sgRNA AND mRNA SEQUENCE MAPPING VIA ENDONUCLEASE DIGESTION AND LC-MS ANALYSIS, USING NOVEL INFORMATICS WORKFLOWS

Catalin Doneanu[1], Alexandre F. Gomes[1], Chris Preston[2], Matt Gorton[2], Tatiana Johnson[1], Bala Addepalli[1], Heidi Gastall[2] and Ying Qing Yu[1]
[1]Waters Corporation, Milford, MA, USA; [2]Waters Corporation, Wilmslow, UK

INTRODUCTION

- The use of an end-to-end workflow for mapping analysis and sequence confirmation of single guide RNAs (sgRNA) and mRNAs is described.
- LC-MS^E data were collected on the ACQUITY™ Premier System coupled with the newest high MS-resolution QTOF Mass Spectrometer from Waters – the Xevo™ MRT MS, which uses the multi-reflectron time-of-flight technology.
- The waters_connect™ SYNTHETIC Library and MAP Sequence applications (Apps), are specialized software tools developed to streamline the mapping and sequence confirmation of nucleic acid based therapeutics, such as sgRNA and mRNA.
- Complete sgRNA sequence coverage (100%) and 98% mRNA coverage were obtained using RapiZyme™ MC1 and RapiZyme™ Cusativin, new endonucleases recently introduced by Waters [1-4], which offer unique cleavage specificity and the opportunity to generate overlapping digestion products.

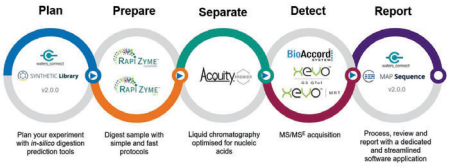

Figure 1. *End-to-end workflow to map RNA sequences, enabled by the RapiZyme Endonucleases, waters_connect SYNTHETIC Library and MAP Sequence applications, and high-resolution LC-MS systems such as the Xevo MRT QTOF Mass Spectrometer coupled with the ACQUITY Premier System.*

Figure 2. *Xevo MRT (multi-reflecting time-of-flight) QTOF Mass Spectrometer with the ACQUITY Premier System.*

METHODS

sgRNA Sample
A 100-mer single-guide RNA (sgRNA), encoding for the HPRT1 human enzyme (hypoxanthine phosphoribosyl-transferases 1), is an 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 XYY sequence in the screenshot from Figure 4), as well as on its last three 3' nucleotides (U*U*U*, sequence YYY in Figure 4) and the asterisk indicates that all these six nucleotides are phosphorothioates.

mRNA Sample
An 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 1813 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 (P/N 186011190) endonuclease was dissolved in a buffer containing 200 mM ammonium acetate pH 8.0, while the RapiZyme Cusativin (P/N 186011192) was dissolved in 200 mM ammonium acetate pH 9.0, to prepare stock solutions with a concentration of 100 units/µL. The detailed digestion protocols with RNase T1 and RapiZyme MC1 and Cusativin 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 and RapiZyme Cusativin are recombinant enzymes from the RNase T2 family, with very different cleavage patterns to the G-based specificity of RNase T1. MC1 cleaves at 5' side of uridine residues, with three primary cleavage sites at [A/U/C]U and secondary cleavages at Cp[A/G]. Cusativin has 3' cytidine specificity cleaving primarily at Cp[A/U/G] and UpA motifs, with secondary cleavages at [A/G]pU and UpU. Their unique specificity combined with miss-cleavage properties, make these two enzymes generate digestion products with longer sequences, which are more likely to have a unique mass.

Figure 3. *Cleavage sites of the RapiZyme MC1 and Cusativin digestion enzymes.*

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, 300 Å, 2.1 x 150 mm (P/N 186010541)
Mobile phases:
A: 0.1% DIPEA (N,N-diisopropylethylamine), 1% HFIP in deionized water, pH 8.5
B: 0.0375% DIPEA, 0.75% HFIP in 65% acetonitrile

Digest separations were performed at a flow rate of 0.4 mL/min with a gradient from 0% to 50% Solvent B in 60 min, at a column temperature of 70°C.
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 m/z. 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 45 to 55 V.

