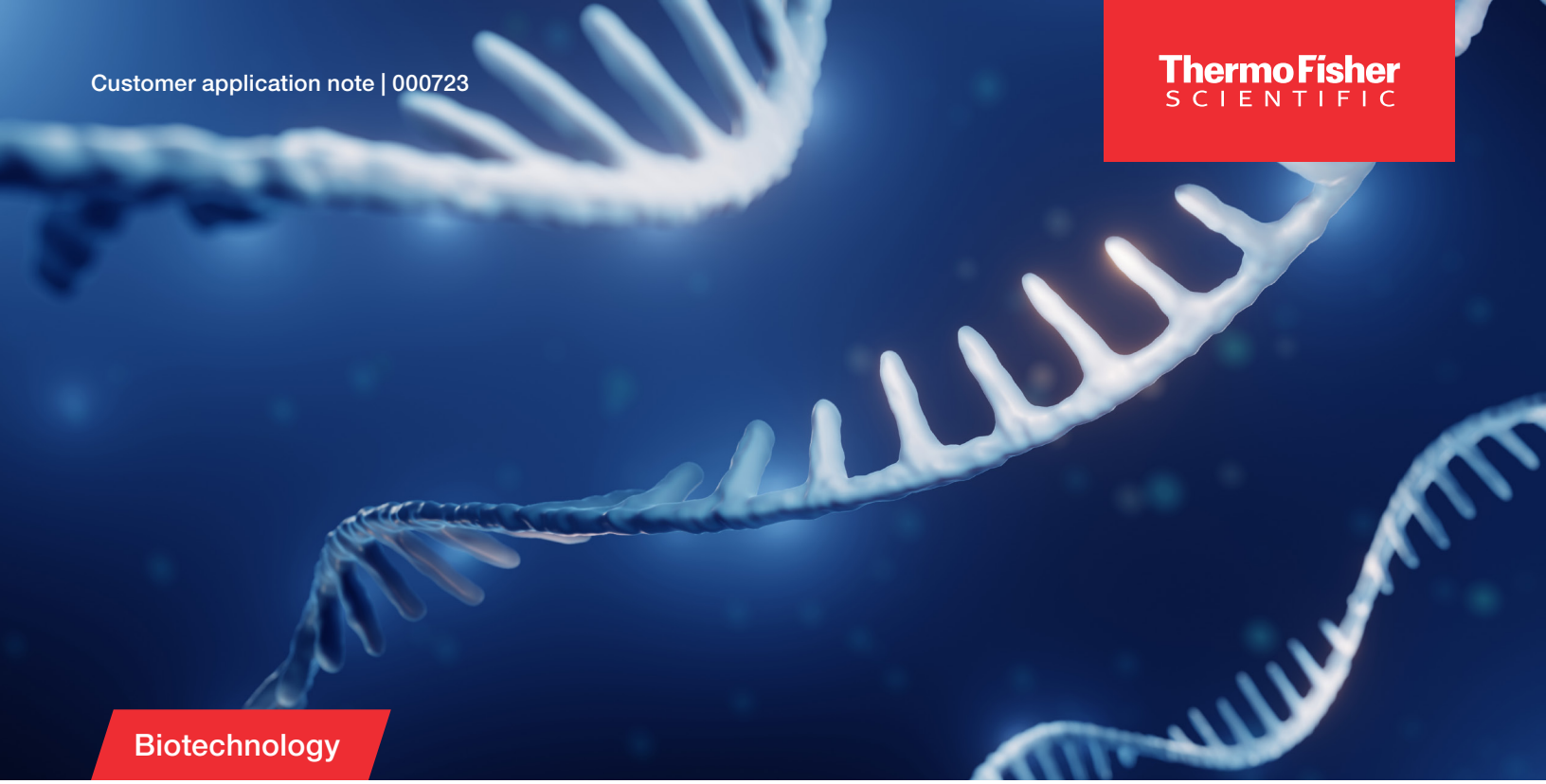
Customer application note | 000723



Biotechnology

mRNA direct sequence mapping using automated partial digestion with magnetic nuclease and LC-HRMS

Authors

Mark Dickman[1], Christina Vanhinsbergh[1], Jon Bardsley[2], Ken Cook[2], Andrew Williamson[2], Jennifer Sutton[3], Keeley Murphy[3]

[1]University of Sheffield, Sheffield, UK
[2]Thermo Fisher Scientific, Hemel Hempstead, UK
[3]Thermo Fisher Scientific, San Jose, CA, USA

