OligoQuest: Oligonucleotide purity analysis with side-product identification and quantitation

Due to recent breakthroughs in delivery methods, there is growing interest in using synthetic oligonucleotides as active pharmaceutical ingredients (APIs) for the treatment of a wide variety of diseases.

Introduction

These drug candidates include multiple classes such as antisense oligonucleotides, small/short interfering RNA, micro-RNA, immunostimulatory oligonucleotides, aptamers, and splice-switching oligonucleotides (1). Accurate analytical characterization of oligonucleotides as APIs is necessary to confirm their identity, to determine their purity, and to identify and quantify synthesis side products.

Determining the molecular weight and confirming the nucleotide sequence of an oligonucleotide are fundamental criteria for establishing the molecule's identity. Oligonucleotide synthesis is a complex process that requires more than 100 sequential chemical reactions to make a single 25-base sequence, and the key to understanding and optimizing this chemistry is the identification of side products. Furthermore, quality of each synthesized oligonucleotide must be evaluated prior to use to ensure that the correct sequence was made, and that purity meets regulatory standards.

This technical note describes the use of the Bruker timsTOF Pro 2 mass spectrometer equipped with the VIP-HESI ion source for the in-depth analysis of oligonucleotides. UHPLC was used to efficiently separate impurities from the full-length product (FLP) and to provide unambiguous confirmation of FLP sequences and synthesis side-products. Their in-depth characterization was carried out by the highly automated OligoQuest™ workflow within Bruker´s BioPharma Compass® software.

Keywords: Oligonucleotides, Sequence confirmation, MS/MS, Side product quantitation, Side product identification, Quality control, High resolution accurate mass spectrometry

Detlev Suckau[1], Eckhard Belau[1], Timo Schierling[2], Stuart Pengelley[1], Julia Schneider[2]; [1]Bruker Daltonics GmbH & Co. KG, Bremen Germany, [2]Axolabs GmbH, Kulmbach, Germany.

Materials and Methods

The 24-mer RNA FLP (dubbed "mod3") with 2´ O-methylation of the ribose at each position was synthesized by Axolabs GmbH (product number X119083K2; sequence: 5OH c a c g c g u g c u u u u g c a c g g c g u g c 3OH). Seven isomers of mod3 were also analyzed (see Figure 1). The sample was diluted with Eluent A to 0.1 µg/µL prior to analysis and 0.4 µg was injected onto the UHPLC system for subsequent MS analysis. A Bruker Elute UHPLC - equipped with an Azura UVD 2.1S UV detector (KNAUER) recording the 260 nm UV chromatogram - was connected to the Bruker timsTOF Pro 2 via the VIP-HESI ion source to separate side-products, concentrate target compounds and remove salts.

For a pure sample such as the 24-mer it is beneficial to use a low-speed auto MS/MS cycle with only one precursor being selected. This ensures that the charge states with the highest intensity will be selected for MS/MS.

Chromatography	
UHPLC column	Waters XBridge Oligonucleotide BEH C_{18}, 130Å 2.5 µm, 2.1 x 50 mm, 70°C column oven temperature
Eluent A (aqueous phase) in deionized water	0.24% (v/v) Triethylamine (TEA), 1.00% (v/v) Hexafluoro-2-propanol (HFIP), 1.00% (v/v) Methanol (MeOH)
Eluent B (organic phase) in Acetonitrile	10% (v/v) Isopropanol (IPA)

Gradient

Time [min]	Flow [mL/min]	% A	% B
0.0	0.25	99	1
1.0	0.25	99	1
3.0	0.25	96	4
16.0	0.25	90	10
16.2	0.25	5	95
16.8	0.25	5	95
17.0	0.25	99	1
23.5	0.25	99	1

VIP-HESI Source Parameters

Nebulizer	Dry gas	Dry temp	Sheath gas temp	Sheath gas flow
4 bar	8 L/min	220°C	450°C	4 L/min

MS Parameters – autoMSMS Acquisition	
AutoMSMS spectra were acquired on a Bruker timsTOF Pro 2 in negative ion mode. MS acquisition parameter in the autoMSMS analysis (timsControl Version 4.1.12) are listed in tabular form below	
Deflection 1 delta	-70 V
Funnel 1 RF	350 Vpp
isCID	0
Funnel 2 RF	400 Vpp
Multipole RF	600 Vpp
Quadrupole ion energy	4 eV
Quadrupole low mass	500 *m/z*
Collision energy	10 eV
Pre-pulse storage	10 µs

