SOFTWARE TOOLS FOR AUTOMATED LC/MS ANALYSIS OF CRITICAL QUALITY ATTRIBUTES OF mRNA MOLECULES

Catalin Doneanu[1], Alexandre F. Gomes[1], Jo-Anne Riley[2], Chris Preston[2], Matt Gorton[2], Tatiana Johnston[1], Bala Addepalli[1], and Ying Qing Yu[1]

[1]Waters Corporation, Milford, MA, USA; [2]Waters Corporation, Wilmslow, UK

INTRODUCTION

- The use of novel software tools for assessing the critical quality attributes (CQAs) of mRNA molecules is described in this poster.
- Three CQAs were investigated: 1) 5'-Capping Efficiency; 2) Poly(A) Tail average length and heterogeneity; 3) mRNA sequence integrity.
- MS^E (DIA) data was collected on the Xevo™ MRT Q-Tof Mass Spectrometer, our newest high MS-resolution instrument using the multi-reflectron time-of-flight technology (MRT).
- The waters_connect™ SYNTHETIC Library v2.0, MAP Sequence v2.0 and INTACT Mass v1.9 applications (Apps), are specialized software tools developed to streamline the analysis of nucleic acid based therapeutics, such as synthetic oligonucleotides, sgRNA and mRNA. They facilitate oligonucleotide mapping and sequence confirmation analyses from sequence entry to report generation.
- Nearly complete mRNA sequence coverage (93.7%) was obtained after combining the sequencing results generated by a panel of four enzymatic digestions, underlining the utility of this approach for comprehensive sequence integrity verification.

Figure 1. Xevo MRT (multi-reflecting time-of-flight) Q-Tof Mass Spectrometer with the ACQUITY™ Premier UPLC System.

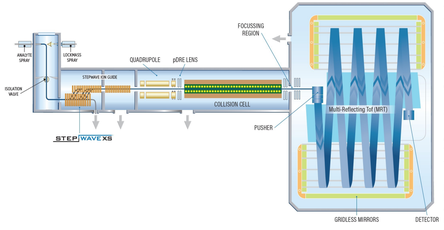


Figure 2. Schematic diagram of the Xevo MRT Q-Tof Mass Spectrometer.

METHODS

mRNA Sample
A custom mRNA construct based on the firefly luciferase (Fluc) sequence was custom-made via IVT (in vitro transcription) synthesis by GenScript (GS, Piscataway, NJ). The mRNA molecule was synthesized with a Cap-1 structure, followed by 1,919 nucleotides and a Poly(A) Tail sequence.

Endonuclease digestions
Chromatographically purified, animal free, ribonuclease T1 (RNase T1, catalogue no IFGRNASET1AFLY500KU, 500kU), isolated from Aspergillus oryzae, was purchased from Innovative Research (Novi, MI). The lyophilized enzyme was dissolved in 5 mL of 100 mM ammonium bicarbonate to prepare a solution containing 100 units/µL. 20 µL of Fluc mRNA (1.4 mg/mL) were mixed with 10 µL of nuclease-free water and incubated at 90°C for 2 min to achieve mRNA denaturation. After cooling the sample to room temperature (RT), 10 µL (1,000 units) of RNase T1 were added and the mRNA sample was digested for 15 min at 37°C.

The RapiZyme™ MC1 endonuclease (P/N 186011190) was dissolved in a buffer containing 200 mM ammonium acetate pH 8.0 at a concentration of 100 units/L. The RapiZyme Cusativin endonuclease (P/N 186011192) was dissolved in a buffer containing 200 mM ammonium acetate pH 9.0 at a concentration of 100 units/L. The detailed digestion protocols with RapiZyme MC1 and RapiZyme Cusativin are provided elsewhere [1-3].

The hRNAse4 enzyme (catalogue no M1284S, 2500 units) was purchased from New England Biolabs (Ipswich, MA). 20 µL of Fluc mRNA were denatured following the same procedure described for RNase T1 digestion, then 10 µL (500 units) of hRNase4 were added and the mRNA sample was digested for 90 min at 37°C.

All digestion mixtures were prepared in QuanRecovery™ MaxPeak™ 300 µL Vials. The digests were analyzed immediately by LC-MS using 5 µL injections.