Keywords

mRNA sequence mapping, oligonucleotide, DNAPac RP column, Orbitrap mass spectrometer, BioPharma Finder software

Application benefits

- Direct sequence analysis of large mRNAs by LC-HRMS
- Reproducible and comprehensive sequence coverage at > 85%
- Simple automated workflow suitable for unmodified and modified mRNA
- Automatic data annotation and mapping for simpler data analysis

Goals

- Develop a simple and reproducible LC-MS based workflow for direct sequencing of mRNA
- Achieve reproducible controlled partial RNase digestion to generate large sequence specific oligonucleotide fragments with magnetic bead technology in an automated workflow
- Optimize separation and identification of complex mRNA digest with unique column technology, robust liquid chromatography, and high-resolution accurate mass (HRAM) mass spectrometry
- Obtain confident sequence identification of oligonucleotide fragments and comprehensive mapping of mRNA with state-of-the-art software

Introduction

Oligonucleotide analysis has gained considerable interest over the last few years with the successful introduction of mRNA vaccines encoding for a modified version of the SARS-CoV-2 spike protein. This has opened the door for other mRNA vaccines to be developed including mRNA vaccines for HIV now in clinical testing.

The popularity in oligonucleotide therapeutics has increased the need for new and improved analytics of these highly charged biological polymers, especially the detailed characterization of mRNA. The unique, highly charged, linear chain with secondary folding has difficult analytical challenges for sample preparation, chromatography, and mass spectrometry detection. Protein sequence verification by LC-MS is a routine test for therapeutic proteins, however, a direct sequencing method has never been developed for large oligonucleotides. Rapid DNA sequencing can be accomplished with new generation sequencers. mRNA, on the other hand, would first have to be converted to DNA with a reverse transcriptase, then the resulting DNA amplified before sequencing could be done. The conversion and amplification would lose any impurities and other information from the initial mRNA product.

In this study, a workflow will be presented for direct sequence analysis of mRNA using UHPLC and HRMS fragmentation. As the RNA is composed of only four different bases, a complete nuclease digestion produces short fragments that are not unique to any position in the long sequence. To overcome this, we developed a reproducible partial RNase digestion using RNase T1 immobilized on magnetic beads. The partial digest was optimized to give enough missed cleavages to produce larger unique sequence fragments that can be mapped to give high sequence coverage of the mRNA. Thermo Scientific™ BioPharma Finder™ software, version 5.0 is used to automate the annotation of the fragment sequence analysis and map these to the known mRNA sequence. Data filters were also designed to prevent false identifications in the analysis. Using this approach > 80% sequence coverage of a range of large RNA products was achieved, including the SARS Co-V2 spike protein, from a single analytical run. The total workflow, including sample preparation, can be done in 90 minutes. The digestion reproducibly produces the same cleavages and missed cleavages to allow a specific chromatographic fragmentation pattern that could be used in a QC environment. The ability to rapidly identify, characterize, and sequence map mRNA therapeutics with high sequence coverage provides important information for identity testing, sequence validation, and impurity analysis.

Experimental

Chemicals

- Deionized water, UHPLC-MS grade, Thermo Scientific™ (P/N W81)
- 1,1,1,3,3,3-Hexafluoro-2-propanol, LC-MS grade, Thermo Scientific™ (P/N AC293410500)
- Triethylamine, extra pure, Thermo Scientific™ (P/N 10621643)
- Methanol, UHPLC-MS, Thermo Scientific™ (P/N A458-1)
- Thermo Scientific™ SMART Digest™ Bulk magnetic RNase T1 Kit (P/N 60120-101)

Instrumentation

- Thermo Scientific™ Orbitrap™ Exploris™ 240 mass spectrometer (P/N BRE725535)
- Thermo Scientific™ Vanquish™ Horizon system consisting of:
  - Vanquish System Base (P/N VH-S01-A)
  - Vanquish Binary Pump H (P/N VH-P10-A)
  - Sampler HT (P/N VH-A10-A)
  - Vanquish Column Compartment H (P/N VH-C10-A-02)
  - Vanquish Diode Array Detector HL (P/N VH-D10-A)
  - Thermo Scientific™ Vanquish™ LightPipe™ Standard flow cell, 10 mm (P/N 6083.0100)
- Thermo Scientific™ KingFisher™ Duo Prime Purification System (P/N 5400110)

