analytical chemistry

pubs.acs.org/ac Letter

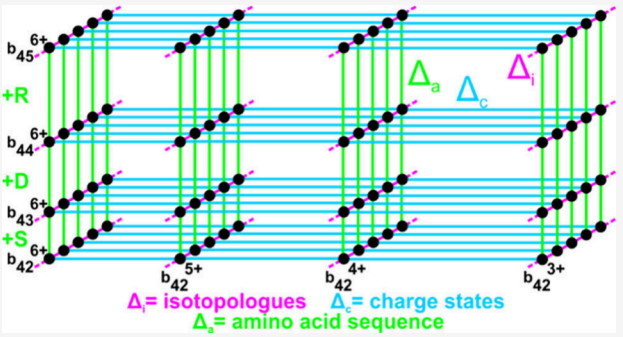
# Mass-Invariant Natural Log-Transformed Mass Spectra Enable Internal Calibration and *De Novo* Sequencing of Intact Proteins

Lissa C. Anderson,* Nathan K. Kaiser, Krishna Saketh Kamadana, and Xian Mallory

Cite This: *Anal. Chem.* 2026, 98, 1839−1845 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

**ABSTRACT:** A key limitation of top-down proteomics is reliance on averagine-based deconvolution to estimate monoisotopic masses, which introduces systematic errors when isotope envelopes are distorted. We present a framework that bypasses averagine by operating directly in natural log−transformed $m/z$ space, where charge-state spacing is mass-invariant and provides an intrinsic reference for internal calibration on both FT-ICR and Orbitrap analyzers. Isotopologue pairing in this domain supports *de novo* sequencing and discriminates near-isobaric residues. By shifting the paradigm from monoisotopic mass estimation to connectivity-driven inference, the approach offers resilience against distorted isotope envelopes and unknown PTMs, establishing a database-independent strategy for discovery-oriented proteoform characterization without known calibrants.






## INTRODUCTION

Intact protein analysis presents persistent challenges due to the complexity of isotope distributions arising from the incorporation of heavy isotopes. As protein mass increases, the number of heavy isotope incorporations rises, diminishing the relative abundance of the monoisotopic peak. Often the monoisotopic peak is weak or undetectable. This complicates the process of determining accurate molecular mass, which is foundational to all top-down proteomic workflows.

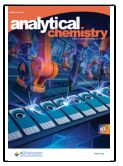
A common workaround is to estimate the monoisotopic mass using the "averagine" model, which approximates an average elemental composition for amino acids to predict expected isotope distributions.[1] While convenient, averagine-based fits can be shifted by one or more isotopologues if the monoisotopic peak is absent or misidentified, introducing systematic mass errors of 1−2 Da or more (Supplemental Figure S1).[2] To accommodate these errors, database search tolerances must be widened, increasing the risk of false positive assignments. Conversely, narrowing mass tolerances improves specificity but excludes true matches when mass estimates are inaccurate due to incorrect isotopic modeling. Additionally, mass measurement accuracy is inherently limited by the difference between the true elemental composition of the analyte and the assumptions built into the averagine model. This issue is further compounded by electrospray ionization (ESI), which generates multiple charge states for each analyte, splitting signal intensity across overlapping mass-to-charge ratio ($m/z$) values, requiring high-resolution and spectral averaging for accurate deconvolution.

Current protein identification strategies—whether based on database search, spectral libraries, or *de novo* sequencing—depend heavily on the accuracy of the monoisotopic mass measurements.[3] In database-driven approaches, experimental monoisotopic masses are compared against *in silico* predictions from protein sequence databases.[4−6] This inherently limits identification to sequences already present in the database. As a result, proteoforms containing unexpected sequence variants or post-translational modifications (PTMs) are not accurately identified or remain unidentified.[7] For nonmodel organisms or novel proteoforms, such as antibodies, *de novo* approaches remain the only option.

However, top-down *de novo* sequencing itself remains limited by the same assumptions: most methods still rely on averagine-based deconvolution to derive fragment and precursor monoisotopic masses prior to sequence inference.[8−11] The inability to reliably determine monoisotopic mass without prior compositional knowledge hinders both identification and accurate interpretation of spectra.

The method presented here provides a conceptual framework for database-independent, *de novo* sequencing of intact proteins that does not require monoisotopic mass assignment or $m/z$-to-mass deconvolution via averagine fitting. The approach is inspired in part by the FLASHDeconv algorithm, which identifies charge-state series in spectra by searching for

Received: October 3, 2025
Revised: December 30, 2025
Accepted: January 8, 2026
Published: January 15, 2026







https://doi.org/10.1021/acs.analchem.5c06165
*Anal. Chem.* 2026, 98, 1839−1845

mass-invariant patterns in natural log-transformed $m/z$ space.[12] While FLASHDeconv uses this principle for spectral decharging, the framework described here uses a related concept to construct mass-difference networks directly from isotopically resolved tandem mass spectrometry (MS/MS) data. These networks enable the identification of sequence-informative relationships between peaks without requiring an accurate precursor monoisotopic mass, opening new possibilities for accurate, untargeted proteoform sequencing. This usage of "mass-difference network" is conceptually related to the "spectrum graph" (or "spectral graph") formulation developed in classical *de novo* sequencing algorithms by Pevzner and colleagues.[13] In those approaches, nodes represent prefix-residue masses from a prefix residue mass (PRM) spectrum and edges correspond to amino-acid mass differences within the peptide- or protein-sequencing framework.[14] Our use of "mass-difference network," however, denotes a more general graph that connects any pair of fragment ions by chemically meaningful mass differences, without the PRM-specific constraints or assumptions about peptide linearity.

