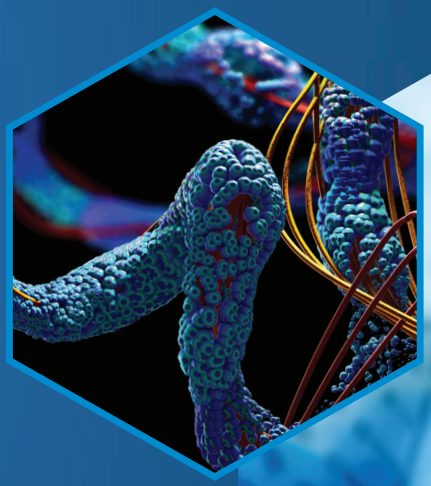
[APPLICATION NOTEBOOK]

LC-MS Analysis Workflows for RNA Therapeutics to Ensure Product Quality and Process Consistency



Waters™

Contents

Attribute Monitoring of RNA: Tools to Ensure Consistent Product Quality

Characterization and monitoring of critical quality attributes with streamlined workflows for CRISPR sgRNA and mRNA

LEARN MORE
waters.com/CGT
waters.com/RapiZymeRNases

EFFICIENCY
Fit for purpose methods and workflows for development, clinical and commercial operations

COMPLIANCE-READY
Inbuilt data integrity and qualification packages for regulated laboratories

STREAMLINED ONBOARDING
Robust, ready to automate procedures for simplified implementation

WORKFLOW DRIVEN
Purposely designed chemical solutions and apps to streamline sample preparation, data acquisition, processing and reporting

OPTIMIZED COLLABORATION
Networkable solutions to share data and methods with enhanced fidelity

Nucleic acid sequences can be delivered to initiate a diverse range of gene-based therapies and vaccine treatments. This includes gene silencing (RNAi), CRISPR sgRNA gene editing, and mRNA-based vaccines.

These complex molecules require compliance-ready analytics for characterization and Chemistry, Manufacturing and Control (CMC) monitoring. Manual, error-prone and laborious workflows often challenge companies from deploying routine CQA assays.

Waters is committed to advancing development and commercialization of RNA therapies with the tools to facilitate sequence confirmation and modification mapping for improved insights on drug design, stability, purity, and molecular integrity.

Integrated Analytical Solutions for Faster and Cost-Effective Identity and CQA Analysis of RNA Therapeutics

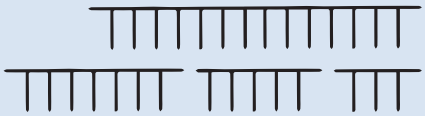
RNA Workflows from Sample to Result

Samples

RNA Samples from discovery to QC

Effortless Reproducibility

Robust, tuneable, automation ready digestion providing unique digest products to confirm sequence and identity



Reliable Oligonucleotide Mapping

Separate digest products with Ion Pairing Reversed Phase (IP-RP) and Hydrophilic interaction chromatography (HILIC) columns on high performance LC systems

Characterization and Monitoring

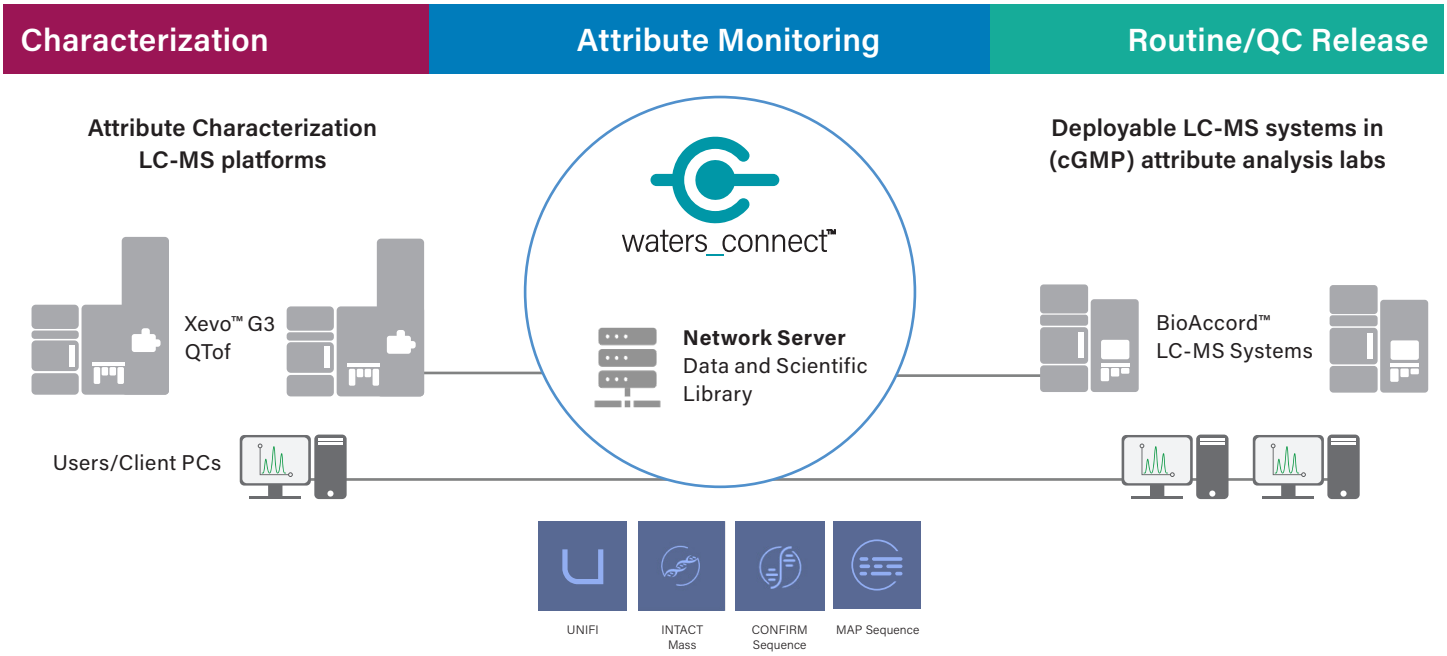
Mass Spectrometry analysis for in depth characterization and monitoring of critical product attributes

Automated Interpretation

Streamline data processing and reporting using compliance ready informatics

Accelerated Development

Confidenct product characterization enables simplified method transfer of routine product and impurity monitoring assays



Introduction

Matthew Lauber, Senior Director, Bioseparations

RNA-based vaccines and therapeutics have seen groundbreaking advancements in recent years, transforming the landscape of modern medicine through new therapeutics and vaccines based on messenger RNA (mRNA) and single guide RNA (sgRNA).

mRNA vaccines work by delivering synthetic mRNA into cells, which then use the genetic instructions to produce a protein that mimics a part of the virus. This protein triggers an immune response, preparing the body to fight the actual virus if encountered. This technology gained significant attention during the COVID-19 pandemic, with the rapid development and deployment of mRNA vaccines by Pfizer-BioNTech and Moderna.

mRNA-based therapeutics offer several advantages, including rapid development, and scalability. Beyond vaccines, mRNA technology holds promise for treating a host of other diseases. mRNA therapeutics can be designed to produce therapeutic proteins, replace defective genes, or modulate immune responses.

sgRNA based vaccines and therapeutics are primarily associated with CRISPR-Cas9 technology. sgRNA is a crucial component of the CRISPR system, guiding the Cas9 enzyme to locations in the genome that enable precise edits to specific genes associated with diseases. This can potentially correct genetic mutations or disrupt viral genomes, providing a novel approach to treating genetic disorders and viral infections.

sgRNA-based therapeutics offer the potential for highly specific and permanent genetic modifications. This precision makes them a powerful tool for treating genetic diseases at their root cause.

Research is ongoing to explore the use of sgRNA-based technology in various therapeutic areas, including cancer, genetic disorders, and infectious diseases. The ability to precisely edit genes opens up new possibilities for personalized medicine and targeted treatments.

Controlling the quality of mRNA and sgRNA products throughout development and manufacture is crucial for safety, efficacy, and to ensure regulatory approval. The analysis of RNA Critical Quality Attributes (CQAs) is a vital component of a robust control strategy. Key considerations include sample preparation, digestion, and automation. Essential measurements such as 5′ cap analysis, 3′ poly A tail analysis, and oligonucleotide mapping for sequence and modification confirmation are critical to verify the integrity, translation efficiency, and stability of mRNA.

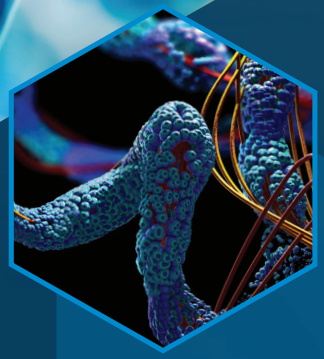
Additionally, intact mRNA drug substance testing using Size Exclusion Chromatography coupled with Multi-Angle Light Scattering (SEC-MALS) provides critical insights into the molecular weight distribution and aggregation state of mRNA. This technique ensures the integrity and homogeneity of the mRNA.

All of these attributes are directly linked to clinical performance and safety and are essential measurements to demonstrate stability and efficacy.

Waters advanced analytics integrate precise digestion enzymes, high purity reagents and standards, and routine software workflows to enhance RNA analysis, ensuring comprehensive and reliable mRNA and gRNA evaluation.

This application notebook summarises various attributes and measurements, highlighting robust techniques and end-to-end methodology to ensure their control throughout product lifecycles

Sample Preparation, Digestion and Automation



Sample Preparation for RNA Analysis

Christian Reidy, Senior Manager, Genetic Medicine

In the rapidly advancing field of nucleic acid therapeutics, single-guide RNA (sgRNA) and messenger RNA (mRNA) have become fundamental tools for gene editing and protein expression. Accurate characterization of these long RNA molecules is critical to ensure their efficacy and safety in therapeutic applications. Liquid Chromatography Mass Spectrometry (LC-MS) is a powerful analytical technique uniquely positioned for this purpose. However, LC-MS analysis of large, intact RNA molecules poses a real challenge due to molecular size and the need for high resolution investigation of minute structural details. A method commonly employed to address this issue is to utilize ribonucleases (RNases) for digestion prior to LC-MS analysis as a bottom-up approach.

RNase digestion simplifies the complex structure of long RNA molecules by cleaving them into smaller oligonucleotides. This process enhances the efficiency and resolution of LC-MS analysis, making it possible to obtain detailed information about the RNA's sequence and modifications. There are a myriad of RNases that can be chosen, each with varying target specificities and digestion properties. By selecting the appropriate RNases, scientists can generate predictable digestion patterns that facilitate the mapping of RNA sequences for identity verification, purity assessment, and modification mapping.

The mechanism of RNase digestion involves the hydrolysis of phosphodiester bonds within the RNA backbone. The enzyme binds to the RNA molecule and catalyzes cleavage at specific sites, determined by its substrate specificity. It is advantageous to digest long RNA into medium-sized products (10–25 nucleotides in length), which preserve sections of unique sequence that can be readily identified without ambiguities. Over-digestion can lead to very small products that are difficult to interpret due to the generation of repeating units and an increased likelihood of sequence isomers, complicating data analysis. Strategies to avoid over-digestion include controlling the digestion time, optimizing the enzyme-to-substrate ratio, and using RNases with cleavage properties that result in partial digestion. Other considerations include ease of use for protocol implementation and whether it is automation friendly. Ultimately, the careful selection and optimization of RNase digestion strategies are pivotal for the reliable and accurate LC-MS analysis of long RNA molecules, thereby advancing the development of effective nucleic acid therapeutics.

Waters™

Application Note

Tunable Digestions of RNA Using RapiZyme™ RNases to Confirm Sequence and Map Modifications

Balasubrahmanyam Addepalli, Tatiana Johnston, Christian Reidy, Matthew A. Lauber

Waters Corporation

Abstract

RNA therapeutics such as sgRNA and mRNA are important modalities for Gene Therapy applications to treat human diseases. Rapid development of these RNA molecules and their deployment require fast tracked development of new analytical methodologies. This methodology aims at determining the identity, purity, and modification(s) of therapeutic RNA through oligo mapping of its sequence by liquid chromatography coupled with mass spectrometry (LC-MS). MS-based sequencing involves direct detection of RNA fragments and localization of the nucleoside modifications in the sequence. Currently available enzymatic tools to produce RNA fragments yield incomplete or partial coverage with a very high ambiguity in sequence interpretation. The availability of enzyme tools with complementary cleavage specificity can overcome this problem and fill sequence gaps while streamlining data interpretation and increasing confidence. This application note evaluates an optimized protocol for RapiZyme MC1 and RapiZyme Cusativin and explores each RNase's reproducibility and complementary specificity with a focus on increased sequence coverage during sgRNA oligo mapping. The new workflows include data processing software, including the MAP Sequence app within waters_connect™ and two supporting MicroApps. These provide a streamlined workflow to facilitate data analysis for routine RNA fragment identification, sequence coverage assessment ,and quantification.

Benefits

- Complete sequence coverage of sgRNA generating more unique masses while maintaining LC-MS compatibility and high signal
- Simple and reproducible protocol that is adaptable for automation without chemical denaturants or enzyme inhibitors
- The ability to perform partial digestions and generate overlapping digestion products
- Automated mRNA digest annotation based on accurate-mass matching as facilitated by *in silico* mRNA digestion calculations by waters_connect applications
- Fully integrated into waters_connect MAP Sequence for streamlined data analysis

Introduction

Background:

Success of mRNA vaccines against SARS-Cov-2 pandemic accelerated the development of RNA therapeutics to treat infectious diseases and provide gene therapies through genome editing involving CRISPR sgRNA. Parallel advancements in analytical technology can accelerate the turnaround of characterization and release testing work being performed in discovery, development ,and QC laboratories. Confirmation of identity, purity ,and modifications in therapeutic mRNA through oligo-mapping and sequencing via LC-MS has been the widely accepted strategy for direct molecular analysis. Similar to a peptide mapping involving bottom-up approaches, RNA is enzymatically digested to generate RNA fragments that can then be separated on a reversed phase (or hydrophilic interaction) column through liquid chromatography (LC). As digestion components are eluted from the column, they can subsequently be ionized and detected by mass spectrometry (MS).[1] To date, RNase T1 has been the most commonly used enzyme for oligo mapping LC-MS. This enzyme cleaves RNA at the 3′ end of guanosine residues (G). As a result, RNase T1 often generates smaller digestion products (monomers, dimers, trimers, etc.), especially when it acts on G-rich regions. These products ambiguously match to more than one location in the sequence leading to data interpretation ambiguity. Further, identical small segments of an RNA molecule are not useful for computing the sequence coverage. This leads to coverage gaps and low confidence sequence information.

To address these limitations, Waters™ has developed a new set of recombinant endonucleases for RNA digestion that exhibit cleavage specificities that are complementary to RNase T1. Recombinant RapiZyme MC1 and RapiZyme Cusativin exhibit optimal activity recognizing and cleaving dinucleotide sites as demonstrated in Figure 1.[2] Furthermore, both RapiZyme RNases also have a propensity for producing partial digestions which produce unique, overlapping products amenable to LC-MS analysis and sequence determination. This application note demonstrates the use of RapiZyme MC1 and RapiZyme Cusativin in simple one-pot protocols to achieve reproducible digestions of a CRISPR single guide (sg)RNA. Use of these enzymes along with IonHance™ HFIP (hexafluoroisopropanaol) and RNA data analysis workflows involving micro apps such as mRNA Cleaver, MAP Sequence, Coverage Viewer and UNIFI processing workflows help enable faster LC-MS characterization of RNA therapeutics.[3] With their reproducibility and compliant ready architecture, these tools are suitable for use in both discovery and QC labs in support of regulatory filings and easy-to-deploy future options for release testing.

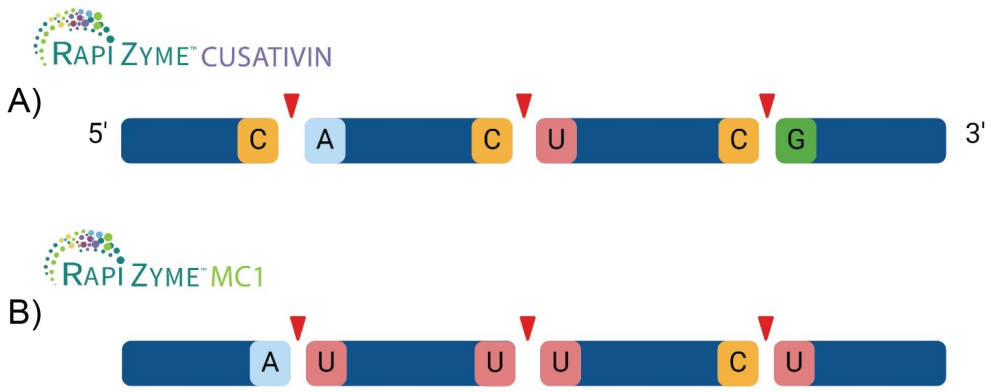


Figure 1. Example illustration showing the primary cleavage location of A) RapiZyme Cusativin Cp[A/U/G] and B) RapiZyme MC1 at [A/U/C]pU. A list of secondary cleavage sites can be found in the RapiZyme RNases Specificity section under Results and Discussion.

Experimental

Sequence			
mG*mA*mU*GAUCUCUCAACUUUAACGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCmU*mU*mU*U			
Elemental composition	Length (nt)	Theoretical MW (monoisotopic)	Theoretical MW (average)
C957H1187N374O690P99S6	100	32,212.16 Da	32,227.55 Da

An asterisk (*) denotes a PS bond and an="m" denotes a 2'-O-methyl (OMe) modification.

RNase free environment: Prior to RNA work, the workspace was made RNase-free by decontaminating the area with 70% isopropyl alcohol (IPA). All reagents used were RNase-free (certified) and prepared with RNase-free water.

Enzyme preparation: One vial (10,000 Units) of RapiZyme MC1 was dissolved in 200 µL of 200 mM ammonium acetate buffer at pH 8 creating a stock solution of 50 U/µL of RNase. One vial (10,000 Units) of RapiZyme Cusativin was dissolved in 200 µL of 200 mM ammonium acetate buffer at pH 9 creating a stock solution of 50 U/µL of RNase.

sgRNA Digestion with RapiZyme MC1

100 pmol (3.2 µg) of Hs.Cas9.HPRT1.2.AA sgRNA (IDT Technologies) was diluted to 10 µL with 200 mM ammonium acetate pH 8 and denatured at 90 °C for 2 minutes followed by rapid cooling at 4 °C. 1 µL of the RapiZyme RNase stock solution (50 Units total) was diluted to 10 µL of 18.2 MΩ water and added to each sample, gently vortexed and centrifuged to ensure mixing. The digestion mixture was incubated for 1 hour at 30 °C. The enzyme was heat inactivated by incubation at 70 °C for 15 minutes.

sgRNA digestion with RapiZyme Cusativin

100 pmol (3.2 µg) of Hs.Cas9.HPRT1.2.AA sgRNA (IDT Technologies) was diluted to 10 µL with 200 mM ammonium acetate pH 9 and denatured at 90 °C for 2 min followed by cooling at 4 °C. 1 µL of the RapiZyme Cusativin stock solution (50 Units total) was diluted to 10 µL of 18.2 MΩ water and added to each sample and gently vortexed and centrifuged to ensure mixing. The digestion mixture was incubated for 1 hour at 30 °C. The enzyme was heat inactivated by incubation at 75 °C for 15 minutes.

sgRNA digestion with RNase T1

100 pmol (3.2 µg) of Hs. Cas9.HPRT1.2.AA sgRNA (IDT Technologies) was diluted to 10 µL with 200 mM ammonium acetate pH 7 and denatured at 90 °C for 2 minutes followed by cooling at 4 °C. 250 units RNase T1 (Worthington Biochemical Corporation) diluted with 10 µL of 18.2 MΩ water was added, gently vortexed and centrifuged to ensure mixing. The digestion mixture was incubated for 1 hour at 37 °C. The enzyme was heat inactivated by incubation at 80 °C for 15 minutes.

sgRNA digestion with RNase 4

10 µg of Hs. Cas9.HPRT1.2.AA sgRNA (IDT Technologies) in 6.25 µL was mixed with 3.75 µL of 8 M urea to achieve a final concentration of 3M urea as recommended by NEB. The RNA was denatured at 90 °C for 5 minutes followed by cooling at 25 °C. The denatured RNA mixture was diluted into 20 µL of 1.5 X NEBuffer™ r1.1 for a final 30 µl volume as recommended by NEB and 1 µL of RNase 4 (NEB #M1284) stock (50U/µL) was added. The mixture was incubated at 37 °C for 1 hour. The reaction was stopped by adding 1 µL of NEB RNase inhibitor (NEB #M0314) followed by incubation at room temperature for 10 minutes.

Each digestion mixture was subsequently transferred to low bind QuanRecovery™ vials (p/n: 186009186 < https://www.waters.com/nextgen/global/shop/vials-containers--collection-plates/186009186-quanrecovery-with-maxpeak-hps-12-x-32-mm-screw-neck-vial-300--l-.html>) and an equivalent amount of RNA digest (50 pmol or 1.6 µg equivalent) for each enzyme was injected for LC-MS analysis.

LC Conditions

Column:	ACQUITY™ Premier Oligonucleotide BEH™ C_{18}, 300 Å, 1.7 µm, 2.1 x 50 mm (p/n: 186010539)
Mobile phase A:	0.1% N,N-diisopropylethylamine (DIPEA) and 1% (v/v) IonHanceHFIP (p/n:186010781) in 18.2 MΩ water
Mobile phase B:	0.0375% DIPEA and 0.075% (v/v) IonHance HFIP (p/n:186010781) in 55:10:35 Acetonitrile:Methanol: 18.2 MΩ

LC system:	ACQUITY Premier BSM System
Detector:	Xevo™ G3 QTof
Wavelength:	254 nm
Flow rate:	0.4 mL/min
Injection:	10 µL
Column temperature:	70 °C
Sample temperature:	4 °C

Gradient Table

Time	%A	%B	Curve
0.0	99.0	1.0	Initial
2.0	99.0	1.0	6
36.0	80.0	20.0	6
41.0	30.0	70.0	6
42.0	99.0	1.0	6
52.0	99.0	1.0	6
53.0	99.0	1.0	11

MS Conditions

Source type:	ESI
Polarity:	Negative
Analyzer mode:	Sensitivity
Capillary voltage:	1.5 kV

Sample cone voltage:	40V
Source temperature:	100 °C
Desolvation temperature:	550 °C
Cone gas:	50 L/h
Desolvation gas:	650 L/h
Scan mass range:	550–2000
Scan time:	1.0 sec
Collision energy:	Low 6 V
High energy ramp:	10–45 V
Intelligent dat capture - intensity threshold:	Medium (10)
Lock correction:	Automatic (30 sec intervals)
Events: LC, sample:	2.0–40.0 min
Lock mass: leucine enkephalin:	554.26202 $(M\text{-}H^{+})^{1-}$
Lock mass: combine width and mass window:	3 scans, 0.5 *m/z*

Results and Discussion

The purpose of this study was to evaluate the batch-to-batch performance of the RapiZyme RNases for RNA digestion and their utility for oligo mapping of therapeutic RNAs. In this study, we investigated the analysis of a CRISPR sgRNA designed for targeting the human HPRT (Hypoxanthine-guanine phosphoribosyltransferase) gene (Hs.Cas9.HPRT1.2.AA).

RapiZyme RNases Specificity

RapiZyme MC1 and Cusativin exhibit a preference for substrate docking and cleavage at specific dinucleotide locations. These preferences are outlined in Table 1. RapiZyme MC1 cleaves RNA at the 5′ end of uridine at ApU, CpU, and UpU sites with a slightly lower preference for CpA and CpG dinucleotide combinations. Incidentally, MC1 does not cleave GpU bonds. RapiZyme Cusativin exhibits slightly broader specificity cleaving RNA at the 3′ end of cytidine at CpA, CpG, CpU sites and with a slightly lower preference at ApU, GpU, UpU, and UpA bonds. Further, Cusativin exhibits little or no cleavage at CpC bonds. Despite having slightly broader specificity, these enzymes can be used to reproducibly achieve partial digestions, and they do not require protective hybridization oligos to achieve such results. This capability provides the opportunity for generating digestion products with overlapping features to increase the sequence coverage and facilitate confident data interpretation. Partial digestion through missed cleavages is a unique feature of both RapiZyme MC1 and Cusativin that creates more diverse digestion components and thus more chances to investigate sequence components and structural details of an RNA analyte. The mass values generated from these longer digestion products reduce the frequency of observed isomers that require more rigorous MS/MS analysis and sequence interpretation.

Informatics for data interpretation: The Waters MicroApp *mRNA Cleaver* predicts the *in silico* mass values of the expected digestion products for common digestion enzymes, including RapiZyme RNases, based on their cleavage specificity. This file is used to match and identify the observed digestion products in the LC-MS analysis through a waters_connect app *MAP Sequence*. The results are used to calculate and visualize the sequence coverage with *Coverage Viewer* for unique or unique and non-unique digestion products. These capabilities are highlighted using HPRT sgRNA as a substrate RNA analyte in the current studies.

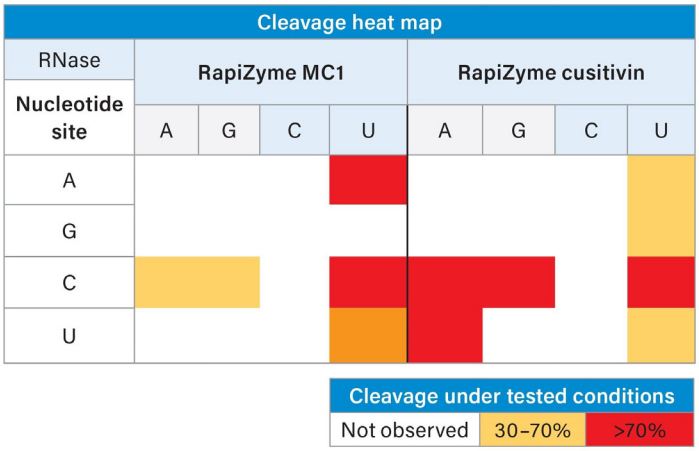


Table 1. Approximation of relative cleavage reaction rates for RapiZyme MC1 and RapiZyme Cusativin at each dinucleotide combination.

RapiZyme RNases Batch to Batch Reproducibility

In oligonucleotide mapping sample preparations, reproducible digestion of the RNA substrate is paramount. Reproducible digestion behavior of RapiZyme MC1 and RapiZyme Cusativin using HPRT sgRNA was evaluated across multiple batches in triplicate. Figure 2 shows LC-UV-MS analysis of the resulting enzyme digests. Both enzymes exhibited excellent reproducibility across UV profiles with minimum differences in peak height, although RapiZyme MC1 showed minor differences in abundance of a subset of digestion products (Figure 2A and 2B). Furthermore, MS signal was evaluated by use of the MAP Sequence app to identify the digestion products and their retention times based on mRNA Cleaver input. *In silico* predicted digestion products were generated with 4 missed cleavages using mRNA Cleaver. The digestion product file so created was used to identify the digestion products in the LC-MS data using the MAP Sequence application. After manual verification of any alternative assignments in the MAP Sequence output, a UNIFI template excel file was prepared denoting the oligonucleotide digestion product, elemental composition, and retention time. To demonstrate and visualize the reproducibility, this file was imported as library components into the waters_connect administration tool so that a UNIFI processing method for data analysis could be applied to all injections. From this MS analysis, extracted ion chromatogram overlays were prepared in the review mode. These are shown in Figure 2C for RapiZyme Cusativin and Figure 2D for RapiZyme MC1.

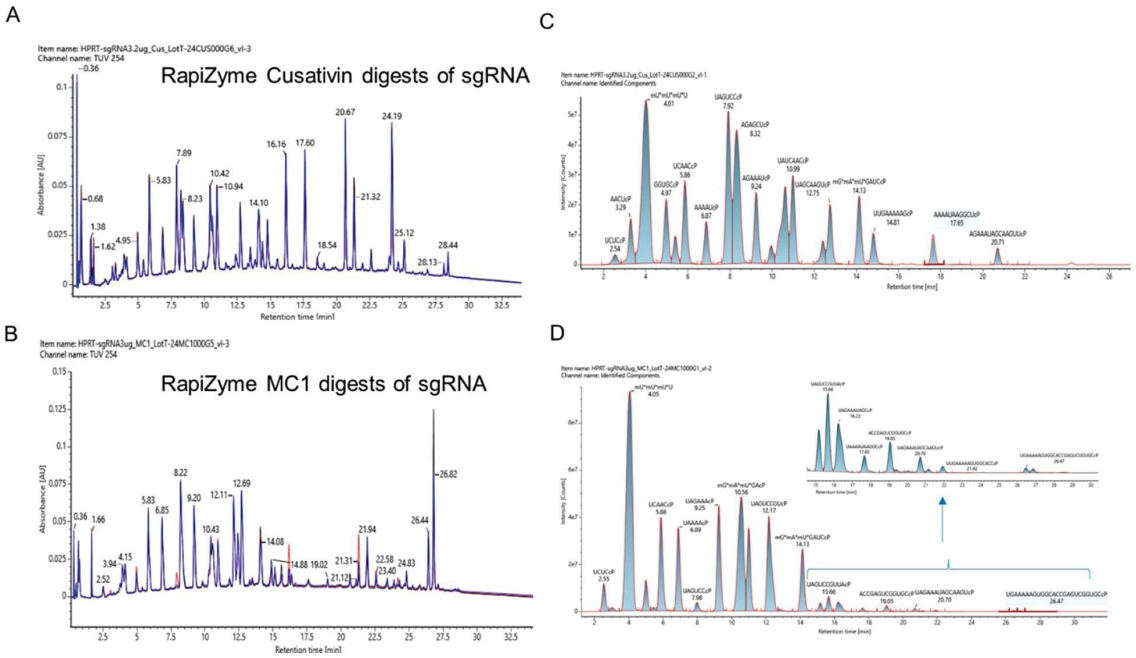


Figure 2. LC-UV-MS analysis of the HPRT sgRNA digested with recombinant RapiZyme Cusativin and MC1. Three independent recombinant enzyme batches of RapiZyme Cusativin (A) and RapiZyme MC1 (B) were used to digest HPRT sgRNA and the three UV chromatograms (black, blue and red traces) are overlayed on top of each other to indicate the reproducible digestion behavior of three preparations. (C) Extracted ion chromatogram overlay of the identified RapiZyme Cusativin digestion products indicating their abundance in the representative TUV trace. (D) Extracted ion chromatogram overlay of the identified RapiZyme MC1 digestion products indicating their abundance in the representative TUV trace. Low abundant digestion product profile of RapiZyme MC1 digestion products are shown in the inset figure.

To compare the digestion product profiles of RapiZyme Cusativin and MC1 to RNase T1 and RNase 4, sgRNA was digested with all four RNases as described in the methods section. The total ion chromatograms (TICs) of the four RNases are shown in Figure 3. A rich diversity of digestion products with abundant MS signal can be easily observed with RapiZyme MC1 and Cusativin digests compared to RNase T1 and RNase 4. Furthermore, the signal intensity was weaker for RNase 4 digest even though a similar mass load (50 pmol equivalent digest) was injected for analysis.

LC-MS data analysis of the RNase T1 and RNase 4 digests revealed the presence of discrete set of digestion products with no overlapping features even though missed cleavages were considered for these enzymes in the mRNA Cleaver Microapp based digestion product predictions. The observed digestion products for each enzyme are separately overlayed onto the HPRT sgRNA sequence as shown in Figure 4 for comparison purposes.

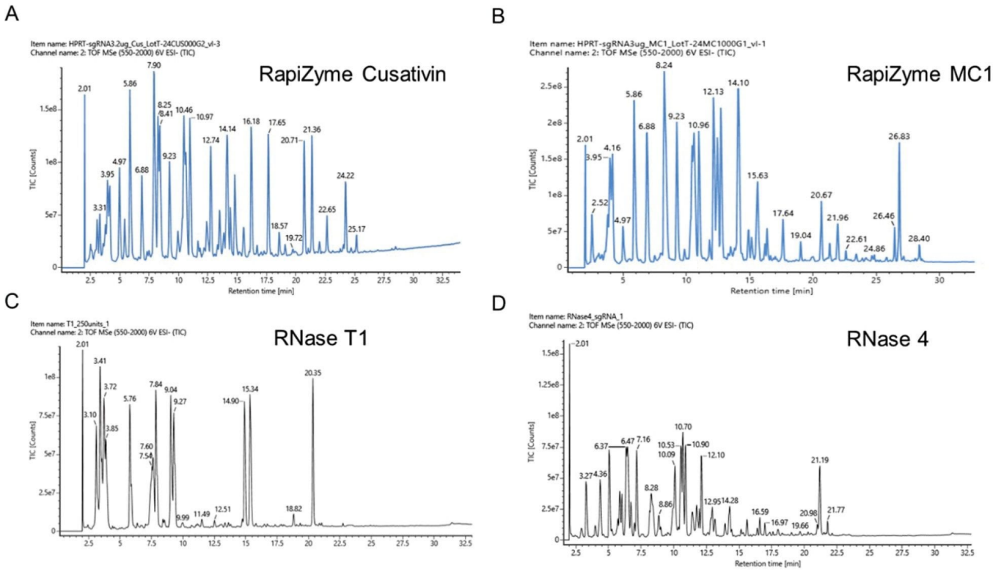


Figure 3. LC-MS analysis of HPRT1 sgRNA digests made with RapiZyme Cusativin (A), RapiZyme MC1 (B), RNase T1 (C) and RNase 4 (D).

