Characterization of mRNA vaccine 5’ and 3’ end products using BioPharma Finder 5.0

Robert L Ross, Qian Qi, Keeley Murphy, Min Du, Thermo Fisher Scientific, Waltham, MA

Thermo Fisher SCIENTIFIC

ABSTRACT

We sought the development of a Thermo Scientific™ BioPharma Finder™ 5.0 software solution for automated detection and characterization of 5’ and 3’ ends of *in vitro* synthesized mRNA therapeutics. For this we utilized enzymatic digestion of intact mRNA substrates followed by HRAM-LC-MS/MS acquisition. For the 5’ cap, an RNase H digestion was performed with molecular probes, followed by LC-MS/MS acquisition and data processing with BioPharma Finder™ 5.0 using the Oligonucleotide Sequencing option. The 3’ poly(A) tail was characterized using ribonuclease T1 where data processing used the Intact Mass Analysis option. Utilizing BioPharma Finder™ 5.0 we were able to quickly detect and characterize both 5’ and 3’ digestion products of mRNA transcripts.

INTRODUCTION

Messenger RNA (mRNA) transcripts represent a new class of therapeutics. By co-opting the cells bio-machinery, mRNA therapeutics can produce the enzymes and proteins necessary to alleviate suffering. Production of mRNA therapeutics necessitates strict quality control at each step of production. Two critical parameters in the production of mRNA, whether *in vitro* or *in vivo* are the 5’ capping and 3’ poly adenylation. Capping and polyadenylation help with the stability of the biopolymer, assists with translation, and helps resist degradation by innate nucleases. Robust workflows are needed to monitor the efficiencies of 5’ and 3’ addition produced *in vitro*,. The use of nucleases for enzymatic digestion followed by LC-MS/MS has been the gold standard for direct measurement of RNA. Instrumentation development continues to improve, what has often lagged are data processing tools to quickly and accurately discriminate the necessary information. Here we report an IPRP-LC-HRAM MS/MS workflow with a dedicated software solution for automated data processing and annotation of 5’ and 3’ end products from mRNA.

MATERIALS AND METHODS

5’ Capping Workup: 500 pmol of a 25 nt long RNA biotinylated probe containing five chimeric DNA bases, was annealed to a complementary sequence at the 5’ end of a synthetic mRNA transcript (100 pmol).Following annealing, $MgCL_2$ was added to a concentration of 3 mM followed by the addition of 10U RNase H was added to the solution and allowed to digest. Digested duplex was removed from solution by complexing with streptavidin coated magnetic beads. Complexed beads were then washed then eluted with 75% EtOH at 90°C, taken to dryness in an oil free speedvac, then analyzed.

Chromatography: Separation was performed with a Thermo Scientific™ Vanquish™ Horizon™ UHPLC using a DNAPac™ RP column. For 5’ capping mobile phase A consisted of LC-MS grade water with an HFIP:TEA modifier. Mobile phase B consisted of LC-MS grade methanol also containing an HFIP:TEA modifier. For 3’ characterization mobile phase A consisted of LC-MS grade water with an HFIP:DBA modifier. Mobile phase B consisted of LC-MS grade ACN also containing an HFIP:DBA modifier.

Mass Spectrometry: Data was acquired using an Thermo Scientific™ Orbitrap Exploris™ 240 mass spectrometer detector operating in negative polarity, application mode: peptide over a 15-minute acquisition window. Full scan resolution was set to 120K over a scan range of 550-2500 (*m/z*); RF lens = 70% with a Normalized AGD target of 100%. Maximum injection time was set to 50 ms with one microscan. Data type was profile. Filters included Intensity (5e3) and Dynamic Exclusion, n=1 for 3 sec, Low and High ppm set to 10. Five $ddMS^2$ scans were performed with an isolation window of 1.6 (*m/z*), and an absolute collision energy of 25V. Orbitrap resolution was set to 60K over a scan range of 200-2500 (*m/z*) using 100% normalized AGC target with a maximum injection time of 200 (ms). One microscan was performed with data type set to profile.