MS/MS Parameters – targeted MS/MS Acquisition	
Targeted MS/MS spectra were acquired in case of the mod3-c1 side product to increase the sequence coverage for that oligo and to locate the loss of c. Here, the mod3-c1 oligo eluted at 12.17 min.	
MS Settings	Scan mode MRM
MS/MS Settings	*m/z* 1912, width (*m/z*) 5, isCID 0, CE (eV) 61.9, x Acq. 1.0, Rt range 11.9-12.3 min

MS/MS Parameters – auto MS/MS Acquisition	
AutoMSMS spectra were acquired on a Bruker timsTOF Pro 2 in negative ion mode. MS acquisition parameters in the autoMSMS analysis (timsCONTROL Version 4.1.12) are listed in tabular form below	
Total cycle time	1 s
MS spectra rate	2 Hz
MS/MS spectra rate (fixed)	2 Hz
No. precursors	1
Normalized threshold	31 counts/1000 scans
Average scans	5
Scan range, isol. width, collision energy	500 -3000 *m/z*, isol. width 3 *m/z*, 15.5-97.6 eV

Data processing

The LC-UV-MS(/MS) data were processed in BioPharma Compass® 2023b using the OligoQuest autoMSMS tutorial workflow method based on the user defined sequences of 24-mer RNA variants, which include residue-specific modifications. Here, 2´ O-methylated nucleotides were used, abbreviated by a, c, g and u in the sequences. OligoQuest enables the automated rapid verification of molecular mass, sequence and the assessment of purity by quantifying chromatographic peaks using the UV and MS signal intensities. In addition, it can identify sequence variants and synthetic impurities based on the input of the target sequence and further workflow parameters, which allows screening for failure sequences, addition of nucleotides or nucleotide exchange variants.

Impurities with incomplete MS/MS coverage were targeted in a second round of analysis with the OligoQuest targetedMSMS workflow method in which only selected *m/z* and Rt ranges were used for MS/MS spectra acquisition. MS precursor ion spectra were not measured in this case.

Results and Discussion

The Full-Length Product

The Result table (Figure 1) provides a quick confirmation that the base peak in the dataset matches the expected molecular weight of the oligo with good mass accuracy and the sample purity. The Expected table shows the analysis result for the currently selected mod3 sample at greater detail listing all identified molecular species, including sequence variants defined in the method´s matching parameters. Here we observe the confirmation of the mod3 FLP, 5 side product candidates: mod3>> [a,g]16, mod3-c1, mod3+g1 and mod3>>[u,c] and mod3>>[a,c]2. The variants with an added g residue (row 8) or the exchange of u-to-c (row 14) were solely identified based on molecular weight without significant support from fragment ion data (MSMS Score = 0). Therefore, the location of these variants in the sequence could not be determined.

In the following study the focus is on the mod3 sample and sequence. The UV-chromatogram (Figure 2) shows mod3 as base peak (93.4% area) at 9.28 min and side products in earlier chromatographic peaks with relative peak areas from 0.7-2.5%. These values reflect the peak areas within the UV chromatogram. The overall purity provided in the Result table (Figure 1) was calculated at 93.4% as well. Purity also includes unrelated MS peaks within the target molecule's chromatographic peak, such as coeluting side-products or unidentified by-products.

Row	Result	Position	Base Peak Mr [Da]	ΔBase Peak Mr [Da]	Mr Ref	Mr Sample	Δ Mr [ppm]	Rt [min]	Purity [%]	Int. [a.u.]	MSMS Score	Sample Name	Sequence
1		3	7969.3751	-0.0114	7969.3865	7969.3751	-1.43	9.44	87.9	5.761E+05	62.29	mod1	5OH g c a c g c g u g c u u u u c g a c g c g u g c 3OH
2		4	7969.3773	-0.0092	7969.3865	7969.3773	-1.15	8.85	94.0	1.075E+06	66.08	mod2	5OH g c a g c c g u g c u u u u g c a c g c g u g c 3OH
3		5	7969.3736	-0.0129	7969.3865	7969.3736	-1.62	9.28	93.4	9.991E+05	58.5	mod3	5OH c a c g c g u g c u u u u g c a c g g c g u g c 3OH
4		9	7985.3481	-0.0155	7985.3636	7985.3481	-1.94	12.99	79.7	6.414E+05	55.51	mod7	5OH g cs a c g c g u g c u u u u g c a c g c g u g c 3OH
5		2	7969.3783	-0.0082	7969.3865	7969.3783	-1.03	13.08	91.3	4.408E+05	61.47	wt	5OH g c a c g c g u g c u u u u g c a c g c g u g c 3OH
6		6	7969.3396	-0.0469	7969.3865	7969.3396	-5.88	8.72	84.4	1.337E+06	63.55	mod4	5OH g c a g c c g u g c u u u u c g a c g c g u g c 3OH
7		7	8001.3213	-0.0194	8001.3408	8001.3213	-2.43	13.07	54.0	2.711E+05	46.48	mod5	5OH g cs a c g c g u g c u u u u g c a c g c g u g cs 3OH
8		8	7985.3478	-0.0158	7985.3636	7985.3478	-1.98	13.00	74.0	4.214E+05	52.34	mod6	5OH g c a c g c g u g c u u u u g c a c g c g u g cs 3OH

