A WORKFLOW FOR PURITY DETERMINATION, INTACT MASS MEASUREMENT AND MS/MS SEQUENCING OF OLIGONUCLEOTIDE IMPURITIES DETECTED IN SYNTHETIC OLIGOS

Catalin Doneanu[1], Jonathan Fox[2], Brad J. Williams[1], Chris Knowles[3], Scott Berger[1] and Ying Qing Yu[1]

[1]Waters Corporation, Milford, MA, USA; [2]Waters Corporation, Wilmslow, United Kingdom; [3]Waters Corporation, Newcastle upon Tyne, United Kingdom

CONNECTING SCIENCE To Solve Problems That Matter

Waters™

OVERVIEW

An automated workflow, suitable for both regulated and non-regulated laboratories, was implemented for impurity analysis of oligonucleotides, involving rapid sequencing and localization of impurity modifications.

INTRODUCTION

- Several LC-MS workflows for oligonucleotide analysis have been recently introduced [1-4].
- INTACT Mass App (software) performs automated, fast deconvolution of oligonucleotide spectra across the entire chromatographic space, providing fast impurity assignments as well as the required metrics (mass accuracy and abundance) to support impurity analysis.
- CONFIRM Sequence App (software) is used for fast processing of both MS/MS and MS^E (no specific precursor selection) fragmentation spectra for verification of sequence coverage. The software displays the relevant matching information (graphically and in table format) and provides statistical analysis on each matched fragment ion. The sequence coverage can be viewed in a "dot-map" form to easily assess the coverage of a predicted sequence, or to locate an impurity modification, and is capable of high-throughput data processing.

INTACT Mass

CONFIRM Sequence

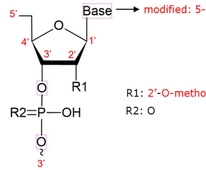
Xevo™ G3 QTof

Figure 1. Xevo™ G3 QTof Mass Spectrometer with the ACQUITY™ Premier UPLC™ System and waters_connect™ Software.

METHODS

Materials

A 21-mer synthetic-oligonucleotide, bearing 13 modified nucleotides, having the sequence: GCC UCA GTC TGC TTC GCA CCT was obtained from ATDBio (Southampton, UK). A 2'-O-(2-methoxyethyl) - (2'-MOE) modification was attached to two guanosines (G) and one adenosine (A). Also, in addition to the attachment of the 2'-MOE functional group to one uridine (U) and six cytidines (C), the nucleobases of these two nucleotides were further modified by the attachment of a 5-Methyl (5-Me) group to produce one 2'-MOE 5-Me uridine (U) and six 2'-MOE 5-Me cytidines (C). Out of the nine cytidine residues present in the 21-mer sequence, three cytidines were modified only at the nucleobase by the addition of a 5-Me functional group (C). Finally, there were eight nucleotides left unmodified: one adenosine (A), three guanosines (G) and five deoxythymidines (T). Stock solutions were prepared in DI water at a concentration of 1 µM (or 2.4 µg/mL), from which a 10 µL volume was injected, which corresponds to loading 10 picomoles of the 21-mer oligonucleotide on-column.



LC Conditions

UPLC System: ACQUITY™ Premier UPLC™ BSM

Eluent A: 7 mM triethylamine (TEA) and 40 mM hexafluoro-2-propanol (HFIP) in Milli-Q water (pH 8.6)

Eluent B: 3.5 mM TEA, 20 mM HFIP in 50% methanol

Column: 2.1 x 150 mm ACQUITY™ Premier OST Column

Flow rate: 300 µL/min

Gradient: 25 min gradient (13 to 23% B), 40-min runtime

UV Detector: TUV, 260 nm

MS conditions

Mass Spectrometer: Xevo™ G3 QTof Mass Spectrometer

Ionization mode: negative ESI

Mass range: 500—5,000 amu

Acquisition rate: 1Hz

Doubly, triply and quadruply charged precursors of the 21-mer and its impurities were fragmented by CID in the collision cell using voltages in the range of 10-70 V. In addition, MS^E (DIA—data independent) datasets were acquired for the same oligonucleotides on the same QTof instrument.

Informatics

Data acquisition and processing was performed using *waters_connect*™ Software. ESI-MS spectra of oligonucleotides were processed automatically using the INTACT Mass App. The oligonucleotide purity assessments were calculated based on the UV response from the UV chromatogram using the same software. Individual MS/MS spectra and MS^E datasets were processed by the CONFIRM Sequence App to establish the optimum collision energy for fragmentation of each oligonucleotide precursor.