LC conditions

Column	Thermo Scientific™ DNAPac™ RP, 100 × 2.1 mm i.d. (P/N 088923)
Mobile phase A	0.2% Triethylamine (TEA), 50 mM 1,1,1,3,3,3-Hexafluoro-2-propanol
Mobile phase B	0.2% Triethylamine (TEA), 50 mM 1,1,1,3,3,3-Hexafluoro-2-propanol, 20% v/v acetonitrile
Column temperature	60 °C
Flow rate	0.2 mL/min

Table 1. LC gradient conditions

Time (min)	A	B
0	98	2
40	75	25
45	75	25
45	98	2
60	98	2

Mass spectrometer settings

Vaporizer temperature	450 °C
Ion transfer tube temperature	350 °C
Source voltage	Negative ion 2,500
Scan mode	DDA
Polarity	(-) negative
MS1 scan	450–3,000
MS2 scan	150–2,000
MS1 resolution	120,000
MS2 resolution	30,000
MS1 Automatic Gain Control	200%
MS2 Automatic Gain Control	50%
Isolation window	4 *m/z*
Collision mode	HCD
Collision energy	Stepped (15, 18, 21)

Controlled mRNA partial digestion

Partial RNase digestions were performed using 20–40 µg of RNA incubated with 1.25–5 µL of immobilized RNase T1 at either 60 °C or 37 °C for 5–15 min in a volume of 50 µL of the SMART Digest RNase buffer. Reactions were stopped by the magnetic removal of the immobilized RNase T1. Automated RNase digestions were performed on a KingFisher Duo Prime system controlled by Thermo Scientific™ BindIt™ software (version 4.0). A 96-deepwell plate was set up with 50 µL of SMART Digest RNase buffer containing 20–40 µg of RNA samples in Row A and 2.5–5 µL RNase T1 immobilized on magnetic beads within 50 µL of SMART Digest RNase T1 buffer in Row G. The KingFisher Duo Prime system was programmed to transfer RNase T1 immobilized magnetic particles to Row A to digest the RNA at 37 °C for 5–15 minutes. Sedimentation of beads was prevented by repeated insertion of the magnetic comb using the mixing speed setting “Fast”. Immediately after incubation, the magnetic beads were collected and removed from the reaction, and the digest solution was actively cooled to 15 °C. Complete RNase digests were performed using 10–20 µg of RNA with the addition of 100 µg RNase T1 at 37 °C for 4 hrs in 0.1 M TEAA. Subsequently, 10–20 µg of digested RNA was analyzed using LC-MS/MS.

LC-MS analysis of mRNA digest

RNA digests were separated by IP-RP-HPLC on a Vanquish binary gradient UHPLC system using a DNAPac RP column. Chromatographic separation details are shown in Table 1. The column temperature was set to 60 °C. UV detection was at a wavelength of 260 nm. Eluted oligonucleotides were analyzed using an Orbitrap Exploris 240 MS instrument Data acquisition was performed using a data-dependent MS/MS experiment.

Data processing and sequence mapping

Data analysis was performed in Thermo Scientific™ BioPharma Finder™ version 5.0 software using the basic default method in the oligonucleotide sequencing module. To identify large fragment ions, the maximum oligonucleotide mass was set to 25,000 Da, minimum confidence at 0.5, and mass accuracy at 10 ppm. The ribonuclease selection was set to RNase T1, specificity level set at “strict”, and phosphate location was set at “none”. Phosphorylation and cyclic phosphorylation were set as variable modifications of the 3' terminal in the sequence manager containing the RNA sequence. Random RNA sequences of the same length and GC content were included in the sequence manager in addition to the correct RNA sequence. For data processing and review, additional filters were included to discount nonspecific identifications to contain MS/MS data in each identification with a confidence score above 90%, best overall structural resolution below 2.0, and delta mass accuracy 20 ppm.

Results and discussion

Complete digestion of mRNA therapeutics using RNases such as RNase T1/A results in the production of a large number of small oligoribonucleotides that map to multiple different locations throughout the RNA sequence and cannot be used for sequence mapping. Many of these oligoribonucleotide fragments will be isomeric and cannot even be identified based on high-resolution accurate mass analysis (HRAM) alone. As such, partial RNase T1 digestion in conjunction with high-resolution LC-MS/MS using an automated RNA sequence analysis method was developed. Partial RNase T1 digests are very difficult to control in solution. Using RNase T1 immobilized onto magnetic beads enables simple control over the digestion time by removing the beads at the end of the reaction time. The magnetic beads also enable automation of the workflow on an automated robotic liquid handling system such as the KingFisher Duo system. Effective removal of the RNase T1 also prevents build-up of the nuclease on the HPLC column, which can potentially further digest the RNA during the chromatography. Following partial RNase T1 digestion, the oligonucleotide fragments were separated using ion pair reverse phase HPLC (IP-RP-HPLC) using TEA/HFIP on a DNApac RP column. The separation was optimized for high resolution between 10 to 50 nt fragment lengths. These are the useful fragment sizes for unique identification with high-resolution mass spectrometry analysis. Identification and sequence analysis were performed using Biopharma Finder software version 5.0. A schematic illustration of the total mRNA sequencing workflow is shown in Figure 1.

