

Quantitative and Comparative Analysis of mRNA Poly(A) Tail Length Distributions Using LC–MS, LC–UV, and Oxford Nanopore Sequencing with Bayesian Harmonization

Melda Zeynep Guray

Lonza Netherlands B.V

Amanda Hughes

Lonza Netherlands B.V

Elisa Calvo-Jiménez

Lonza Netherlands B.V

Meltem Kaya

Lonza Netherlands B.V

Gaston Vondenhoff

`Gaston.vondenhoff@lonza.com`

Lonza Netherlands B.V

Article

Keywords: Poly(A) tail analysis, mRNA, analytical methods, Oxford Nanopore Technology, Mass spectrometry coupled liquid chromatography

Posted Date: August 27th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10680176/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Quantitative and Comparative Analysis of mRNA Poly(A) Tail Length Distributions Using LC–MS, LC–UV, and Oxford Nanopore Sequencing with Bayesian Harmonization

Melda Zeynep Guray, Amanda Hughes, Elisa Calvo-Jiménez, Meltem Kaya, Gaston Vondenhoff*

Analytical Development Services, Lonza Netherlands B.V., Urmonderbaan 20b, 6167 RD Geleen, The Netherlands. *Corresponding author: Gaston.vondenhoff@lonza.com

Abstract

The polyadenylated [poly(A)] tail of messenger RNA (mRNA) is a critical quality attribute governing stability and translational efficiency in mRNA therapeutics. However, analytical methodologies for determining poly(A) tail length distributions—including liquid chromatography–mass spectrometry (LC–MS), liquid chromatography–ultraviolet detection (LC–UV), and Oxford Nanopore Technologies (ONT) sequencing—produce fundamentally different data formats, complicating cross-method comparison and standardized reporting.

Here, we present a harmonized framework for quantitative and qualitative comparison of poly(A) tail distributions across these analytical platforms. Three mRNA constructs (eGFP, fLuc, and hRLuc) with distinct encoded poly(A) lengths were analyzed. Key descriptive parameters—including most abundant length (V_{ab}), and statistically defined minimal (V_{min}) and maximal (V_{max}) lengths—were derived. To enable cross-platform comparability, a Naïve Bayes classification model trained on LC–MS data was applied to LC–UV and ONT datasets.

Despite inherent methodological differences, all techniques consistently identified the most abundant poly(A) species. The Bayesian harmonization approach significantly improved agreement in distribution boundaries across platforms. This study proposes a standardized analytical and reporting approach that enhances comparability and supports regulatory alignment for mRNA characterization.

Keywords

Poly(A) tail analysis, mRNA, analytical methods, Oxford Nanopore Technology, Mass spectrometry coupled liquid chromatography

1. Introduction

Messenger RNA (mRNA)-based therapeutics have rapidly advanced following the success of COVID-19 vaccines, leading to a growing pipeline of RNA-based medicines. A defining structural feature of most linear mRNA molecules is the 3′ poly(A) tail, typically consisting of 80–120 adenosine residues, which plays a central role in mRNA stability^{1, 2} and translational efficiency³.

In vitro transcription (IVT)-derived mRNA typically exhibits a distribution of poly(A) tail lengths rather than a single discrete value. This heterogeneity arises primarily from enzymatic slippage during transcription or variability introduced during post-transcriptional polyadenylation. The polyadenylated [poly(A)] tail is a hallmark structural feature of eukaryotic messenger RNA (mRNA) and plays a critical role in regulating transcript stability, translational efficiency, and cellular localization⁴⁻⁷. In therapeutic mRNA products, the length and heterogeneity of the poly(A) tail directly influence biological performance, making it a key critical quality attribute (CQA) that requires rigorous analytical characterization⁶⁻⁸. Consequently, accurate assessment of not only the average length but also the shape and spread of the poly(A) tail distribution is essential for a comprehensive understanding of product quality.

A variety of analytical techniques have been developed to measure poly(A) tail length distributions, including liquid chromatography–mass spectrometry (LC–MS), liquid chromatography with ultraviolet detection (LC–UV)^{7, 9}, capillary gel electrophoresis (CGE)¹⁰⁻¹², and single-molecule sequencing technologies such as Oxford Nanopore Technologies (ONT)^{13, 14}. While these methods provide valuable and often complementary information, they differ substantially in detection principles, sensitivity, resolution, and data output formats. As a result, comparison of poly(A) tail distributions across platforms remains challenging and often lacks standardization.