Importantly, the mass-invariant charge pattern in natural log-transformed $m/z$ space also provides a powerful means for internal calibration of mass spectra from electrospray-ionized intact proteins without knowledge of known molecular formulas. This is especially valuable in ion trapping instruments with limited charge capacity—such as FT-ICR and Orbitrap mass analyzers—where space charge effects can distort measured frequencies and lead to systematic $m/z$ shifts.[15−17] By adjusting the electric field term in the calibration equation such that the peak positions better match the expected charge-state pattern, it is possible to mitigate space-charge-induced frequency shifts for improved mass analysis. Although previous approaches have proposed strategies for internal calibration based on mass difference analysis,[18−21] they do not leverage the mass-invariant charge pattern in log-transformed $m/z$ space as described here.

## METHODS

The FLASHDeconv algorithm developed by Jeong et al.[12] is open-source, platform-independent software implemented in OpenMS.[22] It consists of three subalgorithms: spectral decharging, deisotoping (via averagine fitting), and feature finding. Only spectral decharging is relevant here.

Briefly, the $m/z$ of an analyte is given by:

$$\frac{m}{z} = \frac{m_N + zm_A}{z} = \frac{m_N}{z} + m_A \therefore \frac{m}{z} - m_A = \frac{m_N}{z} \tag{1}$$

where $m$ is the mass of the charged analyte, $m_N$ is the neutral analyte mass, $z$ is the charge state, and $m_A$ is the charge carrier mass (a proton, 1.007276 Da).

Natural log-transforming results from eq 1 yields:

$$\ln\left(\frac{m}{z} - m_A\right) = \ln\left(\frac{m_N}{z}\right) = \ln(m_N) - \ln(z) \tag{2}$$

For a given $m_N$, the distance between charge states in transformed space forms a universal charge pattern vector, where $z_{min}$ and $z_{max}$ denote the minimum and maximum observed charge states of $m_N$, respectively:

$$U := (-\ln(z_{\min}), -\ln(z_{\min} + 1), ..., -\ln(z_{\max})) \tag{3}$$

This vector represents expected ln-spacing between consecutive charge states, which enables rapid identification of peaks arising from the same mass across charge states.[12]

Since the expected spacing between charge states is precisely known, differences between peaks arising from the same mass can be used to perform an internal calibration of the spectrum in natural log-transformed space. For Fourier-transform ion cyclotron resonance (FT-ICR) mass spectrometry, observed ion cyclotron frequency, $f$, can be converted to $m/z$ via the Ledford equation,[15] given below:

$$\frac{m}{z} = \frac{A}{f} + \frac{B}{f^2} \tag{4}$$

The $A$ term is the magnetic field coefficient, and the $B$ term is related to the electric trapping field. By substituting this expression for $m/z$ in eq 1 and applying the natural log transformation, a calibration relationship can be derived between adjacent charge states ($z_n$ and $z_{n+1}$) of the same mass:

$$\ln(z_{n+1}) - \ln(z_n) = \ln\left(\frac{A}{f_n} + \frac{B}{f_n^2} - m_A\right) - \ln\left(\frac{A}{f_{n+1}} + \frac{B}{f_{n+1}^2} - m_A\right) \tag{5}$$

This equation enables internal calibration based solely on the relative positions of isotopologue peaks across charge states, without requiring known calibrants or prior sequence knowledge.

Unlike methods that convert MS/MS data back to mass space, this framework infers amino acid sequences directly from natural log-transformed $m/z$ values. Following backbone fragmentation, consecutive fragment ions differ by the residue mass (RM) of individual amino acids (or residue mass plus PTMs). For fragments with the same charge state, the natural log-transformed values follow:

$$\ln\left(\frac{m}{z} - m_A\right)_{n\pm1} = \ln\left(\left(\frac{m}{z} - m_A\right)_n \pm \left(\frac{\mathrm{RM}}{z}\right)\right) \tag{6}$$

This relationship enables direct detection of sequence-specific mass differences within the natural log-transformed space, eliminating the need for monoisotopic mass assignment or back-conversion to mass space. For residue-mass hypothesis testing, we considered the top 3−5 most abundant isotopologues per fragment, assuming that at least one pair shares the same number of additional neutrons, enabling accurate ln-space comparisons without prior deisotoping.

Mass spectra were acquired with a custom-built 21 tesla (T) FT-ICR mass spectrometer at the National High Magnetic Field Laboratory (Tallahassee, FL) and an Orbitrap Eclipse Tribrid mass spectrometer (Thermo Fisher Scientific).[23] No phase correction was applied. For calibration of FT-ICR data, ion frequencies were obtained directly from the transient, whereas for Orbitrap data, they were calculated indirectly using the calibration terms.

Proteins were directly ionized by positive electrospray ionization via direct infusion and high-performance liquid chromatography. These methods are described in the Supporting Information.

## RESULTS AND DISCUSSION

To evaluate the performance of this framework, we applied it to isotopically resolved top-down MS/MS spectra acquired from standard proteins apomyoglobin (equine, 17 kDa), Protein G (*Streptococcus*, 21 kDa), and Carbonic Anhydrase II (bovine, 29 kDa) using collision- and electron-based fragmentation. All spectra were analyzed without monoisotopic mass assignment. Instead, peaks were mapped directly into $\ln(m/z - m_A)$ space. In this transformed space, both sequence-informative residue mass differences and charge-state alignment patterns were evaluated for consistency with theoretical expectations.

