 2.1 × 150 mm, 2.7 µm (p/n 227215-702)												
Flow Rate (Method 1)	0.35 mL/min												
Flow Rate (Methods 2 and 3)	0.2 mL/min												
Mobile Phase (Method 1)	A) 200 mM HFIP, 15 mM TEA in water B) Methanol												
Mobile Phase (Methods 2 and 3)	A) 1% HFIP, 0.1% TEA in water B) 1% HFIP, 0.1% TEA in water/methanol (50/50, v/v)												
Gradient (Method 1)	<table> <tr> <th>Time (min)</th><th>%B</th></tr> <tr> <td>0</td><td>3</td></tr> <tr> <td>15</td><td>25</td></tr> <tr> <td>18</td><td>100</td></tr> </table>	Time (min)	%B	0	3	15	25	18	100				
Time (min)	%B												
0	3												
15	25												
18	100												
Gradient (Method 2)	<table> <tr> <th>Time (min)</th><th>%B</th></tr> <tr> <td>0</td><td>1</td></tr> <tr> <td>5</td><td>1</td></tr> <tr> <td>20</td><td>21</td></tr> <tr> <td>60</td><td>50</td></tr> <tr> <td>65</td><td>100</td></tr> </table>	Time (min)	%B	0	1	5	1	20	21	60	50	65	100
Time (min)	%B												
0	1												
5	1												
20	21												
60	50												
65	100												
Gradient (Method 3)	<table> <tr> <th>Time (min)</th><th>%B</th></tr> <tr> <td>0</td><td>1</td></tr> <tr> <td>5</td><td>1</td></tr> <tr> <td>120</td><td>50</td></tr> <tr> <td>125</td><td>100</td></tr> </table>	Time (min)	%B	0	1	5	1	120	50	125	100		
Time (min)	%B												
0	1												
5	1												
120	50												
125	100												
Injection (Method 1)	1 µL												
Injection (Methods 2 and 3)	2 µL												
Needle Wash	Flush port, 100% methanol												
Autosampler Temperature	8 °C												
Column Temperature	65 °C												
Detection DAD	260/4 nm, reference 360/40 nm												

Table 2. MS method parameters for LC/MS analysis using LC methods 2 and 3.

Parameter	Value
Ionization	Negative ESI
Source*	
Drying Gas Temperature	300 °C
Drying Gas Flow	8 L/min
Sheath Gas Temperature	350 °C
Sheath Gas Flow	8 L/min
Nebulizer Pressure	35 psi
Capillary Voltage	4,500 V
Nozzle Voltage	1,000 V
Fragmentor	250 V
Skimmer	65 V
Acquisition – MS	
Acquisition Mode	Extended dynamic range (2 GHz)
Mass Range	<i>m/z</i> 500 to 3,200
Scan Rate	3 spectra/sec
Data Storage	Both profile and centroid
Reference Mass	Disabled
Acquisition – MS/MS	
Acquisition Mode	Auto MS/MS
Mass Range	<i>m/z</i> 100 to 3,200
Scan Rate	1 spectrum/sec
Isolation Width (MS/MS)	Medium (~ 4 amu)
Collision Energy	40 (Fixed)
Auto MS/MS Preferred Table	Included (delta retention time: 4 min)
Max Precursors per Cycle	4
Active Exclusion Enabled	Yes
Active Exclusion Excluded After	2 Spectra
Active Exclusion Released After	0.30 min

* Diverter valve was bypassed to prevent potential nonspecific interactions with stainless-steel components of the valve hardware.

Data analysis

Data were acquired and processed in Agilent OpenLab CDS software (version 2.7), Agilent MassHunter Data Acquisition software (version B.10.00), and Agilent MassHunter Qualitative Analysis software (version B.12.00). Later software versions can be used as well. MassHunter Qualitative Analysis facilitated maximum entropy deconvolution and enabled extracting the smaller and larger sized molecular features by combining information from the MS spectral ions representing different charge states and isotopes of the same oligonucleotide. Further data processing steps were performed in R version 4.2.1 (R Core Team, 2022) including (1) matching the average or monoisotopic molecular weight (MW) of each feature with the corresponding mRNA subsequence(s), taking the cleavage site of the RNase into account, and (2) using the recorded MS/MS spectrum of a feature to annotate the correct subsequence in case multiple isomeric candidate subsequences match the MW of the feature.

Results and discussion

As shown in a previous application note describing the HILIC-MS analysis of 5'-terminal mRNA fragments²⁴, oligonucleotides are readily adsorbed onto the SS surfaces of the LC column, and to a lesser degree, the LC instrument, leading to lower recovery and tailed peak shapes. To evaluate the extent that IP-RPLC-MS analysis of oligonucleotides is similarly affected by metal adsorption, the Agilent RNA resolution standard (consisting of four oligonucleotides: 14-, 17-, 20-, and 21-mer), was repeatedly injected onto an SS-based RPLC column mounted into an 1290 Infinity II Bio LC System using TEA and HFIP as IPR and counterion, respectively. As shown in Figure 2A, the UV absorption response of the four RNA standards gradually increased, plateauing from the eighth injection onwards, suggesting that metal adsorption occurs in IP-RPLC. The hyperbolically changing RNA standard abundances across the subsequent injections arise because the metal surface becomes increasingly passivated with each injection, leading to less RNA adsorption onto the metal surface and, eventually, to a steady state detector response after eight injections of the RNA standard, corresponding to 0.08 pmol of RNA injected.

