


Molecular weight confirmation and impurity analysis of a 70-mer oligonucleotide

Sahana Mollah[1], Kerstin Pohl[1], Remco van Soest[1] and Ling Huang[2]
[1]SCIEX, USA; [2]Integrated DNA Technologies, USA

ABSTRACT

This poster describes the analysis of the integrity and purity of a 70-mer DNA adapter sequence used in next-generation sequencing (NGS). The molecular weight was determined with a mass accuracy of <2 ppm. For the NGS data quality, it is important to be able to differentiate between the loss of the phosphorothioate modified 3' terminal T (*T) and the loss of an A or G nucleotide. It is shown that by using chromatographic separation in combination with a TOF mass spectrometer, the -*T, - A and -G impurities were identified and quantified at levels between 1% and 5% of the full-length 70-mer with excellent precision and accuracy.

INTRODUCTION

Oligonucleotides longer than 70 nucleotides are increasingly used in applications such as gene therapy and next-generation sequencing (NGS). The length and integrity of these long oligonucleotides are critical quality attributes of these products. Oligonucleotide length analysis is also important in measuring poly(A) tail length to ensure the efficacy of mRNA vaccines.[1-2] Compared to short oligonucleotides, such as antisense oligonucleotide drugs, characterizing the length and integrity of long oligonucleotides by mass spectrometry (MS) is challenging because they do not readily fragment by collision-induced dissociation (CID) MS/MS. Identification is mainly performed by determining their accurate mass. Here, we describe a robust workflow that uses a high-speed TOF mass spectrometer to analyze the length, integrity and purity of long oligonucleotides. We determined the Molecular Weight (MW) and analyzed the purity of a 70-mer DNA with a phosphorothioate modified 3' terminal T used in NGS. The presence of 3' dT is important for NGS data quality. The detection and quantitation of the loss of the overhang *T is challenging due to the length of these oligonucleotides. Capillary electrophoresis with optical detection can be used to screen for impurities that differ by 1 nucleotide from the full-length product (FLP).[3-4] However, this technique cannot differentiate between the loss of the *T and the loss of another nucleotide. By using high-resolution TOF MS spectra and the chromatographic separation of the -*T impurity from the FLP and -A and -G impurities—the 2 most challenging impurities due to their very small mass differences with the -*T impurity—positive identification of the impurities and relative quantitation to the FLP was achieved.

MATERIALS AND METHODS

Samples

A 70-mer DNA oligonucleotide with 1 phosphorothioate bond at the 3' end (Seq70) was custom synthesized by Integrated DNA Technologies (IDT). Three impurity structures were also obtained from IDT, including structures with the loss of an internal A (Seq70-A), loss of an internal G (Seq70-G) and loss of the 3' phosporothioate T (Seq70-*T). The Seq70 FLP was desalted, and the impurities were HPLC purified and desalted. Samples were dissolved in a 10 mM Tris, 0.1 mM EDTA solution (IDTE solution, IDT) and diluted with mobile phase A to a 5 µM solution of the FLP with impurities spiked in at 1%, 2%, 3% or 5%.

Chromatography

A Shimadzu Nexera XS inert HPLC system was used with water as mobile phase A and 50/50 (v/v) water/methanol as mobile phase B, both containing 15 mM diisopropylethylamine and 50 mM 1,1,3,3,3-hexafluoroisopropanol. A gradient from 39–44% B in 10 min with a 4 min wash step at 95% B was used with a flow rate of 0.25 mL/min. The column was a Waters ACQUITY PREMIER Oligonucleotide C18 (2.1 × 150 mm, 1.7 µm, 130 Å) kept at 60ºC. The injection volume was set to 5 µL, resulting in a total of approximately 500 ng of oligonucleotides on column.

Mass spectrometry

A ZenoTOF 7600 system (SCIEX) was used in negative polarity mode. Source parameters were optimized for best sensitivity, and MS spectra were acquired from 405 m/z -1,850 m/z with an accumulation time of 0.25 s. Declustering potential was -80 V and a collision energy of -10 V was used.

Data processing

Data were processed using SCIEX OS software 3.0. Mass reconstruction to confirm MW was performed using the Bio Tool Kit module of the SCIEX OS software. Relative quantitation of the spiked-in impurities was performed using the Analytics module with the intact mass quantitation option.

