CONFIRM Sequence - A NEW SOFTWARE TOOL FOR SEQUENCE CONFIRMATION AND IMPURITY ANALYSIS OF SYNTHETIC OLIGONUCLEOTIDES

Catalin Doneanu[1], Mathew Gorton[2], Chris Knowles[2], Ian Morns[3], Chris Preston[3], Ian Reah[3], Mike Broughton[3], Simon Jones[3], Jonathan Rewcastle[3], Rory Mullins[3], Edvinas Karmonas[3], Paul Henesey[3], David Townsley[3], Dave Jackson[3], Pete Reay[3], Gavin Hope[3], Emma Harry[2], Jonathan Fox[2], Jo-Ann Riley[2], Henry Shion[1], Joseph Fredette[1], Ying Qing Yu[1]

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

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.

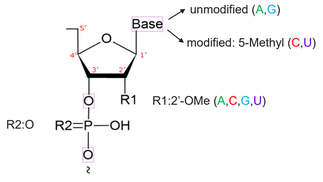
INTRODUCTION

- Synthetic oligonucleotides have emerged in recent years as a powerful alternative to small molecule and protein therapeutics [1].
- Manufacturing and quality control of oligonucleotide therapeutics requires highly selective and sensitive LC-MS methods for impurity detection and quantification.
- The most often used mass spectrometry-based method for oligonucleotide analysis has been reversed-phase chromatography employing a variety of ion-pairing reagents and modifiers in negative ESI-MS mode (IP-RP LC-MS).
- An automated workflow for intact level analysis of synthetic oligonucleotides employing the BioAccord™ LC-MS system was recently described [2-5].
- A critical step for identification of oligonucleotide impurities is mass spectrometry based sequencing [6] and data interpretation.
- Here we introduce a novel software suite: CONFIRM Sequence, developed for fast processing of both MS/MS and MS^E (no specific precursor selection) mass spectra.
- 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 analysis of pre-acquired data.
- The CONFIM Sequence app is capable of finding sequence mutations and sequence omissions.

METHODS

Materials

A 21-mer heavily modified oligonucleotide, containing a 2'-OMe modification on 19 of its nucleotides, having the sequence GUA ACC AAG AGU AUU CCA UTT and the elemental composition C229H306N76O143P20 was purchased from ATDBio (Southhampton, UK). Stock solutions were prepared in DI water at a concentration of 1 µM (or 2.34 µg/mL), from which a 10 µL volume was injected, which corresponds to loading 10 picomoles of the 21-mer oligonucleotide on-column.



Modifications of the 21-mer oligonucleotide

LC Conditions

Oligonucleotide separations were performed on an ACQUITY™ H-Class Bio UPLC system equipped with a 2.1 x 100 mm ACQUITY Premier OST column (P/N 186009485). The column flow rate was 300 µL/min and the column temperature was 60°C. The mobile phase composition was: Solvent A: 7 mM triethylamine (TEA) and 40 mM hexafluoro-2-propanol (HFIP) in Milli-Q water (pH 8.6) and Solvent B: 3.5 mM TEA, 20 mM HFIP in 50% methanol. Separations were performed using a 25-min gradient from 25-35% B and the total runtime was 40 min.

MS conditions

ESI-MS spectra of oligonucleotides were acquired on a Xevo™ G2-XS QTOF instrument in negative ion mode over the m/z range of 500-5,000 with a full scan rate of 1Hz. The optimized ESI source parameters include: capillary voltage 2.5 kV, cone voltage 45V, source temperature 120°C and desolvation temperature 450°C. 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.

Informatics

Data acquisition and processing was performed using waters_connect™ software. Individual MS/MS spectra were processed using the new CONFIRM Sequence app to establish the optimum collision energy for fragmentation of each oligonucleotide precursor. In addition, MS^E (DIA—data independent) datasets were acquired for the same oligonucleotides on the Xevo G2-XS instrument as well as on a BioAccord Tof System with the ACQUITY Premier UPLC system.

RESULTS

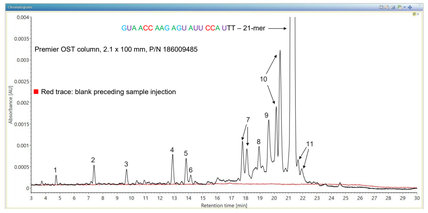


Figure 1. LC-UV chromatogram recorded on an ACQUITY Premier OST column.