### Internal Calibration

In FT-ICR, Coulombic interactions among trapped ions slightly decrease the observed cyclotron frequency, so calculated $m/z$ is underestimated, producing negative mass


https://doi.org/10.1021/acs.analchem.5c06165
Anal. Chem. 2026, 98, 1839−1845

error. Figure 1 demonstrates that the mass-invariant charge pattern in $\ln(m/z - m_A)$ space can be exploited not only for

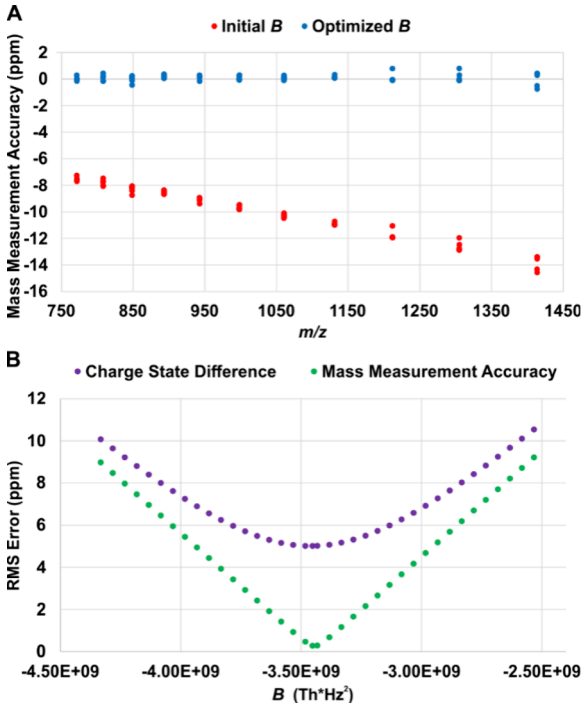



**Figure 1.** Internal recalibration of a 21 T FT-ICR MS[1] spectrum of apomyoglobin with the mass-invariant charge pattern. (A) Mass measurement accuracy versus $m/z$ for the five most abundant consecutive isotopologues of charge states $12^+$ to $22^+$ before (red) and after (blue) optimization of the Ledford B-term. (B) Visual representation that iteration of the B-term in the Ledford equation to minimize charge state difference error (purple) reflects improved consistency in the expected mass-invariant charge pattern, while the concurrent dip in mass error demonstrates enhanced mass measurement accuracy (green). RMS errors improved from 10.22 to 0.27 ppm after optimization.

charge state determination, but also as an intrinsic reference for internal calibration. The systematic slope in mass error observed in the apomyoglobin spectrum prior to correction (Figure 1A, red) reflects a global frequency shift typical of space charge effects in FT-ICR instruments. Because space-charge effects scale with total ion current and can fluctuate across LC gradients, adjacent scans often exhibit correlated frequency shifts. This correlation enables effective scan-by-scan calibration. Iteratively adjusting the B-term restores the expected natural log-domain intervals and flattens the mass error slope (Figure 1A, blue), reducing the RMS error from 10.22 to 0.27 ppm across charge states $12^+$ to $22^+$. The close correspondence between the minima in charge spacing error and overall mass measurement error (Figure 1B) highlights that the two phenomena are coupled, and that correction of the charge-state vector directly translates into improved mass accuracy. In this approach, only the B-term of the Ledford equation—responsible for accounting for electric field perturbations—is adjusted to minimize charge state difference error. The A-term, which has minimal impact on this error, must still be accurately determined to compensate for magnetic field drift over time.

Internal calibration remains beneficial at all magnetic field strengths, as space-charge effects persist even in high-field instruments, though their magnitude decreases with increasing field. The same principle applies across Fourier-transform analyzers, including Orbitrap systems. The ideal Orbitrap relation is $m/z \propto 1/f^2$, but highly charged ion populations in top-down experiments amplify space-charge effects that distort AGC targets and trap fill times, producing a characteristic curvature in the mass-error versus $m/z$ trend.[24] To account for this higher-order frequency dependence, we modeled the calibration as:

$$\frac{m}{z} = \frac{A'}{f^2} + \frac{B'}{f^4} \tag{7}$$

where $A'$ is the primary scale coefficient and $B'$ is a small higher-order correction (not to be confused with the $A$- and $B$-terms of the FT-ICR Ledford equation). The natural log space charge-pattern method applies directly by treating B' as the adjustable parameter, analogous to $B$-term optimization in FT-ICR. Applying this calibration to Orbitrap data (Supplemental Figure S2) yielded comparable improvement in mass measurement accuracy, underscoring that the approach exploits the universal frequency-to-$m/z$ relationship of all trapped ion Fourier-transform analyzers. This accuracy is essential for downstream mass-difference network construction, where even small calibration errors can propagate into incorrect assignments.

## Mass Difference Network Construction

Supplemental Figure S3 demonstrates that the natural log domain charge-pattern intervals are highly precise, validated to six decimal places, and independent of isotope statistics. This precision provides a robust internal reference not only for calibration, but also for sequence inference. Because isotope abundances fluctuate stochastically between acquisitions—particularly in ion-trapping instruments with limited charge capacity—relative intensities cannot be relied upon to assign corresponding isotopologues among consecutive fragment ion peak clusters. Instead, isotopologue pairing in $\ln(m/z - m_A)$ space supports construction of mass-difference networks that remain accurate when monoisotopic peaks are weak or isotope envelopes are distorted.