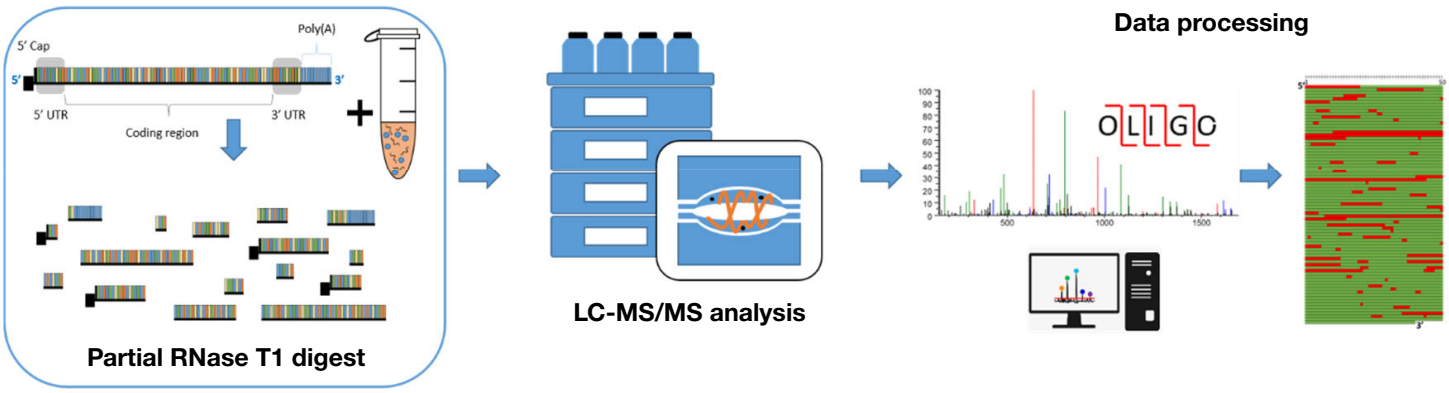


Figure 1. Schematic representation of direct mRNA sequencing workflow

The data analysis software provides automated tools for the identification and mapping of the chromatographic peaks from a mRNA sample digest. The monoisotopic mass and the MS/MS fragmentation pattern of the identified components are compared to the predicted oligonucleotide components of the experimental digest. Identification is based on the evaluation of mass accuracy, isotopic distribution, and charge state determination, as well as through the comparison of experimental and predicted MS/MS fragmentation spectra. A high confidence score indicates a good similarity match and that the probability of obtaining a fragment pattern matching the predicted sequence would be low when compared to a random sequence. In addition, the software automatically calculates an average structural resolution (ASR) value, in the ideal case all bonds between each individual nucleotide residue will be broken and resulting fragment ions matched to the predicted MS/MS spectra. A score of 1.0 indicates that each nucleotide bond in the sequence has been fragmented and matched to the predicted MS/MS spectra of the oligoribonucleotide sequence. The combination of a high confidence score with low delta mass ppm deviation and a low ASR value gives strong confidence in the sequence being correctly matched. The data analysis software also reports the % RNA sequence coverage based on the unique oligoribonucleotides identified and provides powerful visualization tools that show the identified unique oligoribonucleotides mapped to the RNA sequence.

Optimization of the digestion

The digestion is fast and easy to set up using only the magnetic beads, buffer, and sample. The aim would be to produce fragments between 10 and 50 nt long. Digestion optimization can be achieved by altering the time or varying the number of beads introduced. A typical unmodified mRNA of 4,000 bases would optimally digest in 5 to 15 minutes with 2.5 µL of the bulk beads. A modified mRNA as used in the mRNA vaccines would take slightly longer, between 20 to 30 minutes of digestion.