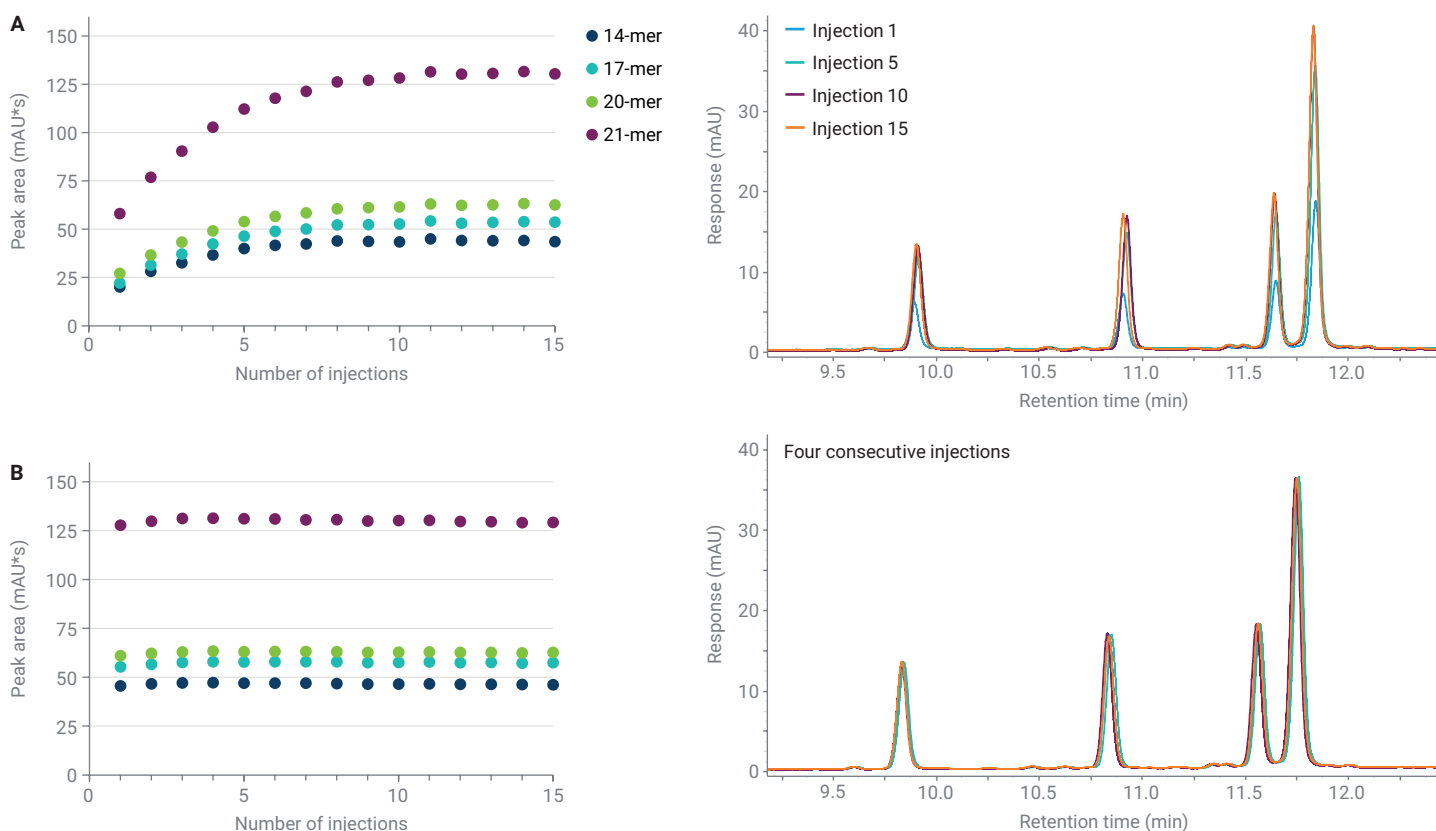


Figure 2. IP-RPLC-UV chromatograms of repeated RNA resolution standard injections using the SS Agilent AdvanceBio Oligonucleotide (A) and the inert Agilent Altura Oligo HPH-C18 (B) column (LC method 1: see Table 1).

The peak abundance of the 21-mer is approximately double those of the three other RNA standards throughout the injections, which is in agreement with the specifications of the RNA standard that states that the amount of the 21-mer is twice the amount of the 14-, 17-, or 20-mer. In contrast with HILIC analysis²⁴, hardly any peak tailing was observed for the oligonucleotides during IP-RPLC analysis, which is due to the partial neutralization of the RNA molecules through phosphodiester-TEA ion-pair formation. Further support for metal adsorption of oligonucleotides in a conventional SS column during IP-RPLC analysis was obtained by analyzing the RNA standard using an Altura RPLC column with Ultra Inert technology (same C18 media). Here, consecutive injections of the standard yielded a stable UV absorption response from the first injection onwards (Figure 2B). By comparing the detector response upon injection of the standard in an LC flow path that either does or does not comprise an LC column, the oligonucleotide amount recovered from the column or the amount that remained on the column can be estimated. When using the Ultra Inert and conventional SS HPLC columns, 4.2 and 6.2% of the total amount of oligonucleotides did not elute from the

column at steady state conditions. The difference between both columns (2.0%) represents the retained proportion of oligonucleotides due to metal adsorption.

The symmetrical peak shapes, good oligonucleotide recovery, and precision achieved across multiple injections support the use of the Altura column for IP-RPLC-MS analyses of complex oligonucleotide mixtures. This study used such a mixture by digesting Fluc mRNA using either RNase T1 or RNase 4. Whereas the former cleaves 3' of G units, the latter recognizes UG and UA dimeric sites and hydrolyzes the intermediate phosphodiester linkage. Both digestions yield a 5'-terminal fragment containing a 3'-terminal phosphate or a cyclic 2',3'-phosphodiester. When profiling both digests with IP-RPLC, the RNase T1 digest mainly shows peaks in the first half of the chromatogram, whereas the RNase 4 digest displays a rich peak profile in the second half of the chromatogram (Figure 3). This indicates that shorter subsequences are generated by RNase T1 than by RNase 4. As depicted in Figure 3, an *in silico* digestion of Fluc mRNA with RNase T1 generates mostly subsequences with a length of 2 to 5 nt, whereas an RNase 4-mediated *in silico* digestion of Fluc mRNA yields readily longer subsequences.

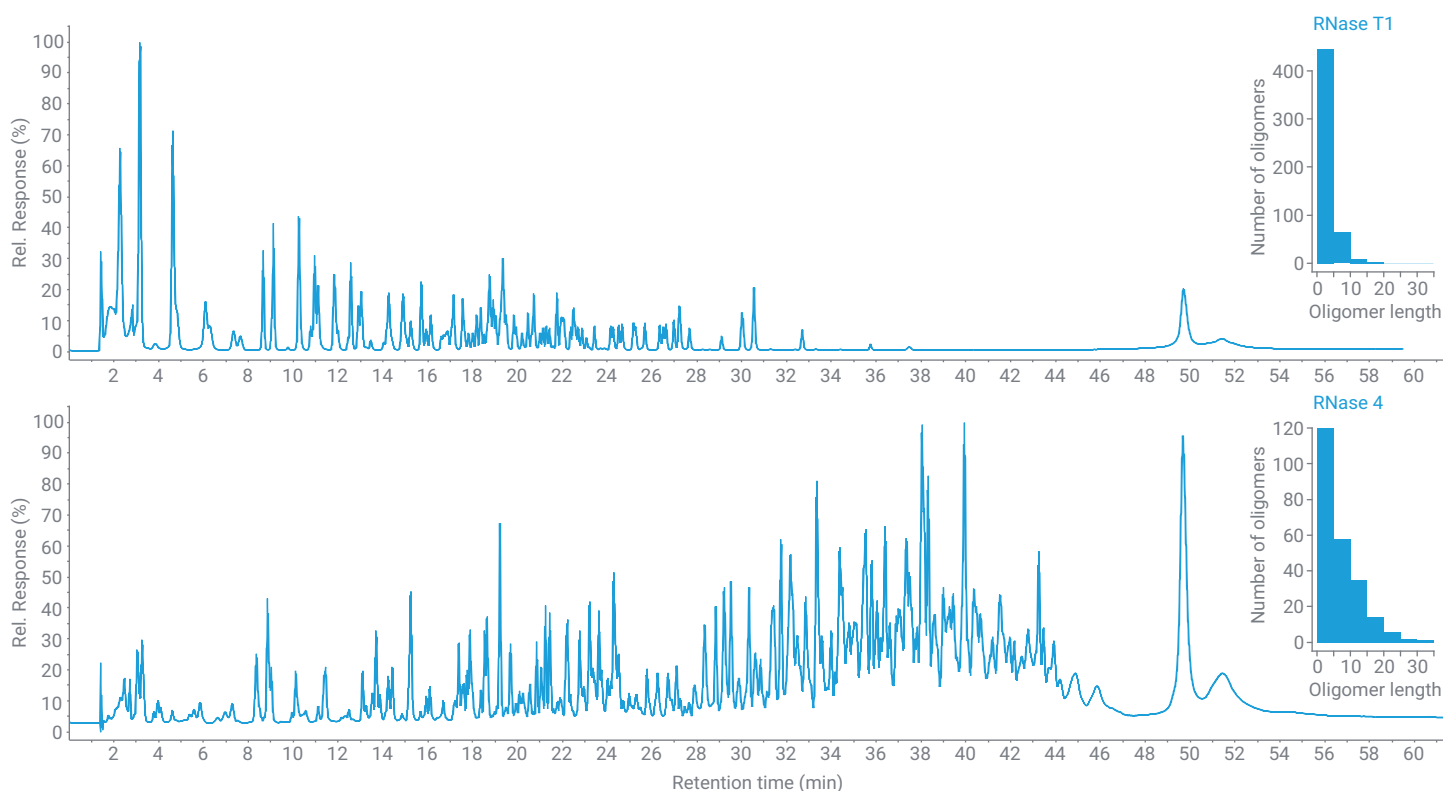


Figure 3. IP-RPLC-UV chromatograms of RNase T1- and RNase 4-digested Fluc mRNA (LC method 2). The *in silico*-generated subsequence length distributions in RNase T1 and RNase 4 digestions are shown to the right of the chromatograms.

