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 °C 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 $^{\circ}$C
Desolvation temperature:	550 $^{\circ}$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.
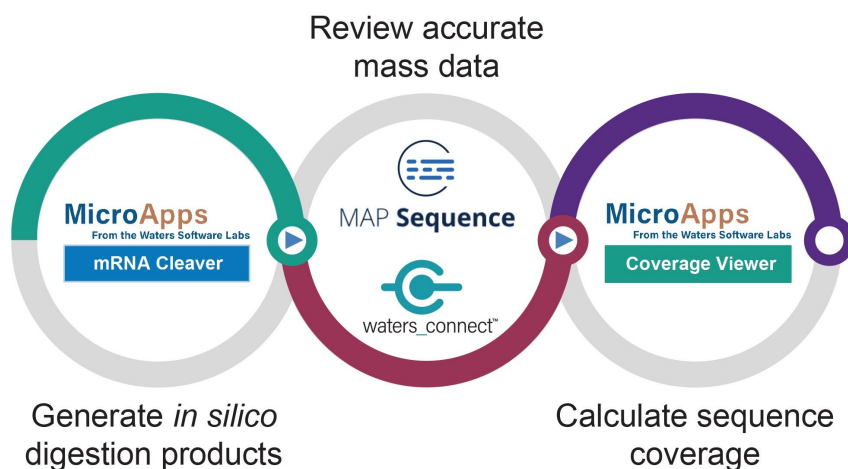
Figure 1. Simple diagram illustrating the workflow introduced for LC-MSE processing of sgRNA/mRNA digests using the mRNA Cleaver MicroApp, MAP Sequence App, and Coverage Viewer MicroApp.

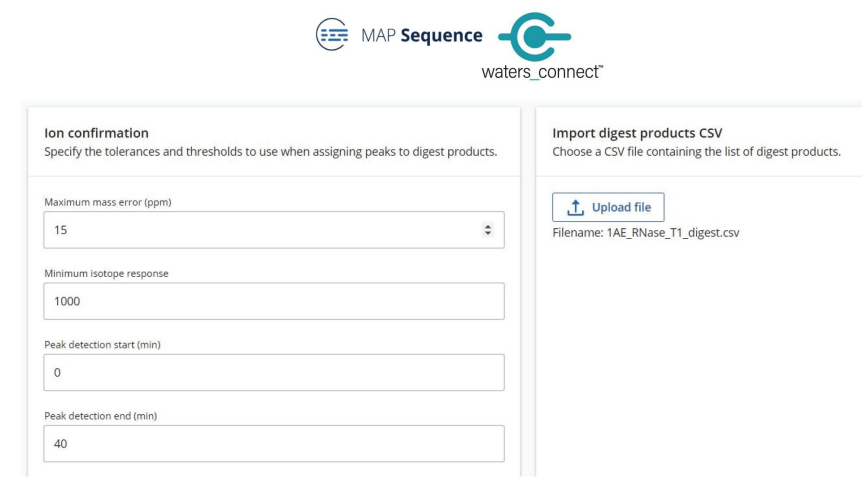*Figure 2. mRNA Cleaver settings used for in-silico prediction of oligonucleotide products resulting from the cleavage of a 100-mer modified sgRNA with RNase T1. The X residue denotes a 2'-OMe phosphorothioated adenosine, while the Y residue corresponds to 2'-OMe phosphorothioated uridine.*