Expected | Unidentified

Row	Confirmed	Sequence	Annotation	Mr Ref	Mr Sample	Δ Mr [ppm]	Δ Mr [Da]	Int. [a.u.]	Rel. Int. [%]	Rt [min]	Intensity Coverage [%]	Sequence Coverage [%]	MSMS Score
1	☑	5OH c a c g c g u g c u u u u g c a c g g c g u g c 3OH	mod3	7969.3865	7969.3736	-1.62	-0.0129	9.991E+05	93.4	9.28	61.0	95.8	58.5
2	☑	5OH c a c g c g u g c u u u u g c g c g g c g u g c 3OH	mod3>>[a,g]16	7985.3814	7985.3740	-0.92	-0.0073	5.135E+04	1.8	8.61	64.7	54.2	1.8
3	☐	5OH c g c g c g u g c u u u u g c a c g g c g u g c 3OH	mod3>>[a,g]2	7985.3814				5.135E+04	1.8	8.61	40.0	20.8	0.43
4	☑	5OH a c g c g u g c u u u u g c a c g g c g u g c 3OH	mod3-c1	7650.3295	7650.3262	-0.44	-0.0034	2.714E+04	1.0	8.93	27.8	13.0	0.1
5	☐	5OH a c g c g u g c u u u u g c a c g g c g u g c 3OH	mod3<-1	7650.3295				2.714E+04	1.0	8.93	27.8	13.0	0.1
6	☐	5OH c a c g g u g c u u u u g c a c g g c g u g c 3OH	mod3-c5	7650.3295				2.714E+04	1.0	8.93	25.4	13.0	0.09
7	☐	5OH c a c g c g u g u u u u g c a c g g c g u g c 3OH	mod3-c9	7650.3295				2.714E+04	1.0	8.93	25.4	13.0	0.09
8	☑	5OH g c a c g c g u g c u u u u g c a c g g c g u g c 3OH	mod3+g1	8328.4495	8328.4227	-3.22	-0.0269	2.299E+04	0.5	9.05	60.9	0.0	0.0
9	☐	5OH c a g c g c g u g c u u u u g c a c g g c g u g c 3OH	mod3+g3	8328.4495				2.299E+04	0.5	9.05	79.2	0.0	0.0
10	☐	5OH c a c g g c g u g c u u u u g c a c g g c g u g c 3OH	mod3+g4	8328.4495				2.299E+04	0.5	9.05	79.2	0.0	0.0
11	☐	5OH c g a c g c g u g c u u u u g c a c g g c g u g c 3OH	mod3+g2	8328.4495				2.299E+04	0.5	9.05	60.9	0.0	0.0
12	☑	5OH c c c g c g u g c u u u u g c a c g g c g u g c 3OH	mod3>>[a,c]2	7945.3752	7945.3663	-1.13	-0.0090	1.198E+04	0.7	8.71	63.0	37.5	0.28
13	☐	5OH c a c g c g u g c u u u u g c c c g g c g u g c 3OH	mod3>>[a,c]16	7945.3752				1.198E+04	0.7	8.71	43.2	8.3	0.04
14	☑	5OH c a c g c g c g c u u u u g c a c g g c g u g c 3OH	mod3>>[u,c]7	7968.4024	7968.3706	-4.00	-0.0318	5.173E+03	0.1	9.05	0.0	0.0	0.0
15	☐	5OH c a c g c g u g c c u u u g c a c g g c g u g c 3OH	mod3>>[u,c]10	7968.4024				5.173E+03	0.1	9.05	0.0	0.0	0.0
16	☐	5OH c a c g c g u g c u c u u g c a c g g c g u g c 3OH	mod3>>[u,c]11	7968.4024				5.173E+03	0.1	9.05	0.0	0.0	0.0
17	☐	5OH c a c g c g u g c u u c u g c a c g g c g u g c 3OH	mod3>>[u,c]12	7968.4024				5.173E+03	0.1	9.05	0.0	0.0	0.0
18	☐	5OH c a c g c g u g c u u u u g c a c g g c g u g c 3OH	mod3	7969.3865				3.412E+03	0.2	8.71	0.0	0.0	0.0