To retrieve more information on the identities of the various peaks resulting from RNase T1 and RNase 4 digestion and to perform an in-depth mRNA characterization, LC/MS analysis was performed. In the LC/MS chromatogram of the RNase T1 digest (Figure 4), the retention time range from 5 to 40 minutes displays a dense peak region, whereas two wider and more abundant peaks appear between 50 and 55 minutes and a narrow peak at approximately 65 minutes. These latter three peaks were highly charged, suggesting high MWs. Following maximum entropy deconvolution using MS profile data, these peaks could be associated with the poly A tail, large mRNA hydrolysates, and the RNase T1 enzyme, respectively.

The deconvoluted MS spectrum of the poly A tail showed a symmetric (supported by a polydispersity close to 1) distribution of poly A tail lengths varying between 118 and 133 A units and with a distribution mode, i.e., the most

abundantly present poly A tail length, at 124 A units, in agreement with the results mentioned in Morreel *et al.*¹⁵ Note the clean deconvoluted spectrum devoid of aberrant adduct formation, which can be attributed to the materials used in the mobile phase/sample flow path (plastic bottles, biocompatible LC material, and Ultra Inert column technology). Additionally, the 5'-terminal cap structure could be traced at approximately 20 minutes in the 5 to 40-minute crowded region by searching for two pseudomolecular ions having the same retention time but differing by 165 amu. The latter mass difference corresponds to an in-source loss of the 7-methylguanine nucleobase. In agreement with Gau *et al.*²², RNase T1 digestion followed by IP-RPLC-MS analysis allowed simultaneous profiling of the 5'-terminal structure, the 3'-poly A tail, and the various subsequences necessary for RNA mapping.

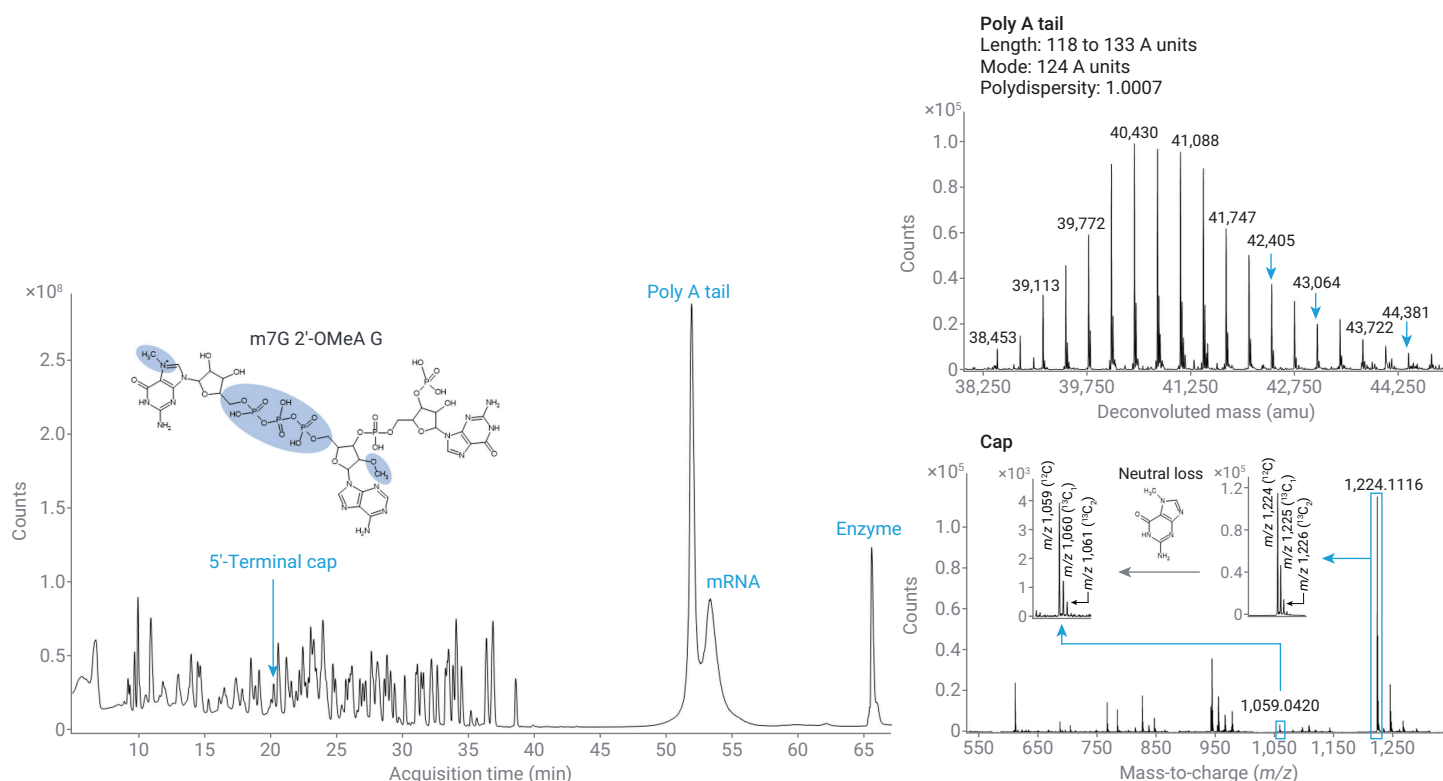


Figure 4. mRNA mapping combined with 5'-cap and 3'-poly A tail analyses of the RNase T1 digest of Fluc mRNA. The displayed cap-containing trimer consists of a 7-methylguanosine (m7G) coupled by a triphosphate bridge to a 2'-O-methyladenosine (2'-OMeA), which is itself linked to a guanosine (G) unit. When the cap structure m7G 2'-OMeA G enters the MS ionization source, a neutral loss of 165 Da corresponding to 7-methylguanine is observed. The mRNA pool displays a variety of poly A tail lengths of which the distribution symmetry around the average length is judged by the polydispersity (LC method 2).

A concurrent analysis procedure of mRNA terminal annotation and mRNA mapping was also obtained upon IP-RPLC-MS analysis of an RNase 4 digest (Figure 5). Here, the poly A tail length distribution was similar to that obtained through RNase T1 digestion. However, opposite of the RNase T1 digest, the capped 5'-terminus was represented by a longer and unique octameric sequence, m7G 2'-O-MeA

G N N N U (N and m7G 2'-O-MeA represent nondisclosed nucleobases and the cap structure, respectively; see the legend of Figure 4 for additional explanations), for which the noncapped counterpart, A G N N N U, could be traced as well (Figure 6). Based on the relative peak areas of the capped and noncapped species, a 99.2% capping efficiency was calculated.