We performed direct RNA sequence mapping on a number of different mRNA samples and long RNA, including mRNA corresponding to modified and unmodified versions of the SARS CoV-2 Spike protein (~3,900 nt). Analysis of three replicate partial RNase T1 digests where all digest conditions were constant is shown in Figure 2A, with three different RNA samples shown in Figure 2B. The results demonstrate that under these conditions similar total ion chromatograms (TICs) were generated across the three replicates and that each RNA sample shows a specific fragmentation pattern. These results indicate that the missed cleavages are not generated randomly. RNase T1 will preferentially cleave at accessible single stranded regions within the RNA and, therefore, will generate fragments from single stranded loop structures first. Cleavage at sites originally protected in the regions of secondary structure will occur later as the RNA unfolds during digestion. In this way the pattern of oligoribonucleotide fragments produced is reproducible and dependent on the sequence and the secondary structure of the large RNA.

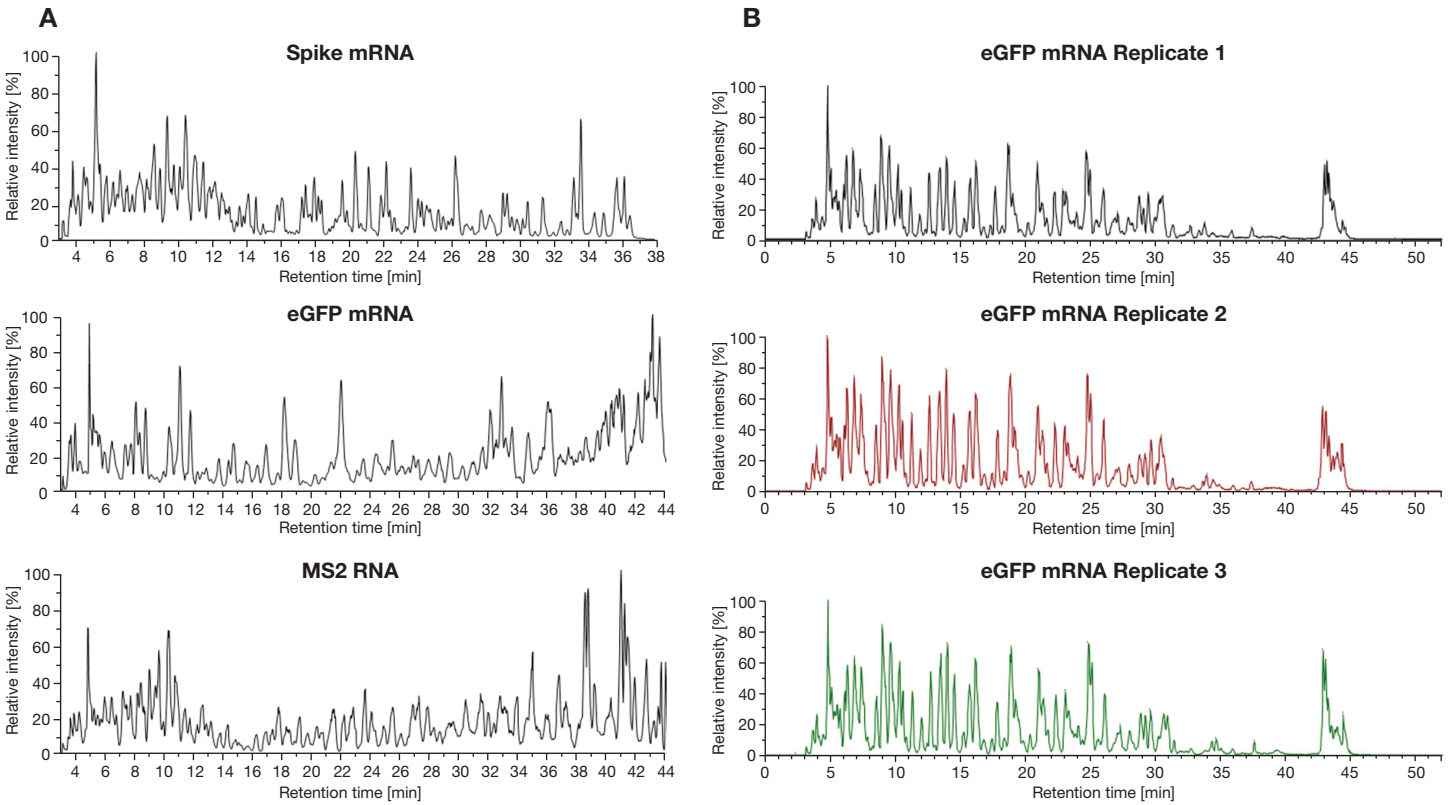
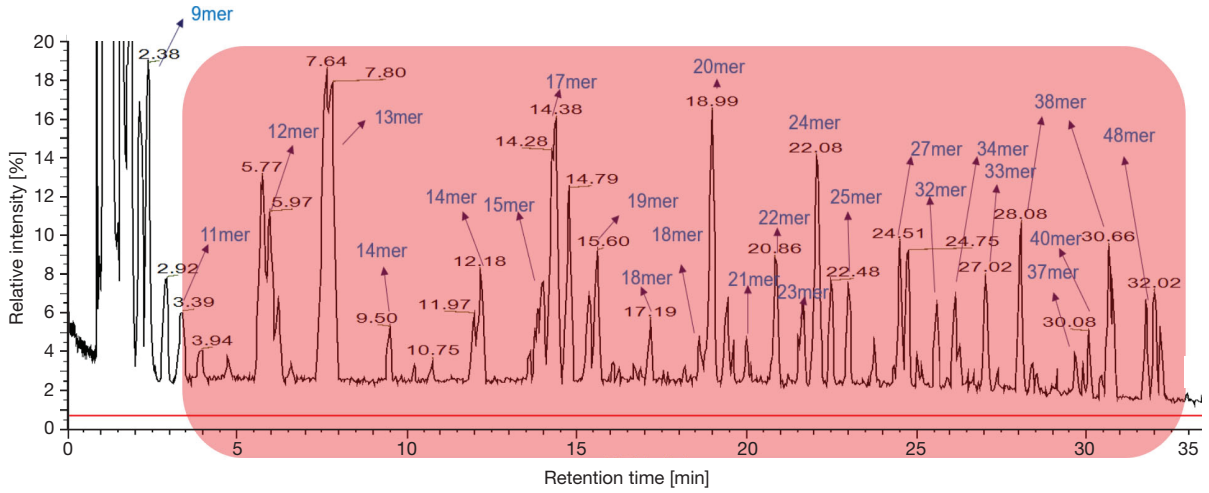


