Waters[™]

응용 자료

RNA CQA Analysis using the BioAccord LC-MS System and INTACT Mass waters_connect Application

Rebecca J. D'Esposito, Catalin E. Doneanu, Heidi Gastall, Scott J. Berger, Ying Qing Yu

연구용으로만 사용되며 진단 절차에는 사용할 수 없습니다.

Abstract

This application note highlights new informatics capabilities to analyze the critical quality attributes (CQAs) of ribonucleic acid (RNA) based therapeutics. Ion-Pair Reversed Phase (IP-RP) Liquid Chromatography-Mass Spectrometry (LC-MS) data was acquired using the BioAccord[™] LC-MS System under the control of the waters_connect[™] informatics platform. Application specific software has been developed including the mRNA Cleaver microApp (to generate *in silico* fragments), the INTACT Mass App (to detect and assign RNA digestion products), and the Coverage Viewer microApp (to visualize digestion data coverage). This workflow was demonstrated for CQAs of both single guide RNAs (sgRNAs) and messenger RNAs (mRNAs), typical of the latest generation of therapeutics and vaccines.

Benefits

- Data analysis workflow for RNA digestion fragment mapping, the mass confirmation of nucleic acid digestion products, using a new version of the INTACT Mass App for data processing and additional microApp tools for *in silico* generation and curation of digest fragments
- · Additional workflows using the INTACT Mass App for measuring mRNA 5' Capping efficiency and 3' poly(A)

tail heterogeneity

Introduction

There has been a surge in the development of RNA-based therapeutics and vaccines in the last decade.¹⁻³ The recent development and approval of the two COVID mRNA-based vaccines has brought RNA-based therapeutics to the forefront of the biopharma industry.⁴⁻⁶ As such, development of analytical methods for monitoring the CQAs of RNA-based therapeutics has become a high priority for ensuring proper control of manufacturing processes. CQAs for nucleic acid therapeutics can include sequence confirmation, 5' capping efficiency (mRNA) and structure, 3' poly(A) tail heterogeneity (mRNA) analysis, localization of modified nucleotides, and purity assessment of the active RNA product.

LC-MS systems have been established as an effective investigative tool when characterizing CQAs of RNA-based therapeutics.⁷⁻¹⁰ However, data processing is a common bottleneck that needs to be addressed. Within this application note, we present a data processing workflow for RNA CQA analysis using tools developed on the waters_connect informatics platform, specifically the INTACT Mass, mRNA cleaver, and Coverage Viewer applications.

Experimental

Reagents and Sample Preparation

The experimental conditions reported here are for the sgRNA digestion and sequence mapping experiments. The appropriate experimental details and sample information for the mRNA CQA data within this application note can be found in the following Waters application notes: mRNA mapping data (Gaye *et al*), 3' poly(A) Tail data (Doneanu *et al*), and the 5' cap data (Ngyuen *et al*).^{13,14,18}

Data processing workflow example:

N,N-diisopropylethylamine (DIPEA, 99.5% purity, catalogue number 387649-100ML), 1,1,1,3,3,3-hexafluoro-2propanol (HFIP, 99% purity, catalogue number 105228-100G), ethanol (HPLC grade, catalogue number 4598282L) and ammonium bicarbonate (LiChropur LC-MS Supelco reagent, catalogue number 5330050050) were purchased from Millipore Sigma (St Louis, MO). Acetonitrile (LC-MS grade, catalogue number 34881-1L) and methanol (LC-MS grade, catalogue number 34966-1L) were 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 digestions was purchased from Thermo Fisher Scientific (Waltham, MA).

100-mer sgRNA was purchased from IDT (Integrated DNA Technologies, Coralville, IA) and reconstituted in 100 µL of nuclease-free water. The sgRNA contained six stability-enhancing modified nucleotides at terminal positions 1, 2, 3, and 97, 98, 99 in the sequence (Figure 1A). The modified nucleotides included were 2'O-methylguanosine, 2'O-methyluridine, and 2'O-methyladenosine and each modification had a thioated phosphate group (Figure 1B). RNase T1 ribonuclease was purchased from Worthington Biochemical Corporation (Lakewood, NJ). For sgRNA digestion, 5 µL of reconstituted sgRNA were mixed with 1 µL digestion buffer (100 mM ammonium bicarbonate), 43 µL of nuclease-free water and 1 µL of the 1 mg/mL RNase T1 solution (prepared in nuclease-free water). The digestion mixture was prepared in a QuanRecovery[™] MaxPeak[™] 300 µL vial and incubated at 37 °C for 15 minutes, then immediately analyzed by LC-MS.

A

#:JGAUCUCUCAACUUUAACGUUUUAGAGCUAGAAAUAGCAAGUUAAAAU AAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCJJJ

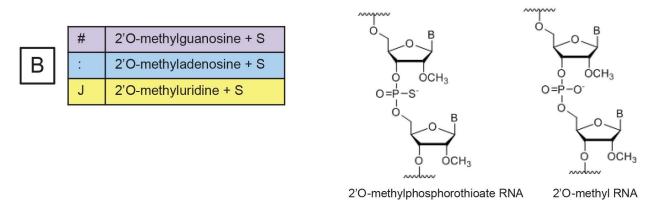


Figure 1. The sequence of the sgRNA analyzed. The '#' represents 2'O-methylguanosine, the ':' represents 2'Omethyladenosine, and the 'J' represents 2'O-methyluridine. All the modified nucleotides included a thioated phosphate group. All datasets were acquired within the waters_connect informatics platform (version 3.2.0) using the UNIFI[™] application (version 3.6.0) and processed using the INTACT Mass application (version 1.6.0). Versions 1.0.0 of the mRNA Cleaver and 1.0.0 of the Coverage Viewer microApps were used.

LC Conditions

LC-MS System:	BioAccord LC-MS System with ACQUITY [™] Premier UPLC [™] (Binary)
Pre-Column:	VanGuard FIT cartridge holder (p/n: 186007949) containing a 2.1 x 5 mm ACQUITY Premier FIT cartridge packed with 1.7 µm BEH™ C18 particles. (p/n: 186009459)
Column:	ACQUITY Premier OST Column 1.7 μm, 130 Å, 2.1 x 150 mm, p/n: 186009486
Column temperature:	60 °C
Flow rate:	300 µL/min
Mobile phases: Solvent A:	8 mM DIPEA (N,N-diisopropylethylamine), 40 mM HFIP (1,1,1,3,3,3-hexafluoroisopropanol), in DI water, pH 8.8
Solvent B:	4 mM DIPEA, 4 mM HFIP in 75% ethanol
Sample temperature:	6 °C
Sample vials:	QuanRecovery MaxPeak HPS vials (p/n: 186009186)
Injection volume:	5 µL

Gradient Table for sgRNA experiments

Time (min)	Flow rate (mL/min)	Composition A (%)	Composition B (%)	Curve
0	0.3	100	0	Initial
45	0.3	85	15	6
50	0.3	85	15	6
50.5	0.3	15	85	6
51.5	0.3	15	85	6
52	0.3	100	0	6

MS Conditions

Acquisition mode:	Full scan
Ionization mode:	ESI (-)
Capillary voltage:	0.8 kV
Cone voltage:	45 V
Source temperature:	120 °C
Desolvation temperature:	500 °C
Desolvation gas (N_2) pressure:	6.5 bar
TOF mass range:	400-5000 <i>m/z</i>
Acquisition rate:	2 Hz
Lock-mass:	waters_connect lockmass solution (p/n: 186009298)

Informatics platform for data acquisition and	waters_connect v.3.2.0
processing:	
Data acquisition:	UNIFI App ver 3.6.0
Data processing:	INTACT Mass App ver 1.6.0

Results and Discussion

Data Processing Workflow for Oligo Mapping

LC-MS based oligo mapping techniques have been developed over the past decades to address the growing need to analyze nucleic acid CQAs.¹⁹⁻²³ Oligo mapping is analogous to peptide mapping as it provides the primary structure (the sequence) of the target nucleic acid through enzymatic digestion and accurate mass determination. Common enzymes used in oligo mapping include RNase T1, RNase A, and MazF. For the enzymatic digestions described here, we used RNase T1 which cuts the oligonucleotide after every guanosine on the 3' end. The mechanism of RNase T1 has been studied extensively and described previously.²⁴⁻²⁷ As we only used one enzyme, we did not expect 100% sequence coverage. Typically, an oligo mapping experiment would use multiple enzymes in parallel as each enzyme has a different specificity and would produce a different set of digestion fragments. By doing so, the resulting digestion fragments from each enzyme should overlap in sequence, providing better sequence coverage. However, to keep this application note concise, we limited our scope to one enzyme.

Another tactic in oligo mapping experiments is to intentionally digest the oligonucleotide partially by altering enzymatic digestion conditions. By using partial digestion, the nucleic acid digestion fragments would be longer on average and therefore, more unique in both sequence and mass. While this approach can improve coverage, partial digestion can be hard to control, adding complexity to an already complex sample through additional, possibly overlapping m/z peaks and coeluting incomplete digestion fragments. For the data shown here, our intention was to completely digest the RNA in question, and we used enzymatic conditions optimized to do so.