3’ Poly(A) tail workup: Supernatant from the RNase H digestion, containing the cleaved mRNA transcript with poly(A) tail, was subjected to a one-hour RNase T1 digest. Digestion products were then isolated using Dynabeads™ Oligo$(dT)_{25}$, taken to dryness in a speedvac, resuspended in MPA and injected onto a Thermo Scientific™ Vanquish™ Horizon UHPLC system.

Mass Spectrometry: Data was acquired using an Thermo Scientific™ Orbitrap Exploris™ 240 mass spectrometer detector operating in negative polarity, application mode: Intact Protein over a 15-minute acquisition window. Full scan resolution was set to 240K over a scan range of 1000-1800 (*m/z*); RF lens = 70% with a Normalized AGD target of 300%. Maximum injection time was set to 200 ms with two microscan.

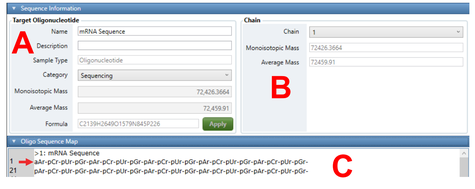

Figure 1. *Sequence Manager* panel. A) Creation of named sequences and selection of processing category. B) Monoisotopic and average mass calculation for each created sequence. C) Sequence map showing each created oligonucleotide.

Data Analysis 5’ Capping

Data analysis was performed by Thermo Scientific™ BioPharma Finder™ 5.0 software. Using the *Sequence Manager* our mRNA sequence is created and modified (**Figs 1-3**). Under Sequence Information, the oligonucleotide name is created and the category for processing is chosen. Here we chose the option “Sequencing”. Once all sequences are added the Sequence Information panel is populated with the monoisotopic and average mass, as well as the formula for each sequence.

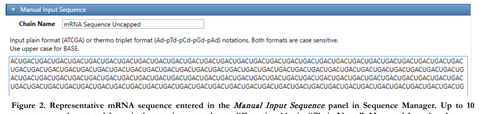

Figure 2. Representative mRNA sequence entered in the *Manual Input Sequence* panel in Sequence Manager. Up to 10 sequences can be created for a single experiment, each one differentiated by its **“Chain Name”**. Mass and formula values are populated in Target Oligonucleotide section, (Fig 1A).

Under *Manual Input Sequence* the chain name is given, and the sequence is copied (**Fig 2**). Up to 10 different sequences can be included. Once all sequences are added the Sequence Information panel is populated with the monoisotopic and average mass, as well as the formula for each sequence.

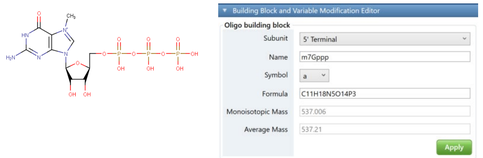

Figure 3. *Building Block and Variable Modification Editor.* Used to create the 5’ cap structure and added as a molecular formula which can be generated through a chemical drawing, as shown on left. For a 5’ capping characterization the modification is added to the **“5’ Terminal”** subunit.

After the sequence has been created a 5’ cap structure is created under the “Building Block and Variable Modification Editor”. From the Subunit dropdown 5’ terminal is chosen, named and its formula is added **(Fig 3).** This automatically populates the Monoisotopic and Average mass boxes. Modifications for 5’, 3’, nucleobase and linker are all created under this panel. After creation of the 5’ terminal modification the *Edit Sequence* panel is accessed **(Fig 4).** It is from this panel that any nucleotide in the sequence can be modified, using either the default modifications in the dropdown menus, or the user created modifications generated in the *Building Block and Variable Modification Editor*. Any change to the sequence in the edit sequence panel is reflected in the *Oligo Sequence Map*. For example, addition of the m7G cap structure generated in the *Building Block* editor to the 5’ terminal of our mRNA sequence is reflected as an addition of a lowercase a to the 5’ end of the mRNA sequence (**Fig 1, red arrow**).The lowercase a is a user defined symbol chosen when the modification is created.

