Myth busting:
"You cannot sequence oligonucleotides over 20 to 30 nucleotides long by LCMS/MS"
Learn how to routinely sequence 100 nt oligonucleotides

Alexander Schwahn[1], Ulrik Mistarz[2], Keeley Murphy[3], and Ken Cook[4]; Thermo Fisher Scientific: 1) Reinach, Switzerland; 2) Copenhagen, Denmark; 3) Massachusetts, USA, 4) Hemel Hempstead, UK

ABSTRACT

Development of a new liquid chromatography high-resolution mass spectrometry (LC-HRMS) method to sequence large oligonucleotides. A chromatography method was used which maintains a high pH to promote higher charge states beneficial to HCD fragmentation of oligonucleotides. The HCD fragmentation was optimized for sequencing of long oligonucleotides in conjunction with the high charge states produced during the chromatography. A Thermo Scientific™ Orbitrap Exploris™ 240 Mass Spectrometer was used for sensitivity, with optimised source conditions known to prevent adduct formation. Thermo Scientific™ BioPharma Finder™ Software was used for identification and sequence mapping of the long oligonucleotides up to 100mer.

INTRODUCTION

The resent boost of oligonucleotide therapeutics has increased the need for new and improved analytics, especially for the characterization of mRNA and longer chain length species such as sgRNA. The highly charged, linear chain structure provides analytical challenges for the chromatography and mass spectrometry for sequence analysis. A successful routine method has been developed for large oligonucleotide sequencing which goes far beyond previously believed limits of 20 to 30 nt lengths. In this study, optimization advice will be presented for sequence analysis of large oligonucleotides up to 100 nt in length with 100% coverage. Evidence is given on how the chromatography conditions are intrinsically linked to provide the MS with the appropriate charge states for successful fragmentation, required for sequence analysis. HCD fragmentation conditions have been studied for short and long oligonucleotides and found to require different conditions. A system suitability test has been developed with a 6-component ssDNA standard to ensure the LC-MS/MS conditions and software are operating at optimum levels. We previously developed a partial digest used for mRNA sequencing which produces large fragments that utilizes these new optimized conditions for sequencing[1]. This enables direct sequence verification of mRNA by LC-MS/MS. Guide and transfer RNA are also at a length which can now be directly sequenced. Commercially available software is used which automates the annotation of the fragment sequence analysis and maps these to known large oligonucleotide sequences. Data filters have also been designed to prevent false identifications in the analysis. These new sequencing techniques will benefit and help advance many areas of oligonucleotide therapeutics. The ability to rapidly identify, characterise and sequence map large oligonucleotide therapeutics with high sequence coverage provides important information for identity testing, sequence validation, and impurity analysis.

MATERIALS AND METHODS

Sample preparation: Synthetic modified RNA oligonucleo-tides were purchased from Thermo Fisher Scientific and were diluted in water to 1 mg/ml.

UHPLC Separation: IPRP separations were performed with a DNAPac RP column (4 µm, 2.1 x 100 mm) using a Thermo Scientific™ Vanquish™ Flex Binary UHPLC system. The eluent system was developed to promote higher charge states while maintaining high resolution:

Eluent A: 20 mM TEA, 80 mM HFIP in water

Eluent B: 20 mM TEA, 80 mM HFIP in acetonitrile or methanol
flow rate 0.3 mL/min; 50ºC with active eluent pre-heating.

Mass Spectrometry:

Characterization assays were performed on an Orbitrap Exploris 240 mass spectrometer. General settings are listed in Table 1

Data Analysis:

Biopharma Finder 5.1 softwarewas used for identification and sequencing. A report was generated with flexible impurity annotation.

Table 1: Orbitrap Exploris 240 MS parameter settings

Global Parameters			
Application Mode:	Intact Protein		
Pressure Mode:	Low pressure		
Negative Ion (V):	2500	ITT Temp. (ºC):	320
Sheath gas (Arb):	35	Vaporizer Temp. (ºC):	300
Aux gas (Arb):	10		
MS1 Full scan			
Orbitrap Resolution:	120k	Scan Range (m/z):	450 – 2500
Microscans (10–55mer):	1	RF Lens (%):	75
Microscans (100mer):	3	Polarity:	Negative
AGC target (%):	300	max IT (ms):	200
ddMS² (Top 2)			
Charge states:	2 - 40	Min. Intensity:	1e4
Dynamic exclusion at fist occurrence for 5 sec within ±10 ppm			
Isolation window (m/z):	4	Collison energy type:	normalized (z=2)
HCD collision energy (%): stepped, values as indicated in the figures			
Orbitrap Resolution:	60k	Scan Range (m/z):	Auto
Microscans (10-55mer):	1	AGC target (%):	200
Microscans (100mer):	5	Max IT (ms):	200