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.
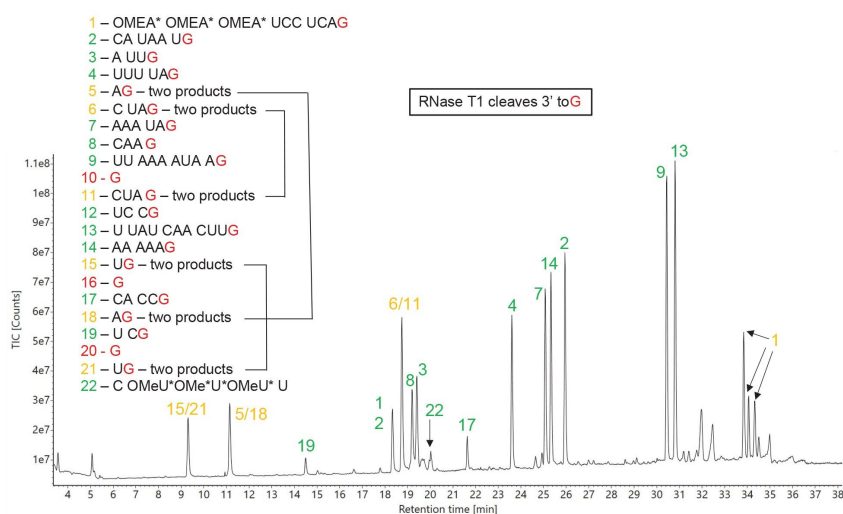
Figure 5. TIC chromatogram recorded for RNase T1 digestion of HPRT1 sgRNA. The predicted oligo products, identified according to their sequence shown the left side of this figure, are used for labeling of all the major chromatographic peaks detected. Peak identifications shown in green indicate unique 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]
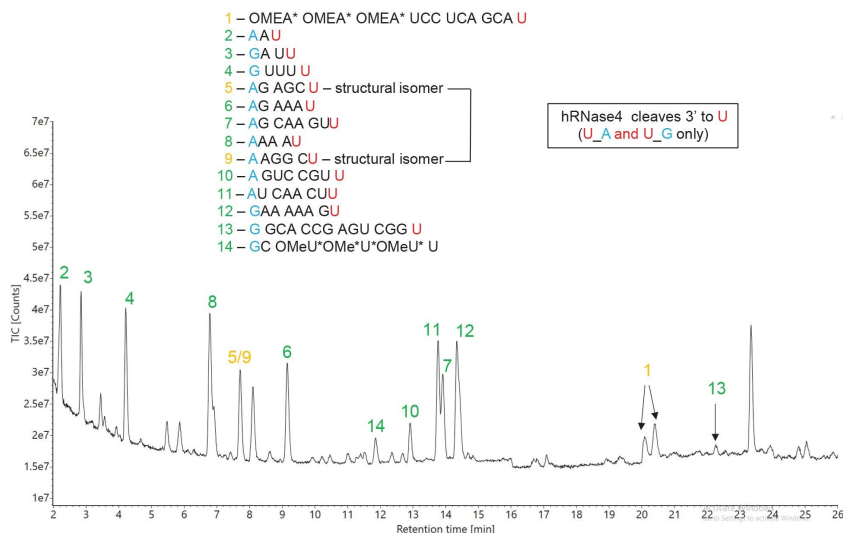
Figure 6. TIC chromatogram recorded for hRNase 4 digestion of HPRT1 sgRNA. The predicted oligo products, identified according to their sequence shown in the center of this figure, are used for labeling of all the major chromatographic peaks detected. Peak identifications shown in green indicate unique 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]
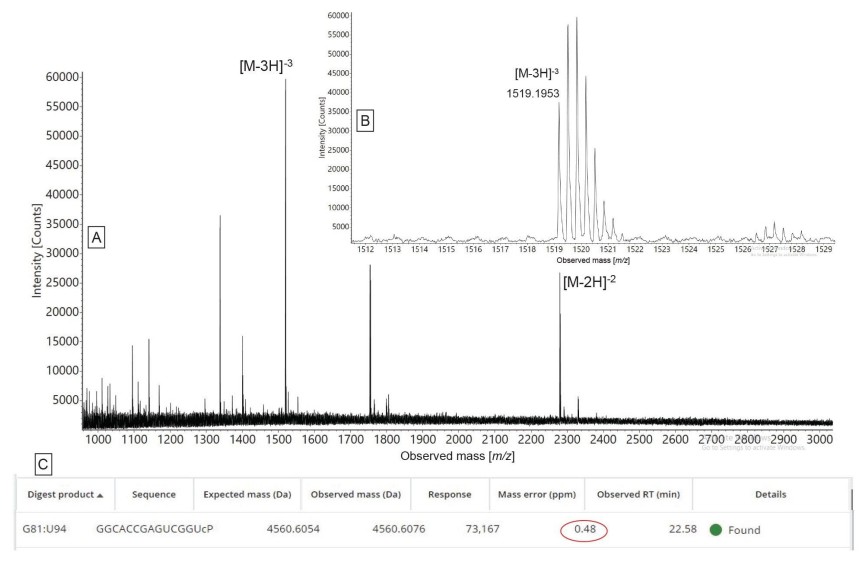

