analytical chemistry

pubs.acs.org/ac Article

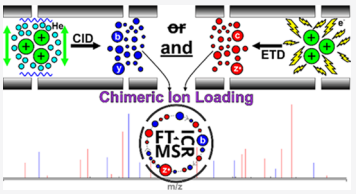
# Increased Single-Spectrum Top-Down Protein Sequence Coverage in Trapping Mass Spectrometers with Chimeric Ion Loading

Chad R. Weisbrod,* Lissa C. Anderson, Joseph B. Greer, Caroline J. DeHart, and Christopher L. Hendrickson

Cite This: *Anal. Chem.* 2020, 92, 12193−12200 Read Online

ACCESS | Metrics & More | Article Recommendations

**ABSTRACT:** Fourier transform mass spectrometers routinely provide high mass resolution, mass measurement accuracy, and mass spectral dynamic range. In this work, we utilize 21 T Fourier transform ion cyclotron resonance (FT-ICR) to analyze product ions derived from the application of multiple dissociation techniques and/or multiple precursor ions within a single transient acquisition. This ion loading technique, which we call, "chimeric ion loading", saves valuable acquisition time, decreases sample consumption, and improves top-down protein sequence coverage. In the analysis of MCF7 cell lysate, we show collision-induced dissociation (CID) and electron-transfer dissociation (ETD) on each precursor on a liquid chromatography-mass spectrometry (LC-MS) timescale and improve mean sequence coverage dramatically (CID-only 15% vs chimeric 33%), even during discovery-based acquisition. This approach can also be utilized to multiplex the acquisition of product ion spectra of multiple charge states from a single protein precursor or multiple ETD/proton-transfer reactions (PTR) reaction periods. The analytical utility of chimeric ion loading is demonstrated for top-down proteomics, but it is also likely to be impactful for tandem mass spectrometry applications in other areas.






## INTRODUCTION

Despite advances in mass spectrometry instrumentation,[1−4] fragmentation techniques,[5−9] and data analysis software,[10−12] intact protein (or top-down) analysis suffers from severe analytical limitations compared to peptide mass analysis. Higher sequence coverage across more proteoforms of larger molecular weight (MW) is required to eliminate current analytical "blind-spots" associated with the approach. To date, bottom-up methods have achieved a higher technological maturity than top-down methodologies.[13] For example, single-shot whole proteome analyses are now capable of identifying 53 000 peptides within a single reversed-phase liquid chromatography with tandem mass spectrometry (LC-MS/MS) experiment.[14] As a result, broader adoption of top-down methods among researchers has been fairly low. However, for select applications, top-down analyses have emerged as the gold standard for difficult tasks such as profiling histone modifications,[15] hemoglobin typing,[16] and rapid bacterial identification.[17,18] Bottom-up analyses remain incapable of distinguishing the combinatoric space of post-translational modifications[19] (PTMs) or full complexity of the proteoforms present within a given sample.[10,20] In fact, when performing bottom-up proteomics, most practitioners have adopted an Occam's razor-like interpretation (principle of maximum parsimony)[20] of the data, such that the fewest number of peptides and protein groups are reported. However, PTMs within the sample may actually be distributed in a far more complex manner as dictated by the biological system.[21] Under the right conditions, top-down analysis can provide an unbiased view of the modifications present and precisely localize them to specific residues along the polypeptide backbone.[22,23] To exercise this clear strength of top-down analysis, high sequence coverage must be obtained from the tandem MS of as many proteoforms as possible, two requirements that are fundamentally at odds with the discovery-based, top-down experimental methodology.

To date, experiments with the highest resulting numbers of proteoform identifications have been made using collision-based activation methods (CID, CAD, and HCD, all of which we refer to here as "CID"), which produce modest sequence coverage.[24−26] The slow heating fragmentation mechanism of CID promotes bond cleavage at the most thermodynamically favored sites along the amide backbone of the protein.[27,28] The resultant spectrum is sparse, with high signal-to-noise (S/N) ratios for relatively few statistically favored, sequence informative *b*/*y* fragment ions. By contrast, electron- and UV photon-based techniques provide access to additional fragmentation pathways and cleavage sites,[29] potentially providing much deeper sequence coverage.[7] However, these

**Received:** March 10, 2020
**Accepted:** August 19, 2020
**Published:** August 19, 2020




https://dx.doi.org/10.1021/acs.analchem.0c01064
*Anal. Chem.* 2020, 92, 12193−12200

additional possible fragment ion types potentially complicate identification. In a recent report, it was found that consideration of up to four independent fragment ion series provides optimal identification.[30] Specifically, with regard to electron-transfer dissociation (ETD),[31] the product ions generated are often complementary to those produced by CID. Moreover, an emerging trend among state-of-the-art top-down analyses is the approach of performing multiple analyses on a given sample with a variety of activation types for tandem mass spectrometry and collating the results before or after fragment ion identification for improved sequence coverage.[32] The aforementioned approach permits the desired improvement in protein sequence coverage; however, it also increases both instrument analysis time and the amount of the sample required.

A variety of hybrid activation schemes have been developed and applied to analyses of intact proteins and peptides. More extensive fragmentation is achieved with EThcD,[33] ETciD, and AI-ETD[34,35] compared to either component fragmentation technique performed on its own. However, these techniques often convert first-generation fragments into subsequent product ion generations (including internal) fragments. For example, conversion of a $c/z\bullet$-type fragment to a $b/y$-type fragment in the cases of both EThcD and ETciD. The result is an inability to observe certain fragments, as well as a compromised detection sensitivity between the precursor fragment and the newly created product fragments. There is little opportunity for control over the size/length of the resultant fragment ions from this process, resulting in a fixed ratio of the fragment ion type and cleavage sites. AI-ETD has shown great promise in improving sequence coverage obtained from top-down sequencing by ETD for lower protein charge states. However, supplemental activation during the ETD reaction with IRMPD produces many new fragment ion pathways ($-H_2O$, $-NH_3$, hydrogen rearrangements, etc.). In addition, the IRMPD power must be carefully metered to achieve the desired release of the ETnoD product ions without overheating to produce $b/y$ and internal fragments.

Here, we describe "chimeric ion loading", a unique approach that combines product ion populations derived from independently isolated and activated precursor ion populations within a single mass spectral acquisition. Importantly, this flexible approach permits formation of fragment ion admixtures from virtually any dissociation method and any set of ion activation parameters. We leverage the ultrahigh magnetic field space charge-capacity available at 21 T and the flexibility of an external multipole storage device (MSD) to allow multiple fills of fragment ions derived from various activation conditions within a single set of ion manipulation steps prior to transient acquisition. In one example, we demonstrate that chimeric ion loading of ETD and CID fragments from all selected precursor ions improves sequence coverage across the board in an online LC-MS/MS analysis of the MCF7 cell lysate (15% CID-only; 23% ETD-only; to 33% chimeric). We also show benefit from complementary $b/y$ ion formation by fragmentation of multiple charge states from the same protein precursor.