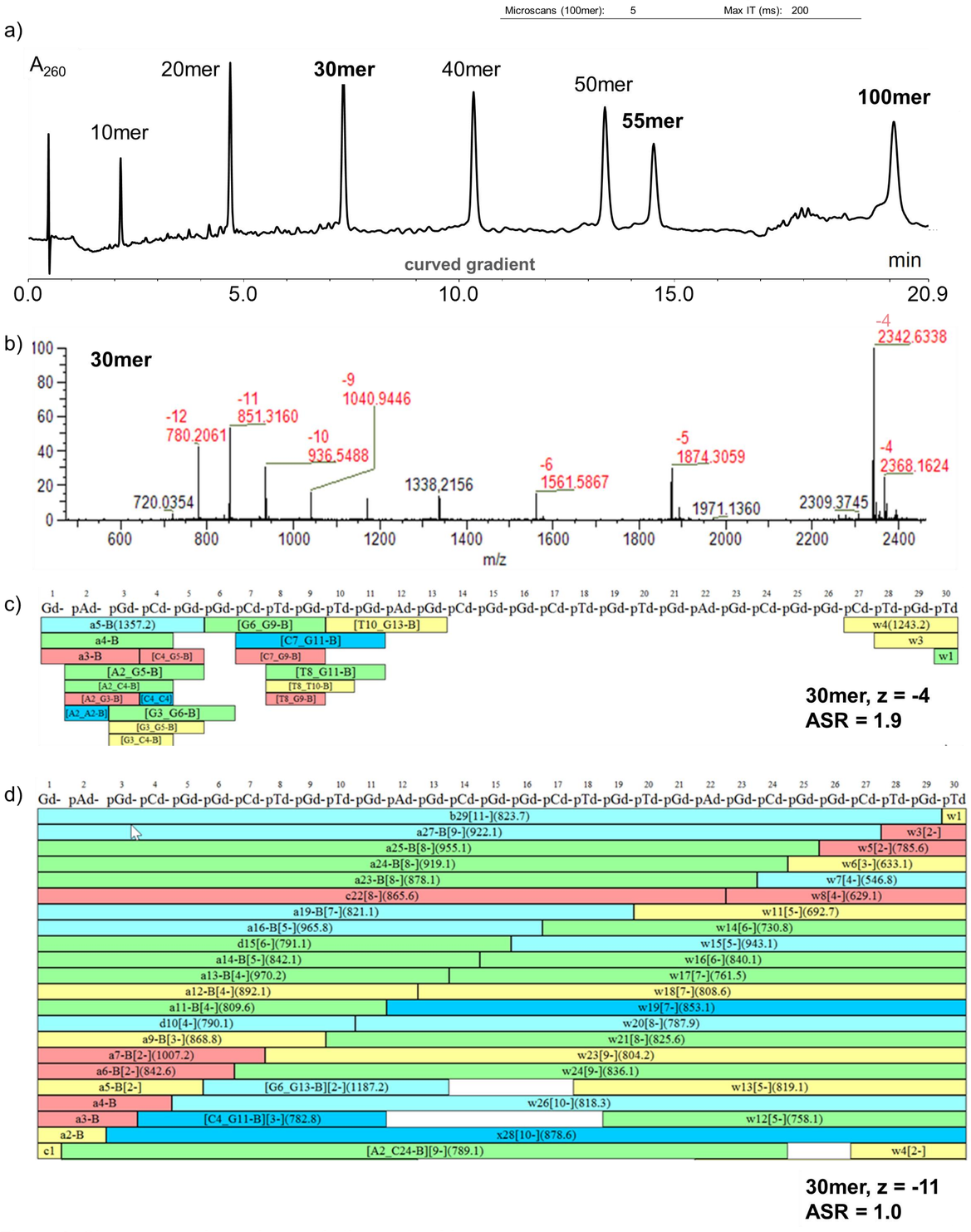

Figure 1: Charge state-dependency of oligonucleotide sequence coverage

UV chromatogram of ssDNA mix (10, 20, 30, 40, 50, 55, and 100 nt), b) mass spectrum for the 30mer ssDNA oligonucleotide, c) and d) fragment coverage map for charge states -4 and -11, respectively; ASR … average structural resolution (ASR = 1.0 indicates all phosphodiester bonds are verified). Colored bars under the oligonucleotide sequence indicate detected fragment ions. Stepped normalized collision energies of 15,18,21 where used.