To address this challenge, there is a growing need for harmonized approaches that enable consistent interpretation and reporting of poly(A) tail distribution data. Such approaches should account for method-specific biases while preserving the underlying biological information. In this work, we evaluate representative orthogonal methods and explore statistical frameworks that support cross-platform comparability, with the aim of establishing a robust and reproducible strategy for poly(A) tail characterization in mRNA therapeutics.

2. Materials and Methods

2.1 Materials

The three mRNA constructs, fLuc, hRLuc and eGFP, with different poly(A) tail lengths were in-house produced at Lonza Netherlands. Synthetic poly(A) standards for LC-UV method were synthesized by Eurofins Genomics (Ebersberg, Germany).

Tip comb for deep-well magnets (97002534B) and deep-well plates (95040450B) for KingFisher Apex, DNAPac RP column (088923) and Smart Digest Magnetic RNase A kit (60120-102) were purchased from Thermo Fisher Scientific. LC-MS grade acetonitrile (100029) and hexafluoro-2-propanol (HFIP, 18127) were

from Merck and dibutylamine (DBA, 471232) was from Sigma-Aldrich. Nuclease-free water was purchased from Invitrogen (AM9937) and the water for mobile phase preparation was produced using Milli-Q system.

RNA sequencing kit (SQK-RNA004) was from Oxford Nanopore Technologies. T4 DNA Ligase (M0202M) and Induro Reverse Transcriptase (M0681S) were purchased from New England Biolabs. The magnetic beads used in ONT sample preparation workflow was from Beckman Coulter (RNAClean XP, A63987).

2.2 LC–MS Measurement

Poly(A) tails are cleaved from mRNA using site-specific cleavage enzyme RNase A. This enzyme cleaves mRNA after every cytosine and uracil residues thereby, the poly(A) remains intact, but the rest of the mRNA is digested into smaller fragments. For that, 40 µg of mRNA was mixed with nuclease-free water and 20 µL of Smart Digest Buffer in a total volume of 50 µL and then digested with 2.5 µL of RNase A magnetic beads for 1 hour at 37°C. The digestion was performed using KingFisher Apex system (Thermo Fisher Scientific, MA, USA) with continuous mixing. The resulting mRNA digest is then mixed with equal volume of nuclease-free water and analyzed via IP-RP chromatography with mass detection. Analysis of the samples were conducted via Vanquish Horizon UHPLC (Thermo Fisher Scientific, MA, USA) and Orbitrap Exploris 120 Mass Spectrometer (Thermo Fisher Scientific, MA, USA) instruments. 10 µL of each sample was injected onto the DNAPac RP column (2.1 x 100 mm) in triplicate. Mobile phase A was 15 mM DBA and 25 mM HFIP in water and mobile phase B was 15 mM DBA and 25 mM HFIP in acetonitrile. The separation of the poly(A) tail was achieved at a flow rate of 0.3 mL/min and at 60°C column temperature with a gradient of 2.5-23%B in 7 min, followed by a gradient of 23-24.5%B in 1 min and 24.5-30%B in 4 min. The column was then washed with 90%B for 2 min and equilibrated to 2.5%B for 4 min. All MS data were recorded in negative ion mode at 2500 V capillary voltage, with a scan range of 570-2500 *m/z* and at a resolution of 15,000. Ion transfer and vaporizer temperatures were 320°C. All the data from UV and MS were acquired using Chromeleon CDS Software (v7.3.2).

2.3 ONT Sequencing

Sequencing libraries were prepared from 300 ng mRNA drug substance following manufacturer's standard protocol for the direct RNA sequencing kit from Oxford Nanopore Technologies.

Briefly, the reverse transcription adapter (RTA) was ligated to the 3' poly(A) tail using concentrated T4 DNA Ligase. Reverse transcription with Induro Reverse Transcriptase was initiated from the RTA. Following purification with SPRI beads, the sequencing adapter (RLA) was ligated to the RNA:cDNA, and the sample was purified with SPRI beads and loaded on an FLO-MIN004RA flow cell. Sequencing was performed with the wf-vax-qc workflow on the GridION Q-Line MK1 (Oxford Nanopore Technologies, Oxford, UK) device. This workflow generates automated reports, including the mean, median, and mode poly(A) tail length as well as a csv file of per-read poly(A) tail lengths.

2.4 LC–UV Measurement

The sample preparation workflow and chromatographic conditions for LC-UV analysis were identical to those used for LC-MS, as described in Section 2.2. Synthetic poly(A) standards of 20, 40, 60, 80 and 100 nucleotides were used as calibrants and 1.5 µL of each standard at a concentration of 100 pmol/µL was injected onto the DNAPac RP column. Following separation, the UV data was collected at 260 nm wavelength. Each standard was injected three times to determine the mean retention time.