Building on this foundation, the framework expands into a a three-dimensional mass-difference network, as illustrated in Supplemental Figure S4. After grouping peaks using predictable $\ln(m/z - m_A)$ spacing, each peak becomes a node in the network. Nodes are linked by edges that encode stable chemically meaningful relationships: (a) charge-state ladders (mass-invariant), (b) isotopic composition differences (charge-invariant), and (c) amino acid residue mass changes. Residue-mass hypothesis testing can be performed entirely in ln-space, avoiding repeated exponentiation back to mass space. When fragment charge is uncertain, plausible $z$ values can be iterated based on the spectrum's charge distribution and tested against predicted ln-space intervals within tolerance. This connectivity transforms the spectrum into a graph structure that captures all relevant relationships among fragment ions. Graph traversal algorithms—such as depth-first or breadth-first search—can then identify clusters or "communities" corresponding to proteoform families or contiguous sequence tags, enabling residue-order inference from the resulting connected paths.

## *De Novo* Sequencing

Figure 2 illustrates the application of the charge-resolved, ln-transformed mass-difference framework to a collision-induced


https://doi.org/10.1021/acs.analchem.5c06165
Anal. Chem. 2026, 98, 1839−1845

| Empirical Data | | | Predicted $\ln(m/z-m_A)_{n-1}$ | | |
|---|---|---|---|---|---|
| $(m/z)_n$ | $(m/z)_{n-1}$ | $\ln(m/z-m_A)$ n-1 | RM=P | RM=V | RM=T |
| 1396.061 | 1387.055 | 7.234212 | 7.234344 | 7.234212 | 7.234082 |
| 1396.153 | 1387.146 | 7.234277 | 7.234410 | 7.234278 | 7.234148 |
| 1396.244 | 1387.237 | 7.234343 | 7.234475 | 7.234343 | 7.234213 |
| | | | RM=E | RM=M | RM=H |
| 1387.055 | 1375.415 | 7.225778 | 7.225712 | 7.225580 | 7.225182 |
| 1387.146 | 1375.506 | 7.225845 | 7.225778 | 7.225646 | 7.225248 |
| 1387.237 | 1375.597 | 7.225911 | 7.225845 | 7.225712 | 7.225314 |
| | | | RM=P | RM=V | RM=T |
| 1375.415 | 1366.138 | 7.219006 | 7.219338 | 7.219204 | 7.219072 |
| 1375.506 | 1366.229 | 7.219072 | 7.219405 | 7.219271 | 7.219139 |
| 1375.597 | 1366.320 | 7.219139 | 7.219472 | 7.219337 | 7.219206 |
| | | | RM=P | RM=V | RM=T |
| 1366.138 | 1357.223 | 7.212453 | 7.212522 | 7.212387 | 7.212254 |
| 1366.229 | 1357.314 | 7.212521 | 7.212589 | 7.212454 | 7.212321 |
| 1366.320 | 1357.405 | 7.212588 | 7.212656 | 7.212521 | 7.212388 |
| | | | RM=E | RM=M | RM=H |
| 1357.223 | 1345.310 | 7.203631 | 7.203766 | 7.203631 | 7.203224 |
| 1357.314 | 1345.402 | 7.203699 | 7.203834 | 7.203699 | 7.203292 |
| 1357.405 | 1345.493 | 7.203767 | 7.203901 | 7.203766 | 7.203359 |
| | | | RM=E | RM=M | RM=H |
| 1345.310 | 1333.488 | 7.194797 | 7.194866 | 7.194730 | 7.194319 |
| 1345.402 | 1333.578 | 7.194866 | 7.194935 | 7.194799 | 7.194388 |
| 1345.493 | 1333.669 | 7.194934 | 7.195003 | 7.194867 | 7.194456 |

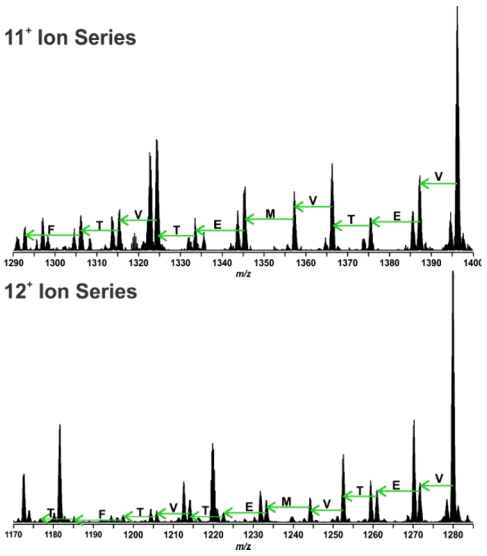



**Figure 2.** Sequence-tag inference from an FT-ICR CID MS/MS spectrum (post-FT average of 3 scans) of the $[M + 19H]^{19+}$ precursor of Protein G (21 kDa). (Left) $m/z$ and $\ln(m/z - m_A)$ values for the three most abundant consecutive isotopologues of seven fragments ($11^+$). Predicted $\ln(m/z - m_A)_{n-1}$ values were calculated from the observed $(m/z)_n$ using eq 6 and compared to observed values to assess isotopologue correspondence. The observed $\ln(m/z - m_A)_{n-1}$ values (yellow) are matched to the predicted values (green). (Right) corresponding fragment series and the derived ten-residue tag (VETVMETVTF), independently confirmed in the $12^+$ series.