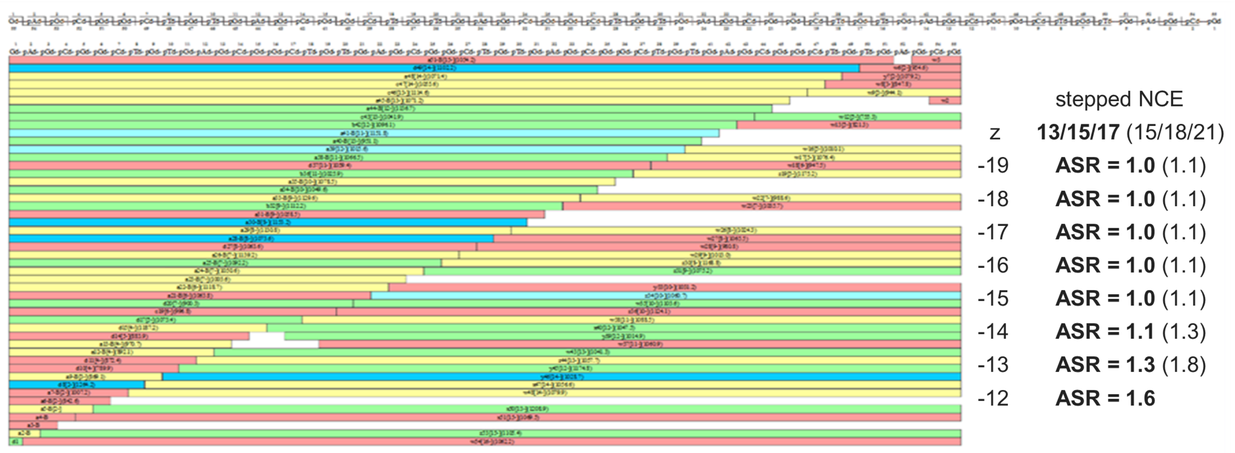

Figure 2: Optimization of fragmentation conditions for 55mer sequencing

Average structural resolution (ASR) improves for lower stepped collision energies (normalization to z = 2) and higher charge states used for fragmentation