RESULTS

Confirmation of MW

Determining the average MW of a larger oligonucleotide from a reconstructed spectrum is possible, but typically results in poor mass accuracy because reconstructed spectra for large oligonucleotides are non-Gaussian. For the 70-mer oligonucleotides studied here, the spectra acquired using the ZenoTOF 7600 system had sufficiently high resolution to determine the masses of the most intense isotopes after reconstruction with excellent mass accuracy. High-resolution TOF MS spectra were acquired for the FLP and for the 3 impurities. The Bio Tool Kit module in SCIEX OS software was used to reconstruct the intact MW. Reconstruction with a step size of 0.01 Da was used to obtain full baseline resolution of the isotopes of the FLP and impurities. The measured isotope pattern was highly correlated to the theoretically calculated pattern, further aiding the confident confirmation of the oligonucleotide structures (Figure 1).

Mass accuracy was calculated by comparing the observed mass to the theoretical mass of the most intense isotope. Excellent mass accuracy (<2 ppm) was observed for each oligonucleotide (Table 1).

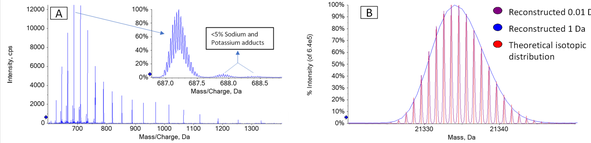

Figure 1. Extracted ion chromatogram (XIC), charge state distribution and reconstructed mass for the Seq70 oligonucleotide. A) Charge state distribution highlighting the -31 charge state of the Seq70 oligonucleotide. B) Reconstructed mass spectrum, shown with both full resolution (0.01 Da step size) and 1 Da resolution. The theoretical isotope distribution matched the reconstructed isotopically resolved spectrum well.

Table 1. Observed mass accuracy of the most intense isotope in the reconstructed spectra of the FLP and impurities (0.01 Da step size).

Analyte	Theoretical average MW (Da)	Theoretical mass of the most intense isotope (Da)	Reconstructed mass of the most intense isotope (Da)	Mass accuracy (ppm)
Seq70	21,334.761	21,334.564	21,334.579	0.70
*Seq70-*T*	21,014.502	21,013.539	21,013.534	-0.23
Seq70-A	21,021.554	21,020.504	21,020.538	1.61
Seq70-G	21,005.555	21,004.509	21,004.538	1.37

Identification of impurities

While the FLP and the most observed impurities have different MWs, the difference in mass between Seq70-*T and Seq70-A of only 7 Da is not sufficient to separate them using mass spectrometry. When the impurities are chromatographically separated, however, the correct reconstructed masses for these 2 impurities can be measured. The low methanol content of mobile phase B (50%) helped to generate the shallow gradient needed to separate the Seq70-*T and Seq70-A impurities. Figure 2 shows the separation of the FLP and the impurities. Figure 3 shows the reconstructed spectra for the 3 impurities spiked in at the 2% level. All 3 separated impurities could be confidently identified based on the TOF MS data.

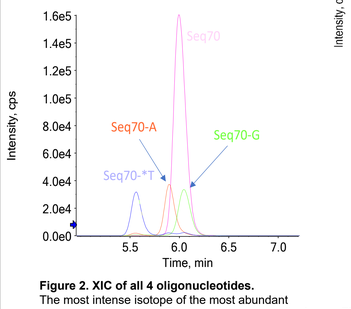

Figure 2. XIC of all 4 oligonucleotides. The most intense isotope of the most abundant charge state was used. The impurities were spiked in at 10% of the FLP. The 2 oligonucleotides with a 7 Da difference in MW, Seq70-*T and Seq70-A, were baseline separated using a shallow gradient.

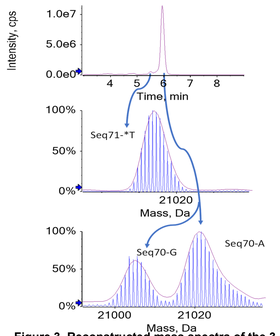

Figure 3. Reconstructed mass spectra of the 3 impurities spiked in at 2% of the FLP concentration. All 3 impurities could be identified based on their intact mass.

