

## Synthetic mRNA Oligo-Mapping Using Ion-Pairing Liquid Chromatography and Mass Spectrometry

---

Maissa M. Gaye, Jonathan Fox, Johannes P.C. Vissers, Ian Reah, Chris Knowles, Matthew A. Lauber

Waters Corporation

---

### Abstract

Messenger RNA has quickly become an important modality for human medicine, as shown with its evaluation for cancer treatment and the FDA approval of mRNA vaccines for COVID-19. The rapid development of mRNA vaccines and other classes of mRNA therapeutics is supported by advances in analytical methodologies. One important aspect of such methodologies is confirming the identity, purity, and modification(s) of a therapeutic mRNA through mapping its sequence by liquid chromatography-mass spectrometry (LC-MS). Mass spectrometry-based sequencing of RNAs has an advantage over templated RNA sequencing by offering direct molecular detection of fragments, which can be used to localize nucleoside impurities and identify important structural attributes (5'-cap and poly-A tail). As such, we propose a workflow for comprehensive bottom-up LC-UV-MS characterization of mRNAs that yields mRNA component-annotated chromatograms derived from accurate-mass matching.

### Benefits

- High chromatographic resolution and MS sensitivity with the use of ion-pairing reversed-phase chromatography in combination with the ACQUITY™ Premier Oligonucleotide BEH™ C<sub>18</sub> 300 Å Column
-

· Automated mRNA digest annotation based on accurate-mass matching as facilitated by *in silico* mRNA digestion calculations and application of waters\_connect/UNIFI™ Scientific Libraries

---

## Introduction

The SARS-CoV-2 pandemic provided the impetus for the fast-paced development of nucleic-acid based medicine, especially synthetic mRNA.<sup>1</sup> Now, 40 years after its discovery in 1961 by Brenner *et al.*<sup>2</sup> [1], mRNA has evolved into being an important modality with massive potential, as shown with the start of an in-human clinical trial for cancer treatment<sup>1-3</sup> and the full approval of two COVID-19 mRNA vaccines by the US Food and Drug Administration in August 2021 and in January 2022, respectively. The rapid development of mRNA vaccines and other classes of mRNA therapeutics are supported by advances in analytical methodologies. One important aspect of such methodologies is confirming the identity, purity, and modification of a therapeutic mRNA through oligo-mapping and sequencing via liquid chromatography hyphenated to mass spectrometry (LC-MS). Nucleic acid sequencing technologies, like Sanger and next-generation sequencing (NGS), provide valuable information to drug developers. However, there is also a heightened level of analysis that can be achieved through the use of LC in combination with tandem MS (LC-MS/MS)<sup>4</sup> or MS<sup>E</sup> (alternating low and high collision energy)<sup>5</sup> based fragmentation. Similar to a proteomics bottom-up approach, LC-MS/MS or MS<sup>E</sup> based sequencing have the advantage of direct molecular detection of RNA fragments, including the detection and localization of nucleoside impurities and important structural attributes, like lipidated nucleobases,<sup>6</sup> endcapped residues, and polyA tail modifications.<sup>4,7</sup> Unlike bottom-up proteomics workflows, where a plethora of data processing solutions exist, options for RNA mapping are limited. We propose a workflow for oligo-mapping based on a bottom-up approach for the characterization of a given synthetic mRNA within a single platform comprising LC, UV detection and MS measurements. Digestion components were processed using an in-house developed, freely available *in silico* digestion library calculator, mRNAalcondemand, in combination with waters\_connect™ to yield an annotated chromatogram. Here, we demonstrate this analytical approach for mRNA sequence mapping using RNase T1 digested luciferase mRNA.

---

## Experimental

---

## Sample Information

Approximately 90 µg of synthetic Cypridina luciferase mRNA (uncapped and not modified with a polyA tail); a gift from Bijoyita Roy (New England Biolabs, Ipswich, MA) was digested using 3'-guanosine specific ribonuclease RNaseT1 (Worthington Biochemical Corporation, Lakewood, NJ). Note that this workflow was repeated using 10 µg of TriLink Biotechnologies (CleanCap® FLuc mRNA, San Diego, CA) firefly luciferase mRNA (untranslated sequences are proprietary) and comparable results were achieved. Luciferase mRNA was denatured prior to digestion using 20 µL urea (8 M) prepared in nuclease-free buffer (10 mM Tris, 0.1 mM EDTA, pH 7.5 in water, Integrated DNA Technologies, Inc, Coralville, IA) at 80 °C for 5 minutes. Next, 24 µg (~10kU) of RNase T1 (Worthington Biochemical Corporation, Lakewood, NJ) resuspended in nuclease-free buffer was added to the denatured mRNA at room temperature and the mixture was then incubated at 37 °C for 30 minutes. Nuclease-free buffer (40 µL) was added at the end of the incubation period, bringing the total sample volume to 80 µL. The final aliquot was transferred to a polypropylene 300 µL autosampler vial (p/n: [186002639 < https://www.waters.com/nextgen/global/shop/vials-containers--collection-plates/186002639-polypropylene-12-x-32-mm-screw-neck-vial-with-cap-and-preslit-pt.html >](https://www.waters.com/nextgen/global/shop/vials-containers--collection-plates/186002639-polypropylene-12-x-32-mm-screw-neck-vial-with-cap-and-preslit-pt.html) ). The resulting digest was subjected to ion-pairing reversed-phase chromatography (IP-RPLC) without any further manipulation prior to MS detection in negative ion mode using the BioAccord™ RDa™ Detector.