RESULTS

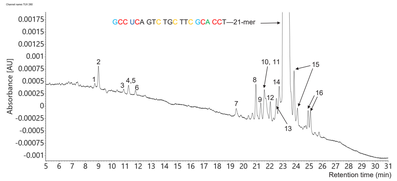


Figure 1. Zoom-in view of LC-UV chromatogram recorded on an ACQUITY Premier OST column.

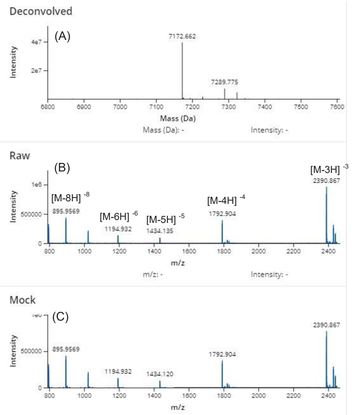


Figure 2. INTACT Mass ESI-MS spectra of the 21-mer heavily modified siRNA oligonucleotide: (A): deconvolved spectrum of the neutral species obtained using the BayesSpray deconvolution algorithm; (B) raw MS spectrum; (C) mock MS spectrum used for assessing the fidelity of the deconvolution process.

Peak ID	Type	Molecule ID	Component	Observed mass (Da)	Expected mass (Da)	Mass error (ppm)	Identity result
7	Impurity	7,172.57	7172.57 2'-MOE-5Me Cytidine,n-dT	6,491.427	6,491.425	0.2	Pass
8	Impurity	7,172.57	7172.57 2'-MOE-5Me Cytidine,5Me-Cytidine	6,492.316	6,492.409	-14.4	Warning
11	Impurity	7,172.57	7172.57 2'-MOE Guanosine	6,769.459	6,769.481	-3.2	Pass
9	Impurity	7,172.57	7172.57 2'-MOE Adenosine	6,785.440	6,785.476	-5.3	Pass
13	Impurity	7,172.57	7172.57 2'-MOE-5Me Uridine	6,794.473	6,794.487	-2.1	Pass
10	Impurity	7,172.57	7172.57 2'-MOE-5Me Cytidine	6,795.455	6,795.471	-2.4	Pass
15	Impurity	7,172.57	7172.57 n-dG	6,843.526	6,843.517	1.3	Pass
12	Impurity	7,172.57	7172.57 n-dT	6,868.558	6,868.524	4.9	Pass
14	Impurity	7,172.57	7172.57 5Me-Cytidine	6,869.450	6,869.508	-8.5	Pass
FLP	Product	7,172.57	7,172.57	7,172.549	7,172.570	-2.9	Pass

Figure 3. Screenshot showing a portion of the processing results table generated by the INTACT Mass App for the 21-mer oligonucleotide and its impurities. The dataset was deconvolved using the BayesSpray charge deconvolution algorithm and 15 oligonucleotide impurities were identified with mass accuracies of under 10 ppm.

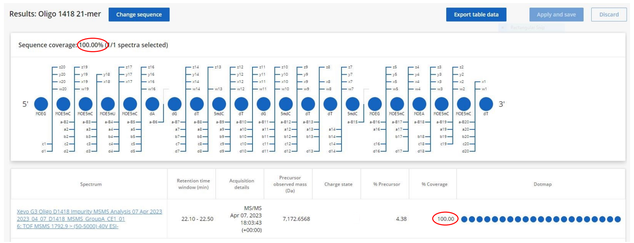


Figure 4. Complete sequence coverage (100%) resulted from the MS/MS fragmentation of the $[M-4H]^{-4}$ precursor (m/z=1792.14) of the 21-mer siRNA oligonucleotide. The precursor was fragmented with an *optimized fixed* collision energy (40 V) in the collision cell of a Xevo G3 QTof mass spectrometer.

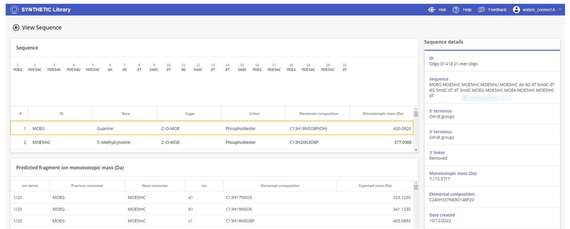


Figure 5. CONFIRM Sequence App screenshot showing the sequence of the 21-mer extensively modified siRNA oligonucleotide as entered in the Synthetic Library.