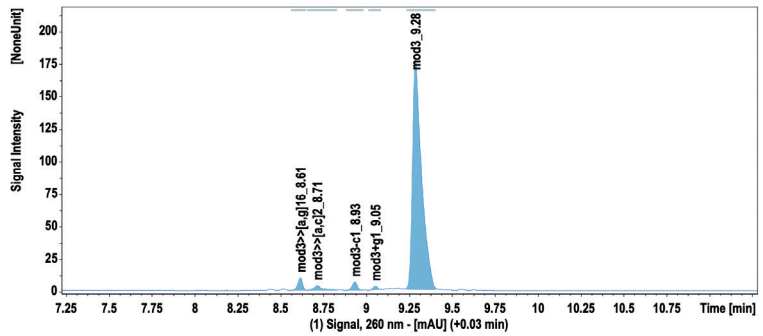
Figure 1

Result summaries of all 8 RNA 24mers (top) from one batch OligoQuest autoMSMS analysis; detailed results for the selected mod3 sample in the Expected table (bottom); deselected entries are excluded from result reports.

They cannot be detected by LC-UV analysis alone but in combination with MS. Adduct peaks from, e.g., sodium or triethylamine are not acknowledged as side products but contribute to the FLP abundance.

For each sample multiple quality attributes are determined, which have been previously defined in the workflow method. For this sample, all reporting attributes are shown in green, confirming that all narrow definitions of acceptance criteria were matched (Figure 3).

The deconvoluted mass spectrum obtained from the mod3 peak in the chromatogram can be overlayed with the theoretical isotope pattern which was calculated based on the elemental composition of the sequence to provide information about the quality of the match (Figure 4). Partial c/u conversions (shift of +1 Da) for example can be detected sensitively by this comparison (see Figure 8). Accurate mass and isotope pattern add credibility to proposed side products for which MS/MS data were not obtained or insufficient.



Chromatogram Peaks

Rt [min]	Int.	Area	Annotation	Rel. Area [%]
8.61	9.520E+00	1.751E+01	mod3>>[a,g]16_8.61	2.5
8.71	3.545E+00	1.129E+01	mod3>>[a,c]2_8.71	1.6
8.93	6.239E+00	1.256E+01	mod3-c1_8.93	1.8
9.05	2.642E+00	4.726E+00	mod3+g1_9.05	0.7
9.28	1.772E+02	6.469E+02	mod3_9.28	93.4

Figure 2
UV 260 nm Chromatogram and the Chromatogram Peaks table used for quantitative assessment of side products.

Multi Attributes (OligoQuest)

Sample Position: 5 | Sample ID: mod3

Method Attribute	Narrow	Wide	Sample Result	Unit
Base peak as expected	≥ 100.0	< 70.0	100.0	Rel. int. [%] vs. base peak
Mass Accuracy [ppm]	< 3.0	≥ 5.0	-1.6	ppm
Intensity Coverage	≥ 40.0	< 30.0	61.0	Rel. int. [%] of all target signals
Sequence Coverage	≥ 30.0	< 20.0	95.8	SC in [%]
Score	≥ 15.0	< 10.0	58.5	Product of IC x SC in [%]

Figure 3
Multi Attributes view, displaying the Narrow and Wide acceptance criteria for 5 quality attributes and the values determined as Sample Result.

