RNA SEQUENCE MAPPING VIA ENDONUCLEASE DIGESTION AND LC-MS ANALYSIS VIA NOVEL INFORMATICS WORKFLOWS

Catalin Doneanu[1], Chris Preston[2], Matt Gorton[2], Alexandre F. Gomes[1], Tatiana Johnston[1], Bala Addepalli[1], Nick Pittman[2] and Ying Qing Yu[1]

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

INTRODUCTION

- The use of a novel software for mapping analysis of single guide RNAs (sgRNA) and mRNAs is described in this poster.
- MS^E (DIA) data was collected on the newest high MS-resolution QTof Mass Spectrometer from Waters using the multi-reflectron time-of-flight technology.
- The waters_connect™ SYNTHETIC Library, MAP Sequence and CONFIRM Sequence 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 through to report generation.
- Complete sgRNA sequence coverage (100%) was obtained through the use of RapiZyme™ MC1, a new endonuclease recently introduced by Waters [1-4], which offers 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 mass fingerprinting-based approach, and opportunities for higher or complete RNA coverage.

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

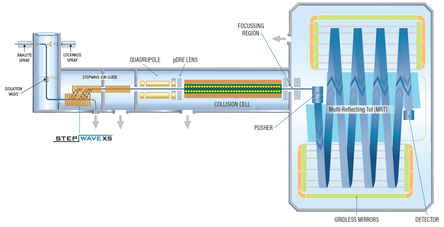


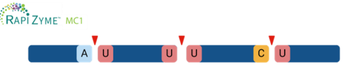
Figure 2. Schematic diagram of the Xevo™ MRT QTof Mass Spectrometer.

METHODS

sgRNA Sample
A 100-mer single-guide RNA (sgRNA), encoding for the mouse GATA2 transcription factor, is a LC-MS oligonucleotide standard recently introduced by Waters (P/N 186011357) with the sequence: 5' - C*U*U* CAA CCA UCU CGA CUC GCG 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 Waters sgRNA LC-MS standard contains a 2'-OMe modification on its first three 5' nucleotides (C*U*U*, denoted as the XYY sequence in the screenshot from Figure 5), as well as on its last three 3' nucleotides (U*U*U*, sequence YYY in Figure 5) and the asterisk indicates that all these six nucleotides are phosphorothioates.

mRNA Sample
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, followed by 1019 nucleotides and no Poly (A) Tail sequence (a tailless mRNA).

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. The detailed digestion protocols with RNase T1 and RapiZyme MC1 are provided elsewhere [1-2]. Both 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:



In addition, MC1 has two minor cleavage sites, with a total of 5 dinucleotide cleavage motifs: A_U / C_U / U_U / C_A / C_G.

LC-MS conditions
UHPLC System: ACQUITY Premier System, Binary with TUV detection
Mass Spectrometer: Xevo MRT QTof MS
Column: ACQUITY Premier OST Column 1.7 µm, 130 Å, 2.1 x 150 mm (P/N 186009486)

Mobile phases:

A: 10 mM DPA (n-dipropyl amine), 40 mM HFIP in deionized water, pH 8.6
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 45 min, at a column temperature of 60 °C. Solvent B.

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 340-4000 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 (in development)
- MAP Sequence App 2.0.0 (in development)
- CONFIRM Sequence App 1.4.0.13

In-silico digestion products as well as the predicted sequence coverages were computed for the Waters sgRNA and GFP mRNA digested with RNase T1 and MC1 (see Figure 6) using the SYNTHETIC Library App. The library can accommodate built-in oligonucleotide modifications as well as custom-editable modifications. The RNase T1 and MC1 digestion products were automatically assigned by the MAP Sequence App using the MS1 data channel (oligonucleotide precursors). Confirmation of these assignment using the elevated energy MS^E fragmentation information is currently under development.

For mRNA digests, the presence of structural isomers, can produce ambiguous assignments for the digestion products. In these situations, the CONFIRM Sequence App [5] can be used to elucidate the correct sequence as shown in the example displayed in Figure 8.

RESULTS

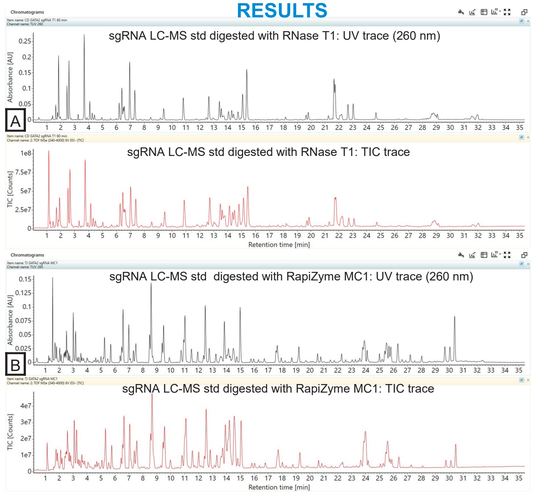


Figure 3. UV and TIC chromatograms recorded for the Waters single guide sgRNA oligonucleotide standard (100-mer) digested with: (A) RNase T1 (Innovative Research); (B) RapiZyme MC1 (Waters).

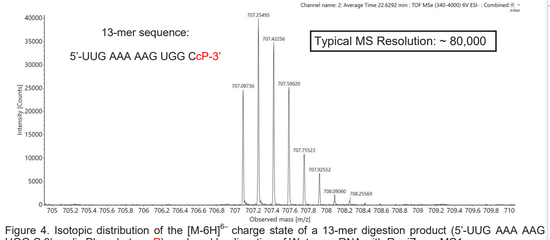


Figure 4. Isotopic distribution of the $[M-6H]^{6-}$ charge state of a 13-mer digestion product (5'-UUG AAA AAG UGG C 3'-cyclic Phosphate - cP) produced by digestion of Waters sgRNA with RapiZyme MC1.