dissociation MS/MS spectrum of Protein G (21 kDa). The two left-most columns of the table list the empirical $m/z$ values of the three most abundant consecutive isotopologues for a series of seven consecutive fragment ions. Predicted $\ln(m/z - m_A)_{n-1}$ were calculated from the observed $(m/z)_n$ values using eq 6. Observed $\ln(m/z - m_A)_{n-1}$ were then compared to predicted values to assess isotopologue correspondence. In all cases, at least two of the three most intense isotopologues matched predicted values within $\pm 1 \times 10^{-6}$, enabling correct isotopologue pairing between consecutive fragment ion peak clusters. When only two of three matched, the most abundant isotopologues differed between adjacent peak clusters. This is expected due to the limited spectral averaging performed here (post-FT average of 3 scans). When a peak is mis-assigned by one isotopologue the resulting $\ln(m/z - m_A)$ difference greatly exceeds the $\pm 1 \times 10^{-6}$ criterion, flagging the mismatch and preventing propagation of an incorrect residue-mass difference. With this approach, a ten-residue sequence tag (VETVMETVTF) was derived from the $11^+$ charge state. The same tag was independently determined from the $12^+$ fragment ion series.

The method also resolved near-isobaric residues—specifically glutamine (Q) and lysine (K)—in a single-scan MS/MS spectrum of Carbonic Anhydrase II (Supplemental Figure S5). Despite the high charge state ($20^+$), poor isotope envelope fidelity, and overlap of neighboring fragment ion signals, the natural log domain framework accurately paired the correct isotopologues and recovered the expected residue-mass difference. These results show that eq 6 identifies residue-mass differences directly from natural log-transformed data, supporting *de novo* sequencing from a single transient without monoisotopic determination.

To visualize the organization of fragment ions, c- and $z^\bullet$-type fragments from a 1500-scan average, 6 ms ETD spectrum of Carbonic Anhydrase II were plotted with $\ln(m/z - m_A)$ on the x-axis and neutral mass on the y-axis (Figure 3).[2] The monoisotopic $\ln(m/z - m_A)$ value of identified fragments is represented with an "x". A total of 492 c-ions and 548 $z^\bullet$-ions were manually validated.[25] The spectrum was subsequently processed with Agilent ExDViewer software (version 4.6.28) to annotate charge states for raw fragment ion isotopologue peaks.[26] No averagine-derived monoisotopic masses were employed. The $m/z$ and charge states of c- and $z^\bullet$-ion peaks matched to the sequence by ExDViewer were used to calculate the neutral masses of individual isotopologues with eq 8:

$$m_N = ze^{\ln(m/z - m_A)} \tag{8}$$

These data are shown as circles ("o"). In this representation—referred to here as a natural log-transformed charge-mass plot—fragment ions form along distinct curved paths determined by their charge state. This provides convenient visual separation of different charge states, which can aid in examining mass differences between fragment peak clusters along each curve. For Carbonic Anhydrase II, the distinct separation of the c- and $z^\bullet$-fragment ion series in the $11^+$ and $5^+$ charge states can be attributed to differences in the rate that these ions gain charge as mass increases (i.e., the number of lysine, arginine, and histidine residues grows faster in the "forward" N- to C-terminal direction). Isotopologues of the same fragment and charge state align vertically, forming clear stacks above their corresponding theoretical monoisotopic positions as shown in Supplemental Figure S6. When multiple charge-state traces overlap in $m/z$, these stacks remain distinct because each charge state follows its own curved trajectory in the transformed coordinates. This separation reduces ambiguity in congested regions of the spectrum and facilitates the construction of sequence tags by providing a reliable way to track which peaks belong together. Sequence tag confidence increases when mass differences are confirmed by more than one isotopologue pair, or when the same residue mass shift is detected in multiple charge states, providing internal consistency that strengthens *de novo* sequence assignments and highlights the utility of natural log-transformed charge-mass plots for interpreting complex top-down MS/MS spectra.

The natural log-transformed charge−mass plot presented here was created to illustrate charge relationships among


https://doi.org/10.1021/acs.analchem.5c06165
Anal. Chem. 2026, 98, 1839−1845

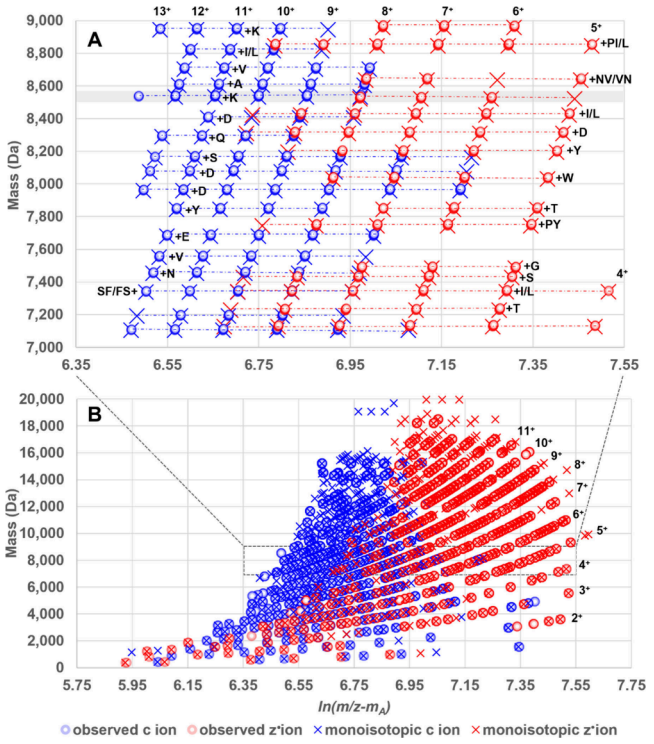



**Figure 3.** Natural Log-transformed charge-mass plot of the $[M + 34H]^{34+}$ precursor of Carbonic Anhydrase II (29 kDa) following ETD (1500 scan average).[2] (A) Expanded view of the boxed region in panel B. Dashed lines connect fragments observed in multiple charge states. An enlarged view of the shaded area in panel A, showing individual isotopologues within each peak cluster, is provided in Supplemental Figure S6. (B) Fragment masses were calculated for charge-assigned fragments with eq 8 and plotted against their $\ln(m/z - m_A)$ values, revealing distinct curves that correspond to the charge states of the fragment ion series. "X" symbols represent theoretical monoisotopic peaks of manually identified fragments. Circles indicate charge-assigned isotopologues matched to the sequence using ExDViewer.[26]