3. Data Analysis Strategy

3.1 Definition of Distribution Parameters

Three key descriptors were defined:

- **V_{ab}**: Most abundant poly(A) length
- **V_{min}** and **V_{max}**: Statistically relevant minimum and maximum poly(A) length
- **Skewness factor (SF)**: Describing distribution symmetry as calculated by Equation 1.

Equation 1.
$$SF = \frac{V_{max} - V_{ab}}{V_{ab} - V_{min}}$$

3.2 LC-MS Data Analysis

The data obtained from MS is processed using BioPharma Finder Software (v5.2) ReSpect deconvolution algorithm to determine the intact mass, as well as the length of each poly(A) tail species. A poly(A) sequence approximately 20 nucleotides longer than the expected poly(A) tail of each construct was generated, and the poly(A) peak from each sample was aligned to this sequence to determine the number of adenosines, hence the poly(A) tail length. The length of each identified tail is plotted against relative abundance of MS response to give the distribution of tail lengths in the sample.

To determine V_{min} and V_{max} reportable lengths, we used JMP statistical software (v19.0) to develop a script that allows for determination of i) the most abundant poly(A) species (V_{ab}), ii) relevant minimal (V_{min}) and maximal (V_{max}) poly(A) lengths using a simple statistical approach. To determine the relevant cut-off from which minimal and maximal lengths should be reported, we made use of the lowest detectable signals. To that end, the mean of the 5% quantile of the lowest signals was determined and a cutoff was calculated using Equation 2.

Equation 2.
$$\text{Cutoff} = (\text{mean 5\% quantile lowest signals}) + 1.653 * \text{SD}$$
, whereby SD is calculated from the 5% quantile data.

Reportable values could next be deduced from this obtained cutoff value by determining the minimal and maximal lengths exhibiting a higher signal than the cutoff value.

3.3 ONT Data Analysis

During sequencing, mRNAs transit the nanopores from 3' to 5' and the resultant electrical current is recorded. The poly(A) tail is evident within this signal as a long homopolymeric stretch after the DNA adapter signal. ONT's basecaller, Dorado (v1.0.2+c758d2f6), estimates the poly(A) tail length from this signal for each sequenced molecule. The wf-vax-qc analysis workflow, developed by ONT, outputs csv files with the poly(A) tail length for each read and plots the distribution of the poly(A) tail lengths in an HTML report. To compare poly(A) tail distributions reported by ONT with UV and MS methods, the per-read poly(A) tail lengths in the csv file were converted to relative abundance of poly(A) tail length, using the relfreq function of the scipy.stats python module.

3.4 LC-UV Data Analysis

This approach relies on synthetically produced poly(A) standards that are used to determine the poly(A) tail length of mRNA samples based on the retention times corresponding to the length of the respective standards. Correlating length with retention time (RT) of these standards, linear fit was applied using JMP (19.0.0) software. From Figure 1, we see that the poly(A) standard with 20 nt falls out of the linear range. Indeed, other (polynomial) calibration models could be applied, however these were found to bias the calculated sizes at the higher end. As most poly(A) tails vary in intended length between 80-150 nt, it was decided to exclude the 20 nt standard and use a linear fit with a range from 40 to 100 nt. This then yielded a calibration curve (R^2 of 0.98) following Equation 3.