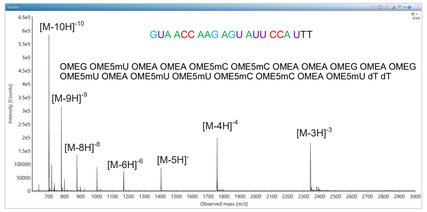


Figure 2. Ion pairing reversed phase ESI-MS spectrum of the 21-mer heavily modified oligonucleotide. The oligonucleotide sequence entered in the CONFIRM Sequence app is listed atop the spectrum along with the conventional sequence.

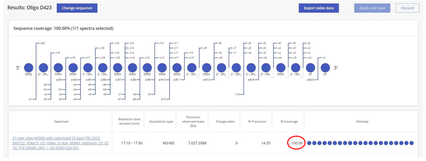


Figure 3. CONFIRM Sequence screenshot showing excellent MS/MS fragmentation coverage (100%) in a dot-map format. The [M-3H]$^{-3}$ precursor of the 21-mer heavily modified oligonucleotide (m/z = 2342.0) was fragmented using an *optimized fixed* collision energy (set at 63 V) in the collision cell of a Xevo G2-XS QTof instrument.

Figure 4. Dot-map sequence coverage (~ 70%) obtained from high energy MS^E untargeted fragmentation. All precursors of the 21-mer heavily modified oligonucleotide (see Figure 2) were fragmented using an *optimized cone voltage ramp* (from 60 to 80 V) applied to the Step Wave of a BioAccord TOF instrument.

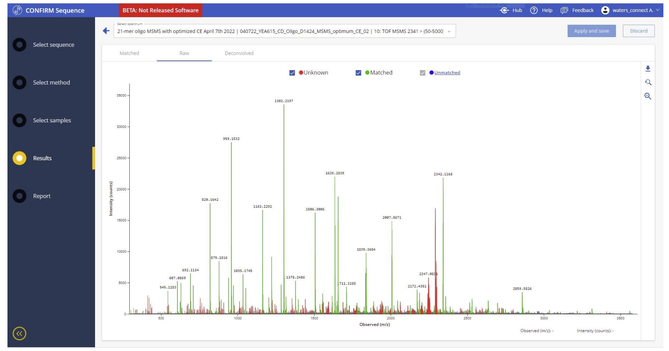


Figure 5. Confirm Sequence screenshot showing an MS/MS spectrum recorded for the 21-mer oligonucleotide. The [M-3H]$^{-3}$ precursor of this oligo (m/z = 2342.0) was fragmented using an *optimized fixed* collision energy (set at 63 V) in the collision cell of a Xevo G2-XS QTof instrument. The fragment ions labeled in green were matched to the oligonucleotide sequence according to the dop-map diagram shown in Figure 3.

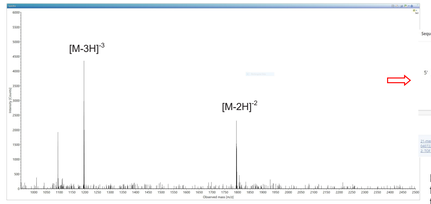


Figure 6. Ion pairing reversed phase ESI-MS spectrum of an 11-mer oligonucleotide impurity which is the least abundant impurity present in the 21-mer oligonucleotide sample (0.2% relative abundance, labeled as peak 1 in the chromatogram shown in Figure 1).

Figure 7. Dot-map sequence coverage (~ 80%) obtained from the MS/MS fragmentation of the [M-3H]$^{-3}$ precursor of the 11-mer oligonucleotide impurity. The precursor was fragmented with an *optimized fixed* collision energy (36 V) in the collision cell of a Xevo G2-XS instrument.