1. Generating in silico digestion components

The data processing workflow for oligo mapping and other nucleic acid CQAs is illustrated in Figure 2. The generation of the *in silico* digestion fragments was accomplished using the mRNA Cleaver microApp. The Graphical User Interface (GUI) of the microApp is shown as Figure 3A. The User Guide is located on the left-hand side of the GUI to improve the user experience with the software. The mRNA Cleaver microApp is a flexible digestion calculator and library generator for oligo mapping investigations.

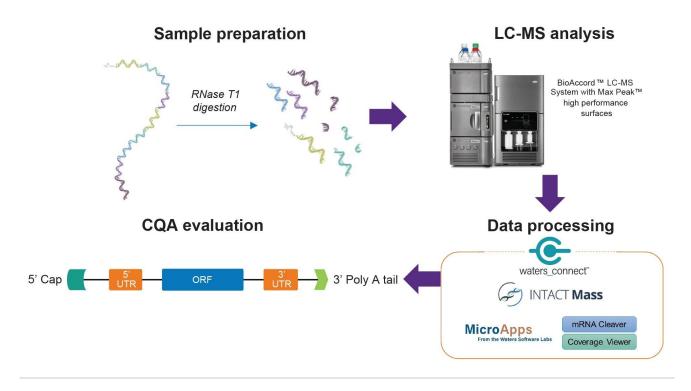


Figure 2. Workflow scheme for the characterization of RNA CQAs.

A user can modify three embedded files to customize the calculator to their specific needs. For example, if a sequence of interest contains modified nucleotides, the masses and chemical compositions can be added, as well as a representative single letter or symbol, to the "residues.csv" file, whose file path can be found in the User Guide. Using the representative letter or symbol of the modified nucleotide will prompt the program to use the adjusted masses and chemical composition to calculate the relevant digestion products. The "enzymes.csv" and "modifications.csv" files can similarly be modified to add different enzymes or mRNA-specific modifications (such as different 5' caps or 3' poly(A) distributions) to fit the user's purpose.

Digestion specific settings are located below the sequence field (Figure 3A) and include adjusting the digestion

product termini, whether a-specific digestion products should be generated, the number of missed cleavages to include, and whether mRNA specific modifications should be considered. Filters for the generated digestion product list are located on the right-hand side, where the generated digestion products are filtered based upon length settings. Additionally, the user can choose whether or not to check the box "Filter Duplicate Sequences". Duplicate sequences are digestion fragments that have the same sequence identity. By default, mRNA Cleaver will have that box checked which means that if there is a duplicate sequence it will be listed once with the alternate locations for that sequence listed in another column. For our workflow, the box is unchecked for use in Coverage Viewer in a subsequent data analysis step. MS specific settings are located below the filters, where the user can indicate the polarity of the expected mass ions, whether the generated list should reference monoisotopic or average masses, the charge state range to be generated for each digestion product, and the *m/z* range to be generated. The digestion product list is generated when the user clicks "Calculate", and the list is output as a comma delimited file (.csv). An example output is shown in Figure 3B. This .csv file will be imported to the INTACT Mass as targeted search components processing.

	Digestion Settings:	Filters:			
mRNA Cleaver 1.0.0	Sequence	Min Length		Max Length	
An in-silico digestion calculator and library generator for Oligo Mapping of mRNA.	#:JGAUCUCUCAACUUUAACGUUUUAGAGCUAGAAAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAAC UUGAAAAAGUGCCACCGAGUCGGUGCLJJU	3	- +	200	
	000444440000440000000000000000000000000	Min a-specific Length		Max a-specific Length	
User guide	Enzyme	3	- +	200	-
See user guide	RNase T1 -	Filter Duplicate Sequences			
Overview	Terminus	MS Settings:			
Enter an mRNA sequence into the Sequence text box, select an enzyme	Enzyme default terminus +			Mass	
from the drop-down and click the	a-specific digestion	O Negative O Positive		O Monoisotopic 🔿 Average	
Calculate button to download a csv file containing predicted digestion	Missed Cleavages	Min Charge		Max Charge	
products and the following properties:	2 - +	1	+	10	
 Item Name: The sequence of the digestion product suffixed with "p" 	Modification	Min m/z		Max m/z	
or "cP" to indicate a linear phosphate or cyclic phosphate	None -	100	- +	4000	
terminus and a modification prefix/suffix where appropriate.	Calculate				
 Formula: The elemental composition of the digestion 		#	2'0-	methylguanosine + :	S
product.			_	methyladenosine +	
 Location: Location in the original sequence of the first occurrence of 		:		methyluridine + S	2

Figure 3. (A) The mRNA Cleaver microApp User Interface (GUI).

A	8	c	D	E	F		H	1	J	K	L	M	N	0	P	Q	R	5
1 Item Name	Formula	Location	Tag "Fragment#"	Tag "Precursor"	Additional	neutral mass (Monoisotopic)	start	end	lengt	1 1·	2-	3-	4-	5-	6-	7-	8-	9-
2 #JGp	C42H55N17O26P4S3	#1:G4	R1	R1	-	1433.161675	1	1	4	4 1432.	715.57	476.71	-	-	-	-	-	- 1
3 AUCUCUCAACUUUAACGp	C159H200N57O121P17	A5:G21	R2	R2	-	5369.678846	5	5 3	21	17 -	2683.8	1788.9	1341.4	1072.9	893.9	766.1	670.2	2 595
4 UUUUAGp	C56H70N18O46P6	U22:G27	R3	R3	-	1916.211727	22	2 3	27	6 1915.	957.1	637.73	478.05	382.24	-	•	-	-
5 CUAGp	C38H49N15O29P4	C30:G33	R5	R5	-	1303.177108	30) :	33	4 1302.	650.58	433.39	- 1	-	-	-	-	-
6 AAAUAGp	C59H73N27O40P6	A34:G39	R6	R6	-	1985.293381	34	1 3	39	6 1984.	991.64	660.76	495.32	396.05	-	-	-	-1
7 CAAGp	C39H50N18O27P4	C40:G43	R7	R7	-	1326.204326	40) 4	43	4 1325.	662.09	441.06	-	-	-	-	-	-
8 UUAAAAUAAGp	C97H119N41068P10	U44:G53	R8	R8	-	3255.449025	44	1 !	53	10 3254.	1626.7	1084.1	812.85	650.08	541.6	464.1	405.9	360
9 CUAGp	C38H49N15O29P4	C55:G58	R10	R10	-	1303.177108	55	5 !	58	4 1302.	650.58	433.39	-	-	-	-	-	-
10 UCCGp	C37H49N13O30P4	U59:G62	R11	R11	-	1279.165875	59	9 (52	4 1278.	638.58	425.38	-	-	-	-	-	-
11 UUAUCAACUUGp	C103H129N36O80P11	U63:G73	R12	R12	2	3490.424643	63	3	73	1 3489.	1744.2	1162.5	871.6	697.08	580.7	497.6	435.3	3 386
12 AAAAAGp	C60H74N30O38P6	A74:G79	R13	R13	2	2008.320599	74	1	79	6 2007.	1003.2	668.43	501.07	400.66	2	-	-	20
13 CACCGp	C47H62N19O35P5	C83:G87	R16	R16	2	1607.23438	83	3 1	37	5 1606.	802.61	534.74	400.8	-	2	-	-	20
14 UCGp	C28H37N10O23P3	U90:G92	R18	R18	-	974.124588	90	9	92	3 973.1	486.06	-	-	-	-	-	-	-
15 CJJJU	C48H63N11O34P4S3	C96:U100	R21	R21	-	1557.165149	96	5 10	00	5 1556.	777.58	518.05	388.28	-	-		-	

Figure 3. (B) Example output from mRNA Cleaver.

2. Oligo Fragment Mass Confirmation

Once the *in silico* digestion product list is generated and the experimental data is collected on the digested sample, the next step in the workflow is to process the data through the INTACT Mass App within waters_connect. The output from the mRNA Cleaver microApp provides the expected masses of the digestion products and their molecule IDs which can be copied directly into the "Expected Masses" column within the application. Modifiers can be added to include common adducts and sequence modifiers. Figure 4A contains the results screen after the analysis has finished processing while Figure 4B illustrates a representative deconvoluted spectrum output. The INTACT Mass results, MS, UV, and TIC tables can be downloaded using the download button on the right-hand side of the table.

sgRNArerun_04	Sample list	030223_C	D_8mM_DI	PEA_VanGuard	6uM_1AA_sg	RNA_digest_45m	in_5uL_01				View 54	ample method	Vew report
View dashboard	030223_CD_8mM_DIPEA_VanGuard 600 ¥ 64M_1AA.sgRNA_digest_45min_SuL _01 Replikates: 1/1 Sample position: 1:E,2	TIC (1997 - 2017 - 107 -		4 ¢ 2 1 1 • 2 Å	Peak 6 🖬 7	10 □ 14 ■ 0 ◆ 12 ◆ △ 9 0 13 ●	19 Å 18 Å 18 Å	26 1 25 0 23 A 24 0	30 0 39 🛦 43 🖌			,Åv,	
View SSTs and QCs	030225_CD_8mM_DIPEA_VanGuard_ (Fol) & 6uM_1AB_sgRNA_digest_45min_5uL _01	0			10	15	20	25	30 Recention Time (min)	ži do		so sis	(a):-
View blank results	Replicates: 1/1 Sample position: 1:E,1	TUV 260											
View Pass results	030223.20.8mth.DIPEA.VanGuard. Em & GudL1AE.gRNA.digest_45min_SuL_01 Replicates: 1/1 Sample postion: 1:E,3	About		ļ	iò.	k	20	2	」 Retention Time (min)	i i	å. Raterior	so sis	2: xce (ALI): -
View Warning results		Results	TIC	TUV 260 Purity:	MS								8 I
View Fail results		Y	Type *	Molecule ID	n F	Component	4. 4. 7	Observed mass (Da)	 Expected mass (Da) Y 	% Mass error (ppm) % ↓	v (mins)	T % Observed UV RT y (mins)	t LC area
View untargeted		2 Pro		\$1:G4	#1:G4 A5:G21		△	1,433.151				5.71	115,2
results		4 Pro	oduct (J22.G27	U22:G27				1,916.212	Fail			
Create summary report		5 Pro	oduct	C30:G33	C30:G33	All Forms				Pas			

Figure 4. (A) The results dashboard of the INTACT Mass Application.