Equation 3. $\text{Size} = -277.6 + 36.38 \cdot \text{RT}$

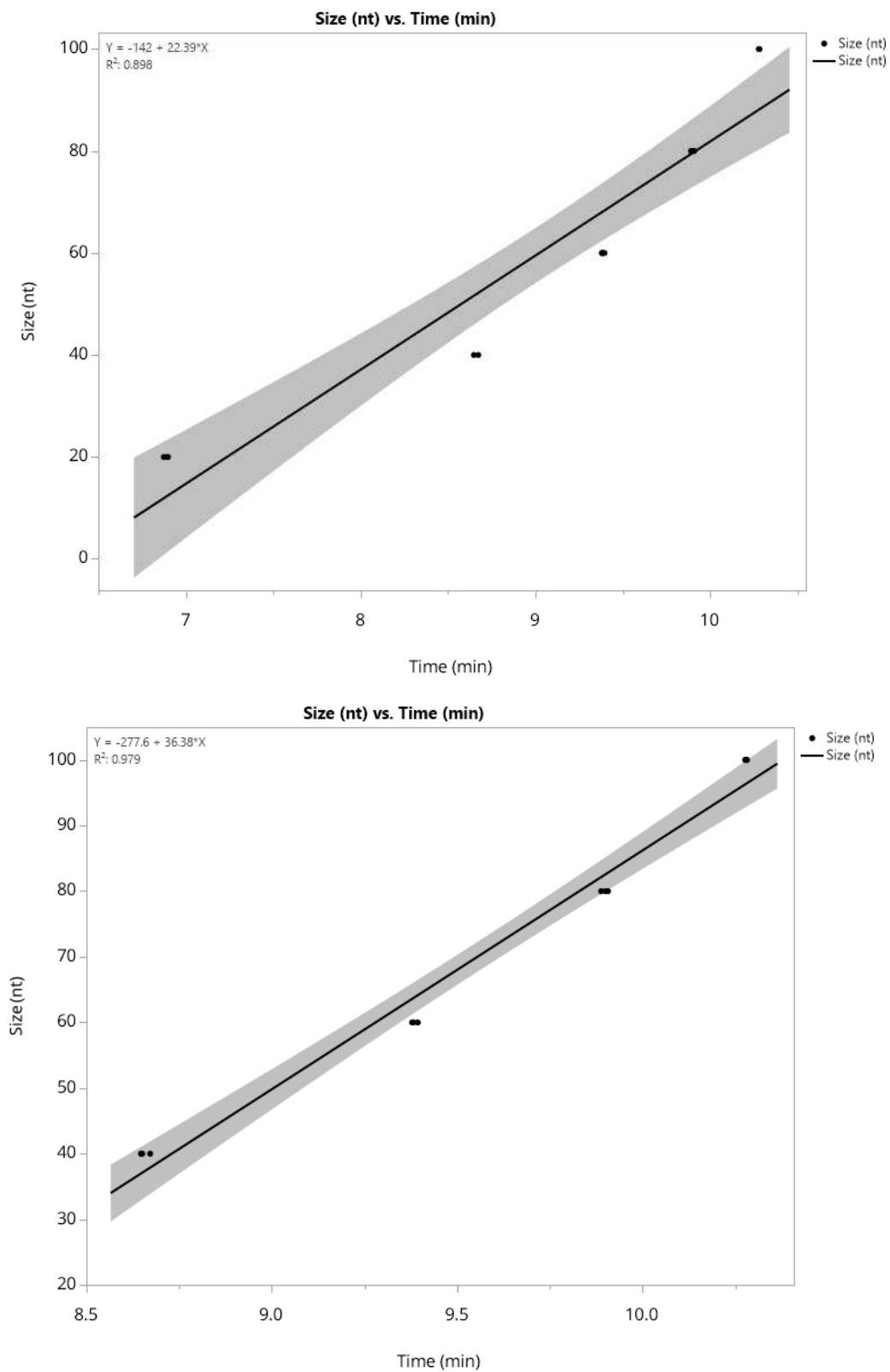


Figure 1. Calibration curve generated by poly(A) standards via UHPLC-UV method. Upper panel with 5 standards between 20-100 nt. Lower panel with 4 standards between 40-100 nt.

From Table 1 we see that back calculated accuracy of the calibrants lies between 93-106% and that analytical precision is ≤ 1.6 % CV.

Table 1. Evaluation of accuracy and precision of back calculated results.

Size (nt)	Mean Backcalculated Size (nt)	Mean Accuracy (%)	Precision (%CV)
40	37	93	1.6
60	64	106	0.5
80	83	103	0.6
100	96	96	0

4. Results and Discussion

4.1 Identification of the Most abundant Poly(A) Length

In terms of identification of the most abundant poly(A) species we find that all three technologies are able to, with some minor relative differences, identify the most abundant poly(A) species for all three tested mRNA molecules, see *Table 4*.

4.1 Limitations of Direct Comparison of all Three Approaches in Describing Poly(A) Tail Distribution

While for LC-MS data, each datapoint represents the relative abundance of an identified poly(A) species, for LC-UV data identification is solely based on UV-absorbance and retention time. Therefore, confidence in the presence of very small or very large poly(A) species is low with both methods. This is different for ONT data, where molecules are individually sequenced through the pore and their poly(A) tail lengths measured. The sequencing quality of each molecule is determined from the mean of the quality of all its nucleotides. Only molecules that pass a mean quality threshold are used in subsequent analyses.

In Figure 2 normalized abundancies for each of the identified poly(A) species as measured by the three respective methods is plotted against the poly(A) length for the fLuc construct. From this we see that for both LC methods display a comparatively narrow distribution that is slightly right skewed. ONT seems to identify a much broader distribution of poly(A) species.