## EXPERIMENTAL SECTION

**Cell Culture.** MCF7 human breast adenocarcinoma cells were grown at 37 °C and 5% $CO_2$ in RPMI 1640 growth medium (Corning-Mediatech, Manassas, VA) supplemented with 10% fetal bovine serum (Sigma-Aldrich Corp., St. Louis, MO) and 0.5% penicillin/streptomycin (Corning-Mediatech). The cells were washed twice on-dish in 10 mL of ice-cold 1× Dulbecco's phosphate-buffered saline (Life Technologies) to remove residual medium and serum proteins, harvested by scraping, and pelleted by centrifugation at 200*g* for 10 min at 4 °C. The cell pellets, each comprising 24 confluent 10 cm dishes (approximately 2E8 cells), were stored dry at −80 °C prior to lysis and protein quantitation.

**Sample Preparation.** The cell pellets were thawed on ice and resuspended in 24 mL of 20 mM Tris, pH 7.5, containing 100 mM sodium chloride, 1% (w/v) *N*-lauroylsarcosine, and 1× final concentration of HALT protease/phosphatase inhibitor cocktail (EDTA-free) (Thermo Fisher Scientific, San Jose, CA). The lysates were incubated on ice for 20 min. Magnesium chloride was added to a final concentration of 1 mM, followed by 750 units of benzonase nuclease (Sigma-Aldrich Corp.). The lysates were then incubated at 25 °C for 20 min, chilled on ice, and centrifuged at 16 800*g* for 15 min at 4 °C to pellet cellular debris. Total lysate protein concentration was determined by microplate BCA assay (Thermo Fisher Scientific). Following protein quantitation, 300 μg of protein from each lysate was precipitated in acetone, incubated at −80 °C overnight, and then washed in ice-cold acetone by centrifugation at 16 800*g* for 10 min at 4 °C. The pellets were then reconstituted in 150 μL of 1% (w/v) sodium dodecyl sulfate (SDS) containing 50 μM dithiothreitol and 1× Tris−acetate sample buffer (Expedeon Inc., San Diego, CA). The samples were incubated at 95 °C for 5 min and centrifuged at 16 800*g* for 10 min at room temperature to pellet any remaining debris. The supernatants were loaded into an 8% acrylamide gel-eluted liquid-fraction entrapment electrophoresis (GELFrEE) cartridge and resolved into 12 fractions according to the manufacturer's protocol (GELFrEE 8100 Fractionation System, Expedeon Inc.). Aliquots (10 μL) from each GELFrEE fraction were resolved by SDS-polyacrylamide gel electrophoresis and visualized by silver nitrate stain[36] to evaluate total protein content. The eluted fractions were stored at −80 °C. Directly prior to LC-MS/MS analysis, the fractions were concentrated, desalted, and cleaned of SDS by precipitation in a mixture of methanol, chloroform, and water.[37] After two additional methanol washes, the pellets were immediately reconstituted in 50 μL of ice-cold HPLC Solvent A (0.3% formic acid and 5% acetonitrile (v/v) in water; all mass spectrometry grade) with gentle pipetting.

**Instrumentation.** A custom hybrid dual-cell linear radio-frequency (RF) ion trap 21 T Fourier transform-ion cyclotron resonance mass spectrometer (FT-ICR MS) (Figure 1A) was used to collect all data.[4] The linear RF ion trap is a highly modified Velos Pro[2] ion trap mass spectrometer (ITMS) (Thermo Fisher Scientific, San Jose, CA). The ITMS is equipped with a commercial Orbitrap Fusion[3] atmospheric pressure ionization inlet/frontend electron-transfer dissociation (FETD) reagent ion source (Thermo Fisher Scientific). The glow discharge FETD source enables generation of reagent ions for both ETD and proton-transfer reactions (PTR).[9] An external multipole storage device (MSD),[38] situated between the ITMS and the FT-ICR, accepts multiple ion fills prior to FT-ICR detection. The MSD imposes an axial electric field for sequestration of ions near the exit of the MSD. The ions are delivered to the ICR cell by mass-dependent ejection with an auxillary RF pseudopotential.[39] Many changes to the underlying ion trap control language (ITCL) code were required to enable the instrument to perform chimeric ion loading, which allows looped scan events while replacing the


https://dx.doi.org/10.1021/acs.analchem.0c01064
Anal. Chem. 2020, 92, 12193−12200

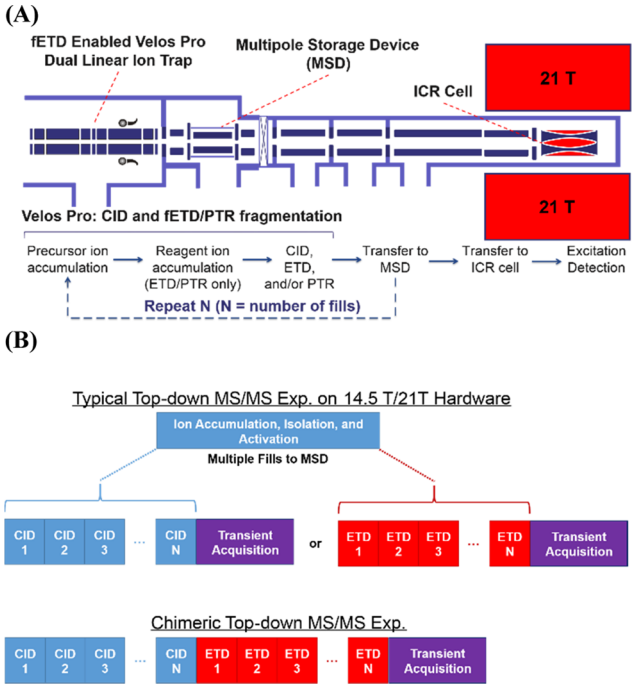



**Figure 1.** (A) Schematic diagram of the 21 T FT-ICR MS system. The reagent ion source-equipped dual linear ion trap enables flexible and rapid ion manipulation and fragmentation. The multipole storage device accumulates large ion populations for analysis. The custom dynamically harmonized cell produces highly ideal fields, while the magnet allows for high charge capacity. (B) (top) Typical individual ETD or CID activation experiments in which $N$ ion fills are conducted prior to transient acquisition. In contrast (bottom), chimeric ion loading permits multiple ion activation approaches prior to transient acquisition.

underlying properties within each scan event (i.e., ion activation type, reaction period, and precursor $m/z$) with those desired at each ion manipulation step. Several additional modifications were made to allow the control of chimeric ion loading directly through the Xcalibur method editor. These modifications were performed by disabling the neutral loss scan type within ITCL and replacing underlying variables with chimeric ion loading variables, which permits access from the existing GUI interface with an indexed map of parameters.