Oligo length	Peak label	Retention time (min)	Oligonucleotide sequence	Elemental composition	Most abundant precursor monoisotopic mass	Charge state	Optimum Collision Energy (V)	MSMS Sequence Coverage (%)	TOTAL Seq Coverage (%)	TUV Area Percentage (%)
11-mer	1	4.73	GU AUU CCA UTT	C119 H161 N33 O77 P10	1196.8952	3	36	81.8	95.1	0.18
12-mer	2	7.39	AGU AUU CCA UTT	C130 H175 N38 O83 P11	1967.3805	2	59	83.3	92.3	0.51
14-mer	3	9.67	AG AGU AUU CCA UTT	C152 H203 N48 O96 P13	1545.2950	3	46	85.7	85.7	0.45
15-mer	4	12.89	AAG AGU AUU CCA UTT	C163 H217 N53 O102 P14	1659.6510	3	49	86.7	86.7	0.89
16-mer	5	13.83	C AAG AGU AUU CCA UTT	C174 H233 N56 O109 P15	1770.6752	3	51	75.0	83.2	0.85
17-mer	6	14.15	CC AAG AGU AUU CCA UTT	C184 H261 N59 O116 P16	1881.7307	3	-	-	-	0.34
18-mer	7	17.8/18.1	CC AAG AGU AUU CCA UTT	C184 H261 N59 O116 P16	1996.0555	3	-	-	-	2.58
21-mer	8	18.94	AUA ACC AAG AGU AUU CCA UTT	C229 H306 N76 O142 P20	2336.1198	3	51	78.2	78.2	1.02
21-mer	9	19.62	GU* ACC AAG AGU AUU CCA UTT	C218 H292 N71 O137 P19	2227.0954	3	59	85.6	85.6	2.42
20-mer	10	20.1/20.4	AUA A*C AAG AGU AUU CCA UTT	C218 H290 N73 O136 P19	2230.4272	3	59	82.5	82.5	7.13
21-mer	MAIN Peak	21.28	GUA ACC AAG AGU AUU CCA UTT	C229 H306 N76 O143 P20	2341.4514	3	63	100.0	100	83.03
21-mer	11	21.6/21.9	GUA ACC AAG AGU AUU CCA UTT	C229 H307 N76 O143 P20	2341.7873	3	-	-	-	0.60

Table I. Eleven oligonucleotide impurities were identified in a 21-mer extensively modified oligonucleotide. Eight impurities and the full length product (FLP) were sequenced using a Xevo G2-XS instrument 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, an 11-mer oligonucleotide was sequenced with 95% sequence coverage, while the sequence of the FLP (21-mer oligonucleotide) 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.

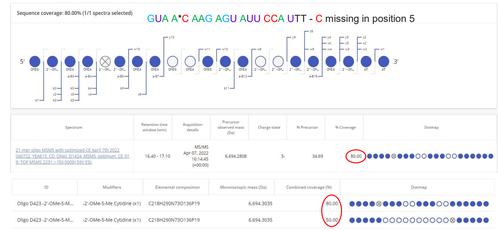


Figure 8. Confirm Sequence screenshot showing the sequence coverage of a 20-mer oligonucleotide impurity belonging to the peak dublet 10 shown in Figure 1. The most abundant isomer of this impurity, eluting just before the FLP, was identified as an oligonucleotide missing a 2'-OMe 5 Me cytidine residue in position 5. This oligonucleotide sequence displayed a significantly higher sequence coverage (80%) than an alternative sequence which is missing the same residue at position 16. There are 4 modified cytidines in the 21-mer sequence, but because they are located as pairs, there are only two possible 20 -mer impurities missing a single 2'-OMe 5 Me cytidine. The software looked for a missed 2'-OMe 5 Me cytidine from the sequence of the 21-mer and found strong evidence (80% sequence coverage) for assigning this impurity to the sequence: GUA A*C AAG AGU AUU CCA UTT.

CONCLUSIONS

- A new software application - CONFIRM Sequence, was developed for fast processing of both MS/MS and MS^E (no specific precursor selection) mass spectra
- The MS/MS spectra acquired for eight 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
- The workflow described here demonstrates the capability of the CONFIRM Sequence app to achieve very high sequence coverage (~ 95%) even for low abundance oligonucleotide impurities, down to ~ 0.2% abundance levels.

References

1. Sharma VK, Watts JK Oligonucleotide therapeutics: chemistry, delivery and clinical progress, *Future Med Chem*, 2015, 7(16), 2221-2242.
2. An Automated Compliance-Ready LC-MS Workflow for Intact Mass Confirmation and Purity Analysis of Oligonucleotides, 2020, Waters application note, P/N 720006820EN.
3. Intact Mass Confirmation Analysis on the BioAccord LC-MS System for a Variety of Extensively Modified Oligonucleotides, 2020, Waters application note, P/N 720007028EN.
4. Analysis of Oligonucleotide Impurities on the BioAccord System with ACQUITY Premier, 2021, Waters application note, P/N 720007301EN.
5. 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.
6. McLukey SA, Van Berkel GJ, Glish GL Tandem Mass Spectrometry of Small, Multiply Charged Oligonucleotides, *J Am Soc Mass Spectrom*, 1992, 3, 60-70.

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