Direct application of Equation 2 on datasets coming off these three methods yields very different $V_{\max}$ and $V_{\min}$ cutoff values. For example, ONT and LC-UV would obviously show a very large distribution, almost indifferent from all measured species, see Table 2. Hence, direct comparison of length distribution of LC-MS with ONT or LC-UV data is not possible or very arbitrary.

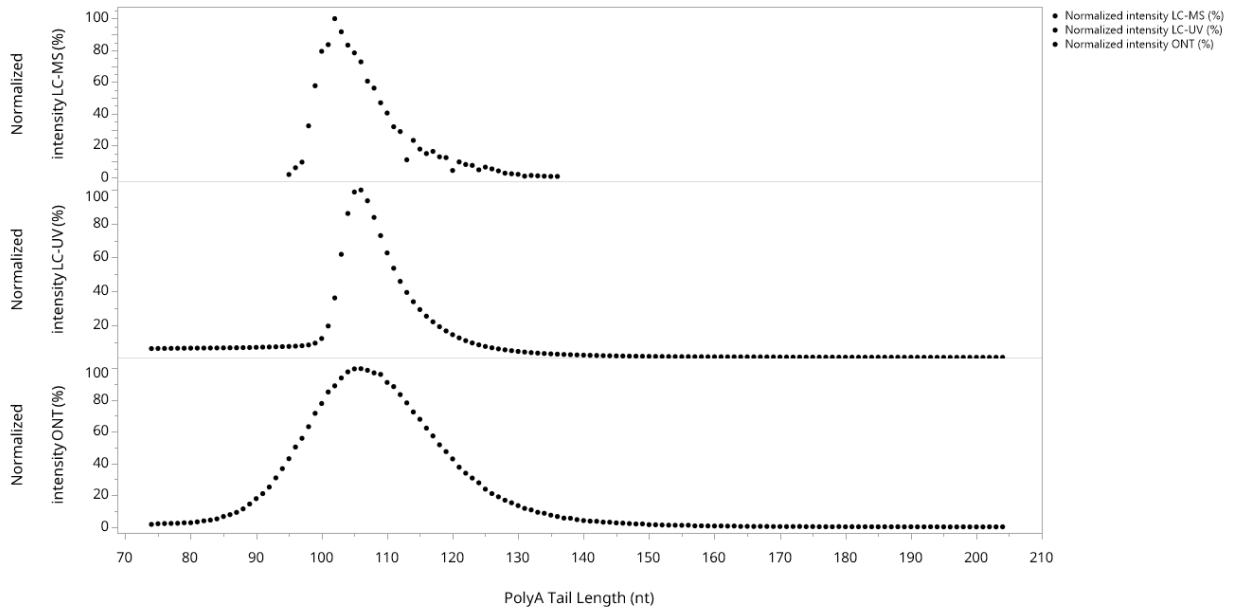


Figure 2. Visual comparison of poly(A) tail distributions as measured by LC-MS (upper panel), LC-UV (middle panel), and ONT (lower panel) for fLuc construct.

Table 2. Exemplary data analysis applying Equation 2 on all three methods.

Construct	Plasmid Encoded Poly(A) Tail Length	Observed Poly(A) Tail Length			
			LC-MS	ONT	LC-UV
fLuc	99	$V_{\max}$	130	194	206
		$V_{\min}$	94	7	15
		V_{ab}	102	106	106
eGFP	120	$V_{\max}$	146	352	206
		$V_{\min}$	109	7	15
		V_{ab}	124	123	124
hRLuc	80	$V_{\max}$	119	270	206
		$V_{\min}$	65	7	15
		V_{ab}	82	85	87

4.2 Naïve Bayes Harmonization

To enable cross-method comparison, a Naïve Bayes classifier was implemented using LC-MS data as a training dataset. For both the ONT and LC-UV dataset, poly(A) lengths were classified as reportable (R) or non-reportable (NR) based on posterior probabilities derived from the respective method-specific signal distributions. Prior probabilities ($P(R)$ and $P(NR)$) were calculated using LC-MS identified poly(A) species. In case of ONT datasets, only poly(A) species were considered in the calculation where an uninterrupted

stretch of poly(A) species was found. For instance, in case of a discontinuity in the retrieved lengths, the next highest or lowest value would be considered for the Priors calculation. Consecutive poly(A) distributions and priors calculations are summarized in Table 3.

Table 3. Prior probabilities, based on LC-MS training data.