## LC Conditions

LC system:	ACQUITY UPLC™ Premier BSM System (as part of the BioAccord System)
Detector:	ACQUITY UPLC TUV Detector
Wavelength:	260 nm
Column:	ACQUITY Premier Oligonucleotide BEH C18, 2.1 X 150 mm, 300 Å, 1.7 µm (p/n:186010541)
Column Temperature:	70 °C
Sample Temperature:	4 °C

Injection: 5  $\mu$ L

Flow Rate: 0.4 mL/min

Mobile phase A: 0.1% N,N-diisopropylethylamine (DIPEA) as the IP reagent and 1% 1,1,1,3,3,3-hexafluoroisopropanol (HFIP) in deionized water

Mobile phase B: 0.0375% DIPEA and 0.075% HFIP in 65:35 acetonitrile/water

## Gradient Table

Time (min)	mL/min	A (%)	B (%)	Curve
Initial	0.4	97	3	*
60	0.4	70	30	6
60.5	0.4	5	95	6
61	0.4	97	3	6
70	0.4	97	3	6

Run time = 70 minutes

## MS Conditions

MS system: BioAccord LC-MS System

Detector: ACQUITY RDa Detector

Mode: Full scan with fragmentation

Polarity:	Negative
Cone voltage:	40 V
Fragmentation cone voltage:	80–200 V
Mass range:	High (400–5000 <i>m/z</i> )
Scan rate:	2 Hz
Capillary voltage:	0.80 kV
Desolvation temperature:	400 °C

---

## Results and Discussion

Ion-pairing reversed-phase chromatography (IP-RPLC) with a C<sub>18</sub> stationary phase has become a tried-and-true approach for the analysis of oligonucleotides.<sup>4,7</sup> The mobile phase contains an ion pairing reagent, commonly an alkylamine, which adsorbs onto the C<sub>18</sub> stationary phase<sup>8,9,10</sup> and thereby introduces a mixed mode like retention mechanism.<sup>8–10</sup> The N,N-diisopropylethylamine (DIPEA)/ 1,1,1,3,3,3-hexafluoroisopropanol (HFIP) mobile phase system used in this application is compatible with both optical UV detection and negative ion mode mass spectrometry.<sup>4,7,–10</sup> HFIP is used to enhance electrospray ionization.<sup>8</sup>

RNase T1 digested luciferase mRNA was injected onto an ACQUITY Premier Oligonucleotide BEH C<sub>18</sub> (2.1 x 150 mm, 300 Å, 1.7 µm) Column and gradients were developed with an ACQUITY Premier Binary LC equipped with an ACQUITY UPLC TUV Detector. The ACQUITY Premier Oligonucleotide BEH C<sub>18</sub> Column used in this work is similar to ACQUITY Premier Oligonucleotide BEH C<sub>18</sub> 130 Å Column, but with a wider pore size, providing better resolution for longer oligonucleotide species. Data were acquired in triplicate using negative ion mode mass spectrometry with the ACQUITY RDa Detector of a BioAccord Benchtop LC-MS system. Moreover, the ACQUITY RDa Detector was programmed to acquire MS<sup>E</sup> data such that every other scan produced a high energy

---

fragment ion spectrum that could be later used to corroborate LC peak identifications. Figure 1 depicts total ion chromatograms (TIC) for an RNase T1 control sample (top trace), mRNA control sample (bottom trace) and the digested mRNA (middle trace). Oligonucleotide fragments resulting from the digestion of luciferase mRNA with RNase T1 were readily separated according to a separation with a 4-sigma peak capacity of 613. Overall, chromatographic peaks were sharp and symmetrical with  $\sim 0.01$  minutes variation in retention time (RT) over triplicate injections. Digest components eluted between 2 and 23 minutes on a 60 minute gradient time method; some incompletely digested mRNA was seen to elute around 29 minutes and intact RNaseT1 eluted around 54 minutes. As can be seen on the top trace of Figure 1, an RNase T1 control sample showed signal only after a 50 minute retention time. This confirmed that RNase T1 would not introduce interference within the retention window of the mRNA digestion components. (Figure 1, middle trace). Likewise, the bottom trace of Figure 1 shows that intact luciferase mRNA elutes at approximately 38 minutes, confirming that the peak observed at 29 minutes in the digested sample (Figure 1, middle trace) corresponds to incompletely digested mRNA. We note here that in the case of synthetic mRNA comprising 5'-cap and poly-A tail structures, peaks at  $\sim 29$  and 37 minutes are observed post digestion. The slight shift from  $\sim 38$  minutes to  $\sim 37$  minutes might be indicative of the undigested polyA structure (an investigation into this chromatographic behavior will be described in a future application note).