**Liquid Chromatography.** Reconstituted protein from GELFrEE fraction 1 was optionally diluted up to 5-fold in ice-cold HPLC solvent A (based on silver nitrate stain intensity) and analyzed by reverse-phase LC-MS/MS. For each injection, 5 $\mu$L was loaded onto an in-house-fabricated 360 $\mu$m o.d. × 150 $\mu$m i.d. fused-silica microcapillary trap column packed 2−3 cm with PLRP-S resin (5 $\mu$m particle, 1000 Å pore, Agilent Technologies, Palo Alto, CA). The nano-HPLC system (ACQUITY M-Class, Waters, Milford, MA) was operated at 2.5 $\mu$L/min during trapping and washed with 95% A for 10 min. Analytical columns were packed in-house using 360 $\mu$m o.d. × 75 $\mu$m i.d. fused-silica microcapillary with PLRP-S resin (same as corresponding trap columns) to 15 cm length. Elution gradients were applied using a 0.3 $\mu$L/min flow rate with a gradient of 5−20% B in 5 min, 20−35% B in 20 min, 35−60% B in 75 min, 60−75% B in 15 min, and 75−95% B in 5 min (120 min total length). Following separation, the samples were directly ionized by nanoelectrospray ionization using a 15 $\mu$m fused-silica PicoTip (New Objective, Woburn, MA) emitter packed with 2−3 mm of the same PLRP-S resin as previously described. A UWPR nanospray source was utilized for application of the ionization voltage, fixturing the column, and providing fine adjustment for the ESI emitter (http://proteomicsresource.washington.edu/protocols05/nsisource.php).

**Data Handling.** Data were acquired using 1.524 s (precursor ion spectra) and 0.762 s (product ion spectra) transient durations, which correspond to 1.2 M and 600 k resolving power at $m/z$ 200. All data were processed in raw file format (Thermo Fisher Scientific) in reduced profile mode (noise baseline-subtracted). Tandem ms data acquired on fractionated MCF7 proteoforms in LC-MS mode were all single transient spectra (no averaging of scans or microscans). Data were further reduced to monoisotopic decharged masses using xTract. Data were searched using TDPortal (v1.3), a high-throughput top-down proteomics search platform hosted by the National Resource for Translational and Developmental Proteomics (available for academic collaborators here: http://nrtdp.northwestern.edu/tdportal-request/) using a three-pronged search strategy employing tight absolute mass (2.2 Da window), wide absolute mass (200 Da window), and biomarker (10 ppm) searches; all with a product ion mass tolerance of ±10 ppm.[24] Data acquired in direct infusion mode on carbonic anhydrase II is the average of 10 transients. These data were also processed with xTract and fragment ions were identified using Prosight Lite[40] with a product ion mass tolerance of ±10 ppm.

## RESULTS AND DISCUSSION

Recent work has demonstrated that combination of the spectra acquired with multiple ion activation techniques provides deep

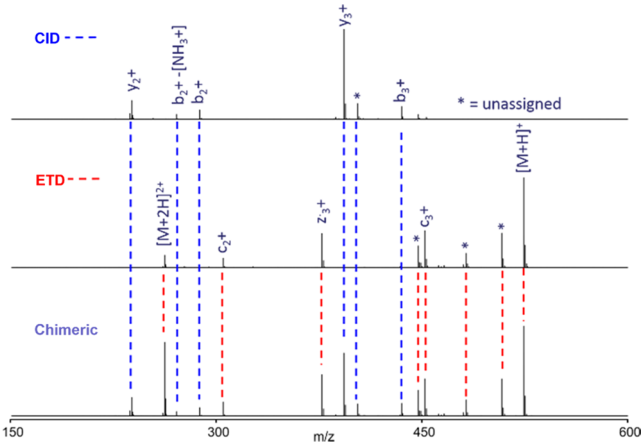



**Figure 2.** Product ion spectra acquired on MRFA $[M + 2H]^{2+}$ with CID (top), ETD (middle), and chimeric (bottom) activation modes. Product ions present within the spectra produced using ETD or CID alone are all present within the spectrum produced using the chimeric activation mode.

sequence coverage for intact proteins throughout a range of molecular weight.[8,41] However, combining data across activation types and LC-MS/MS acquisitions, including hundreds to thousands of transients averaged to produce a single composite spectrum, significantly increases sample consumption and acquisition period. This observation, coupled with the unique performance characteristics of the NHMFL 21 T FT-ICR,[4] provided the underlying motivation for the development of chimeric ion loading. A similar term,


https://dx.doi.org/10.1021/acs.analchem.0c01064
Anal. Chem. 2020, 92, 12193−12200