Figure 7. MAP Sequence App automated assignment of a low-intensity oligonucleotide product generated following the digestion of HPRT1 sgRNA with hRNase4 enzyme: (A) combined ESI-MS spectrum of the oligonucleotide product G GCA CCG AGU CGG U; (B) inset showing the isotopic distribution of the triply charge 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 four 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.
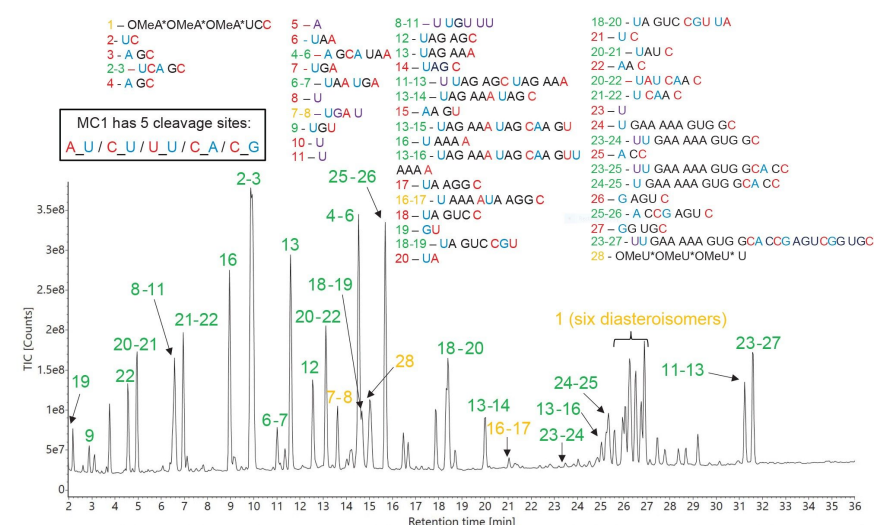
Figure 8. TIC chromatogram recorded for RapiZyme MC1 digestion of HPRT1 sgRNA. The predicted oligo products, identified according to their sequence shown in the two panels located on the right side of this figure, are used for labeling of all the chromatographic peaks detected. Peak identifications shown 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]
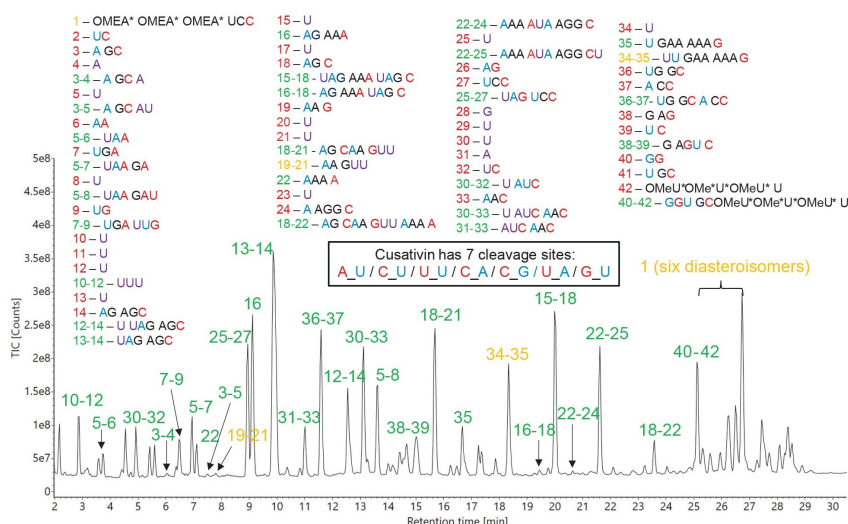
Figure 9. TIC chromatogram recorded for RapiZyme Cusativin digestion of HPRT1 sgRNA. The predicted oligo products, identified according to their sequence shown in the four panels located at the top of this figure, are used for labeling of all the major chromatographic peaks detected. Peak identifications 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.