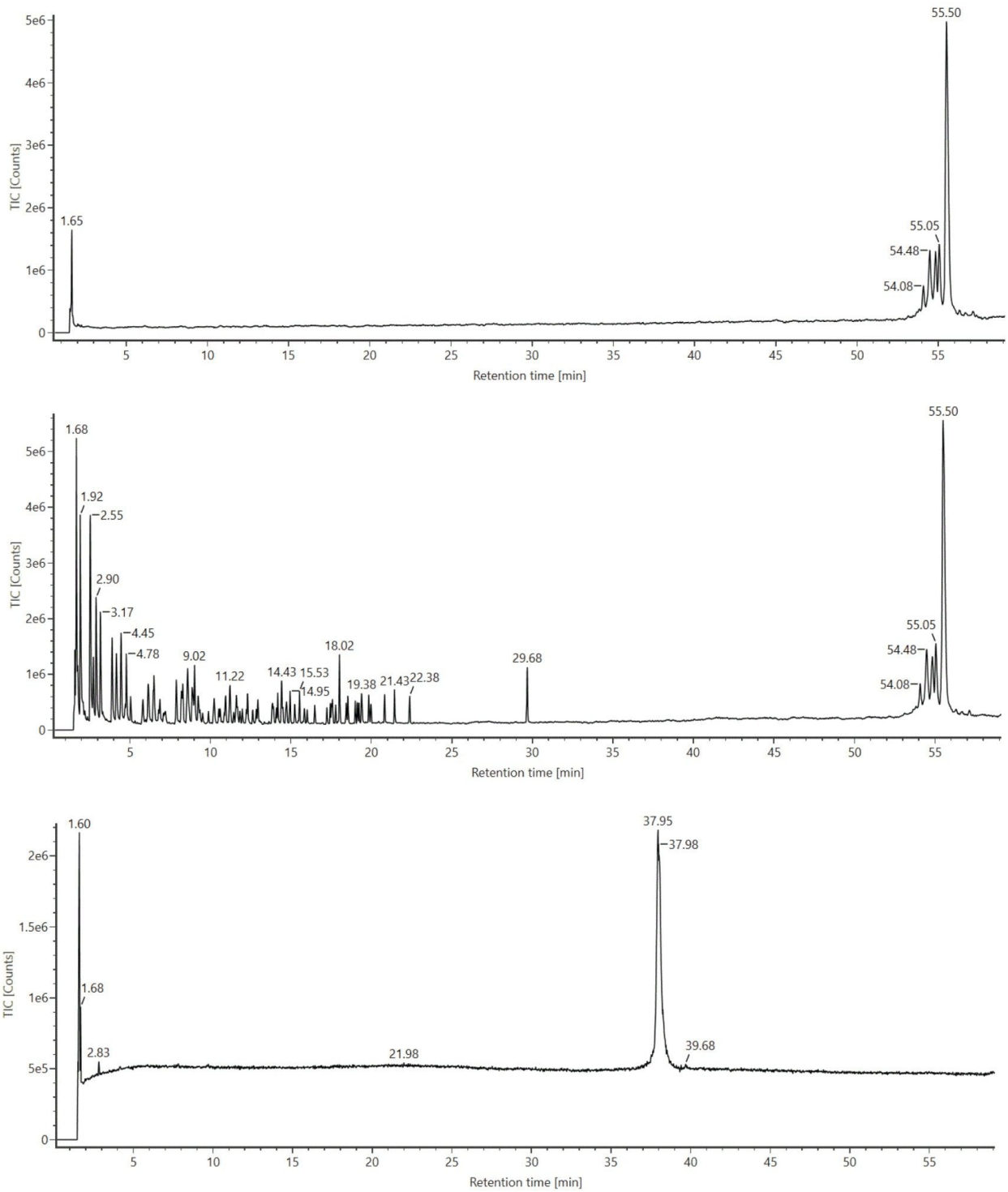


Figure 1: TIC for RNase T1 control sample (top trace), mRNA control sample (bottom trace) and the digest

---

*(middle trace) obtained from ion-pairing reversed-phase chromatography (IP-RPLC) of luciferase mRNA digested with Rnase T1 and analyzed using the ACQUITY UPLC I-Class System (ACQUITY Premier Oligonucleotide BEH C<sub>18</sub> Column, 2.1 x 150 mm, 300 Å, 1.7 μm) and the BioAccord ACQUITY RDa Detector in negative ion mode.*

The graphical user interface (GUI) of the *in-silico* digestion mRNA calculator, mRNAalcondemand, is shown in Scheme 1. Next to the base sequence, a number of digestion parameters are specified, such as modification(s), enzyme, and missed cleavages. Based on this input, the calculator defaults to a number of MS specific settings, including charge state and *m/z* ranges, as well as the ability to conduct calculations based on monoisotopic or average mass. The generated output, in the form of a flat text csv file, can be utilized in UNIFI or waters\_connect software, or used for complementary downstream analysis.



mRNA MASS Calculator

### Sequence and Modifications

```

      AUGGAGGACGCCAAGAACAUCAAGAAGGGCCCCCGCCCCUUCUACCCCCUGGAGGAC
      GGCACCGCCGGGAGCAGCUGCACAAGGCCAUGAAGCGGUACGCCUGGUGCCCGGC
      ACCAUCGCCUUCACGACGCCCAUCAGAGGGGACAUACCUACGCCGAGUACUUC
      GAGAUAGCGUGCGGUGGCCGAGGCCAUGAAGCGGUACGCCUGAACACCAACCAC
      CGGAUCGUGGUGGACAGCAGAGAAGCCUUCAGUUCUUAUGCCCGUGCUGGGGCGC
      CCUGUUAUCGCGGUGGCCGUGGCCCGCCCAACGACAUCAACAAGCAGCGGGAGCU
      GCUGAACAGCAUGGGCAUCAGCCAGCCACCGUGGUGUUCGUGAGCAAGAAGGGCC
      UGCAGAAGAUCCUGAACGUGCAGAAGAAGCUGCCCAUCAUCCAGAAGAUCAUCA
      UGGACAGCAAGACCGACUACAGGGCUUCAGAGCAUGUACACCUUCGAGCCAGCC
      ACCUGCCCCCGGCUUCAACGAGUACGACUUCGUGCCGAGAGCUUCGACCCGGGACA
    
```

Clear

Modifications/Fragments:

Modification:

Enzyme:

a-specific digestion:  OFF    non-redundant:  OFF

Missed cleavages:

Adducts:

### MS settings

Length

Min Length:     Max Length:

Length a-specific fragment

Min Length:     Max Length:

Charge

Min Charge:     Max Charge:

Polarity:  negative

m/z range

Min m/z range:     Max m/z range:

Mass:  monoisotopic

Reset    Calculate

digested\_sequence....csv ^

Scheme 1: mRNA calculator GUI for *in silico* mRNA digest mass calculation.

Next, a library was created where the digest components are considered to be individual analytes, which was achieved by importing the output of the calculator as a spreadsheet into a UNIFI Scientific Library. The created libraries can be utilized in an HRMS Screening Analysis Method, targeting the digest compounds post-acquisition, using user-specified tolerances. The use of mass tolerance-based library matching is demonstrated below. As a result of the library search, an annotated chromatogram is automatically generated (Figure 2). A close-up, corresponding to a zoomed view of the 17 to 20 minute retention time window is represented in Figure

2.