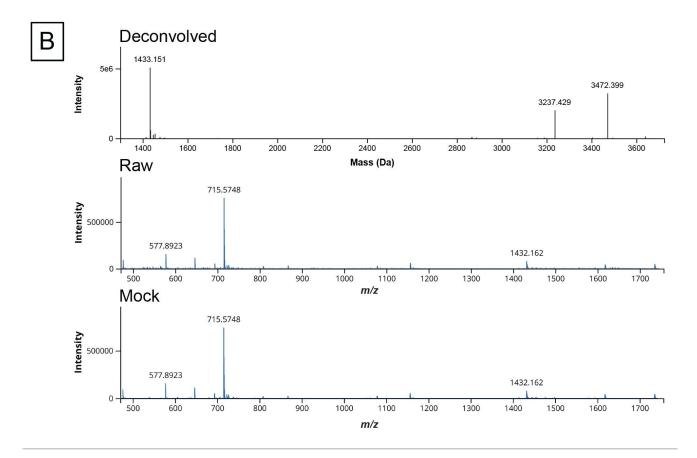


Figure 4. (B) An example of the deconvoluted spectra output.

While the tabular data produced by the INTACT Mass App is the key result for reporting fragment mapping data, the Coverage Viewer microApp can be used for visualization of these results against the sequence, and calculation of sequence coverage. The GUI of the Coverage Viewer microApp is featured in Figure 5. The sequence of the RNA molecule of interest is entered into the sequence field and the mRNA Cleaver output generated in the first step of the workflow can be imported to populate the target digest fragment list (bottom left-hand corner). The INTACT Mass results file can be imported using the "Import" button on the left-hand side. The sequence coverage is visualized on the right-hand side of the GUI and the coverage percentage is calculated as targets are confirmed. If the INTACT Mass results file is imported, digestion fragments that have the keyword "Pass" will be considered "Confirmed". Manually assigned products can also be checked off by the user to be considered "Confirmed" as well. The red colored area under the relevant sequence will change to green to indicate that specific part of the sequence has been confirmed by accurate mass measurement. Once finished, the user can then export the target digested fragment section with the "Confirmed" column as a .csv for

later use. Additionally, the coverage map image can be copied by clicking "Capture image to clipboard" and saved in another program.

