MAP SEQUENCE - A NOVEL SOFTWARE APP FOR RAPID AND AUTOMATED SEQUENCE MAPPING OF mRNA MOLECULES

Catalin Doneanu [1], Alexandre F Gomes [1], Tatiana Johnston [1], Chris Preston [2], Matt Gorton [2], Bala Addepalli [1], Christian Reidy [1], Guillaume Bechade [2] and Ying Qing Yu [1]

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

INTRODUCTION

- The use of a novel software App—MAP Sequence for sequence mapping of messenger RNAs (mRNAs) is described in this poster.
- MS^E data (data independent acquisition—DIA) was collected on the newest high MS-resolution QTof Mass Spectrometer from Waters, using the multi-reflectron time-of-flight technology - Xevo™ MRT.
- The waters_connect™ SYNTHETIC Library v2.0 and MAP Sequence v2.0 applications (Apps), are specialized software tools developed to streamline the analysis of nucleic acid based therapeutics, such as synthetic oligonucleotides, sgRNAs and mRNAs. They facilitate oligonucleotide mapping and sequence confirmation analyses from sequence entry through to report generation.
- Near complete mRNA sequence coverage (95.4%) was obtained through the use of two recently introduced endonucleases [1-5], RapiZyme™ MC1 and RapiZyme™ Cusativin, which offer unique cleavage specificity and opportunity to generate overlapping digestion products
- Combining the results from a panel of enzymatic digestions improves overall confidence in the accuracy of the sequence mapping approach, opening opportunities for complete mRNA coverage.

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

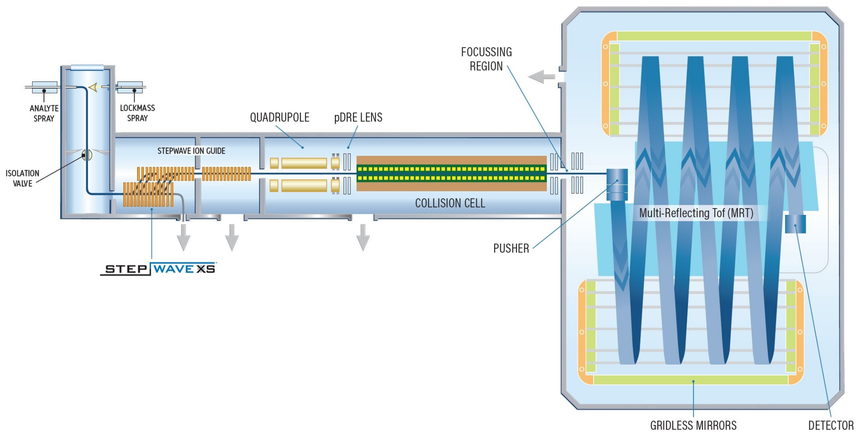
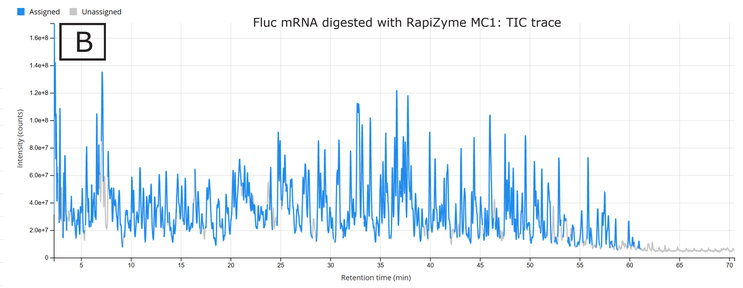


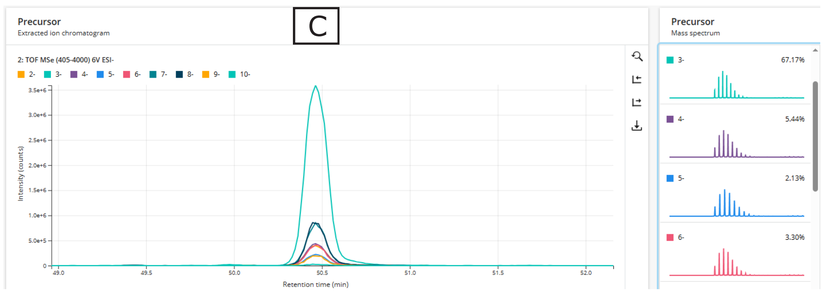
Figure 2. Schematic diagram of the Xevo™ MRT QTof Mass Spectrometer.