Construct	Consecutive Stretch in ONT	Prior P(R) - LC-MS Based	Prior P(NR) - LC-MS Based
eGFP	7- 307 nt	38/301 = 0.12624	1-0.12624 = 0.87375
fLuc	7-239 nt	38/233 = 0.16309	1-0.16309 = 0.83691
hRLuc	6- 263 nt	54/258 = 0.20930	1-0.20930 = 0.79070

Using Bayes' rule: $P(R | x) \propto P(R)P(x | R)$, with x being the intensity, or response, of a poly(A) species as measured by the respective method. The Gaussian (Normal) distribution is calculated using

Equation 4: $P(x | R) = N(x | \mu_R, \sigma_R^2)$, with

Equation 5: $P(R) = N(x | \mu_R, \sigma_R^2)$, and

Equation 6: $P(NR) = N(x | \mu_{NR}, \sigma_{NR}^2)$ as the unnormalized posterior probabilities.

Calculation of the Gaussian likelihood for each measured intensity value:

$$\text{Equation 7. } N(x | \mu_R, \sigma_R^2) = \frac{1}{\sqrt{2\pi\sigma_R^2}} \exp \frac{-(x-\mu_R)^2}{2\sigma_R^2}$$

$$\text{Equation 8. } N(x | \mu_{NR}, \sigma_{NR}^2) = \frac{1}{\sqrt{2\pi\sigma_{NR}^2}} \exp \frac{-(x-\mu_{NR})^2}{2\sigma_{NR}^2}$$

The normalized posterior probabilities are then calculated as

$$\text{Equation 9. } P(x | R) = \frac{P(R)N(x | \mu_R, \sigma_R^2)}{P(R)N(x | \mu_R, \sigma_R^2) + P(NR)N(x | \mu_{NR}, \sigma_{NR}^2)}, \text{ and}$$

$$\text{Equation 10. } P(x | NR) = 1 - P(x | R)$$

Subsequently, poly(A) lengths were next classified using *Equation 11*.

$$\text{Equation 11. If } \frac{P(x | R)}{P(x | R) + P(x | NR)} > 0.5, \text{ then classify as "R", else as "NR"}$$

Using JMP distribution fitting, it is found that the poly(A) tail lengths, and consequently the Gaussian likelihoods as well, are not normally distributed. Following JMP fitting functionality, a Sinh-Arcsinh (SHASH) distribution describes the likelihoods better. Therefore, the Gaussian likelihoods were SHASH transformed and the lengths classified anew. From Table 4, it is observed however that this transformation does not greatly change the description of the distribution. Therefore, a more general Gaussian distribution may be used to align ONT with LC-MS data.

Using Naïve Bayes for ONT and LC-UV data, with LC-MS data as the training dataset, we find that the distributions are in good agreement. Also, as an indication, the SF for the classified data is similar between all methods per construct with exception of the fLuc construct measured by LC-MS. Not surprisingly, we find a much higher SF when compared to the other two methods, indicative of more response of larger species.

Although size calculation for the eGFP construct with a poly(A) tail of ~120 nt is outside of the calibrated range in the LC-UV method, the determined size of 124 nt is the same as determined by LC-MS and ONT. This suggests that even with only four calibration points that do not fully cover the range of the assess poly(A) tail variants determined, a good approximation of $V_{\max}$, $V_{\min}$ and V_{ab} may be obtained.

The main drawback for ONT and LC-UV data is that without prior knowledge (e.g. via an LC-MS training data set) it is difficult to set a consistent threshold to define minimal and maximal lengths that define how the distribution is reported.

Table 4. Combined results poly(A) tail distribution for all three constructs evaluated by three methods.

Construct	Method	LC-MS		ONT						LC-UV	
	Length	All identified	SF	All identified	SF	Naïve Bayes classified	SF	SHASH transformed	SF	Naïve Bayes classified	SF
fLuc	$V_{\max}$	130	3.50	580	4.79	128	1.38	126	1.31	138	1.52
	$V_{\min}$	94		7		90		91		85	
	V_{ab}	102		106		106		106		106	
eGFP	$V_{\max}$	146	1.47	593	4.05	144	1.11	145	1.11	144	2.00
	$V_{\min}$	109		7		104		103		114	
	V_{ab}	124		123		123		123		124	
hRLuc	$V_{\max}$	119	2.18	746	8.37	123	1.65	112	1.59	115	1.53
	$V_{\min}$	65		6		62		68		67	
	V_{ab}	82		85		85		85		86	

In determining which method is most suitable for its intended purpose, several factors need to be considered. Applicable to any laboratory, including the requirements of quality control laboratories, is of course the availability of equipment. Starting with the most affordable solution is naturally an LC-UV. This rather standard equipment is used for many attributes from small to large molecule analysis and may be adapted to each need. In terms of consumables, reagents and maintenance, prices are low. It's relatively easy to operate the instrument, interpret data and perform validation. The drawback of this method however is the need for reference standards. As calibration relies on the retention time, any fluctuation gives a bias to the determined poly(A) tail length.