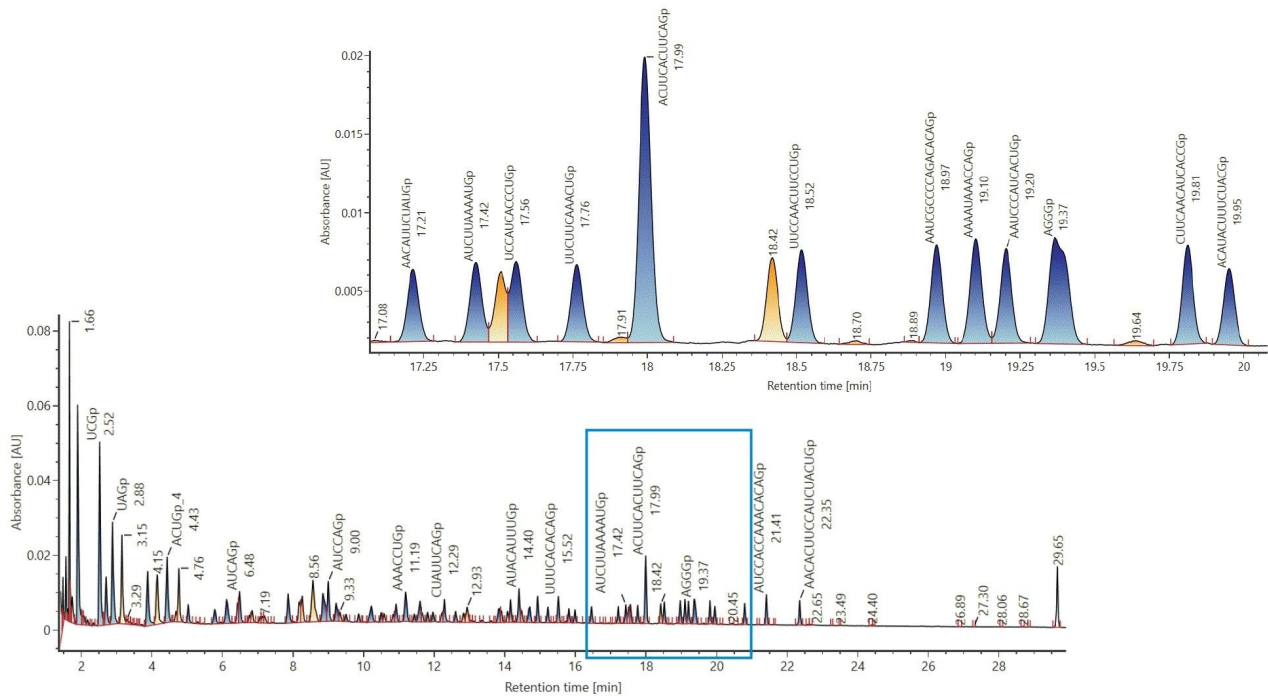


Figure 2: Annotated TIC of luciferase mRNA digest generated after matching, based on accurate mass, to a target component library. Luciferase mRNA was digested with Rnase T1 and analyzed using the ACQUITY Premier BSM LC (ACQUITY Premier Oligonucleotide BEH C<sub>18</sub> Column, 2.1 x 150 mm, 300 Å, 1.7 μm) and the BioAccord ACQUITY RDa Detector in negative mode. Target components were calculated using the mRNA MASS calculator.

A total of 436, 428, and 441 potential identifications (IDs) of digestion components were produced from each technical replicate upon screening data against an *in silico* library created with an allowance for up to 2 missed cleavages. Several criteria were considered for manual validation. 40 out of 441 identified components were rejected based on abundance and peak shape, e.g., low abundant chromatographic peak shoulders. The majority of these rejected IDs (27 out of 40) were situated between 24 and 60 minutes. Overall, ~60% of the identified and validated components were within 10 ppm mass error (261 digest components). The RNase T1 control sample (Figure 1, top trace) was subjected to the same query and as expected did not yield any identifications. In addition to peak shape and abundance, we further validated the results based on high-confidence interpretation

of isotopic distributions to deduce charge assignments, which yielded 139 digest components (65% within 5 ppm mass error). Lastly, some assignments were not included in the analysis because they appeared to be redundant assignments triggered by repeated detection on broadened and shoulder-containing chromatographic peaks. The number of identifications was thereby reduced by another 16 components. In a couple instances, however, analyte masses showed up at multiple retention times to produce two unique IDs corresponding to two distinct, well-defined chromatographic peaks. Here, we also took note of the presence of isomeric IDs (species having the same chemical composition but different sequences) as well as isobaric IDs (species having different chemical compositions [different exact mass], yet similar nominal mass). Ultimately, 90 unique components could be identified based on accurate-mass matching, and these are reported in Table 1, including isomeric nucleotide sequences and isobaric ions (cells highlighted in grey in Table 1).



---

based on accurate-mass matching. Cells corresponding to isomeric sequences or isobaric ions are highlighted in grey.

Figure 3 shows example data that produced an identification of digestion component UCCACUCUAUGp. This component eluted at 16.49 minutes as seen on the top left chromatogram and was identified from the *in silico* digestion library based on 5 ions carrying 2 to 6 negative charges (Figure 3, left bottom trace) at  $m/z$  576.5732 ( $[M-6H]^{6-}$ ), 692.0876 ( $[M-5H]^{5-}$ ), 865.3603 ( $[M-4H]^{4-}$ ), 1154.1438 ( $[M-3H]^{3-}$ ) and 1731.7095 ( $[M-2H]^{2-}$ ). Isotopic distributions for  $[M-5H]^{5-}$  and  $[M-2H]^{2-}$  ions are illustrated to show data in support of charge state assignments (Figure 3, right).