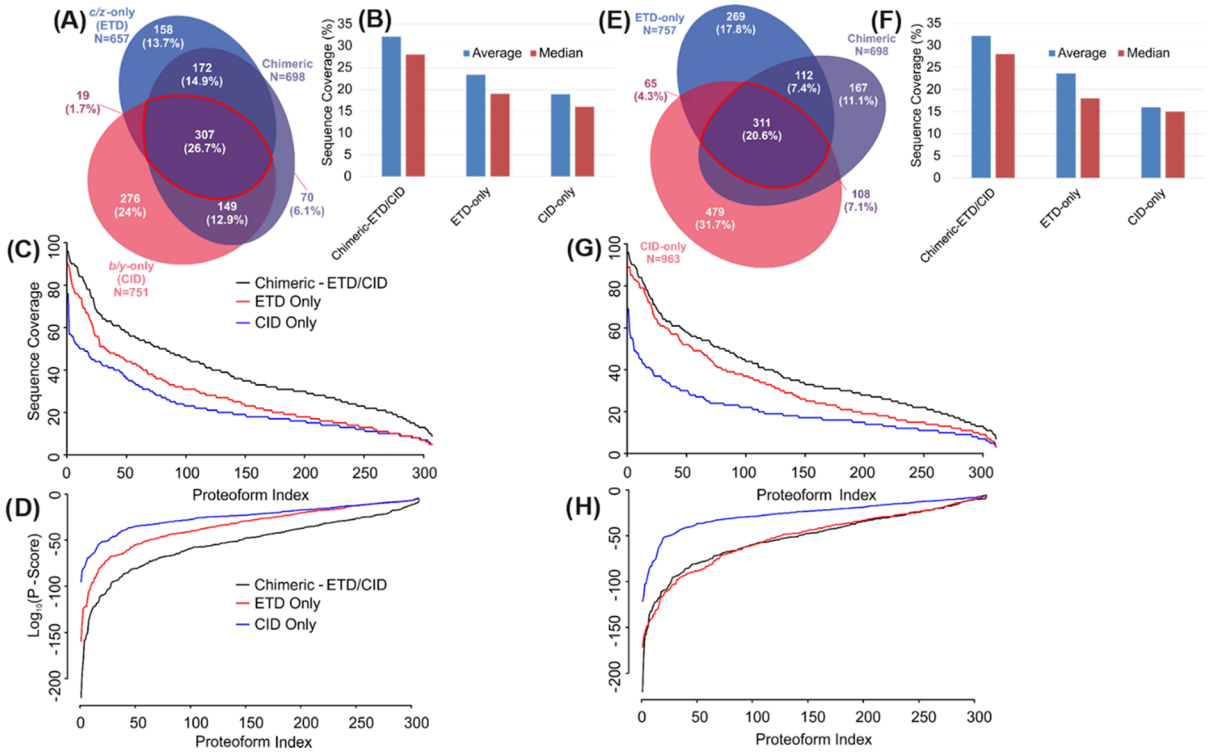



**Figure 3.** Global discovery-based interrogation of intact proteins from the GELFrEE fraction 1 of the MCF7 cells (*H. sapiens*). All LC-MS data were acquired in technical triplicate. The control data (A–D) were generated by searching LC-MS chimeric ion loading runs with TDPortal and querying all spectra for either *b*/*y* (CID), *c*/*z*• (ETD), or *b*/*c*/*y*/*z*• (Chimeric). (A) Euler diagram contains the number of proteoform identifications from each query. (B) Accompanying bar chart shows the mean and median sequence coverage for all identified proteoforms for each query. The commonly identified proteoforms (307) in all three queries outlined in red within the Euler diagram are shown in the plots directly below. (C) Sequence coverage as a function of the proteoform index is displayed for each of the three queries. Sequence coverage for chimeric ion data is improved over data acquired with each individual activation type. (D) Bottom plot shows the P-score for each identification across query type. The data in the neighboring column (E–H) shows the comparison of chimeric ion loading against the optimized LC-MS methods for CID-only or ETD-only experiments.

"chimeric" spectra,[42] is used in mass spectrometry and should not be confused with chimeric ion loading. A chimeric spectrum is the result of an ion isolation experiment in which two or more species remain within the isolation window and the resultant product ion spectrum contains fragment ions from all species present. These spectra provide unique challenges for identification of peptides and proteins because most conventional search engines (Sequest,[43] Mascot,[44] etc.) presume that the tandem ms spectra contain product ions from a single purely isolated precursor. Whether or not chimeric spectra are acquired depends solely on the performance characteristics of the isolation and how closely in $m/z$ two isotopic envelopes overlap. In contrast, chimeric ion loading is the deliberate creation of an admixture of product ions acquired under different acquisition conditions. These conditions could include, but are not limited to, activation type, activation (reaction) period, and precursor charge state. For the purposes of this manuscript, anything described as chimeric refers to spectral acquisition in which chimeric ion loading was used.

Targeted or discovery-based analysis of intact proteins requires appropriate consideration of the acquisition parameters. For sequence analysis at 21 T, typically activation type is selected and the number of ion fills varied for optimized S/N and sequence coverage.[45,46] Scaling the number of ion fills as a function of MW of the protein has been shown to be advantageous.[47] Multiple ion fills are conducted by looping through a set of steps during each fill, as shown in Figure 1A: precursor ion accumulation, isolation, reagent ion accumulation (ETD/PTR only), ion activation, and finally, transfer to a multipole storage device (MSD). This set of steps constitutes a single ion fill and when repeated $N$ times accumulates large ion populations for improved signal-to-noise (S/N). In addition, multiple ion fills permit operation of the ion trap under more ideal conditions for ion manipulation and activation. Each ion fill is regulated by automatic gain control (AGC)[48] within the linear ion trap and each fill meets the desired AGC target for $MS^n$ acquisitions. The cumulative ion target (CIT) is simply the AGC target multiplied by the number of fills (example: $N$ = 10; AGC target = 2E5; CIT = 10 × 2E5 ions = 2E6 ions). This permits the entire loop of $N$ fills to be carried out using a single AGC determination. The underlying assumption is that the composition of the ion beam is not changing rapidly enough to affect the accuracy of the number of charges delivered during each event. This approximation holds true for both direct infusion and LC-MS analysis as long as an excessively large number of ion fills is not selected. In fact, this process has been shown to be very linear over a broad range of CIT's[47] (>1E7 CIT). While not presently implemented, incorporating multiple AGC scans interleaved regularly between ion fills, then utilizing only the most recent AGC determination to inform injection times for subsequent ion fills would be optimal to improve linearity and accuracy. Figure 1B illustrates the typical experimental


https://dx.doi.org/10.1021/acs.analchem.0c01064
Anal. Chem. 2020, 92, 12193–12200

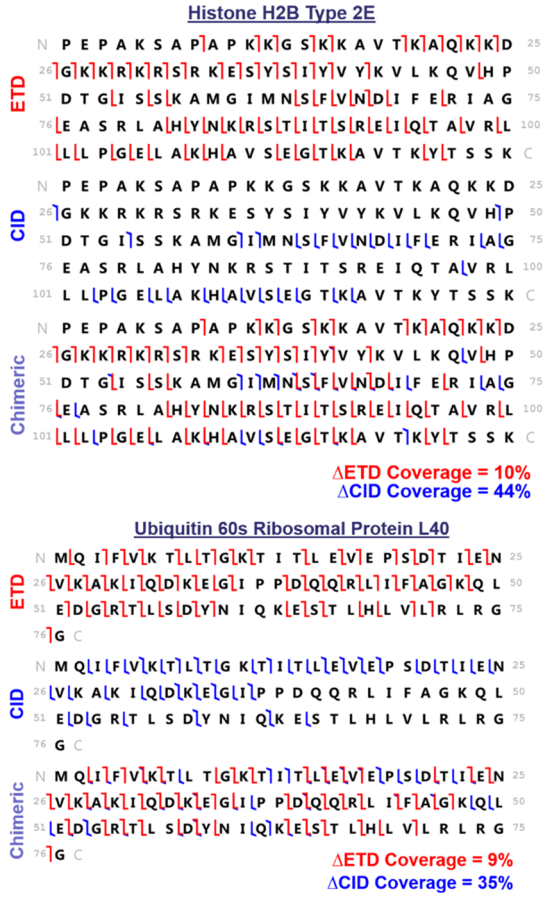



**Figure 4.** Sequence coverage maps for both histone H2B type 2-E (Q16778) and ubiquitin 60s ribosomal protein L40 (P62987) constructed from data acquired using either ETD, CID, or chimeric modes. These sequence coverage maps illustrate the complementarity of CID vs ETD and also reveal that this complementarity is preserved during chimeric ion loading. Improvements in sequence coverage using chimeric ion loading over either ETD or CID are reported.

approach when conducting CID or ETD. Chimeric ion loading allows us to create an admixture of CID and ETD product ions, each with a separate number of ion fills, *N*.

A simple example of a chimeric ETD/CID spectrum is shown in Figure 2. The analyte is the tetramer MRFA $[M + 2H]^{2+}$. The expected number of peptide backbone fragments for any polypeptide is $n - 1$ (where $n$ = AA length). Therefore, not including side chain and water loss peaks, we expect a maximum of 3 $b/y$ pairs (CID) and 3 $c/z\bullet$ pairs (ETD) of the MRFA fragment ions. Using this analyte, the spectral complexity is appropriately constrained such that it is evident that all fragment ions contained within the spectra acquired from each individual activation method (CID or ETD) are contained in the chimeric ion loading spectrum (indicated by the color-coded dashed lines).