The RNase T1 ribonuclease has 3'-guanosine specificity, cleaving after guanosine residues, regardless of the identity of the next nucleotide residue: G_A / G_C / G_G / G_U.
RapiZyme MC1 is a recombinant enzyme from the RNase T2 family, which display 5'-uridine specificity, cleaving at three primary cleavage sites: A_U / C_U / U_U / and two minor cleavage sites: C_A / C_G, for a total of 5 cleavage sites.
RapiZyme Cusativin is a recombinant enzyme from the same family, which display 3'-cytidine specificity, cleaving primarily at: C_A / C_G / C_U / with other four minor cleavage sites at: U_A / A_U, / G_U / U_U , for a total of 7 cleavage sites.
Human RNase4 (hRNAse4) is an endonuclease with 3'-uridine specificity, cleaving only at two cleavage sites: U_A and U_G.

LC-MS conditions
UHPLC System: ACQUITY Premier Binary System with TUV detector (260 nm)
Mass Spectrometer: Xevo MRT Q-Tof
Column: ACQUITY Premier OST Column 1.7 µm, 300 Å, 2.1 x 150 mm (P/N 186010541)

Mobile phases:

Eluent A: 10 mM DPA (n-dipropyl amine), 40 mM HFIP in deionized water, pH 8.6
Eluent B: 10 mM DPA, 40 mM HFIP in 50% methanol

Digest separations were performed at a flow rate of 0.3 mL/min with a gradient from 0% to 50% Solvent B in 90 min, at a column temperature of 60 °C.

Data-independent acquisitions (DIA) were performed in MS^E mode on a Xevo MRT Q-Tof Mass Spectrometer operated by waters_connect software. Data was acquired with 0.5 s scans over a mass range of 400-4000 Da. Low-energy MS^E scans were acquired with a CE (collision energy) of 6 V, while the high-energy fragmentation scans used CE ramping from 30 to 55 V.

Informatics:

- waters_connect Informatics Platform 4.1.0.17
- SYNTHETIC Library App 2.0.0
- MAP Sequence App 2.0.0
- INTACT Mass App v1.9

All three mRNA CQAs (Capping Efficiency, Average Length and Heterogeneity of the Poly(A) Tail, and Sequence Integrity) were measured following a single LC-MS assay optimized to provide all the required information.

For the 5'-Capping Efficiency assay, the MS^E data was processed with the MAP Sequence App after entering the two Cap-1 structures in the SYNTHETIC Library App.

For PolyA) Tail analysis, the MS^E data was processed with the INTACT Mass App to deconvolve and measure the ten different polyadenosine species detected ranging from a 98-mer - $C(A)_{97}$ to a 107-mer - $C(A)_{106}$. The Average Poly(A) Tail length is 102.3 -mer.

In-silico digestion products were computed for Fluc mRNA digested with RNase T1, hRNase4, RapiZyme MC1 and RapiZyme Cusativin using the SYNTHETIC Library App based on the corresponding nucleotide sequences entered in the library. The library can accommodate built-in oligonucleotide modifications as well as custom-editable modifications. Following data processing using the MAP Sequence App, digestion products are automatically assigned using the low-energy MS^E data channel (oligonucleotide precursors), as well as the elevated energy MS^E data channels (oligonucleotide fragments).

RESULTS

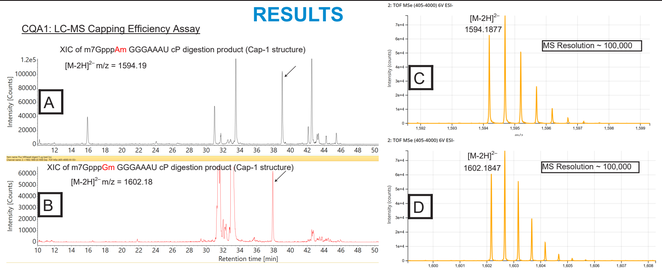


Figure 3. 5'-Cap analysis of the GenScript Fluc mRNA via a single LC/MS assay:

(A,B) Extracted ion chromatograms corresponding to the doubly charged digestion products containing two Cap-1 structures, m7GpppAm (panel A) and m7GpppGm (panel B), attached to the same 7-mer hRNase4 digestion product with the sequence GGG AAA U cP;

(C,D) MAP Sequence isotopic distributions of the doubly charged precursors corresponding to the 9-mer Cap-1 digestion products generated following hRNase4 digestion. Both Cap-1 structures were detected in the hRNAse4 digest, but the free digestion product (GGG AAA U cP, m/z= 1163.13, -2) was not detected (data not shown). In addition, no Cap-related impurities (truncations of Cap-0/Cap-1/Cap-2 structures) were detected upon processing the data with the Map Sequence App.