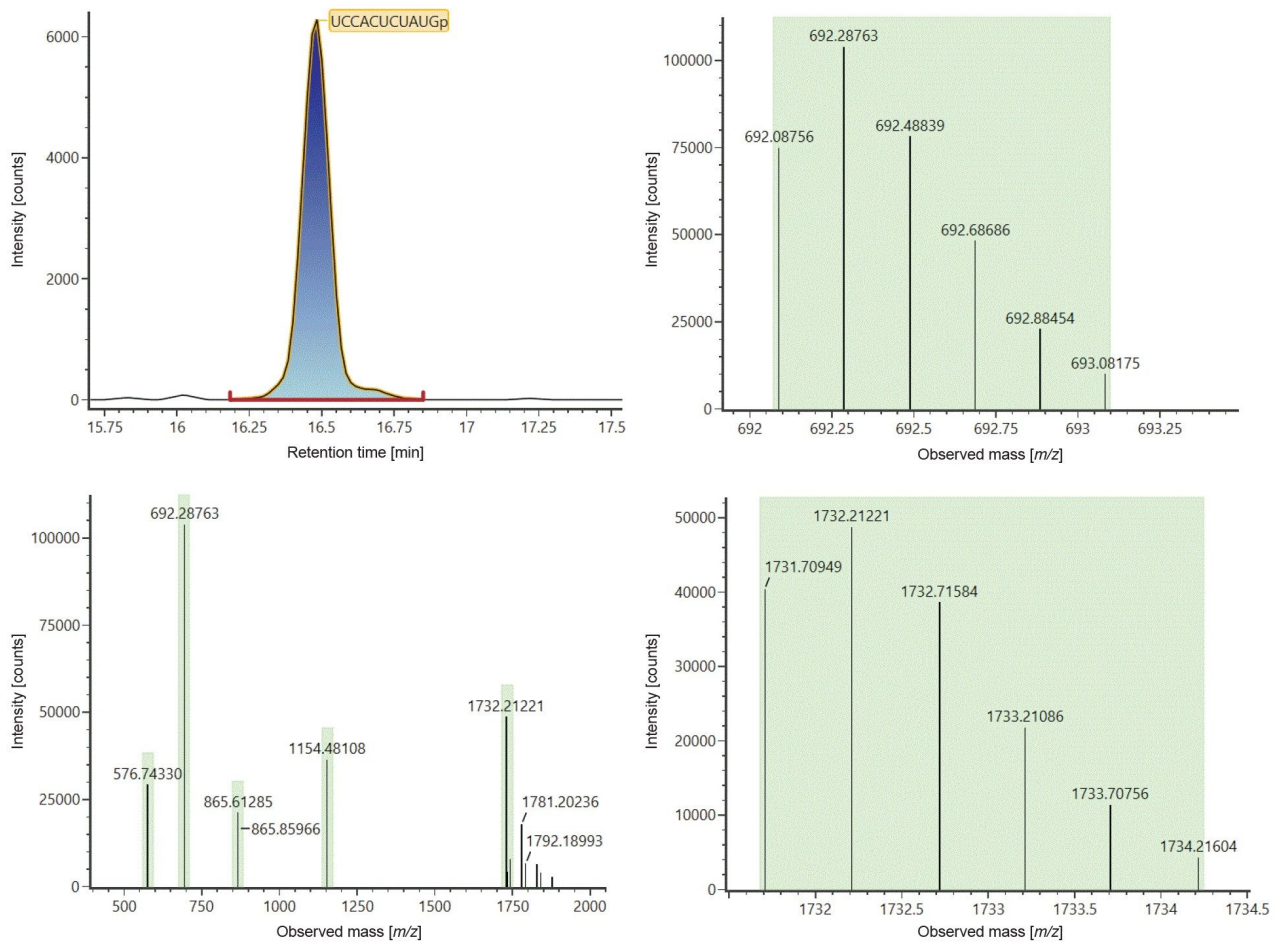


Figure 3: Identification of digest component UCCAUCACCCUGp eluting at 16.49 minutes (left top trace). The component was identified from the generated *in silico* digest based on 5 ions carrying 2 to 6 negative charges (left bottom trace) at  $m/z$  576.5732 ( $[M-6H]^{6-}$ ), 692.0876 ( $[M-5H]^{5-}$ ), 865.3603 ( $[M-4H]^{4-}$ ), 1154.1438 ( $[M-3H]^{3-}$ ) and 1731.7095 ( $[M-2H]^{2-}$ ). The experimentally observed isotopic distributions for  $[M-5H]^{5-}$  and  $[M-2H]^{2-}$  ions are depicted on the right top and bottom traces respectively.

Ambiguities resulting from the presence of isomeric or isobaric ions were resolved using the waters\_connect CONFIRM Sequence™ application for the interpretation of MS<sup>E</sup> spectra as reported in Table 2. Isomeric sequences at position 623–631 (ACAUCCUCGp) and 551–559 (UCACCAUCGp) are predicted from the RNase T1 *in silico* digestion and both are assigned to the same retention time of 13.87 minutes. The correct assignment cannot be made using the intact mass analysis. MS<sup>E</sup> data from the same injection were used to elucidate the

correct sequence for this assignment using the waters\_connect™ CONFIRM Sequence application, wherein high energy fragment ions are predicted for each sequence and matched to isotope clusters of the integrated raw data via a bespoke algorithm. Confirmed fragment ions are represented on a Dot-Map allowing a rapid assessment of the sequence coverage (Figure 4, B): full sequence coverage was obtained for UCACCAUCGp such that it could be readily validated as the correct assignment (Table 2, index #63).

Neutral precursor masses obtained for manually validated digestion components (Table 1) spanned from 973.1406 Da (CCGp, RT 2.54 minutes) to a 20-mer nucleotide at 6366.7830 Da (UCAUUGAGUUCUCAAACUGp, RT 19.39 minutes). The earliest eluting component was UGUGp, and it was identified with an observed neutral mass of 1320.1560 and a retention time of 1.64 minutes. The last luciferase mRNA digest component to elute within the manually validated set of IDs (Table 1) was CAGGp (1342.1992Da), and it was observed to elute at 22.38 minutes. Elution of CAGGp at such a late retention time was unexpected since the closely related sequence UAAGp eluted at ~4 minutes. In order to address the issue of false positives, we used the waters\_connect CONFIRM Sequence application to further characterize components IDs. Validated based on MS data with MS<sup>E</sup> data. As it can be seen in Table 2, 34/90 components, including CAGGp, did not generate an adequate number of MS<sup>E</sup> fragments to further validate the sequence assignment, although accurate-mass matching supported the assignment. Another interesting example is the observation within the same dataset of assigned sequences AAAACAUGUUGCCGp (4860.6678 Da, 15-mer, 9 purines) and AUACAUUUGACAAAGp (4845.6568 Da, 15-mer, 9 purines) with identical purine content but eluting 10 minutes apart. Using MS<sup>E</sup> data, we were able to rule out AAAACAUGUUGCCGp as a false positive and confirm the identification of AUACAUUUGACAAAGp. This demonstrates the importance of strategically using both accurate-mass matching and fragmentation spectra to aid the unambiguous identification of components resulting from digested mRNA sequences. Additionally, our observations have led us to recognize that the retention of digested mRNA components may not be as predictable as first thought and that there is an urgent need for the interactions between oligonucleotides and chromatographic stationary phases to be further studied and modeled.