Inherent to this approach is the ability to tailor the admixture of ions such that the S/N of fragment ions generated from each condition is optimized. Prior experimental optimization for discovery-based, data-dependent acquisition (DDA) on the 21 T FT-ICR MS with a single activation type per acquisition has informed selection of ETD parameters of 15 fills × 2E5 ions (CIT 3E6) and CID parameters of 2 fills × 5E5 ions (CIT 1E6). These individual targets were adopted for MCF7 cell lysate discovery-based analysis (3E6 CIT-ETD + 1E6 CIT-CID = 4E6 CIT-chimeric). Ion storage capacity and ETD reaction capacity constraints imposed by the linear RF ion trap result in a significant difference in the effective MS/MS duty cycle for ETD vs CID. ETD suffers from a 30% lower MS/MS acquisition rate, and this is reflected in the total proteoform identifications obtained using ETD-only and chimeric ion loading. However, with further development of ETD methodology and optimization, we expect the ETD tandem MS repetition rate to more closely match our CID capabilities. The 21 T FT-ICR at NHMFL is capable of handling >3E7 charges before nonlinearity in signal magnitude is detected.[47] Therefore, operating at 4E6 CIT is well below this practical limit for our hardware. In principle, this experiment could be conducted on an Orbitrap mass analyzer; however, space charge and coalescence within the spectrum could prevent analytical utility even with the highest field Orbitrap (D20). Charge capacity at 21 T is well over an order of magnitude greater than that feasible on commercially available Orbitrap systems.[49,50] A mixture where 75% of the CIT originates from ETD fragment ions and 25% CIT from CID fragments is well-suited to the analysis of proteins from 5 to 30 kDa. Further optimization for application to a broader intact protein MW range or even real-time application of varying the percentage of the CIT between the two activation modes may be advantageous. Finally, while not done in this study, ion accumulation during transient acquisition allows for greater overall instrument acquisition efficiency. Simultaneous acquisition and ion accumulation/manipulation is currently being pursued.

The impact of chimeric ion loading on the analysis of fractionated MCF7 (*Homo sapiens*) lysate with respect to proteoforms identified, sequence coverage, and P-score is shown in Figure 3A–H. A "control" database search was performed against the chimeric ion loading runs (technical triplicate) with UniProt *H. sapiens* database (20 155 protein entries, 42 317 isoforms, and 319 264 isoform features queried) containing the theoretical spectra with $b/y$ (CID), $c/z\bullet$, or $b/c/y/z\bullet$ ion series. This strategy was employed to ascertain the effect of total proteoform identifications at a fixed FDR[51] and P-score with the chimeric ion loading data. Performing queries against this database containing only ion series pairs ($b/y$ or $c/z\bullet$) illustrated that chimeric ion loading data did not introduce any artefactual P-score or FDR issues within this informatics platform. In fact, within this data set we identified 657 proteoforms by $b/y$ query, 751 proteoforms by $c/z\bullet$ query, and 698 proteoforms by $b/c/y/z\bullet$ query (Figure 3A). It was expected that the total number of proteoforms identified would be similar for all queries provided the S/N of product ions resulting from both CID and ETD were similar. The average and median sequence coverage is reported in Figure 3B. These statistics are calculated across all proteoforms identified for each query type. Chimeric ion loading achieved the highest median (28%) and average (32%) sequence coverage, which was more than 10% better than single activation data (nonchimeric). Sequence coverage was compared across all commonly identified proteoforms (outlined in red within Figure 3A) and sorted in order of descending sequence coverage (Figure 3C). The commonly identified species were the focus due to the pseudostochastic sampling nature of data-dependent acquisition (DDA). Here, it is evident that combining ions using chimeric ion loading produces significant improvement in sequence coverage for this subset of commonly identified proteoforms. Figure 3D


https://dx.doi.org/10.1021/acs.analchem.0c01064
Anal. Chem. 2020, 92, 12193−12200

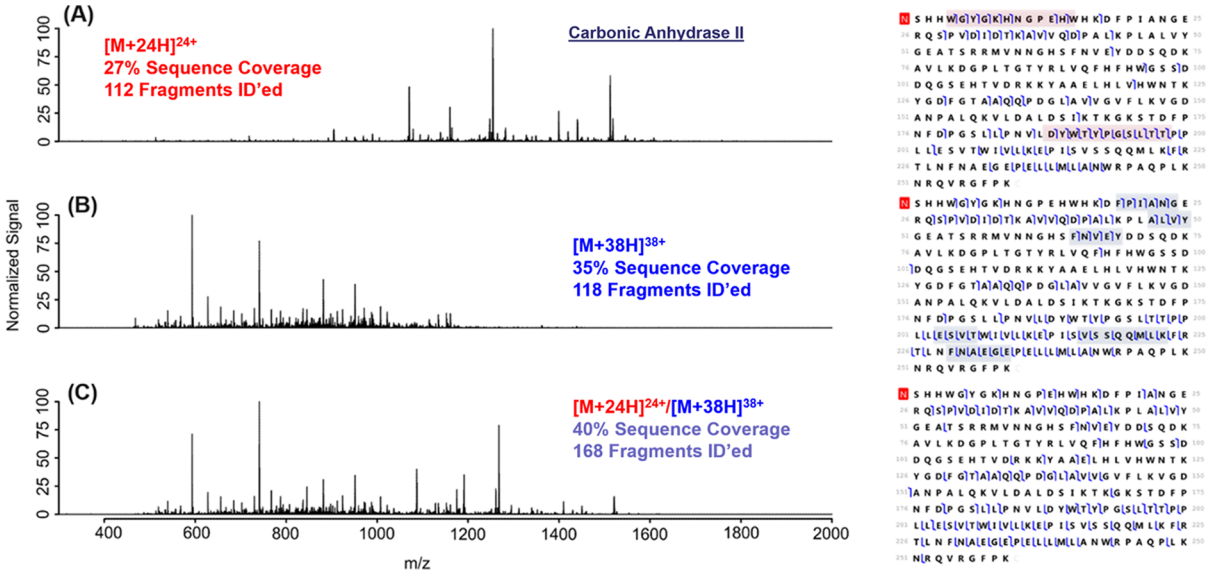



**Figure 5.** Product ion spectra and corresponding sequence coverage maps collected using chimeric ion loading to leverage the propensity for disparate protein charge states to yield complementary *b*/*y* ions (carbonic anhydrase II; *Bos taurus*). The top row (A) includes the product ion spectra (CID) of the $[M + 24H]^{24+}$ with the corresponding sequence coverage map. The same data is shown for the $[M + 38H]^{38+}$ (CID) and the $[M + 24H]^{24+}/[M + 38H]^{38+}$ (Chimeric—CID/CID) on the middle (B) and bottom rows (C), respectively. Within the sequence coverage maps, the regions of the protein which exhibit complementary fragment ions are highlighted.

