Full Spectral Confirmation of Multiply Charged RNA Molecules and Their Modifications with a Single Quadrupole LC/MS System

Yongdong Wang[1], Don Kuehl[1], Stacey Simonoff[1], Yue Su[2,3], Ning Zhang[2,3], Shundi Shi[3] and Shenglong Zhang[2]
[1]Cerno Bioscience, Las Vegas, NV; [2] New York Institute of Technology, NY, NY; [3]Columbia University, NY, NY

ThP 239

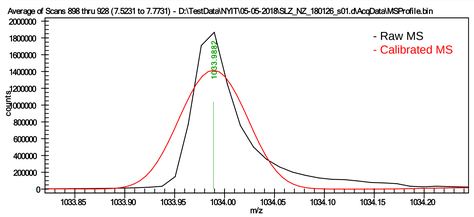
Introduction

Biomolecules are typically measured as multiply charged ions with MS, requiring post-processing to convert the measured m/z into the original mass through an iterative, typically nonlinear, and often ill-understood process involving trial values for the original mass. Recognizing that MS measures not just single masses but entire distributions of all relevant masses and their abundances, we propose a different approach by eliminating the error- and artifact-prone original mass estimation step from the process altogether while achieving biomolecule identification and relative quantitation through full spectral calibration and analysis. This new approach was tested here with synthetic RNA molecules measured from a higher resolution LC-TOF and compared to unit resolution single quadrupole LC-MS for the identification and relative quantitation of RNA molecules and their modifications. Accurate identification of RNA molecules, their fragments, and modifications are of great scientific and therapeutic importance[1].

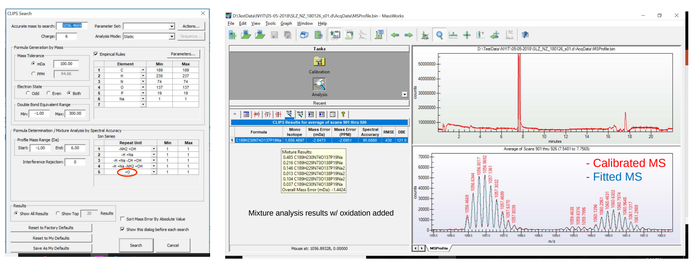
Method

We break the process into three discrete but related steps: (1) delineate and mathematically account for the physical, chemical, and instrumental factors, namely charge state, isotope distribution, MS instrument line shape, and spectral overlaps arising from adducts or modifications; (2) apply the model along with a known standard (either singly or multiply charged) to calibrate the profile mode MS data for both mass and spectral accuracy[2] with the MassWorks software tool[3]; (3) Acquire LC/MS data of RNA samples under high and unit mass resolution for the identification and relative quantitation, directly in the original mass spectral (m/z) space, utilizing all multiply charged signals but without any explicit multiple charge deconvolution step. We call this approach **SAMMI**, short for **S**pectrally **A**ccurate **M**odeling of **M**ulti-charged **I**ons[4], currently implemented with MATLAB.

LC-TOF MS Identification of RNA Molecules and Possible Ion Species



A 20-nt RNA measured on Agilent LCqTOF instrument with an internal single charge reference standard contained in the mobile phase at m/z 1034. A full spectral calibration called TrueCal™ within MassWorks software allows for the TOF peak shape to be calibrated into a known functional form and thereby calibrating for both mass (m/z) accuracy and spectral accuracy.

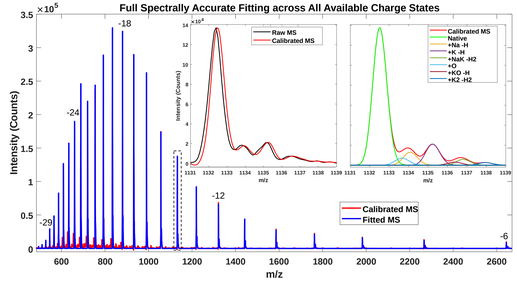


As previously reported[4], MassWorks CLIPS (Calibrated Lineshape Isotope Profile Search) known elemental composition search with the mixture mode analysis turned on correctly identified the 20-nt RNA sodium adduct in the presence of modifications and other adducts, to the spectral accuracy of 95.7%, accounting for a high percentage of the mass spectral signals observed in the given mass spectral range. The relative concentrations of various adducts and modifications were also obtained during the same spectral accuracy CLIPS analysis process, including the identification of a biochemically plausible modification (+O) and its relative quantitation at 3.7%.

Once these possible adducts and modifications have been identified and verified on the higher resolution TOF system, we proceeded to analyze even longer RNA strands using a more affordable and widely available LC/MSD single quadrupole system to demonstrate its potential as a QA/QC confirmation tool of RNA molecules in the presence of adducts or modifications.