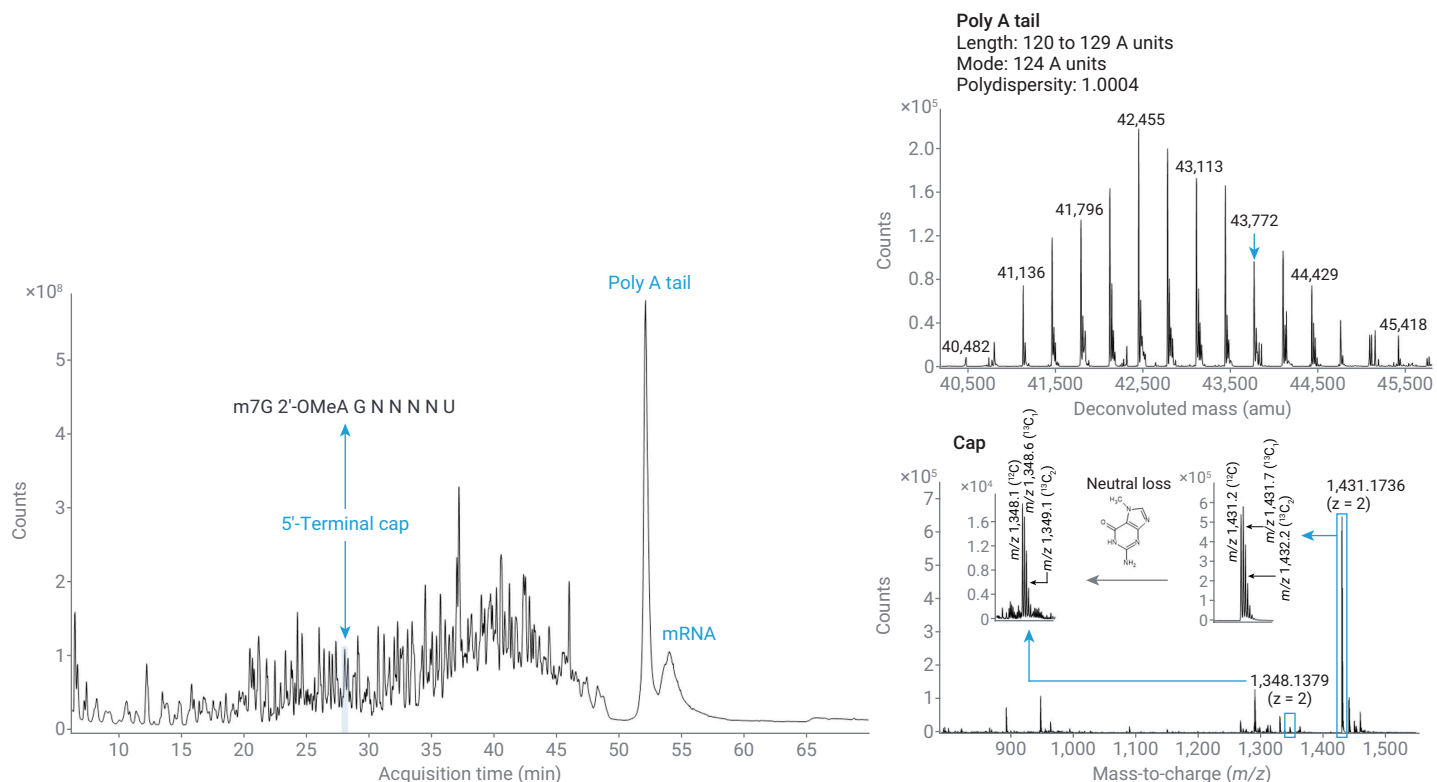


Figure 5. mRNA mapping combined with 5'-cap and 3'-poly A tail analyses of the RNase 4 digest of Fluc mRNA (LC method 2). The cap-containing oligonucleotide is much longer (8 units) than observed for the RNase T1 digest. See Figure 4 for further information.

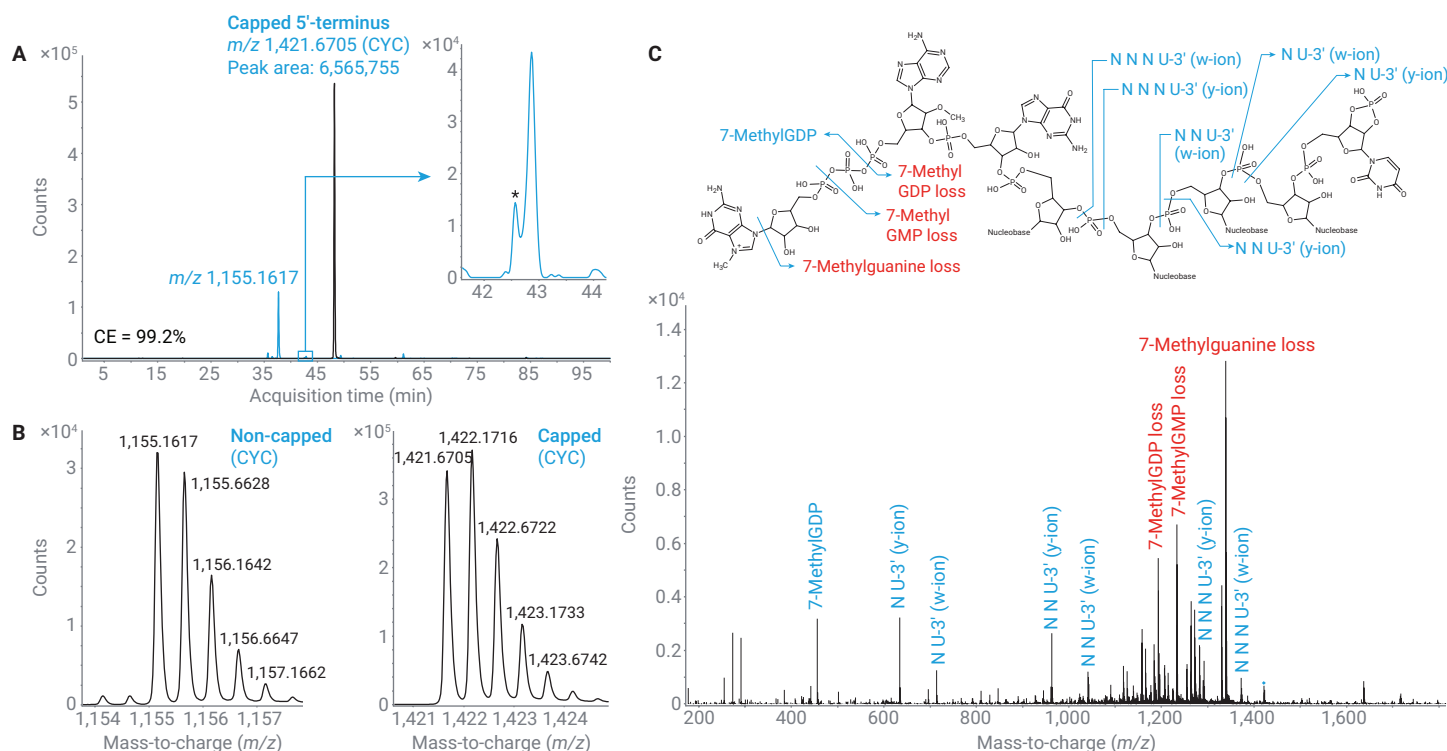


Figure 6. 5'-Terminal mRNA sequence analysis and capping efficiency computation (LC method 3) of the FLuc mRNA digest obtained with RNase 4. (A) Extracted ion chromatograms (EICs) at m/z 1,155.1617 (blue) and 1,421.6705 (black), corresponding to the noncapped (A G N N N U) and capped 5'-termini (m7G 2'-OMeA G N N N U; see Figure 5) having a 2',3'-cyclic-phosphodiester (CYC) 3'-terminus. The abundant peak labeled m/z 1,155.1617 could be distinguished from the isomeric CYC-terminated noncapped species by MS/MS spectral interpretation. The blue chromatogram in the inset represents a zoom of the m/z 1,155.1617-based EIC. (B) m/z Isotope envelope of the doubly charged CYC-terminated capped and noncapped species. (C) MS/MS spectrum of the CYC-terminated capped species with indication of the cleavage positions. See Figure 7 for the cleavage positions leading to the w- and y-type ions. Annotation labels based on the structure of the product ion or on that of the corresponding neutral loss are indicated in blue or red, respectively. *Contaminating peak represents a ^{13}C isotope of an internal mRNA subsequence. Note that an extended LC gradient (method 3) was used to better resolve the latter contaminant from the noncapped 5'-terminus, facilitating accurate capping efficiency determination.

For further sequence confirmation, the IP-RPLC-MS data were subjected to small and large-sized molecular feature extraction returning monoisotopic or average MW values associated with all peaks. These MWs were then searched among the many theoretical MWs derived from an *in silico* digestion of the mRNA, enabling subsequence annotation of the recorded MWs. Occasionally, the same MW is obtained for multiple LC/MS peaks. As such peaks represent different isomeric subsequences, MS/MS fragmentation is necessary

to distinguish between all possible isomers. Collision-induced dissociation of an RNA oligomer occurs typically through phosphodiester cleavage, yielding 5'-terminal a and c ions and 3'-terminal w and y ions³⁰, which can be used to distinguish the different isomers (Figure 7). This procedure of using full MS supported by MS/MS to annotate the LC/MS peaks yielded the RNase T1 fragment annotations shown for the more abundant peaks in Figure 8.