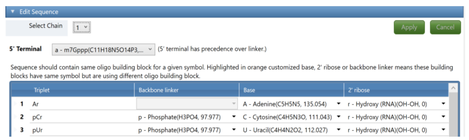

Figure 4. *Edit Sequence* panel. Each individual nucleotide in the individual sequences can be customized. For the 5’ capping, the 5’ terminal dropdown menu is accessed and the cap structure that we created under building blocks is added to the sequence. Addition of the 5’ terminal cap, once applied, is then added as a separate character in the oligo sequence map (red arrow, Fig 1).

Once the sequence(s) are created and saved, the sequence manager is closed, and the Oligonucleotide Analysis option is chosen from the home screen. The processing analysis is named, and the data file is chosen. Next the oligonucleotide sequence that was just created is chosen along with Basic Default Method. Depending on chromatographic and digestion conditions, parameters such as signal to noise, peak width, etc are defined, as well as nuclease. For this method, a “Nonspecific” nuclease is chosen with no 3’ phosphate. The method is then saved under a new name and the data processed.

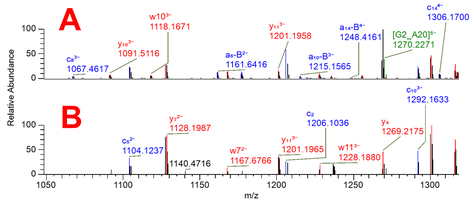

Figure 5. Tandem mass spectra of an RNase H digestion product. Panel A is the theoretical fragmentation pattern. Panel B is the actual fragmentation. Color coded annotated MSMS spectrum is automatically generated for each oligonucleotide identification. Matched spectra lends confidence to the identification

MSMS fragment identification is a critical component of LCMS mRNA mapping. In addition to identification by full scan analysis and the mass deconvolution of multiple charge states; fragmentation identification provides the confirmation of sequence order as well as information pertaining to potential modifications and the modification site within the sequence. MSMS fragmentation identification and annotation was automatically generated, and color coded by the data analysis software. In addition to fragment annotation, the observed fragmentation spectrum is automatically compared to a predicted fragmentation model generated for each identified sequence (**Fig 5**).

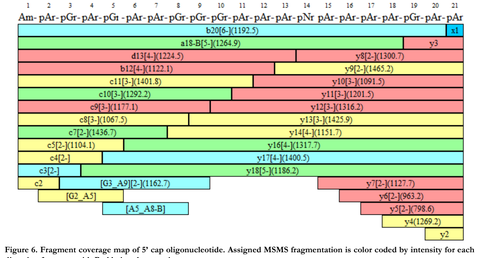

Figure 6. Fragment coverage map of 5’ cap oligonucleotide. Assigned MSMS fragmentation is color coded by intensity for each digestion fragment with Red being the most intense.

MSMS fragmentation identification was also evaluated using the fragment coverage map tool, which provides a visual representation of all assigned MSMS fragmentation and is color coded by intensity for each digestion fragment (**Fig 6**).

Characterization of the mRNA 5’ cap is accomplished through annealing a biotinylated molecular probe complementary to the 5’ end of the mRNA. RNase H will cleave the RNA strand in a DNA:RNA duplex, so DNA nucleotides (minimum of 5) are added to the biotinylated probe. When annealed to the mRNA a DNA:RNA duplex is created resulting in a cleavage motif for RNase H. **Figure 7** shows the mapping result of a 5’ cap RNase H digestion.

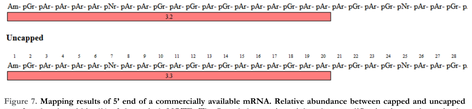

Figure 7. Mapping results of 5’ end of a commercially available mRNA. Relative abundance between capped and uncapped was found to be within 4% of theoretical. NOTE: The 7-methylguanosine triphosphate modification is not shown in the mapping results.