Both RapiZyme MC1 and Cusativin digestion products exhibited overlaps which provide extra sequence location and context, thereby increasing the confidence in data interpretation. These sequence overlaps are possible because of the reproducible partial digestion of RNA by RapiZyme MC1 and Cusativin. Moreover, complete (100%) sequence coverage was achieved when observed digestion products were mapped to the sgRNA sequence for both RapiZyme MC1 and Cusativin. In contrast, RNase T1 yielded partial sequence coverage (~78%) with gaps in coverage and ambiguity where a single digestion product can be mapped to at least two regions of the sequence. Although RNase 4 digest provided complete sequence coverage, its digestion products are discrete with no sequence overlaps.

Regarding the terminal phosphate type on the digestion products, RapiZyme MC1 and Cusativin digests exhibited predominantly cyclic phosphates while the RNase 4 exhibited equivalent amounts of cyclic and linear phosphates, which helps explain the lower intensities of signal seen across the example RNase 4 data. On the

other hand, linear phosphate is the predominant form for RNase T1 digestion products (Table 2). The length of digestion products also varied with RapiZyme MC1 exhibiting the longest digestion product (27 nt) compared to other enzymes.

Figure 4. Depiction of identified HPRT sgRNA digestion products for RNase T1, RNase 4, RapiZyme MC1, and RapiZyme Cusativin. Bolded bases signify areas of the sequence where residues were detected in overlapping digestion products. Sequences that end in p (e.g. AACGUUUUp) signifies the presence of linear phosphate product detected in high abundance.

HPRT sgRNA results						
RNase	**Total preparation time (min)**	**Number of digestion products**	**Sequence coverage**	**Sequence overlaps**	**Longest digestion product (nt)**	**3′ Product formed**
RapiZyme MC1	90	37	100%	Yes	27	Cyclic phosphate (>95%)
RapiZyme Cusativin	90	51	100%	Yes	16	Cyclic phosphate (>75%)
RNase T1	75	13	72%	No	17	Linear phosphate
RNase 4	90	12	100%	No	17	50% Cyclic 50% Linear

Table 2. HPRT sgRNA Results.

Conclusion

In this work, we have established a robust analytical workflow involving sample preparation and data analysis of synthetic RNA therapeutics using IP-RP-LC-MS.

- Synthetic sgRNA was reproducibly digested using RapiZyme MC1 and RapiZyme Cusativin with as little as 100 pmol or 3.2 µg RNA consumed in the sample preparation protocol and only 50 pmol (1.6 µg) of the RNA digest injected for analysis
- High chromatographic resolution as well as excellent UV and MS signal was achieved for both short and long digestion products using ion-pairing reversed phase chromatography performed with an ACQUITY Premier Binary System, ACQUITY Premier Oligonucleotide BEH C18 Column, and IonHance HFIP mobile phase reagent
- RNA digest fragments were identified and mapped to the sequence based on accurate-mass matching as facilitated by in silico RNA digestion calculations through mRNA Cleaver, MAP Sequence, and Coverage Viewer
- Complete coverage with sequence overlaps was obtained with both RapiZyme RNases to provide confident data interpretation and output

Momentum for newer and more advanced RNA therapeutics, including mRNA vaccines, CRISPR sgRNA, and synthetic oligonucleotides, continues to build. To help address the complexities related to characterization, new solutions for sample preparation, chromatographic separation, informatics for data analysis, and data interpretation are needed. The methodology outlined in this application note with RapiZyme MC1 and RapiZyme Cusativin demonstrates how RNA therapeutics such as sgRNA can be reproducibly characterized and monitored through the generation of unique digestion products and LC-MS analysis for confident and complete sequence coverage. This work demonstrates the generation of orthogonal and additive sequence mapping information for deployment of multiple nucleases with different cleavage preferences.

More information detailing the waters_connect MAP Sequence workflow for RNA characterization can be found in RNA Digestion Product Mapping Using an Integrated UPLC-MS and Informatics Workflow, Waters Application Note - 720008553.

References

1. Gau, B.; Dawdy, A.; Wang, H.L. *et al.* (2023) Oligonucleotide Mapping Via Mass Spectrometry to Enable Comprehensive Primary Structure Characterization of an mRNA Vaccine Against SARS-Cov-2. *Scientific Reports* 13, 9038.

2. Thakur P, Atway J, Limbach PA, Addepalli B. RNA Cleavage Properties of Nucleobase-Specific RNase MC1 and Cusativin are Determined by the Dinucleotide-Binding Interactions in the Enzyme-Active Site. *Int J Mol Sci.* 2022 Jun 24;23(13):7021.

3. Gaye MM, Fox J, Vissers JPC, Reah I, Knowles C, Lauber M (2022) Synthetic mRNA Oligo Mapping Using Ion Pairing Liquid Chromatography and Mass Spectrometry. Waters Application Note – 720007669. June 2022.

Featured Products

ACQUITY Premier System <https://www.waters.com/nextgen/global/products/chromatography/chromatography-systems/acquity-premier-system.html>

Xevo G3 QTof <https://www.waters.com/nextgen/global/products/mass-spectrometry/mass-spectrometry-systems/xevo-g3-qtof.html>

RapiZyme RNases <https://www.waters.com/nextgen/global/products/standards-and-reagents/rapizyme-rnases.html>

IonHance MS-Grade Mobile Phase Additives and Buffers <https://www.waters.com/nextgen/global/products/standards-and-reagents/ionhance-ms-grade-mobile-phase-additives-and-buffers.html>

waters_connect Software Solutions <https://www.waters.com/nextgen/global/products/informatics-and-software/waters_connect.html>

UNIFI Scientific Information System <https://www.waters.com/waters/nav.htm?cid=134801359>

720008539, September 2024

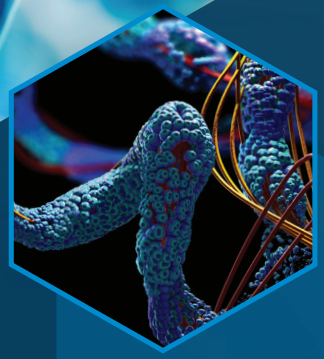
RNases for LC-MS Oligo-mapping

Generate unique digestion products and obtain complete sequence coverage with RapiZyme MC1 and Cusativin's novel specificities

LEARN MORE
waters.com/RapiZymeRNases

Waters™

RNA Oligo Mapping



RNA Oligo Mapping

Elizabeth Foley, Marketing Project Specialist, Cell & Gene Therapy, Consumables & Lab Automation

RNA oligo mapping is a critical process in the characterization and validation of mRNA sequences, which play a pivotal role in determining the amino acid sequences of proteins. The integrity of the mRNA sequence is essential, as any mutations or errors can significantly impact the function, efficacy, and safety of the resulting protein products. Regulatory bodies mandate rigorous testing and validation of mRNA sequences to ensure compliance with safety and efficacy standards prior to clinical use.

Recent advancements in analytical techniques, particularly Liquid Chromatography-Mass Spectrometry (LC-MS), have emerged as powerful tools for RNA sequence confirmation. This section's application notes discuss an integrated UPLC-MS and informatics workflow designed for digested RNA oligonucleotide mapping, with examples for sgRNA and mRNA. This approach not only confirms the RNA sequence but also allows for the detection of modified bases within the RNA, providing deeper insights into the product's structure. Such capabilities are crucial for maintaining high-quality standards in gRNA and mRNA production, as they enable manufacturers to monitor and control the integrity of their products effectively.

The use of LC-MS in a compliant environment enhances the reliability of RNA analysis, making it an attractive option for laboratories focused on mRNA therapeutics. By employing this technology, researchers can achieve greater sequence coverage and specificity, which is vital for ensuring that the RNA is free from any alterations that could compromise its intended function. The ability to map RNA digestion products accurately not only supports regulatory compliance but also fosters confidence in the safety and efficacy of RNA-based therapies and vaccines.

In summary, RNA oligo mapping through advanced LC-MS techniques represents a significant advancement in the field of molecular biology. It ensures that RNA sequences are rigorously validated, thereby safeguarding the quality of therapeutic products and aligning with the stringent requirements set forth by regulatory authorities. As the demand for RNA therapeutics continues to grow, the importance of reliable RNA mapping techniques will only increase, paving the way for innovative solutions in biotechnology and medicine.

Waters™

Application Note

RNA Digestion Product Mapping Using an Integrated UPLC-MS and Informatics Workflow

Catalin E. Doneanu, Chris Preston, Matt Gorton, Tatiana Johnson, Bala Addepalli, Ying Qing Yu

Waters Corporation

Abstract

LC-MS analysis is an indispensable tool for single-guide RNA (sgRNA) analysis due to its sensitivity, specificity, and ability to detect a wide array of oligonucleotide modifications. This application note highlights new informatics tools designed for fast data processing of LC-MS datasets acquired following the enzymatic digestion of sgRNAs. The new workflows encompass data processing software, including the waters_connect™ MAP Sequence App. Enzymatic digested oligonucleotide products are generated *in-silico* using the mRNA Cleaver MicroApp, that the MAP Sequence App matches to the experimental data, and the results are visualized using the Coverage Viewer MicroApp. The workflow described here highlights sample preparation using an array of RNase T1, T2 (RapiZyme™ MC1 and Cusativin), and hRNase4 ribonucleases, through LC-MS analysis and data processing. Unique cleavage sites offered by various enzyme specificities combined with new informatics tools (using accurate mass measurement) enable better sgRNA sequence confirmation in a quick and streamlined fashion, increasing the efficiency of drug development and the release of RNA based products to clinic or customers.

Benefits

- New compliance-ready informatics streamlines RNA Product Mapping using UPLC-MS
- RNase T2 enzymes offer unique cleavage sites and more opportunity for overlapping digest products, compared to conventional RNase T1 digests
- Combining the results from a panel of enzymatic digestions improves overall confidence in the accuracy of the mass fingerprinting-based approach, and opportunities for higher or complete sequence coverage

Introduction

Single guide RNAs (sgRNAs) were first described in the original CRISPR paper published in 2012 when the two RNA molecules required for cleavage of double-stranded DNA (the tracRNA – Cas9 enzyme scaffold and the crRNA – involved in DNA target recognition) were fused into a single RNA construct.[1] sgRNAs are a key component of the CRISPR-Cas9 gene editing system.[1–2] The sgRNA molecule directs the Cas9 nuclease to create double-stranded breaks in DNA, opening the door for genetic editing. The discovery of the CRISPR technology, which was recognized with a Nobel Prize for Chemistry in 2020, revolutionized the field of gene editing applications. One of its major applications was the development of simple and rapid COVID-19 tests during the 2020 global pandemic, but there are many other applications in research and personalized medicine for CRISPR technology including the treatment of genetic diseases, cancer, and infectious diseases.[3] sgRNAs are typically produced by solid-phase oligonucleotide synthesis, and their analytical characterization involves verification of their intact molecular weight, as well as complete sequence verification.[4,5]

Current workflows for sgRNA digest product mapping are laborious and time consuming, involving manual data analysis. RNA digestion strategies are not yet optimized and often result in generation of large numbers of ambiguous oligonucleotide digest product,which provide little utility in the confirmation of the sequence of RNA molecules. In a 2023 published application note, we discussed the UPLC-MS and informatics workflow we developed to automatically map sgRNAs and mRNAs sequences following RNase T1 digestion.[6] This current application note details a revised workflow utilizing multiple digestion enzymes, in independent reactions, to increase the sequence coverage through cumulative unambiguous digest product assignments. Here we are reporting digest product mapping results from an array of enzyme digestions, including RNase T1, T2 (RapiZyme MC1 and Cusativin), and hRNase4. Another major enhancement of this workflow, comes from the introduction of a new waters_connect App (MAP Sequence), used for automated assignment of digested RNA product maps, in a quick (less than 3 minutes) and streamlined manner.

Experimental

Reagents and Sample Preparation

Dipropylethylamine (DPA, 99% purity, catalogue number D214752–500ML) and 1,1,1,3,3,3-hexafluoro-2-propanol (HFIP, 99% purity, catalogue number 105228–100G) were purchased from Millipore Sigma (St Louis, MO). Methanol (LC-MS grade, catalogue number 34966–1L) was obtained from Honeywell (Charlotte, NC). HPLC grade Type I deionized (DI) water was purified using a Milli-Q system (Millipore, Bedford, MA). Mobile phases were prepared fresh daily. Ultrapure nuclease-free water (catalogue number J71786.AE) for sgRNA/mRNA digestions was purchased from Thermo Fisher Scientific (Waltham, MA).

Ten nanomoles of a 100-mer sgRNA oligo encoding for the HPRT1 enzyme (hypoxanthine phosphoribosyl-transferarase 1) were purchased from Integrated DNA Technologies (Coralville, IA). The sequence of the sgRNA oligonucleotide is: 5′ - A*A*A* UCC UCA GCA UAA UGA UUG UUU UAG AGC UAG AAA UAG CAA GUU AAA AUA AGG CUA GUC CGU UAU CAA CUU GAA AAA GUG GCA CCG AGU CGG UGC U*U*U* U-3′. The RNA oligonucleotide contains a 2′-OMe modification on its first three 5′ nucleotides (A*A*A*), as well as on its last three 3′ nucleotides (U*U*U*) and the asterisk indicates that all these six nucleotides are phosphorothioated. Stock solutions of the sgRNA oligonucleotide were prepared in DI water at a concentration of 5 µM.

Fungus-derived animal free purified ribonuclease T1 (catalogue no IFGRNASET1AFLY 500KU) was ordered from Innovative Research (Novi, MI) and the lyophilized enzyme was dissolved in 5 mL of 100 mM ammonium bicarbonate (catalogue no 5.33005–50G, Millipore Sigma) to prepare a solution containing 100 units/µL. hRNase4 digestion enzyme (catalogue number M1284S) was purchased from New England Biolabs (NEB, Ipswich, MA). RapiZyme MC1 (p/n: 186011190, 10000 units/tube) and RapiZyme Cusativin (p/n: 186011192, 10000 units/tube) are two novel RNA digestion enzymes recently introduced by Waters Corporation.[7]

For sgRNA digestion with RNase T1, 2 µL of 5 µM sgRNA were mixed with 28 µL of nuclease-free water and 10 µL of RNase T1 enzyme and the digestion was allowed to proceed at 37 °C for 15 min. The digestion mixture was prepared in a QuanRecovery MaxPeak 300 µL vial. The digest was analyzed immediately by LC-MS using 5 µL injections.

For sgRNA digestion with hRNase4, 2 µL of 5 µM sgRNA in 3M urea were heat denatured at 90 °C for 5 min, followed by quick cooling to 25 °C. The denatured sgRNA was mixed with 4 µL of enzyme buffer (10X conc), 33 µL of nuclease-free water and 1 µL of hRNase4 enzyme (50 units/µL) and the digestion was allowed to proceed

at 37 °C for 60 min. The hRNase4 digestion was stopped with 1 µL of NEB RNase inhibitor (NEB catalogue number M0314S), followed by incubation at room temperature for 10 minutes.

The digestion protocols used for RapiZyme MC1 and Cusativin are very similar. For RapiZyme MC1, the sgRNA (10 µL, 2–5 µM solution) was denatured at 90 °C for 2 min in a buffer containing 200 mM ammonium acetate pH 8.0. For RapiZyme Cusativin, the sgRNA (10 µL, 2–5 µM solution) was denatured at 90 °C for 2 min in a buffer containing 200 mM ammonium acetate pH 9.0. Both samples were cooled on ice and spun in a microcentrifuge to collect the sample droplets. After adding 50 units of digestion enzyme (1 µL of either RapiZyme MC1 or Cusativin) and 8 µL of nuclease-free water to obtain a final volume of ~20 µL, the sgRNA was digested at 37 °C for 60 min in an Eppendorf thermomixer. The enzymatic digestion was stopped by exposure to 70 ºC for 15 min, to inactivate the enzymes. The digest was analyzed immediately by LC-MS using 5 µL injections.

All datasets were acquired with UNIFI™ Scientific Information System App version 3.6.0.21 and subsequently processed using the MAP Sequence App within the waters_connect informatics platform, assisted by the mRNA Cleaver and Coverage Viewer MicroApps.

LC Conditions

LC-MS system:	Xevo™ G3 QTof LC-MS with ACQUITY™ Premier System
Column:	ACQUITY™ Premier Oligonucleotide BEH C_{18} FIT Column 130 Å, 1.7 µm, 2.1 x 150 mm, (p/n: 186009487)
Column temperature:	60 °C
Flow rate:	300 µL/min
Mobile phases:	Solvent A: 10 mM DPA (dipropylethylamine), 40 mM HFIP (1,1,1,3,3,3-hexafluoroisopropanol) in DI water, pH 8.5 Solvent B: 10 mM DPA, 40 mM HFIP in 50% methanol

Sample temperature:	8 °C
Sample vials:	QuanRecovery MaxPeak HPS vials (p/n: 186009186)
Injection volume:	5 µL
Wash solvents:	Purge solvent: 50% MeOH Sample Manager wash solvent: 50% MeOH Seal wash: 20% acetonitrile in DI water

Gradient Table

Time (min)	Flow rate (mL/min)	Solvent A composition (%)	Solvent B composition (%)	Curve profile
0.00	0.3	100	0	Initial
45.00	0.3	50	50	6
45.10	0.3	15	85	6
49.10	0.3	15	85	6
50.00	0.3	100	0	6
60.00	0.3	100	0	6

MS Conditions

MS system:	Xevo™ G3 QTof Mass Spectrometer
Ionization mode:	ESI(-)
Acquisition mode:	MS^E
Acquisition rate:	1 Hz
Capillary voltage:	2.5 kV
Cone voltage:	40 V

Source offset:	60 V
Source temperature:	120 °C
Desolvation temperature:	550 °C
Cone gas flow:	50 L/h
Desolvation gas flow:	600 L/hr
TOF mass range:	340–4000 (MS^E acquisition)
Low energy CE:	6 V
High energy CE ramp:	25 to 50 V
Lock-mass:	50 pg/µL Leu Enk
Data acquisition:	waters_connect 3.6.0.21
Data processing:	waters_connect 3.6.0.21
Data processing:	mRNA Cleaver 2.0 MAP Sequence v 1.0 Coverage Viewer v 2.0

Results and Discussion

LC-MS based approaches for RNA sequence confirmation are becoming more commonly adopted by biotherapeutics organizations. Improvements in chromatographic reproducibility and resolution and the increased usability of high mass resolution MS, along with improved sensitivity and accuracy of MS systems, allow for confident sequence confirmation and detection of base modifications. Most importantly, new oligonucleotide informatics tools have made the daunting task of data analysis for RNA mapping much easier.

Workflow Features

Figure 1 highlights the workflow introduced for processing of the LC-MS^E datasets acquired for sgRNAs digested with a variety of ribonucleases. In the first step, the mRNA Cleaver MicroApp is used for generating *in-silico* digested oligo products. The second step of the workflow uses the waters_connect MAP Sequence App to automatically assign the predicted oligo digest products to LC-MS data. In the final step of Figure 1 workflow, the sequence coverage obtained for the sgRNA digested with multiple enzymes is summarized using the Coverage Viewer MicroApp.

Figure 2 shows an example of the GUI (graphical user interface) of the mRNA Cleaver MicroApp with the settings used for RNase T1 digestion of the modified sgRNA. Several other enzymes are available by default in the mRNA Cleaver MicroApp, including MazF, RNase A, hRNase4, Colicin E5, as well as two recently introduced enzymes from Waters Corporation: RapiZyme™ MC1 and RapiZyme™ Cusativin.

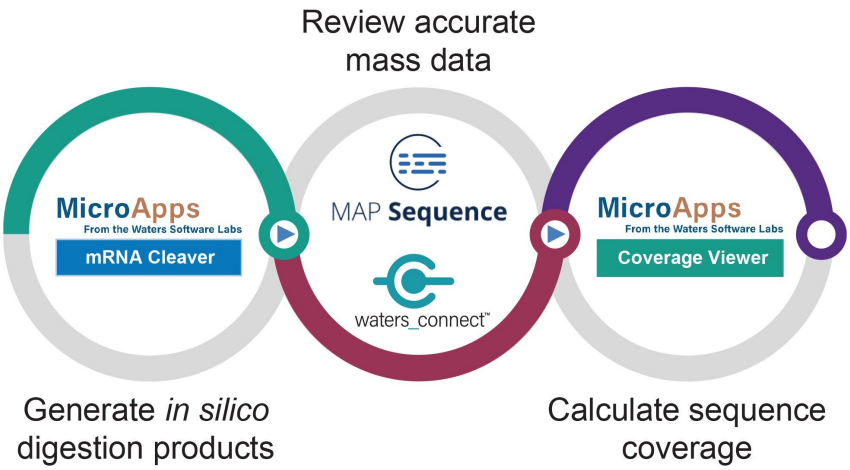


Figure 1. Simple diagram illustrating the workflow introduced for LC-MSE processing of sgRNA/mRNA digests using the mRNA Cleaver MicroApp, MAP Sequence App, and Coverage Viewer MicroApp.

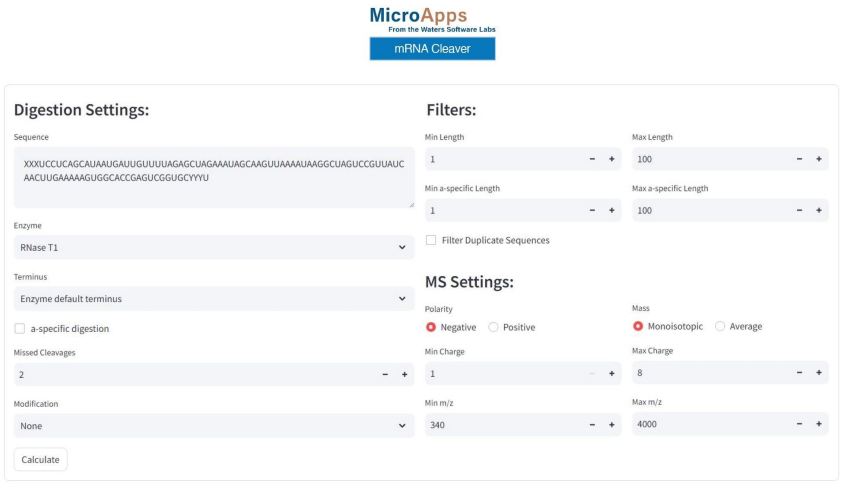


Figure 2. mRNA Cleaver settings used for in-silico prediction of oligonucleotide products resulting from the cleavage of a 100-mer modified sgRNA with RNase T1. The X residue denotes a 2'-OMe phosphorothioated adenosine, while the Y residue corresponds to 2'-OMe phosphorothioated uridine.

As mentioned in the Experimental Section, the RNA oligonucleotide contains a 2′-OMe modification and phosphorothioated linkers on its first three 5′ nucleotides (A*A*A*) denoted by letter XXX in the RNA sequence shown in Figure 2. In addition, the last three 3′ nucleotides (U*U*U*) denoted by letters YYY in the sequence listed in Figure 2, are phosphorothioated and modified with a 2′-OMe functional group. The *in-silico* digestion products predicted using the settings listed in Figure 2, assumed up to two missed cleavages for the RNase T1 enzyme, and generation of oligo products containing a linear phosphate at the 3′-position, and considered products as short as a single nucleotide. The addition of modified nucleotides is directly supported in the mRNA Cleaver MicroApp.

The simple processing parameters required for the MAP Sequence App (Figure 3) data processing include the chromatographic Retention Time (RT) range, minimum isotope response and the mass accuracy tolerance used for associating predicted and observed oligo products. A portion of MAP Sequence results obtained post-processing is highlighted within Figure 4. Unique (green circles) and ambiguous (amber triangles) oligo products assignments were obtained after matching the experimentally measured monoisotopic oligonucleotide products against the predicted monoisotopic masses.

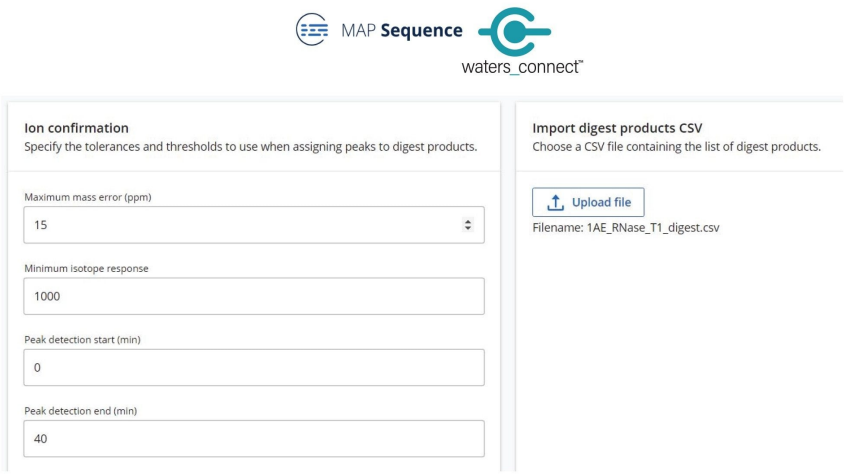


Figure 3. MAP Sequence App processing parameters. This app uses the in-silico predicted oligo products from mRNA Cleaver (see the csv file on the right side) to match them against the LC-MS^E product map dataset recorded for the digested sample.

Digest product	Sequence	Expected mass...	Observed mass (...	Response	Mass error (pp...	Observed RT (min)	Details ▲
C40:G43	CAAGp	1326.2043	1326.1995	3,326,684	-3.68	19.21	Found
U90:G92	UCGp	974.1246	974.1213	834,312	-3.38	14.50	Found
U22:G29	UUUUAGAGp	2590.3117	2590.3115	2,573	-0.06	27.02	Found
A28:G33	AGCUAGp	1977.2771	1977.2628	5,340	-7.22	29.61	Found
C30:G39	CUAGAAAUA...	3270.4599	3270.4655	643,450	1.72	33.85	Found
A34:G43	AAAUAGCAA...	3293.4871	3293.4767	2,837	-3.16	30.76	Found
C55:G62	CUAGUCCGp	2564.3324	2564.3341	3,045	0.67	26.83	Found
A74:G81	AAAAAGUGp	2659.3933	2659.3871	4,449	-2.33	28.28	Found
U90:G95	UCGGUGp	1970.2448	1970.2285	4,712	-8.24	14.50	Found
C11:G17	CAUAAUGp	2267.3075	2267.2986	5,968,973	-3.89	25.96	5 alternative peaks
A18:G21	AUUGp	1304.1611	1304.1572	3,737,728	-3.04	19.42	1 alternative peak
U22:G27	UUUUAGp	1916.2117	1916.2061	4,427,716	-2.91	23.61	3 alternative peaks
A34:G39	AAAUAGp	1985.2934	1985.2837	5,370,444	-4.88	25.09	2 alternative peaks

Figure 4. MAP Sequence App screenshot indicating unique (green circles) and ambiguous (amber triangles) oligo product assignments based on matching the experimentally measured monoisotopic oligonucleotide precursors against the predicted monoisotopic masses.

Mapping Results From a Panel of Enzymes

The total ion chromatogram (TIC) recorded for the RNase T1 sgRNA digest is displayed in Figure 5. The mRNA Cleaver MicroApp predicted 22 oligo products, listed on this figure, assuming all possible enzymatic cleavages which produce oligo products containing a terminal guanosine (G) residue. As demonstrated in this figure, there are 3 pairs of identical oligo products, highlighted by chromatographic peaks labeled in amber (sequences: AG - peaks 5/18, UG - peaks 15/21 and CUAG - peaks 6/11), that produce ambiguous sequence assignments. This shortcoming illustrates a potential limitation of RNase T1: the enzyme lacks the desired specificity to produce completely unique products, even for a relatively short substrate - like a 100-mer sgRNA. Clearly, cleavage after every G residue (a theoretical 25% chance of cleaving considering that there are only 4 unmodified nucleotides in RNAs) generates too many predicted oligo products with a high likelihood of identical oligo product sequences as the RNA sequence length increases.

Figure 5 summarizes the results of MAP Sequence assignment: all chromatographic peaks belonging to unique oligo products are labeled in green, the ambiguous sequences are labeled in amber, and the missing (unmatched) products are listed in red within the sgRNA sequence shown on the left side of this figure. Most of the abundant chromatographic peaks detected in this chromatogram are assigned to predicted products. The only other ambiguous product (besides the 3 identical pair mentioned above) is the first (5′-end) oligo product (OMEA* OMEA* OMEA* UCC UCA G) which potentially contains multiple diastereoisomers due to the presence of sulfur chiral centers in the phosphorothioated linkers of the first three nucleotides. The only missing (unmatched) oligonucleotide products are single G residues (product no 10,16,20) which cannot be retained on column under the IP-RP conditions used. The single nucleotide G residues were the only residues not detected in the LC-MSE dataset out of the entire 100-nt composition of the sgRNA and the inclusion of up to two missed RNase T1 cleavages did not improve that result. However, remarkably, the DPA ion-pairing reagent allowed for the separation of two dinucleotide species (UG – pair 15/21 vs AG – pair 5/18) as illustrated in Figure 5.

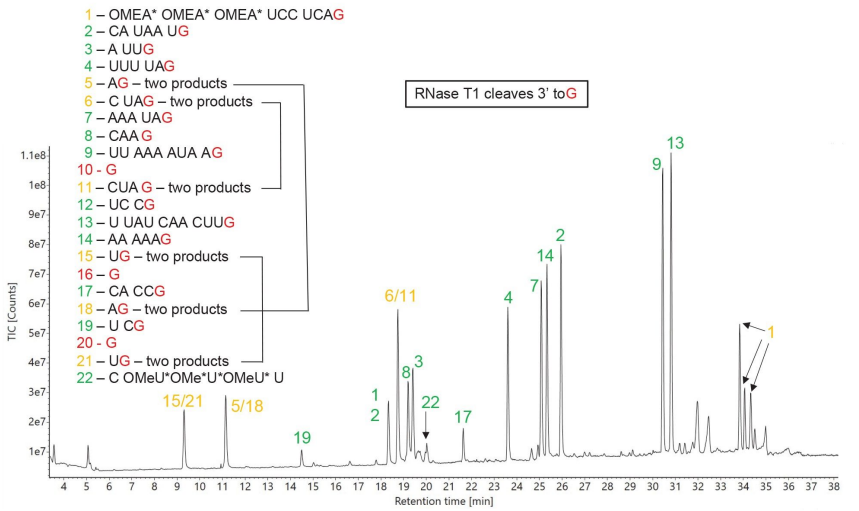


Figure 5. TIC chromatogram recorded for RNase T1 digestion of HPRT1 sgRNA. The predicted oligo products, identified according to their sequence shown the left side of this figure, are used for labeling of all the major chromatographic peaks detected. Peak identifications shown in green indicate unique sequences, while peak labels shown in amber indicate the presence of ambiguous (identical) sequences. Only three single nucleotide predicted products (G only products 10, 16, 20) were not detected out of the entire 100-mer sgRNA sequence.