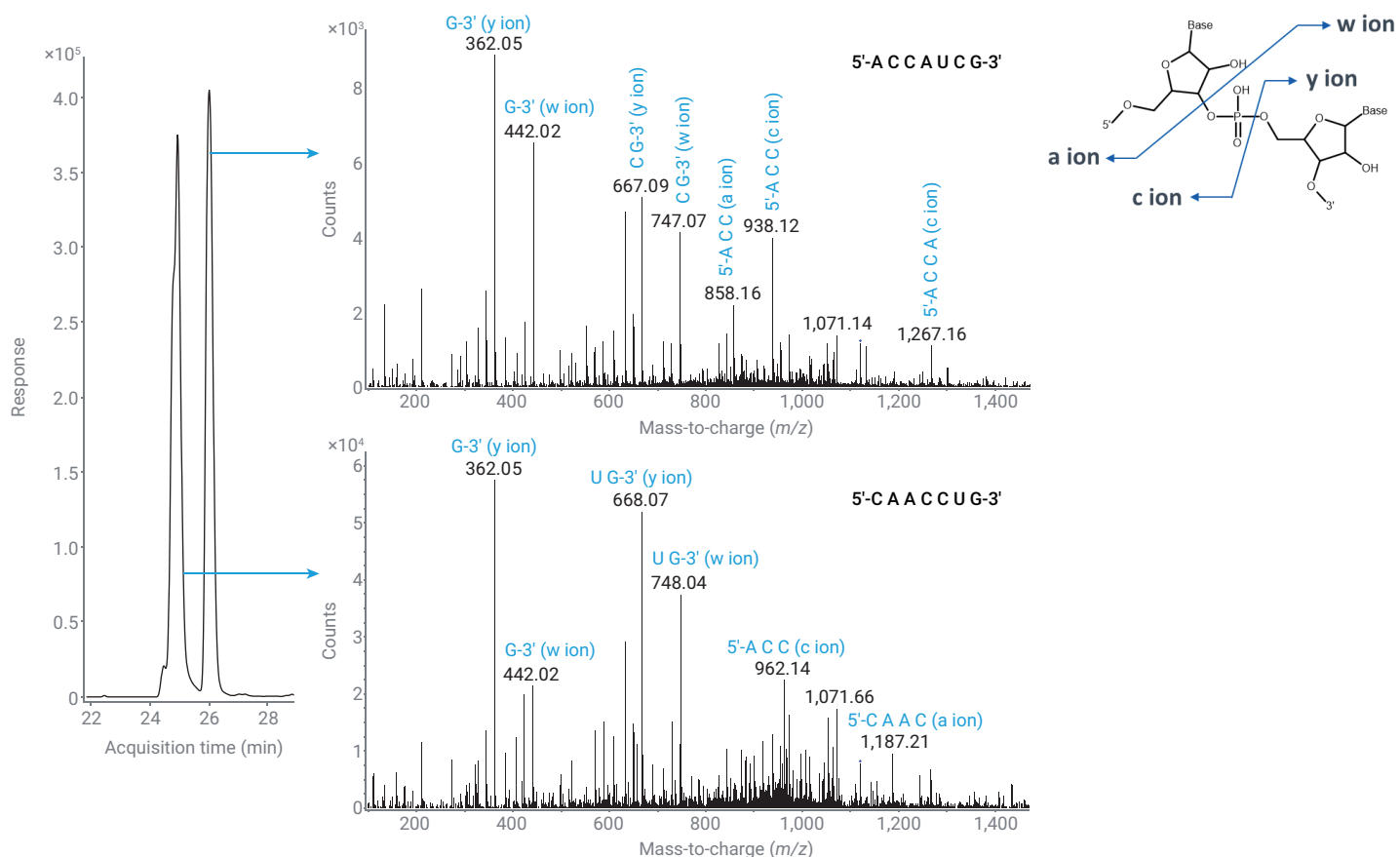


Figure 7. IP-RPLC-MS peaks (EIC) representing two isomeric subsequences derived from the RNase T1 digest of Fluc mRNA (LC method 2). The more abundant MS/MS product ions represent the a-, c-, y-, and w-type phosphodiester cleavages yielding 5'- and 3'-terminal, as well as internal fragments.

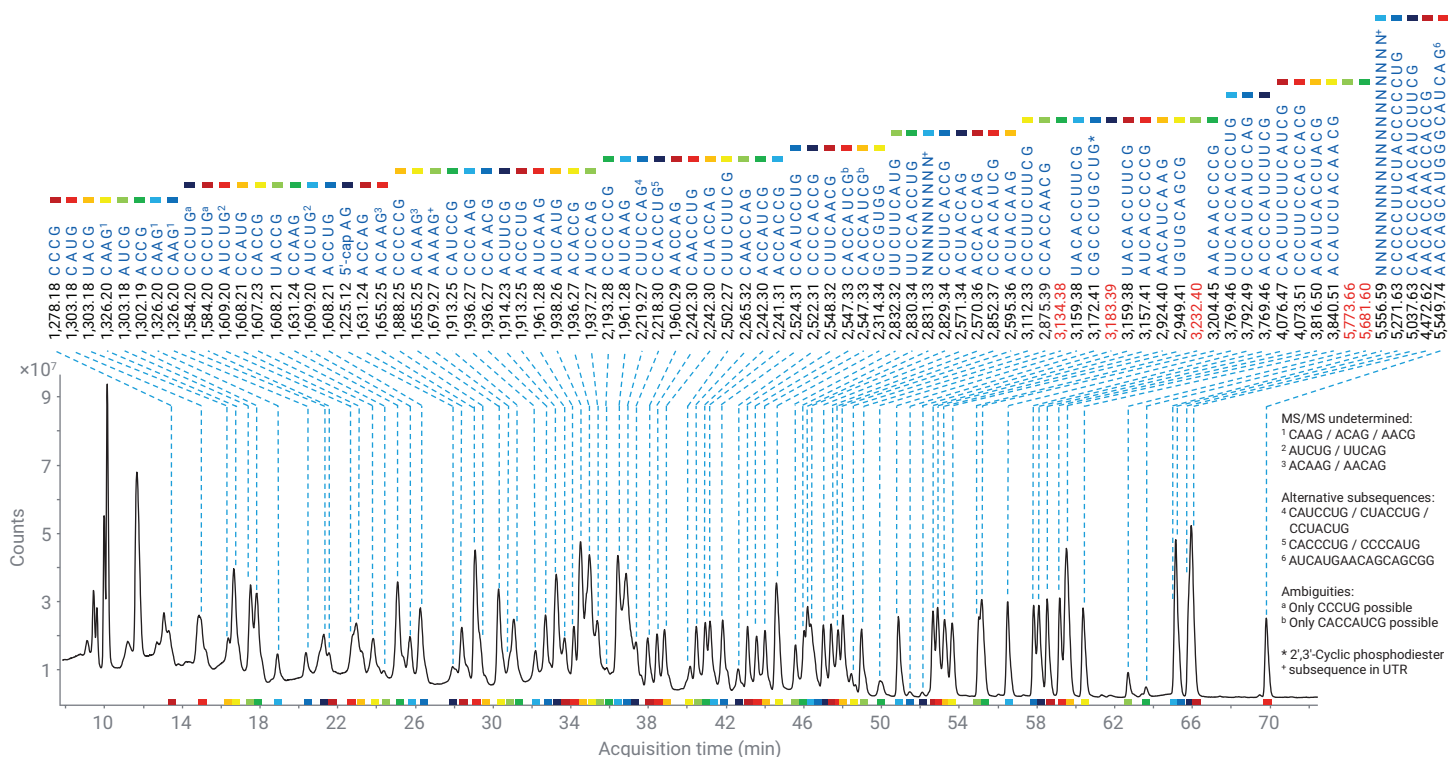


Figure 8. mRNA mapping of Fluc mRNA RNase T1 digest with indication of subsequence annotation based on MS/MS spectral elucidation (LC method 3). Colored bars above the annotations and below the chromatogram peaks improve the visual tracking of the subsequence annotation for each peak. Each subsequence is preceded by its MW as recorded through MS. Red annotations correspond to unidentified peaks.