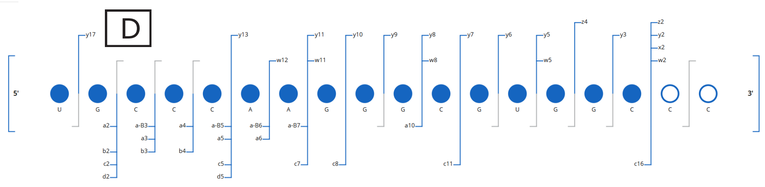
RESULTS

A

Digestion product	Sequence	Length	Expected mass (Da)	Observed mass (Da)	Response	Mass error (ppm)	Observed RT (min)	
U693:U702	UGCGUGCGGU-cP	10	3253.3956	3253.3952	8,125,430	-0.12	28.84	Found
A705:A723	AGCCACGCCCGGGACCCCA-cP	19	6091.8601	6091.8595	2,961,787	-0.10	50.85	Found
G189:A198	GAGGUGGACA-cP	10	3323.4613	3323.4611	2,546,850	-0.08	31.74	Found
U1802:C1810	UGAGUGGGC-cP	9	2971.3816	2971.3814	4,599,541	-0.08	25.47	Found
A1426:C1436	AGCACCCCAAC-cP	11	3491.5052	3491.5051	4,480,043	-0.04	34.98	Found
U151:C170	UGGUGCCCGGCACCAUCGCC-cP	20	6391.8371	6391.8370	4,917,172	-0.01	51.85	Found
U262:C273	UGAACACCAACC-cP	12	3821.5418	3821.5418	3,192,089	0.01	37.50	Found
G1470:C1481	GACGACGCCGGC-cP	12	3908.5486	3908.5487	557,339	0.01	36.22	Found
U223:C234	UGAGCGUGCGGC-cP	12	3926.5116	3926.5117	8,686,020	0.02	35.31	Found
G114:C123	GGCGAGCAGC-cP	10	3298.4661	3298.4662	2,347,953	0.03	30.20	Found
A174:A186	ACCGACGCCCACA-cP	13	4141.5940	4141.5941	1,611,327	0.04	40.68	Found







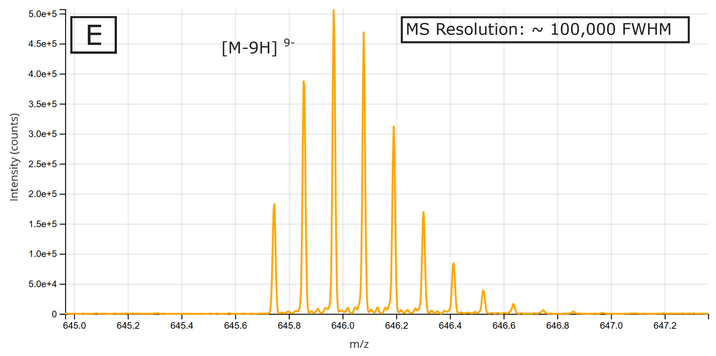


Figure 3. MAP Sequence processing results:

(A) Portion of the digestion product table generated for the 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.
(B) 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.
(C) Overlaid XICs for all the charge states (from -2 to -10) detected for a large(18-mer) MC1 digestion product U658:C675, containing two missed cleavages.
(D) Dot-map diagram showing the elevated energy (MS^E) fragmentation of the same 18-mer digestion product.
(E) Resolved isotopic distribution of the $[M-9H]^{9-}$ precursor of the U658:C675 digestion product.
(F) Comparison of sequence coverage obtained for digestion of the Fluc mRNA with three enzymes - RapiZyme MC1, RapiZyme Cusativin and RNase T1 - assuming up to four missed cleavages. 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 assignment ambiguities. RapiZyme MC1 and RapiZyme Cusativin enzymes, can produce longer unique digestion products compared to RNase T1, through a combination of unique cleavage specificity and intentionally generated missed cleavages by regulating enzyme amounts and digestion time [1-5]. In this way, RapiZyme MC1 and Cusativin (RNase T2 enzymes) achieve increased sequence coverage compared to RNase T1 [1,2].