fragments, but it does not include fragment ion intensities or all measured ions. Adding these elements could provide additional context, such as highlighting highly abundant fragments or showing ions observed across multiple charge states. However, incorporating intensity overlays or every detected ion may introduce complexity that reduces interpretability. Future work could explore selective visualization strategies, such as highlighting ions supported by multiple charge states or interactive formats that allow toggling intensity information. To provide a more complete view of the data set, Supplemental Figure S7 shows *all* charge-state-assigned ions from Figure 3A, including fragments that correspond to the sequence but that are not c- or z•-type (e.g., a•-, y-, and neutral losses), as well as ions that were not matched to the sequence.

There are several practical considerations that should be acknowledged regarding the proposed framework. First, the approach performs best when spectra contain a broad range of charge states, as the calibration relies on detecting consistent charge-spacing patterns across multiple states. Spectra with narrow charge state distributions—such as those from native ESI—offer fewer reference points, reducing calibration accuracy. Second, incomplete fragmentation limits the number of sequence-informative mass differences available for constructing reliable tags, which can result in shorter sequence tags and reduced confidence in *de novo* assignments. This challenge is particularly evident in single-scan acquisitions or when fragmentation efficiency is low. Third, low signal-to-noise ratios—where fragment ion peaks are barely distinguishable from background noise—can impair accurate pairing in $\ln(m/z - m_A)$ space, especially for minor fragments or low-abundance proteoforms. Noise-driven mismatches may propagate errors in mass-difference networks if not properly filtered. Although the ln-space pairing criterion minimizes off-by-one isotope errors, misassignment of near-isobaric peaks is theoretically possible—for example, the M+1 isotopologue of glutamine (Gln) is nearly isobaric with the monoisotopic peak of glutamic acid (Glu)—if lower-abundance isotopologues are considered. These issues are not unique to this method; they represent common challenges faced by all deconvolution and *de novo* sequencing approaches. Implementing robust statistical filters and confidence scoring for isotopologue pairing can help prevent error propagation in network-based sequence inference. Future work will focus on integrating these strategies and exploring machine-learning approaches to improve performance under challenging conditions such as complex mixtures or PTM-rich samples.

Although the examples presented here use canonical amino acid masses, the framework is readily extensible to incorporate PTMs or other chemical alterations by expanding the residue-mass difference set. This flexibility enables accurate interpretation of modified proteoforms without prior knowledge of their composition.

## CONCLUSIONS

This work demonstrates that the mass-invariant charge pattern in natural log-transformed space provides a universal internal reference for calibration that is inherently present in every electrospray-generated intact-protein spectrum. By exploiting this property, systematic frequency shifts in FTMS analyzers can be corrected without external calibrants or prior sequence knowledge. Because the calibration approach relies only on fundamental charge state spacing, it can be implemented as an automated, scan-by-scan routine to mitigate ion-population variability and maintain subppm mass accuracy over extended acquisitions. The same strategy can also be retroactively applied to previously collected data sets.

Beyond calibration, the framework enables true database-independent top-down *de novo* sequencing without averagine-based *m/z*-to-*m* deconvolution. Representing peaks as nodes and chemically meaningful mass differences as edges transforms complex spectra into navigable networks, where graph traversal algorithms can recover contiguous sequence tags or families of related proteoforms. This interpretation shifts the paradigm from monoisotopic mass estimation toward connectivity-driven inference, offering resilience against distorted isotopic envelopes and unknown PTMs.

Looking forward, this approach naturally lends itself to automated workflows. Machine-learning models trained on ln-space networks could learn characteristic connectivity signatures for amino-acids and PTMs directly from experimental data. Moreover, the underlying principles are not limited to proteins; the approach should be adaptable to other isotopically resolved biopolymers, including modified RNA and other oligonucleotides, where accurate mass-difference mapping is equally critical. Together, these advances point toward an integrated workflow where ln-space calibration and network-based sequencing deliver discovery-oriented biomo-


https://doi.org/10.1021/acs.analchem.5c06165
Anal. Chem. 2026, 98, 1839−1845

lecular characterization, from single-transient acquisition to comprehensive proteoform analysis.

## ■ ASSOCIATED CONTENT

### Data Availability Statement

Raw data files (.dat and .raw), along with R code and Excel spreadsheets used for B-term calibration iteration, are available for download from the Open Science Framework (OSF):http://doi.org/10.17605/OSF.IO/5AR6H.

### Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.analchem.5c06165.

Methods, list of experiments (Table S1), figures (PDF)

## ■ AUTHOR INFORMATION


### Corresponding Author

**Lissa C. Anderson** − *National High Magnetic Field Laboratory, Tallahassee, Florida 32310, United States; FSU Department of Chemistry & Biochemistry, Tallahassee, Florida 32306, United States;* orcid.org/0000-0001-8633-0251; Email: anderson@magnet.fsu.edu

### Authors

**Nathan K. Kaiser** − *National High Magnetic Field Laboratory, Tallahassee, Florida 32310, United States*

**Krishna Saketh Kamadana** − *FSU Department of Computer Science, Tallahassee, Florida 32306, United States*

**Xian Mallory** − *FSU Department of Computer Science, Tallahassee, Florida 32306, United States*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.analchem.5c06165


### Notes

The authors declare the following competing financial interest(s): Lissa C. Anderson is pursuing a patent related to the methods and technologies described here.