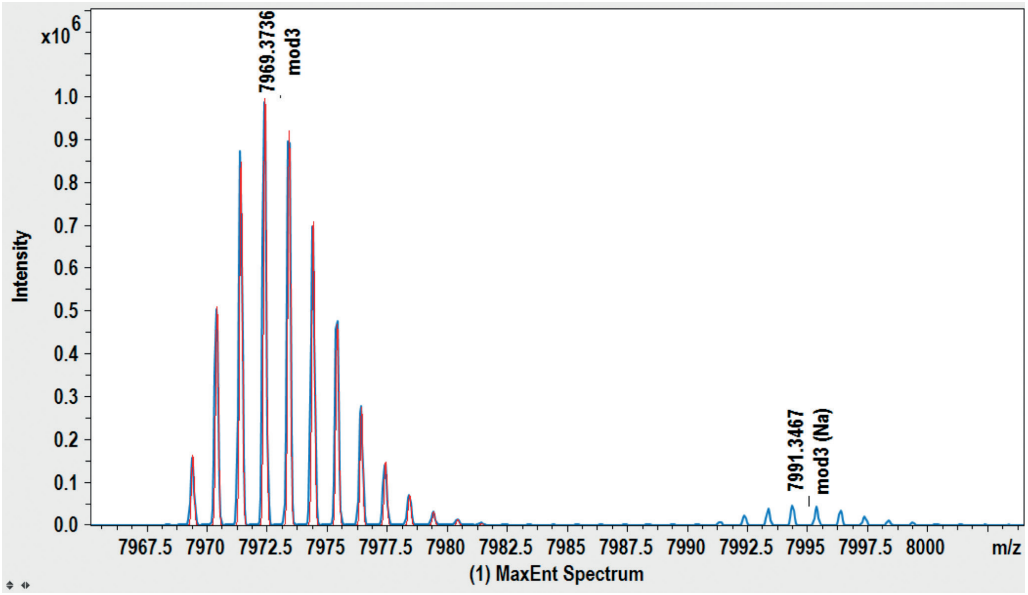


Figure 4

Overlay between experimental (blue) and calculated (red) isotope pattern of the mod3 FLP allows to assess, e.g., the absence of C/U conversion.

The Sequence Map (Figure 5) supplies a clear overview about the quality of the match between MS/MS data and the selected sequence in the Expected table. The 5´-fragment assignment matches are shown in red bricks, 3´-fragments are shown in blue bricks (ppm errors inside). The sequence with index numbers is counted from 5´-end above and from 3´-end below. The green bottom line uses more stringent sequence validation criteria, serving to confirm only residues that are bracketed by 3' and 5' fragment ions: the number inside reports the redundancy in the validation for each individual residue. The intensity of the fragment ions is color coded in 3 intensity ranges to guide the manual validation of weak fragment ions.

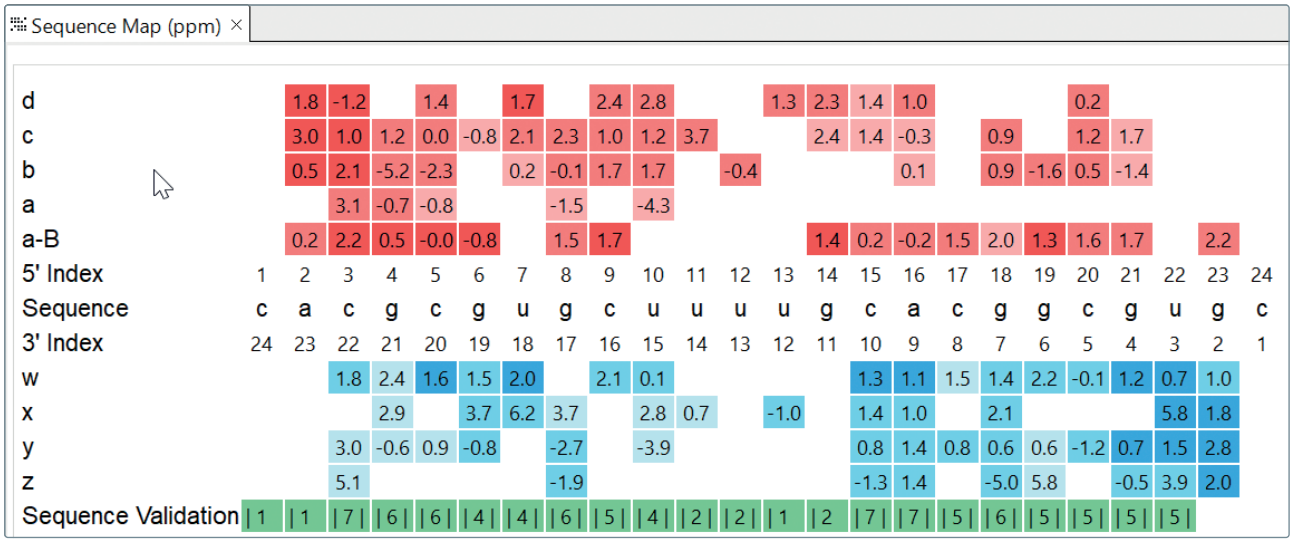


Figure 5

Sequence Map shows the sequence coverage (green bar) of the MS/MS fragment ions for the mod3 24mer.

To verify the individual matches in the Sequence Map, fragment ions and charge states can be selected in BioPharma Compass and the respective profile spectrum of the fragment is shown together with its theoretical isotope pattern (Figure 6).