Indeed, LC-MS gives a second line of identification of the tail length, and, given the uniqueness of the poly(A) molecule in terms of m/z , identification of the species is less likely to be biased. In terms of consumables, an MS-based determination is similar to an LC-UV based approach. However, high-resolution instruments are expensive due to their high initial costs and maintenance expenses. Moreover, there is a greater dependence on expertise of the analyst for operation of the instrument and interpretation of the

data. The validation of the methods are also not as straightforward as LC-UV methods. Therefore, LC-MS methods are more suitable for characterization purposes rather than release.

In terms of sensitivity, accurate determination and its potential to build multi-attribute methods, ONT is unparalleled. We have shown accurate determination of the most abundant poly(A) species, which are fully orthogonal to results obtained via LC-MS and LC-UV. While instrument costs are similar to an LC and less maintenance is required, the consumables are relatively expensive. Also, complex data processing via onboard softwares (e.g. Dorado) is required, which in turn comes with the necessity of controlling these softwares in quality control laboratories.

Therefore, we recommend to use a LC-UV or LC-MS instrument for high-throughput measurements to support research projects or process development studies. When multiple attributes like RNA sequence, integrity and poly(A) tail need to be measured for a given product, ONT is found to be the most efficient and affordable technology. For routine release testing in quality control laboratories, LC-UV currently appears to be a suitable option due to its simplicity, throughput and ease of validation.

Table 5. Holistic comparison of LC-MS, LC-UV, and ONT for characterization of poly(A) tail length and distribution.

Method	Ease of Use	Material Costs	Hands-on Time Required (hrs)	Total Equipment Runtime (hrs)	Throughput	Accuracy and Precision	Specificity	Sensitivity	Multi-attribute with Other Assays	QC Suitability
LC-MS	Moderate	€	3 hrs/ 96 samples	0.3 hrs/ sample	Sample preparation: 96 samples/ plate 4 plates/day	High	High	Moderate	Yes, with modifications of current protocol ¹⁵	Moderate
LC-UV	Easy	€	3 hrs/ 96 samples	0.3 hrs/ sample	Sample preparation: 96 samples/ plate 4 plates/day	Limited	Low	High	No	High
ONT	Very easy	€€€	3.5 hrs/ 5 samples	3 hrs sequencing/5 samples	5 samples/device	High	High	Highest	Yes	Moderate

5. Conclusion

This study demonstrates that all three analytical platforms reliably identified the most abundant poly(A) species across constructs. LC–MS provided high specificity and confident species identification, while LC–UV offered accessibility and high-throughput advantages. ONT enabled single-molecule resolution with high sensitivity.

This study demonstrates that poly(A) tail length distributions obtained from fundamentally different analytical platforms can be harmonized through a unified statistical framework. The use of Naïve Bayes classification, trained on LC–MS data, enables consistent definition of reportable ranges across LC–UV and ONT datasets.

We propose adoption of standardized descriptors (V_{ab} , V_{min} , V_{max} , SF) and statistical classification approaches to improve comparability, reproducibility, and regulatory interpretation of poly(A) tail analytics.

This framework supports the development of robust analytical strategies for mRNA therapeutics and facilitates alignment across laboratories and technologies, which could also facilitate analytical method transfer from a non-GMP analytical laboratory to a QC environment. For example, orthogonally characterized constructs in non-GMP may serve as reference materials in Quality Control where they can be used to perform validation and release of clinical batches.

Each method presents trade-offs. While LC–UV is cost-effective and may provide high-throughput, it naturally exhibits lower specificity. Additionally, this method relies on calibration, which means that previously characterized (ideally certified) standards must be available. Nevertheless, we showed that accurate identifications of poly(A) species beyond the calibration range is possible. LC–MS offers high specificity and confidence. While consumable costs are low, instrumentation cost are high. Compared to ONT, LC-MS does not offer high sensitivity, allowing for identification of low-abundant species. ONT offers the highest sensitivity and therefore is most suitable to fully characterize the drug substance. However, consumable costs are higher. We therefore suggest to use LC–UV/MS for process development work where high throughput is needed, while for selection of final candidates and to establish a process in advance of GMP work and subsequent release, ONT would be the better alternative.

Data availability

The datasets analyzed during the current study are available in the Zenodo repository:
<https://zenodo.org/records/21979942>

Author Contributions

All authors contributed to study design, data acquisition, analysis, and manuscript preparation.

Declaration of Competing Interest

This work was funded by Lonza and partially supported by Oxford Nanopore Technologies.