Figure 2. Analysis of three replicate partial RNase T1 digests where all digest conditions were constant. (A) Total ion current for partial T1 digests for spike protein, eGFP, and MS2 mRNA. All digests were obtained using the conditions described in the methods section with 40 g mRNA in each digest of 200 µL. (B) Chromatograms and sequence coverage of three replicate eGFP mRNA digests. Highly reproducible separation and sequence coverage were obtained.

To further examine the reproducibility of the partial RNase T1 digests, the sequence coverage using unique oligoribonucleotide identifications was compared. The results in Figure 2B show that across the replicate eGFP RNase T1 digests, the mean sequence coverage was 70.6% (RSD 1.7%) under the conditions used. These results highlight the reproducibility of the oligoribonucleotide digestion and data analysis.

The high-resolution chromatography also provides some chromatographic separation of isomeric oligonucleotides. This aids in the accurate identification of the digestion fragments by enabling the collection of diagnostic MS/MS fragmentation ions without interference from overlapping isomers. Figure 3 shows the typical resolution from an mRNA digestion of the unmodified spike protein mRNA sequence that is used as the basis of the mRNA COVID vaccines. The main peaks are indicated with the fragment length [blue]. The separation is mainly based on size, but different sequences of the same length are also separated. The table shows the sequence coverage at near 90% after applying the filters described in the methods section. The sequence coverage is based only on unique sequence matches. Non unique matches from small fragments[*] were at 0%. The unidentified peaks were found almost exclusively in the early eluting small fragments or late eluting larger fragments that could not be sequenced.



Oligonucleotide	number of MS peaks	MS peak area	Sequence coverage	Abundance (mol)
1: Spike	3387	28.6%	89.3%	100%
1: Spike*	3387	28.6%	0%	100%
Unidentified	7755	71.4%	--	--

Figure 3. The TIC trace of a partial magnetic T1 nuclease digestion. The digest was obtained using the conditions described in the methods section with 40 µg mRNA, 2.5 µL of beads in 200 µL of SMART Digest buffer. The nucleotide fragment length is indicated in blue and the pink shaded area donates the useful chromatographic region for unique oligonucleotide sequence information. The table shows the sequence coverage using unique and non-unique[*] identifications.