Side Products Characterization

As shown in Figure 1 for mod3, 5 side product candidates were returned from the OligoQuest autoMSMS analysis: mod3-c1, mod3+g1, mod3>> [a,g]16, mod3>>[u,c] and mod3>>[a,c]2. They were obtained by screening the mod3 sequence 5OH c a c g c g u g c u u u u g c a c g g c g u g c 3OH for residue losses (-c1), additions (+g) or exchanges ([a,g]16, [a,c]2 and [u,c]). The residue-specific assignment was omitted here if the MS/MS score did not suggest a clear location of the sequence variation.

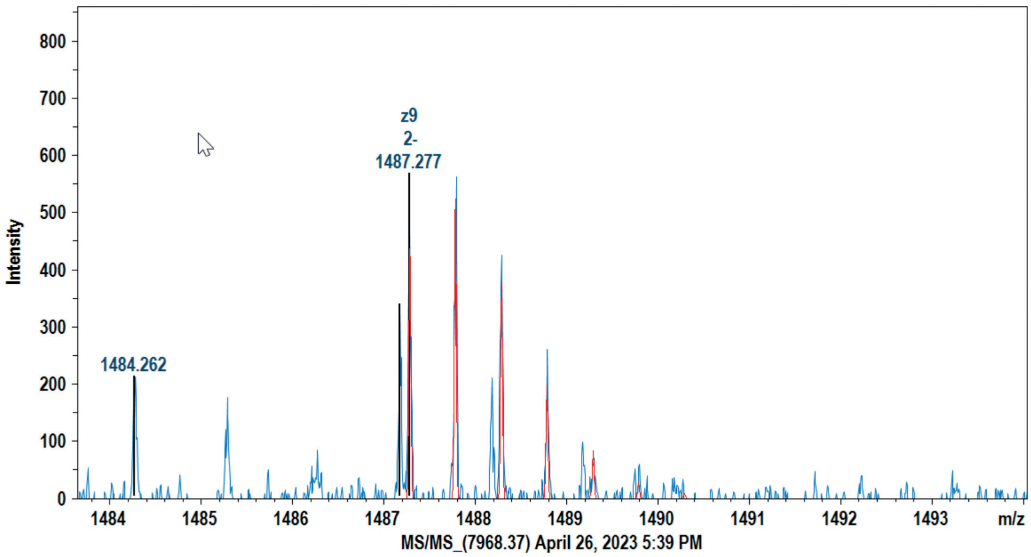


Figure 6
The z9 (-2) fragment in the MS/MS spectrum of mod3. The theoretical pattern is overlayed automatically in red; another -1 fragment at m/z 1487.17 is also present.