Relative Quantitation of 5’ cap

The Orbitrap Exploris™ 240 mass spectrometer’s sensitivity and scan speed make it an ideal choice for quantification of biomolecules. For the RNase H assay, a 4:1 molar mixture of capped to uncapped mRNA was subjected to the RNase H digestion then analyzed. **Figure 8** shows extracted ion chromatograms for both capped and uncapped oligonucleotides from the RNase H digest. Co-elution was observed for the two products, and not unexpected given the difference between the two digestion products is a single GDP. Extracted ion chromatograms were generated using a mass range from the top 5 isotopic peaks of the most abundant charge state (8th and 7th, respectively) in both capped (893.4994 – 893.9995 Da, **Fig 8B**) and uncapped (958.5666 – 959.1395 Da, **Fig 8C**) at a mass tolerance of 5 ppm. Using the same strategy, the top three charge state areas from both capped and uncapped were averaged and compared. **Comparison of capped to uncapped peak area returned a ratio of 24%, within 4% variance from theoretical.**

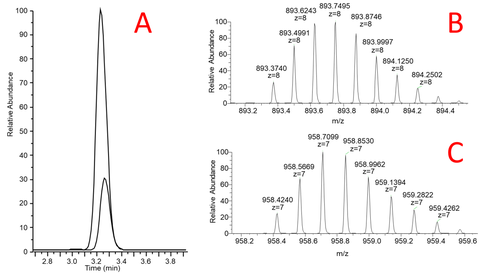

Figure 8 LC-MS data from RNase H digestion. (A) Overlay of extracted ion chromatograms for both capped and uncapped oligonucleotides. (B) Isotopic peaks for the most abundant charge state(8^{+}) in the capped oligonucleotide detected in the sample.(C) Isotopic peak envelope for the most abundant charge state(7^{+0}) in the uncapped digestion products detected in the sample. The top five peaks were averages to generate the extracted ion chromatograms used to measure abundance. Chromatogram and spectra were generated in Freestyle 1.8.2

Data Analysis 3’ poly(A) Characterization

Data analysis was performed by Thermo Scientific™ Biopharma Finder™ 5.0 software using the Xtract™ Deconvolution Algorithm. In *Sequence Manager*, a poly(A) oligonucleotide was created under the Deconvolution Category. The average length of the poly(A) tail in the sample falls between 120 and 130 adenosines. Using the workflow described under 5’ capping data analysis, we created in *Sequence Manager* a poly(A) sequence with a length of 140 adenosines. BFP 5.0 allows for detection of truncated species in the sample. Using this feature, we can assign length to the deconvoluted masses based on their mass, relative to the 140-mer.

Intact Mass Analysis option. The Default Oligo Xtract™ Processing Method was chosen for the deconvolution. Source Spectra Method was set to *Average Over Selected Retention Time* as our chromatographic trace showed a poly(A) elution of between 10 and 10.3 minutes. Other detection parameters were output mass range 35-47K, S/N of 3, Rel. abundance Threshold % of 4, a Charge Range of 5-50.

Table 1. Detailed list of deconvoluted masses identified in the sample. Note: not all table columns are shown.

Modification	Monoisotopic Mass	Theoretical Mass (Da)	Matched Mass Error (ppm)	Relative Abundance	Number of Charge States	Charge State Distribution	Average Charge
A+	m	m	m	%	m	m	m
18x3Truncation,1xA	40411.683	40411.505	4.4	100.00	19	21 - 39	28.53
19x3Truncation,1xA	40082.625	40082.452	4.3	92.04	19	21 - 39	28.53
17x3Truncation,1xA	40740.734	40740.557	4.4	89.82	20	21 - 40	28.77
16x3Truncation,1xA	41069.788	41069.610	4.3	79.84	19	22 - 40	29.18
20x3Truncation,1xA	39753.571	39753.400	4.3	73.54	19	20 - 38	27.90
15x3Truncation,1xA	41398.835	41398.662	3.4	72.02	20	21 - 40	29.13
14x3Truncation,1xA	41727.880	41727.715	4.0	54.11	19	22 - 40	29.81
13x3Truncation,1xA	42056.947	42056.767	4.3	38.40	18	23 - 41	30.56
21x3Truncation,1xA	39424.503	39424.347	4.0	35.57	15	22 - 38	29.08
12x3Truncation,1xA	42385.993	42385.820	4.1	22.24	12	24 - 40	29.52
22x3Truncation,1xA	39095.472	39095.295	4.5	18.94	14	24 - 38	29.71
11x3Truncation,1xA	42715.073	42714.872	4.7	17.43	12	27 - 41	31.98
10x3Truncation,1xA	43044.181	43043.925	6.0	10.16	10	30 - 41	34.47
12x3Truncation,1xA	42386.093	42385.820	6.4	6.79	5	34 - 39	36.04
9x3Truncation,1xA	43373.222	43372.977	5.6	5.00	8	34 - 42	37.96