## ■ ACKNOWLEDGMENTS

This work was performed at the Ion Cyclotron Resonance User Facility at the National High Magnetic Field Laboratory at Florida State University, which is supported by the National Science Foundation (NSF) Divisions of Materials Research and Chemistry (DMR-2128556) and by the State of Florida. Additional support was provided by the NSF Division of Computing and Communications Foundations CCF 2523717 to Xian Mallory. Portions of the manuscript text and Supporting Information were drafted or refined with the assistance of AI-based language tools (e.g., ChatGPT, OpenAI), under the direct supervision and final review of the authors.

## ■ REFERENCES


(1) Senko, M. W.; Beu, S. C.; McLafferty, F. W. Determination of Monoisotopic Masses and Ion Populations for Large Biomolecules from Resolved Isotopic Distributions. *J. Am. Soc. Mass Spectrom.* **1995**, *6* (4), 229−233.

(2) Weisbrod, C. R.; Kaiser, N. K.; Syka, J. E. P.; Early, L.; Mullen, C.; Dunyach, J.-J.; English, A. M.; Anderson, L. C.; Blakney, G. T.; Shabanowitz, J.; Hendrickson, C. L.; Marshall, A. G.; Hunt, D. F. Front-End Electron Transfer Dissociation Coupled to a 21 T FT-ICR Mass Spectrometer for Intact Protein Sequence Analysis. *J. Am. Soc. Mass Spectrom.* **2017**, *28* (9), 1787−1795.

(3) Nesvizhskii, A. I. A Survey of Computational Methods and Error Rate Estimation Procedures for Peptide and Protein Identification in Shotgun Proteomics. *Journal of Proteomics* **2010**, *73* (11), 2092−2123.

(4) Eng, J. K.; McCormack, A. L.; Yates, J. R. An Approach to Correlate Tandem Mass Spectral Data of Peptides with Amino Acid Sequences in a Protein Database. *J. Am. Soc. Mass Spectrom.* **1994**, *5* (11), 976−989.

(5) Perkins, D. N.; Pappin, D. J. C.; Creasy, D. M.; Cottrell, J. S. Probability-Based Protein Identification by Searching Sequence Databases Using Mass Spectrometry Data. *Electrophoresis* **1999**, *20* (18), 3551−3567.

(6) Taylor, G. K.; Kim, Y.-B.; Forbes, A. J.; Meng, F.; McCarthy, R.; Kelleher, N. L. Web and Database Software for Identification of Intact Proteins Using "Top Down" Mass Spectrometry. *Anal. Chem.* **2003**, *75* (16), 4081−4086.

(7) Smith, L. M.; Thomas, P. M.; Shortreed, M. R.; Schaffer, L. V.; Fellers, R. T.; LeDuc, R. D.; Tucholski, T.; Ge, Y.; Agar, J. N.; Anderson, L. C.; Chamot-Rooke, J.; Gault, J.; Loo, J. A.; Pasa-Tolic, L.; Robinson, C. V.; Schluter, H.; Tsybin, Y. O.; Vilaseca, M.; Vizcaino, J. A.; Danis, P. O.; Kelleher, N. L. A Five-Level Classification System for Proteoform Identifications. *Nat. Methods* **2019**, *16* (10), 939−940.

(8) Horn, D. M.; Zubarev, R. A.; McLafferty, F. W. Automated de Novo Sequencing of Proteins by Tandem High-Resolution Mass Spectrometry. *Proc. Natl. Acad. Sci. U. S. A.* **2000**, *97* (19), 10313−10317.

(9) Liu, X.; Lennard; Wu, S.; VanDuijn, M. M.; Luider, T. M.; Tolić, Nikola; Kou, Q.; Dvorkin, M.; Alexandrova, S.; Vyatkina, K.; Paša-Tolić, Ljiljana; Pevzner, P. A. De Novo Protein Sequencing by Combining Top-down and Bottom-up Tandem Mass Spectra. *J. Proteome Res.* **2014**, *13* (7), 3241−3248.

(10) Vyatkina, K.; Wu, S.; Dekker, L. J. M.; VanDuijn, M. M.; Liu, X.; Tolić, N.; Dvorkin, M.; Alexandrova, S.; Luider, T. M.; Paša-Tolić, L.; Pevzner, P. A. De Novo Sequencing of Peptides from Top-down Tandem Mass Spectra. *J. Proteome Res.* **2015**, *14* (11), 4450−4462.

(11) Vyatkina, K.; Wu, S.; Lennard; VanDuijn, M. M.; Liu, X.; Tolić, Nikola; Luider, T. M.; Paša-Tolić, Ljiljana; Pevzner, P. A. Top-down Analysis of Protein Samples by de Novo Sequencing Techniques. *Bioinformatics* **2016**, *32* (18), 2753−2759.

(12) Jeong, K.; Kim, J.; Gaikwad, M.; Hidayah, S. N.; Heikaus, L.; Schluter, H.; Kohlbacher, O. FLASHDeconv: Ultrafast, High-Quality Feature Deconvolution for Top-down Proteomics. *Cell Systems* **2020**, *10* (2), 213−218.e6.