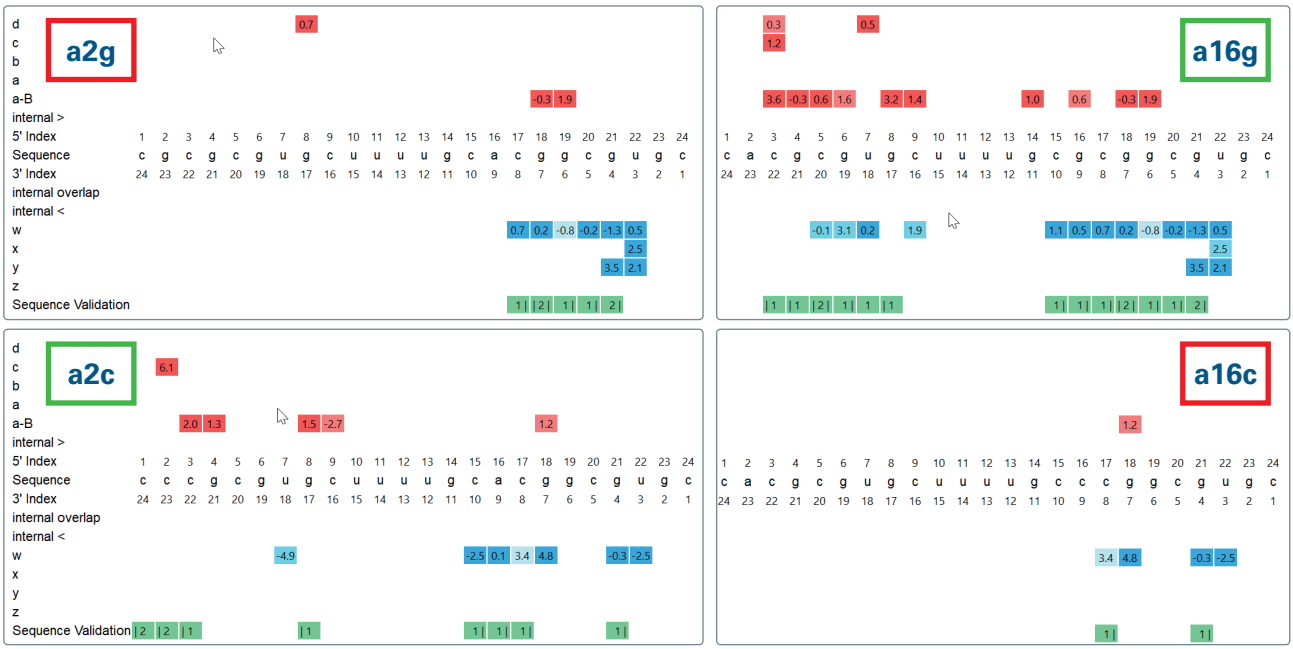
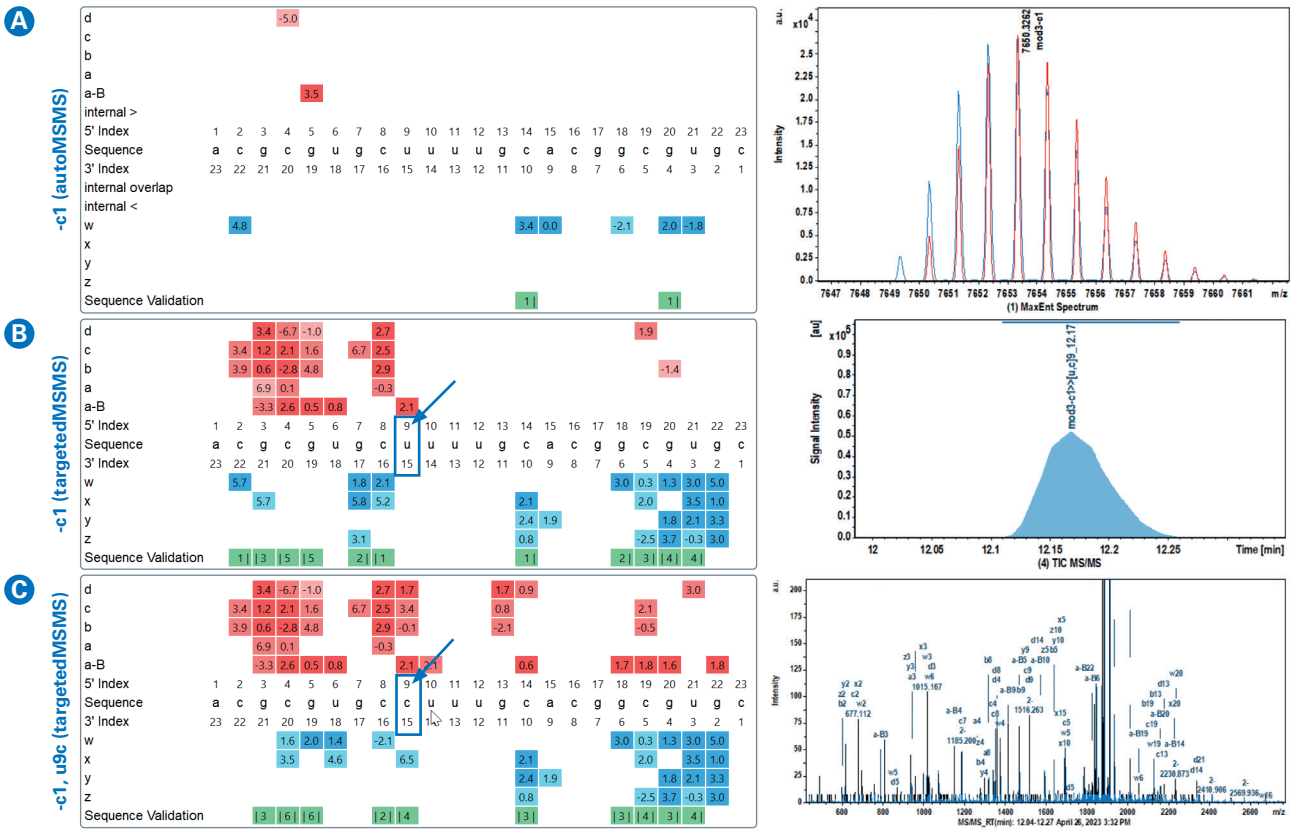


Figure 7
Sequence matches for the a/g (top) and a/c (bottom) exchanges at positions 2 and 16 in mod3 side products.

In Figure 7 the sequence matches for the exchange variants a/c and a/g are shown. The variants a16g and a2c are the best matches and were, therefore, marked "Confirmed" in the Expected table (Figure 1).