Based on these results, the GenScript Fluc mRNA Capping Efficiency is 100%, with 62.7% of mRNA bearing an m7GpppAm Cap-1 and 37.3% of mRNA having an m7GpppGm Cap-1 structure. This is an unexpected result, as typical 5'-mRNA Caps have either a methylated adenosine-Am or a methylated guanosine-Gm Cap-1 structure, *but not both*.

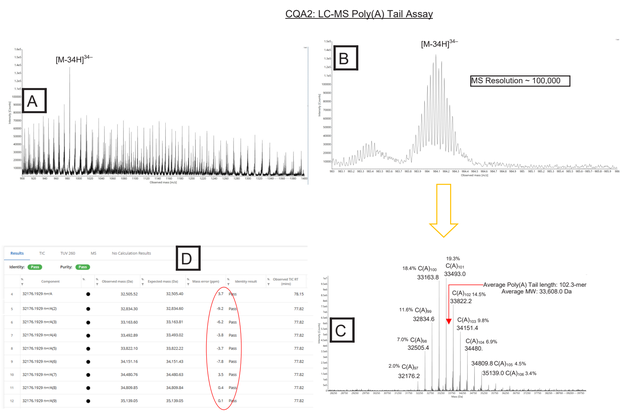
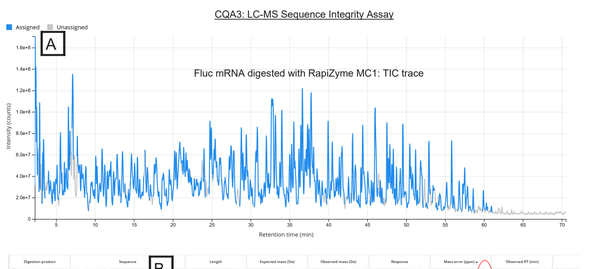
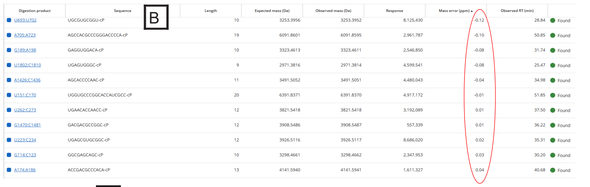
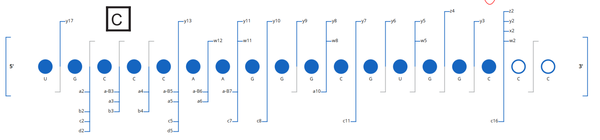


Figure 4. Poly(A) Tail analysis of the GenScript Fluc mRNA via a single LC/MS assay:

(A) combined ESI-MS spectrum of all the polyadenosine variants detected;
(B) isotopically resolved [M-34H]$^{34-}$ charge state of the most abundant 102-mer Poly(A) Tail oligonucleotide with the sequence $C(A)_{101}$, resulted from RNase T1 cleavage of GS Fluc mRNA.
(C) deconvolved ESI-MS spectrum showing the ten Poly(A) Tail variants detected along with their relative abundances calculated based on the ESI-MS signal intensity. The calculated average Poly(A) Tail length is 102.3-mer after taking into account the relative abundances of each Poly(A) species detected above 2% relative abundance.
(D) screenshot showing the INTACT Mass App processing results. Very good mass accuracy (< 10 ppm) was obtained on the Xevo MRT QTof for the measurement of all ten detected Poly(A) Tail species.