In addition to RNase T1 digestion, the HPRT1 sgRNA was individually digested with hRNase4, an enzyme that recently became commercially available.[7] Unlike RNase T1, hRNase4 is a uridine (U) endoribonuclease which cleaves after U residues only if they are followed by purines (A and G) residues. The two cleavage sites of hRNase4 are defined in the mRNA Cleaver MicroApp as U_A and U_G, accounting for a theoretical chance of 12.5% for cleaving RNA substrates, which is half of the theoretical cleavage probability of RNase T1. This observation is exemplified by the TIC chromatogram from Figure 6 recorded for the hRNase4 digest of the same sgRNA substrate. The number of predicted oligo products is significantly decreased compared to the RNase T1 digestion (14 vs 22) and it matches well with the expected number of oligo products predicted for cleavage of a random 100-mer RNA sequence (12–13 products). While there are no identical predicted sequences, as was the case for RNase T1, hRNase4 produced a pair of 6-mer structural isomers (pair 5/9, AG AGC U vs A AGG CU) which contributed to ambiguous sequence assignments, because they were not resolved chromatographically under the IP-RP conditions used (which were identical to those used for separation of RNase T1 digests). This ambiguous assignment can be solved by adjusting the chromatographic conditions to separate the two 6-mer oligo products and by using their corresponding fragments (recorded in the elevated energy MS^E fragmentation data) to assign the correct sequences using the waters_connect CONFIRM Sequence App..[8]

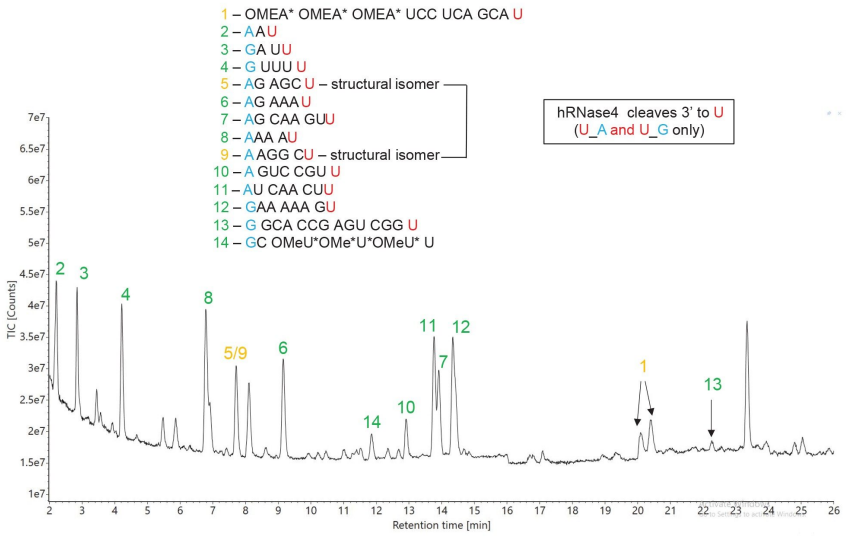


Figure 6. TIC chromatogram recorded for hRNase 4 digestion of HPRT1 sgRNA. The predicted oligo products, identified according to their sequence shown in the center of this figure, are used for labeling of all the major chromatographic peaks detected. Peak identifications shown in green indicate unique sequences, while peak labels shown in amber indicate the presence of ambiguous (identical) sequences. Only a single predicted oligonucleotide product (peak no 13) was not detected.

The only other ambiguous assignment was caused by a late eluting peak doublet corresponding to the first 5′-end oligo product (OMEA* OMEA* OMEA* UCC UCA GCA U) which also displayed heterogeneity due to its multiple chiral centers. An ESI-MS spectrum generated by a low-abundance oligo product (0.6% abundance relative to the intensity of the ESI-MS spectrum produced by the most abundant product) is shown in Figure 7. The doubly and triply charged oligo precursors detected in this spectrum confirmed the presence of a 14-mer predicted oligo product (sequence: G GCA CCG AGU CGG U) which was detected after lowering the ion isotope response threshold down to 1000 counts in MAP Sequence. As a final remark regarding RNase T1 and hRNase4 biological activity, both enzymes are cleaving after their preferred residues (3′-side of G/U residues) and while RNase T1 adds a linear phosphate to the 3′-end of the digested oligo product, hRNase4 produces digestion products containing roughly equivalent levels of both 3′ linear and cyclic phosphates detected by the software. Both RNase T1 and hRNase4 cleave the RNA substrate completely, as there is no evidence in the TIC chromatograms of undigested sgRNA (data not shown).

RapiZyme MC1 and RapiZyme Cusativin are two new enzymes recently introduced by Waters Corporation. MC1 has high uridine specificity, while Cusativin has high cytidine specificity.[9,10]

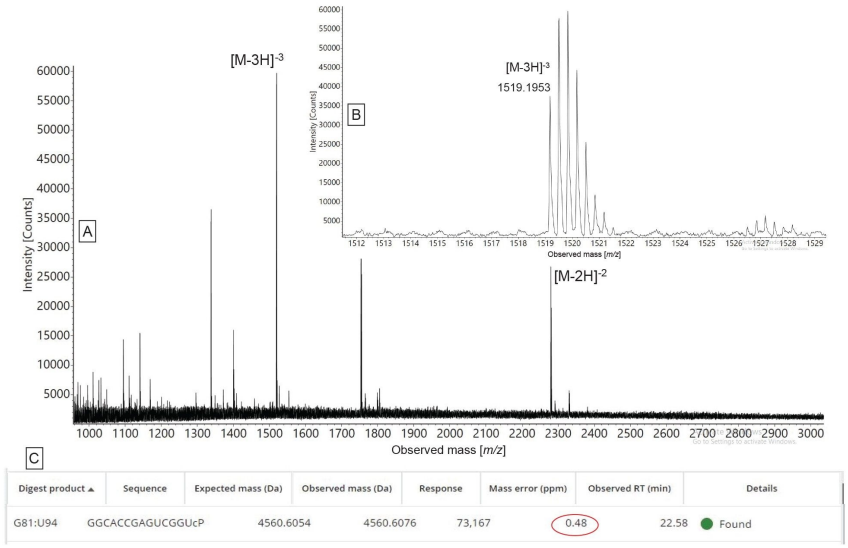


Figure 7. MAP Sequence App automated assignment of a low-intensity oligonucleotide product generated following the digestion of HPRT1 sgRNA with hRNase4 enzyme: (A) combined ESI-MS spectrum of the oligonucleotide product G GCA CCG AGU CGG U; (B) inset showing the isotopic distribution of the triply charge oligonucleotide precursor; (C) MAP Sequence App result indicating the assignment of the doubly and triply charged precursors from the ESI spectrum to this oligonucleotide product with a mass error of under 1 ppm.

The TIC chromatogram from the LC-MS analysis of the MC1 digest of the HPRT1 sgRNA substrate is displayed in Figure 8, along with the corresponding sequence assignments that resulted after MAP Sequence processing. A relatively high number of predicted oligo products was produced by the mRNA Cleaver App as a result of five different expected cleavage sites for MC1 (A_U / C_U / U_U / C_A / C_G), while also considering up to two missed cleavages. Many of the predicted short oligo products (with several single nucleotides and dinucleotides among them) were not detected in the experimental data, while most of the assigned sequence coverage utilized oligonucleotide products containing missed cleavages. This is analogous to digestion of proteins with trypsin (high-specificity) vs pepsin (low-specificity). The former produce simpler peptide maps, and the latter generate the potential for maps with overlapping sequence coverage.

It is also worth mentioning that, unlike RNase T1 and hRNase4, MC1 is cleaving in front of their specific targets (to the 5′-end of U residues) and is adding predominantly a cyclic phosphate to the 3′-end of the oligo digested product. Cusativin cleaves mainly after cytidines (3′- to C residues), but it has four other minor cleavage sites. Overall, there are seven different cleavage sites expected for Cusativin, defined in the mRNA Cleaver MicroApp as: C_U / C_A / C_G / U_U / A_U / U_A / G_U). Like MC1, Cusativin is predominantly adding a cyclic phosphate to the 3′-end of each oligo digested product, although both forms (cyclic and linear) were detected by the software.

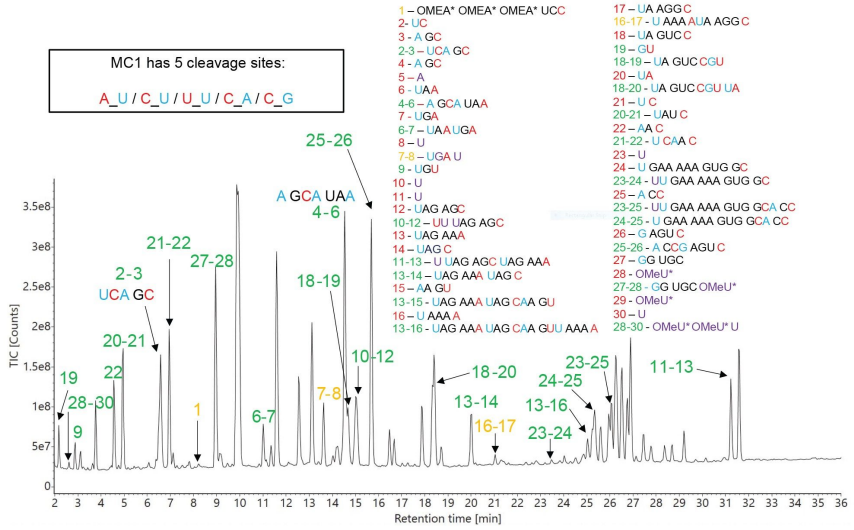


Figure 8. TIC chromatogram recorded for RapiZyme MC1 digestion of HPRT1 sgRNA. The predicted oligo products, identified according to their sequence shown in the two panels located on the right side of this figure, are used for labeling of all the chromatographic peaks detected. Peak identifications shown in green indicate unique sequences, while peak labels shown in amber indicate the presence of ambiguous (identical) sequences. Most of the sequence assignments are made from oligonucleotide products with one or two missed cleavages.

Cusativin specificity is reflected in the TIC chromatogram recorded for the HPRT1 sgRNA digest (Figure 9), as many assigned oligonucleotide products contain missed cleavages. It is clearly advantageous to have a relatively high number of missed cleavages in order to cover parts of the sgRNA sequence that might otherwise not be detected, since the corresponding products without missed cleavages are too short (single nucleotides, dinucleotides or trinucleotides) and difficult to separate by IP-RP UPLC. Both MC1 and Cusativin cleave the RNA substrate completely, as there is no evidence in the TIC chromatograms of undigested sgRNA (data not shown). A separate application note focused on RapiZyme MC1 and RapiZyme Cusativin provides more details on these unique enzymes and their use for RNA digestion.[11]

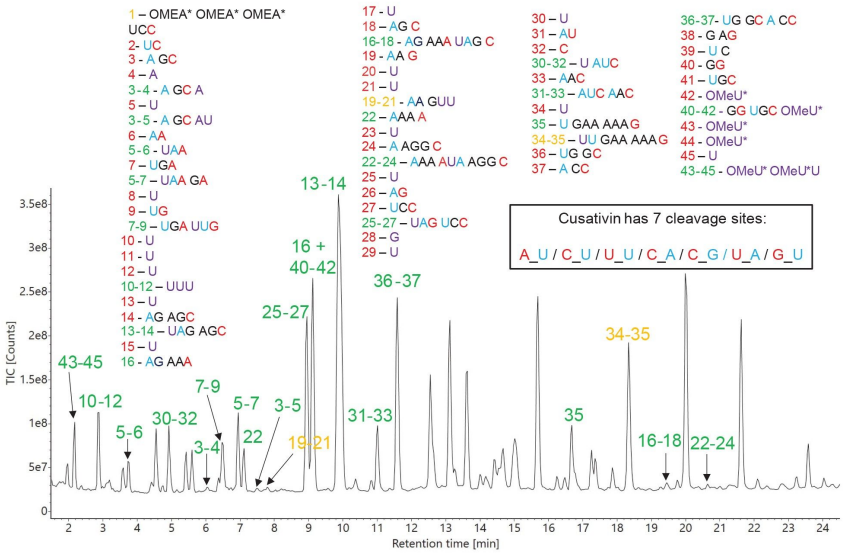


Figure 9. TIC chromatogram recorded for RapiZyme Cusativin digestion of HPRT1 sgRNA. The predicted oligo products, identified according to their sequence shown in the four panels located at the top of this figure, are used for labeling of all the major chromatographic peaks detected. Peak identifications shown in green indicate unique sequences, while peak labels shown in amber indicate the presence of ambiguous (identical) sequences. Most of the sequence assignments are made from oligonucleotide products with one or two missed cleavages.

Based on the above observations, the number of MC1 and Cusativin missed cleavages was increased to allow for up to three missed cleavages, in an attempt to obtain complete sequence coverage for these enzymes. In addition, the cleavage sites defined in the mRNA Cleaver were expanded, to include cleavages around several modified residues present in the sgRNA sequence. For example, in the case of MC1, three more cleavage sites were added: C_Y / Y_Y / Y_U, where Y denotes a 2′-OMe phosphorothioated urdine). As a result of these implementation, a newly predicted oligo product, having the sequence YYU was detected following MAP Sequence processing in both digests. The identity of this product was confirmed after processing the elevated-energy fragmentation spectrum using the CONFIRM Sequence App.[8] The isotopic distribution of its singly charged precursor along with the dot-map diagram generated by the CONFIRM Sequence App are displayed as Figure 10. The sequence of this trinucleotide was verified, indicating that both MC1 and Cusativin cleave between modified uridine residues (a Y_Y motif cleavage) which is a unique and useful enzymatic feature. The hRNase4 enzyme cannot cleave after modified uridines, but MC1 and Cusativin are not affected by 2′ uridine modifications and therefore are able to produce unique digestion products.[7]

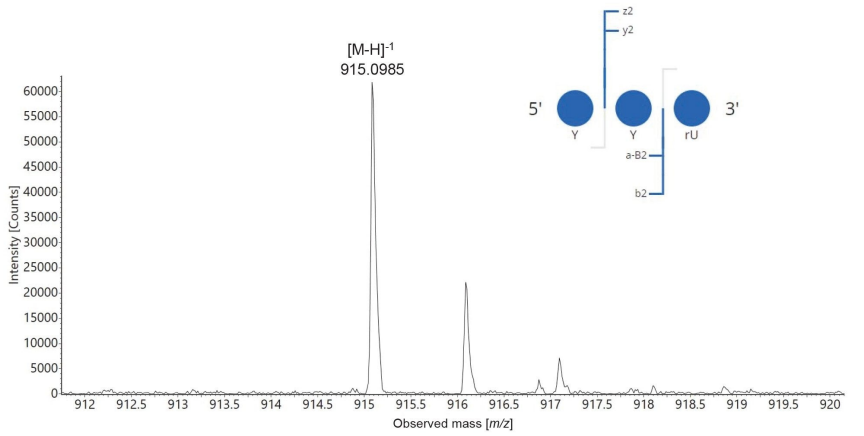


Figure 10. Isotopic distribution of a singly charged ion of the trinucleotide YYU detected in the RapiZyme MC1 digest of the modified sgRNA. The inset displays a screen capture from the CONFIRM Sequence App, showing a dot-map fragmentation diagram that confirms the sequence of this oligo digestion product as YYU (residues no 98–100 in the sequence of the 100-mer sgRNA). These results confirm the unique ability of MC1 to cleave around 2′-modified uridines, a feature that is missing for other ribonucleases.

In the final step of the RNA analysis workflow (Figure 1), a summary of the sequence coverage obtained for the same sgRNA digested with RNase T1, hRNase4, RapiZyme MC1, and RapiZyme Cusativin was produced using the Coverage Viewer MicroApp (Figure 11). The unique sequence coverage obtained individually for RNase T1 and hRNase4 is in the range of 80–90%, while MC1 and Cusativin digestion achieved 100% coverage for this molecule. Overall, these results argue that greater confidence in mapping the oligo products is realized when combining the results from multiple enzymes.

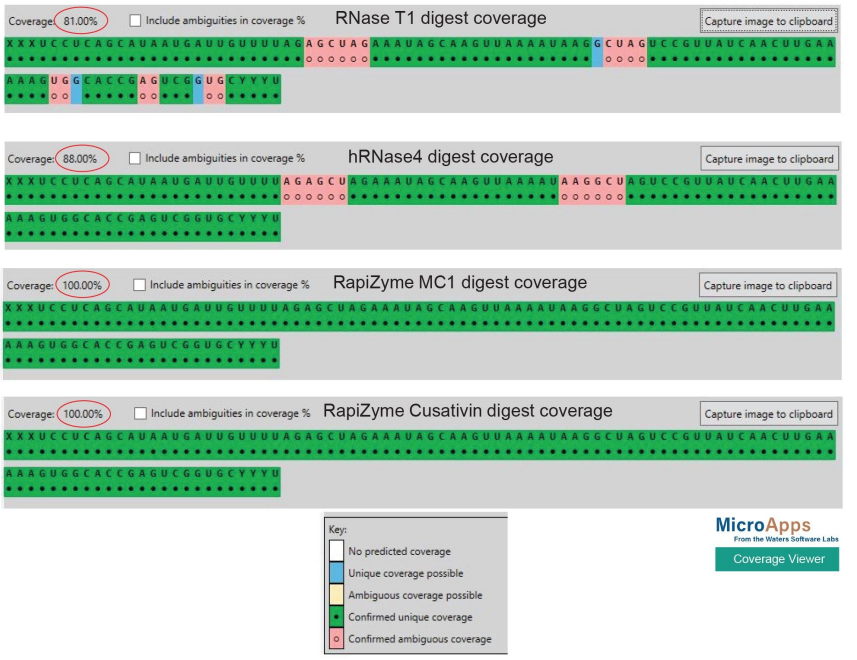


Figure 11. Coverage Viewer MicroApp results, summarizing the unique sequence coverage obtained following RNase T1, hRNAase4, RapiZyme MC1, and RapiZyme Cusativin digestions.

The use of mass spectrometry for sequence confirmation of large RNA molecules (sgRNA, mRNA) is still an area of active workflow development. The development of software automation for processing these complex datasets, and the larger toolset of digestion enzymes explored here will be key to enhancing the utility of this emerging methodology.

Conclusion

- New software tools are demonstrated for efficient RNA digestion product mapping using UPLC-MS data obtained from multiple enzymatic digests of an sgRNA. The new RNA digestion product mapping application (MAP Sequence) was used to successfully process the sgRNA digest data from each of the enzymes to generate product maps that enabled confirmation of the predicted sequence.

- The two novel RNase T2 enzymes (RapiZyme MC1 and RapiZyme Cusativin) show a greater number of oligo digestion products and missed cleavages, with unique digestion specificity. These enzymes have the potential to be useful for obtaining oligo product maps with regions of overlapping sequences.

- High sequence coverage is achieved for all enzymes, with MC1 and Cusativin providing 100% sequence confirmation at the oligo digestion product level for this molecule. In reality, this will not be obtainable for all sgRNAs molecules, and for larger mRNA sequences. Combining the results from a panel of enzymatic digestions improves overall confidence in the accuracy of the mass fingerprinting-based approach, and provides the greatest opportunities for complete sequence coverage for RNA based therapeutics.

- While oligo mapping will predominantly be used in development today, the integrated UPLC-MS data acquisition and data processing workflows on the compliance-ready waters_connect informatics platform, enables the potential use of these workflows in manufacturing and quality functions.

References

1. Jinek M, Chylinsky K, Fonfara I, Hauer M, Doudna JA, Charpentier E. A Programmable Dual-RNA-Guided DNA Endonuclease in Adaptive Bacterial Immunity, *Science*, 2012, 337, 816–821.

2. Jiang F, Doudna JA. CRISPR-Cas9 Structures and Mechanisms, *Annu Rev Biophys*, 2017, 46, 505–529.

3. Ganbaatar U, Liu C. CRISPR-Based COVID-19 Testing: Toward Next Generation Point of Care Diagnostics, *Front Cell Infect Microbiol*, 2021, 11. https://doi.org/10.3389/fcimb.2021.663949 < https://www.frontiersin.org/journals/cellular-and-infection-microbiology/articles/10.3389/fcimb.2021.663949/full>

4. Catalin E. Doneanu, Patrick Boyce, Henry Shion, Joseph Fredette, Scott J. Berger, Heidi Gastall, Ying Qing Yu. LC-MS Analysis of siRNA, Single Guide RNA and Impurities using the BioAccord System with ACQUITY Premier System and New Automated INTACT Mass Application. Waters Application Note. 720007546. 2022

5. Goyon A, Scott B, Kurita K, Crittenden CM, Shaw D, Lin A, Yehl P, Zhang K. Full Sequencing of CRISPR/Cas9 Single Guide RNA (sgRNA) via Parallel Ribonuclease Digestions and Hydrophilic Interaction Liquid Chromatography High-Resolution Mass Spectrometry Analysis, *Anal Chem*, 2022, 93, 14792–14801. doi: 10.1021/acs.analchem.1c03533 <https://pubs.acs.org/doi/full/10.1021/acs.analchem.1c03533>

6. Rebecca J. D'Esposito, Catalin E, Doneanu Heidi Gastall, Scott J. Berger, Ying Qing Yu. RNA CQA Analysis using the BioAccord LC-MS System and INTACT Mass waters_connect. Waters Application Note. 720008130. 2023.

7. Wolf EJ, Grunberg S, Dai N, Chen T-H, Roy B, Yigit E, Correa IR. Human RNase 4 Improves mRNA Sequence Characterization by LC-MS/MS, *Nucleic Acid Res*, 2022, 50, e106. DOI:10.1093/nar/gkac632 < https://doi.org/10.1093/nar/gkac632>

8. Catalin E. Doneanu, Chris Knowles, Matt Gorton, Henry Shion, Joseph Fredette, Ying Qing Yu. CONFIRM Sequence: A waters_connect Application for Sequencing of Synthetic Oligonucleotide and Their Impurities. Waters Application Note. 720007677. 2022.

9. Grunberg S, Wolf EJ, Jin J, Ganatra MD, Becker K, Ruse C, Taron CH, Correa IR, Yigit E. Enhanced Expression and Purification of Nucleotide-specific Ribonucleases MC1 and Cusativin, *Protein Expr Purif Acid Res*, 2022, 190, 105987. doi:10.1016/j.pep.2021.105987 < https://www.sciencedirect.com/science/article/pii/S1046592821001704>

10. Thakur P, Atway J, Limbach PA, Addepalli B. RNA Cleavage Properties of Nucleobase-Specific RNase MC1 and Cusativin Are Determined by the Dinucleotide-Binding Interactions in the Enzyme-Active Site, *Int J Mol Sci*, 2022, 23, 7021.

11. Balasubrahmanyam Addepalli Tatiana Johnston, Christian Reidy, Matthew A. Lauber. Tunable Digestions of RNA Using RapiZyme™ RNases to Confirm Sequence and Map Modifications. Waters Application Note. 720008539. 2024.

Featured Products

ACQUITY Premier System <https://www.waters.com/nextgen/global/products/chromatography/chromatography-systems/acquity-premier-system.html>

Xevo G3 QTof <https://www.waters.com/nextgen/global/products/mass-spectrometry/mass-spectrometry-systems/xevo-g3-qtof.html>

RapiZyme RNases <https://www.waters.com/nextgen/global/products/standards-and-reagents/rapizyme-rnases.html>

waters_connect Software Solutions <https://www.waters.com/nextgen/global/products/informatics-and-software/waters_connect.html>

UNIFI Scientific Information System <https://www.waters.com/waters/nav.htm?cid=134801359>

720008553, September 2024

Waters™

Application Note

Oligo Mapping of mRNA Digests Using a Novel Informatics Workflow

Catalin E. Doneanu, Chris Preston, Matt Gorton, Tatiana Johnson, Bala Addepalli, Ying Qing Yu

Waters Corporation

Abstract

Recent advances in LC-MS and informatics technologies are supporting scientists in their efforts to design innovative strategies for mRNA critical attribute analysis. This application note describes a workflow for mRNA sequence mapping, from sample preparation using RNase T2 enzymes, to data acquisition using UPLC™-QTof MS, and novel informatics for data analysis. The combined use of a novel RNA digestion enzyme and automated software processing for digested oligo mapping of Green Fluorescent Protein (GFP) mRNA are discussed in detail. Improved sequence coverage based on unique accurate masses of oligonucleotide digest products is illustrated. In addition to the sequence confirmation, we also demonstrated that 5′ capping efficiency analysis can benefit from use of the novel RNase T2 digestion approach.

Benefits

- A new informatics workflow featuring the waters_connect™ MAP Sequence App streamlines oligonucleotide mapping of mRNA digest data acquired using UPLC QTof MS
- RNase T2 enzymes (MC1 and Cusativin) offer unique cleavage specificity and opportunity to generate overlapping digestion products, achieving higher sequence coverage, compared to conventional RNase T1 digestion

Introduction

The recent development and approval of two COVID mRNA-based vaccines has brought RNA-based therapeutics to the forefront of the biopharma industry.[1–3] The resulting need for rapid product development has necessitated development of analytics for the precise characterization of mRNA critical quality attributes (CQAs), including sequence and modification integrity. Two workflows were previously developed by Waters™ for directly assessing mRNA CQAs, namely 5′ capping efficiency measurement and Poly(A) Tail heterogeneity analysis, utilizing the waters_connect INTACT Mass App, while a different workflow utilizing the MAP Sequence App, was recently demonstrated for oligo mapping of 100-mer single guide RNAs.[4,5,6]

Conventional oligonucleotide sequencing techniques, like Sanger sequencing and next-generation sequencing (NGS) have been applied for sequence analysis of mRNAs, due to their cost effectiveness and throughput, but lack the ability for assessment of the many modifications that can arise during production and degradation of mRNA molecules. Alternatively, LC-MS based approaches known for their exceptional specificity, sensitivity, and quantification performance are becoming more popular, and readily address base and backbone modifications.[7–10] Overall, for characterization of complex biologics, orthogonal techniques are desired to obtain a holistic view of product quality.

LC-MS workflows for RNA digest oligo mapping have historically been laborious and time consuming, involving significant manual data analysis and curation. In a recent application note, we discussed a UPLC-MS and informatics workflow for automatically mapping sgRNAs sequences following individual digestion by an array of enzymes, including RNase T1 and RNase T2 enzymes (RapiZyme MC1 and Cusativin), and hRNase4.[6] This application note extends the utility of the workflow (highlighted in Figure 1) to larger mRNA molecules digested by multiple RNA digestion enzymes, including two recently launched RNase T2 enzymes (RapiZyme MC1 and RapiZyme Cusativin), that have unique cleavage patterns that promote higher sequence coverage.[6,11] The rapid (<2 minutes) automated assignment of sgRNA digestion products and streamlined interface for data processing and review provided a significant advantage for sgRNAs and should provide even greater efficiencies for the more complex mRNA characterization exercise. In addition, digestion by RapiZyme MC1 and RapiZyme Cusativin was shown to provide superior results for the co-measurement of mRNA capping efficiency compared with the widely used RNase T1 digestion approach.

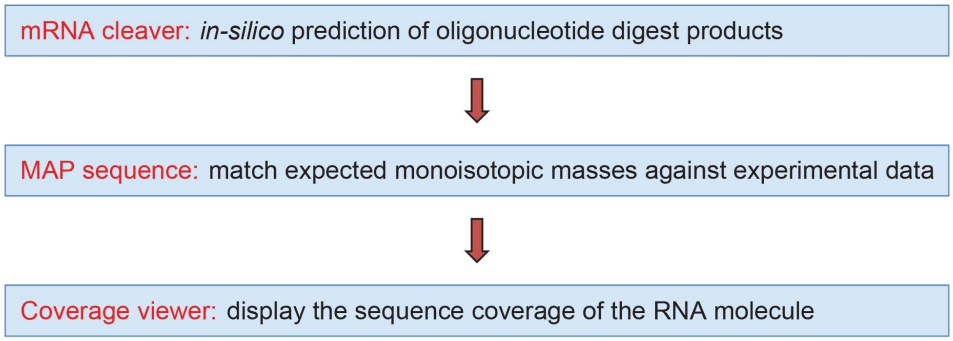


Figure 1. Workflow diagram for UPLC-MS processing of mRNA digests using the mRNA Cleaver MicroApp, MAP Sequence App and Coverage Viewer MicroApp.

Experimental

Reagents and Sample Preparation

Dipropylethylamine (DPA, 99% purity, catalogue number D214752–500ML) and 1,1,1,3,3,3-hexafluoro-2-propanol (HFIP, 99% purity, catalogue number 105228–100G) were purchased from Millipore Sigma (St Louis, MO). Methanol (LC-MS grade, catalogue number 34966–1L) was obtained from Honeywell (Charlotte, NC). HPLC grade Type I deionized (DI) water was purified using a Milli-Q system (Millipore, Bedford, MA). Mobile phases were prepared fresh daily. Ultrapure nuclease-free water (catalogue number J71786.AE) for mRNA digestions was purchased from Thermo Fisher Scientific (Waltham, MA).

An mRNA construct based on the jellyfish green fluorescent protein (GFP) sequence was custom-made via IVT (in vitro transcription) synthesis by Biosynthesis (Lewisville, TX). The mRNA molecule was synthesized with a Cap1 structure (with the sequence: 7MeGpppA(2′-OMe) with elemental composition: C32H42N15O26P5), followed by 1019 nucleotides and no Poly(A) Tail sequence.

Chromatographically purified, animal free, ribonuclease T1 (catalogue no LS01490, 500kU), isolated from Aspergillus oryzae, was purchased from Worthington Biochemical Corporation (Lakewood, NJ). The lyophilized enzyme was dissolved in 5 mL of 100 mM ammonium bicarbonate (catalogue no 5.33005–50G, Millipore Sigma) to prepare a solution containing 100 units/µL. For mRNA digestion with RNase T1, 5 µL of 5 µM GFP mRNA were

mixed with 25 µL of nuclease-free water and 10 µL of RNase T1 enzyme (1000 units) and the digestion was allowed to proceed at 37 °C for 15 minutes. The digestion mixture was prepared in a QuanRecovery™ MaxPeak™ 300 µL vial. The digest was analyzed immediately by LC-MS using 5 µL injections.

RapiZyme MC1 (p/n: 186011190 <https://www.waters.com/nextgen/global/shop/standards--reagents/186011190-rapizyme-mc1-10000-units.html> , 10000 units/tube) and RapiZyme Cusativin (p/n: 186011192 < https://www.waters.com/nextgen/global/shop/standards--reagents/186011192-rapizyme-cusativin-10000-units.html> , 10000 units/tube) are two novel RNA digestion enzymes recently introduced by Waters Corporation.[6,7] The GFP mRNA digestion protocols used for RapiZyme MC1 and Cusativin are very similar. For RapiZyme MC1, the GFP mRNA (10 µL, 5 µM solution) was denatured at 90 °C for 2 minutes in a buffer containing 200 mM ammonium acetate pH 8.0. For RapiZyme Cusativin, the mRNA (10 µL, 5 µM solution) was denatured at 90 °C for 2 minutes in a buffer containing 200 mM ammonium acetate pH 9.0. Both samples were cooled on ice and microcentrifuged to collect the sample droplets. After adding 50 units of digestion enzyme (1 µL of either RapiZyme MC1 or RapiZyme Cusativin) and 8 µL of nuclease-free water to obtain a final volume of ~20 µL, the mRNA was digested at 37 °C for 60 minutes in an Eppendorf thermomixer. The enzymatic digestion was stopped by heating to 70 ºC for 15 minutes, to inactivate the enzymes. The digest was analyzed immediately by LC-MS using 5 µL injections.

All datasets were acquired with waters_connect UNIFI™ App version 3.6.0.21 and subsequently processed using the MAP Sequence App, assisted by the mRNA Cleaver and Coverage Viewer MicroApps.

LC Conditions

LC-MS system:	Xevo™ G3 QTof LC-MS with ACQUITY™ Premier UPLC (Binary) System
Column:	ACQUITY Premier Oligonucleotide BEH™ C_{18} FIT Column 130 Å, 1.7 µm, 2.1 x 150 mm, (p/n: 186009486)
Column temperature:	60 °C
Flow rate:	300 µL/min

Mobile phases:	Solvent A: 10 mM DPA (dipropylethylamine), 40 mM HFIP (1,1,1,3,3,3-hexafluoroisopropanol) in DI water, pH 8.5 Solvent B: 10 mM DPA, 40 mM HFIP in 50% methanol
Sample temperature:	8 °C
Sample vials:	QuanRecovery MaxPeak HPS vials (p/n: 186009186)
Injection volume:	5 µL
Wash solvents:	Purge solvent: 50% MeOH Sample Manager wash solvent: 50% MeOH Seal wash: 20% acetonitrile in DI water

Gradient Table

Time (min)	Flow rate (mL/min)	Solvent A composition (%)	Solvent B composition (%)	Curve profile
0.00	0.3	100	0	Initial
45.00	0.3	50	50	6
45.10	0.3	15	85	6
49.10	0.3	15	85	6
50.00	0.3	100	0	6
60.00	0.3	100	0	6

MS Conditions

MS system:	Xevo™ G3 QTof Mass Spectrometer
Ionization mode:	ESI(-)

Acquisition mode:	MS^E
Acquisition rate:	1 Hz
Capillary voltage:	2.5 kV
Cone voltage:	40 V
Source offset:	60 V
Source temperature:	120 °C
Desolvation temperature:	550 °C
Cone gas flow:	50 L/h
Desolvation gas flow:	600 L/hr
TOF mass range:	340–4000 (MS^E acquisition)
Low energy CE:	6 V
High energy CE ramp:	25 to 50 V
Lock-mass:	50 pg/µL Leu Enk
Data acquisition:	waters_connect 3.6.0.21
Data processing:	waters_connect 3.6.0.21
Data processing:	mRNA Cleaver MicroApp v1.1.0

Results and Discussion

As biotherapeutics organizations advance their product development pipelines, increasing adoption for LC-MS based methods for RNA sequence analysis helps to achieve more thorough initial characterization and reliable ongoing sequence confirmation. Improvements in chromatographic reproducibility and resolution, and the increased usability of high mass resolution MS, along with improved sensitivity and accuracy of MS systems, allow for confident sequence confirmation and detection of base modifications. Most importantly, new oligonucleotide informatics tools have made the daunting task of data analysis for RNA map data analysis faster and easier.