The elucidation of the -c1 side product benefits from the high isotopic fidelity of the data: the MS spectrum shows the presence of an additional sequence (Mr 7649.346), which is 1 Da lighter than the -c1 variant, and the match of the autoMSMS data to the sequence is not good enough to confirm the loss of c1 and exclude any loss of c at another position (Figure 8A). A targeted analysis of the most abundant charge state -4 at 12-13 min yielded a much better coverage of the sequence (Figure 8B) which allowed to assign the loss of c to residue 1. However, automatic screening for further sequence variants of mod3-c1 yielded the u9c-variant (mod3-c1,u9c) as best match (Figure 8C and D). This helps explain the isotope pattern of the mod3-c1. In fact, one can quantify the composition of that isotope pattern´s constituents based on the isotope pattern: the chromatographic peak at 12.17 min is composed of 22% mod3-c1 and 78% mod3-c1,u9c; both species are coeluting with this UHPLC separation and are only 1 Da apart.



Row	Result	Position	Mr Ref	Rt [min]	MSMS Score	Sample Name	Sequence
1		2	7650.3295	12.14	12.58	mod3-c1	5OH a c g c g u g c u u u u g c a c g g c g u g c 3OH

D

Expected | Unidentified

Row	Confirmed	Annotation	Mr Ref	Rt [min]	Intensity Coverage [%]	Sequence Coverage [%]	MSMS Score	Sequence
1	☑	mod3-c1>>[u,c]9	7649.3455	12.14	59.5	65.2	38.82	5OH a c g c g u g c c u u u g c a c g g c g u g c 3OH
2	☐	mod3-c1>>[u,c]11	7649.3455	12.14	57.9	60.9	35.23	5OH a c g c g u g c u u c u g c a c g g c g u g c 3OH
3	☑	mod3-c1	7650.3295	12.14	22.2	56.5	12.58	5OH a c g c g u g c u u u u g c a c g g c g u g c 3OH
4	☐	mod3-c1>>[u,c]12	7649.3455	12.14	57.9	60.9	35.23	5OH a c g c g u g c u u u c g c a c g g c g u g c 3OH
5	☐	mod3-c1>>[u,c]10	7649.3455	12.14	57.9	60.9	35.23	5OH a c g c g u g c u c u u g c a c g g c g u g c 3OH

Figure 8

Ⓐ Analysis of the -c1 variant by autoMSMS; the intact mass spectrum shows the presence of an added species at -1 Da. (Ⓑ, Ⓒ) The targetedMSMS analysis finds that variant as undergoing an additional u9c conversion with a high quality MS/MS spectrum, a conclusive sequence match and the highest MSMS score. Ⓓ Summary of the targetedMSMS analysis of mod3-c1.

Conclusion

- The OligoQuest workflow in Biopharma Compass enables the fast identification and characterization of oligonucleotide samples based on LC-UV-MS and MS/MS data. Meaningful reports can be generated automatically reducing analysis turnaround time. The reliability of the data can be visually verified through individual fragment ion inspection in the interactive interface for time-efficient review.
- Side products are initially characterized by accurate intact mass and isotope pattern, and by matching fragment ion patterns with those calculated from candidate sequences using data dependent acquisitions in the OligoQuest autoMSMS workflow. They can be quantified by concomitant UV-chromatogram records as part of the workflow down to lower than 1%.
- Lower abundant side products or those with unspecific MS/MS results can be further analyzed to a greater depth using the OligoQuest targetedMSMS workflow. Here, the side product candidate can be analyzed in an MRM-like approach and high-quality MS/MS spectra can be obtained in these more challenging cases. In this study, coeluting u-to-c exchange variants of such side products at -1 Da mass difference were found and quantified based on the high isotopic fidelity available on Bruker timsTOF and QTOF systems.
- The examples also illustrate the benefits of combining autoMSMS and targetedMSMS analyses: the former providing high quality intact mass and precursor isotopic information, the latter even generating sufficiently high quality MS/MS data to identify 2nd level "side-product of another side-product" species.

References

[1] Muslehiddinoglu J et al. (2020) *Technical Considerations for Use of Oligonucleotide Solution API*, Nucleic Acid Therapeutics, **30**/ 4, DOI: 10.1089/nat.2020.08

Bruker Switzerland AG

Fällanden · Switzerland
Phone +41 44 825 91 11

Bruker Scientific LLC

Billerica, MA · USA
Phone +1 (978) 663-3660

ms.sales.bdal@bruker.com – www.bruker.com

Learn more at **www.bruker.com/biopharma-compass**