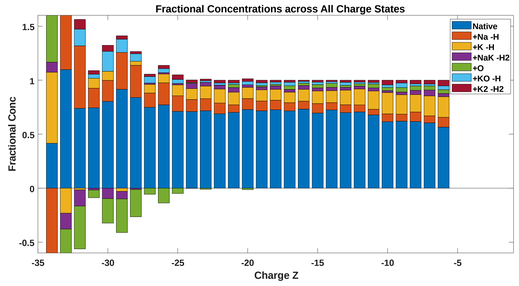
LC/MSD Analysis of 50-nt RNA

Seq: 5′-HO-**UAUUCAAGUUACACUCAAGAAGGAAUAAUUUCUAAACCGUUACCAUUACU**-OH-3′
Neutral formula: $C_{474}H_{585}N_{182}O_{345}P_{49}$, Exact Monoisotopic Mass =15,862.0969



Through a separate loop injection of the Agilent tune solution standard, both accurate mass and spectral accuracy calibration are achieved all the way to m/z 3,000. This enables the exact modeling of all possible ion species across all possible charge states. These ion species include not only what is typically expected, the negatively charged RNA molecules after the loss of one or more protons but also Na or K adducts and their combinations with each other and other modifications, for a total of 7 possible ion species for each charge state. With more than 20 charge states observable within the m/z range, a total of more than 140 ion species are fitted to the calibrated profile mode mass spectral data through Multiple Linear Regression (MLR) to obtain their respective relative concentrations.

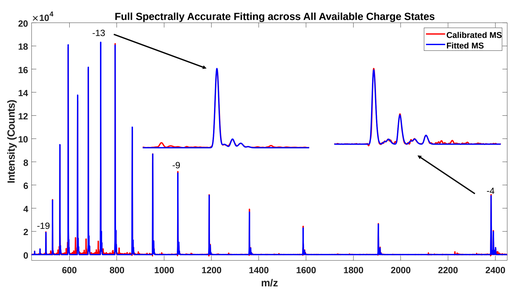
It is interesting to note that abundance distribution across the charges is not monotonic but seems to be multimodal, indicating multiple major forces at play during ionization. At unit mass resolution, the mass spectral signals contributed by some underlying ion species amount to no more than a small bump or shoulder in the corresponding mass spectral profile, further underscoring the importance of spectral accuracy and the full spectral calibration that helps achieving it.



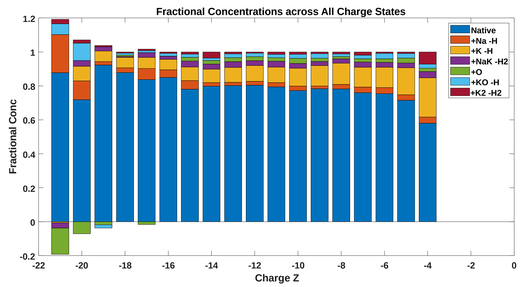
From the more than 140 relative concentrations obtained, corresponding to the more than 140 ion species possibly present, the relative concentrations can be grouped and normalized to unity under each charge state to obtain the fractional concentrations of various ion species. While these fractional concentrations show that the native form $(M\text{-}H)^-$ dominates with more than half of the spectral signal, other forms such as Na and K adducts also exist in significant fractions and they do change from one charge state to another, approaching 2x variation across the full range of charges observed. The negative fractional concentrations under high charges are likely caused by the lack of signal to noise and/or possible absence of these ion species and the presence of other ion species not already included in the MLR fitting.

LC/MSD Analysis of 30-nt RNA

Seq: 5′-HO-**UCGACUCUAGAGGAUCCCCACGUACGAUAC**-OH-3′
Neutral formula, $C_{284}H_{355}N_{112}O_{206}P_{29}$, Exact Monoisotopic Mass = 9,527.3136



While the monoisotopic mass is not directly observable, the various ion species do contribute to the overall mass spectral response and make a difference in the overall profile. Between the charge of -4 and -13 shown in the above inserts, it is spectrally clear that the relative concentrations of these ion species are different, indicating that the ionization mechanism that gives rise to different ion charges is also contributing to the different relative amounts of these ion species.



Conclusions

- It is feasible and advantageous to perform multiply charged ion species analysis directly in the original m/z space. The key advantage is that the results are directly and spectrally interpretable and any errors, either from the measurement itself or the algorithm applied, can be directly traced and understood in a scientifically transparent way.
- Full mass spectral calibration w.r.t. both the m/z and spectral accuracy is critical for accurate and direct spectral analysis.
- The fractional concentration of a given ion species does vary, sometime quite significantly, across charge states. A multiple charge deconvolution algorithm that makes the implicit or explicit assumption of constant fractions will be severely impacted with unexplained systematic errors in the results without warning.
- In the SAMMI approach presented, each ion species at any given charge state is treated as an independent unknown variable to be determined without any implicit or explicit assumptions. The amount of each ion species can be summed across different charge states afterwards for total quantitation.
- While high resolution systems such as LC TOF MS can serve as powerful tools in establishing the ion species to be considered for SAMMI, the more readily available and workhorse unit mass resolution systems such as LC/MSD can be employed for the fast and accurate routine QC/QA analysis of biomolecules.

Acknowledgement

We would like to thank Dr. Barney Yoo from Hunter College - CUNY for the LC/TOF MS data and Mr. Doug Postl and Mr. Dat Phan from Agilent Technologies for the LC/MSD data. The New York Institute of Technology and Columbia University team would like to acknowledge the support from NIH grants 1R21HG009576 and 1R56HG011099-01. The Cerno Bioscience team would like to acknowledge the support from NIH grant 1R56HG011099-01.

References

1. Björkbom, A.; Lelyveld, V.S.; Zhang, S.; Zhang, W.; Tam, C. P.;Blain, J.C.;Szostak, J.W., J. Am. Chem. Soc., **2015**, *137*, 45, 14430.
2. Y. Wang, M. Gu., Anal. Chem., **2010**, *82*, 17, 7055.
3. MassWorks, Cerno Bioscience, Las Vegas, NV, USA.
4. Zhang, N. et al, 67th ASMS Conference Poster, TP 662, June 2-6, 2019, Atlanta, GA.