Lastly, we manually estimated sequence coverage by comparing the matched, manually validated digest components to the mRNA sequence. A preliminary estimate of sequence coverage for the 401 initial matches produced a coverage value of approximately 76%. When the rigorously validated matches were checked against the mRNA sequence, a coverage value of ~22% was obtained. Many of the observed digestion components mapped to more than one location in the luciferase mRNA sequence (Table 2) and the redundancy was expected because there are only four unique residues in a modified or fully modified nucleic acid sequence.

Index	Validated luciferase mRNA digest components using accurate mass matching (10 ppm mass error)	Observed RT (min)	Position on mRNA sequence	Additional position(s) on mRNA sequence	Validation using MS/MS data (CONFIRM Sequence)			
					coverage (%)			
					Yes / No	Peak 1	Peak 2	Peak 3
1	UCUGp_1	1.64	217-229	1553-1556	yes	50		
2	CCUUGGCCGp	1.65	886-905	NA	no	0		
3	UCUACUGCCGp	1.7	71-79	NA	no	0		
4	CCGGUCAAGp	1.93	1520-1529	NA	no	10		
5	GAAp	2.1	378-381	445-448, 635-638, 712-715	yes	25		
6	CCGp_6	2.54	59-61	246-248, 284-286, 838-840, 855-857, 1073-1075, 1076-1078, 1223-1225, 1520-1522	no	0		
7	GAGACACCUAGp	2.54	1667-1678	NA	no	0		
8	CUCGp	2.55	25-29	NA	no	0		
9	CUg	2.55	177-179	422-424, 427-429, 452-454, 512-514, 560-562, 852-854, 906-908, 940-942, 965-967, 1430-1432, 1493-1495	yes	50		
10	GAGACAUUUg	2.55	455-466	NA	no	16.67		
11	UAUUGAUAGCAGp	2.55	165-176	NA	no	0		
12	ACGp	2.73	779-781	809-811, 920-922, 958-960	no	0		
13	CAGp	2.73	174-176	273-275, 287-289	yes	50		
14	AUGp	2.92	209-211	291-293, 407-409, 571-573, 709-711, 806-808, 1100-1102, 1193-1195, 1271-1273, 1367-1369, 1415-1417, 1427-1429, 1467-1469, 1581-1583, 1688-1690	yes	100		
15	AAGp_4	3.18	43-45	146-148, 155-157, 379-381, 446-448, 636-638, 713-715, 1105-1107, 1139-1141, 1220-1222, 1304-1306, 1433-1435, 1475-1477, 1490-1492, 1679-1681	yes	100		
16	ACACCCUAGAAg	3.44	1670-1681	NA	no	16.67		
17	UGAUUUg	3.9	1530-1537	NA	yes	50		
18	CCCGp	3.91	902-905	NA	no	0		
19	CCUGp	3.91	1439-1442	1439-1442, 1564-1567, 1599-1602	yes	25		
20	CUCUGUUg	3.91	1307-1314	NA	no	12.5		
21	GUUUUUg	3.91	1048-1055	NA	no	12.5		
22	AAGCCGUGp	4.17	1220-1227	NA	no	12.5		
23	UAAGp_1	4.79	644-647	1371-1374	yes	50		
24	AAAGp	5.06	312-315	1147-1150, 1645-1648	yes	100		
25	CUCUGp	5.81	1307-1311	NA	yes	60		
26	UCCCGp	5.83	1232-1236	NA	no	0		
27	CACCGp	6.18	1470-1474	NA	no	0		
28	AUCUGp	6.51	639-643	1445-1449	yes	100		
29	AUCCGp	6.53	30-34	NA	yes	20		
30	AAAAACAUGUUGCCGp	6.52	234-248	NA	no	0		
31	CAAAg	6.79	822-826	NA	yes	40		
32	CCUCUGp	7.89	687-692	NA	yes	50		
33	CUCCCGp	7.91	968-973	NA	yes	33.33		
34	CCCCAGp	8.29	653-658	NA	no	16.67		
35	AUCCAGp	8.58	734-739	NA	yes (coelution)	33.33	33.33	
36	CAUUGp	8.58	65-70	NA	yes	100		
37	AUUCUGp	8.63	1204-1209	NA	yes	66.67		
38	AAUCAGp	8.88	1155-1160	NA	yes	66.67		
39	UACAGp	8.88	841-846	NA	yes	100		
40	AUCCAGp	8.89	734-739	NA	yes	33.33	33.33	
41	AACUGp	9.03	110-115	332-337, 400-405, 984-989	yes	100		
42	ACACAGp	9.26	659-664	NA	no	16.67		
43	ACAAAGp	9.52	1025-1030	NA	yes	33.33		
44	UCUUCUGp	9.89	828-834	NA	yes	71.43		
45	AUCUCUGp	10.23	679-685	NA	no	14.29		
46	UCUACUGp	10.23	71-77	NA	yes	57.14	28.57	
47	AUUCAGp	10.53	318-324	NA	no	14.29		
48	CAUACUGp	10.53	815-821	NA	yes	57.14		
49	AUUAUGp	10.95	1408-1412	NA	yes	71.43		
50	AAACCUg	11.22	1483-1489	NA	yes	71.43		
51	ACUACAGp	11.22	947-953	1285-1291	yes	71.43		
52	CCACAAg	11.25	1557-1563	NA	yes	14.29		
53	UCUCUUCGp	11.69	1512-1519	NA	yes	37.5		
54	AAAACAGp	11.84	1197-1203	NA	yes	71.43		
55	UCCUACGp	11.99	102-109	NA	yes	37.5		
56	CCACUAGp	12.32	35-42	NA	yes	25		
57	AAAUCCUGp	12.63	798-805	NA	no	12.5		
58	ACAUACUGp	12.63	1277-1284	NA	yes	14.29	62.5	40
59	CUGCCGp	12.64	852-857	NA	no	0		
60	GGCGp	13.37	367-370	NA	no	0		
61	CCUCACACGp	13.84	929-937	NA	yes	22.22		
62	ACAUCCUCGp	13.87	623-631	NA	yes	33.33		
63	UCACCAUCGp	13.87	551-559	NA	yes	100	33.3	
64	AUUCUUUUg	13.93	1418-1426	NA	yes	55.56		
65	ACAUUUUUg	14.19	458-466	NA	yes	22.22		
66	AUUCUUUUg	14.19	1063-1071	NA	yes	44.44		
67	UGAUGAUUCUUUUg	14.19	1413-1426	NA	no	7.14		
68	AUACAUUUg	14.43	1018-1024	NA	yes	55.56		
69	ACACCCUAGp	14.44	1670-1678	NA	no	11.11		
70	AACAACUCGp	14.72	740-748	NA	yes	22.22		
71	AAACAUCGp	14.96	1162-1170	NA	yes	55.56		
72	CAACAAGp	14.96	685-673	NA	yes	77.78		
73	AAAAACAUGp	15.25	234-242	NA	yes	77.78		
74	CUGACCCUACUACGp	15.25	512-526	NA	no	0		
75	UUUCACACAGp	15.55	1109-1118	NA	yes	80		
76	UACAUUUUg	15.84	1348-1355	NA	yes	80		
77	AUACAUUUUg	16.02	1016-1030	NA	yes	20		
78	CUACAUAAAAGp	16.02	1653-1662	NA	yes	50		
79	GGCUGp	16.02	420-424	NA	no	0		
80	UCCACUCUACUg	16.49	786-796	NA	yes	72.73		
81	AACAUCUACUg	17.24	300-310	NA	yes	63.64		
82	AUCUUAUUUg	17.45	698-708	NA	yes	63.64		
83	ACAUACUACUACGp	18.45	197-208	NA	yes	50		
84	CUUGUGp	18.45	177-182	NA	no	0		
85	GUGGp	18.55	506-509	914-917, 1312-1315	no	0		
86	AGGp	19.39	1603-1608	NA	no	0		
87	GAGGp_1	19.39	632-635	1214-1217	no	0		
88	UCAUUGAGUUCUCAAACUg	19.39	593-612	NA	no	0		
89	UAACUACAACCAg	19.43	1392-1404	NA	yes	69.23		
90	CAGGp	22.38	287-290	NA	no	0		