shows the P-score profile for the commonly identified proteoforms sorted in ascending P-score order. These data show that P-score is improved dramatically when the query is made against all four product ion series present as opposed to omission of a series pair and is consistent with the findings of others.[30] Figure 3E−H displays the results obtained from three sets of optimized triplicate discovery-based data acquisitions: ETD-only (3E6 CIT), CID-only (1E6 CIT), and chimeric ion loading (25% CID/75% ETD, 4E6 CIT). In this comparison, each acquisition was individually optimized to produce the best possible results for each activation type. This is readily apparent in the numbers of total proteoforms identified in each case (757 ETD-only; 963 CID-only; 698 chimeric ion loading, Figure 3E). The cost of performing chimeric ion loading is that additional ion fills are required to attain the desired CIT of 4E6 ions (see above), with on average 100 ms required for ETD and 20−30 ms required for CID ion fills. This reduces the tandem MS acquisition rate and represents the real "cost" of performing chimeric ion loading. However, this cost in time is low compared to the cost of entirely separate LC-MS runs or spectral acquisitions for each activation type. In Figure 3F, the average and median sequence coverages illustrate that chimeric ion loading is the method of choice for obtaining higher overall sequence coverage without sacrificing a significant number of total identifications compared with ETD-only. In Figure 3G,H, the sequence coverage is improved, as observed with the control queries, and the P-scores were unchanged using chimeric ion loading for all commonly identified species. Additionally, comparing the control database search (Figure 3A−D) vs the actual data (Figure 3E−H) illustrates that performing chimeric ion loading does not induce a deleterious effect on the sequence coverage of the individual activation type ion populations (ETD or CID) even though ions were mixed, stored, and analyzed together.

The improvement in sequence coverage for two MCF7 proteins is illustrated in Figure 4. These example proteins were selected because they represent average gains in sequence coverage obtained by chimeric ion loading. Fragment ion spectra used for proteoform identification and sequence coverage maps shown in Figure 4 were obtained with single transients, during global discovery-based top-down proteomics experiments. As a result, there is some variance in the observed fragment ions between chimeric ion loading and ETD-only or CID-only due to small differences in the timing of automated MS/MS spectral acquisition during the LC elution. In the case of histone H2B type 2-E, the ETD-only tandem mass spectrum produced 57% sequence coverage. The CID-only tandem mass spectrum yielded 23% sequence coverage. Therefore, chimeric ion loading produced improvements of 44% over CID-only and 10% over ETD-only generated spectra (67% sequence coverage under chimeric ion loading conditions). Similar improvements were also observed for ubiquitin 60s ribosomal protein L40. Such complementarity, or the ability for a given fragmentation technique to generate product ions that are unique, is truly what drives the improvement in coverage when operating in chimeric ion loading mode. We expect further benefits from inclusion of additional activation approaches (i.e., chimeric ion loading with CID/ETD/UVPD). Currently, our hardware supports CID, HCD, ETD, PTR, UVPD (193 nm), and IRMPD. Activation types that yield high numbers of unique fragments relatively rapidly will be the focus of future work.

Another viable application of chimeric ion loading is interrogation of multiple disparate charge states of the same protein using a single activation mode. CID typically generates the most thermodynamically favored bond cleavages along the peptide backbone. However, protein secondary and tertiary structures also influence the likelihood of amide bond cleavage and thus provide the opportunity for exploitation.[52] Protein ions are known to adopt an unfolded gas-phase conformation when they are highly saturated with charge carriers.[53,54] A compact, more folded conformation is adopted as charge state decreases. In Figure 5A,B, the propensity for disparate charge states of the same protein precursor ions to yield unique


https://dx.doi.org/10.1021/acs.analchem.0c01064
Anal. Chem. 2020, 92, 12193−12200

fragment ions (carbonic anhydrase II; 24+ vs 38+) is exploited with chimeric ion loading. Instead of mixing product ion fills from separate activation techniques, 10 CID product ion fills from the $[M + 24H]^{24+}$ were mixed with 10 CID product ion fills from the $[M + 38H]^{38+}$. Each precursor ion was isolated within a 5 Da wide window. Figure 5A shows the CID spectrum generated from the $[M + 24H]^{24+}$ of carbonic anhydrase II (CAII), which yields 27% sequence coverage (112 fragments identified). Figure 5B shows the CID spectrum generated from the $[M + 38H]^{38+}$ of CAII with sequence coverage of 35% (118 fragments identified). The chimeric ion loading spectrum containing 10 ion fills each of $[M + 24H]^{24+}$ and $[M + 38H]^{38+}$ produced 40% total sequence coverage (168 fragments identified). The accompanying fragment ion maps show the regions where unique fragment ions contribute to the overall gain of 5% sequence coverage from the admixture of the ions in the chimeric spectrum. This approach was conducted in direct infusion mode and represented a proof-of-principle worthy of further study. Optimization of the degree of separation in charge states selected, relative number of ion fills, and automated selection of each charge state remain areas in which further impact could be achieved, especially with respect to intact protein LC-MS experiments.

Chimeric ion loading is defined by the ability to create admixtures of fragment ions of the same species, albeit with separate acquisition characteristics. We intend to explore this space further to "design" ion populations containing large, medium, and small MW fragment ions from a single protein precursor ion via ETD such that sequence coverage is fully optimized within a single spectrum. In addition, with the use of PTR, we can deliberately "program" our spectrum such that the ion fills occupy certain analytical $m/z$ regions within a single spectral acquisition based on the elapsed reaction period. In principle, chimeric ion loading can be applied to lower field ICR and Orbitrap but will be more constrained by limited ion capacity.

## AUTHOR INFORMATION


### Corresponding Author

**Chad R. Weisbrod** − *Ion Cyclotron Resonance Program, National High Magnetic Field Laboratory, Tallahassee, Florida 32310, United States;* orcid.org/0000-0001-5324-4525; Email: weisbrod@magnet.fsu.edu

### Authors

**Lissa C. Anderson** − *Ion Cyclotron Resonance Program, National High Magnetic Field Laboratory, Tallahassee, Florida 32310, United States;* orcid.org/0000-0001-8633-0251

**Joseph B. Greer** − *National Resource for Translational and Developmental Proteomics, Northwestern University, Evanston, Illinois 60208, United States*

**Caroline J. DeHart** − *NCI RAS Initiative, Frederick National Laboratory for Cancer Research, Frederick, Maryland 21702, United States*

**Christopher L. Hendrickson** − *Ion Cyclotron Resonance Program, National High Magnetic Field Laboratory, Tallahassee, Florida 32310, United States; Department of Chemistry and Biochemistry, Florida State University, Tallahassee, Florida 32306, United States*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.analchem.0c01064


### Author Contributions

The manuscript was written through the contributions of all authors. All authors have given approval to the final version of the manuscript.

### Notes

The authors declare no competing financial interest.

Raw data and TDportal results available upon request to the corresponding author. These data exceed the supplement file limit for the journal.

## ACKNOWLEDGMENTS


The authors thank Prof. Galit Lahav of Harvard Medical School for the kind gift of the MCF7 human breast cancer cells. A portion of this work was performed at the Ion Cyclotron Resonance User Facility at the National High Magnetic Field Laboratory, which is supported by the National Science Foundation Division of Materials Research and Division of Chemistry through Cooperative Agreement No. DMR-1644779 and the State of Florida. TDPortal is a publicly available resource provided by the National Resource for Translational Developmental Proteomics based at Northwestern University, which is supported by a grant from the National Institute of General Medical Sciences, National Institutes of Health P41 GM108569 (Neil L. Kelleher, P.I.) with additional support provided by the Sherman Fairchild Foundation. The authors thank John P. Quinn for the design and maintenance of the FT-ICR MS instruments. The authors cordially acknowledge John E. P. Syka and Micheal W. Senko for their helpful discussion.