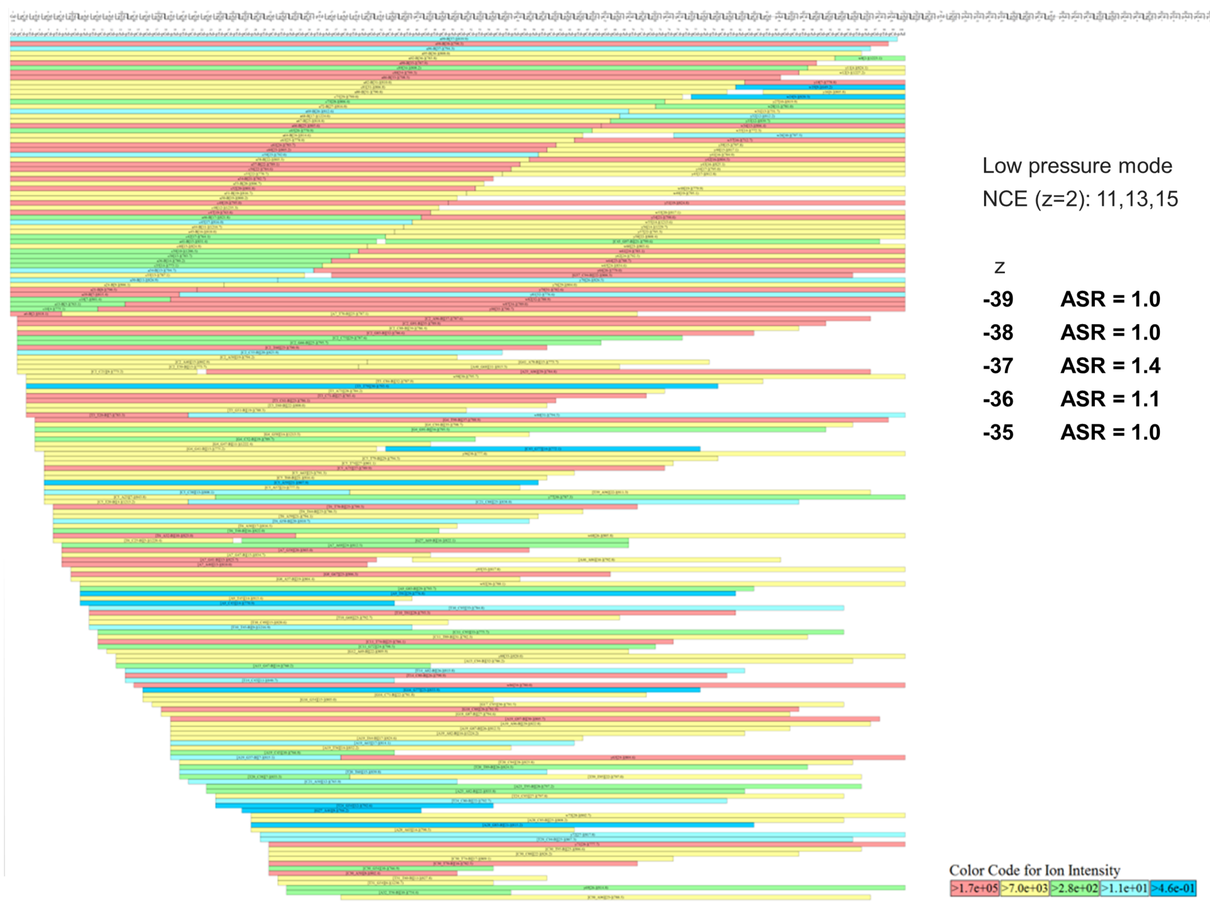

Figure 3: Complete sequence coverage of a 100 nt ssDNA oligonucleotide

RESULTS

Method optimization

Large synthetic oligonucleotides are relatively new therapeutics which have proven difficult to sequence by MS and fully characterize. The 6-component standard contains 10, 20, 30, 40, 50, and 55mer ssDNA. This allows optimization of the chromatography as well as the HCD fragmentation. A simple experiment, using different stepped normalized collision energies (NCE) energies, allows to identify fragmentation energies that result in good sequence coverage over the size range of the oligonucleotides, and also which charge states yield optimum sequence coverage. Selected results are shown in figure 1. The chromatography gives baseline separation of all components. The 30mer shows the most abundant charge state to be -4 yet the best information for sequencing are derived from the higher charge states. Charge state -4 gives an average structural resolution (ASR) of 1.9. Charge state -11 has an ASR of 1.0, which identifies all nucleotide bases present and confirms the sequence.

ASR values provides an indication of the level of fragmentation and identification of the oligonucleotide. In the optimum case, all bonds between individual nucleotides will be broken giving identified fragments for the entire sequence. This results in an ideal score of 1.0. The results shown for the 30mer were obtained using HCD fragmentation energies of 15,18,21.

Smaller 10 and 20mer standards were relatively easy to sequence with most fragmentation energies, however, slightly better results were obtained with elevated energies up to 21,23,25. The larger oligonucleotide standards of 50 and 55mers required lower fragmentation energies to give good sequence data. This is to the extent that the use of higher fragmentation energy results in too much fragmentation of these long oligonucleotides with no useful fragments being found to unambiguously match against the target sequence. This on its own may be a reason for many reports of difficulties in sequencing oligonucleotide fragments over 20 to 30nt long.

Figure 2 shows the data from 2 different fragmentation energies used with the 55mer standard. The best sequence coverage arises from the lower fragmentation energy of 13,17,19. This resulted in an ideal ASR fragmentation score of 1.0 for 5 different charge states. Taking the fragmentation up to 15,18,21 starts to reduce the quality of fragments for sequence coverage. If higher fragmentation energies of 21/23/25 are used, which would work fine with a 20mer, then there are very few fragments even identified for sequence analysis.

In addition to the need for lower energy fragmentation, it is also apparent from figures 1 and 2 that the higher charge states give better sequence data. The ASR score drops as the charge state is reduced. This shows a link between the chromatography conditions and the sequencing by LCMS. If high pH eluents are used (e. g. a lower concentration of HFIP), this promotes the higher charge states that give the superior sequence data. The eluents used here are all between pH 9.5 and 10.0. This requires thought concerning the chromatography column used. The column used here is a polymeric phase (stabile form pH 0-14) that can withstand the high pH required for this analysis. Silica columns would not survive long at these alkaline conditions.

Using the optimization techniques described here, a 100mer was exposed to direct sequence analysis. This is now in the size realm of guide and transfer RNA. As it was previously thought that this size of oligonucleotide could not be directly sequenced by MS, most prevalent analytical methods require prior digestion to smaller fragments. This pre-requisite is completely abolished using the method described here, yielding a perfect ASR of 1.0 for 3 different charge states (Fig. 3.)

CONCLUSIONS

High-resolution mass spectroscopy using an Orbitrap Exploris 240 system with HCD fragmentation, gives the sensitivity and resolution required to sequence large oligonucleotides previously deemed too large for direct MS-sequencing.

Optimization of the chromatography and fragmentation energy makes this analysis routine and available to therapeutic oligonucleotides of large lengths.

The optimized new method using LC/HRMS is robust, simple and shows considerable advantages to methods requiring digestion into smaller fragments.

The use of BioPharma Finder software fully annotates the results and finds sequence failures present in the sample.

REFERENCES

1. Vanhinsbergh et al. (2022), Characterization and Sequence Mapping of Large RNA and mRNA Therapeutics Using Mass Spectrometry, Analytical Chemistry 94(20):7339-7349

TRADEMARKS/LICENSING