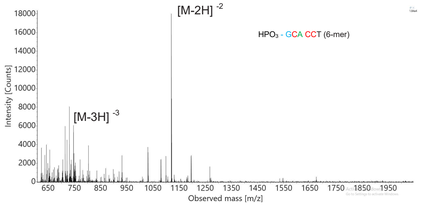


Figure 6. Ion pairing ESI-MS spectrum of a 6-mer 5'-phosphorylated oligonucleotide impurity, which is the least abundant oligonucleotide impurity present in the 21-mer FLP (0.05% relative abundance according to UV peak areas), labeled as peak 1 in the chromatogram shown in Figure 1.

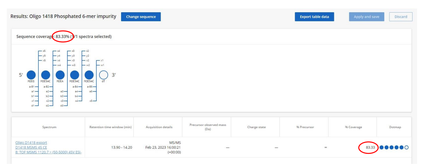


Figure 7. Dot-map sequence coverage (83%) resulted from the MS/MS fragmentation of the $[M-2H]^{-2}$ precursor (m/z=1120.7) of the 5'-phosphorylated 6-mer least abundant oligonucleotide impurity, labeled as peak no 1 in Figure 1. The precursor was fragmented with an *optimized fixed* collision energy (45 V) in the collision cell of a Xevo G3 instrument.

Oligo length	Peak label	Retention time (min)	Oligonucleotide sequence	Elemental composition	Delta mass (mass difference from FLP)	Molecular weight (neutral monoisotopic mass)	Most abundant precursor monoisotopic mass / charge state	Second precursor/ charge state	Optimum Collision Energies (V)	MSMS Sequence Coverage (%)	COMBINED Seq Coverage (%)	TUV Area Percentage (%)
6-mer	1	8.7	5'-phosphorylated GCA CCT	C75 H111 N21 O47 P6	- 4929 Da (loss of GCCUCAGTCTGCTTC)	2243.5367	1120.7611 (-2)	746.8383 (-3)	45 / 26	83.3	89.5	0.05
5-mer	2	9.0	CA CCT	C62 H92 N16 O36 P4	- 5412 Da (loss of GCCUCAGTCTGCTTCG)	1760.4811	879.2332 (-2)	1759.4738 (-1)	36 / 70	100.0	100.0	0.27
6-mer	3	10.9	tGCA CCT	C75 H110 N21 O44 P5	- 5009 Da (loss of GCCUCAGTCTGCTTC)	2163.5704	1080.7779 (-2)	720.1828 (-3)	40 / 28	74.2	82.4	0.06
7-mer	4	11.2	c GCA CCT	C85 H124 N24 O50 P6	- 4706 Da (loss of GCCUCAGTCTGCTT)	2466.6324	1232.3089 (-2)	821.2053 (-3)	38 / 28	65.1	76.8	0.09
8-mer	5	11.3	T GCA CCT	C95 H137 N26 O57 P7	- 4402 Da (loss of GCCUCAGTCTGCT)	2770.6784	1384.3319 (-2)	922.5522 (-3)	38 / 30	65.7	75.2	0.07
9-mer	6	11.8	TT GCA CCT	C105 H150 N28 O64 P8	- 4098 Da (loss of GCCUCAGTCTGC)	3074.7245	1536.3496 (-2)	1023.8825 (-3)	38 / 32	68.6	78.3	0.08
19-mer	7	19.5	GC* UCA G*- TG- TT- GCA CCT	C217 H304 N64 O133 P18	- 681 Da (2'-MOE-5Me C & dT)	6491.4269	2162.8017 (-3)	1621.8495 (-4)	-	-	-	0.14
19-mer	8	20.9	tC* UCA GT- TG- TT* GCA CCT	C217 H303 N63 O134 P18	- 680 Da (2'-MOE-5Me C & 5MeC)	6492.4109	2163.1297 (-3)	1622.0955 (-4)	-	-	-	0.45
20-mer	9	21.3	CC UCA GT- TG- TT- GC* CCT	C227 H319 N64 O141 P19	- 387 Da (2'-MOE A)	6785.4774	2260.8119 (-3)	1695.3621 (-4)	-	-	-	0.29
20-mer	10	21.6	tC* UCA GT- TG- TT- GCA CCT	C229 H321 N64 O140 P19	- 377 Da (2'-MOE-5Me C)	6795.4981	2264.1588 (-3)	1697.8673 (-4)	-	-	-	0.58
20-mer	11	21.7	*CC UCA GT- TG- TT- GCA CCT	C227 H319 N64 O140 P19	- 403 Da (2'-MOE G)	6769.4824	2255.4869 (-3)	1691.3633 (-4)	-	-	-	0.25
20-mer	12	22.0	GCC UCA G*- TG- TT- GCA CCT	C230 H324 N67 O141 P19	- 304 Da (dT)	6868.5257	2288.5013 (-3)	1716.1242 (-4)	-	-	-	0.32
20-mer	13	22.5	GCC *CA GT- TG- TT- GCA CCT	C227 H318 N67 O139 P19	- 378 Da (2'-MOE-5Me U)	6794.4889	2263.8227 (-3)	1697.6149 (-4)	-	-	-	0.45
20-mer	14	22.7	GCC UCA GT- TG- TT* GCA CCT	C230 H326 N66 O142 P19	- 303 Da (5Me C)	6869.5097	2288.8293 (-3)	1716.3702 (-4)	-	-	-	0.65
21-mer	FLP	23.1	GCC UCA GT- TG- TT- GCA CCT	C240 H337 N69 O148 P20		7172.5717	1792.1357 (-4)	2389.8500 (-3)	40 / 55	100.0	100.0	94.51
20-mer	15	23.8 / 24.1	GCC UCA *T- TG- TT- GCA CCT	C230 H325 N64 O142 P19	-329 Da (dG)	6843.5192	2280.1658 (-3)	1709.8725 (-4)	-	-	-	1.17
21-mer	16	24.9 / 25.1	GCC UCA GT- TG- TT- GCA CCT	C240 H342 N69 O151 P20	+53 Da (HSO3)	7225.5956	2407.5246 (-3)	1805.3916 (-4)	-	-	-	0.57
											TOTAL	100.00