Peaks where the annotated subsequence occurs multiple times along the mRNA sequence (multisite subsequences) prevent a one-to-one correspondence of the peak with a particular site on the mRNA sequence and were discarded from further analysis. By way of example, the trimer ACG occurs 13 times in the ORF of Fluc mRNA, thereby covering >2% of the sequence. For RNase T1 digestion of Fluc mRNA, the remaining site-specific peaks could be split into those representing a subsequence containing a 3'-terminal phosphate group (60 peaks) and those characterized by a 3'-terminal 2',3'-cyclic phosphodiester moiety (52 peaks). Among this total of 112 subsequences, those located in the ORF covered 49.4% of the latter mRNA region (Figure 9). This value approaches the upper border of the 12% to 56% range based on published sequence coverages of Fluc mRNA computed using site-specific subsequences derived from LC/MS analyses of complete RNase T1 digests.^{21,22,25,26}

By performing the same RNA mapping procedure on the RNase 4 digest, 159 site-specific subsequences were annotated on the full mRNA sequence, which represented

83 and 76 subsequences bearing a 3'-phosphate and a 2'-3'-cyclic-phosphodiester at the 3'-terminus, respectively. These single-site subsequences covered 88.6% of the Fluc mRNA ORF (Figure 9), a value surpassing the previously published 78%.²⁶ Clearly, the use of RNase 4 allowed a much higher sequence coverage than RNase T1 digestion used in the longer subsequences obtained by the enzyme.

When combining the identified site-specific subsequences resulting from both the RNase T1 and RNase 4 digestions (Figure 9), 94.7% of the Fluc mRNA ORF is covered, which is similar to the value reported by Vanhinsbergh *et al.*²⁰ where partial RNase T1 digestion was performed. Furthermore, the resulting 94.7% coverage is close to the 96.9% coverage obtained when including subsequences that occur multiple times along the mRNA sequence (Table 3). This observation results from the long and unique subsequences generated by RNase 4 digestion, eliminating the often occurring need to consider multisite subsequences to further improve mRNA sequence coverage.

RNA mapping of RNase T1 digest (sequence coverage: 49.4%)

```

1 AUGGAGGACG CCAAGAACAU CAAGAAGGGC CCCGCCCCU UCUACCCCUU GGAGGACGGC ACCGCCGGCG AGCAGCUGCA CAAGGCCAUG AAGCGGUACG
101 CCCUGGUGCC CGGCACCAUC GCCUUCACCG ACGCCACAU CGAGGUGGAC AUCACCUACG CCGAGUACUU CGAGAUAGAG GUGCGGCUUG CGAGGCCAU
201 GAAGCGGUAC GGCCUGAACA CCAACCACCG GAUCGUGGUG UGCAGCGAGA ACAGCCUGCA GUUCUUAUG CCCGUGCUGG GCGCCUGUU CAUCGGCGUG
301 GCCGUGGCC CCGCCAACGA CAUCUACAAC GAGCGGGAGC UGCUGAACAG CAUGGGCAUC AGCCAGCCCA CCGUGGUGUU CGUGAGCAAG AAGGGCCUGC
401 AGAAGAUCCU GAACGUGCAG AAGAAGCUC CCAUCAUCCA GAAGAUCAUC AUCAUGGACA GCAAGACCGA CUACCAAGGC UUCAGAGCA UGUACACCUU
501 CGUGACGAGC CACCUGCC CCGGCUUCA CGAGUACGAC UUCGUGCCCG AGAGCUUCGA CCGGACAAG ACCAUCGCC UGAUCAUGAA CAGCAGCGGC
601 AGCACC GGCC UGCCAAGGG CGUGGCCUG CCCACCGGA CCGCCUGCGU GCGGUUCAGC CACGCCCGGG ACCCAUCUU CGGCAACCGA AUCAUCCCG
701 ACACCGCCAU CCUGAGCGUG GUGCCCUUCC ACCACGGCUU CGGCAUGUUC ACCACCCUGG GCUACCUGAU CUGCGGCUUC CGGGUGGUGC UGAUGUACCG
801 GUUCGAGGAG GAGCUGUUC UGCGGAGCCU GCAGGACUAC AAGAUCGAGA GCGCCUGCU GUGGCCACC CUGUUCAGCU UCUCGCCAA GAGCACCUG
901 AUCGACAAGU ACGACCUAG CAAACUGCAC GAGAUCCGA GCGGCGGCG CCCCUAGC AAGGAGGUG GCGAGGCCGU GGCCAAGCGG UUCCACCUGC
1001 CCGGCAUCCG GCAGGGCUAC GGCUGACCG AGACCAACAG CGCCAUCUG AUCACCCCGG AGGGCGACGA CAAGCCCGGC GCCGUGGGA AGGUGUGGCC
1101 CUUCUUCGAG GCCAAGGUG UGACCUUGGA CACCGGCAAG ACCUGGGCG UGAACGAGCG GGGCGAGCUG UGCGUGCGGG GCCCAUGAU CAUGAGCGGC
1201 UACGUGAACA ACCCGAGGC CACCAACGCC CUGAUCGACA AGGACGGCG GCUGCACAGC GGCACAUCC CUACUGGGA CGAGGACGAG CACUUCUUA
1301 UCGUGGACCG GUGAAGAGC CUGAUCAAGU ACAAGGGCUA CAGGUGGCC CCCGCCGAGC UGAGAGCAU CCUGCUGCAG CACCCCAACA UCUCGAGCG
1401 CGGCGUGGCC GGCUGGCCG ACGACGACGC CGGCGAGCUG CCGCCGCGCG UGUGUGUGCU GAGCAGCGG AAGACCAUGA CGGAGAAGGA GAUCGUGGAC
1501 UACGUGGCCA GCGAGGUGAC CACCGCCAAG AAGCUGCGGG GCGGCGUGGU GUUCGUGGAC GAGGUGCCA AGGGCCUGAC CGGCAAGCUG GACGCCCGGA
1601 AGAUCCGGGA GAUCCUGAUC AAGGCCAAGA AGGGCGGCAA GAUCGCCUG UGA