Relative quantitation of impurities

Using the intact quantitation option in the SCIEX OS software, the areas of the impurity mass peaks were calculated relative to the FLP for 5 µM samples of FLP spiked with 0%, 1%, 2%, 3% and 5% of the 3 impurities. Each sample was analyzed in triplicate. Figure 4 shows the area of the extracted ion chromatogram (XIC) that was reconstructed, the average spectrum and the integrated reconstructed mass spectrum for the 2% spike-in sample. The measured values found were in good correlation with the spiked-in amounts (Table 2). The Seq70-A impurity was present in the FLP, causing the values for this impurity to be higher than the spiked-in amount. Table 3 summarizes the average accuracy and precision for the different spiked-in amounts using a linear calibration. Accuracies were excellent and precision was better than 15%, except for the 3% Seq70-*T spike-in level, which had a precision of 21%.

Table 3. Quantitation summary

		Seq70-*T		Seq70-A		Seq70-G	
Spike-in %	Actual concentration (µM)	Average accuracy (%)	CV (%)	Average accuracy (%)	CV (%)	Average accuracy (%)	CV (%)
1	*0.05*	*90.5*	*2.15*	*81.8*	*10.3*	*96.3*	*3.63*
2	*0.10*	*110*	*2.68*	*119*	*12.5*	*103*	*6.79*
3	*0.15*	*104*	*21.3*	*107*	*15.3*	*103*	*10.5*
5	*0.25*	*95.4*	*6.52*	*91.7*	*2.19*	*97.8*	*2.91*

Spike-in % (vs. FLP)	Seq70-*T (%)	Seq70-A (%)	Seq70-G (%)
0	*0.026*	*1.5*	*0.067*
1	*1.1*	*2.6*	*0.84*
2	*1.9*	*4.0*	*1.7*
3	*3.0*	*5.8*	*3.0*
5	*4.7*	*8.3*	*5.2*

Table 2. Measured impurity levels (vs. FLP) for different spiked-in levels. Impurity levels are determined from the relative areas of the average MW peaks after mass reconstruction in the Analytics module of the SCIEX OS software.

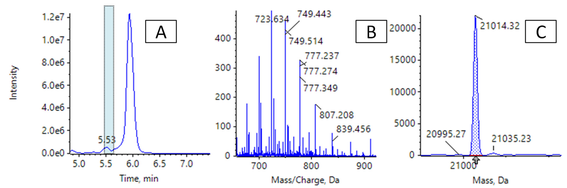

Figure 4. Intact quantitation processing of the Seq70-*T impurity at the 2% spiked-in level using SCIEX OS software. A) The area of the XIC that was used for reconstruction. B) The average spectrum. C) The integrated area of the reconstructed spectrum.

CONCLUSIONS

- The MWs of a 70-mer oligonucleotide and 3 of its (n-1) impurities were determined using the ZenoTOF 7600 system
- Full isotopic resolution and reconstructed mass accuracy of <2 ppm were achieved using high-resolution TOF MS data
- Impurities that could not be distinguished by mass spectrometry only were separated chromatographically and correctly identified and quantified
- Accurate and precise relative quantitation of 3 spiked (n-1) impurities in the range from 1%-5% were achieved using the easy-to-use intact quantitation feature of SCIEX OS software
- The workflow described for an NGS adapter can be used to analyze the length and integrity of poly(A) tails for mRNA vaccines and for single guide RNA for CRISPR

REFERENCES

1 Jalkanen AL, Coleman SJ, Wilusz, J. Determinants and implications of mRNA poly(A) tail size – Does this protein make my tail look big? Semin Cell Dev Biol. 2014; 34:24-32.
2 Schlake T, Thess A, Fotin-Mleczek M et al. Developing mRNA-vaccine technologies. RNA Biol. 2012; 9(11):1319-30.
3 RNA Analysis for CRISPR by capillary electrophoresis with laser-induced fluorescence detection. SCIEX technical note, RUO-MKT-02-10432-B.
4 Capillary electrophoresis of oligonucleotides. Integrated DNA Technologies, Inc., 2011.

TRADEMARKS/LICENSING

To receive a copy of this poster:
- Scan the code with your phone camera
- Complete the form