quence:									
GAUCUCUCZ	ACULUAA	CGUUUUAGA	AGCUAGAAAUAGCAAGUUAAAAUAAGG	UAGUCCGUUAUCAACUUK	GAAAAAG	UGGCA	CEGAGUEGGUGEIUU		Coverage: 91.00% Capture image to
GROCOCOCA		~~~~~			0.000010	U GOLA			• ; 2 GAU CUC UCA CU UVA ACE UVU VAG ACE VAG AAA VAG CAA UVU AAA AVA AGE CUA GUC COU VAU CAA CUU GAA AAA GUG GCA CGG VGC 3J3 U
Load target lis	t C:\Use	rs\usrdeso\C	OneDrive - Waters Corporation\Documer	nts\projects\CONFIRM-sequ	uencesupp	port/rot	ntestcase\1aa_2missed.csv		
			-	nts\projects\CONFIRM-sequ	uencesupp	port'y ot	ntestcase\1aa_2missed.csv	Clear all confirmations	
onfirmations	Import.			nts\projects\CONFIRM-sequ			otestcase\1aa_2missed.csv		
onfirmations	Import.	Export. Description			Fragme		rt End Average neutral mas	s Monoisotopic neutra 6784.8299569	
Confirmations	import.	Export. Description	Item name #JGAUCUCUCAACUUUAACGp #JGAUCUCUCAACUUUAACGUUUUAGp	Formula C201H253N74O146P215 C257H321N92O191P275	Fragme 3 R1-2 3 R1-3	nt Sta 1 1	rt End Average neutral mas 21 27	s Monoisotopic neutra 6784.8299569 8683.0311199	
confirmations	: Import. ocation 1 :G21 1:G27 1:G4	Export. Description	Item name #JGAUCUCCAACUUUAACGp #JGAUCUCCCAACUUUAACGUUUUAGp #JGp	Formula C201H253N74O146P215: C257H321N92O191P275: C42H55N17O26P4S3	Fragme 3 R1-2 3 R1-3 R1	nt Sta 1 1	rt End Average neutral mas 21 27 4	s Monoisotopic neutra 6784.8299569 8683.0311199 1433.1616749	
Confirmations	import.	Export.	Item name #JGAUCUCUCAACUUUAACGp #JGAUCUCUCAACUUUAACGUUUUAGp #JGp AGCUAGp	Formula C201H253N74O146P21S3 C257H321N92O191P27S3 C42H55N17O26P4S3 C58H73N25O42P6	Fragme 3 R1-2 3 R1-3 R1 R4-5	nt Sta 1 1 1 28	rt End Average neutral mas 21 27 4 33	s Monoisotopic neutra 6784.8299569 8683.0311199 1433.1616749 1977.277063	
Confirmations	cation :G21 :G27 :G4 28:G33 28:G39	Export. Description	Item name #JGAUCUCUCAACUUUAACGp #JGAUCUCUCAACUUUAACGUUUUAGp #JGp AGCUAGA AGCUAGAAUAGp	Formula C201H253N74O146P2153 C257H321N92O191P2753 C42H55N17026P453 C58H73N25O42P6 C117H144N52O81P12	Fragme 3 R1-2 3 R1-3 R1 R4-5 R4-6	nt Sta 1 1 1 28 28	rt End Average neutral mas 21 27 4 33 39	Monoisotopic neutra 6784.8299569 8683.0311199 1433.1616749 1977.277063 3944.559880000001	
Confirmations	Import. 000000000000000000000000000000000000	Export. Description	Item name #JGAUCUCUCACUUUAACGP #JGAUCUCUCACUUUAACGP #JGp AGCUAGAAAUAGP AGCUAGAAAUAGP AAUAGP	Formula C201H253N74O146P2153 C257H321N92O191P2753 C42H55N17026P453 C58H73N25O42P6 C117H144N52081P12 C59H73N27O40P6	Fragme 3 R1-2 3 R1-3 R1 R4-5 R4-6 R6	nt Sta 1 1 28 28 34	rt End Average neutral mas 21 27 4 33 39 39	Monoisotopic neutra 6784.8299569 8683.0311199 1433.1616749 1977.277063 3944.559880000001 1965.293381	
onfirmations Confirmed La # 2 # 2 # 2 # 2 # 2 # 2 # 2 # 2 # 2 #	Import. 0cation 1:G21 1:G27 1:G4 28:G33 28:G39 34:G39 34:G39 34:G43	Export.	Item name #JGAUCUCUCAACUUUAACGp #JGAUCUCUCAACUUUAACGp #JGAD AGCUAGAACG AGCUAGAAUAGp AAAUAGp AAAUAGp AAAUAGp	Formula C201H253N740146P215: C257H321N920191P275; C24H55N17026P453 C58H73N25042P6 C117H144N52081P12 C98H73N27040P6 C98H121N45066P10	Fragme 3 R1-2 3 R1-3 R1 R4-5 R4-6 R6 R6-7	nt Sta 1 1 28 28 34 34	t End Average neutral mas 21 27 4 33 39 99 39 43	 Monoisotopic neutra 6784.8299569 8683.0311199 1433.1616749 1977.277063 3944.559880000001 1985.293381 3293.4871430000007 	
Confirmed La	Import. Cation I:G21 I:G27 I:G4 28:G33 28:G39 34:G39 34:G53	Export.	Rem name #JGAUCICUCAACUUUAACGp #JGAUCICUCAACUUUAACGUUUIAGG #JGG AGCUAGA AGCUAGAAUAGG AAGUAGA AAAUAGCAAGG AAAUAGCAAGG	Formula C01H253N740146P215; C257H321N920191P275; C42H55N17026P453 C58H73N25042P6 C171H144N52081P12 C59H73N27040P6 C98H211N45066P10 C195H231N4506P10	Fragme 3 R1-2 3 R1-3 R1 R4-5 R4-6 R6 R6 R6-7 R6-8	nt Sta 1 1 28 28 34 34 34 34	t End Average neutral mas 21 27 4 33 39 39 39 39 53	 Monoisotopic neutra 6784.8299569 8683.0311199 1433.1616749 1977.277063 3944.559880000001 1985.293381 3293.4871430000007 6530.925604000001 	
onfirmations Confirmed La # # Q # Q # Q # Q # Q # Q # Q # Q A Q A Q A A A Q A A A Q A	import. ccation 1 I:G21 I:G27 I:G4 28:G33 28:G39 34:G39 34:G39 34:G33 34:G53 5:G21	Export.	Item пате #IGAUCIUCICAACUUUAACGp #IGAUCIUCICAACUUUAACGUUUUAG #IGA AGCUAGA AGCUAGA AGCUAGA AAUUAGCAG AAAUUGCAAGUUAAAUUAGG AAUUGCCAAGUUAAAGUAAAG	Formula C201H253N740146P215: C257H221N920191P275; C42H55N17026P453 C58H73N25042P6 C117H144N52081P12 C99H73N27040P6 C99H121N45068P10 C195H230R960133P20 C159H230R570121P17	Fragme 3 R1-2 3 R1-3 R1 R4-5 R4-6 R6 R6-7 R6-8 R2	nt Sta 1 1 28 28 34 34 34 5	Average neutral mas 21 27 4 33 39 39 39 39 31 32 33 34 35 32	s Monoisotopic neutral 6784.8299569 6883.0311199 1433.1616749 1977.277063 3944.559880000001 1985.293381 3293.4871430000007 6530.92564000001 53369.5678846000001 5369.678846000001	
confirmations Confirmed La #**	import. cation i isc21 isc27 isc4 28:633 28:639 24:639 24:639 24:633 34:653 35:621 35:627	. Export. Description	Item name #ISANJCUCICACUUUAACGp #ISGA #ISGA ACCUNGA ACCUNGA ACCUNGAANUAGP AAAUAGCASUUAAANAAGP AAAUAGCASUUAAANAAGP AUAUCICACUUUAACUUAAGGUUUAAGP	Formula C201H253N74O146P215: C257H321N92O191P275; C42H55N17026P453 C58H73N25042P6 C117H144N52081P12 C59H72N42504P6 C195H22N4506F10 C195H228N86O133P20 C159H220N570121P17 C159H220N570121P17	Fragme 3 R1-2 3 R1-3 R1 R4-5 R4-6 R6 R6-7 R6-8 R2 R2 R2-3	nt Sta 1 1 28 28 34 34 34 5 5 5	rt End Average neutral mas 21 27 4 33 39 39 39 43 53 21 27	s Moneisotopic neutra 6784.8299569 688.0311199 1433.1616749 1977.27063 3944.559880000001 1985.293381 3293.4871430000007 6530.925604000001 5369.678846000001 5369.678846000007 7267.8800000	
Image: Confirmations Confirmed Ltt	import. cation i isc21 isc27 isc4 28:633 28:633 28:639 34:633 34:633 5:621 5:627 5:629	Export.	Item name #JGAUCUCUCACUUUAACGP #JGAUCUCACUUCAACUUUACGP #JGA ACUMCANAGO AAUNGCANGO AAUNGCANGU AAUNGCANGUAACUUUAGG AUNCUCCACUUUAACGUUUAGG AUCUCCACUUUAACGUUUAGG	Formula C201H253N740146P215: C257H221N920191P275; C24H55N17026P433 C58H73N25042P6 C1117H144N52081P12 C59H73N27040P6 C99H721N4506P10 C159H220N570121P17 C159H220N570121P17 C159H220N57012P17	Fragme 3 R1-2 3 R1-3 R1 R4-5 R4-6 R6 R6-7 R6-8 R2 R2-3 R2-4	nt Sta 1 1 1 28 28 34 34 34 5 5 5 5	tt End Average neutral mas 21 27 4 4 39 39 39 43 53 53 21 27 27 27 28	s Monoisotopic neutra 6784.8299569 - 8683.0311199 - 1433.1616749 - 1977.277063 - 3944.5598800000001 - 1985.293381 - 3293.487743000007 - 6530.925604000001 - 5369.67884600001 - 7267.880009 - 7941.979964	
Image: Second	import. ccation 1 isG21 isG21 isG27 isG4 28:G33 28:G39 34:G53 34:G53 34:G53 35:G27 5:G27 5:G29 74:G79	Export Description	Item name #ISANJCUCICACUUUAACGp #ISGA #ISGA #ISGA ACCUAGA ACCUAGA ACCUAGAAUNAGA AAAUAGCA AAAUAGCAAGUUAAAANAAGP AAAUAGCAAGUUAAAANAAGP AUXCUCCACUUUAACGUUUHAGA AUXCUCCACUUUAACGUUUHAGA	Formula C201H253N74O146P215: C257H231N92O191P275: C24F5SN17026P453 C59H72N25O42P6 C179H14N52081P12 C59H72N82C66P10 C159H220N55C013P20 C159H220 C159H20 C159H220 C159H20	Fragme 3 R1-2 3 R1-3 R1 R4-5 R4-6 R6-7 R6-8 R6-7 R6-8 R2 R2-3 R2-4 R13	nt Sta 1 1 28 28 34 34 34 34 5 5 5 74	rt End Average neutral mas 21 27 4 33 39 39 39 43 43 53 21 27 27 29 79	 Monoisotopic neutra 6784.8299569 6883.0311199 1433.1616749 1977.277063 3944.559880000001 1965.293381 3293.4871430000007 6530.02560400001 7267.880009 7941.797964 2008.320599 	
Image: Second	import. ccation 1 isG21 isG21 isG27 isG4 28:G33 28:G39 24:G39 24:G33 34:G53 35:G21 35:G27 74:G79 74:G81	Export Description	Item name escalucion/caculuua/cgp escalucion/caculuua/caculuua/caculuua/caculuua/caculuua/caculuua/caculuua/caculuua/caculuua/caculuua/caculuua/caculuua/caculuua/caculua/caculuua/caculua/caculuua/caculuua/caculua/caculua/caculua/caculu	Formula C20114253N740146P215 C257H231N740146P215 C2457H21N820191P2C C44555N170204P65 C44555N17024046 C39H712N74046 C39H712N74046 C39H712N74046 C39H712N74046 C39H712N74067314 C159H22N856013920 C459H24000367 C159H24000366 C459H2400364 C459H240004 C459H240000000000000000000000	Fragme 3 R1-2 3 R1-3 3 R1 R1 R4-5 R4-6 R6 R6 R6-7 R6-8 R2 R2 R2-3 R2-4 R13 R13-14	nt Sta 1 1 28 28 34 34 34 5 5 5 5 74 74 74	t End Average neutral mas 21 27 4 33 39 39 39 39 21 27 29 79 81	 Monoisotopic heutra 6784.8299569 8683.0311199 1433.1616749 1977.277063 3944.559880000001 1985.293381 3293.4871.430000007 6530.92560400001 5369.678846000001 5369.678846000001 7367.880009 741.979964 2008.320599 2659.3933360000005 	
Image: Second	import. ccation i i:521 i:527 i:54 28:633 28:639 34:633 34:633 34:653 35:621 35:621 35:627 35:629 74:669 74:6681 74:682	Export.	Item name #SAUCUCICACUUUAAGp #SAUCUCICACUUUAAGU #Saucucicacuuuagu #SAUCUCICACUUUAAGU #Saucucicacuuuagu #SAUCUCICACUUUAAGU #Saucucicacuuuagu #Saucucicacuuuagu #Saucucicacuuuagu #Saucucicacuuuagu #Saucucicacuuuagu #Saucucicacuuuagu #Saucucicacuuuagu #Saucucicacuuuagu #Saucucicacuuuagu #Saucucicacuuuuagu #Saucucicacuuuugu #Saucucicacuuuuagu #Saucucicacuuuugu #Saucucicacuuuugu #Saucucicacuuuuuuuuuuuuuuuuuuuuuuuuuuuuuu	Formula C20142530740146P215 C2574231740201912757 C44555117026453 C5487382504296 C1974144452058192 C59472384205492 C59472384205492 C594724384205492 C15942084750121917 C215442084750121917 C215442084750121917 C215442084750121917 C215442084750121917 C215442084750121917 C215442084750121917 C215442084750121917 C215442084750121917 C215442084750121917 C215442084750121917 C215442084750121917 C21544208475012197 C2154420847501200000000000000000000000000000000000	Fragme 3 R1-2 3 R1-3 3 R1 R1 R4-5 R4-6 R6 R6 R6-7 R6-8 R2 R2-3 R2-3 R2-4 R13 R13-14 R13-15	nt Sta 1 1 1 28 28 34 34 34 34 34 5 5 5 5 74 74 74	rt End Average neutral mas 21 27 4 33 39 39 43 53 21 27 27 29 27 29 27 29 27 29 53 21 27 27 29 53 21 27 27 28 53 27 27 28 27 28 28 28 28 28 28 28 28 28 28	Moncisotopic neutra 678-48,299569 8683,0311199 1433,161749 1977,277063 394-4559800000001 1985,293381 2823,4877430000007 6530,925644000001 7267,880009 7267,980009 7267,98000005 7267,98000005 7267,98000005 7267,98000005 7267,98000005 7267,98000005 7267,98000005 7267,98000005 7267,98000005 7267,98000005 7267,9800005 7267,9800005 7267,9800005 7267,9800005 7267,9800005 7267,9800005 7267,9800005 7267,9800005 7267,9800005 7267,9800005 7267,9800005 7267,9800005 7267,9800005 7267,9800005 7267,9800005 7267,9800005 7267,9800005 7267,9900005 7267,9900005 7267,9900005 7267,9900005 7267,9900005 7267,9900005 7267,9900005 7267,9900005 7267,9900005 7267,900005 7267,900005 7267,900005 7267,900005 7267,900005 7267,900005 7267,90	
Confirmations Confirmed # ## # 2 # A2 A2 A2 A2 A2 A2 A2 A2 A3 A2 A4 A2 A4 A4	import. cation 1:521 1:527 1:54 2:8:633 2:8:639 3:4:639 3:4:643 3:4:643 3:4:643 3:4:653 3:527 3:529 7:4:679 7:4:681 3:8:692	Export.	Ітет нате вісанисисисансициансор вісанисисисансициансор вісански санкина вісански санкина вісански санкина вісански санкина вісански санкина вісански санкина вісански санкинанськи вісански санкинанськи вісански санкинанськи вісански санкинанськи вісански санкина вісански санкина вісански санкина вісански санкина вісански санкинанськи вісански сан	Formula C201142307402148P215 C201142307402148P215 C201123742214020191P275 C2011237421020492 C20112144452008192 C201121445006910 C201121445006910 C20112145006910 C20112145006910 C20112145006910 C20112145006910 C201121450003895	Fragme 3 R1-2 3 R1-3 R1 R4-5 R4-6 R6 R6-7 R6-8 R2 R2-3 R2-4 R13 R13-14 R13-15 R17-18	nt Sta 1 1 28 28 34 34 34 34 5 5 5 5 74 74 74 88	Average neutral mas 21 27 4 33 39 31 32 33 24 25 26 27 4 33 33 34 35 37 38 39 39 31 21 22 31 32 33 34 35 36 37 38 39 39 31 32 33 34 35 36 37 38 39 39 31 32 33 34 35 36 37	4 Monoisotopic neutre 6784,8299569 8683,011199 1433,1616749 1972,27063 3944,57986000001 1982,23381 3293,4871430000001 7367,88000 7441,379964 2003,32599 2003,32599 3004,440771000001 1646,22454 3004,44077100001	
Confirmations Confirmed # # # # # # # # # # A:	import. cation i:G21 i:G27 i:G4 28:G33 28:G39 24:G39	Export.	Nminume	Formula C01142331740146P2155 C42H5591710264453 C42H5591710264453 C43H73812504296 C43H73812504296 C43H74845204976 C43H74845204976 C43H428478450139270 C43H4284785015646P23 C43H4284780515466P23 C43H4284780515466P23 C43H4148203495 C43H4148203495 C43H4148203495	Fragme 3 R1-2 3 R1-3 R1 R4-5 R6- R6- R6- R6- R2-3 R2-4 R13 R13-14 R13-15 R17-18 R17-19	nt Sta 1 1 28 28 34 34 34 34 34 34 34 34 34 34 34 34 34	End Average neutral mas 21 21 27 4 33 39 43 53 27 27 43 21 79 61 62 29 93 21	Monolisotopic neutra (778.429559 868.301119 1433.1616749 1197.27063 394.455980000001 1985.23381 235.487143000007 653.92564000001 7267.88009 7941.97964 2008.32599 2659.33336000005 300.4440771000001 1648.22483	
Confirmations Confirmed Ir #" #" #" #" A: A:	import. oction 1 IG21 IG21 IG24 IG4 IG4 IG4 IG39 IG4 IG4 IG4 IG4 IG4 IG4 IG4 IG4	Export.	Item name escalucio(CAACUUUAACGP) escalucio(CAACUUUAACGP) escalucio(CAACUUUAACGP) escalucio(CAACUUUAACGP) escalucio(CAACUUAACGP) escalucio(CAACUUAAC	Formula C201142307400146P215 C401459117026445 C401459117026445 C491734220409 C117811446520819 C491734220444520819 C491734220449 C491734220449 C4917342049 C4917442049 C4917444444049 C491744444444444444444444444444444444444	Fragme 3 R1-2 3 R1-3 R1 R4-5 R4-6 R6-7 R6-8 R2 R2-3 R2-4 R13 R13-14 R13-15 R17-18 R17-19 R5	nt Sta 1 1 28 28 34 34 34 34 5 5 5 74 74 88 88 30	t End Average neutral mas 21 27 4 39 99 99 99 90 91 91 92 91 92 92 92 93 93 94 94 95 95 95 95 95 95 95 95 95 95	Monolisotopic neutra (774.429569 888.0.01199 1433.161749 1977.27763 1944.53980000001 1995.23381 1239.4714.3000007 7267.80009 7267.80009 7267.80009 7267.80009 7267.80009 7364.079964 2008.320599 2008.40771000001 1648.22453 1993.21978 1303.177708	
Confirmations Confirmed Is # # # # # # # # # # # # #	import. cation i:G21 i:G27 i:G4 28:G33 28:G39 24:G39	Export.	Nminume	Formula C01142331740146P2155 C42H5591710264453 C42H5591710264453 C43H73812504296 C43H73812504296 C43H74845204976 C43H74845204976 C43H428478450139270 C43H4284785015646P23 C43H4284780515466P23 C43H4284780515466P23 C43H4148203495 C43H4148203495 C43H4148203495	Fragme 3 R1-2 3 R1-3 R1 R4-5 R6- R6- R6- R6- R2-3 R2-4 R13 R13-14 R13-15 R17-18 R17-19	nt Sta 1 1 28 28 34 34 34 34 5 5 5 74 74 88 88 30	End Average neutral mas 21 21 27 4 33 39 43 53 27 27 43 21 79 61 62 29 93 21	Monolisotopic neutra (778.429559 868.301119 1433.1616749 1197.27063 394.455980000001 1985.23381 235.487143000007 653.92564000001 7267.88009 7941.97964 2008.32599 2659.33336000005 300.4440771000001 1648.22483	
onfirmations Confirmed #"	import. oction 1 IG21 IG21 IG24 IG4 IG4 IG4 IG39 IG4 IG4 IG4 IG4 IG4 IG4 IG4 IG4	Export.	Item name escalucio(CAACUUUAACGP) escalucio(CAACUUUAACGP) escalucio(CAACUUUAACGP) escalucio(CAACUUUAACGP) escalucio(CAACUUAACGP) escalucio(CAACUUAAC	Formula C201142307400146P215 C401459117026445 C401459117026445 C491734220409 C117811446520819 C491734220444520819 C491734220449 C491734220449 C4917342049 C4917442049 C4917444444049 C491744444444444444444444444444444444444	Fragme 3 R1-2 3 R1-3 R1 R4-5 R4-6 R6-7 R6-8 R2 R2-3 R2-4 R13 R13-14 R13-15 R17-18 R17-19 R5	nt Sta 1 1 28 28 34 34 34 34 34 34 5 5 74 74 74 88 88 80 30	t End Average neutral mas 21 27 4 39 99 99 99 90 91 91 92 91 92 92 92 93 93 94 94 95 95 95 95 95 95 95 95 95 95	Monolisotopic neutra (774.429569 888.0.01199 1433.161749 1977.27763 1944.53980000001 1995.23381 1239.4714.3000007 7267.80009 7267.80009 7267.80009 7267.80009 7267.80009 7364.079964 2008.320599 2008.40771000001 1648.22453 1993.21978 1303.177708	