While LC-MS for digested peptide mapping of biopharmaceutical proteins is routine, employing equivalent approaches to RNA is more challenging. Unlike proteins, which are composed of 20 different amino acids, mRNAs contain only four building blocks (A, C, G, T), making the generation of oligonucleotide digestion products that have unique masses mappable to mRNAs more challenging. High frequency cleavers, like RNase T1 with G-cleavage specificity, generate many isomeric or even identical mRNA digestion products that result in ambiguous assignments. Enzymes with lower frequency cleavage sites are able to produce longer digestion products, especially when they generate a population of missed cleavage oligonucleotides. The novel RNase Type2 enzymes employed in this work (RapiZyme MC1 and RapiZyme Cusativin), have a greater chance at producing assignable digestion products with unique masses.

An informatics workflow (Figure 1) featuring the waters_connect MAP Sequence App for data processing and sequence assignments was designed to facilitate fast and automated data processing of LC-MS datasets acquired following the enzymatic digestion and UPLC-MS data acquisition.[6] In the first step, the mRNA Cleaver MicroApp is used for generating the predicted in-silico digested oligonucleotide products of the specified mRNA molecule. In the second step, the waters_connect MAP Sequence App processes the UPLC-MS data and matches the predicted neutral monoisotopic masses of digested oligos to the experimental MS1 data. In the final step, the sequence coverage obtained for the mRNA digest is summarized and visualized using the Coverage Viewer MicroApp.

As the workflow utilizes the accurate mass measurement of RNA digestion products for oligo mapping, it is critical to generate unambiguous, unique products through enzymatic digestion, to reduce reliance on user intervention with MS2 fragmentation data to resolve assignment ambiguities. RapiZyme MC1 and RapiZyme Cusativin enzymes can produce longer unique digestion products through a combination of unique cleavage specificity and intentionally generated missed cleavages by regulating enzyme amounts and digestion time. Data

independent MS^E acquisition is used for confirming oligonucleotide digestion product assignments. The elevated energy MS^E fragmentation information will be incorporated into future releases of the App, particularly to address some remaining ambiguous assignments.

mRNA Oligo Mapping

Three independent LC-MS^E oligo map datasets were acquired for GFP mRNA sample digested with three different ribonucleases: RNase T1, RapiZyme MC1 and RapiZyme Cusativin. The three UPLC-MS^E datasets were processed with the MAP Sequence App using the parameters highlighted in Figure 2 and the results were exported to the Coverage Viewer Micro App for quick visualization of sequence coverage. The three TIC chromatograms recorded for: RNase T1 (Figure 3A), Rapizyme MC1 (4A) and RapiZyme Cusativin (5A) are shown. MC1 and Cusativin are known to produce a significantly larger number of misses cleavages compared to RNase T1, and this enzymatic feature is clearly reflected in the complexity of chromatograms shown in Figures 4A and 5A.[6–9]

A comparison of the sequence mapping results (Figures 3B, 4B, and 5B) indicates that RNase T1 produced considerably lower unique coverage (~60%, 3B), compared to the MC1 (~97%, 4B) and Cusativin (~85%, 5B) analyses, because the number of ambiguously assigned digestion products was much higher. This is a direct result of the much broader specificity of RNase T1 (cleavage after every G residue), in contrast with MC1 and Cusativin cleavage specificity, which relies on very specific dinucleotide motifs.[6–9] MC1 cleaves at the 5′-end of uridine residues, with three major cleavage sites (A_U / C_U / U_U) and two minor cleavage sites (C_A / C_G). Cusativin cleaves at the 3′-end of cytidine residues, with four major cleavage sites (C_A / C_G / C_U / U_A) and three minor cleavage sites (A_U / G_U / U_U). While RNase T1 adds a linear phosphate to the 3′-end of all its digestion products, both MC1 and Cusativin produce mainly digestion products with a 3′ cyclic phosphate.

Compared to RNase T1, MC1, and Cusativin also have a higher capacity to generate controlled missed cleavages, therefore up to four missed cleavages were allowed for these two enzymes in the mRNA Cleaver prediction tool, while only two missed cleavages were applied for RNase T1. With more missed cleavages, both MC1 and Cusativin produced longer digestion products compared to RNase T1. These longer oligonucleotide products are more likely to have unique masses, reducing the chance for ambiguous assignments. As a result of these enzymatic features, the mapping results of MC1 and Cusativin exbibit better coverage values.

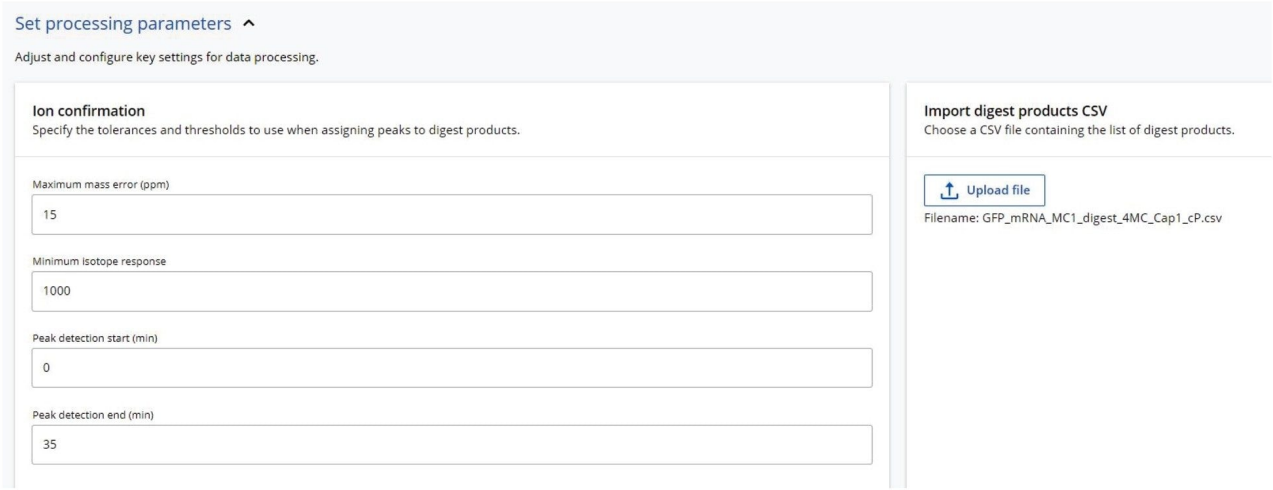
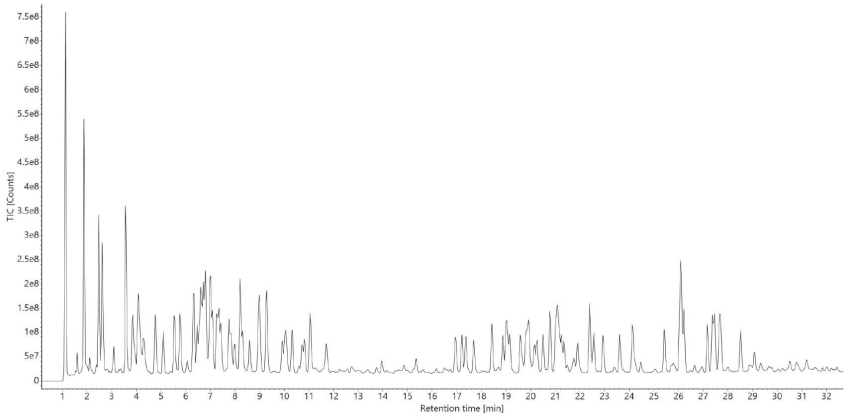


Figure 2. MAP Sequence processing parameters. This app uses the in-silico predicted digestion products from mRNA Cleaver (see the csv file on the right side) to match them against the UPLC-MS dataset recorded for the digested sample.

3A



3B

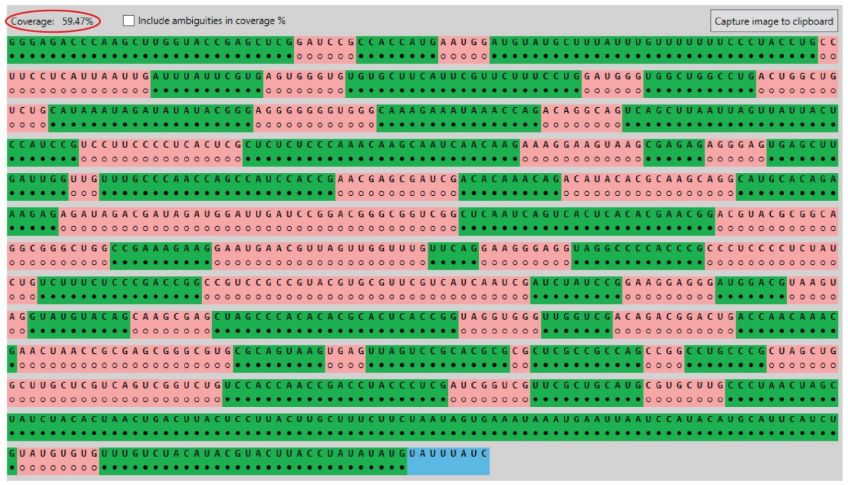
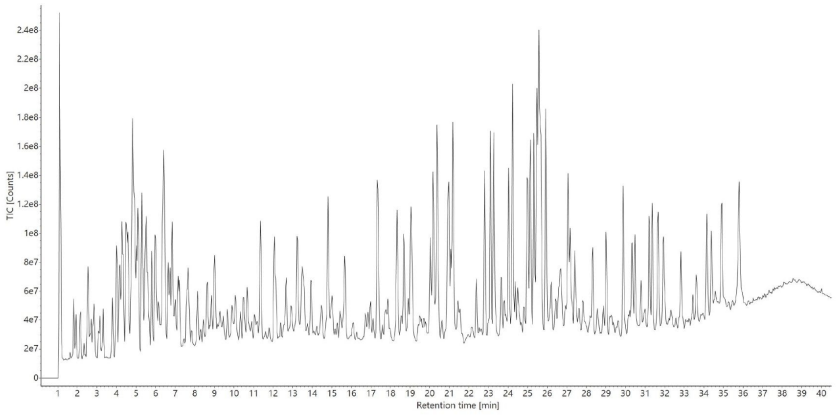


Figure 3. (A) TIC chromatograms recorded for the GFP mRNA digested with RNase T1; (B) Coverage Viewer Micro App results, summarizing and visualizing the unique sequence coverage oligonucleotides assigned following RNase T1 digestion analysis.

4A



4B

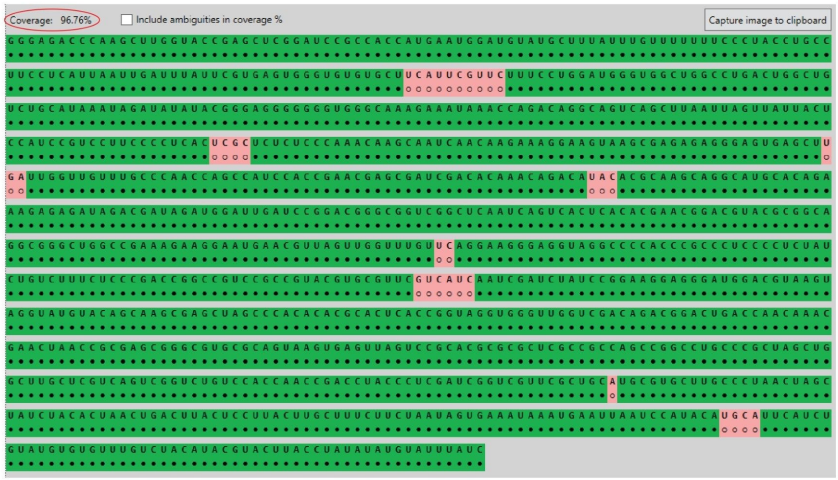
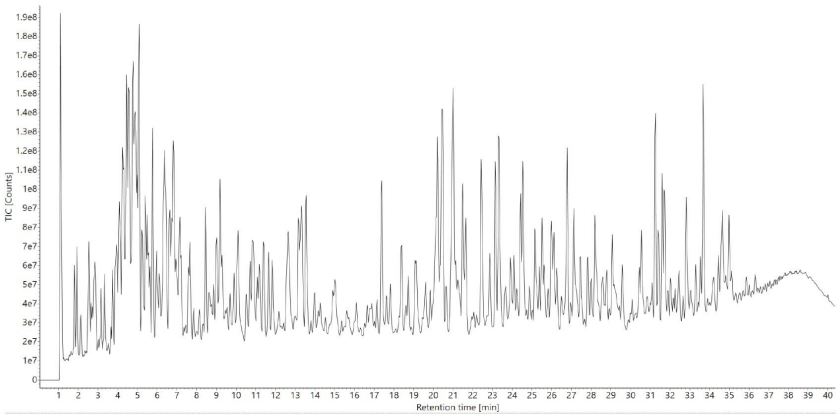


Figure 4. (A) TIC chromatograms recorded for the GFP mRNA digested with RapiZyme MC1; (B) Coverage Viewer Micro App results, summarizing and visualizing the unique sequence coverage oligonucleotides assigned following RapiZyme MC1 digestion analysis.

5A



5B

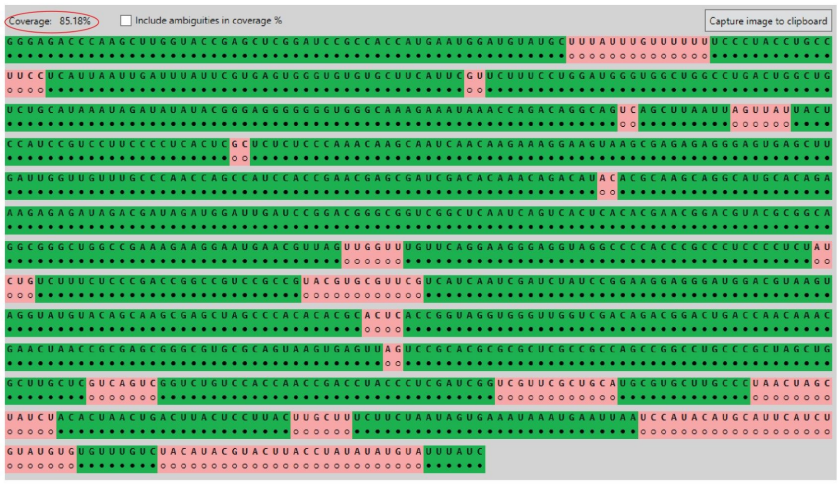


Figure 5. (A) TIC chromatograms recorded for the GFP mRNA digested with RapiZyme Cusativin; (B) Coverage Viewer Micro App results, summarizing the unique sequence coverage oligonucleotides assigned following RapiZyme Cusativin digestion analysis.

In the case of ambiguous sequence assignments, the CONFIRM Sequence waters_conect App could be used to further investigate the elevated energy MS^E data to assign the fragmentation ions to the most probable oligonucleotide sequences.[12] To illustrate this approach, we selected a pair of 16-mer digestion products, produced by RNase T1 digestion of GFP mRNA. The pair is formed by the structural isomers U251:G266 (sequence UCC UUC CCC UCA CUC G) and C555:G570 (CCC UCC CCU CUA UCU G) with the same nucleotide

composition. As shown by the inset (Figure 6 Top), MAP Sequence automatically found these two sequences based on multiple precursors (charge states 2–9) and labeled them as "alternative product assignments". The extracted mass chromatogram of the corresponding triply charged precursor (*m/z*=1654.86), also showed in Figure 6, indicates the presence of a single, high-abundance, chromatographic peak which could belong to one or both oligonucleotide products. CONFIRM Sequence was used to automatically assign the fragment ions detected in the elevated energy ESI-MS^{E} spectrum (Figure 7) recorded for this oligonucleotide digestion product to confidently reveal a match for the C555:G570 sequence with 100% sequence coverage.

Figure 6. Extracted mass chromatogram of a triply charged precursor (m/z-1654.86) that proves ambiguous as it could potentially be assigned to isomeric digestion products U251:G266 or C555:G570 resulting from RNase T1 cleavage of the GFP mRNA. The top inset shows the MAP Sequence assignment of these two structural isomers, while the right-side inset displays the charge states (2–9) detected by MAP Sequence for the ESI-MS spectrum of this oligonucleotide product.

Figure 7. CONFIRM Sequence screenshots showing the dot-map sequence coverage obtained for two possible 16-mer digestion products produced by RNase T1 digestion of GFP mRNA. The chromatographic peak detected at 26.09 minutes in Figure 4 was produced by the C555:G270 16-mer oligonucleotide (CCC UCC CCU CUA UCU G), given the 100% sequence coverage obtained. Only a few terminal fragments ions could be assigned to the alternative isomeric sequence (UCC UUC CCC UCA CUC G).

Several examples of mRNA Cap analysis have been described in previously published application notes from Waters.[5,13] The mRNA cap is added to the 5′-end of an mRNA and it typically consists of a fully matured 7 N-methyl guanosine that is 5′-5′-linked to the mRNA via a triphosphate ester (m7Gppp). In the case of GFP mRNA, a variant Cap1 structure was used, involving the 2′-O-methylation of an adenosine residue that is then attached to the mRNA. The Cap1 structure of the GFP mRNA is therefore denoted as m7GpppAm.

Both the uncapped and capped GFP 5′ mRNA digestion products were detected in the MC1 mRNA digest, as demonstrated by the extracted mass chromatograms displayed in Figure 8. Following MC1 digestion, the uncapped version (*m/z*=1419.5302, -3), eluted at 22.0 min in the top trace (Figure 8A) and the Cap1 modified species (*m/z*=1700.8962, -3), eluted at 24.7 min in the bottom trace (Figure 8B). Using the integrated peak area obtained from these triply charged state XIC peaks, the capping efficiency was calculated to be 79.8%.

The identity of these oligonucleotide species is further confirmed by the isotopic distributions (Figure 9) recorded for the most abundant charge states detected ([M-3H]3-). The monoisotopic peak of the free (uncapped) 13-mer MC1 digest oligonucleotide with the sequence GGA AGA CCC AAG C, is highlighted in Figure 8A, while the monoisotopic peak of the Cap1 modified 13-mer oligonucleotide is shown in the bottom panel (Figure 8B). In this case, both MC1 and Cusativin produced the same digestion product (sequence GGA AGA CCC AAG C) after cleaving the 5′-terminus of GFP mRNA, thus enabling this measurement of the mRNA capping efficiency. However, in the case of RNase T1 digestion, because the first mRNA residue at the 5′ end is a guanosine, the uncapped digestion product was too short (a single nucleotide) to be retained under the IP-RP conditions employed. For this reason, only the Cap1 digested product was detected (sequence m7GpppAmGp, most abundant monoisotopic ion: *m/z*=602.5482, -2, data not shown), and it was not possible to measure the mRNA capping efficiency. Therefore, in the case of GFP mRNA, only MC1 and Cusativin could be used successfully for capping efficiency calculations.

These results highlight the current utility of mass spectrometry for sequence confirmation of large mRNA molecules, but there are still gains to be realized for workflow efficiency and automation, and it remains an area of active workflow development. Continued development of software automation for processing these complex datasets, along with the expansion of the digestion enzyme toolset, will be crucial for enhancing the utility of this emerging methodology.

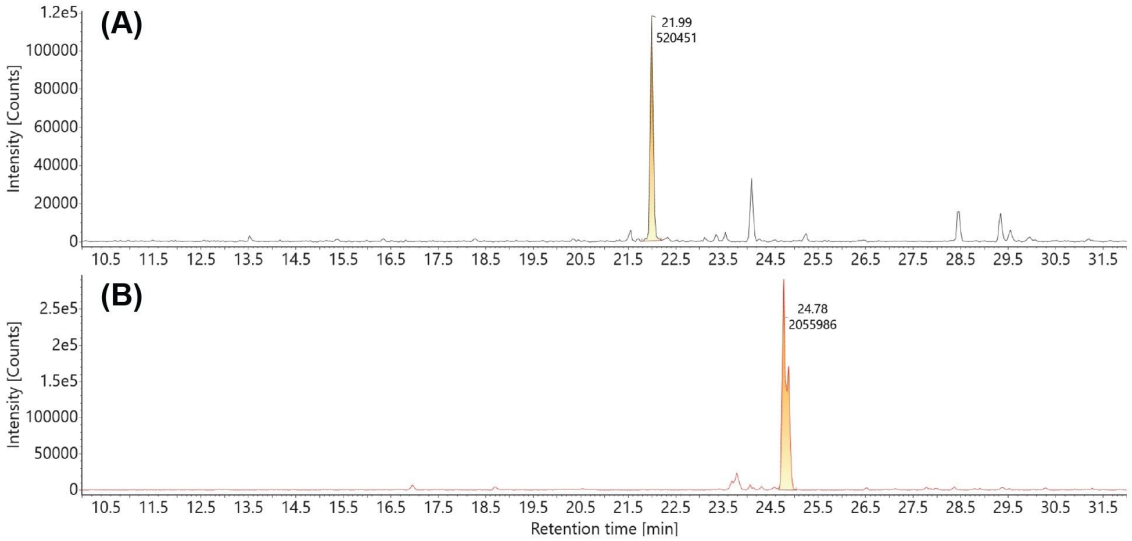


Figure 8. Extracted mass chromatograms of the triply charged precursors of the uncapped (A) and capped (B) version of the same 13-mer digestion product (GGA AGA CCC AAG C) produced by MC1 digestion: (A) m/z=1419.51 (-3) uncapped GGA AGA CCC AAG C oligonucleotide; (B) m/z=1700.87 (-3) Cap1 modified GGA AGA CCC AAG C oligonucleotide.

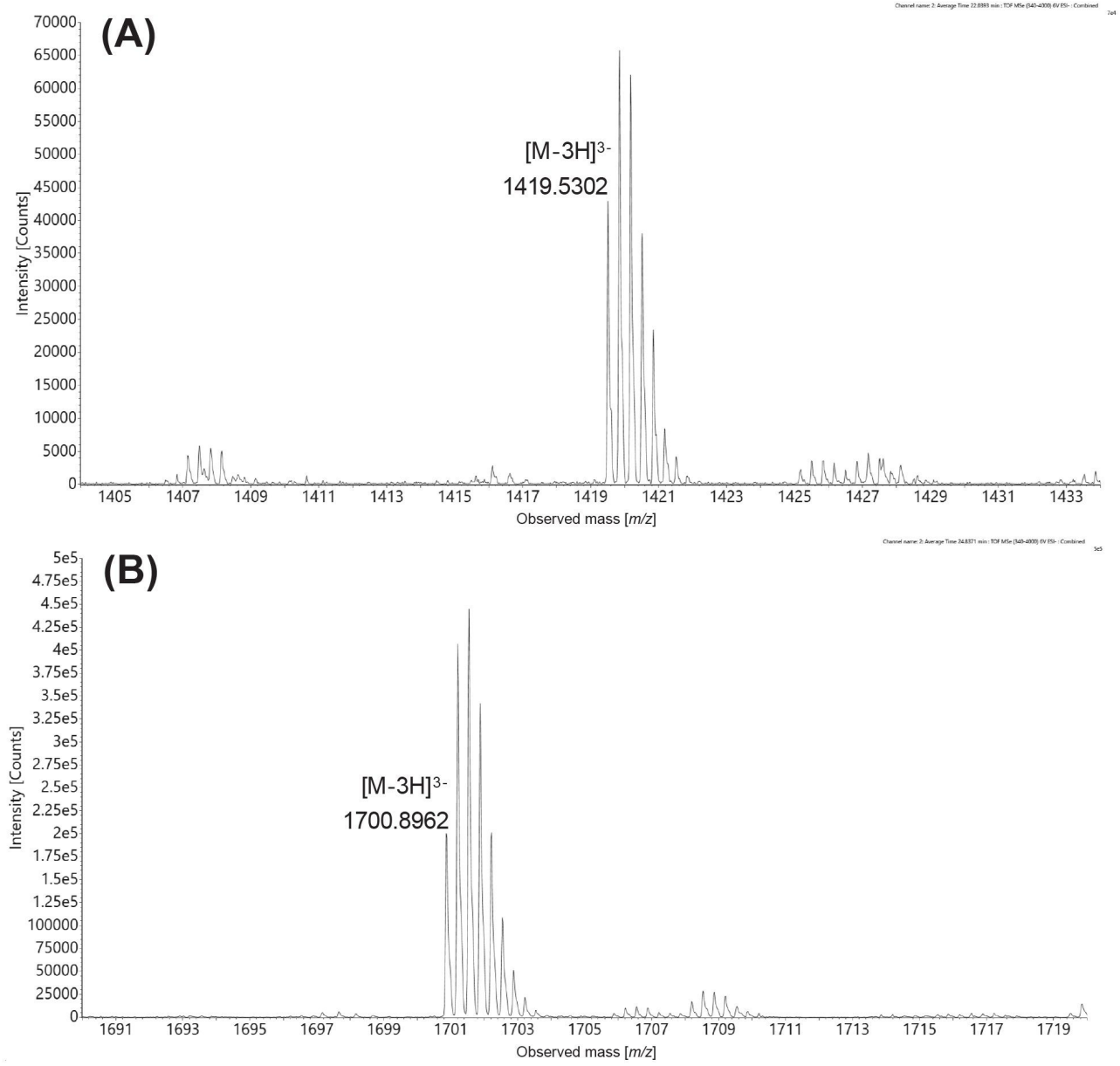


Figure 9. Isotopic distributions of the triply charged ions of the uncapped (A) and Cap1 modified (B) versions of a 13-mer oligonucleotide with the sequence GGA AGA CCC AAG C, produced after MC1 digestion of GFP mRNA. The monoisotopic peaks of the uncapped and capped oligonucleotides are labeled on the corresponding ESI-MS spectra.

Conclusion

- A new informatics workflow was demonstrated, featuring the waters_connect MAP Sequence waters App, which facilitated MS1 oligo mapping of mRNA digests using UPLC-MS acquired data
- Two novel RNase T2 enzymes (RapiZyme MC1 and RapiZyme Cusativin) generated a greater population of fully digested oligonucleotides and missed cleavage oligos, with their unique digestion specificity compared to RNase T1, resulting in higher and more confident overlapping sequence coverage.
- Isomeric-digestion oligonucleotide products can be differentiated using the waters_connect CONFIRM Sequence App to match elevated energy fragment ions to sequences of isomeric oligos

References

1. Xu S, Yang K, Li R, Zhang L. mRNA Vaccine Era-Mechanisms, Drug Platform and Clinical Prospection, *Intl J Mol Sci Chem*, 2020, 21 (18), 6582, doi: 10.3390/ijms21186582 <https://www.mdpi.com/1422-0067/21/18/6582> .

2. Verbeke R, Lentacker I, De Smedt SC, Dewitte H. The dawn of mRNA vaccines: The COVID-19 case. *J Controlled Release*, 2021, 333, 511–520, doi: 10.1016/ j.jconrel.2021.03.043 < https://www.sciencedirect.com/science/article/pii/S0168365921001541?via%3Dihub> .

3. Jackson NA, Kester KE, Casimiro D, Gurunathan S, DeRosa F. The Promise of mRNA Vaccines: A Biotech and Industrial Perspective. *npj Vaccines*, 2020, 5, https://doi.org/10.1038/s41541-020-0159-8 < https://www.nature.com/articles/s41541-020-0159-8> .

4. Synthetic mRNA Oligo-Mapping Using Ion-Pairing Liquid chromatography and Mass Spectrometry, 2022, Waters application note. June, 2022. 720007669.

5. RNA CQA Analysis using the BioAccord LC-MS System and INTACT Mass waters_connect Application, 2023, Waters application note. November, 2023. 720008130.

6. RNA Digestion Product Mapping Using an Integrated UPLC-MS and Informatics Workflow, 2024, Waters application note. September, 2024. 720008553.

7. Jiang T, Yu N, Kim J, Murgo JR, Kissai M, Ravichandran K, Miracco E, Presnyak V, Hua S. Oligonucleotide Sequence Mapping of Large Therapeutic mRNAs via Parallel Ribonuclease Digestions and LC-MS/MS, *Anal Chem*, 2019, 91, 8500–8506, doi:10.1021/acs.analchem.9b01664 <https://doi.org/10.1021/acs.analchem.9b01664> .

8. Vanhinsbergh CJ, Criscuolo A, Sutton JN, Keely M, Williamson AJK, Cook K, Dickman M. Characterization and Sequence Mapping of Large RNA and mRNA Therapeutics using Mass Spectrometry, *Anal Chem*, 2022, 94, 7339–7349, doi:10.1021/acs.analchem.2c00765 <https://doi.org/10.1021/acs.analchem.2c00765> .

9. Gau B, Dawdy AW, Wang HL, Bare B, Castaneda CH, Friese OV, Thompson MS, Lerch TF, Cirelli DJ, Rouse JC. Oligonucleotide Mapping via Mass Spectrometry to Enable Comprehensive Primary Structure Characterization of an mRNA Vaccine Against SARS CoV-2, *Sci Rep*, 2023, 13, No 9038, doi:10.1038/s41598-023-36193-2 <https://doi.org/10.1038/s41598-023-36193-2> .

10. Tang S, Liu GY, Yan Y, Wang S, Li N. Development of a Flow Through-Based Limited Digestion Approach for High-Throughput and High-Sequence Coverage Mapping of Therapeutic mRNAs, *Anal Chem*, 2024, 96, 16944–17003, doi:10.1021/acs.analchem.4c04384 <https://pubs.acs.org/doi/10.1021/acs.analchem.4c04384> .

11. Tunable Digestion of RNA Using RapiZyme RNases to Confirm Sequence and Map Modifications, 2024, Waters application note. September, 2024. 720008539.

12. CONFIRM Sequence: A waters_connect Application for Sequencing of Synthetic Oligonucleotide and Their Impurities, 2022, Waters application note. July, 2022. 720007677.

13. Rapid Analysis of a Synthetic mRNA Cap Structure Using Ion-Pairing RPLC with the BioAccord LC-MS System, 2021, Waters application note. August, 2021. 720007329.

Featured Products

ACQUITY Premier System <https://www.waters.com/nextgen/global/products/chromatography/chromatography-systems/acquity-premier-system.html>

Xevo G3 QTof <https://www.waters.com/nextgen/global/products/mass-spectrometry/mass-spectrometry-systems/xevo-g3-qtof.html>

waters_connect Informatics <https://www.waters.com/nextgen/global/products/informatics-and-software/waters_connect.html>

720008677, January2025

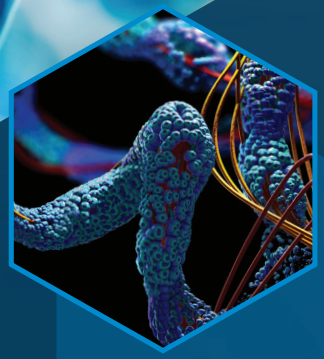
Accelerate RNA analysis with the waters_connect Informatics Platform

Monitor Critical Quality Attributes with streamlined tools for routine analysis

LEARN MORE
waters.com/waters_connectBiopharma

Waters™

5' Capping Efficiency & 3' Poly(A) Tail Heterogenity



The Importance of Well Characterized mRNA Therapeutics and Vaccines

Nick Pittman, Marketing Manager, Global Biopharmaceutical Business

Ensuring the quality of mRNA-based vaccines and therapeutics is paramount for safety, efficacy, and to obtain regulatory approval. The analysis of mRNA Critical Quality Attributes (CQAs) is central to a robust control strategy, with several key considerations

Risk Reduction and Regulatory Compliance: Early identification of CQAs in the product lifecycle can reduce later stage development surprises and delays. Regulatory agencies emphasize the need for robust CQA assessment to ensure product safety and efficacy. An early, comprehensive understanding of attributes impacting your product allows for a control strategy that is well aligned with regulatory expectations.