Sequence confirmation by MS/MS

The Orbitrap Exploris mass spectrometer in combination with BioPharma Finder software will generate fragmentation spectra from multiple charge states from each oligonucleotide. This increases the confidence of the oligoribonucleotide identifications, as the multiple higher charge states yield more informative fragments for sequencing. Using this process, routine sequencing of fragments up to 50 nucleotides in length is possible. The larger the fragment increases the likelihood of it being uniquely matched in the total sequence. Isomeric oligonucleotides containing the same base composition were typically separated during the chromatography and the MS/MS fragmentation data with the automated sequence annotation enabled identification of the sequence isomers as shown with one example in Figure 4. Two 16 nucleotide fragments of identical base composition were separated during the chromatography and correctly sequenced, identified, and mapped to the original sequence.

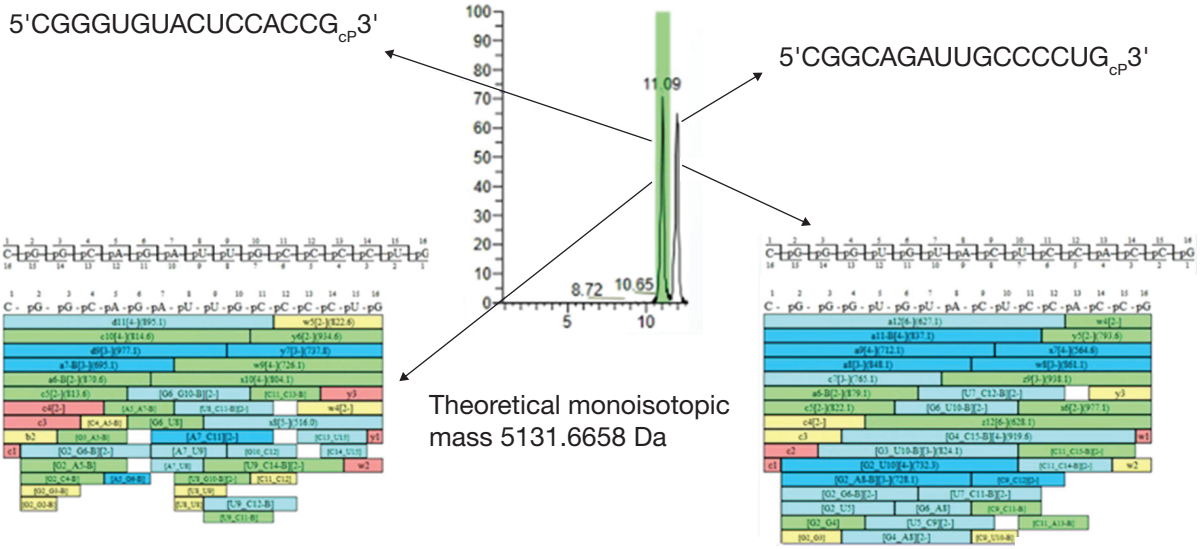


Figure 4. Extracted ion chromatogram showing the mass 5131.6658 fragments in a SARS CoV-2 spike protein mRNA sequencing experiment. The sequence information is directly from BioPharma Finder software, which also shows the position in the original mRNA sequence, the full scan spectra, and the MS/MS spectra compared to the predicted spectra (not shown).

Inbuilt database searching was performed against the correct RNA sequence and a random RNA sequence of the same GC content and size for each corresponding RNA. The % sequence coverages show that analysis of partial RNase T1 digests revealed sequence coverage of SARS CoV-2 spike protein and eGFP mRNA of 89.3 and 86.2%, respectively. Sequence coverage of the eGFP mRNA, not including the polyA tail, was 95.9%. For each data set analyzed there was very low or no matches against a random RNA sequence of the same GC content and size. This is another indication of the validity of the identifications. The results show that the partial RNase T1 digests in conjunction with LC-MS/MS analysis can identify, with high sequence coverage, mRNA of multiple different samples in a single analysis.

The vaccines made by Moderna and Pfizer–BioNTech use mRNA that has been chemically modified to increase stability and efficacy. Therefore, further work was performed using unmodified and chemically modified mRNA Fluc sequence. The modified mRNA contained 5-methoxyuridine (replacing all uridines), and both samples were analyzed using the workflow previously described. For data analysis, 5-methoxyuridine was added to the modified sequence using the sequence editor in BioPharma Finder software to generate a new sequence where all uridines were replaced by 5-methoxyuridine. This was then used as the database in the sequence analysis for the modified mRNA. The LC-MS/MS analysis resulted in > 90% sequence coverage of the unmodified and modified Fluc mRNA from a single analysis, demonstrating the ability to generate high sequence coverage of chemically modified mRNA. Figure 5 shows the MS/MS spectra from both the unmodified and modified mRNA of the same corresponding oligoribonucleotide from the partial mRNA digests. The specific fragment ions corresponding to sites of the 5-methoxyuridine modifications are highlighted in red. The sequence analysis of the MS/MS spectra successfully identifies the modified uridines.