Figure 5. The Coverage Viewer microApp User Interface.

3. Test case: sgRNA digestion and mapping

The CRISPR-based gene editing process has become popular in recent years, largely due to its simplicity.¹⁶ The process uses two main components, guide RNA and the CRISPR-associated (Cas) nuclease.¹⁶ The guide RNA is a specific RNA sequence that recognizes the target DNA of interest and directs the Cas nuclease to the correct site for editing .¹⁶ The Cas nuclease is a non-specific nuclease which depends entirely on the guide RNA to act upon a specific DNA locus.¹⁶ Originally, guide RNA was composed of two distinct oligonucleotides, the Crispr RNA, which is complementary to the target, and the tracr RNA, which served as a binding scaffold for the Cas nuclease.¹⁶ As the technology evolved the crispr RNA and the tracr RNA were fused together with a region called the "linker loop" and became single guide RNA or sgRNA.¹⁶ The key CQA for synthetically generated sgRNA is confirming its sequence, including the modified residues engineered into specific sites on the molecule.

By performing a RNase T1 digestion and analyzing the digestion products using our workflow, we can identify the digestion products and confirm, at the MS1 level (without additional MS fragmentation), the map coverage of the sgRNA sequence. First, the sgRNA modified nucleotides were added to the residues.csv file to allow mRNA Cleaver to correctly determine the *in silico* digestion products of the sgRNA. Then, we adjusted the digestion and MS settings of the mRNA Cleaver microApp to reflect the experimental conditions, specifically that RNase T1 was the enzyme used and that we would like mRNA Cleaver to generate digestion fragments with two missed cleavages. Even in a complete digestion, there is still a chance the enzyme will not cleave at every G and by increasing the missed cleavage variable to two, we account for that possibility. The in silico digestion product list was then generated and exported into a .csv file. Two ambiguous sequences, 4mer oligos with the sequence CUAG, were identified. Ambiguous sequences are digestion fragments that are isomeric and/or isobaric, which means the m/z peaks and charge state envelopes for these sequences will overlap. Ambiguous sequences that cover a portion of the RNA that is critical to analysis may require further experimental steps to confirm sequence identity such as fragmentation in the gas phase. Otherwise, the other 19 digestion fragments have unique expected masses and sequence identities, which should be discernable in our mass spectra. We base this assumption upon the expected mass accuracy generated by the BioAccord, and the *in silico* generated m/zrange for each digestion fragment from the mRNA Cleaver microApp.