Acknowledgements

The authors thank colleagues from Lonza Process Development for providing mRNA constructs.

References

- (1) Wang, Z.; Kiledjian, M. Functional Link between the Mammalian Exosome and mRNA Decapping. *Cell* 2001, 107, 751–762.
- (2) Moqtaderi, Z.; Geisberg, J. V.; Struhl, K. Secondary structures involving the poly(A) tail and other 3' sequences are major determinants of mRNA isoform stability in yeast. *Microb. Cell* 2014, 1 (4), 137–139.
- (3) Xiang, K.; Bartel, D. P. The molecular basis of coupling between poly(A)-tail length and translational efficiency. *eLife* 2021, 10, e66493. <https://doi.org/10.7554/eLife.66493>.
- (4) Kuhn, A. N.; Diken, M.; Kreiter, S.; Vallazza, B.; Türeci, Ö.; Sahin, U. Determinants of Intracellular RNA Pharmacokinetics: Implications for RNA-Based Immunotherapeutics. *RNA Biol.* 2011, 8 (1), 35–43. <https://doi.org/10.4161/rna.8.1.13767>.
- (5) Belgrad, J.; Fakih, H.; Khvorova, A. Nucleic Acid Therapeutics: Successes, Milestones, and Upcoming Innovation. *Nucleic Acid Ther.* 2024, 34 (2), 52–72.
- (6) United States Pharmacopeia (USP). Draft Guidelines on Analytical Procedures for mRNA Vaccine Quality; USP: Rockville, MD, 2023.
- (7) Gilar, M.; Doneanu, C.; Gaye, M. Liquid Chromatography Methods for Analysis of mRNA Poly(A) Tail Length and Heterogeneity. *Anal. Chem.* 2023, 95, 14308–14316.
- (8) Ross, R.; England, J.; Bassignani, R.; Murphy, K.; Zhang, Y.; Du, M. Characterization of In Vitro-Transcribed (IVT) mRNA Poly(A) Tail by LC-HRAM-MS and BioPharma Finder 5.0 Software. Thermo Fisher Scientific Application Note, 2022.
- (9) Beverly, M.; Hagen, C.; Slack, O. Poly(A) Tail Length Analysis of In Vitro Transcribed mRNA by LC-MS. *Anal. Bioanal. Chem.* 2018, 410, 1667–1677.
- (10) Warzak, D. A.; Pike, W. A.; Luttgeharm, K. D. Capillary Electrophoresis Methods for Determining the IVT mRNA Critical Quality Attributes of Size and Purity. *SLAS Technol.* 2023, 28 (5), 369–374. <https://doi.org/10.1016/j.slast.2023.06.005>.
- (11) Su, W.-C.; Yu, Z.; Lieu, R.; Freund, E.; Hokeness, K.; Gu, Y.; Kohler, J.; Chen, T.; Fan, Y.; Zhang, K. Developing a High-Throughput Capillary Gel Electrophoresis Workflow for mRNA Integrity and Poly(A) Tail Analysis. *J. Chromatogr. A* 2025, 1761, 466387.

- (12) Di Grandi, D.; Dayeh, D.; Kaur, K.; Chen, Y.; Henderson, S.; Moon, Y.; Bhowmick, A.; Ihnat, P.; Fu, Y.; Muthusamy, K.; Palackal, N.; Pyles, E. A Single-Nucleotide Resolution Capillary Gel Electrophoresis Workflow for Poly(A) Tail Characterization in the Development of mRNA Therapeutics and Vaccines. *J. Pharm. Biomed. Anal.* 2023, 236, 115701.
- (13) Ogami, K.; Oishi, Y.; Hoshino, S. Protocol for Analyzing Intact mRNA Poly(A) Tail Length Using Nanopore Direct RNA Sequencing. *STAR Protoc.* 2023, 4, 102393.
- (14) Chang, J. J.; Yang, X.; Teng, H.; Zhang, J.; Reames, B.; Zhang, S.; Corbin, V.; Coin, L. J. M. Using Synthetic RNA to Benchmark Poly(A) Length Inference from Direct RNA Sequencing. *Gigascience* 2025, 14, giaf098. <https://doi.org/10.1093/gigascience/giaf098>.
- (15) Gau, B. C.; Dawdy, A. W.; Wang, H. L.; et al. Oligonucleotide Mapping via Mass Spectrometry to Enable Comprehensive Primary Structure Characterization of an mRNA Vaccine against SARS-CoV-2. *Sci. Rep.* 2023, 13, 9038. <https://doi.org/10.1038/s41598-023-36193-2>.