A database search performed on the chemically modified mRNA partial digest against the unmodified Fluc mRNA sequence only identified oligoribonucleotides that did not contain uridine.

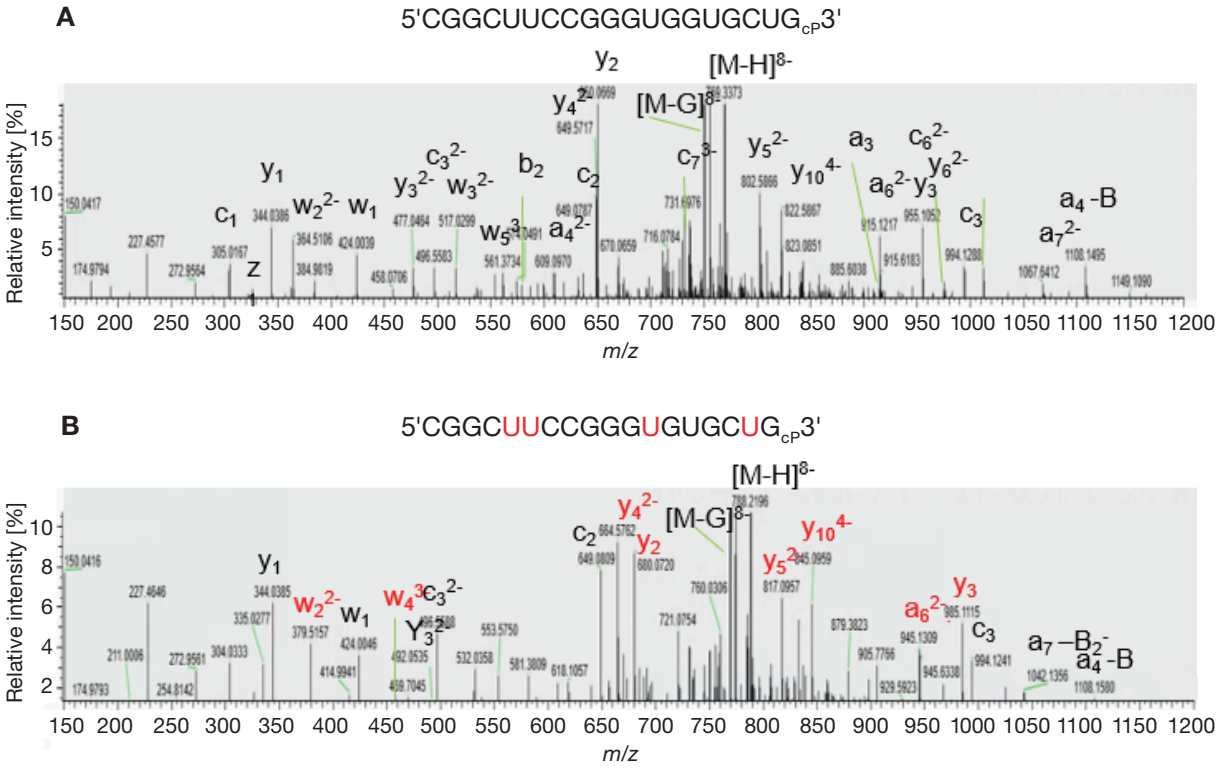


Figure 5. MS/MS spectra of the oligoribonucleotide CGGCUUCCGGGUGGUGCUG$_{cP}$ unmodified (A), and the corresponding oligoribonucleotide where the uridines are replaced with 5-methoxyuridines (B). The corresponding fragment ions are highlighted and those fragment ions specific for the 5-methoxyuridine oligoribonucleotide are highlighted in red.

Conclusion

A complete semi-automated direct mRNA sequencing workflow was developed. It enables the following:

- Reproducible and controlled partial mRNA digestion to generate sequence specific fragments using immobilized RNase T1 on magnetic beads.
- High-resolution separation of 10 to 50 nt oligonucleotide fragments using the DNApac RP column with a TEA/HFIP eluent system.
- Accurate sequence identification through automated annotation of high quality HRAM MS and MS/MS spectra in BioPharma Finder version 5.0 software.
- Sequence mapping of large modified and unmodified mRNA using BioPharma Finder version 5.0 software.
- Fast analysis of 90 minutes per sample.
- Further development opportunities for simple methods to go into QC testing of mRNA vaccines.

Learn more at **thermofisher.com**