The *in vitro* digestion was performed as described in the experimental methods and the data for both the digested sgRNA and the non-digested intact sgRNA was collected on the BioAccord LC-MS system using waters_connect. Figure 6A depicts the total ion chromatograms for the intact sgRNA and the subsequent digestion. The digested fragments eluted between five and 40 minutes of the gradient. The sgRNA has been completely digested as the intact sgRNA elutes at 35.6 minutes and there is no corresponding peak in the digestion sample trace.

The LC-MS data and the list of *in silico* digestion components were both submitted to the INTACT Mass App to assign the various digestion components based on a mass match with a 10 ppm match tolerance. We then imported the INTACT Mass results into the Coverage Viewer microApp to visualize the sequence coverage, shown in Figure 6B. We were able to achieve 91% MS1 sequence coverage. For this test case, we did not perform collision induced fragmentation to ascertain sequence identity as our ambiguous sequences were the same sequence identity so it would be superfluous. However, one may want to exclude ambiguous sequences from calculating sequence coverage when there are many ambiguous sequences or ambiguous sequences can be confirmed by fragmentation.²⁸ To achieve 100% sequence coverage, one could use another endoribonuclease such as MasF or RNase A that have differing cleavage specificity to RNase T1 to cover the unidentified portion of

the sgRNA sequence.

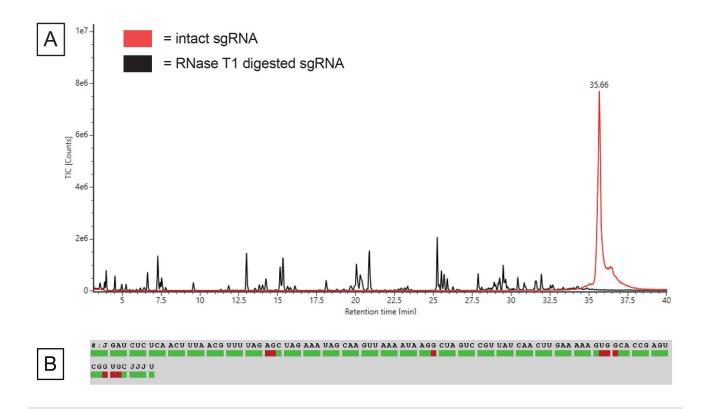


Figure 6. (A) TIC of sgRNA digested (black trace) and the undigested intact sgRNA (red trace), (B) Sequence coverage afforded by LC-MS via the Coverage Viewer microApp.

mRNA digestion and mapping

This workflow can be applied to larger RNAs, such as messenger RNA (mRNA). However, mapping larger RNAs at the MS1 level is ambitious as there will be many additional ambiguous sequences and coeluting species, exponentially increasing the data complexity. Gaye *et al* has previously released an application note regarding methodology to optimally perform mRNA sequence mapping using IP-RP LC-MS and the waters_connect platform.¹⁸ Here, we have reprocessed the data using the updated workflow.

Figure 7A contains the TIC chromatogram from the RNase T1 mRNA digest from Gaye *et al* as well as the Coverage Viewer results.¹⁸ The inset shows the extracted mass spectrum(XIC) from the peak eluting at 8.58 minutes. This representative XIC highlights the complexity of the data as multiple species are coeluting under the

peak. Following the workflow, the mRNA Cleaver microApp was used to generate two in silico enzymatic digestion product lists: one including the ambiguous sequences and one excluding them. These list and the acquired LC-MS data were submitted to the INTACT Mass App for data processing. The results from INTACT Mass were then imported into the Coverage Viewer microApp.

The results from INTACT Mass were then imported into the Coverage Viewer microApp. The MS1 level digestion fragment maps indicated 38.87% unique sequence coverage and 75.18% overall sequence coverage (ambiguous sequences included). These maps are illustrated in Figure 7B and 7C respectively. While this sequence coverage can be improved using another endoribonuclease in parallel, the complexity of the data (coeluting species, ambiguous sequences, etc.) will continue to be a challenge for data interpretation. Data processing time for these more complex samples scales with the number of peaks analyzed and can become limiting to analysis throughput.

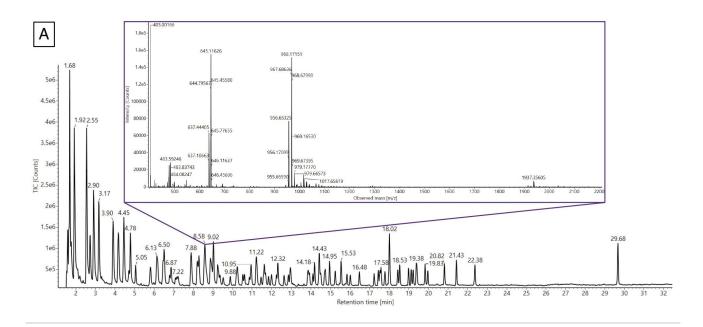


Figure 7. (A) TIC of the mRNA digest from Gaye et al.⁽¹⁸⁾ The inset spectrum is a representative XIC from the peak eluting at 8.58 min that contains several oligonucleotide species that are coeluting.

38.87% unique coverage at MS1 level

В

GG	G	AGA	ccc	AAG	CUU	GGU			cuc											GCC					UAC	UGC	GCC	ACU	GUU	CAU
																													AGC	UGU
			-	_		-	_	-	-	-	_	-	_	_			-			A.C.A.		-	_	_	-	_			AUU	GDJ
		100		ACG	AGA															ACA							0A0	GUA	AUG	GAP
UG	C 1	AGA	GUA	GAG	GCC				UUU									CAG			GAA	ccu	GGU	ACA	UAC	GUG	UUG	GGU	CAA	GGI
AC	c i	AAG	GGC	GGC	GAC																								GAA	GUC
GC	:U (GA	GAC	AUC	AUU			-								-				GGA	-						-		UAC	ACO
AU					ACC							CCA	GGC	UUC	AAC	AUC	ACC	GUC	AUU	GAG	UUC	UUC	AAA	CUG	AUC	GUG	AUC	GAC	AUC	CUC
																													GAU	
			-	-					_	_	_	-			-	-	_					_		_			_		GAU	-
-													_																	
GU																													GCC	
																													ACC	
		JUG	υсυ	GGA	CAU	ACU	uuc	UAC	GAU	ACA	ບບບ	GAC	ААА	GCA	AGA	UAC	САА	UUC	CAG	GGU	ccc	UGC	AAG	GAG	AUU	cuu	AUG	GCC	GCC	GA
																													CAA	
																													CAA	
AC	:U (UCC								ACU	UCC	AU
UA	C I																								САА	CUG	cuc	GUC	GUA	CA
AU	UU 2															_				UUC		_	-	_	UUU	GAU	GCU	GAA	GGA	GC
UG	U	GAU	CUG	ACC	ccc	AAC														CGA							GCC	GGU	CAA	AG
GA	υ																			UAC								CAG	GGU	υυ
UG	:U (_						GAA								GC		

Figure 7. (B) Unique sequence coverage (no ambiguous sequences included) afforded by MS1 via the Coverage Viewer microapp.

16

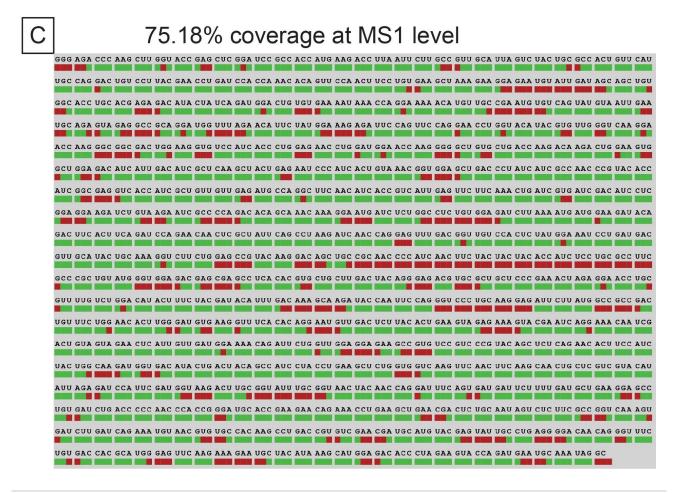


Figure 7. (C) Overall sequence coverage afforded by MS1 via the Coverage Viewer microapp.

Further sequence elucidation of ambiguous fragments can be done using the CONFIRM Sequence App if data independent (MSE) fragmentation data is collected as part of the mapping experiment on the BioAccord. An example of this type of analysis from Gaye *et al* is shown in Figure 8.¹⁸ Workflow automation of this additional stage of analysis is ongoing.