Building robustness and confidence into complex processes: mRNA production involves many intricate steps. Process changes can impact product integrity, clinical outcomes, and regulatory approval, so the impact of any changes must be fully understood. Close monitoring of product quality, and linking process inputs with quality outputs, supports development of robust, scalable and effective manufacturing processes.

Safety and efficacy: Unresolved quality issues can lead to reduced mRNA effectiveness, poor clinical trial outcomes and costly delays. Early characterization and routine monitoring of CQAs enhances product understanding to demonstrate the delivery of safe and efficacious therapeutics and vaccines.

Two key mRNA attributes have significant impact on the mRNA product. The mRNA 5′ cap regulates translation to allow efficient protein synthesis and prevents mRNA degradation, enhancing in vivo stability. It also helps to avoid immune responses that may degrade the mRNA and facilitates nuclear export to the cytoplasm, where translation occurs. With a crucial role in both function and stability, it is important to fully understand and monitor the capping efficiency of the 5′ cap.

Like the 5′ cap, the 3′ poly-A tail helps enable nuclear export and also enhances stability by preventing premature degradation. The poly-A tail interacts with the 5′ cap to stimulate translation. A transcript lacking a poly-A tail exhibits low translation levels and is susceptible to decapping. For these reasons, understanding and controlling 3′ poly-A tail heterogeneity and length are critical for effective product development.

To support successful development and production of mRNA based products, Waters have developed dedicated, enzymes, reagents, workflows and protocols, enabling routine understanding throughout development and manufacture to ensure safe and efficacious vaccines and therapeutics.

Waters™

Application Note

Rapid Analysis of Synthetic mRNA Cap Structure Using Ion-Pairing RPLC with the BioAccord LC-MS System

Jennifer M. Nguyen, Siu-Hong Chan, Bijoyita Roy, Martin Gilar, Brett Robb, Weibin Chen, Matthew A. Lauber

Waters Corporation, New England Biolabs Inc.

Abstract

Due to the crucial need to fight the SARS-CoV-2 pandemic, the development of mRNA vaccines progressed rapidly throughout 2020. This has led to the emergency use authorization of two highly efficacious mRNA vaccines.[1] These vaccines contain synthetic mRNA encapsulated in a lipid nanoparticle for delivery, and they use the recipient's body to conduct *in vivo* translation of a stabilized version of the viral spike protein and elicit the immune response.

The mRNA in these vaccines is produced by an enzymatic process known as *in vitro* transcription (IVT). Much like cellular mRNAs, synthetic mRNAs also need to be modified on their 5′ ends to include a 7-methyl guanosine nucleotide in a 5′-5′-linkage that is referred to as a 5′ cap. The vaccine mRNA is also modified to include a chain of adenosines on the 3′ end, called the poly(A) tail (typically ~120 nucleotides). These modifications are critical features of mRNA structure, playing important roles in recognition of the mRNA by cellular factors, stability of the synthetic mRNA, and translational efficiency of the synthetic mRNA molecule.[2] As such, determining the nature of the 5′ cap and the measurement of the poly(A) tail length by LC-MS analysis is vital to understand the product quality and ensure the safety and efficacy of these new vaccines.[3]

In this application note, we developed a rapid and sensitive LC-MS method applicable to synthetic mRNA capping analysis. Previously, a LC-MS method has been reported to analyze pre-defined 5′ fragments of synthetic mRNA.[4] Here, we demonstrate that ACQUITY Premier Columns and their MaxPeak High Performance Surfaces (HPS), previously shown to enhance oligonucleotide recovery, can also improve the performance of mRNA fragment analysis. This, in combination with an easy-to-use, compliance-ready BioAccord System, provides a fit-for-purpose platform for the rapid analysis of 5′ capping of IVT mRNA preparations. A less than 5-minute method was established, and linearity was demonstrated to detecting product-related impurities down to less than 0.1% of the target 5′ capped fragment.

Benefits

- Fast, quantitative analysis of synthetic mRNA 5′ analysis
- Minimal to no column conditioning required for ACQUITY Premier Columns with MaxPeak High Performance Surfaces
- Linearity, reproducibility, and robustness suitable in both development and QC testing
- High quality MS spectra produced with DIPEA (diisopropylethylamine) mobile phase

Introduction

The success of the Pfizer-BioNTech and Moderna SARS-CoV-2 vaccines has brought a surge of interest to mRNA molecules while also placing new demands on analytics to better support the development and manufacturing of a new modality. Among critical quality attributes (CQAs) such as percentage of full-length mRNA, untranslated regions (UTRs), 3′ poly(A) tail length, the sequence, structure, and chemical modifications of the mRNA,[5] the presence of a proper 5′ cap structure is important to ensuring maximal gene expression, to evading innate recognition mechanisms against foreign RNA, and to increasing resistance towards exonuclease degradation.[6]

In eukaryotes, the modified 5′ end of the mRNA is an N7-methyl-guanosine connected to the 5′ nucleotide by a 5′ to 5′ triphosphate group. Such a modified structure is known as the cap. In lower eukaryotes such as yeast, the first nucleotide of the capped mRNA does not have any 2′-O modifications, and the cap is known as Cap-0. In higher eukaryotes such as humans, the first nucleotide of the capped mRNA contains a methyl group at the 2′-O position. This mammalian cap structure is known as Cap-1.[6]

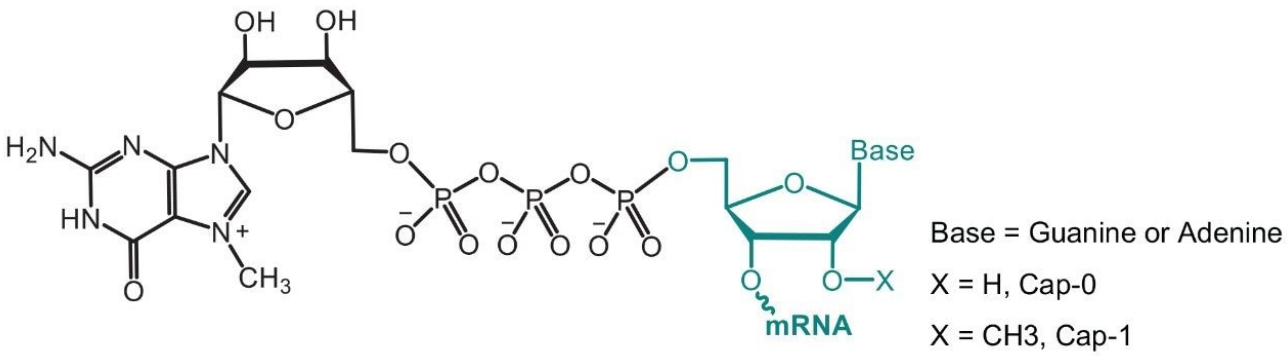


Figure 1. mRNA cap structure.

Since proper cap structure plays a critical role in mRNA translation and innate immune response, effective manufacturing of capped RNA is required. The Cap-1 structure can be added to the synthetic mRNA after IVT enzymatically using vaccinia mRNA capping enzyme (VCE) (Reactions 1.1, 1.2 and 1.3, Figure 2) in concert with vaccinia cap 2′-O-methyltransferase (Reaction 2, Figure 2), or during IVT by using an appropriate cap analog. Verifying the extent of RNA capping, however, is not trivial. RNA capping is the addition of a single nucleotide to an RNA that is often kilobases in length. For example, the BioNTech/Pfizer BNT162b2 is 4,284 nt-long. To date, there is no effective analytical method available to detect and quantify such a small change in physical property of an intact RNA molecule. Beverly and co-workers from the Novartis Institutes of Biomedical Research reported a liquid chromatography coupled to mass spectrometry (LC-MS) method that can be used to identify 5′ cap structures and quantify the capping efficiency of synthetic mRNA.[4] The method employs biotinylated probes to direct RNase H to cleave off a pre-defined fragment from the 5′ end of the synthetic mRNA. After enrichment via streptavidin beads, the 5′ cleavage fragments can be analyzed by LC-MS to examine the extent of 5′ capping.

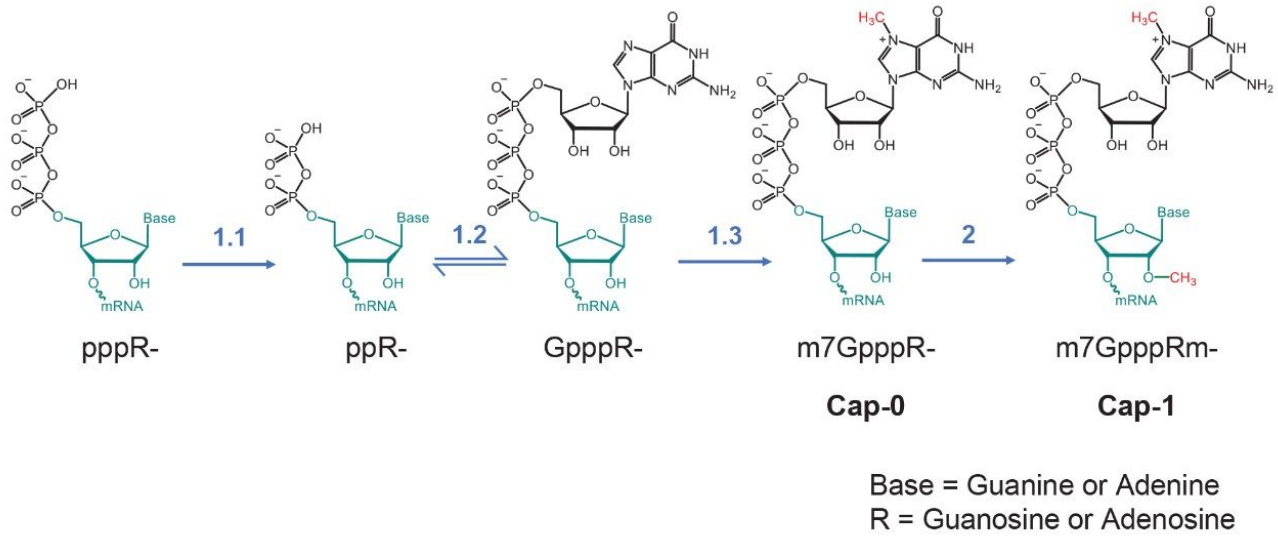


Figure 2. Enzymatic capping of RNA.

In this application note, we demonstrate the advantages of LC-MS to evaluate the 5′ capping efficiency of synthetic mRNA using an easy-to-use, compliant-ready BioAccord LC-MS System and ACQUITY Premier Oligonucleotide BEH C_{18} Columns, which feature MaxPeak High Performance Surfaces Technology. This hybrid organic/inorganic surface technology has been shown to minimize adsorptive losses by blocking interactions to metal surfaces.[7, 8] Here, the performance of the MaxPeak Premier Technology versus conventional columns greatly improves recovery of the mRNA 5′ fragments from the first injection. This, coupled with the BioAccord LC-MS System, can facilitate the analysis of 5′ cap structure of synthetic mRNA through MS quantitation.

Experimental

Sample Preparation

In this application note, we demonstrate the advantages of LC-MS to evaluate the 5′ capping efficiency of synthetic mRNA using an easy-to-use, compliant-ready BioAccord LC-MS System and ACQUITY Premier Oligonucleotide BEH C_{18} Columns, which feature MaxPeak High Performance Surfaces Technology. This hybrid organic/inorganic surface technology has been shown to minimize adsorptive losses by blocking interactions to

metal surfaces.[7, 8] Here, the performance of the MaxPeak Premier Technology versus conventional columns greatly improves recovery of the mRNA 5′ fragments from the first injection. This, coupled with the BioAccord LC-MS System, can facilitate the analysis of 5′ cap structure of synthetic mRNA through MS quantitation.

Figure 3. RNA oligonucleotides used in this study.

An equimolar solution of Cap-1 and the targeting oligo at 2.5 pmol/µL concentrations was prepared in water. A 1:10, 1:100, and 1:1000 dilution series of the four precursor oligonucleotides versus Cap-1 at a constant 2.5 pmol/µL concentration was prepared in water.

Data Treatment and Analysis:

LC Conditions

LC system:	BioAccord LC-MS
Detection:	UV detection at 260 nm

Vials:	Polypropylene vials (p/n: 186002639)
Column(s):	ACQUITY Premier Oligonucleotide BEH C_{18}, 130 Å, 1.7 µm, 2.1 x 50 mm (p/n: 186009484); ACQUITY UPLC Oligonucleotide BEH C18, 130 Å, 1.7 µm, 2.1 x 50 mm (p/n: 186003949)
Column temp.:	60 °C
Sample temp.:	4 °C
Injection volume:	5.0 µL (sample)
Flow rate:	0.3 mL/min
Mobile phase A:	1.0% hexafluoroisopropanol (HFIP), 0.1% N,N-diisopropylethylamine (DIPEA) in water
Mobile phase B:	0.75% HFIP, 0.0375% DIPEA in 65:35 acetonitrile:water
Gradient:	5–25% B in 5, 10, or 20 min

MS Conditions

MS system:	BioAccord LC-MS
Ionization mode:	Full scan, ESI negative
Acquisition range:	400–5000 *m/z*
Capillary voltage:	0.8 kV

Cone voltage:	40 V
Desolvation temp.:	400 °C

Data Management

Chromatography and MS software:	waters_connect with UNIFI v1.9.12

Results and Discussion

Initial Column Performance

Before developing a LC-MS method for 5′ capping efficiency, we evaluated the chromatographic recovery of the mRNA fragments from a LC-MS analysis. As described in literature and prior application notes, we have demonstrated that oligonucleotides as well as phosphorylated compounds can adsorb to metal surfaces such as stainless steel. This unwanted effect can be mitigated with the use of MaxPeak Premier Columns.[9-11] For RNA 5′ cap analysis, we investigated the LC-UV-MS separations of the Cap-1 species, consisting of the m7GpppGm group on the 5′ end, using either a conventional ACQUITY UPLC Oligonucleotide BEH C_{18} Column or an ACQUITY Premier Oligonucleotide BEH C_{18} Column packed with the same lot of stationary phase. It was confirmed that the ACQUITY Premier Column with its MaxPeak High Performance Surfaces confers benefits like improved recoveries and out-of-the-box performance without the need for conditioning or passivation.

For oligonucleotide LC-MS analysis, ion-pairing reversed-phase (IP-RP) separations with amine mobile phase additives is the preferred mode of chromatography given their resolving power and amenability to MS. However, amines with acidic counter ions, such as acetate or bicarbonate salts, produce too much ion suppression at the concentrations required for effective separations, though reducing the concentration of IP reagent can help. Since its introduction in 1997 by Apffel and co-authors, hexafluoroisopropanol (HFIP) has been preferred for IP-RP separations that are hyphenated to MS technology.[12] HFIP is a weak acid that can help buffer amine containing mobile phases without significantly impairing ionization efficiency. Additionally, it has been theorized that alkylamines are more likely to adsorb onto the stationary phase to act as a better IP system in the presence of HFIP and that it positively affects the desolvation of oligonucleotides in ESI-MS.[13] Here, we used the amine N,N-diisopropylethylamine (DIPEA) buffered with HFIP, as it has been shown to greatly increase MS signal intensity and peak shape versus triethylamine.[14]

Recovery of the Cap-1 fragment as based on peak area from the first injections were evaluated for the conventional ACQUITY UPLC Column versus the ACQUITY Premier Column. Representative UV chromatograms resulting from the first injection on the ACQUITY Premier Column and from the first to fifth injection of the conventional ACQUITY UPLC Column are shown in Figure 4. No peak representing the Cap-1 species was seen upon the initial use of the conventional column (Figure 4). In contrast, the ACQUITY Premier Column gave a high intensity peak on its first injection and reproducible peak areas from subsequent injections. Even after successive injections of Cap-1 on the conventional column, by injection 5, the peak area of the Cap-1 fragment was still only 37% of that of the ACQUITY Premier Column.

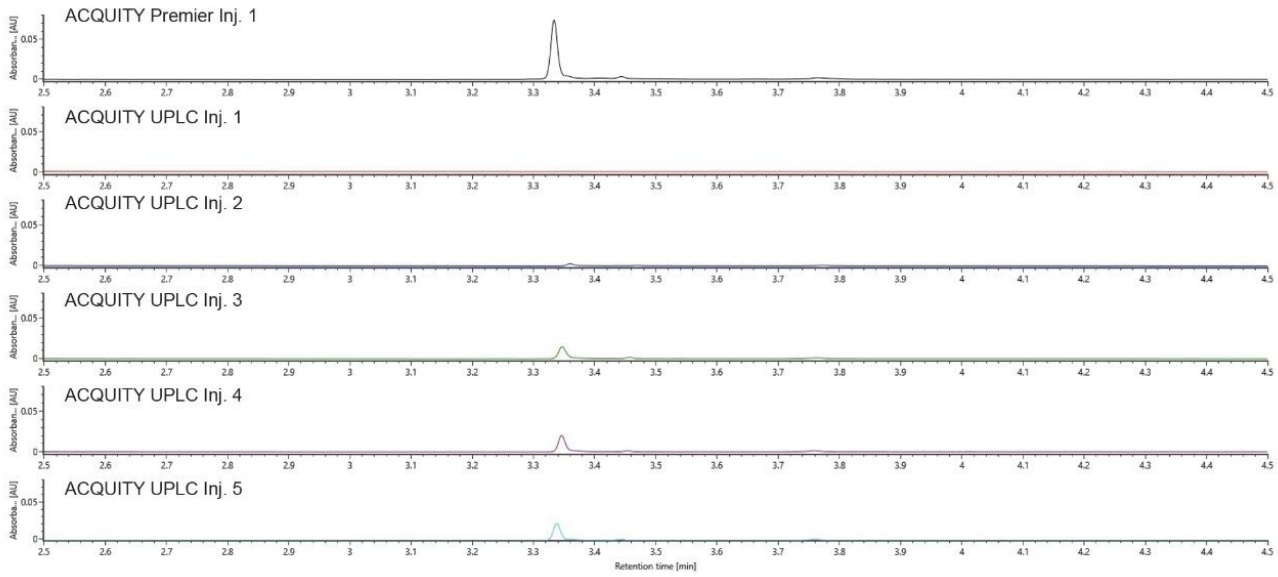


Figure 4. UV chromatograms of the first injection of Cap-1 fragment obtained using an ACQUITY Premier Oligonucleotide BEH C_{18}, 1.7 µm, 2.1 x 50 mm Column and from the first to fifth injections of an ACQUITY UPLC Oligonucleotide BEH C_{18}, 1.7 µm, 2.1 x 50 mm Column. Separations were performed with a BioAccord LC-MS System using a flow rate of 0.3 mL/min, column temperature of 60 °C, DIPEA-HFIP modified mobile phases, and 2.5 pmol mass loads.

Development of a Fast LC-MS Method

We sought to develop a LC-MS method for 5′ cap quantification and mass confirmation on the BioAccord System using the ACQUITY Premier Columns. As an easy-to-use LC-optical-MS platform, the BioAccord System is designed for comprehensive analysis of biotherapeutics with reduced complexity in MS instrument operation.[15-17] With the BioAccord LC-MS System, robust and efficient LC-optical-MS methods can be developed for oligonucleotide analysis, reducing the cost and time for gene-based therapy development without compromising product quality. While quantitation of the 5′ cap fragments via UV is often favored for some analysts without access to MS instrumentation, we chose to optimize the method of quantitation based on MS signals using the BioAccord, because of the benefits to shorten run times while also taking advantage of accurate MS detection.

Using this strategy, we developed a fast LC-MS method using a short 50 mm ACQUITY Premier Column. Our aim was to rapidly elute the species and use the high-quality mass spectra for mass confirmation and relative quantitation. In Figure 5, it is demonstrated that a method of less than 5 minutes can be employed with little or no loss in recovery of the Cap-1 species. Moreover, Figure 5A suggests even a shorter run time of less than 5 minutes could be achieved; however, the lower abundant species would have become less resolved from the main peak when a run time is shorter, which could interfere with the interpretation of the mass spectra.

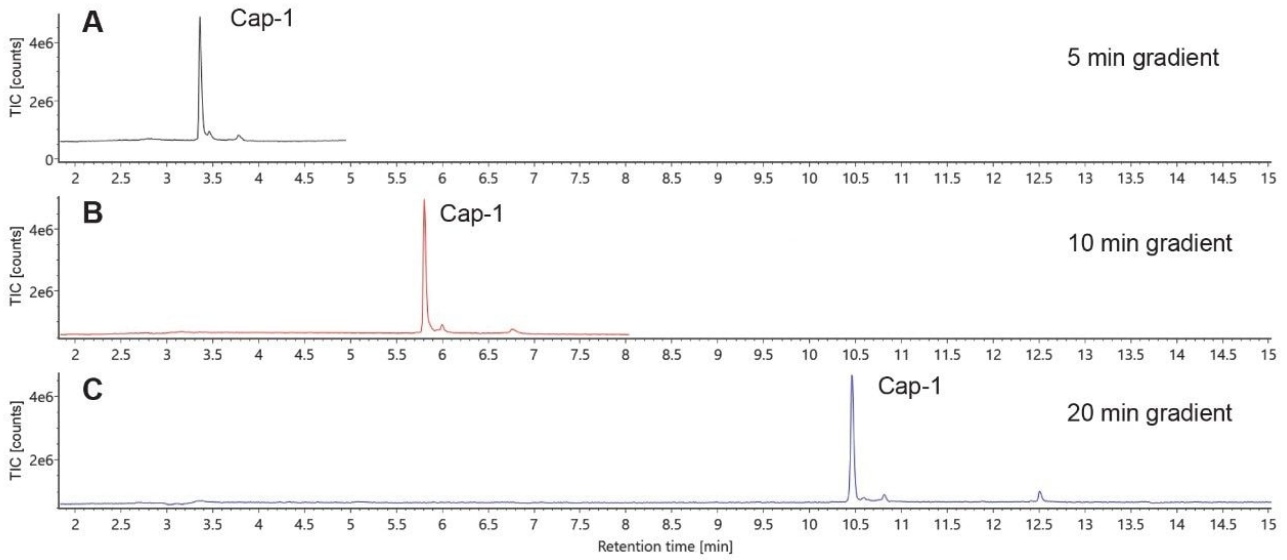


Figure 5. TIC chromatograms of Cap-1 fragment obtained using an ACQUITY Premier Oligonucleotide BEH C_{18}, 1.7 µm, 2.1 x 50 mm Column. Separations were performed with a BioAccord LC-MS System using a flow rate of 0.3 mL/min, column temperature of 60 °C, DIPEA-HFIP modified mobile phases, 2.5 pmol mass loads, and a gradient from 5–25% B in (A) 5, (B) 10, or (C) 20 min.

Relative Quantitation of 5′ Cap Fragment Variants for the Evaluation of 5′ Capping Efficiency

As previously mentioned, the percentage of Cap-1 mRNA molecules generated from IVT is an important CQA and must be determined accurately to understand and predict the efficacy of the synthetic mRNA preparation. To demonstrate the capability of the BioAccord LC-MS System for this measurement, we created a series of solutions, in which the ratios between the four precursor species to Cap-1 is 1:10, 1:100, and 1:1000, respectively, while the concentration of Cap-1 remained constant in all solutions. The charge state profiles as obtained from the solution series are displayed in Figure 6.

In Figure 6, the most abundant charge state for Cap-1 was the $[M\text{-}10H]^{10-}$ charge state. When running samples containing each individual mRNA fragment (data not shown), the 10 times deprotonated ion was also the most abundant for each fragment, and these charge states can be seen visually in the 1:10 dilution sample. Therefore, we chose this charge state for data analysis and quantification using extracted ion chromatograms (XICs).

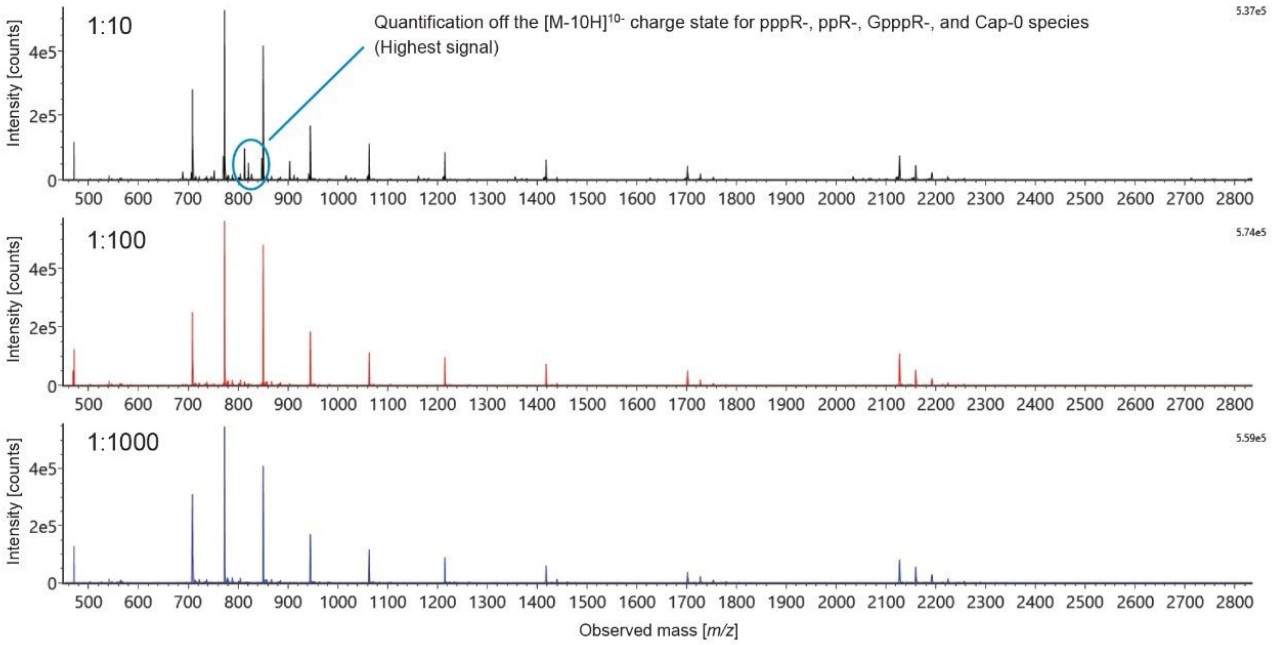


Figure 6. Mass spectra of Cap-1 fragment and its product related impurity fragments (pppG, ppG, GpppG, and m7GpppG, or Cap-0) obtained using an ACQUITY Premier Oligonucleotide BEH C_{18}, 1.7 µm, 2.1 x 50 mm Column. Separations were performed with a BioAccord LC-MS System using a flow rate of 0.3 mL/min, column temperature of 60 °C, DIPEA-HFIP modified mobile phases, and a dilution series of 1:10, 1:100, and 1:1000 versus a 12.5 pmol/µL concentration of Cap-1.

Representative XICs for the $[M-10H]^{10-}$ charge state of Cap-0 are shown in Figure 7 for each dilution solution. For each dilution, the isotopic distribution for the charge state can be clearly observed, and XICs were taken from the entire isotope mass at a mass error of 10 ppm. Even at a 1:1000 dilution, corresponding to a mass load on column of only 12.5 fmol of the product related impurities, the XIC of the Cap-0 species could be readily generated and integrated.

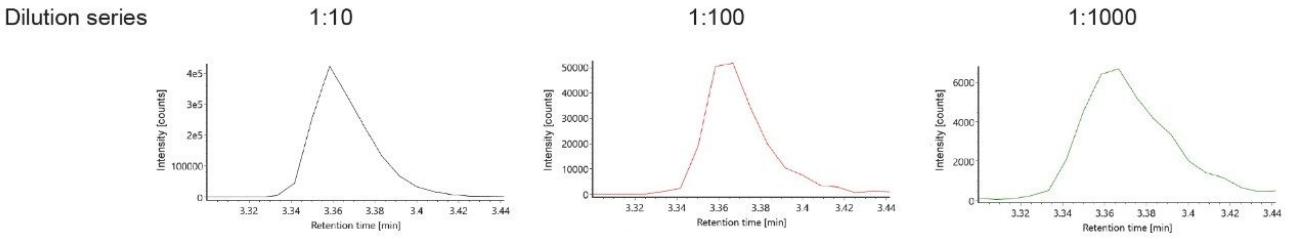


Figure 7. XICs as obtained from the [M-10H]$^{10-}$ charge state of the Cap-0 fragment using an ACQUITY Premier Oligonucleotide BEH C_{18}, 1.7 µm, 2.1 x 50 mm Column. Separations were performed with a BioAccord LC-MS System using a flow rate of 0.3 mL/min, column temperature of 60 °C, DIPEA-HFIP modified mobile phases, and a dilution series of 1:10, 1:100, and 1:1000 versus a 12.5 pmol/µL concentration of Cap-1.

From the XICs of the dilution series, calibration curves could be generated and are displayed in Figure 8 for each of the precursor fragments. Here, the linear regression for each chart produces an R squared value of 1.000, indicating the capability of this rapid LC-MS method for evaluating the extent of 5′ capping.

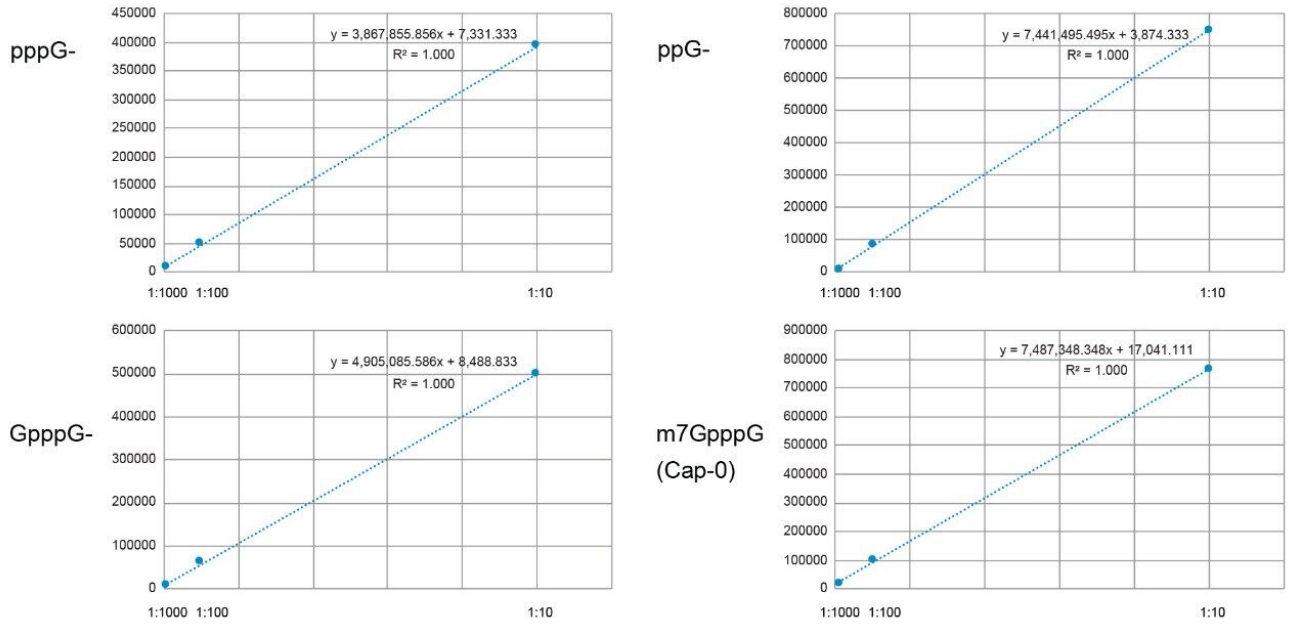


Figure 8. Calibration curves generated from a 1:10, 1:100, and 1:1000 dilution series of Cap-1 at 12.5 pmol/µL versus its product-related impurity fragments (pppG, ppG, GpppG, and m7GpppG, or Cap-0).

Separation of the RNase H Probe (Targeting Oligo) and RNA 5′ Fragments

An important consideration when analyzing RNase H-based mRNA fragment analysis, is the LC separation of the RNase H probe, or targeting oligo, from the analyte species. In this example, the RNase H probe is a 3′ desthiobiotin affinity tag-modified RNA:DNA chimera approximately the same length as the cleaved 5′ fragment. The probe is designed to be a complement of the target 5′ fragment such that the two hybridize in solution to form a double-stranded RNA-DNA duplex structure. After RNase H cleavage at a pre-defined site, the probe and the 5′ fragment remain associated as a duplex in solution and thus, it is necessary to resolve the species of interest from the RNase H oligo probe that is expected to be present in the sample mix.