## REFERENCES


(1) Hu, Q.; Noll, R. J.; Li, H.; Makarov, A.; Hardman, M.; Graham Cooks, R. *J. Mass Spectrom.* **2005**, *40*, 430−443.

(2) Pekar Second, T.; Blethrow, J. D.; Schwartz, J. C.; Merrihew, G. E.; MacCoss, M. J.; Swaney, D. L.; Russell, J. D.; Coon, J. J.; Zabrouskov, V. *Anal. Chem.* **2009**, *81*, 7757−7765.

(3) Senko, M. W.; Remes, P. M.; Canterbury, J. D.; Mathur, R.; Song, Q.; Eliuk, S. M.; Mullen, C.; Earley, L.; Hardman, M.; Blethrow, J. D.; Bui, H.; Specht, A.; Lange, O.; Denisov, E.; Makarov, A.; Horning, S.; Zabrouskov, V. *Anal. Chem.* **2013**, *85*, 11710−11714.

(4) 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. *J. Am. Soc. Mass Spectrom.* **2015**, *26*, 1626−1632.

(5) Riley, N. M.; Mullen, C.; Weisbrod, C. R.; Sharma, S.; Senko, M. W.; Zabrouskov, V.; Westphall, M. S.; Syka, J. E.; Coon, J. *J. Am. Soc. Mass Spectrom.* **2016**, *27*, 520−531.

(6) Riley, N. M.; Sikora, J. W.; Seckler, H. S.; Greer, J. B.; Fellers, R. T.; LeDuc, R. D.; Westphall, M. S.; Thomas, P. M.; Kelleher, N. L.; Coon, J. J. *Anal. Chem.* **2018**, *90*, 8553−8560.

(7) Shaw, J. B.; Li, W.; Holden, D. D.; Zhang, Y.; Griep-Raming, J.; Fellers, R. T.; Early, B. P.; Thomas, P. M.; Kelleher, N. L.; Brodbelt, J. S. *J. Am. Chem. Soc.* **2013**, *135*, 12646−12651.

(8) Shaw, J. B.; Malhan, N.; Vasil'ev, Y. V.; Lopez, N. I.; Makarov, A.; Beckman, J. S.; Voinov, V. G. *Anal. Chem.* **2018**, *90*, 10819−10827.

(9) Earley, L.; Anderson, L. C.; Bai, D. L.; Mullen, C.; Syka, J. E.; English, A. M.; Dunyach, J. J.; Stafford, G. C.; Shabanowitz, J.; Hunt, D. F.; Compton, P. D. *Anal. Chem.* **2013**, *85*, 8385−8390.

(10) Schaffer, L. V.; Millikin, R. J.; Miller, R. M.; Anderson, L. C.; Fellers, R. T.; Ge, Y.; Kelleher, N. L.; LeDuc, R. D.; Liu, X.; Payne, S. H.; Sun, L.; Thomas, P. M.; Tucholski, T.; Wang, Z.; Wu, S.; Wu, Z.; Yu, D.; Shortreed, M. R.; Smith, L. M. *Proteomics* **2019**, *19*, No. 1800361.

(11) Toby, T. K.; Fornelli, L.; Srzentic, K.; DeHart, C. J.; Levitsky, J.; Friedewald, J.; Kelleher, N. L. *Nat. Protoc.* **2019**, *14*, 119−152.