```

RNA mapping of RNase 4 digest (sequence coverage: 88.6%)

```

1 AUGGAGGACG CCAAGAACAU CAAGAAGGGC CCCGCCCCU UCUACCCCUU GGAGGACGGC ACCGCCGGCG AGCAGCUGCA CAAGGCCAUG AAGCGGUACG
101 CCCUGGUGCC CGGCACCAUC GCCUUCACCG ACGCCACAU CGAGGUGGAC AUCACCUACG CCGAGUACUU CGAGAUAGAG GUGCGGCUUG CGAGGCCAU
201 GAAGCGGUAC GGCCUGAACA CCAACCACCG GAUCGUGGUG UGCAGCGAGA ACAGCCUGCA GUUCUUAUG CCCGUGCUGG GCGCCUGUU CAUCGGCGUG
301 GCCGUGGCC CCGCCAACGA CAUCUACAAC GAGCGGGAGC UGCUGAACAG CAUGGGCAUC AGCCAGCCCA CCGUGGUGUU CGUGAGCAAG AAGGGCCUGC
401 AGAAGAUCCU GAACGUGCAG AAGAAGCUC CCAUCAUCCA GAAGAUCAUC AUCAUGGACA GCAAGACCGA CUACCAAGGC UUCCAGAGCA UGUACACCUU
501 CGUGACGAGC CACCUGCC CCGGCUUCA CGAGUACGAC UUCGUGCCCG AGAGCUUCGA CCGGACAAG ACCAUCGCC UGAUCAUGAA CAGCAGCGGC
601 AGCACC GGCC UGCCAAGGG CGUGGCCUG CCCACCGGA CCGCCUGCGU GCGGUUCAGC CACGCCCGGG ACCCAUCUU CGGCAACCGA AUCAUCCCG
701 ACACCGCCAU CCUGAGCGUG GUGCCCUUCC ACCACGGCUU CGGCAUGUUC ACCACCCUGG GCUACCUGAU CUGCGGCUUC CGGGUGGUGC UGAUGUACCG
801 GUUCGAGGAG GAGCUGUUC UGCGGAGCCU GCAGGACUAC AAGAUCGAGA GCGCCUGCU GUGGCCACC CUGUUCAGCU UCUCGCCAA GAGCACCUG
901 AUCGACAAGU ACGACCUAG CAAACUGCAC GAGAUCCGA GCGGCGGCG CCCCUAGC AAGGAGGUG GCGAGGCCGU GGCCAAGCGG UUCCACCUGC
1001 CCGGCAUCG GCAGGGCUAC GGCUGACCG AGACCAACAG CGCAUCCUG AUCACCCCG AGGGCGACGA CAAGCCCGGC GCCGUGGGA AGGUGGUGCC
1101 CUUCUUCGAG GCCAAGGUGG UGACCUUGGA CACCGGCAAG ACCUGGGCG UGAACGAGCG GGGCGAGCUG UGCGUGCGGG GCCCAUGAU CAUGAGCGGC
1201 UACGUGAACA ACCCGAGGC CACCAACGCC CUGAUCGACA AGGACGGCG GCUGCACAGC GGCACAUCC CUACUGGGA CGAGGACGAG CACUUCUUA
1301 UCGUGGACCG GUGAAGAGC CUGAUCAAGU ACAAGGGCUA CAGGUGGCC CCCGCCGAGC UGAGAGCAU CCUGCUGCAG CACCCCAACA UCUCGAGCG
1401 CGGCGUGGCC GGCUGGCCG ACGACGACGC CGGCGAGCUG CCGCCGCGCG UGUGUGUGCU GAGGACGGC AAGACCAUGA CGGAGAAGGA GAUCGUGGAC
1501 UACGUGGCCA GCGAGGUGAC CACCGCCAAG AAGCUGCGGG GCGGCGUGGU GUUCGUGGAC GAGGUGCCA AGGGCCUGAC CGGCAAGCUG GACGCCCGGA
1601 AGAUCCGGGA GAUCCUGAUC AAGGCCAAGA AGGGCGGCAA GAUCGCCUG UGA

```

RNA mapping of RNase T1 + RNase 4 digests (sequence coverage: 94.7%)

```

1 AUGGAGGACG CCAAGAACAU CAAGAAGGGC CCCGCCCCU UCUACCCCUU GGAGGACGGC ACCGCCGGCG AGCAGCUGCA CAAGGCCAUG AAGCGGUACG
101 CCCUGGUGCC CGGCACCAUC GCCUUCACCG ACGCCACAU CGAGGUGGAC AUCACCUACG CCGAGUACUU CGAGAUAGAG GUGCGGCUUG CGAGGCCAU
201 GAAGCGGUAC GGCCUGAACA CCAACCACCG GAUCGUGGUG UGCAGCGAGA ACAGCCUGCA GUUCUUAUG CCCGUGCUGG GCGCCUGUU CAUCGGCGUG
301 GCCGUGGCC CCGCCAACGA CAUCUACAAC GAGCGGGAGC UGCUGAACAG CAUGGGCAUC AGCCAGCCCA CCGUGGUGUU CGUGAGCAAG AAGGGCCUGC
401 AGAAGAUCCU GAACGUGCAG AAGAAGCUC CCAUCAUCCA GAAGAUCAUC AUCAUGGACA GCAAGACCGA CUACCAAGGC UUCCAGAGCA UGUACACCUU
501 CGUGACGAGC CACCUGCC CCGGCUUCA CGAGUACGAC UUCGUGCCCG AGAGCUUCGA CCGGACAAG ACCAUCGCC UGAUCAUGAA CAGCAGCGGC
601 AGCACC GGCC UGCCAAGGG CGUGGCCUG CCCACCGGA CCGCCUGCGU GCGGUUCAGC CACGCCCGGG ACCCAUCUU CGGCAACCGA AUCAUCCCG
701 ACACCGCCAU CCUGAGCGUG GUGCCCUUCC ACCACGGCUU CGGCAUGUUC ACCACCCUGG GCUACCUGAU CUGCGGCUUC CGGGUGGUGC UGAUGUACCG
801 GUUCGAGGAG GAGCUGUUC UGCGGAGCCU GCAGGACUAC AAGAUCGAGA GCGCCUGCU GUGGCCACC CUGUUCAGCU UCUCGCCAA GAGCACCUG
901 AUCGACAAGU ACGACCUAG CAAACUGCAC GAGAUCCGA GCGGCGGCG CCCCUAGC AAGGAGGUG GCGAGGCCGU GGCCAAGCGG UUCCACCUGC
1001 CCGGCAUCCG GCAGGGCUAC GGCUGACCG AGACCAACAG CGCAUCCUG AUCACCCCG AGGGCGACGA CAAGCCCGGC GCCGUGGGA AGGUGGUGCC
1101 CUUCUUCGAG GCCAAGGUGG UGACCUUGGA CACCGGCAAG ACCUGGGCG UGAACGAGCG GGGCGAGCUG UGCGUGCGGG GCCCAUGAU CAUGAGCGGC
1201 UACGUGAACA ACCCGAGGC CACCAACGCC CUGAUCGACA AGGACGGCG GCUGCACAGC GGCACAUCC CUACUGGGA CGAGGACGAG CACUUCUUA
1301 UCGUGGACCG GUGAAGAGC CUGAUCAAGU ACAAGGGCUA CAGGUGGCC CCCGCCGAGC UGAGAGCAU CCUGCUGCAG CACCCCAACA UCUCGAGCG
1401 CGGCGUGGCC GGCUGGCCG ACGACGACGC CGGCGAGCUG CCGCCGCGCG UGUGUGUGCU GAGGACGGC AAGACCAUGA CGGAGAAGGA GAUCGUGGAC
1501 UACGUGGCCA GCGAGGUGAC CACCGCCAAG AAGCUGCGGG GCGGCGUGGU GUUCGUGGAC GAGGUGCCA AGGGCCUGAC CGGCAAGCUG GACGCCCGGA
1601 AGAUCCGGGA GAUCCUGAUC AAGGCCAAGA AGGGCGGCAA GAUCGCCUG UGA

```

Figure 9. Fluc mRNA ORF sequence coverage. Multisite subsequences are excluded. See text for further explanation.

Table 3. Fluc mRNA ORF sequence coverage.

	RNase T1	RNase 4	RNase T1 + 4
Only Site-Specific Subsequences	49.4%	88.6%	94.7%
All Subsequences (Length ≥ 3)	75.7%	88.9%	96.9%

Conclusion

This application note investigates mRNA mapping using IP-RPLC-MS on a low adsorption flow path, composed of an Agilent Altura HPLC column with Ultra Inert technology and Agilent 1290 Infinity II Bio LC System. When applied on a complex oligonucleotide mixture resulting from the RNase T1 and RNase 4 digestion of Fluc mRNA, a low-adsorption flow path, enabled (1) reaching high sequence coverage, (2) displaying the poly A tail length distribution, (3) verifying the 5'-terminal cap structure, and (4) calculating the capping efficiency. Strikingly, by combining the results from parallel digestions with a low-frequent (RNase 4) and a high-frequent (RNase T1) cutter, 94.7% of the mRNA sequence could be covered.