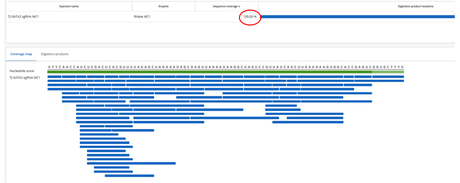


Figure 5. Coverage map generated by the MAP Sequence App for the Waters sgRNA LC-MS standard (100-mer oligo) digested with RapiZyme MC1. Full sequence coverage (100%) was confirmed following the assignment of multiple digestion products precursors originating from the same region of the sgRNA sequence, as highlighted by the blue rectangle bars.

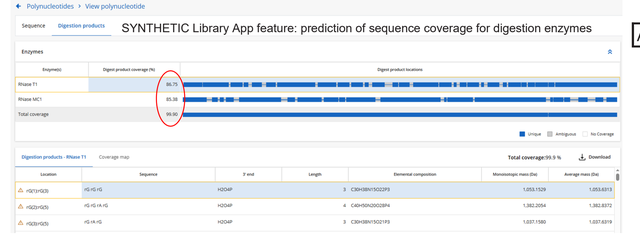


Figure 6. SYNTHETIC Library screenshot displaying the predicted sequence coverage generated for the digestion of GFP mRNA (1019 nucleotides) with two endonucleases: RNase T1 (86.75% predicted coverage) and Rapizyme MC1 (85.38%) assuming up to 2 missed-cleavages for each enzyme. There is unique sequence coverage for each enzyme, therefore the total coverage that could be potentially obtained by combining the individual enzyme coverages is 99.9%. Using multiple enzymes for individual digestion of the same mRNA substrate could significantly increase the sequence coverage.

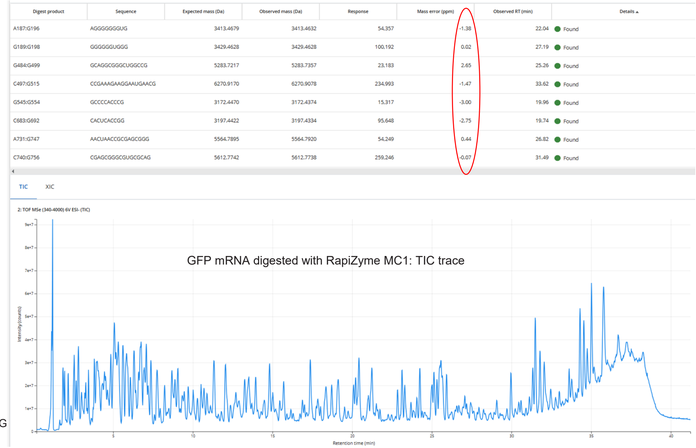
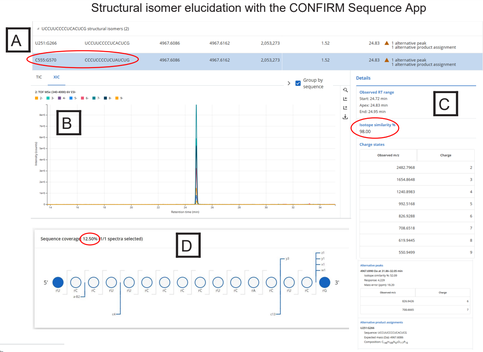


Figure 7. MAP Sequence results generated by the MAP Sequence App for the GFP mRNA (1019 nucleotides) digested with RapiZyme MC1. High sequence coverage (>85%) was obtained following the assignment of digestion products precursors. The MS^E dataset was search with a 15 ppm mass tolerance for the digestion product precursors. Typical mass accuracy for the Xevo MRT instrument is in the range of -3 to +3 ppm. 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 [1-4]. In this way, RapiZyme MC1 and Cusativin (RNase T2 enzymes) achieve increased sequence coverage compared to RNase T1 [1,2].



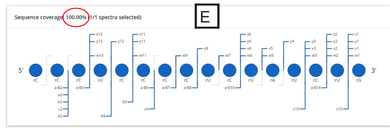


Figure 8. (A), (B), (C) are screen-captures from the MAP Sequence App: (A) pair of structural isomers identified by MAP Sequence as *alternative product assignments*; (B) overlaid extracted mass chromatograms of 8 oligonucleotide precursors that could be assigned to isomeric digestion products U251:G266 or C555:G570 resulting from RNase T1 cleavage of the GFP mRNA; (C) side panel displaying the charge states (2-9) detected by MAP Sequence for the ESI -MS spectrum of this digestion product.
(D) and (E) are CONFIRM Sequence App screenshots showing the dot-map sequence coverage obtained for two possible 16-mer digestion products. The chromatographic peak detected at 24.8 min, shown in panel (B), was produced by the C555:G270 16-mer oligonucleotide (CCC UCC CCU CUA UCU G), given the 100% sequence coverage displayed in panel (E). Only a few terminal fragments ions could be assigned to the alternative isomeric sequence (UCC UUC CCC UCA CUC G) - panel (D).

CONCLUSIONS

- MS^E data acquisition on the new Xevo MRT QTof Mass Spectrometer provides improved mass resolution, mass accuracy and sensitivity for analysis of RNA digests
- MAP Sequence and SYNTHETIC Library Apps deliver improved usability for mRNA sequence mapping
- Complete sgRNA sequence coverage (100%) was obtained through the use RapiZyme MC1, a new endonuclease recently introduced by Waters [1-4]
- 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. CONFIRM Sequence: A waters_connect Application for Sequencing of Synthetic Oligonucleotide and Their Impurities , 2022, Waters application note P/N 720007677EN.

720008892EN