(13) Dančík, V.; Addona, T. A.; Clauser, K. R.; Vath, J. E.; Pevzner, P. A. De Novo Peptide Sequencing via Tandem Mass Spectrometry. *Journal of Computational Biology* **1999**, *6* (3−4), 327−342.

(14) Kou, Q.; Wu, S.; Tolic, N.; Pasa-Tolic, L.; Liu, Y.; Liu, X. A mass graph-based approach for the identification of modified proteoforms using top-down tandem mass spectra. *Bioinformatics* **2017**, *33*, 1309−1316.

(15) Ledford, E. B.; Rempel, D. L.; Gross, M. L. Space Charge Effects in Fourier Transform Mass Spectrometry. II. *Mass Calibration. Analytical Chemistry* **1984**, *56* (14), 2744−2748.

(16) Chen, S.; Comisarow, M. B. Simple Physical Models for Coulomb-Induced Frequency Shifts and Coulomb-Induced Inhomogenous Broadening for like and Unlike Ions in Fourier Transform Ion Cyclotron Resonance Mass Spectrometry. *Rapid Commun. Mass Spectrom.* **1991**, *5* (10), 450−455.

(17) Easterling, M. L.; Mize, T. H.; Amster, I. J. Routine Part-Per-Million Mass Accuracy for High- Mass Ions: Space-Charge Effects in MALDI FT-ICR. *Anal. Chem.* **1999**, *71* (3), 624−632.

(18) Bruce, J. E.; Anderson, G. A.; Brands, M. D.; Pasa-Tolic, L.; Smith, R. D. Obtaining More Accurate Fourier Transform Ion Cyclotron Resonance Mass Measurements without Internal Standards Using Multiply Charged Ions. *J. Am. Soc. Mass Spectrom.* **2000**, *11* (5), 416−421.

(19) Williams, D. K.; Muddiman, D. C. Parts-Per-Billion Mass Measurement Accuracy Achieved through the Combination of Multiple Linear Regression and Automatic Gain Control in a Fourier Transform Ion Cyclotron Resonance Mass Spectrometer. *Anal. Chem.* **2007**, *79* (13), 5058−5063.
(20) Wu, S.; Kaiser, N. K.; Meng, D.; Anderson, G. A.; Zhang, K.; Bruce, J. E. Increased Protein Identification Capabilities through Novel Tandem MS Calibration Strategies. *J. Proteome Res.* **2005**, *4* (4), 1434−1441.
(21) Savory, J. J.; Kaiser, N. K.; McKenna, A. M.; Xian, F.; Blakney, G. T.; Rodgers, R. P.; Hendrickson, C. L.; Marshall, A. G. Parts-Per-Billion Fourier Transform Ion Cyclotron Resonance Mass Measurement Accuracy with a "Walking". *Calibration Equation.* **2011**, *83* (5), 1732−1736.
(22) Rost, H. L; Sachsenberg, T.; Aiche, S.; Bielow, C.; Weisser, H.; Aicheler, F.; Andreotti, S.; Ehrlich, H.-C.; Gutenbrunner, P.; Kenar, E.; Liang, X.; Nahnsen, S.; Nilse, L.; Pfeuffer, J.; Rosenberger, G.; Rurik, M.; Schmitt, U.; Veit, J.; Walzer, M.; Wojnar, D.; Wolski, W. E; Schilling, O.; Choudhary, J. S; Malmstrom, L.; Aebersold, R.; Reinert, K.; Kohlbacher, O. OpenMS: A Flexible Open-Source Software Platform for Mass Spectrometry Data Analysis. *Nat. Methods* **2016**, *13* (9), 741−748.
(23) Hendrickson, C. L.; Quinn, J. P.; Kaiser, N. K.; Smith, D. F.; Blakney, G. T.; Chen, T.; Marshall, A. G.; Weisbrod, C. R.; Beu, S. C. 21 T Fourier Transform Ion Cyclotron Resonance Mass Spectrometer: A National Resource for Ultrahigh Resolution Mass Analysis. *J. Am. Soc. Mass Spectrom.* **2015**, *26* (9), 1626−1632.
(24) Gorshkov, M. V.; Good, D. M.; Lyutvinskiy, Y.; Yang, H.; Zubarev, R. A. Calibration Function for the Orbitrap FTMS Accounting for the Space Charge Effect. *J. Am. Soc. Mass Spectrom.* **2010**, *21* (11), 1846−1851.
(25) Mikawy, N. N.; Rojas Ramírez, C.; DeFiglia, S. A.; Szot, C. W.; Le, J.; Lantz, C.; Wei, B.; Zenaidee, M. A.; Blakney, G. T.; Nesvizhskii, A. I.; Loo, J. A.; Ruotolo, B. T.; Shabanowitz, J.; Anderson, L. C.; Håkansson, K. Are Internal Fragments Observable in Electron Based Top-down Mass Spectrometry? *Molecular & cellular proteomics: MCP* **2024**, *23* (9), No. 100814.
(26) Beckman, J. S.; Voinov, V. G.; Hare, M.; Sturgeon, D.; Vasil'ev, Y.; Oppenheimer, D.; Shaw, J. B.; Wu, S.; Glaskin, R.; Klein, C.; Schwarzer, C.; Stafford, G. Improved Protein and PTM Characterization with a Practical Electron-Based Fragmentation on Q-TOF Instruments. *J. Am. Soc. Mass Spectrom.* **2021**, *32* (8), 2081−2091.


https://doi.org/10.1021/acs.analchem.5c06165
*Anal. Chem.* 2026, 98, 1839−1845