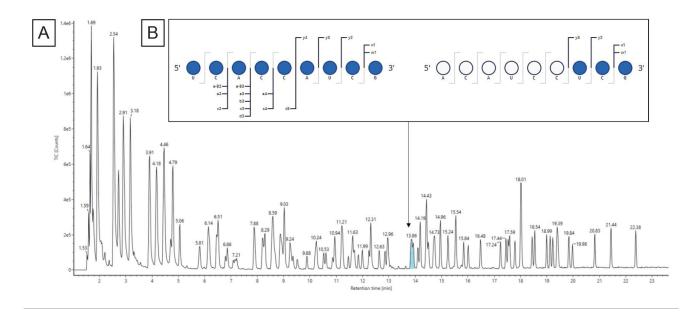


Figure 8. (A) Digested fragment components at position 623–631(ACAUCCUCGp) and 551–559 (UCACCAUCGp) are predicted and assigned to the same RT peak in the TIC from Gaye et al. It is not possible to determine the correct assignment using intact mass information. (B) MS^E data from the same injection can be used to elucidate the correct sequence for this assignment. Using the waters_connect CONFIRM Sequence application, high energy fragment ions are predicted 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.

Measuring 5' capping effectiveness

A CQA specific to mRNA is to assess the 5' capping efficiency and structure. The 5' cap of a mRNA molecule is essential for several biological functions including: regulation of nuclear export, prevention of degradation by exonucleases, promotion of translation, and promotion of 5' proximal intron excision.¹⁴ In a previous application note, Nguyen *et al.* described a method for the rapid analysis (under five minutes) of synthetic mRNA cap structure with the BioAccord LC-MS System.¹⁴

Using our workflow, we were able to detect and assign the 5' cap impurities based on the deconvoluted spectra as shown in Figure 9A. To analyze this data, we applied custom charge deconvolution, a new feature of the INTACT Mass App (as of release version 1.6.0). Auto deconvolution uses optimized settings for peak detection and deconvolution processing and is appropriate for simplifying routine analysis of many samples of varying levels of complexity. However, this data had a large dynamic range of analytes, including several low intensity impurities near lower detection limits of the analysis that we wanted to robustly identify, and we found that the auto deconvolution algorithm settings were not appropriate in this case. Using custom deconvolution allows scientists to optimize the spectral deconvolution settings and better represent the full population of 5' cap structures present on the molecule.

For this data, we applied the BayesSpray charge deconvolution algorithm and tailored the retention time range, the input *m/z* range, the output expected mass range, and charge range to our data. The INTACT Mass data for the 5' cap analysis is shown in Figure 9B. We were able to identify all of the impurities (ppG, pppG, GpppG, and m7GpppG (Cap-0)) in the 1:10 dilution sample and three of the aforementioned impurities (excluding pppG) in the 1:100 dilution sample from Ngyuen *et al.*¹⁴

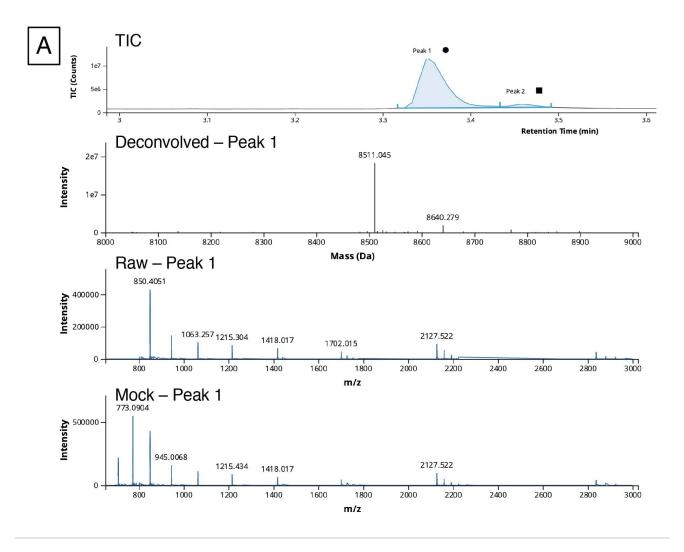


Figure 9. (A) TIC, Deconvoluted, Raw, and Mock spectra for the peak at 3.37 minutes (circle).

Resul	Its TIC TUV 260	MS											₹
Identi	ty: Pass Purity: Pass											$\sum_{i=1}^{\infty}$	t≣ IIV
	€ Component ▼	е. Т	Expected mass (Da)	* *	Mass error (ppm)		*. *	Identity result	*. T	MS response	*. T	Molecule %	
1							Pass						
2	uncapped+ All Forms						Pass			19,074	4,941		100
3	uncapped+ 5' ppG		8,138.02	9		-8.6	Pass			318	3,414		1.7
4	uncapped+ 5' GpppG		8,483.07	6		-8.3	Pass			183	8,164		1
5	uncapped+ 5' m7GpppG		8,497.09	2		-7.6	Pass			314	4,099		1.6
6	uncapped+ 5' m7GpppGm Ca	p-1	8,511.10	7		-7.4	Pass			18,255	9,263		95.7

Figure 9. (B) INTACT Mass App results table for the Cap-1 fragment and its product related impurity fragments.

Effective translation is critically dependent upon the 5' cap of mRNA and therefore the percentage of correctly capped mRNA product is a closely monitored CQA in the manufacturing of mRNA therapeutic products. The assay developed by Nguyen *et al* and the data analysis workflow using custom deconvolution settings is sensitive and robust enough to routinely identify unwanted impurities arising from an imperfect capping process.

Achieving reliable 3' poly(A) tail length and distribution measurements

Another CQA specific to mRNA is to determine the 3' poly(A) tail length. The 3' poly(A) tail length is directly correlated with mRNA stability.¹⁷ Therefore, determining the length, or rather the average length and size distribution, can impart valuable insight into the polyadenylation process. A relatively fast (15 min runtime) LC-MS assay was developed by Doneanu *et al* for the analysis of 3' poly(A) Tail oligonucleotides produced from Firefly Luciferase (FLuc) mRNA, following RNase T1 digestion.¹³ As shown in Figure 10A, the 3' poly(A) Tail elutes as a well resolved, abundant, late-eluting chromatographic peak in the TIC. The corresponding combined ESI-MS spectrum, containing a population of coeluting 3' poly(A) species, was submitted to the workflow and deconvolved using the MaxEnt1 charge deconvolution algorithm available in the INTACT Mass App. Custom charge deconvolution parameters, such as the input *m/z* range, the output mass range, and retention time range, were adjusted to produce the deconvolved spectrum shown in the inset of Figure 10A. This spectrum indicates a broad 3' poly(A) heterogeneity, with oligonucleotides ranging in size from 122 to 132 mers differing by the characteristic 329.2059 Da adenosine monophosphate residue mass. The INTACT Mass results are shown in Figure 10B and the ESI-MS signals and their corresponding intensities of all detected oligo components are

summarized in a plot displayed in Figure 10C. This graph was used to calculate the weight-averaged mass of the 3' poly(A) Tail of the FLuc mRNA. The measured average mass corresponds to the length of a 126.5 mer, and this measurement was confirmed from orthogonal SEC measurements.¹⁵ Following INTACT Mass processing, the eleven 3' poly(A) variants were measured with typical mass accuracies better than 45 ppm (Figure 10B).

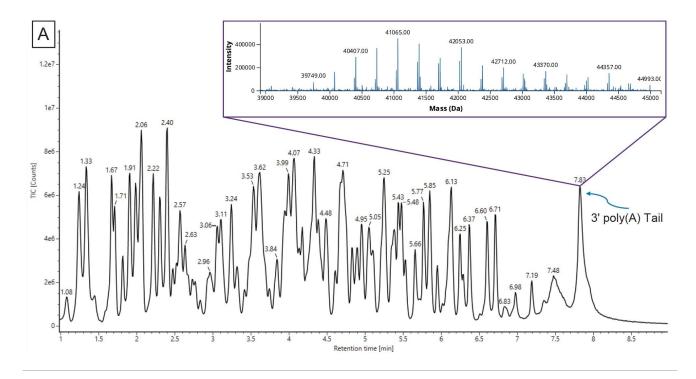
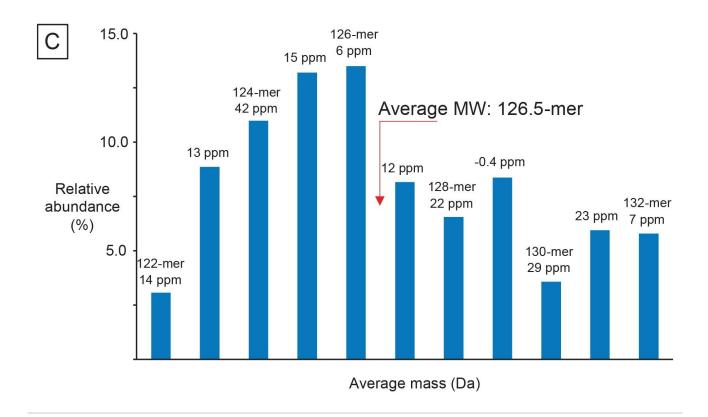
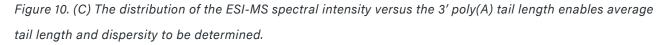


Figure 10. (A) TIC chromatogram of the 3' poly (A) Tail oligonucleotide mixture derived from RNase T1 digested FLuc mRNA. The inset illustrates the MaxEnt1 charge deconvolution from the INTACT Mass Application.

