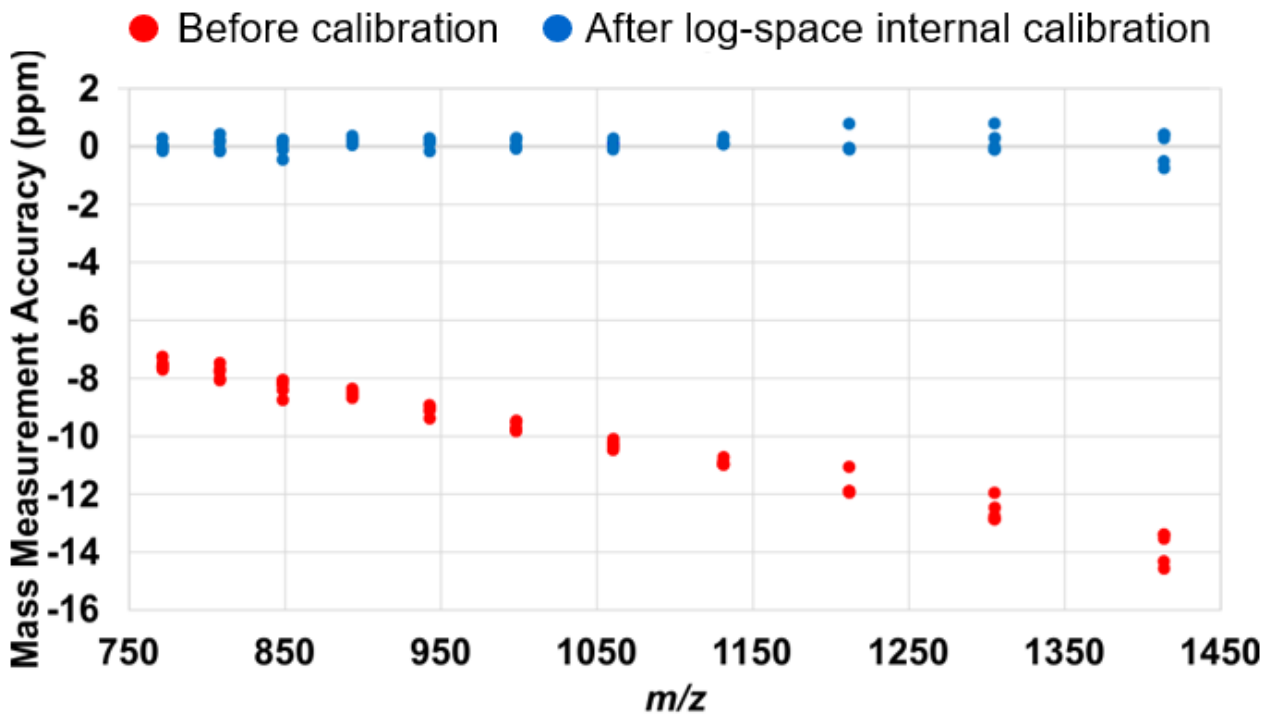
Connectivity-Driven Sequencing of Intact Proteins

Lissa C. Anderson[1,2], Nathan K. Kaiser[1], Krishna Saketh Kamadana[3], Xian Mallory[3]
1. National High Magnetic Field Laboratory, FSU; 2. Department of Chemistry and Biochemistry, FSU; 3. Department of Computer Science, FSU
Funding Grants: K. M. Amm (NSF DMR-2128556); Xian Mallory (NSF CCF-2523717)

Accurate determination of intact protein mass remains a central challenge in top-down proteomics, particularly when isotope patterns are distorted or monoisotopic peaks are weak or absent. Many workflows rely on averagine-based models to estimate monoisotopic masses, an approximation that can introduce systematic errors and propagate uncertainty into downstream protein identification and sequencing.

Here we introduce a framework that bypasses averagine assumptions by operating directly in natural log–transformed mass-to-charge space, where charge-state spacing is mass-invariant. This intrinsic property enables internal calibration of Fourier-transform mass spectra without external calibrants or prior knowledge of protein composition. Using tandem mass spectrometry data acquired on a 21 tesla Fourier transform ion cyclotron resonance (FT-ICR) instrument at the National High Magnetic Field Laboratory and on Orbitrap analyzers, we demonstrate correction of space-charge–induced mass errors and construction of mass-difference networks directly from measured spectra.

By shifting from monoisotopic mass estimation to connectivity-driven inference, this approach enables database-independent *de novo* sequencing of intact proteins and discrimination of near-isobaric residues from single-scan spectra. Because spectra are represented as connectivity-based networks rather than fixed mass assignments, the framework naturally lends itself to future artificial-intelligence and machine-learning approaches for automated interpretation of complex top-down mass spectrometry data.



Internal recalibration of intact protein mass spectra using mass-invariant charge-state spacing in natural log–transformed mass-to-charge space. Correcting space-charge–induced mass errors provides the accurate, consistent mass relationships required for connectivity-driven sequencing of intact proteins.

Facilities and instrumentation used: Ion Cyclotron Resonance Facility, 21 tesla Fourier-transform ion cyclotron resonance mass spectrometer

Citation: Anderson, L.C.; Kaiser, N.K.; Kamadana, K.S.; Mallory, X., Mass-Invariant Natural Log-Transformed Mass Spectra Enable Internal Calibration and De Novo Sequencing of Intact Proteins, Analytical Chemistry, 98 (3), 1839-1845 (2026) doi.org/10.1021/acs.analchem.5c06165