Table I. Sixteen oligonucleotide impurities were identified in a 21-mer siRNA modified oligonucleotide. Six impurities and the full length product (FLP) were sequenced using a Xevo G3 mass spectrometer and the individual MS/MS spectra fragmented with optimum collision energies were processed using the CONFIRM Sequence App. The MS/MS sequence coverage for the FLP and its impurities were above 75%. The lowest abundance impurity, measured at 0.05% relative abundance, is a 6-mer 5'-phosphorylated oligonucleotide which was sequenced with 83% sequence coverage. The sequence of the FLP (21-mer oligonucleotide, 94.51% purity) was confirmed with 100% coverage. The total sequence coverage corresponds to the combined sequence obtained from the MS/MS fragmentation of two precursors of each oligonucleotide.

CONCLUSIONS

- INTACT Mass App provides quick, automated assignment of putative impurities detected in synthesized oligonucleotides with both UV and MS detectors
- Mass accuracy obtained for intact mass measurements of oligos and their impurities on the Xevo G3 QTof mass spectrometer is better than 10 ppm
- The MS/MS spectra acquired for six oligonucleotide impurities along with the MS/MS spectra acquired for a 21-mer extensively modified FLP were processed using the CONFIRM Sequence App to confirm their expected sequences
- The CONFIRM Sequence App is capable of finding sequence omissions, insertions or sequence scrambling, as impurities of extensively modified synthetic oligonucleotides
- The workflow described here demonstrates the capability of the CONFIRM Sequence App to achieve very good sequence coverage (83.3%) even for low abundance oligonucleotide impurities, down to ~ 0.05% relative abundance levels.

References

1. An Automated Compliance-Ready LC-MS Workflow for Intact Mass Confirmation and Purity Analysis of Oligonucleotides, 2020, Waters application note, P/N 720006820EN.

2. Analysis of Oligonucleotide Impurities on the BioAccord System with ACQUITY Premier, 2021, Waters application note, P/N 720007301EN.

3. LC-MS Analysis of siRNA, Single Guide RNA and Impurities using the BioAccord System with ACQUITY Premier System and New Automated INTACT Mass Aplication, 2022, Waters application note, P/N 720007546EN.

4. CONFIRM Sequence: A waters_connect Application for Sequencing of Synthetic Oligonucleotides and Their Impurities, 2022, Waters application note, P/N 720007677EN.

TO DOWNLOAD A COPY OF THIS POSTER, VISIT WWW.WATERS.COM/POSTERS