To validate the LC separation of the RNase H probe and the analyte fragments, an RNase H probe that could be used to generate the 25-nt RNA oligonucleotide from a longer RNA was designed and synthesized. Figure 9 shows that the rapid LC-MS method can separate the Cap-1 RNA oligonucleotide and the RNase H probe in an equimolar solution of the two species.

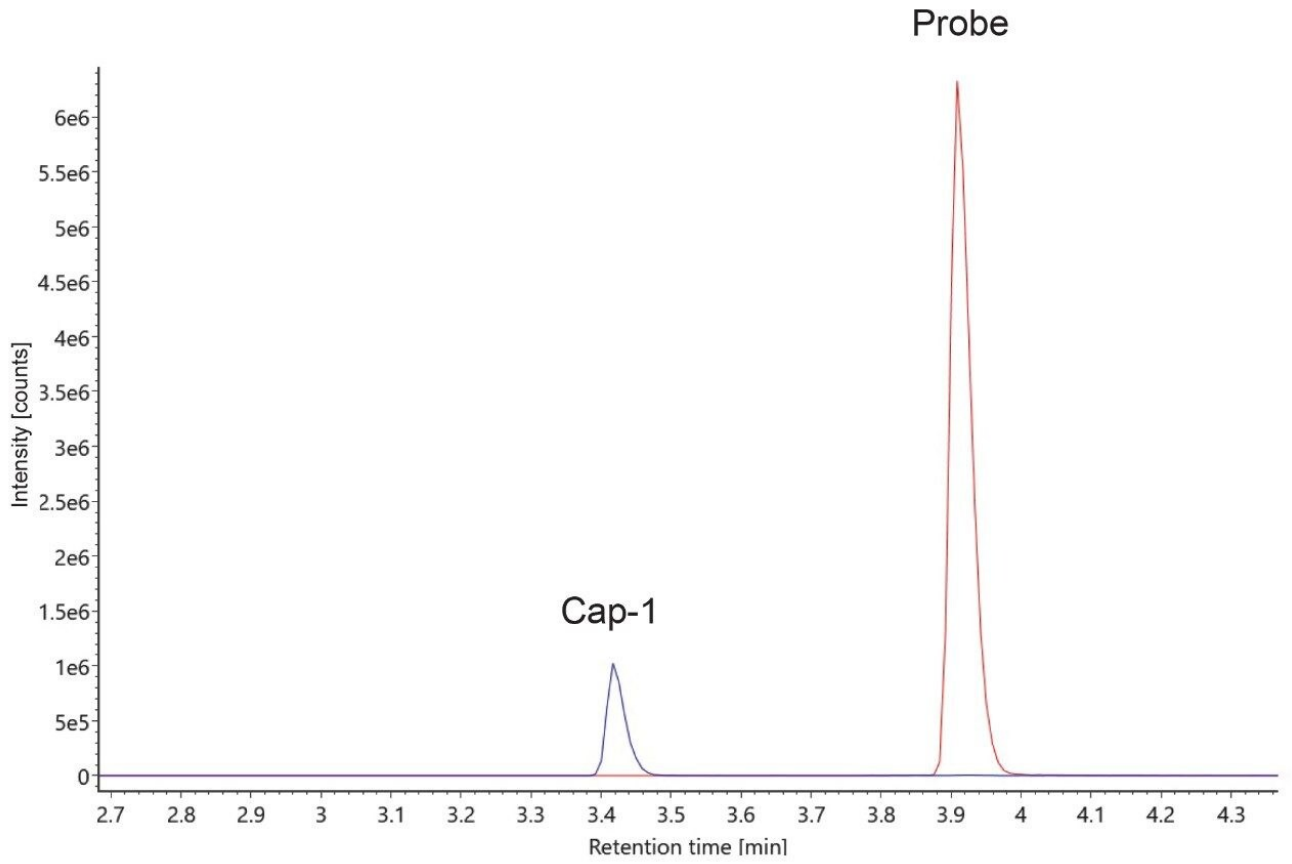


Figure 9. Overlaid XICs of Cap-1 and a representative probe from an equimolar solution using an ACQUITY Premier Oligonucleotide BEH C_{18}, 1.7 µm, 2.1 x 50 mm Column. Separations were performed with a BioAccord LC-MS System using a flow rate of 0.3 mL/min, column temperature of 60 °C, DIPEA-HFIP modified mobile phases, and 12.5 pmol mass loads.

Conclusion

The value of mRNA as a vaccine modality is now proven by the high effectiveness of the mRNA-based vaccines against COVID-19 that have now been administered to over 25% of the world population. Analytical techniques to ensure their proper design, development, and reproducible manufacture are important. In this application note, we demonstrate that a rapid LC-MS method applicable to evaluating the extent of 5′ capping of synthetic mRNA,

an important CQA for synthetic mRNAs, using an ACQUITY Premier Oligonucleotide BEH C_{18} Column couple with a BioAccord LC-MS System. The ACQUITY Premier Column with MaxPeak High Performance Surfaces Technology could provide considerable improvements in RNA recovery upon first injection. Paired with the BioAccord LC-MS, this combination of technologies allows for accurate quantitation, even at low limits of detection, such that it is possible to validate the manufacturing of Cap-1 mRNA molecules, and potential presence of product precursor related impurities. Moreover, these results highlight the potential of MS-based quantitation for high throughput assays, which could help accelerate the development of mRNA modalities.

References

1. Padda, I. S.; Parmar, M. COVID (SARS-COV-2) Vaccine. StatPearls. Treasure Island (FL), 2021.
2. Jackson, N. A. C.; Kester, K. E.; Casimiro, D.; Gurunathan, S.; DeRosa, F. The Promise of mRNA Vaccines: A Biotech and Industrial Perspective. *NPJ Vaccines* 2020, 5:11. doi:10.1038/s41541-020-0159-8.
3. Muttach, F.; Muthmann, N.; Rentmeister, A. Synthetic mRNA Capping. *Beilstein J. Org. Chem.* 2017, 13, 2819–32. doi:10.3762/bjoc.13.274.
4. 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, 5021–30. doi:10.1007/s00216-016-9605-x.
5. Alelyunas, Y.; Shion, H.; Wrona, M. High Sensitivity Intact Monoclonal Antibody (mAb) HRMS Quantification. Waters Application Note, 720006222 <https://www.waters.com/nextgen/us/en/library/application-notes/2018/high-sensitivity-intact-monoclonal-antibody-hrms-quantification.html> , 2018.
6. Ramanathan, A.; Robb, G. B.; Chan, S. H. mRNA Capping: Biological Functions and Applications. *Nucleic Acids Research* 2016, 44, 7511–26. doi:10.1093/nar/gkw551.
7. Lauber. M.; Walter, T. H.; Gilar, M.; DeLano, M.; Boissel, C. A.; Smith, K.; *et al*. Low Adsorption HPLC Columns Based on MaxPeak High Performance Surfaces. Waters White Paper, 720006930EN < https://www.waters.com/waters/library.htm?cid=511436&lid=135074404&lcid=135074403> , 2020.
8. Smith, K. M.; Wilson, I. D.; Rainville, P. D. Sensitive and Reproducible Mass Spectrometry-Compatible RP-UHPLC Analysis of Tricarboxylic Acid Cycle and Related Metabolites in Biological Fluids: Application to Human Urine. *Anal. Chem.* 2020. doi:10.1021/acs.analchem.0c03863.

9. DeLoffi, M.; Nguyen, J. M.; Izzo, G. S.; Lauber, M.; Savaria, M. Improved Chromatographic Performance with a Premier Peptide C_{18} Column Versus a Titanium-Lined C_{18} Column Technology. Waters Application Brief, 720007022 <https://www.waters.com/nextgen/us/en/library/application-notes/2020/improved-chromatographic-performance-with-a-premier-peptide-c18-column-versus-a-titanium-lined-c18-column-technology.html> , 2020.

10. DeLano, M.; Walter, T. H.; Lauber, M. A.; Gilar, M.; Jung, M. C.; Nguyen, J. M.; *et al*. Using Hybrid Organic-Inorganic Surface Technology to Mitigate Analyte Interactions with Metal Surfaces in UHPLC. *Anal. Chem. 2021*. doi:10.1021/acs.analchem.0c05203.

11. Gilar, M.; DeLano, M.; Gritti, F. Mitigation of Analyte Loss on Metal Surfaces in Liquid Chromatography. *J. Chromatogr., A* 2021; 1650, 462247. doi:10.1016/j.chroma.2021.462247.

12. Apffel, A.; Chakel, J. A.; Fischer, S.; Lichtenwalter, K.; Hancock, W. S. Analysis of Oligonucleotides by HPLC-Electrospray Ionization Mass Spectrometry. *Anal. Chem.* 1997, 69, 1320–5. doi:10.1021/ac960916h.

13. Gilar, M.; Fountain, K. J.; Budman, Y.; Holyoke, J. L.; Davoudi, H.; Gebler, J. C. Characterization of Therapeutic Oligonucleotides Using Liquid Chromatography with On-Line Mass Spectrometry Detection. *Oligonucleotides* 2003, 13, 229–43. doi:10.1089/154545703322460612.

14. Gong. L.; McCullagh, J. S. Comparing Ion-Pairing Reagents and Sample Dissolution Solvents for Ion-Pairing Reversed-Phase Liquid Chromatography/Electrospray Ionization Mass Spectrometry Analysis of Oligonucleotides. *Rapid Commun. Mass Spectrom*. 2014, 28, 339–50. doi:10.1002/rcm.6773.

15. Li, W.; Lauber, M. Comprehending COVID-19: Preliminary Examination of the SARS-CoV-2 Spike Protein by Peptide Mapping. Waters Application Note, 720006909 < https://www.waters.com/nextgen/us/en/library/application-notes/2020/comprehending-covid-19-preliminary-examination-of-the-sars-cov-2-spike-protein-by-peptide-mapping.html> , 2020.

16. Doneanu, C.; Fox, J.; Harry, E.; Knowles, C.; Yu, Y. Q.; Fredette, J.; *et al*. Intact Mass Confirmation Analysis on the BioAccord LC-MS System for a Variety of Extensively Modified Oligonucleotides. Waters Application Note, 720007028 <https://www.waters.com/nextgen/us/en/library/application-notes/2020/intact-mass-confirmation-analysis-on-the-bioaccord-lc-ms-system-for-a-variety-of-extensively-modified-oligonucleotides.html> , 2020.

17. Shion, H.; Berger, S. J.; Yu, Y. Q. Application of a Mass Confirmation Workflow for Biotherapeutics Screening. Waters Application Note, 720007027 <https://www.waters.com/nextgen/us/en/library/application-notes/2020/application-of-a-mass-confirmation-workflow-for-biotherapeutics-screening.html> , 2020.

Acknowledgments

Jennifer Nguyen, Weibin Chen, Matthew Lauber, Martin Gilar (Waters Corporation, Milford, MA, USA).

Siu-Hong Chan, Bijoyita Roy, Brett Robb (New England Biolabs Inc., Ipswich, MA, USA).

Featured Products

ACQUITY UPLC I-Class PLUS System <https://www.waters.com/134613317>

ACQUITY UPLC Tunable UV Detector <https://www.waters.com/514228>

BioAccord LC-MS System for Biopharmaceuticals <https://www.waters.com/waters/nav.htm?cid=135005818>

UNIFI Scientific Information System <https://www.waters.com/134801648>

ACQUITY Premier Columns <https://www.waters.com/waters/nav.htm?&cid=513206>

720007329, August 2021

Waters™

Application Note

Ion-Pairing Reversed Phase LC-MS Analysis of Poly(A) Tail Heterogeneity Using the BioAccord LC-MS System

Catalin E. Doneanu, Joseph Fredette, Ying Qing Yu

Waters Corporation

Abstract

This application note demonstrates the advantages of the BioAccord LC-MS System for the analysis of Poly(A) Tail mRNA heterogeneity.

Benefits

- An LC-MS assay developed on the BioAccord LC-MS System running under waters_connect informatics platform enables robust analysis of mRNA Poly(A) Tail length and heterogeneity, including average molecular weight information for the Poly(A) Tail oligonucleotide region common to therapeutic and prophylactic mRNAs

Introduction

Messenger RNAs (mRNAs) have recently emerged as a new class of biopharmaceuticals, as demonstrated by the successful development and approval of two highly efficacious vaccines based on the mRNA sequence encoding for the SARS-CoV-2 spike protein.[1–2] The development of analytical methods for monitoring the critical

attributes of mRNA molecules is essential for proper control of the manufacturing process, but the development of such methods for the analysis of large mRNAs (>1,000 nucleotides), including mRNA vaccines, is challenging. One of the functional structures common to all therapeutic mRNAs is the Poly(A) Tail, a long (50–250 nucleotides) sequence at the 3′-end of the molecule that contains only adenosine residues.[3–6] The 3′-end oligonucleotide is typically isolated from the mRNA molecule for analysis using the RNase T1 ribonuclease enzyme, which cleaves after guanosine residues.[7–8]

Here we investigated the capabilities of the BioAccord LC-MS System for intact mass analysis of a Poly(A) Tail oligonucleotide fragment.

Experimental

Reagents and Sample Preparation

N,N-diisopropylethylamine (DIPEA, 99.5% purity, catalogue number 387649-100ML), 1,1,1,3,3,3-hexafluoro-2-propanol (HFIP, 99% purity, catalogue number 105228-100G), ethanol (HPLC grade, catalogue number 459828-2L), and ammonium bicarbonate (LiChropur LC-MS Supelco reagent, catalogue number 5330050050) were purchased from Millipore Sigma (St Louis, MO). Acetonitrile (LC-MS grade, catalogue number 34881-1L) and methanol (LC-MS grade, catalogue number 34966-1L) were obtained from Honeywell (Charlotte, NC). HPLC grade Type I deionized (DI) water was purified using a MilliQ system (Millipore, Bedford, MA). Mobile phases were prepared fresh daily.

Ultrapure nuclease-free water (catalogue no J71786.AE) for mRNA digestions was purchased from Thermo Fisher Scientific (Waltham, MA).

Twenty nanomoles of a 120-mer polyadenosine mRNA oligonucleotide (Poly A Tail) were purchased from Integrated DNA Technologies (Coralville, IA). The oligonucleotide sequence contained a cytidine incorporated at position 28, a uridine incorporated at position 56, a guanosine incorporated at position 84 and another uridine at position 112 as indicated by the sequence 5′- AAA AAA AAA AAA AAA AAA AAA AAA AAA CAA AAA AAA AAA AAA AAA AAA AAA AUA AAA AAA AAA AAA AAA AAA AAA AAA AAG AAA AAA AAA AAA AAA AAA AAA AAA AAA UAA AAA AAA -3′. These four residues were inserted in the Poly A Tail sequence in order to make the 120-mer RNA nucleotide more stable during chemical synthesis. The full length 120-mer oligo product (FLP) was

not further purified by HPLC in order to prevent its potential degradation during a large scale chromatographic purification process. The elemental composition of the synthetic Poly A Tail oligonucleotide is C1197 H1439 N592 O724 P119 and its average molecular weight is 39,388.64 Da.

Stock solutions of the synthetic Poly A Tail oligonucleotide were prepared in DI water at a concentration of 10 µM, from which a 5 µL volume was injected, which corresponds to loading 50 picomoles of the 120-mer oligonucleotide on-column.

Firefly luciferase mRNA (Fluc-beta mRNA, catalogue no M1436/1000-C1-A120-NM-PO, with a concentration of 1.63 mg/mL) was obtained from Amp Tec (Hamburg, Germany). The mRNA was digested with RNase T1 ribonuclease from *Aspergillus oryzae* (purchased from Millipore Sigma: catalogue no R1003-100KU). For mRNA digestion, 20 µL of Fluc mRNA were mixed with 5 µL digestion buffer (100 mM ammonium bicarbonate), 13 µL of nuclease-free water and 2 µL of a 1:100 diluted solution of the RNase T1 ammonium sulfate solution (prepared also in nuclease-free water). The digestion mixture was prepared in a QuanRecovery MaxPeak 300 µL vial and incubated at 37 ^{o}C for 15 min, then immediately analyzed by LC-MS.

All datasets were acquired in UNIFI App ver. 2.1.2.4 and processed using the INTACT Mass App ver 1.6.0.18 within the waters_connect informatics platform.

The MaxEnt 1 charge deconvolution algorithm was used for processing of oligonucleotide ESI-MS spectra to produce accurate intact mass measurements.

LC Conditions

LC-MS system:	BioAccord LC-MS System with ACQUITY Premier UPLC (Binary)
Pre-Column:	VanGuard FIT cartridge holder (p/n: 186007949) containing a 2.1 x 5 mm ACQUITY Premier FIT cartridge packed with 1.7 µm BEH C_{18} particles (p/n: 186009459)
Column:	ACQUITY Premier OST Column 1.7 µm, 130 Å, 2.1 x 50 mm, (p/n: 186009484)
Column temperature:	60 ^{o}C

Flow rate:	300 µL/min
Mobile phases:	Solvent A: 8 mM DIPEA (N,N-diisopropylethylamine), 40 mM HFIP (1,1,1,3,3,3-hexafluoroisopropanol), in DI water, pH 8.8 Solvent B: 4 mM DIPEA, 4 mM HFIP in 75% ethanol
Sample temperature:	6 °C
Sample vials:	QuanRecovery MaxPeak HPS vials (p/n: 186009186)
Injection volume:	5 µL

Wash Solvents

Purge solvent:	50% MeOH
Sample manager wash solvent:	50% MeOH
Seal wash:	20% acetonitrile in DI water

Gradient Table

Time (min)	Flow (mL/min)	Solvent A composition (%)	Solvent B composition (%)	Curve profile
0.00	0.3	100	0	Initial
10.00	0.3	85	15	–
10.50	0.3	15	85	6
11.00	0.3	15	85	6
11.10	0.3	100	0	6
15.00	0.3	100	0	6

MS Conditions

Acquisition mode:	Full scan
Ionization mode:	ESI(-)
Capillary voltage:	0.8 kV
Cone voltage:	45 V
Source temperature:	120 ^{o}C
Desolvation temperature:	500 ^{o}C
Desolvation gas (N_2) pressure:	6.5 bar
TOF mass range:	400–5000
Acquisition rate:	2 Hz
Lock-mass:	waters_connect lockmass solution (p/n: 186009298)

Informatics platform for data acquisition and processing:	waters_connect
Data acquisition:	UNIFI App ver 2.1.2.4
Data processing:	INTACT Mass App ver 1.6.0.18

Results and Discussion

Since its discovery in 1961, mRNA has been proposed as therapeutic modality for treatment of a variety of diseases.[2] Several structural elements are common to all mRNA molecules designed for therapeutic and prophylactic use, including the Cap sequence located at the 5′-end of the molecule, two untranslated regions, the open reading frame sequence (ORF) - encoding for the gene of interest and the Poly(A) tail located at the 3′-end of the molecule. These structural elements are depicted in a cartoon shown as Figure 1. Similar in structure to a naturally occurring cellular mRNA, a synthetic mRNA is engineered to contain a modified Cap sequence (~50 nucleotides)[9] and a long chain of adenosines (50–250 nt) for the Poly(A) tail.[3–6] These two modified regions are critical quality attributes (CQAs) of therapeutics mRNAs, playing an important role in mRNA recognition by the cellular machinery responsible for in vitro transcription (IVT). In addition, the 5′-Cap and the 3′-Poly(A) tail improve the efficiency of the IVT process.[1] Finally, because the Cap and the Poly(A) regions are located at the termini of the mRNA molecule, they play critical roles for stability of synthetic mRNAs.

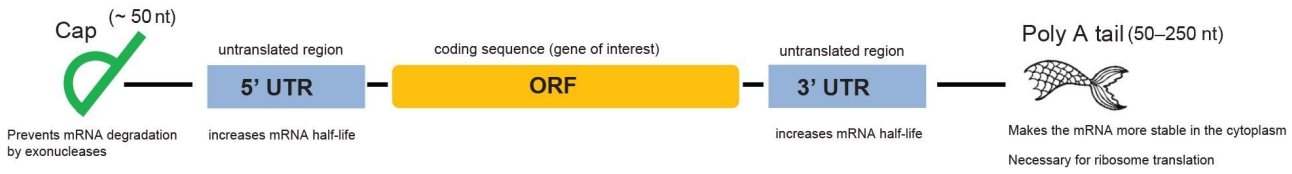


Figure 1. Schematic diagram showing the structural elements of an mRNA molecule. The Poly(A) Tail oligonucleotide, located at the 3′-termini, is a typical cleavage product generated by treatment with RNAse T1 ribonuclease.

LC-MS Method Optimization for a Synthetic Poly(A) Tail Oligonucleotide

Several LC-MS assays have been previously reported for the analysis of intact 5′-Cap oligonucleotides as well as intact Poly(A) tail variants.[10–13,14,15] Compared to the analysis of 5′-Cap oligonucleotides, the LC-MS characterization of Poly(A) tails is more challenging for several reasons. The Poly(A) tail contains a longer nucleotide sequence, therefore in negative ESI-MS studies it is more difficult to ionize relative to the much shorter Cap sequence. Also, the Poly(A) tail exhibits higher heterogeneity and reduced stability after preparation, being highly prone to enzymatic degradation. In addition, the biological processes responsible for synthesis of the Poly(A) tails in different species are not fully understood, therefore it is difficult to predict exactly all the different sequence variants that might be produced following the IVT process.[3–6] For these reasons, for the initial experiments, we decided to use a chemically synthesized analogue of a Poly(A) tail because the oligo sequence could be controlled better and the 120-mer oligonucleotide can be engineered to be more stable after solubilization in DI water, allowing for long-term LC-MS studies (up to four weeks after sample solubilization). The synthetic Poly(A) tail was first analyzed using an ion-pairing mobile phase suitable for the separation of shorter oligonucleotides, such as resolving siRNA oligos (typically 20 to 25 nucleotides) and their impurities.[16–18] The UV chromatogram showing the separation of the crude (unpurified) synthetic 120-mer Poly(A) oligomer is displayed in Figure 2. The TEA (triethylamine) ion pairing reagent is clearly not suitable for separation of such larger oligos as demonstrated by the broad (~0.3 min wide) chromatographic peak shown in this figure.

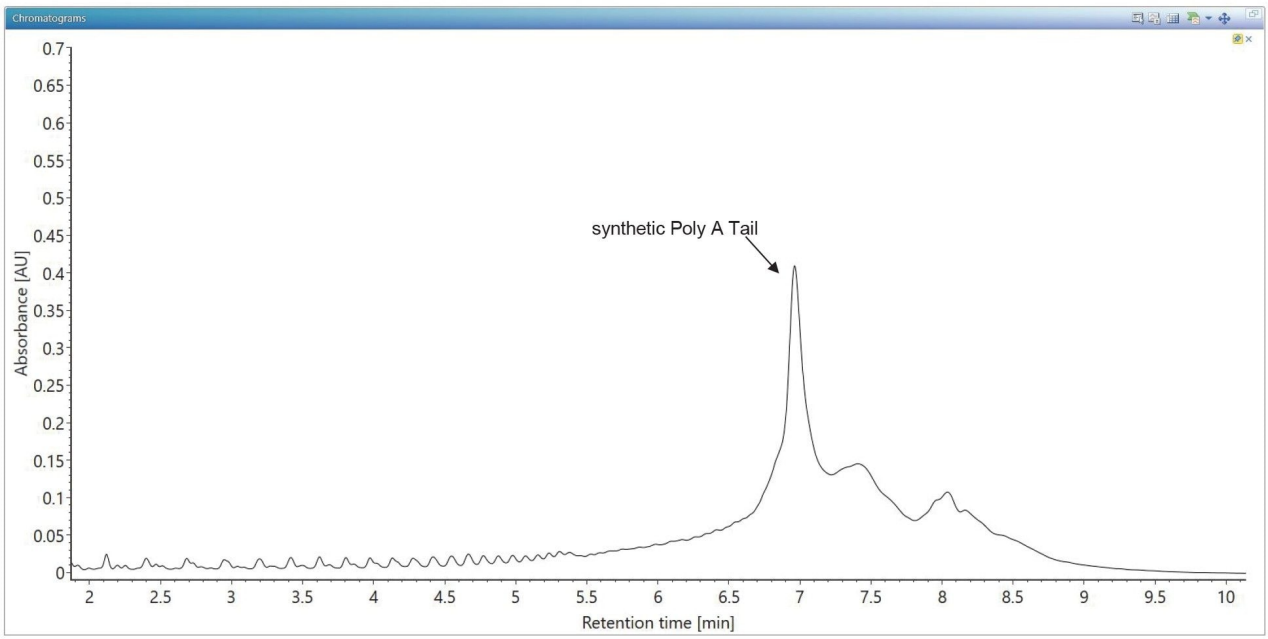


Figure 2. TUV chromatogram showing the IP-RP separation of the 120-mer synthetic Poly(A) Tail oligonucleotide using TEA (triethylamine) as the ion pairing reagent. Solvent A contained 7 mM TEA and 40 mM HFIP in deionized water, while Solvent B contained 3.5 mM TEA, 20 mM HFIP in 50% methanol. TEA is not suitable for separation of large oligos (>40-mers) as it produces rather broad chromatographic peaks (0.3 min peak width) for these species.

A branched ion pairing reagent – DIPEA (N,N-diisopropylethylamine) is recommended for separation of large oligos, including single guide RNA oligos (sgRNAs) that are typically 100-mers.[7,19] As shown by the overlaid UV chromatograms shown in Figure 3, the chromatographic shape improves considerably (peak width ~0.1 min) for the Poly(A) Tail in the presence of DIPEA. Three different organic solvents were investigated as mobile phase B eluents, including methanol, acetonitrile and ethanol, in order to optimize the intensity of the ESI-MS spectrum for the 120-mer oligo. Methanol is a commonly used solvent for TEA separation of short oligos, but in the case of large oligos, the other two organic solvents (ACN and EtOH) provide slightly sharper Poly(A) Tail peaks, as indicated by chromatograms in Figure 3.[16–18]

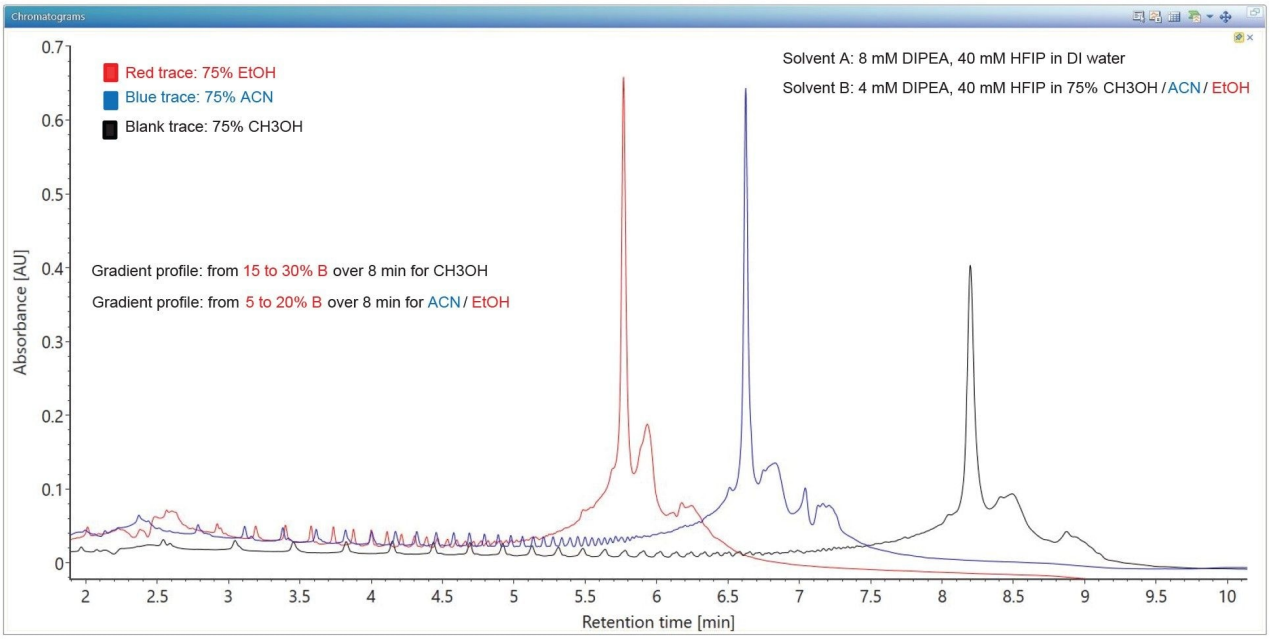


Figure 3. Overlaid TUV chromatogram showing the IP-RP separation of the 120-mer synthetic Poly(A) Tail oligonucleotide using DIPEA (N,N-diisopropylethylamine) as the ion pairing reagent in the presence of three different organic solvents. Solvent A contained 8 mM DIPEA and 40 mM HFIP in deionized water, while Solvent B contained 4 mM DIPEA, 4 mM HFIP in 75% organic solvent, including: 1) methanol (black trace); 2) acetonitrile (blue trace) and 3) ethanol (red trace). DIPEA perform much better compared to TEA, producing sharper chromatographic peaks (~0.1 min peak width) for large oligos like the Poly(A) tail 120-mer oligonucleotide.

While methanol provided ~3 fold increase in ESI-MS signal intensity in the presence of DIPEA vs TEA (data not shown), the acetonitrile and ethanol further improved Poly(A) Tail oligonucleotide MS response, as demonstrated by the combined ESI-MS spectra displayed in Figure 4. Acetonitrile is typically the preferred organic solvent for IP/RP LC-MS of large oligos, but ethanol was recently proposed as an alternative for Poly(A) Tail analysis.[7,15,19] The combined ESI-MS spectra presented in Figure 4 confirms that ethanol can provide an improved ESI-MS signal over acetonitrile.

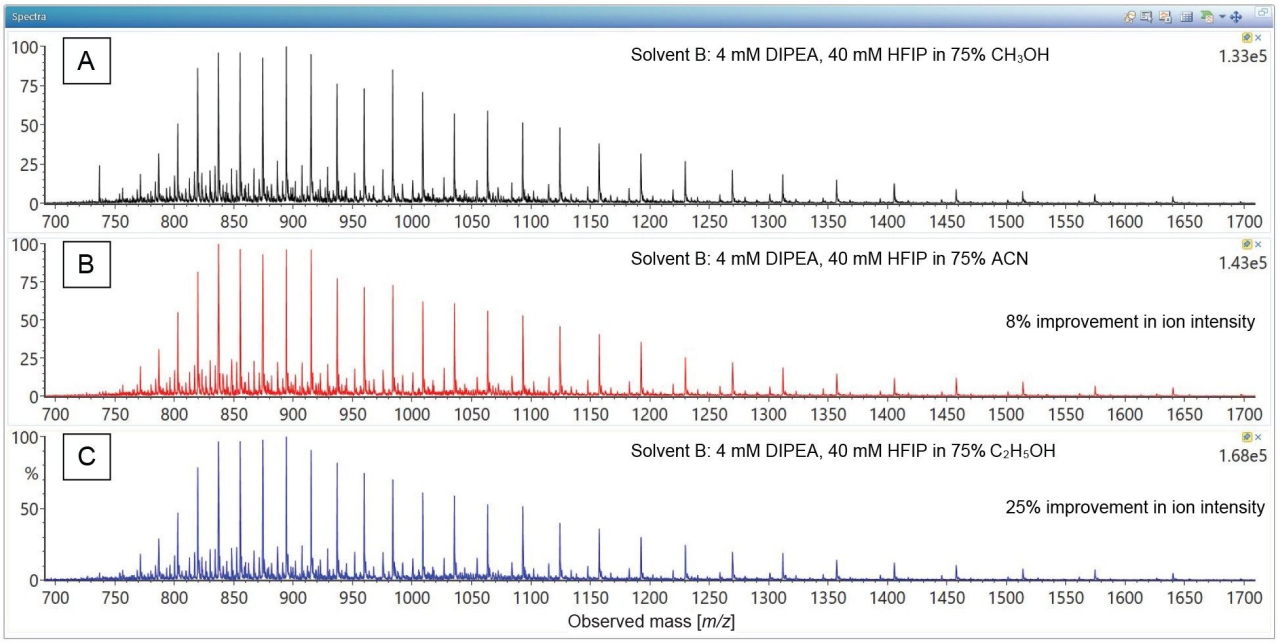


Figure 4. Combined ESI-MS spectra recorded for the Poly(A) Tail oligonucleotide following the IP-RP separations displayed in Figure 4. Solvent A contained 8 mM DIPEA and 40 mM HFIP in deionized water, while Solvent B contained 4 mM DIPEA, 4 mM HFIP in 75% organic solvent, including: methanol (panel A), acetonitrile (panel B), and ethanol (panel C). Ethanol was the preferred solvent since it produced the strongest ESI-MS signal.