					×.				
	ф. Т	Component	*: *	Observed mass (Da)	* Exp	pected mass (Da)	*	Mass error (ppm)	
1	122mer	122mer		40,077.6	4	40,077.	13		12
2	122mer n+rA	123mer		40,406.8	5	40,406.1	34		1
3	122mer n+rA(2)	124mer		40,737.2	4	40,735.	54		4
4	122mer n+rA(3)	125mer		41,065.1	7	41,064.	75		1
5	122mer n+rA(4)	126mer		41,394.5	6	41,393.	95		1
6	122mer n+rA(5)	127mer		41,723.6	6	41,723.	16		1
7	122mer n+rA(6)	128mer		42,053.1	3	42,052.1	37		1
8	122mer n+rA(7)	129mer		42,381.7	3	42,381.5	57		
9	122mer n+rA(8)	130mer		42,712.1	в	42,710.	78		3
10	122mer n+rA(9)	131mer		43,040.4	8	43,039.5	98		1
11	122mer n+rA(10)	132mer		43,369.5	5	43,369.	19		

Figure 10. (B) The INTACT Mass Application results table illustrating the wide dispersity of the 3' poly(A) tail feature, with up to eleven adenosine masses added to the first detected tail variant, a 122 -mer. Eleven oligonucleotide species, ranging from 122- to 132-mers were identified with mass accuracies better than 45 ppm using INTACT Mass processing with custom deconvolution parameters.





Both the 3' poly(A) dispersity as well as the average mass measurement are important CQAs for therapeutic mRNA molecules, measuring process robustness, and potential clinical effect. Thus, the LC-MS assay developed by Doneanu *et al* and our data analysis workflow using the INTACT Mass App proves effective for analysis of 3' poly(A) Tails enzymatically cleaved from prophylactic/therapeutic mRNAs.

Conclusion

 CQAs for longer oligos and modern RNA therapeutics can be analyzed by approaches including sequence map confirmation (oligos, sgRNA, and mRNA), 5' capping effectiveness (mRNA), and 3' poly(A) tail analysis (mRNA) via workflows built on the BioAccord LC-MS System and the waters_connect informatics platform.

- Informatics enhancements to the INTACT Mass application were produced to support these workflows. The new custom deconvolution capability enables routine automated data processing, even with samples of low signal intensity and high dynamic range for the analytes of interest.
- These enhancements to the INTACT Mass application are complemented by additional microApps developed for RNA data analysis:
 - mRNA Cleaver that generates an *in silico* digestion list with user editable residues, enzymes, and mRNA modifiers
 - · Coverage Viewer that allows for easy visualization and calculation of digestion fragment map sequence coverage

References

- Zhu, Y., Zhu, L., Wang, X. & Jin, H. RNA-based Therapeutics: An overview and prospectus. *Cell Death & Disease* 13, (2022).
- Crooke ST, Witztum JL, Bennett CF, Baker BF. RNA-Targeted Therapeutics. *Cell Metab.* 2018 Apr 3;27(4):714-739. doi: 10.1016/j.cmet.2018.03.004. Erratum in: *Cell Metab.* 2019 Feb 5;29(2):501. PMID: 2961764
- 3. Damase, T. R. *et al.* The Limitless Future of RNA therapeutics. *Frontiers in Bioengineering and Biotechnology* 9, (2021).
- 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
- Verbeke, R., Lentacker, I., De Smedt, S. C. &; Dewitte, H. The dawn of mRNA vaccines: The COVID-19 case. Journal of Controlled Release 333, 511–520 (2021).
- Jackson, N. A., Kester, K. E., Casimiro, D., Gurunathan, S. &; DeRosa, F. The Promise of mRNA vaccines: A biotech and industrial perspective. *npj Vaccines* 5, (2020).
- Goyon, A. *et al.* Full sequencing of CRISPR/Cas9 single guide RNA (sgRNA) via parallel ribonuclease digestions and hydrophilic interaction liquid chromatography-high-resolution mass spectrometry analysis. *Analytical Chemistry* 93, 14792–14801 (2021).
- 8. Ivleva, V. B., Yu, Y.-Q. &; Gilar, M. Ultra-performance liquid chromatography/tandem mass spectrometry (UPLC/MS/MS) and UPLC/MS^e analysis of RNA oligonucleotides. *Rapid Communications in Mass*

Spectrometry 24, 2631–2640 (2010).

- Jiang, T. *et al.* Oligonucleotide sequence mapping of large therapeutic mRNAs via parallel ribonuclease digestions and LC-MS/MS. *Analytical Chemistry* 91, 8500–8506 (2019).
- Ross, R., Cao, X. & Limbach, P. Mapping post-transcriptional modifications onto transfer ribonucleic acid sequences by liquid chromatography tandem mass spectrometry. *Biomolecules* 7, 21 (2017).
- Shion, H., Boyce, P., Berger, S. & Yu, Y. Q. INTACT Mass[™] A Versatile waters_connect[™] application for Rapid Mass Confirmation and Purity Assessment of Biotherapeutics. Waters Application Note, 720007547, February 2023.
- Jackson, N. A., Kester, K. E., Casimiro, D., Gurunathan, S. & DeRosa, F. The promise of mRNA vaccines: A biotech and industrial perspective. *npj Vaccines* 5, (2020).
- Doneanu, C., Fredette, J. & Yu, Y. Q. Ion-Pairing Reversed Phase LC-MS Analysis of Poly(A) Heterogeneity Using the BioAccord LC-MS System, Waters Application Note, 720007925, 2023.
- Nguyen, J. M. *et al.* Rapid analysis of synthetic mRNA cap structure using ion-pairing RPLC with the BioAccord LC-MS System. Waters Application Note, 720007329, October 2023.
- Gilar, M. Size-exclusion chromatography method for poly(A) tail analysis of mRNA. Waters Application Note, 720007853, October 2023.
- Hu, L., Li, Y., Wang, J., Zhao, Y. & Wang, Y. Controlling CRISPR-Cas9 by guide RNA engineering. WIREs RNA 14, (2022).
- Pelletier, J.; Sonenberg, N. The Organizing Principles of Eukaryotic Ribosome Recruitment. Annu. Rev. Biochem. 2019, 88, 307–335
- Gaye, M. M. *et al.* Synthetic mRNA oligo-mapping using ion-pairing liquid chromatography and mass spectrometry. Waters Application Note, 720007669, October 2023.
- 19. Fekete, S *et al*. Challenges and emerging trends in liquid chromatography-based analyses of mRNA pharmaceuticals. *Journal of Pharmaceutical and Biomedical Analysis*. 224 (2022).
- 20. Fountain, K. J., Gilar, M. & Gebler, J. C. Analysis of native and chemically modifed oligonucleotides by tandem ion-pair reversed phase high-performance liquid chromatography/electrospray ionization mass spectrometry.

Rapid Commun. Mass Spectrom. 17, 646-653. https://doi.org/10.1002/rcm.959 (2003).

- Zhang, G., Lin, J., Srinivasan, K., Kavetskaia, O. & Duncan, J. N. Strategies for bioanalysis of an oligonucleotide class macromolecule from rat plasma using liquid chromatography-tandem mass spectrometry. *Anal. Chem*. 79, 3416–3424. https://doi.org/10.1021/ac0618674 (2007).
- Deng, P., Chen, X., Zhang, G. & Zhong, D. Bioanalysis of an oligonucleotide and its metabolites by liquid chromatography-tandem mass spectrometry. *J. Pharm. Biomed. Anal.* 52, 571–579. https://doi.org/10.1016/j.jpba.2010.01.040 (2010).
- 23. Gau *et al*. Oligonucleotide mapping via mass spectrometry to enable comprehensive primary structure characterization of an mRNA vaccine against SARS-CoV-2. *Nature Scientific Reports*. 13:9038 2023.
- Loverix S, Winqvist A, Stromber R, Steyaert J. Mechanism of RNase T1: concerted triester-like phosphoryl transfer via a catalytic three-centered hydrogen bond *Chem Biol*. 2000 Aug, 7(8):651-8 doi: 10.1016/s1074-5521(00)00005-3.
- Loverix S, Steyaert J. Deciphering the Mechanism of RNase T1. *Methods in Enzymology*, Academic Press, Vol. 341, 2001, 305-323, doi: 10.1016/S0076-6879(01)41160-8
- Masachika Irie. 3- RNase T1/RNase T2 Family RNases. *Ribonucleases, Academic Press*, 1997. 101-130, doi: 10.1016/B978-012588945-2/50004-2.
- Sorrentino, Salvatore. The Eight Human "Canonical" Ribonucleases: Molecular Diversity, catalytic properties, and special biological actions of the enzyme proteins. *FEBS Letters*. Vol 584, Issue 11, 2010, 2194-2200, doi: 10.1016/j.febslet.2010.04.018
- Gaye M.M. *et al.* CRISPR Single Guide RNA Characterization by IP-RP-LC-MS with a Premier Oligonucleotide BEH 300Å C18 Column. Waters Application Note, 720007897, October 2023.

Featured Products

BioAccord LC-MS System with ACQUITY Premier for Routine Biotherapeutic Analysis < https://www.waters.com/waters/nav.htm?cid=135087537>

waters_connect <https://www.waters.com/waters/nav.htm?cid=135040165>

27

UNIFI Scientific Information System <https://www.waters.com/waters/nav.htm?cid=134801648>

720008130, November 2023

 \wedge

© 2023 Waters Corporation. All Rights Reserved. 이용 약관 개인정보 처리방침 상표 채용정보 쿠키 쿠키 기본 설정