In an attempt to obtain complete sequence coverage for MC1 and Cusativin, the cleavage sites defined in the mRNA Cleaver were expanded, to include cleavages between unmodified cytidine residues and modified uridines (a C_Y motif, where Y denotes a 2'-OMe phosphorothioated uridine). As a result of this implementation, a newly

predicted oligo product, having the sequence YYYU was detected following MAP Sequence processing in both digests. The isotopic distribution of its singly charged precursor, along with the dot-map fragmentation diagram generated by the CONFIRM Sequence App are displayed in Figure 10. The sequence of this tetranucleotide was verified, indicating that both MC1 and Cusativin cleave between unmodified and modified residues, which is a unique and useful enzymatic feature. The hRNAse4 enzyme cannot cleave when modified residues are present in the RNA sequence, but MC1 and Cusativin can cleave bonds between unmodified and modified residues and therefore are able to produce unique digestion products.[7]
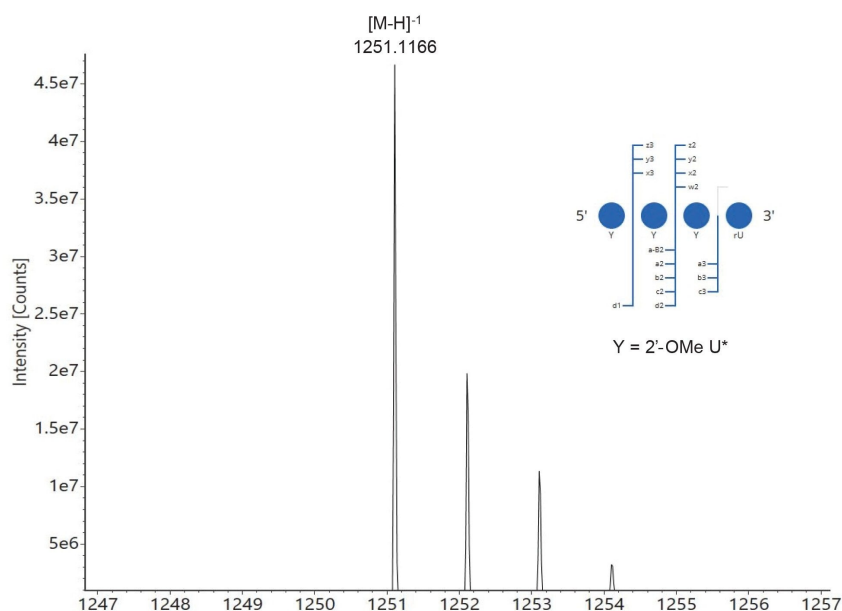
Figure 10. Isotopic distribution of the singly charged tetranucleotide YYYU detected in the RapiZyme MC1 digest of the modified sgRNA. The inset displays a screen capture from the CONFIRM Sequence App, showing a dot-map fragmentation diagram that confirms the sequence of this oligo digestion product as YYYU (residues no 97-100 in the sequence of the 100-mer sgRNA).

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

∧