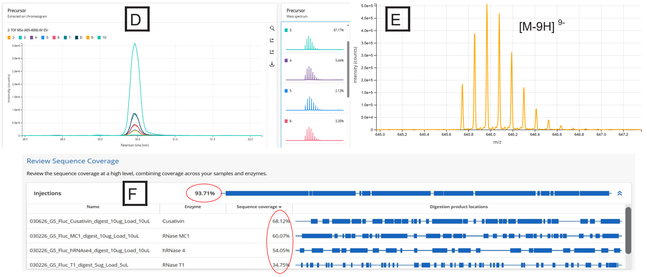


Figure 5. MAP Sequence App processing results: (A) TIC chromatogram of the Fluc mRNA digest, indicating that the majority of the ESI-MS response belongs to matched/assigned digestion product precursors, highlighted by the blue trace. The unassigned chromatographic peaks/regions are displayed in gray in the same chromatogram. (B) Portion of the digestion product table generated for the GS Fluc mRNA digested with RapiZyme MC1. The MS^E dataset was search with a 5 ppm mass tolerance for digestion product precursors and fragments. The typical mass accuracy for the Xevo MRT instrument for measuring both precursors and fragments is in the range of –1 to +1 ppm (sub-ppm range) as highlighted in this table. (C) Dot-map diagram showing the elevated energy (MS^E) fragmentation of a large (18-mer) MC1 digestion product U658:C675 containing two missed cleavages. (D) Overlaid XICs for all the charge states (from -2 to -10) detected for the 18-mer digestion product. (E) Resolved isotopic distribution of the [M-9H]$^{9-}$ precursor of U658:C675 nucleotide. (F) Comparison of *unambiguous* sequence coverage obtained for digestion of the GS Fluc mRNA with four enzymes - RapiZyme MC1, RapiZyme Cusativin, hRNase4 and RNase T1 - assuming up to four missed cleavages in each case. The *combined sequence coverage* demonstrates that the sequence coverage can be significantly increased through the *multi-enzyme digestion approach* outlined in this workflow. It is critical to generate unambiguous, unique products through enzymatic digestion, to reduce reliance on user intervention with MS2 fragmentation data to resolve sequence assignment ambiguities.

Automatic elucidation of structural isomers with the MAP Sequence App

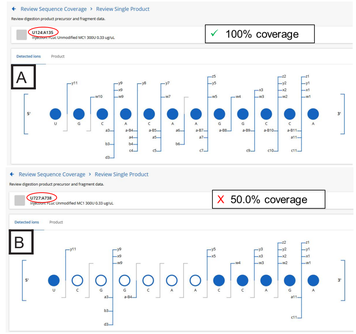


Figure 6. An example of automatic elucidation of structural isomers using user-specified acceptance criteria: (A) the 12-mer U124:A135 digestion product displays full (100%) sequence coverage at the fragment ion level, therefore this digestion product is confirmed; (B) its isomeric counterpart, the 12-mer U727:A738 digestion product has only 50.0% fragment ion coverage, therefore this assignment is rejected.

CONCLUSIONS

- MS^E data acquisition on the Xevo MRT Q-Tof Mass Spectrometer provides excellent mass resolution, mass accuracy and sensitivity for analysis of mRNA CQAs.
- SYNTHETIC Library, MAP Sequence and INTACT Mass Apps deliver excellent usability for assessing the mRNA CQAs.
- The Capping Efficiency of GS Fluc mRNA was 100%, based on two 5'-Cap1 structures detected.
- The Average Poly(A) Tail length calculated for the GS Fluc mRNA is 102.3-mer, after taking into account the ten most abundant species.
- Near complete unambiguous mRNA sequence coverage (93.7%) was obtained through the use of the *combined digestion approach*, using four different endonucleases.
- Structural isomers of digestion products were automatically differentiated by the MAP Sequence App 2.0 algorithm after using the elevated energy fragment ions to differentiate isobaric and isomeric digested oligonucleotides.

References

1. Tunable Digestion of RNA Using RapiZyme RNases to Confirm Sequence and Map Modifications, 2024, Waters application note P/N 720008539EN.
2. RNA Digestion Product Mapping Using an Integrated UPLC-MS and Informatics Workflow, 2024, Waters application note P/N 720008553EN.
3. Sequence Mapping of mRNA Digests Using the Xevo MRT Mass Spectrometer and waters_connect MAP Sequence 2.0 Application, 2025, Waters application note P/N 720009171EN.

720009414EN