References

- Szabó, G. T.; Mahiny, A. J.; Vlatkovic, I. COVID-19 mRNA Vaccines: Platforms and Current Developments. *Mol. Ther.* **2022**, *30*(5), 1850–1868.
- Fortner, A.; Schumacher, D. First COVID-19 Vaccines Receiving the US FDA and EMA Emergency Use Authorization. *Discoveries* **2021**, *9*(1).
- Heinz, F. X.; Stiasny, K. Distinguishing Features of Current COVID-19 Vaccines: Knowns and Unknowns of Antigen Presentation and Modes of Action. *NPJ Vaccines* **2021**, *6*(1), 104.
- Park, J. W.; Lagniton, P. N.; Liu, Y.; Xu, R. H. mRNA Vaccines for COVID-19: What, Why and How. *Int. J. Biol. Sci.* **2021**, *17*(6), 1446.
- Qin, S.; Tang, X.; Chen, Y.; Chen, K.; Fan, N.; Xiao, W.; Zheng, Q.; Li, G.; Teng, Y.; Wu, M.; et al. mRNA-Based Therapeutics: Powerful and Versatile Tools to Combat Diseases. *Signal Transduct. Target Ther.* **2022**, *7*(1), 166.
- Xiao, Y.; Tang, Z.; Huang, X.; Chen, W.; Zhou, J.; Liu, H.; Liu, C.; Kong, N.; Tao, W. Emerging mRNA Technologies: Delivery Strategies and Biomedical Applications. *Chem. Soc. Rev.* **2022**, *51*(10), 3828–3845.
- Kwon, S.; Kwon, M.; Im, S.; Lee, K.; Lee, H. mRNA Vaccines: the Most Recent Clinical Applications of Synthetic mRNA. *Arch. Pharmacol. Res.* **2022**, *45*(4), 245–262.
- Sahin, U.; Karikó, K.; Türeci, O. mRNA-Based Therapeutics—Developing a New Class of Drugs. *Nat. Rev. Drug Discov.* **2014**, *13*(10), 759–780.
- Pardi, N.; Hogan, M. J.; Porter, F. W.; Weissman, D. mRNA Vaccines—a New Era in Vaccinology. *Nat. Rev. Drug Discov.* **2018**, *17*(4), 261–279.
- Kariko, K.; Weissman, D. Naturally Occurring Nucleoside Modifications Suppress the Immunostimulatory Activity of RNA: Implication for Therapeutic RNA Development. *Curr. Opin. Drug Discov. Dev.* **2007**, *10*(5), 523.
- Minnaert, A. K.; Vanluchene, H.; Verbeke, R.; Lentacker, I.; De Smedt, S. C.; Raemdonck, K.; Sanders, N. N.; Remaut, K. Strategies for Controlling the Innate Immune Activity of Conventional and Self-Amplifying mRNA Therapeutics: Getting the Message Across. *Adv. Drug Deliv. Rev.* **2021**, *176*, 113900.
- Callaway, E.; Naddaf, M. Pioneers of mRNA COVID Vaccines Win Medicine Nobel. *Nature* **2023**, *622*(7982), 228–229.
- Blenke, E. O.; Oernskov, E.; Schoeneich, C.; Nilsson, G. A.; Volkin, D. B.; Mastrobattista, E.; Almarsson, O.; Crommelin, D. J. A. The Storage and In-Use Stability of mRNA Vaccines and Therapeutics: Not a Cold Case. *J. Pharm. Sci.* **2023**, *112*(2), 386–403.
- Packer, M.; Gyawali, D.; Yerabolu, R.; Schariter, J.; White, P. A Novel Mechanism for the Loss of mRNA activity in Lipid Nanoparticle Delivery Systems. *Nat. Commun.* **2021**, *12*(1), 6777.
- Morreel, K.; t'Kindt, R.; Debyser, G.; Jonckheere, S.; Sandra, P.; Sandra, K. Diving into the Structural Details of In Vitro Transcribed mRNA Using Liquid Chromatography–Mass Spectrometry-Based Oligonucleotide Profiling. *LCGC Europe* **2022**, *35*(2022), 220–236.

16. Beverly, M.; Dell, A.; Parmar, P.; Houghton, L. Label-Free Analysis of mRNA Capping Efficiency Using RNase H Probes and LC-MS. *Anal. Bioanal. Chem.* **2016**, *408*, 5021–5030.
17. 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.
18. Rapid Analysis of mRNA 5' Capping with High Resolution LC/MS. *Agilent Technologies application note*, 5994-3984EN, **2021**.
19. Analysis of mRNA poly-A Sequence Variants by High-Resolution LC/MS. *Agilent Technologies application note*, 5994-3005EN, **2021**.
20. Vanhinsbergh, C. J.; Criscuolo, A.; Sutton, J. N.; Murphy, K.; Williamson, A. J.; Cook, K.; Dickman, M. J. Characterization and Sequence Mapping of Large RNA and mRNA Therapeutics Using Mass Spectrometry. *Anal. Chem.* **2022**, *94*(20), 7339–7349.
21. Goyon, A.; Scott, B.; Yehl, P.; Zhang, K. Online Nucleotide Mapping of mRNAs. *Anal. Chem.* **2024**, *96*(21), 8674–8681.
22. Gau, B. C.; Dawdy, A. W.; Wang, H. L.; Bare, B.; Castaneda, C. H.; Friese, O. V.; Thompson, M. S.; Lerch, T. F.; Cirelli, D. J.; Rouse, J. C. C. Oligonucleotide Mapping Via Mass Spectrometry to Enable Comprehensive Primary Structure Characterization of an mRNA Vaccine Against SARS-CoV-2. *Sci. Rep.* **2023**, *13*(1), 9038.
23. Xiong, J.; Wu, J.; Liu, Y.; Feng, Y. J.; Yuan, B. F. Quantification and Mapping of RNA Modifications. *Trends Anal. Chem.* **2024**, 117606.
24. Determining mRNA Capping with HILIC-MS on a Low-Adsorption Flow Path. *Agilent Technologies application note*, 5994-7118EN, **2024**.
25. Jiang, T.; Yu, N.; Kim, J.; Murgo, J. R.; Kissai, M.; Ravichandran, K.; Miracco, E. J.; Presnyak, V.; Hua, S. Oligonucleotide Sequence Mapping of Large Therapeutic mRNAs Via Parallel Ribonuclease Digestions and LC-MS/MS. *Anal. Chem.* **2019**, *91*(13), 8500–8506.
26. Wolf, E. J.; Grünberg, S.; Dai, N.; Chen, T. H.; Roy, B.; Yigit, E.; Corrêa Jr, I. R. Human RNase 4 Improves mRNA Sequence Characterization by LC-MS/MS. *Nucleic Acids Res.* **2022**, *50*(18), e106–e106.
27. Evaluation of Different Ion-Pairing Reagents for LC/UV and LC/MS Analysis of Oligonucleotides. *Agilent Technologies application note*, 5994-2957EN, **2021**.
28. Guimaraes, G. J.; Bartlett, M. G. Managing Nonspecific Adsorption to Liquid Chromatography Hardware: a Review. *Anal. Chim. Acta.* **2023**, *1250*, 340994.
29. Lardeux, H.; Goyon, A.; Zhang, K.; Nguyen, J. M.; Lauber, M. A.; Guillarme, D.; D'Atri, V. The Impact of Low Adsorption Surfaces for the Analysis of DNA and RNA Oligonucleotides. *J. Chromatogr. A* **2022**, *1677*, 463324.
30. McLuckey, S. A.; Van Berkel, G. J.; Glish, G. L. Tandem Mass Spectrometry of Small, Multiply Charged Oligonucleotides. *J. Am. Soc. Mass Spectrom.* **1992**, *3*(1), 60–70.