Table 2: Identification and validation of luciferase mRNA digest components based on accurate-mass matching



and further validation using the waters\_connect CONFIRM Sequence application and collected MS<sup>E</sup> spectra.

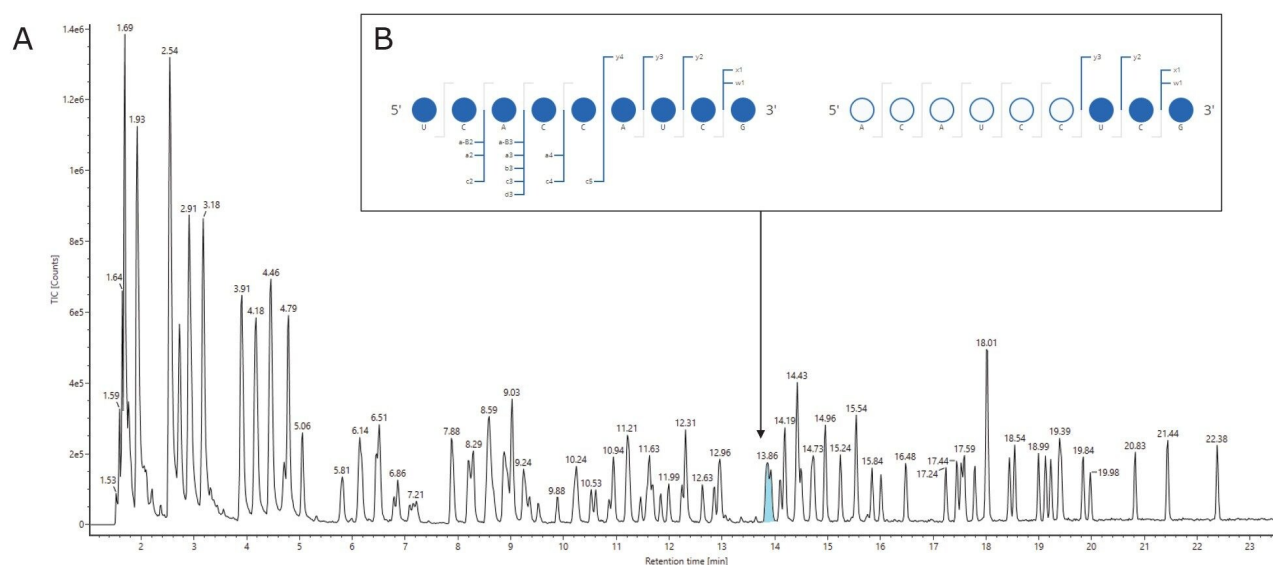


Figure 4: (A) Digested fragment components at position 623–631 (ACAUCCUCGp) and 551–559 (UCACCAUCGp) are predicted from RNase T1 digest and are assigned to the same RT peak in the TIC. It is not possible to determine the correct assignment using intact mass information. (B) MS<sup>E</sup> data from the same injection can be used to elucidate the correct sequence for this assignment. Using the waters\_connect CONFIRM Sequence application, high energy fragment ions are predicted using McLucky annotation<sup>11</sup> for each sequence and matched to isotope clusters of the integrated raw data via a bespoke algorithm. The software presents confirmed fragment ions on a Dot-Map to quickly assess the sequence coverage.

## Conclusion

In the present work, we established a robust analytical workflow for oligo-mapping of synthetic mRNA using IP-RPLC and MS.