The intact mass ESI-MS spectrum recorded using 8 mM DIPEA 40 mM HFIP in 75% ethanol as eluent B is displayed (Figure 5B) and the deconvolved spectrum is shown in the panel above (Figure 5A). The ESI-source parameters (ESI voltage, cone voltage, source temperature, desolvation temperature) were optimized to increase the sensitivity of the ESI-MS signal and minimize the levels of sodium adducts. It is also worth mentioning that the choice of the ion pairing reagent, the amount and chemical purity of all the other reagents, as well as the mobile phase preparation procedures are critical in generating cleaner, low-adduct ESI-MS spectra which facilitates the sensitive detection of lower-abundance oligonucleotide species.

In addition to the main target of the chemical synthesis - the 120-mer oligonucleotide, the crude sample contains some shorter length oligonucleotide impurities clearly visible, eluting before the major peak in the chromatograms presented in Figures 2 and 3. It is possible that slightly longer or shorter Poly(A) species, exceeding or being just under the 120-mer target, are present in the crude sample as well, but the RP-IP chromatography does not have the capability to resolve such closely related species. However, it has been

shown that size-exclusion chromatography (SEC) is able to better resolve these types of mixtures.[20]

For co-eluting oligonucleotides, the waters_connect INTACT Mass application performs automatic deconvolution of all detected oligonucleotides (above a user specified intensity threshold) regardless of their retention times or chromatographic profile.[19,21] This approach (automated deconvolution, non-RT based) is especially well suited for the analysis of co-eluting oligonucleotide impurities distributed around a parent oligonucleotide species. All the ESI-MS spectra belonging to oligonucleotides eluting between 5.5 and 6.5 min in the red trace chromatogram shown in Figure 3 were automatically deconvolved using the MaxEnt1 algorithm, with the software self-optimizing the deconvolution settings.[22]

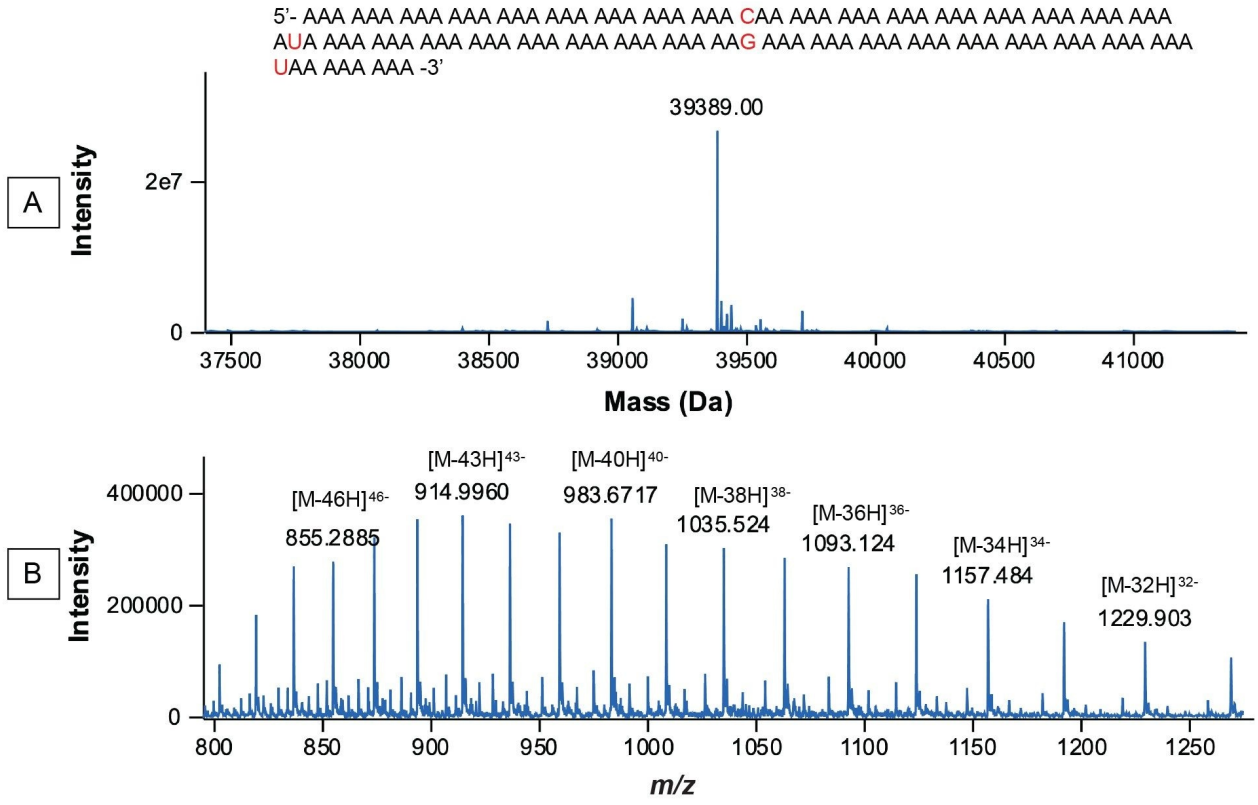


Figure 5. INTACT Mass spectra recorded for the synthetic 120-mer Poly(A) Tail oligonucleotide: (A) MaxEnt1 charge deconvolved spectrum; (B) raw ESI-MS spectrum. The sequence of the 120-mer shown in panel A contains a cytidine incorporated at position 28, a uridine incorporated at position 56, a guanosine incorporated at position 84, and another uridine at position 112. These four residues are used for increasing the chemical stability of the Poly(A) Tail.

In the next step, the deconvolved spectra were searched automatically against a list of 20 possible modifications including the addition and subtraction of up to 10 adenosine residue masses (329.2059 Da) as shown in the INTACT Mass screenshot displayed in Figure 6. The deconvolved spectral results from the automatically generated report are shown in Figure 7. As indicated by this report, seven Poly(A) Tail species were putatively identified, with mass accuracy better than 25 ppm. In addition to the expected 120-mer Poly(A), three species containing more adenosine residues (121,122, and 123-mers), as well as three species containing fewer adenosines (117, 118, and 119-mer) were identified. The lowest identified species, a 123-mer was detected at 1.5% abundance level according to its ESI-MS response.

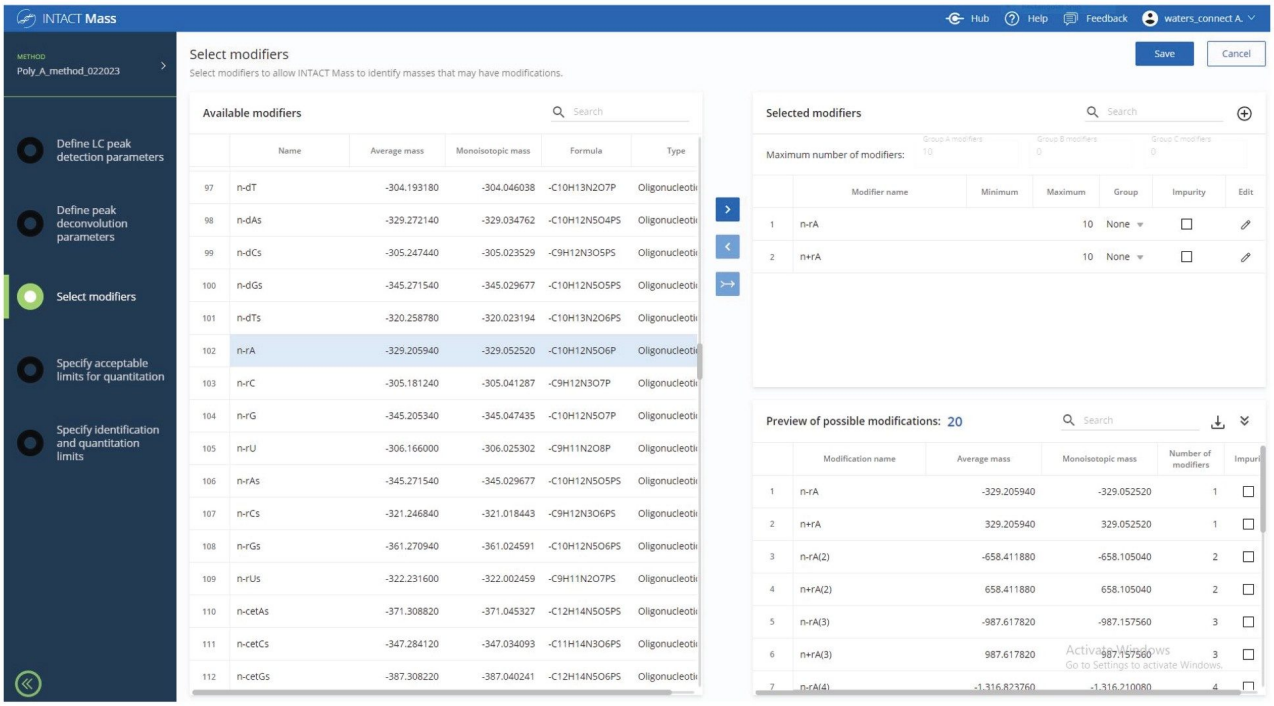


Figure 6. INTACT Mass App screenshot showing twenty possible oligonucleotide modifications, including the addition and subtraction of up to 10 adenosine masses (329.2059 Da), that could be present in the crude 120-mer synthetic Poly(A) Tail sample.

Results TIC TUV 260 MS

Identity: Pass Purity: Pass

	Molecule ID	Component		Observed mass (Da)	Expected mass (Da)	Mass error (ppm)	Identity result	Observed TIC RT (mins)	Observed UV RT (mins)
1							Pass		
2	39,388.64	39388.64 All Forms					Pass		
		Poly(A) Tail species:							
3	39,388.64	39,388.64	120-mer	39,389.17	39,388.64	13.4	Pass	6.75	6.70
4	39,388.64	39388.64 n-rA	119-mer	39,060.32	39,059.43	22.8	Pass	6.75	6.70
5	39,388.64	39388.64 n+rA	121-mer	39,718.44	39,717.85	14.8	Pass	6.75	6.70
6	39,388.64	39388.64 n-rA(2)	118-mer	38,730.98	38,730.23	19.4	Pass	6.75	6.70
7	39,388.64	39388.64 n+rA(2)	122-mer	40,047.76	40,047.05	17.6	Pass	6.75	6.70
8	39,388.64	39388.64 n-rA(3)	117-mer	38,401.57	38,401.02	14.3	Pass	6.75	6.70
9	39,388.64	39388.64 n+rA(3)	123-mer	40,377.12	40,376.26	21.3	Pass	6.75	6.70

Figure 7. Section of the INTACT Mass App report displaying the processing results obtained for the analysis of the 120-mer sgRNA oligonucleotide. Seven oligonucleotide species, ranging from 117- to 123-mers were identified with mass accuracies under 25 ppm.

Poly(A) Tail Analysis of an mRNA Digest

The same LC-MS assay developed for the synthetic Poly(A) Tail was applied to a RNAse T1 digested Fluc (firefly luciferase) mRNA. A previous report proposed an isolation procedure of the Poly(A) Tail oligonucleotide from the complex mRNA digest mixture, using $(dT)_{25}$ magnetic beads, in order to enrich this oligonucleotide fraction before subsequent LC-MS analysis.[14] However, in our case, the separation of the digestion mixture, shown in Figure 8, exhibited a well resolved, abundant, Poly(A) Tail eluting as the last eluting peak in the UV and TIC chromatograms.

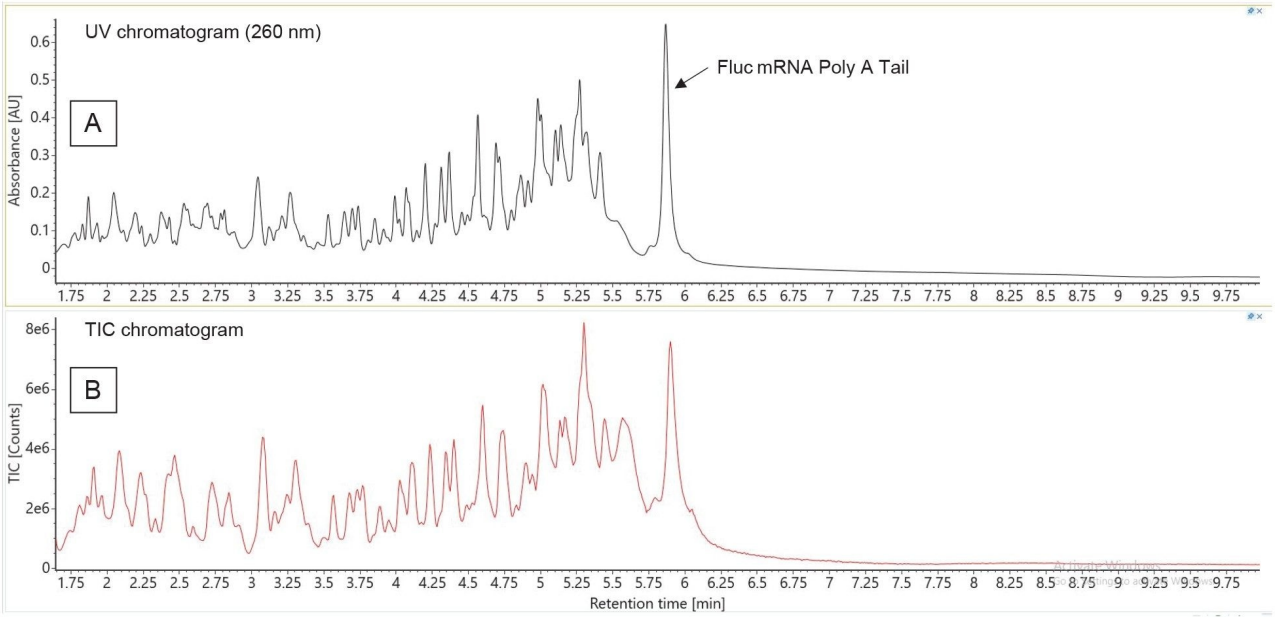


Figure 8. Separation of the RNase T1 digested Fluc mRNA: (A) UV chromatogram recorded at 260 nm; (B) TIC chromatogram recorded on the BioAccord LC-MS System.

The corresponding deconvolved ESI-MS spectrum (Figure 9), indicates a relatively high Poly(A) heterogeneity, with oligonucleotides in the range of 122–132-mers, based on the presence of many signals spaced by one adenosine mass (329.2 Da). Following INTACT Mass processing with Custom deconvolution parameters, almost all of these eleven Poly(A) variants were measured with mass accuracies better than 30 ppm, as shown in the results displayed in Figure 10.

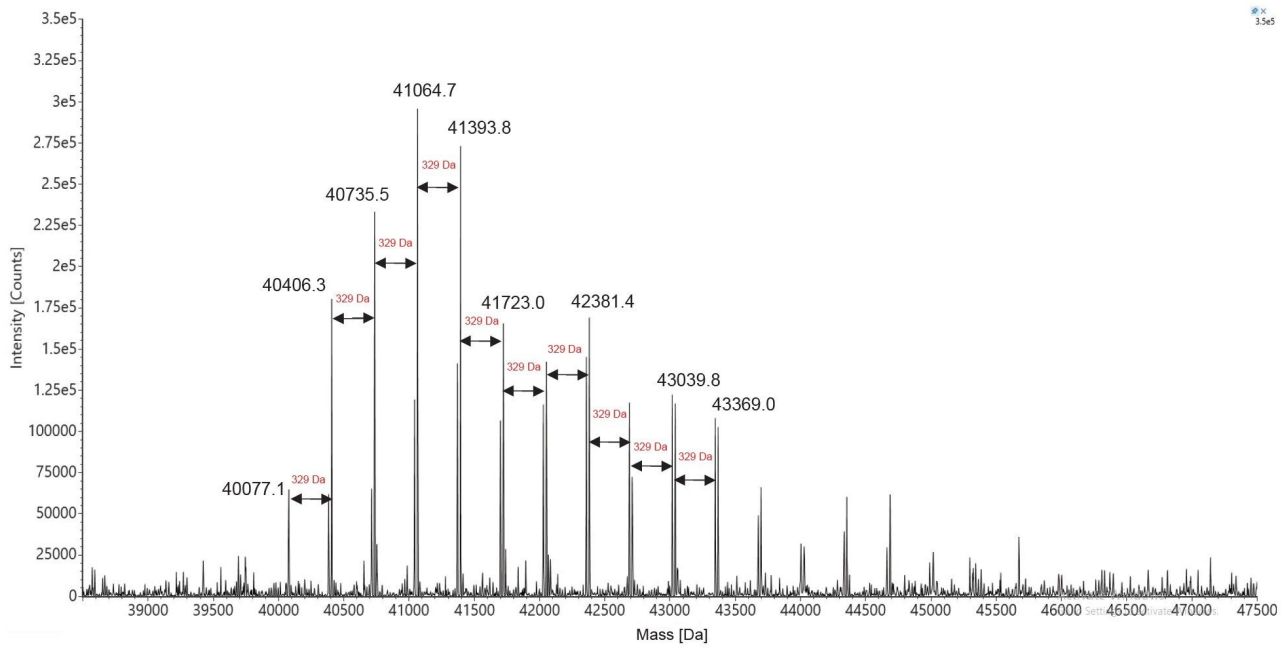
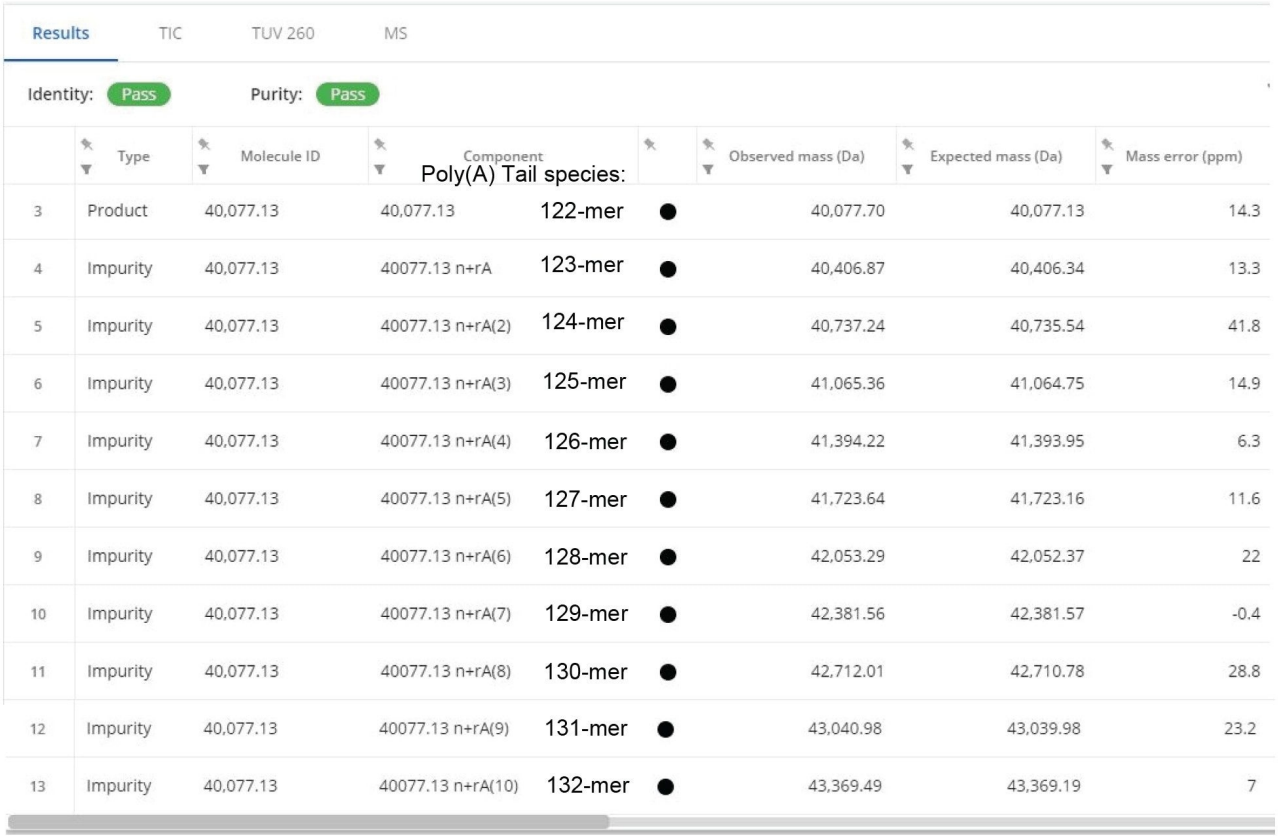


Figure 9. MaxEnt1 charge deconvolved spectrum of the Poly(A) Tail oligonucleotide mixture resulted from RNase T1 digestion of the Fluc mRNA. The Poly(A) Tail oligonucleotide has a wide dispersity, with up to eleven adenosine masses added to the first detected species, a 122-mer.



Results | TIC | TUV 260 | MS

Identity: Pass Purity: Pass

	Type	Molecule ID	Component	Poly(A) Tail species:		Observed mass (Da)	Expected mass (Da)	Mass error (ppm)
3	Product	40,077.13	40,077.13	122-mer	●	40,077.70	40,077.13	14.3
4	Impurity	40,077.13	40077.13 n+rA	123-mer	●	40,406.87	40,406.34	13.3
5	Impurity	40,077.13	40077.13 n+rA(2)	124-mer	●	40,737.24	40,735.54	41.8
6	Impurity	40,077.13	40077.13 n+rA(3)	125-mer	●	41,065.36	41,064.75	14.9
7	Impurity	40,077.13	40077.13 n+rA(4)	126-mer	●	41,394.22	41,393.95	6.3
8	Impurity	40,077.13	40077.13 n+rA(5)	127-mer	●	41,723.64	41,723.16	11.6
9	Impurity	40,077.13	40077.13 n+rA(6)	128-mer	●	42,053.29	42,052.37	22
10	Impurity	40,077.13	40077.13 n+rA(7)	129-mer	●	42,381.56	42,381.57	-0.4
11	Impurity	40,077.13	40077.13 n+rA(8)	130-mer	●	42,712.01	42,710.78	28.8
12	Impurity	40,077.13	40077.13 n+rA(9)	131-mer	●	43,040.98	43,039.98	23.2
13	Impurity	40,077.13	40077.13 n+rA(10)	132-mer	●	43,369.49	43,369.19	7

Figure 10. Section of the INTACT Mass App report displaying the processing results obtained for the analysis of the Fluc mRNA Poly(A) Tail. Eleven oligonucleotide species, ranging from 122- to 132-mers were identified with mass accuracies better than 30 ppm.

In a different graphical representation, the ESI-MS signals and their corresponding intensities were plotted in Figure 11 to calculate the average mass of the Poly(A) Tail of the Fluc mRNA which corresponds to a 126.5-mer. Both the Poly(A) dispersity as well as the average mass measurement are important quality attributes for therapeutic mRNA molecules. The average Poly(A) Tail length from the IP-RP LC-MS data presented here correlated well with the value obtained from the SEC-UV method.[20] Thus, the assay developed on the synthetic Poly(A) Tail oligonucleotide was demonstrated to be effective for analysis of Poly(A) Tails enzymatically cleaved from prophylactic/therapeutic mRNAs.

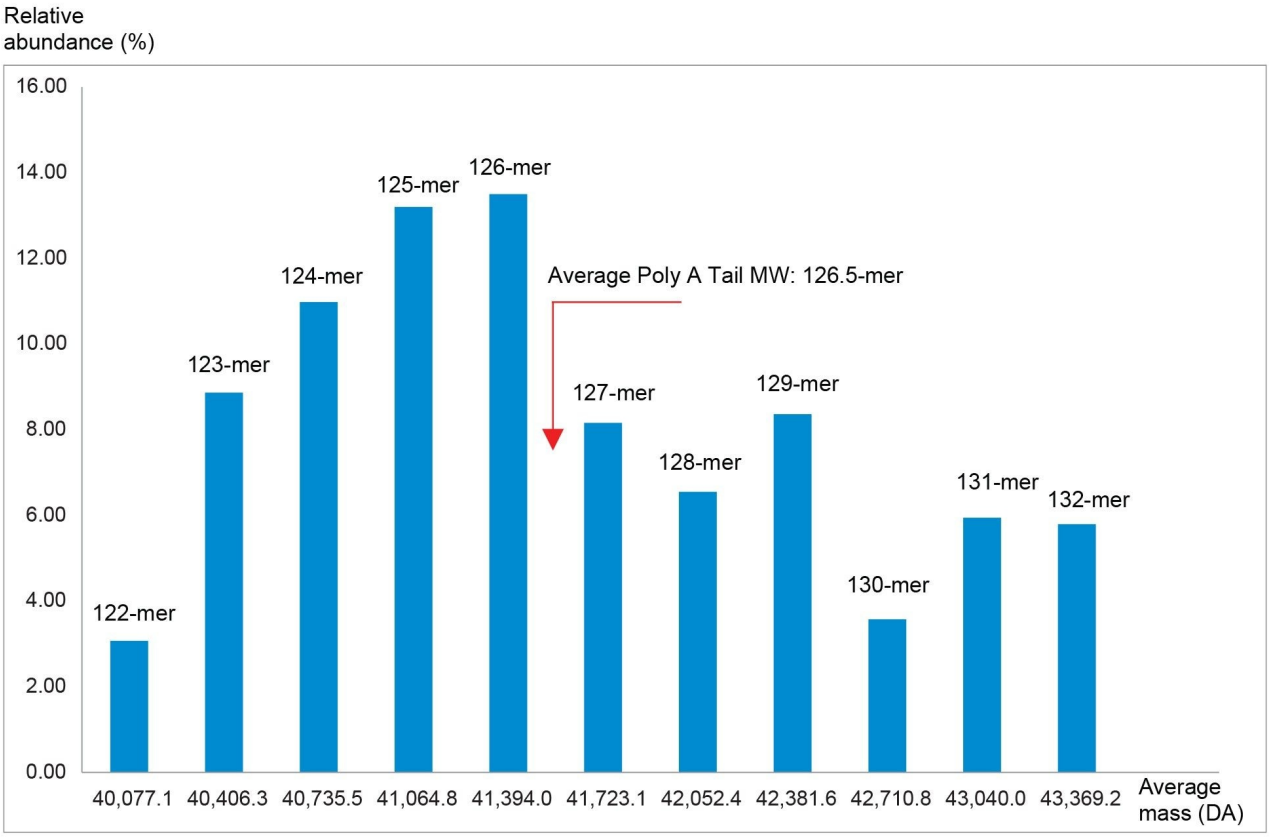


Figure 11. Distribution of the average ESI-MS spectral intensity versus the Poly(A) Tail length for the Fluc mRNA.

Conclusion

- An IP-RP LC-MS assay developed on the BioAccord LC-MS System provides length and heterogeneity information, as well as the average molecular weight for the Poly(A) Tail oligonucleotide region of prophylactic/therapeutic mRNAs
- Various co-eluting Poly(A) Tail oligonucleotides can be measured accurately with mass accuracies of better than 25 ppm on the bench-top BioAccord LC-MS System operated under the compliance-ready waters_connect informatics platform
- An LC-MS assay developed for a synthetic Poly(A) analogue, can be successfully applied for the analysis of real Poly(A) 3′-end fragments produced following the enzymatic digestion of prophylactic/therapeutic mRNAs mRNAs

References

1. Jackson NAC, Kester KE, Casimiro D, Gurunathan S, DeRosa F. The Promise of mRNA Vaccines: A Biotech and Industrial Perspective, *npj Vaccines,* 2020, 11, 1–6.

2. Sahin U, Kariko K, Tureci O. mRNA-Based Therapeutics – Developing a New Class of Drugs, *Nat Rev*, 2014, 13, 759–780.

3. Sheets MD, Wickens W. Two Phases in the Addition of a Poly(A) Tail, *Genes & Develop*, 1989, 3, 1401–1412.

4. Dreyfus M, Regnier P. The Poly(A) Tail of mRNAs: Bodyguard in Eukaryotes, Scavenger in Bacteria, *Cell*, 2002, 111, 611–613.

5. Chang H, Lim J, Ha M, Kim VN. TAIL-seq: Genome-Wide Determination of Poly(A) Tail Length and 3′ End Modifications, *Mol Cell*, 2014, 53, 1044–1052.

6. Nicholson AL, Pasquinelly AE. Tales of Detailed Poly(A) Tails, *Trends Cell Biol*, 2019, 29, 191–200.

7. Jiang T, Yu N, Kim J, Murgo JR, Kissai M, Ravichandran K, Miracco EJ, Presnyak V, Hua S. Oligonucleotide Sequence Mapping of Large Therapeutic mRNAs via Parallel Ribonuclease Digestions and LC-MS/MS, *Anal Chem,* 2019, 91, 8500–8506.

8. Vanhinsbergh C, Criscuolo A, Sutton JN, Murphy K, Williamson AJK, Cook K, Dickman MJ. Characterization and Sequence Mapping of Large RNA and mRNA Therapeutics Using Mass Spectrometry, *Anal Chem,* 2022, 94, 7339–7349.

9. Muttach F, Muthmann N, Rentmeister A. Synthetic mRNA Capping, *Beilstein J Org Chem,* 2017, 13, 2819–2832.

10. Beverly M, Dell A, Parmar P, Houghton L. Label-Free Analysis of mRNA Capping Efficiency Using RNse H probes and LC-MS, *Anal Bioanal Chem*, 2016, 408, 5021–5030.

11. Nguyen J, Chan SH, Roy B, Gilar M, Robb B, Chen W, Lauber MA. Rapid Analysis of Synthetic mRNA Cap Structure Using Ion-Pairing RPLC with the BioAccord LC-MS System, Waters Application Note, 720007329, 2021.

12. Muthmann N, Spacek P, Reichert D, van Dulmen M, Rentmeister A. Quantification of mRNA Cap-Modifications by Means of LC-QqQ-MS, *Methods*, 2022, 203, 196–206.

12. Muthmann N, Spacek P, Reichert D, van Dulmen M, Rentmeister A. Quantification of mRNA Cap-Modifications by Means of LC-QqQ-MS, *Methods*, 2022, 203, 196–206.

13. Chan SH, Whipple JM, Dai N, Kelley TM, Withers K, Tzertzinis G, Correa Jr IR, Robb GB. RNase-H-based Analysis of Synthetic mRNA 5′ Cap Incorporation, *RNA*, 2022, 28, 1144–1155.

14. 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.

15. Strezsak SR, Pimentel AJ, Hill IT, Beuning PJ, Skizim NJ. Novel Mobile Phase to Control Charge States and Metal Adducts in the LC/MS for mRNA Characterization Assays, *ACS Omega Reviews*, 2022, 7, 22181–22191.

16. Doneanu C E, Fox J, Harry E, Knowles C, Yu Y Q, Fredette J, Chen W. An Automated Compliance-Ready LC-MS Workflow for Intact Mass Confirmation and Purity Analysis of Oligonucleotides, Waters Application Note, 720006820, 2020.

17. Doneanu C E, Fox J, Harry E, Knowles C, Yu Y Q, Fredette J, Chen W. Intact Mass Confirmation Analysis on the BioAccord LC-MS System for a Variety of Extensively Modified Oligonucleotides, Waters Application Note, 720007028, 2020.

18. Doneanu C E, Knowles C, Fox J, Harry E, Yu Y Q, Fredette J, Chen W. Analysis of Oligonucleotide Impurities on the BioAccord System with ACQUITY Premier, Waters Application Note, 720007301, 2021.

19. Doneanu C E, Boyce P, Shion H, Fredette J, Berger S J, Gastall H, Yu Y Q. LC-MS Analysis of siRNA, Single Guide RNA and Impurities Using the BioAccord™ System with ACQUITY™ Premier and New Automated INTACT Mass Application, Waters Application Note, 720007546, 2022.

20. Gilar M. Size-Exclusion Chromatography Method for Poly(A) Tail Analysis of mRNA, Waters Application Note, 720007853, 2023.

21. Shion H, Boyce P, Berger S J, Yu Y Q. INTACT Mass™ - a Versatile waters_connect™ Application for Rapid Mass Confirmation and Purity Assessment of Biotherapeutics, Waters Application Note, 720007547, 2022.

22. Ferrige AG, Seddon MJ, Green BN, Jarvis SA, Skilling J. Disentangling Electrospray Spectra With Maximum Entropy, *Rapid Commun. Mass Spectrom.*, 1992, 6, 707–711.