Informatics
- SYNTHETIC Library App 2.0.0
- MAP Sequence App 2.0.0
- waters_connect Informatics Platform 4.1.0

RESULTS

This poster presents a new generation of LC-MS workflow purposefully designed for the characterisation and quality attribute monitoring of nucleic-acid based therapeutics and vaccines (Figure 1).

As illustrated in Figure 4, unique advanced digestion-strategy planning tools within the SYNTHETIC Library app allow scientists to optimise their experimental design. These tools assist in selecting the most effective digestion strategy to maximize sequence coverage while minimizing ambiguous digestion products.

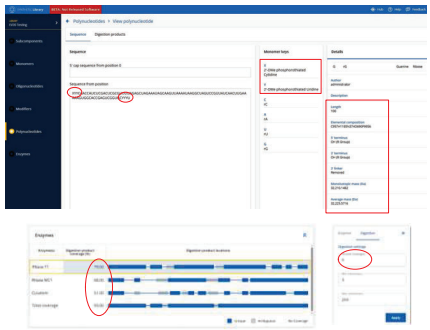

Figure 4. *SYNTHETIC Library screenshot displaying (A) entry of the sequence of the Waters sgRNA standard, including the modified 3' and 5' nucleotides. Theoretical characteristics of the sequence are automatically calculated. (B) The predicted sequence coverage generated with three endonucleases: RNase T1, RapiZyme MC1 and RapiZyme Cusativin assuming no missed-cleavages for each enzyme. Using multiple enzymes for individual digestion of the same sgRNA substrate could significantly increase the sequence coverage.*

The TIC chromatogram resulting from the digestion of Waters sgRNA with RapiZyme MC1 is shown Figure 5A. Assigned digestion fragments are clearly highlighted in blue, facilitating rapid assessment of data quality and the suitability of the processing method.

A total of 77 theoretically predicted digestion products were matched to the experimental data within the method-defined tolerances: 5 ppm mass error for precursors and fragments, 70% isotopic similarity, and a minimum of 50% coverage per digestion product (Figure 5B).

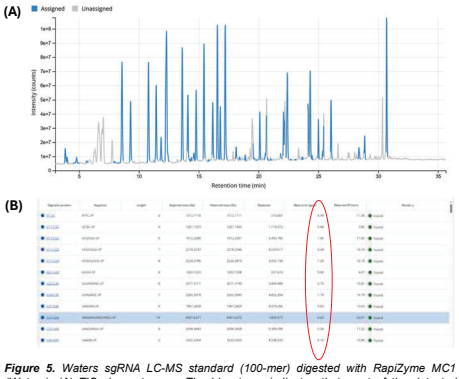

Figure 5. *Waters sgRNA LC-MS standard (100-mer) digested with RapiZyme MC1 (Waters). (A) TIC chromatogram. The blue trace indicates that most of the detected chromatographic peaks were assigned to predicted digestion products. (B) Results table excerpt from the MAP Sequence App. Unique digestion products (indicated by green circles) were confidently identified by matching the experimentally-measured monoisotopic masses (MS1) with the in-silico predicted monoisotopic masses. The sequences of these unique digestion products were further validated through fragment ion assignments. Additional details for the highlighted 14-mer digestion product (blue) are provided in Figure 6. The average mass accuracy for all assigned MC1 digestion products is <1 ppm.*

sgRNA and mRNA can produce multiple digest products which have the same mass (isobaric), the same elemental composition (isomeric), or the same sequence. This can make the assignment of the measured masses ambiguous. Therefore, the oligonucleotide digest products were fragmented in the mass spectrometer (LC-MS^E data-independent MS2 fragmentation) to produce higher confidence assignment and resolve isomeric and isobaric ambiguities (Figure 6).

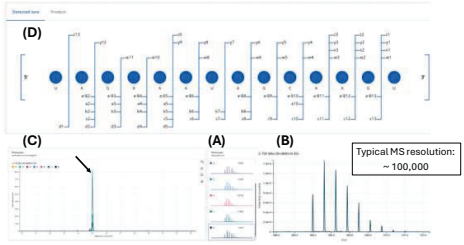

Figure 6. *Detailed analysis of the U31:U44 MC1 digestion product (14-mer) provided by the MAP Sequence App: (A) Isotopic envelopes for all seven detected charge states (from –2 to –8); (B) zoomed view of the [M–8H]$^{8-}$ charge state; (C) overlaid XICs (extracted ion chromatograms) for all charge states - this digestion product elutes at 25.0 min; (D) dot-map fragmentation diagram displaying all assigned fragment ions, demonstrating that full (100%) sequence coverage was achieved following elevated-energy MS^E fragmentation. The Xevo MRT MS instrument provided a resolution of 100,000, fully resolving the isotopic distribution of the [M–8H]$^{8-}$ charge state.*

Using complementary digestion enzymes, 100% sequence coverage of the sgRNA was achieved, with automated confirmation of both linear sequence and modifications (Figure 7A). These results were compiled into a comprehensive, automatically-generated report (Figure 7B).