- Synthetic mRNAs were reproducibly digested using RNase T1, starting with as little as 10 µg of material, and

injected without additional sample clean-up onto an ACQUITY Premier Oligonucleotide BEH C<sub>18</sub> (2.1 x 150 mm, 300 Å, 1.7 µm) Column

- High chromatographic resolution was achieved using ion-pairing reversed-phase chromatography on an ACQUITY Premier LC such that digest components could be readily separated from incompletely digested mRNA and residual enzyme and efficiently detected with a BioAccord ACQUITY RDa Detector
- Annotated mRNA digest chromatograms were generated based on accurate-mass matching as facilitated by *in silico* mRNA digestion calculations and the application of waters\_connect/UNIFI Scientific Libraries
- Assigned sequences for digested components were further validated based on MS<sup>E</sup> spectra using the waters\_connect CONFIRM Sequence application. In addition, Dot-Map visualization was used to quickly check the fragment ion coverage of potential assignments

With this work, it was our intent to establish the chromatographic, detection, and data interpretation approaches that would be needed to facilitate the bottom-up characterization of an mRNA molecule. RNase T1 digestion was applied only as a first example and a proof of concept on establishing a workflow for data collection and analysis. That said, there is ample opportunity to more comprehensively probe a given mRNA structure by (1) using multiple, different nucleases to generate orthogonal and additive sequence mapping information and (2) adopting a multiplexing approach for data acquisition. These aspects, aimed at achieving comprehensive sequence coverage, will be explored in future work.

---

## References

1. Xu, S.; Yang, K.; Li, R.; Zhang, L., mRNA Vaccine Era—Mechanisms, Drug Platform and Clinical Prospection. *International Journal of Molecular Sciences* 2020, *21* (18), 6582.
2. Brenner, S.; Jacob, F.; Meselson, M., An Unstable Intermediate Carrying Information From Genes to Ribosomes for Protein Synthesis. *Nature* 1961, *190* (4776), 576–581.
3. Weide, B.; Pascolo, S.; Scheel, B. Derhovannessian, E.; Pflugfelder, A.; Eigentler, T. K. Pawelec, G.; Hoerr, I.; Rammensee, H. G.; Garbe, C., Direct Injection of Protamine-Protected Mrna: Results of a Phase 1/2 Vaccination Trial in Metastatic Melanoma Patients. *J Immunother* 2009, *32* (5), 498–507.

4. Jiang, T.; Yu, N.; Kim, J.; Murgo, J.-R.; Kissai, M.; Ravichandran, K.; Miracco, E. J.; Presnyak, V.; Hua, S., Oligonucleotide Sequence Mapping of Large Therapeutic mRNAs via Parallel Ribonuclease Digestions and LC-MS/MS. *Anal. Chem.* 2019, **91** (13), 8500–8506.
5. Plumb, R. S.; Johnson, K. A.; Rainville, P.; Smith, B. W.; Wilson, I. D.; Castro-Perez, J. M.; Nicholson, J. K., UPLC/MS<sup>E</sup>; A New Approach for Generating Molecular Fragment Information for Biomarker Structure Elucidation. *Rapid Communications in Mass Spectrometry* 2006, **20** (13), 1989–1994.
6. Packer, M.; Gyawali, D.; Yerabolu, R.; Schariter, J.; White, P., A Novel Mechanism for the Loss of mRNA Activity in Lipid Nanoparticle Delivery Systems. *Nat. Commun.* 2021, **12** (1), 6777.
7. Goyon, A.; Scott, B.; Kurita, K.; Maschinot, C.; Meyer, K.; Yehl, P.; Zhang, K., On-line Sequencing of CRISPR Guide RNAs and Their Impurities via the Use of Immobilized Ribonuclease Cartridges Attached to a 2D/3D-LC-MS System. *Anal. Chem.* 2021.
8. Guo, L.; Worth, A. J.; Mesaros, C.; Snyder, N. W.; Glickson, J. D.; Blair, I. A., Diisopropylethylamine/Hexafluoroisopropanol-Mediated Ion-Pairing Ultra-High-Performance Liquid Chromatography/Mass Spectrometry for Phosphate and Carboxylate Metabolite Analysis: Utility for Studying Cellular Metabolism. *Rapid Commun Mass Spectrom* 2016, **30** (16), 1835–1845.
9. Birdsall, R. E.; Gilar, M.; Shion, H.; Yu, Y. Q.; Chen, W., Reduction of Metal Adducts in Oligonucleotide Mass Spectra in Ion-Pair Reversed-Phase Chromatography/Mass Spectrometry Analysis. *Rapid Commun Mass Spectrom* 2016, **30** (14), 1667–1679.
10. Fountain, K.; Gilar, M.; Budman, Y.; Gebler, J., Purification of Dye-Labeled Oligonucleotides by Ion-Pair Reversed-Phase High-Performance Liquid Chromatography. *Journal of chromatography. B, Analytical technologies in the biomedical and life sciences* 2003, **783**, 61–72.
11. McLuckey, S. A.; Van Berkel, G. J.; Glish, G. L., Tandem Mass Spectrometry of Small, Multiply Charged Oligonucleotides. *J. Am. Soc. Mass Spectrom.* 1992, **3** (1), 60–70.

## Acknowledgments

We would like to thank our colleagues at Waters for their valuable contribution to this work: Ana-Maria Rotaru, Emanuela Petreanu, Claudia Florea, Dave Jackson and Simon Jones. Many thanks to our collaborators at New England Biolabs: Bijoyita Roy, Siuhong Chan, Ivan R. Corrêa Jr., Erbay Yigit, and G. Brett Robb for providing

luciferase mRNA and many thoughtful discussions.

---

## Featured Products

[ACQUITY UPLC I-Class PLUS System <https://www.waters.com/134613317>](https://www.waters.com/134613317)

[BioAccord LC-MS System for Biopharmaceuticals <https://www.waters.com/waters/nav.htm?cid=135005818>](https://www.waters.com/waters/nav.htm?cid=135005818)

[ACQUITY UPLC Tunable UV Detector <https://www.waters.com/514228>](https://www.waters.com/514228)

[ACQUITY RDa Detector <https://www.waters.com/waters/nav.htm?cid=135077027>](https://www.waters.com/waters/nav.htm?cid=135077027)

[UNIFI Scientific Information System <https://www.waters.com/134801648>](https://www.waters.com/134801648)

[waters\\_connect <https://www.waters.com/waters/nav.htm?cid=135040165>](https://www.waters.com/waters/nav.htm?cid=135040165)

720007669, June 2022

© 2023 Waters Corporation. All Rights Reserved.

[Terms of Use](#) [Privacy](#) [Trademarks](#) [Sitemap](#) [Careers](#) [Cookies](#) [Preferências de cookies](#)