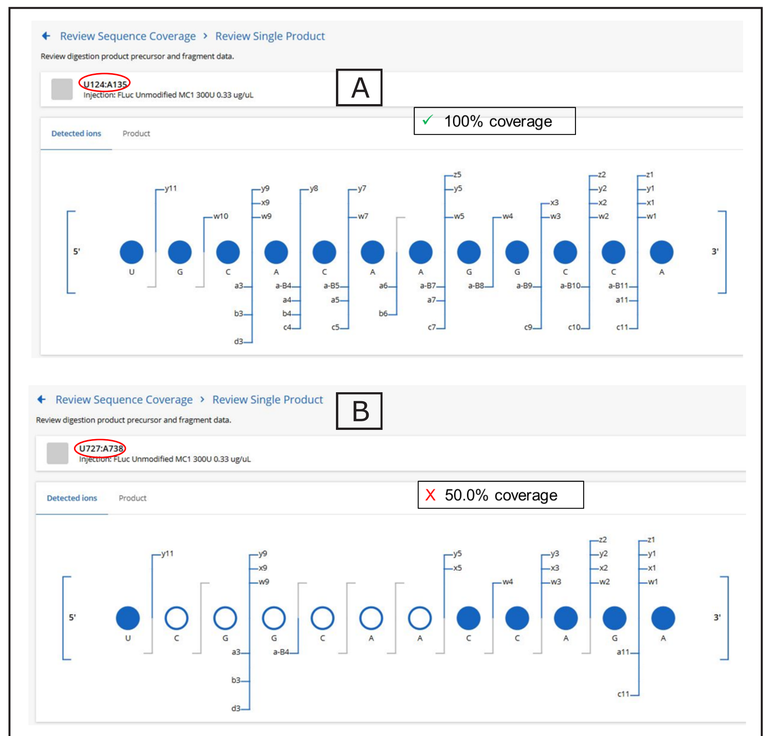
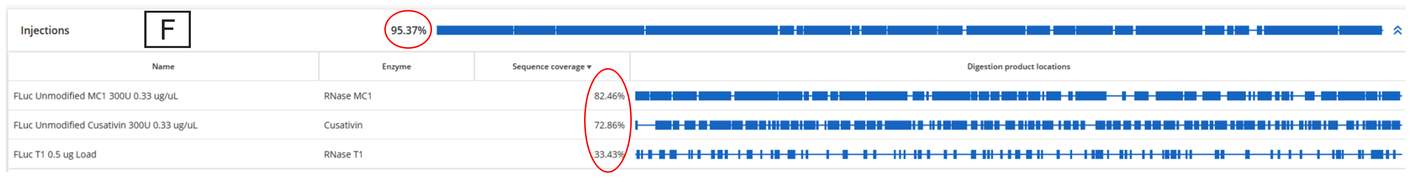


Figure 4. An example of automatic elucidation of structural isomers using user-specified acceptance criteria: (A) the 12-mer U124:A135 digestion product displays full (100%) sequence coverage at the fragment ion level, therefore this digestion product is confirmed; (B) its isomeric counterpart, the 12-mer U727:A738 digestion product has only 50.0% fragment ion coverage, therefore this assignment is rejected.



METHODS

Experimental Section

mRNA Sample
A custom mRNA construct based on the firefly luciferase (Fluc) sequence was custom-made via IVT (in vitro transcription) synthesis by GenScript (Piscataway, NJ). The mRNA molecule was synthesized with a Cap1 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.
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 RNase T1, RapiZyme MC1 and RapiZyme Cusativin are provided elsewhere [1-2]. All digestion mixtures were prepared in QuanRecovery™ MaxPeak™ 300 µL Vials. The digests were analyzed immediately by LC-MS using 5 µL injections.
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.

LC-MS experimental conditions
An ACQUITY™ Premier UPLC™ BSM System equipped with a 2.1 x 150 mm ACQUITY™ Premier OST Column (P/N 186010541) was used for all oligonucleotide separations. A IP-RP mobile phase containing 10 mM DPD (dipropylamine) and 40 mM HFIP was used as Solvent A, while the composition of Solvent B was 10 mM DPA, 40 mM HFIP in 50% methanol. Gradient separations were performed at a flow rate of 0.4 mL/min, from 0% B to 50% B over 60 min, at a column temperature of 60°C. UV data was acquired at a fixed wavelength of 260 nm.

Data-independent acquisitions (DIA) were performed in MS^E mode on a Xevo MRT 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

In-silico digestion products were computed for the mRNA digested with RNase T1, RapiZyme MC1 and RapiZyme Cusativin using the SYNTHETIC Library App based on the corresponding Fluc mRNA nucleotide sequence 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). For mRNA digests, the presence of structural isomers, can produce ambiguous assignments for the digestion products. In these situations, the MAP Sequence App can be used to elucidate the correct sequence as shown in the example displayed in Figure 4.

CONCLUSIONS

- MS^E data acquisition on the new Xevo MRT QTof Mass Spectrometer provides excellent mass resolution (> 100,000 FWMH), mass accuracy (< 1ppm) and sensitivity for analysis of mRNA digests
- MAP Sequence and SYNTHETIC Library Apps deliver excellent usability for mRNA sequence mapping
- Near complete mRNA sequence coverage (95.4%) was obtained through the use of RapiZyme MC1 and RapiZyme Cusativin, two new endonucleases recently introduced by Waters [1-4]
- 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
- The Apps employed for UHPLC-MS data acquisition and data processing on the compliance-ready waters_connect informatics platform, enable the potential use of these workflows in manufacturing and quality functions

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. 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
4. 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.
5. 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.

720009345EN