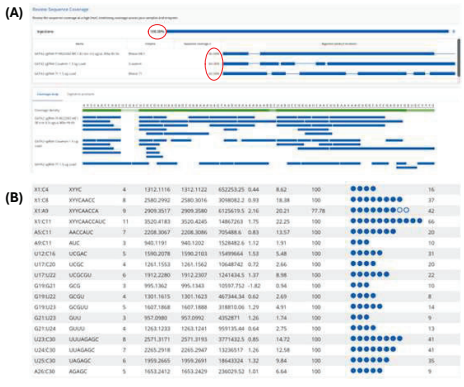

Figure 7. *(A) Comparison of sequence coverage of the Waters sgRNA obtained with three enzymes - RapiZyme MC1, RapiZyme Cusativin, and RNase T1 - assuming up to four missed cleavages. The combined sequence coverage demonstrates that complete sequence characterization (100%) of the Waters sgRNA can be achieved through a multi-enzyme digestion approach. The lower panel shows which digestion products were experimentally verified for each sample, and how they map to the sequence. (B) Excerpt from the MAP Sequence report showing results from the RapiZyme Cusativin digestion, confirming that all digested product assignments were extensively verified through corresponding fragment ion data.*

The same workflow was successfully applied to longer nucleic acid molecules such as mRNA. The TIC chromatogram of FLuc mRNA digested with RapiZyme MC1 is shown in Figure 8A. Despite the chromatogram's complexity, a significant proportion of peaks were confidently identified and assigned to theoretical digestion products (blue trace). With a combination of three enzymes, the linear sequence was confidently confirmed with 98% coverage (Figure 8B).

Figure 8. *(A) TIC chromatogram of FLuc mRNA (~2000 nt) digested with RapiZyme MC1 (Waters). Most of the detected chromatographic peaks were assigned to predicted digestion products (blue). (B) Sequence coverage of FLuc mRNA when digested with RNase T1, RapiZyme Cusativin and RapiZyme MC1. Each enzyme brings unique contributions to the overall sequence coverage (lower panel), producing a combined sequence coverage of 98%.*

CONCLUSIONS

- LC-MS mapping of long-chain RNA sequences has been streamlined through the development of new endonucleases; excellent UPLC™ Separation Technology; high-performing mass spectrometers; and bespoke RNA mapping software. These tools facilitate the workflow from beginning to end.
- RapiZyme MC1 and RapiZyme Cusativin endonucleases [1-4] generate complementary digest products to RNase T1. They were used in a multi-enzyme approach to produce complete sequence coverage (100%) for a 100-mer sgRNA, and 98% sequence coverage for FLuc mRNA (~2000 nt).
- MSE data acquisition on the new Xevo MRT QTOF Mass Spectrometer provides improved mass resolution, mass accuracy and sensitivity for analysis of RNA digests. Combined with the ACQUITY Premier Column and LC System, whose performance for oligonucleotide separation is well established [5], it produced comprehensive data to map the sgRNA and mRNA.
- MAP Sequence and SYNTHETIC Library Apps enable users to plan their digestion strategies; and process, review and report their results. These dedicated applications have multiple features to support mapping experiments, including combining coverage from multiple samples, and handling ambiguous digestion products.
- The Apps and LC-MS instrumentation are upheld by the compliance-ready waters_connect Informatics Platform, enabling 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. Synthetic mRNA Oligo-Mapping Using Ion-Pairing Liquid Chromatography and Mass Spectrometry, 2022, Waters application note P/N 720007669.

720009329EN

Xevo, ACQUITY, UPLC, RapiZyme, QuanRecovery, MaxPeak, Waters, BioAccord and waters_connect are trademarks of Waters Technologies Corporation.
©2024 Waters Corporation