Featured Products

ACQUITY UPLC I-Class PLUS System <https://www.waters.com/134613317>

BioAccord LC-MS System for Biopharmaceuticals <https://www.waters.com/waters/nav.htm?cid=135005818>

Biopharmaceutical Platform Solution with UNIFI <https://www.waters.com/10195515>

ACQUITY UPLC Tunable UV Detector <https://www.waters.com/514228>

720007925, August 2023

BioAccord™ SYSTEM

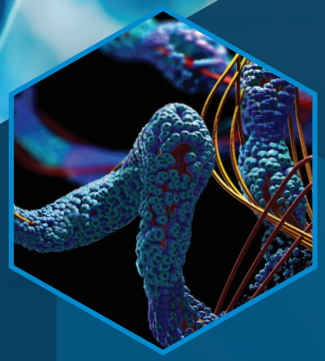
The BioAccord LC-MS System with ACQUITY Premier UPLC

Access RNA critical quality attributes quickly with a compact, integrated, and easy-to-use system for biopharma applications

LEARN MORE
waters.com/BioAccord

Waters™

Intact mRNA Drug Substance Testing with SEC MALS



Intact mRNA Analysis

Lavelay Kizekai, Principal Scientist, Evaluations & App Science

Integrity analysis of long RNA molecules is essential for validating product quality in advanced therapeutics and vaccines. RNA's functional integrity hinges on two critical factors—stability and solubility. Degradation, the enzymatic breakdown of transcripts, compromises gene expression accuracy and yields truncated RNAs that fail to produce the intended protein products. Aggregation limits RNA accessibility for translation and can illicit an immune response. By examining intact RNA rather than relying exclusively on digests or calibration-based approaches, researchers can confirm purity, molecular weight, and stability at each stage of development.

Size-exclusion chromatography (SEC) coupled with multi-angle light scattering (MALS) addresses this complex analytical need by directly measuring molecular weight distributions and the presence of aggregated or truncated species. Wide-pore SEC columns are particularly well-suited for large RNA constructs, as they accommodate broad hydrodynamic radii while minimizing unwanted secondary interactions. Combining SEC with MALS provides robust data on molar mass, shape, and purity—critical attributes that guide process optimization and ensure product consistency.

This application note demonstrates a systematic method development approach—employing design of experiments (DoE)—to refine SEC-MALS for both mRNA and dsDNA ladders. By tuning mobile phase composition, ionic strength, and column temperature, consistent separation and robust quantitation are demonstrated. The multi-attribute data generated by simultaneous MALS, UV, and refractive index detection provide a thorough profile of molecular weight, aggregation state, and structural integrity. From process development to final quality control, SEC-MALS offers a scalable workflow suited for monitoring RNA therapeutics, ensuring safer and more effective products.

Waters™

Application Note

Optimized Method Conditions for mRNA Characterization by SEC-MALS With GTxResolve™ Premier SEC 1000 Å 3 µm Columns

Lavelay Kizekai, Bretton Fletcher, Szabolcs Fekete, Balasubrahmanyam Addepalli, Sophia Kendrick, Michelle Chen, Matthew A. Lauber

Waters Corporation, Wyatt Technology

Abstract

Detailed characterization of the therapeutic drug substance, such as mRNA payload, is essential to ensure the safety and efficacy of gene therapy products. In this application note, we evaluated the efficiency of GTxResolve Premier SEC 1000 Å 3 µm Columns for size-exclusion chromatography coupled with multi-angle light scattering (SEC-MALS) to assess critical quality attributes of mRNAs. The hydrophilic MaxPeak™ High-Performance Surfaces (h-HPS) column hardware and high efficiency packed beds consisting of 3 µm bridged ethylene polyethylene oxide (BE-PEO) surface-modified silica exhibited enhanced sensitivity and low MALS noise with little to no particle shedding.[1]

Accurately assessing key quality attributes such as molar mass, size, and aggregation is essential for mRNA drug substances.[2] This application note demonstrates the benefits of method development through systematic design of experiments (DoE) to ensure appropriate separation and accurate quantitation of molar mass, size, and aggregation of mRNA drug substances. The optimized SEC-MALS method employed in current investigations can yield quality molar mass measurements, aggregation profiling, structural, and conformational insights with robust, reproducible analyses.

Benefits

- The 1000 Å pores offer appropriate and well-matched fractionation range for a wide range of mRNA sizes
- Bridged Ethylene Polyethylene Oxide (BE-PEO) surface-modified silica provides low MALS noise suitable for high-confidence biophysical measurements of mRNA
- The particle size (3 µm) is suitable for resolving smaller and larger components of complex dsDNA ladder
- Low absorption surfaces provide flexible and expedient method development

Introduction

Significant interest has been shown in advancing and utilizing messenger ribonucleic acid (mRNA) technologies for therapeutic products, particularly highlighted by the development of COVID-19 vaccines. mRNA molecules are transient biopolymers that convert genetic information into functional proteins.[3] This mRNA technology has proven to be faster in development and manufacturing than other vaccine platforms, making it a promising solution for addressing future pandemics and other infectious diseases like rabies, Zika virus, and cytomegalovirus infections.[4] The theoretical risk of integrating the target gene into the host genome is minimized when using mRNA-based vaccines, relative to viral vectors, due to their transient nature and their degradation after protein translation. On the other hand, stability and storage concerns can be problematic for mRNA therapeutics requiring ultra-cold conditions. Thus, it is essential to quantify key quality and stability-indicating attributes that potentially affect potency and efficacy.

Size-exclusion chromatography coupled with multi-angled light scattering (SEC-MALS) with an optimized method has been shown to effectively measure biophysical attributes of mRNA, such as molar mass, aggregate percentage, radius of gyration (R_g), hydrodynamic radius (R_h), and R_g/R_h ratios, for comprehensive insights into the size, conformation, and purity of mRNA molecules. These attributes are critical not only during the development process but also in the final therapeutic product. In this application note, we demonstrate the utility of systematic optimization (DoE) of mobile phase composition for improved separation of a typical mRNA (Cas9) molecule using a GTxResolve Premier SEC 1000 Å 3 µm 4.6 x 150 mm Column. The high-efficiency column was also used to highlight the size-based separation of a dsDNA ladder and mRNA products and to demonstrate the necessity of molar mass quantitation by MALS in order to provide the identity of peaks that differ in conformation and elution properties.

Experimental

The mobile phase for all SEC separations was 50 mM Tris-HCl, 150 mM NH_4Cl pH 7.5, filtered to 0.1 µm.

Waters dsDNA 50 to 1350 Ladder was prepared by reconstituting the lyophilized sample with 100 µL of 18.2 MΩ water and gently aspirating up and down with a 100 µL pipette to obtain a 1 mg/mL working solution.

Porcine thyroglobulin (PTG) (Sigma, p/n: T1126–100 mg) was prepared by dissolving in PBS at a 1.0 mg/mL concentration and filtered to 0.1 µm.

eSpCas9 mRNA (N1-Methylpseudouridine/m1Ψ) (Cas9-mRNA) (Neat) was obtained from GenScript and stored at -80 °C. The sample was thawed to room temperature and placed in an LC sample manager maintained at 6 °C.

eSpCas9 mRNA (N1-Methylpseudouridine/m1Ψ) (Cas9-mRNA) was annealed by incubating the thawed sample in a 1.5 mL microcentrifuge™ tube for 2 minutes at 75 °C. The sample was reannealed by immediately cooling at 6 °C in an LC instrument sample manager maintained at 6 °C.

TriLink™ CleanCap® FLuc (5 moU) 100 µg (p/n: L-7202–100) was obtained from TriLink BioTechnologies and stored at -80 °C. The sample was thawed to room temperature and placed in an LC sample manager at 6 °C.

TriLink™ CleanCap® EPO-mRNA (5 moU) 100 µg (p/n: L-7209–100) was obtained from TriLink BioTechnologies and stored at -80 °C. The sample was thawed to room temperature and placed in the LC sample manager at 6 °C.

LC Conditions

LC system:	Arc Premier™ with Quaternary Solvent Manager (QSM) and Flow Through Needle Sample Manager (SM-FTN)
Vials:	Max Recovery Vials and Caps (Waters p/n: 186000327C) and Waters 300 µL polypropylene screw neck vial (Waters p/n: 186004112)

Column(s):	▪ GTxResolve Premier SEC 1000 Å, 3 µm, 7.8 x 300 mm Column (Waters p/n: 186010738) ▪ GTxResolve Premier SEC 1000 Å, 3 µm, 4.6 x 150 mm Column (Waters p/n: 186010735)
Column temperature:	40 °C
Sample temperature:	6 °C
Sample manager washes:	18.2 MΩ water
Seal wash:	10 % HPLC grade methanol / 90% 18.2 MΩ water (v/v)
Samples and injection volume:	▪ Porcine thyroglobulin (PTg): 50 µL ▪ Waters dsDNA 50 to 1350 Ladder: 8 µL ▪ TriLink CleanCap™ FLuc mRNA: 8 µL ▪ TriLink CleanCap EPO-mRNA: 8 µL ▪ GenScript™ eSPCas9 mRNA (N1-Methylpseudouridine/m1Ψ): 8 µL)
Flow rate:	0.30 mL/min
Mobile phase A:	50 mM Tris, 150 mM NH_4Cl pH 7.5
Sample(s):	Porcine thyroglobulin (PTg) (Sigma, p/n: T1126–100 mg); Waters dsDNA 50 to 1350 Ladder (Waters p/n: 186010778); TriLink CleanCap FLuc (5moU) 100 µg (p/n: L-7202–100); TriLink CleanCap EPO-mRNA (5moU) 100 µg (p/n: L-7209–100); GenScript eSpCas9(N1) mRNA 1 mg/mL (p/n: SC2325–1 mg)

Gradient:	Isocratic
LC system control:	HPLC CONNECT™ 4 Software

Detector conditions

Multi-angle light scattering (MALS):	Wyatt DAWN™ MALS Detector with a WyattQELS ™ embedded online Dynamic Light Scattering (DLS) Module
UV/Vis:	Arc Premier W2998 PDA Detector Channel 1: 280 nm with a resolution at 4.8 nm; Channel 2: 260 nm with a resolution at 4.8 nm
Refractive Index (dRI):	Wyatt Optilab™ dRI Detector
Data Acquisition and analysis:	ASTRA™ 8 Software

Results and Discussion

Exploring method conditions

Taking full advantage of SEC requires method optimization. This is because the entropy-based distribution of solute between internal pores and interstitial column volume is often influenced by unwanted secondary interactions (electrostatic, H-bonding, hydrophobic), which adversely impact analyte solubility, retention, and structure and compromise the separation quality and component recovery of the drug substance. Factorial designs involving rational Design of Experiments (DoEs) can help identify the effects of different variables and help optimize the most influential factors to understand the relationship between mobile phase input variables and the observed responses.[4] A 2-level design for the screening phase followed by 2^3 full factorial designs were

run to screen for both quantitative and categorical factors. The individual blocks examined the effect of co-solvent (organic modifier), ionic strength, and temperature of the mobile phase (quantitative factors). The effect of categorical factors (type of buffer and additives, including PBS and Tris buffers and ammonium and magnesium salts) was studied in separate experimental blocks. The monitored responses include (1) pre-peak area % (high molecular weight-HMW amount), (2) total peak area (recovery), (3) the elution time of the monomer peak, (4) the width of elution window (including all observed peaks), and (5) the width of the monomer peak. Cas9 mRNA was used as test species because of the presence of a significant amount of HMW species.

We hypothesized that the routine analysis adopted for proteins is not suitable for nucleic acids because of the differences in physicochemical properties. After evaluation of the input factor effects, three influential mobile phase factors were identified as affecting the separation performance: (1) the nature of the additive (chaotropic or "water structure breaking" vs. kosmotropic or "water structure forming"), (2) the ionic strength, and (3) the column temperature. Organic co-solvents showed no significant effects. Interestingly, the cation of the additive seemed to have a greater effect than the anion. Ammonium salts were found to be beneficial. It is worth mentioning that $MgCl_2$ as an additive may result in high selectivity and resolution at 10 mM concentration, but a huge negative effect was always observed on mRNA recovery. Therefore, we decided to avoid the use of magnesium salts. In general, separation improved up to 200–250 mM mobile phase total ionic strength, but further increases did not result in separation improvement. Increased temperature also significantly improved selectivity and resolution.

To conclude, ammonium chloride or ammonium sulfate as an additive in combination with Tris buffer and elevated temperature were found to be good conditions for high resolution separation and high recovery. Our recommendation is to start with mobile phase consisting of 50 mM Tris (pH 7.5) and 250 mM NH_4Cl, and column temperature of 40–50 °C. Figure 1 illustrates our DoE approach and shows a comparison between a historically set (non-ideal) condition and optimized condition when injecting Cas9 mRNA sample.

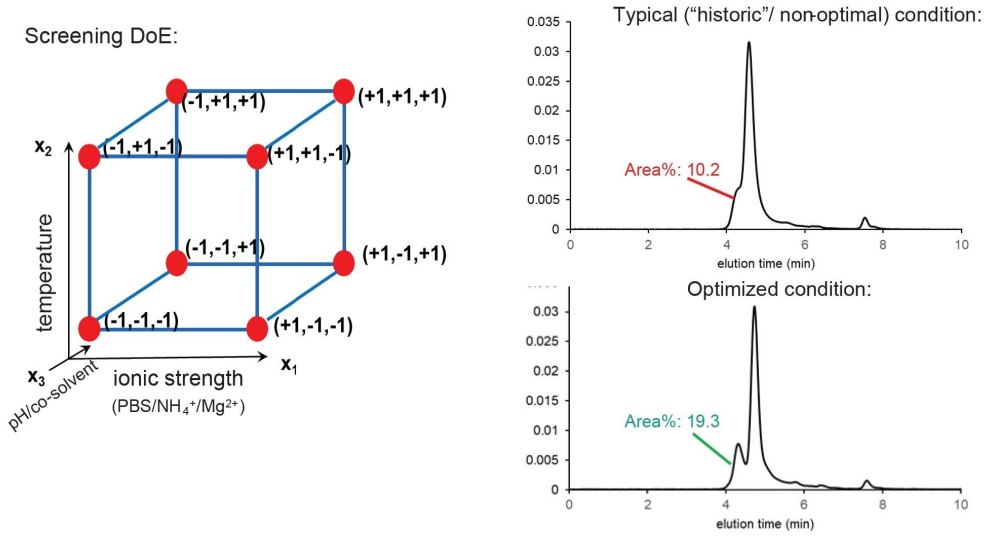


Figure 1. Schematic illustration of screening DoE approach (left) and comparison of Cas9 mRNA chromatograms obtained using non-optimized and optimized conditions (right). The optimized condition corresponds to: Mobile phase: 50 mM TRIS (pH 7.5) + 250 mM NH_4Cl, Temperature: 50 °C, F=0.25 mL/min, Column: GTxResolve Premier SEC 1000 Å, 3 µm, 4.6 x 150 mm Column (Waters p/n: 186010735).

SEC-MALS Analysis of mRNA using Optimized Method Conditions

Using the DoE method conditions established in the previous section, multi-attribute quantitation of several mRNAs and a dsDNA ladder were performed by size-exclusion chromatography coupled with UV, MALS, and dRI detection (SEC-MALS). SEC-MALS was performed using 50 mM TRIS (pH 7.5) +250 mM NH_4Cl with temperature set at 40 °C and column flow rate set at 0.30 mL/min. HPLC CONNECT software was developed to run synergistically with Arc™ Premier system and ASTRA software for SEC-MALS method development, data acquisitions, and processing. Porcine thyroglobulin was used to determine system parameters, including normalization, band broadening, and alignment, (results not shown here) before mRNA evaluations.

In addition to characterizations that standard SEC-UV methods provide, simultaneously collecting UV, MALS, and RI signals in an SEC method enables the absolute measurement of molar mass, R_g and R_h, and UV extinction coefficients, providing a simple and efficient workflow for elucidating multiple quality attributes simultaneously. This combination of information enables researchers to identify each eluting peak, independent of elution time. In

addition, evaluating size (via R_g and/or R_h) with molar mass provides key conformational information about each eluting peak and can inform whether an unexpected elution time is the result of column interactions or differences in molecular structure relative to a traditional size standard. Finally, the directly measured extinction coefficient can be used in orthogonal offline measurements and is critical for routine quantitative analyses, quality control, and dosing.

Waters 50–1350 dsDNA standard was used for system suitability to validate column performance for SEC-MALS (Figure 2). The Waters dsDNA ladder is a mixture of 17 double-stranded DNA species ranging from 50 to 1350 base pairs and is designed and certified for use with SEC columns of pore size greater than 450 Å. Figure 2A shows the UV chromatogram for dsDNA ladder (8 µL) acquired on GTxResolve Premier SEC 1000 Å, 3 µm, 7.8 x 300 mm Column, with an overlay of the measured molar mass for each species as determined by MALS. All 17 species are well-distinguishable with some fully separated, with the peak profile corresponding to the relative abundance provided for the ladder. The measured weight-averaged molar masses (M_w) for each peak were in excellent agreement with the expected molar masses for each DNA species, as shown in Fig. 2B, where linear regression analysis yields a slope of 1.03 with 95% confidence interval between 1.00 and 1.05.

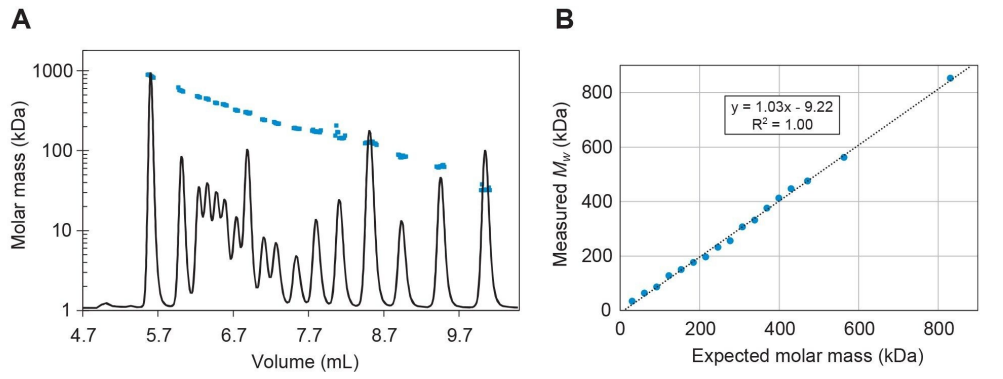


Figure 2. The molar mass of dsDNA Ladder superimposed on UV chromatograms acquired at 260 nm (A) and plot of weight-averaged molar mass for each eluting dsDNA species against expected molar mass for each species (B). Separation was performed using a GTxResolve Premier SEC 1000 Å, 3 µm, 7.8 x 300 mm Column (Waters p/n: 186010738) using 50 mM TRIS (pH 7.5) +250 mM NH_4Cl, Temperature: 40 °C, Flow rate=0.30 mL/min.

SEC-MALS can provide key insights on the identity of complex and coeluting mixtures, which cannot be provided by SEC-UV alone. To demonstrate this, we analyzed two Cas9 mRNA samples with varying degrees of high molecular weight species (HMWS) and low molecular weight species (LMWS) using the same conditions discussed in the DoE experiments above. Figure 3 shows the measured molar mass overlaid onto the UV chromatograms for the two Cas9 mRNA samples. As expected, the UV chromatogram for the "neat" Cas9 mRNA sample (black curve), where the sample was thawed on ice from storage conditions and placed directly into the sample manager, shows significant UV absorption ahead of the main elution peak, indicating the presence of multiple high molecular weight species (HMWS). The weight-averaged molar masses (M_w) measured for the first and second HMWS peaks correspond roughly to trimer and dimer Cas9 mRNA species, respectively (M_w=4690 kDa and 3050 kDa, respectively). The M_w measured across the main peak (1530 kDa) is in agreement with theoretical monomer molar mass (1460 kDa) but is slightly elevated due to the coelution of HMWS. Denaturing aggregates with a short heat treatment followed by re-annealing removes nearly all the HMWS. Although the identity of the HMWS remains the same (mostly dimer and trimer), the mass fraction decreases from ~33% to ~2% (Table 1). Similarly, the decrease in coeluting HMWS results in the measured M_w of the monomer peak (1470 kDa) to within 1% of the theoretical monomer molar mass. Absolute identification of each eluting species can potentially provide insights into molecular interactions leading to the formation of high- and low-molecular weight molecules, sample integrity, and sample purity.

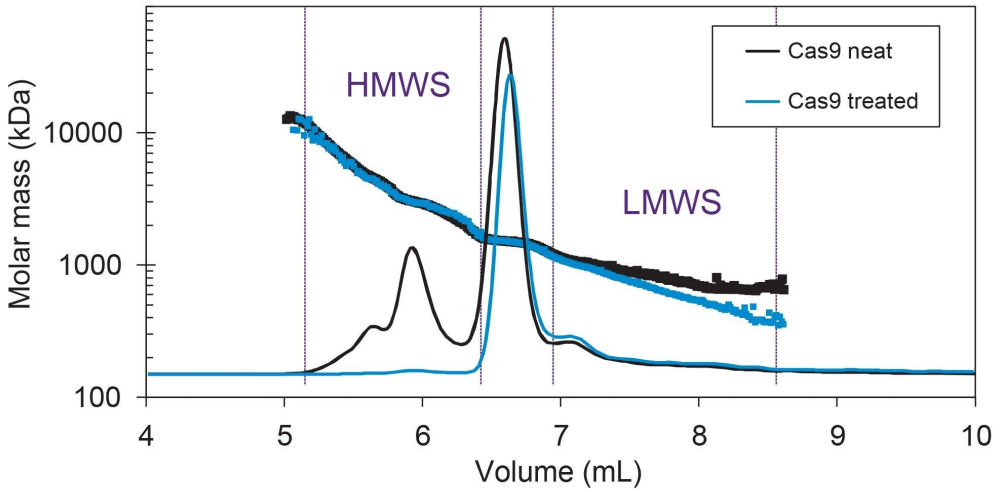


Figure 3. Molar mass of Cas9 mRNA samples injected neat (black) and after denaturing and reannealing (blue), superimposed on UV chromatograms acquired at 260 nm. Separation performed with a GTxResolve Premier SEC 1000 Å, 3 µm, 7.8 x 300 mm Column (Waters p/n: 186010738) using 50 mM TRIS (pH 7.5) +250 mM NH_4Cl, Temperature: 40 °C, F=0.30 mL/min.

	HMWS		Main peak		LMWS	
	M_w (kDa)	Mass. frac.	M_w (kDa)	Mass. frac.	M_w (kDa)	Mass. frac.
Neat	3660 ± 20	33.1 ± 0.6%	1530 ± 1	55.5 ± 0.3%	909 ± 13	11.5 ± 0.4%
Denatured	3390 ± 60	1.9 ± 0.1%	1470 ± 1	74.4 ± 0.4%	781 ± 1.4	23.7 ± 0.5%

Table 1. Weight-averaged molar mass (M_w) and mass fraction, measured for three elution regions corresponding to high-molecular weight species (HMWS), mRNA monomer, and low-molecular weight species (LMWS). All values are average and standard deviation of three replicate injections.

Finally, we analyzed the elution time and measured molar mass for three mRNA samples compared to the dsDNA ladder: heat-treated and re-annealed Cas9 mRNA, firefly luciferase mRNA (FLuc), and erythropoietin mRNA (EPO). Figure 4 shows the measured molar masses overlaid onto the UV chromatograms for three

different mRNA species run on the same HPLC setup and same conditions as the dsDNA ladder. Again, the absolute measurement of molar mass by MALS detector allows us to identify mRNA products with measured M_w values of Cas9 (blue), FLuc (green), and EPO (magenta) in excellent agreement with expected molar masses (summarized in Table 2).

In addition to molar masses and mass fractions, online UV-MALS-RI detection enables the measurement of other biophysical properties and sample attributes like size, UV extinction coefficient, and mRNA purity (A260/A280) with excellent reproducibility. Table 2 summarizes some of these measurements for each of the different mRNA samples used (n=3). Extinction coefficients measured at 260 nm are in near-perfect agreement with expected values for single-stranded RNA (~25 mL mg-1 cm-1, which corresponds to a conversion factor of 40 µg/mL). Similarly, A260/A280 measurements match the expected ratio of 2.1 for pure RNA (1.86–1.88 for dsDNA).[5]

The relationship between size—either hydrodynamic radius or radius of gyration—and molar mass is correlated with the conformation of the analyte. For large biomolecules, like mRNA, the molecular conformation can be assessed by plotting R_g versus molar mass on a log-log plot and evaluating the slope of line of best fit. For highly compact molecules, that slope is expected to be close to 1/3, for a random coil polymer 1/2, and for a rigid rod 1. Doing this conformation analysis for the dsDNA ladder and three mRNA produces slopes of 0.75 and 0.43, respectively, suggesting that the dsDNA forms a long, extended structure whereas the mRNA exists as a compact linear coil. These observations are in good agreement with the known structure of these molecules, whereby the double-helix of dsDNA locks the molecule into an extended structure that grows with a prescribed geometry, as compared to mRNA which is more flexible.

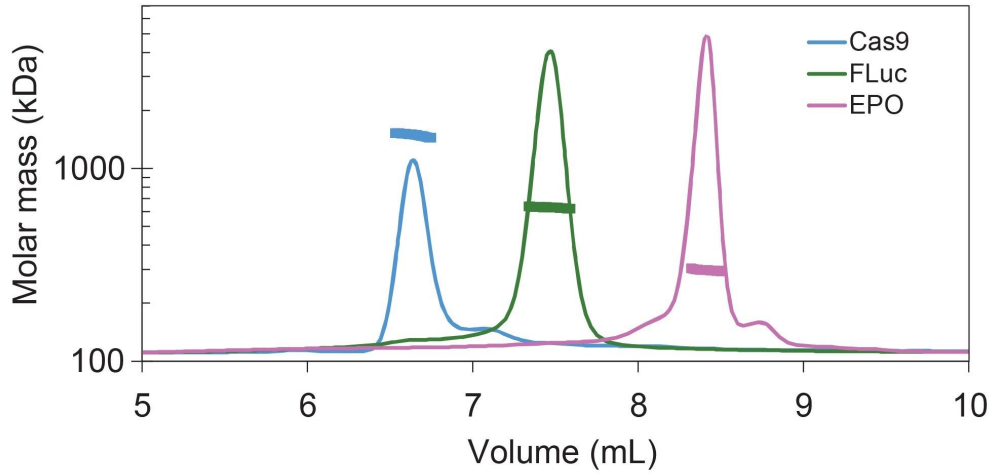


Figure 4. Molar mass of Cas9 (blue), Firefly Luciferase (FLuc) (green), and erythropoietin (EPO) (magenta), superimposed on UV chromatograms acquired at 260 nm. Separation performed using GTxResolve Premier SEC 1000 Å, 3 µm, 7.8 x 300 mm Column (Waters p/n: 186010738) using 50 mM TRIS (pH 7.5) +250 mM NH_4Cl, Temperature: 40 °C, F=0.30 mL/min.

	Number of bases	M_w (kDa)	R_g (nm)	$\varepsilon_{260\ nm}$ (mL mg^{-1} cm^{-1})	$\varepsilon_{280\ nm}$ (mL mg^{-1} cm^{-1})	A260/A280
Cas9	4471	1500 ± 1	23.8 ± 0.0	24.7 ± 0.1	12.5 ± 0.0	1.97 ± 0.00
FLuc	1922	631 ± 0	17.9 ± 0.1	24.1 ± 0.7	11.1 ± 0.3	2.18 ± 0.00
EPO	859	298 ± 0	11.9 ± 0.2	25.0 ± 0.2	12.8 ± 0.1	1.96 ± 0.00

Table 2. Summary of measured attributes of Cas9, FLuc, and EPO mRNA. ε is the measured extinction coefficient for each sample and wavelength, and A is the UV peak area (AU min) for each sample and wavelength, calculated across the monomer peak in the sample. All results are average and standard deviation of three injections.

The molecular conformation insights provided by the MALS quantitation support for absolute measurements like SEC-MALS over a traditional calibration curve method with SEC-UV alone. The applicability of calibration curves is heavily impacted by differences in chemical and structural properties between calibration standards and user samples. Indeed, Figure 5B shows a typical calibration plot with measured M_w as a function of elution volume for both the dsDNA samples and the mRNA samples, illustrating the potential shortcomings of using a calibration curve for mass or size determination. (Scaling the y-axis in Fig. 5B to number of nucleotides does not improve the estimation of mRNA size, data not shown.) While the calibration curve indicates that all 17 dsDNA species are well-separated and elute in the linear range of the column, indicating great column performance and suitability, the use of dsDNA as a calibrant would have led to significant underestimation of the length and molar mass of the mRNA samples.

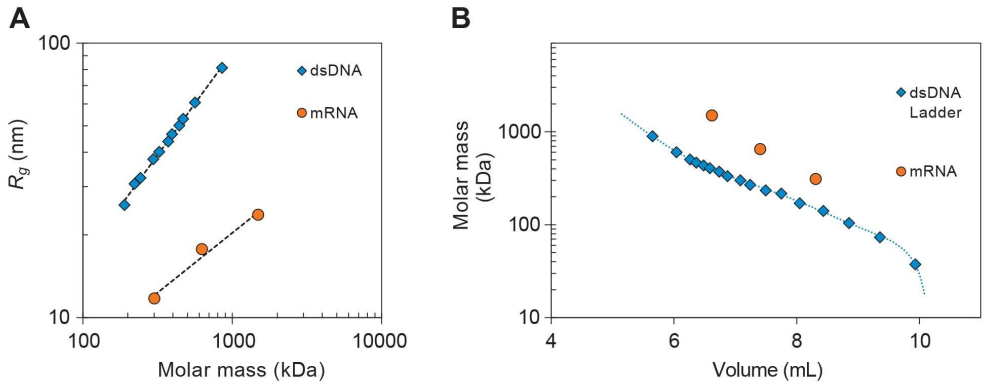


Figure 5. (A) R_g versus molar mass for select peaks from the dsDNA ladder and from the monomer peak of the Cas9, FLuc, and EPO mRNA samples. The position and slope of the data reflect the difference in conformation between dsDNA and mRNA. (B) Plot of measured weight-averaged molar masses of dsDNA ladder and mRNAs to show varying elution behavior.

Conclusion

In this application note, we illustrate the utilization of GTxResolve Premier SEC 1000 Å 3 µm Column technology, and the significance of optimized method conditions to give excellent resolution of mRNA biomolecules and quality MALS analysis for adequate biophysical attribute characterization. The appropriately matched pore size and inert particle of the packing material efficiently separated mRNA aggregates and fragments from monomer and dsDNA Ladder resolution to enable mRNA integrity analysis to ensure purity, identity, safety, and the efficacy of mRNA drug therapeutics.

References

1. Camacho, K. J., Tchoul, O., Xu, Y., Finny, A. S., Kizekai, L., Byrd, S., Addepalli, B., Xu, M., Lauber, M., & McLaughlin, J. (2024). Bridged Ethylene Polyethylene Oxide Surfaces To Improve Packing Materials For Widepore Size Exclusion Chromatography. *Journal of Separation Science*, 47(20). https://doi.org/10.1002/jssc.202400541 <https://doi.org/10.1002/jssc.202400541>

2. Wang, P., Akula, R., Chen, M., & Legaspi, K. (2023). SEC-MALS Method for Characterizing mRNA Biophysical Attributes. AN1616. *Wyatt Technology Corporation*.

3. USP. (2023). Draft Guidelines: 3rd Edition Analytical Procedures for Quality of mRNA Vaccines and Therapeutics. *United States Pharmacopeia*. Retrieved from https://www.usp.org/sites/default/files/usp/document/about/usp-commitment-to-confidentiality.pdf < https://www.usp.org/sites/default/files/usp/document/about/usp-commitment-to-confidentiality.pdf>

4. S. Fekete, M.A. Lauber, A Systematic Quick Method Development Approach to Optimize Protein Size Exclusion Chromatography, Waters application note. 720007790, November, 2022.

5. Giron Koetsier, Ph.D., and Eric Cantor, Ph.D. A Practical Guide to Analyzing Nucleic Acid Concentration and Purity with Microvolume Spectrophotometers. *New England Biolabs.* 2019.

Featured Products

Arc Premier System <https://www.waters.com/nextgen/global/products/chromatography/chromatography-systems/arc-premier-system.html>

720008669, January 2025

Lighting the Way for mRNA Analysis

Achieve up to 50% higher resolution for mRNA, LNPs, and viral vectors with wide pore SEC-MALS for biophysical characterization.

LEARN MORE
waters.com

Waters™

waters.com/RNA

waters.com/LNP

waters.com/RapiZymeRNases

Waters™

Waters Corporation
34 Maple Street
Milford, MA 01757 U.S.A.
T: 1 508 478 2000
F: 1 508 872 1990
waters.com