Table 1 lists the detected features of the deconvoluted spectrum. Modification refers to the theoretical sequence (140-mer) matched with the deconvolution result ± Delta Mass and listed as a truncation. Each of the matched modifications can be referenced to its Relative Abundance. This information shows the length and intensity distribution in the poly(A) sample. The Number of Charge States column can be used as a reference value for reliability of the results. Matches of more than a dozen charge states means a high confidence in the deconvolution results.

Deconvoluted peaks are annotated based off of the difference in the deconvoluted mass and a theoretical mass of 46067 Da, the calculated mass of 140 x 330.063. Each loss of an adenosine from the theoretical mass will be a loss of 330.063 Da (AMP – OH), so using the truncation feature in BioPharma Finder we are able to annotate the deconvoluted masses as a truncated 140 mer. The annotated deconvoluted masses for a T1 digest of mRNA are shown in **Figure 8**.

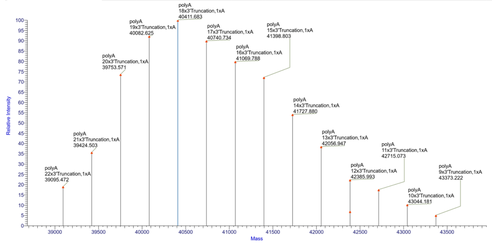

Figure 9. Deconvolution result of poly(A) sample with BioPharma Finder™ 5.0 software. Identification is a measure of the loss of the mass of AMP from a theoretical 140-mer and listed as a truncation

Figure 10 shows a graph of the deconvoluted monoisotopic mass, listed in **Table 1**, divided by the mass of AMP, versus relative abundance. From the mRNA template the length of the theoretical poly(A) tail is known. Graphing the output of the Intact Analysis shows a clustering of lengths between 126 and 121 adenosines, results that agree with the theoretical length.

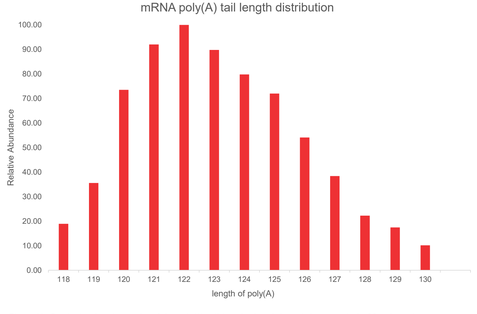

Figure 10. Graph showing the results of the poly(A) tail lengths found through intact analysis. Monoisotopic masses are divided by the mass of AMP - OH (330.063 Da). This is then plotted against the relative abundance to show length and abundance within the sample.

CONCLUSIONS

BioPharma Finder 5.0 software is a powerful tool in the characterization of mRNA vaccines. We show its utility for the characterization of both the 5’ and 3’ ends of *in vitro* produced mRNA and give customers confidence in characterization of their therapeutic products.

ACKNOWLEDGEMENTS

The authors would like to thank all our colleagues at Thermo Fisher Scientific for critical feedback and discussions. We especially want to thank Debbie Fraleigh for her patience in producing this work

TRADEMARKS/LICENSING

© 2022 Thermo Fisher Scientific Inc. All rights reserved. All trademarks are the property of Thermo Fisher Scientific and its subsidiaries. This information is not intended to encourage use of these products in any manner that might infringe the intellectual property rights of others.

PO66161-EN0422S