(12) Park, J.; Piehowski, P. D.; Wilkins, C.; Zhou, M.; Mendoza, J.; Fujimoto, G. M.; Gibbons, B. C.; Shaw, J. B.; Shen, Y.; Shukla, A. K.; Moore, R. J.; Liu, T.; Petyuk, V. A.; Tolic, N.; Pasa-Tolic, L.; Smith, R. D.; Payne, S. H.; Kim, S. *Nat. Methods* **2017**, *14*, 909−914.
(13) Chait, B. T. *Science* **2006**, *314*, 65−66.
(14) Hebert, A. S.; Thoing, C.; Riley, N. M.; Kwiecien, N. W.; Shiskova, E.; Huguet, R.; Cardasis, H. L.; Kuehn, A.; Eliuk, S.; Zabrouskov, V.; Westphall, M. S.; McAlister, G. C.; Coon, J. J. *Anal. Chem.* **2018**, *90*, 2333−2340.
(15) Zheng, Y.; Fornelli, L.; Compton, P. D.; Sharma, S.; Canterbury, J.; Mullen, C.; Zabrouskov, V.; Fellers, R. T.; Thomas, P. M.; Licht, J. D.; Senko, M. W.; Kelleher, N. L. *Mol. Cell. Proteomics* **2016**, *15*, 776−790.
(16) He, L.; Rockwood, A. L.; Agarwal, A. M.; Anderson, L. C.; Weisbrod, C. R.; Hendrickson, C. L.; Marshall, A. G. *Clin. Chem.* **2019**, *65*, 986−994.
(17) Arnold, R. J.; Reilly, J. P. *Rapid Commun. Mass Spectrom.* **1998**, *12*, 630−636.
(18) Cain, T. C.; Lubman, D. M.; Weber, W. J., Jr.; Vertes, A. *Rapid Commun. Mass Spectrom.* **1994**, *8*, 1026−1030.
(19) Aebersold, R.; Agar, J. N.; Amster, I. J.; Baker, M. S.; Bertozzi, C. R.; Boja, E. S.; Costello, C. E.; Cravatt, B. F.; Fenselau, C.; Garcia, B. A.; Ge, Y.; Gunawardena, J.; Hendrickson, R. C.; Hergenrother, P. J.; Huber, C. G.; Ivanov, A. R.; Jensen, O. N.; Jewett, M. C.; Kelleher, N. L.; Kiessling, L. L.; Krogan, N. J.; Larsen, M. R.; Loo, J. A.; Ogorzalek Loo, R. R.; Lundberg, E.; MacCoss, M. J.; Mallick, P.; Mootha, V. K.; Mrksich, M.; Muir, T. W.; Patrie, S. M.; Pesavento, J. J.; Pitteri, S. J.; Rodriguez, H.; Saghatelian, A.; Sandoval, W.; Schluter, H.; Sechi, S.; Slavoff, S. A.; Smith, L. M.; Snyder, M. P.; Thomas, P. M.; Uhlen, M.; Van Eyk, J. E.; Vidal, M.; Walt, D. R.; White, F. M.; Williams, E. R.; Wohlschlager, T.; Wysocki, V. H.; Yates, N. A.; Young, N. L.; Zhang, B. *Nat. Chem. Biol.* **2018**, *14*, 206−214.
(20) Nesvizhskii, A. I.; Aebersold, R. *Mol. Cell. Proteomics* **2005**, *4*, 1419−1440.
(21) Strahl, B. D.; Allis, C. D. *Nature* **2000**, *403*, 41−45.
(22) Sze, S. K.; Ge, Y.; Oh, H.; McLafferty, F. W. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 1774−1779.
(23) Siuti, N.; Kelleher, N. L. *Nat. Methods* **2007**, *4*, 817−821.
(24) Anderson, L. C.; DeHart, C. J.; Kaiser, N. K.; Fellers, R. T.; Smith, D. F.; Greer, J. B.; LeDuc, R. D.; Blakney, G. T.; Thomas, P. M.; Kelleher, N. L.; Hendrickson, C. L. *J. Proteome Res.* **2017**, *16*, 1087−1096.
(25) McCool, E. N.; Lubeckyj, R. A.; Shen, X.; Chen, D.; Kou, Q.; Liu, X.; Sun, L. *Anal. Chem.* **2018**, *90*, 5529−5533.
(26) Tran, J. C.; Zamdborg, L.; Ahlf, D. R.; Lee, J. E.; Catherman, A. D.; Durbin, K. R.; Tipton, J. D.; Vellaichamy, A.; Kellie, J. F.; Li, M.; Wu, C.; Sweet, S. M.; Early, B. P.; Siuti, N.; LeDuc, R. D.; Compton, P. D.; Thomas, P. M.; Kelleher, N. L. *Nature* **2011**, *480*, 254−258.
(27) Huang, Y.; Triscari, J. M.; Tseng, G. C.; Pasa-Tolic, L.; Lipton, M. S.; Smith, R. D.; Wysocki, V. H. *Anal. Chem.* **2005**, *77*, 5800−5813.
(28) Cobb, J. S.; Easterling, M. L.; Agar, J. N. *J. Am. Soc. Mass Spectrom.* **2010**, *21*, 949−959.
(29) Julian, R. R. *J. Am. Soc. Mass Spectrom.* **2017**, *28*, 1823−1826.
(30) Fornelli, L.; Srzentic, K.; Toby, T. K.; Doubleday, P. F.; Huguet, R.; Mullen, C.; Melani, R. D.; Dos Santos Seckler, H.; DeHart, C. J.; Weisbrod, C. R.; Durbin, K. R.; Greer, J. B.; Early, B. P.; Fellers, R. T.; Zabrouskov, V.; Thomas, P. M.; Compton, P. D.; Kelleher, N. L. *Mol. Cell. Proteomics* **2020**, *19*, 405−420.
(31) Syka, J. E.; Coon, J. J.; Schroeder, M. J.; Shabanowitz, J.; Hunt, D. F. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 9528−9533.
(32) Fornelli, L.; Srzentic, K.; Huguet, R.; Mullen, C.; Sharma, S.; Zabrouskov, V.; Fellers, R. T.; Durbin, K. R.; Compton, P. D.; Kelleher, N. L. *Anal. Chem.* **2018**, *90*, 8421−8429.
(33) Brunner, A. M.; Lossl, P.; Liu, F.; Huguet, R.; Mullen, C.; Yamashita, M.; Zabrouskov, V.; Makarov, A.; Altelaar, A. F.; Heck, A. J. *Anal. Chem.* **2015**, *87*, 4152−4158.
(34) Riley, N. M.; Westphall, M. S.; Coon, J. J. *Anal. Chem.* **2015**, *87*, 7109−7116.
(35) Riley, N. M.; Westphall, M. S.; Coon, J. J. *J. Proteome Res.* **2017**, *16*, 2653−2659.
(36) Merril, C. R.; Goldman, D.; Sedman, S. A.; Ebert, M. H. *Science* **1981**, *211*, 1437−1438.
(37) Wessel, D.; Flugge, U. I. *Anal. Biochem.* **1984**, *138*, 141−143.
(38) Wilcox, B. E.; Hendrickson, C. L.; Marshall, A. G. *J. Am. Soc. Mass Spectrom.* **2002**, *13*, 1304−1312.
(39) Kaiser, N. K.; Savory, J. J.; Hendrickson, C. L. *J. Am. Soc. Mass Spectrom.* **2014**, *25*, 943−949.
(40) Fellers, R. T.; Greer, J. B.; Early, B. P.; Yu, X.; LeDuc, R. D.; Kelleher, N. L.; Thomas, P. M. *Proteomics* **2015**, *15*, 1235−1238.
(41) Riley, N. M.; Westphall, M. S.; Coon, J. J. *J. Am. Soc. Mass Spectrom.* **2018**, *29*, 140−149.
(42) Houel, S.; Abernathy, R.; Renganathan, K.; Meyer-Arendt, K.; Ahn, N. G.; Old, W. M. *J. Proteome Res.* **2010**, *9*, 4152−4160.
(43) Eng, J. K.; McCormack, A. L.; Yates, J. R. *J. Am. Soc. Mass Spectrom.* **1994**, *5*, 976−989.
(44) Perkins, D. N.; Pappin, D. J.; Creasy, D. M.; Cottrell, J. S. *Electrophoresis* **1999**, *20*, 3551−3567.
(45) Schaub, T. M.; Hendrickson, C. L.; Horning, S.; Quinn, J. P.; Senko, M. W.; Marshall, A. G. *Anal. Chem.* **2008**, *80*, 3985−3990.
(46) Rose, C. M.; Russell, J. D.; Ledvina, A. R.; McAlister, G. C.; Westphall, M. S.; Griep-Raming, J.; Schwartz, J. C.; Coon, J. J.; Syka, J. E. *J. Am. Soc. Mass Spectrom.* **2013**, *24*, 816−827.
(47) 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. *J. Am. Soc. Mass Spectrom.* **2017**, *28*, 1787−1795.
(48) Schwartz, J. C.; Zhou, X.; Bier, M. E. Method and apparatus of increasing dynamic range and sensitivity of a mass spectrometer. U.S. Patent US5572022A1996.
(49) Eliuk, S.; Makarov, A. *Annu. Rev. Anal. Chem.* **2015**, *8*, 61−80.
(50) Zubarev, R. A.; Makarov, A. *Anal. Chem.* **2013**, *85*, 5288−5296.
(51) LeDuc, R. D.; Fellers, R. T.; Early, B. P.; Greer, J. B.; Shams, D. P.; Thomas, P.; Kelleher, N. L. *Mol. Cell. Proteomics* **2019**, *18*, 796−805.
(52) Chanthamontri, C.; Liu, J.; McLuckey, S. A. *Int. J. Mass Spectrom.* **2009**, *283*, 9−16.
(53) Shelimov, K. B.; Clemmer, D. E.; Hudgins, R. R.; Jarrold, M. F. *J. Am. Chem. Soc.* **1997**, *119*, 2240−2248.
(54) Shelimov, K. B.; Jarrold, M. F. *J. Am. Chem. Soc.* **1997**, *119*, 2987−2994.


https://dx.doi.org/10.1021/acs.analchem.0c01064
*Anal. Chem.* 2020, 92, 12193−12200