Published in final edited form as:

Anal Chem. 2018 July 17; 90(14): 8553–8560. doi:10.1021/acs.analchem.8b01638.

The Value of Activated Ion Electron Transfer Dissociation for High-Throughput Top-Down Characterization of Intact Proteins

Nicholas M. Riley^{1,2}, Jacek W. Sikora⁵, Henrique S. Seckler⁵, Joseph B. Greer⁵, Ryan T. Fellers⁵, Richard D. LeDuc⁵, Michael S. Westphall¹, Paul M. Thomas⁵, Neil L. Kelleher⁵, and Joshua J. Coon^{1,2,3,4,*}

¹Genome Center of Wisconsin, University of Wisconsin-Madison, Madison, WI, 53706, USA

²Department of Chemistry, University of Wisconsin-Madison, Madison, WI, 53706, USA

³Department of Biomolecular Chemistry, University of Wisconsin-Madison, Madison, WI, 53706, USA

⁴Morgridge Institute for Research, Madison, Wisconsin, 53706, USA

⁵National Resource for Translational and Developmental Proteomics, Departments of Chemistry and Molecular Biosciences, Northwestern University, Evanston, Illinois 60208, USA

Abstract

High-throughput top-down proteomic experiments directly identify proteoforms in complex mixtures, making high quality tandem mass spectra necessary to deeply characterize proteins with many sources of variation. Collision-based dissociation methods offer expedient data acquisition, but often fail to extensively fragment proteoforms for thorough analysis. Electron-driven dissociation methods are a popular alternative approach, especially for precursor ions with high charge density. Combining infrared photo-activation concurrent with electron transfer dissociation (ETD) reactions, i.e., activated ion ETD (AI-ETD), can significantly improve ETD characterization of intact proteins, but benefits of AI-ETD have yet to be quantified in high-throughput top-down proteomics. Here we report the first application of AI-ETD to LC-MS/MS characterization of intact proteins (<20 kDa), highlighting improved proteoform identification the method offers over higher energy collisional dissociation (HCD), standard ETD, and ETD followed by supplemental HCD activation (EThcD). We identified 935 proteoforms from 295 proteins from human colorectal cancer cell line HCT116 using AI-ETD compared to 1,014

*Corresponding author: jcoon@chem.wisc.edu.

Supporting Information

The .RAW files, .tdReport files, and other files with result metrics are available online at the Chorus Project (Project ID: 1444). Supporting Information is available free of charge online, including an expanded Materials and Methods section and the following figures:

Figure S1. Chromatograms from top-down LC-MS/MS proteomic experiments.

Figure S2. Comparing laser powers (18W, 24W, and 30W) for AI-ETD analyses.

Figure S3. Comparing search strategies for EThcD and AI-ETD.

Figure S4. Overlap in proteoform identifications between dissociation methods.

Figure S5. Proportion of fragments that can be matched to expected product ion types.

Figure S6. Sequence coverage differences for HCD, ETD, and EThcD.

Figure S7. Distribution of sequence coverage values for all four dissociation methods.

Figure S8. HCD spectrum for the $z = 7$ precursor ion from the human mitochondrial acyl carrier protein proteoform (UniProt accession O14561) from Figure 5 in the main text.

proteoforms, 915 proteoforms, and 871 proteoforms with HCD, ETD, and EThcD, respectively. Importantly, AI-ETD outperformed each of the three other methods in MS/MS success rates and spectral quality metrics (e.g., sequence coverage achieved and proteoform characterization scores). In all, this four-method analysis offers the most extensive comparisons to date and demonstrates that AI-ETD both increases identifications over other ETD methods and improves proteoform characterization via higher sequence coverage, positioning it as a premier method for high-throughput top-down proteomics.

Graphical Abstract

Liquid chromatography coupled to tandem mass spectrometry (LC-MS/MS) is the centerpiece of rapid, large-scale protein sequencing, with the large majority of analyses relying on identification of proteolytically derived peptides to study the proteome (i.e., shotgun proteomics).^{1–3} Although this approach has enabled rapid characterization of many thousands of proteins in a single experiment,^{4–9} the reliance on measuring surrogate peptides rather than proteins can leave shotgun methods blind to some degrees of biological context. Top-down proteomics takes a complementary approach, aiming to characterize all the proteoforms present in a system through analysis at the intact level.^{10–14} As such, top-down methods offer the potential to reveal combinatorial patterns of post-translational modifications (PTMs), splice variants, and genetic mutations.^{15,16}

Top-down LC-MS/MS analyses can identify hundreds of proteins and thousands of proteoforms in a given experiment,^{17–22} an achievement built on recent improvements in MS instrumentation, separation methods for intact proteins, and informatics tools. One challenge in top-down experiments is achieving robust dissociation of intact protein ions to sufficiently identify and characterize various proteoforms on a LC timescale. Beam-type collisional activation, referred to as higher energy-collisional dissociation (HCD) on Orbitrap instrument platforms, is widely used in most top down experiments because it is relatively straight-forward to implement, enables quick tandem MS acquisition, and is a thoroughly-studied approach. Despite its benefits, however, HCD often generates only moderate protein sequence coverage with known sites of preferential cleavage,^{23–27} limiting its utility in proteoform characterization. Alternative dissociation methods, including electron-driven dissociation (e.g., electron capture dissociation, ECD, and electron transfer dissociation, ETD)^{28–30} and photo-dissociation (e.g., ultraviolet photodissociation, UVPD)^{27,31,32} methods have been developed for improved sequence coverage and PTM localization in

intact protein analyses and have been successfully applied to LC-MS/MS top-down protein characterization.^{33–36}

Of these alternative dissociation methods, ETD is the most widely available and is often used in conjunction with HCD in top-down experiments to improve proteoform characterization.^{37–43} ETD is particularly well-suited for dissociating highly charged protein cations, but dissociation efficiency suffers as precursor ion charge density decreases, restricting its effectiveness.^{44,45} Several strategies have been investigated to improve the sequence information generated from ETD spectra by disrupting gas-phase structure of protein cations via collisional activation, increased temperatures, higher-energy electrons, and photo-activation.^{45–66} ETD followed by HCD of all product ions, i.e., EThcD,⁶⁷ and infrared photoactivation concurrent with ETD reactions, called activated ion ETD (AI-ETD),⁶⁸ are especially beneficial for characterizing intact proteins. We recently implemented AI-ETD on a quadrupole-Orbitrap-linear ion trap (Orbitrap Fusion Lumos, q-OT-LIT) and improved top-down characterization of intact proteins, including near-complete sequence coverage (for proteins <25 kDa) and improved phosphosite localization in intact phosphoproteins.^{26,69–71} Despite its clear advantages, AI-ETD has only been coupled to online separations (capillary zone electrophoresis) for top-down proteomics in one pilot study⁷² and has yet to be integrated with liquid chromatography for LC-MS/MS characterization of intact proteins.

Here we demonstrate the first coupling of AI-ETD with online reversed phase separations for high-throughput top-down proteome characterization. Using the human colorectal cancer cell line HCT 116 as a model system, we compare the performance of AI-ETD to HCD, ETD, and EThcD. AI-ETD outperforms the other two ETD-based methods, especially in terms of spectral quality and proteoform sequence coverage, and the minimal difference in identifications between AI-ETD and HCD (935 vs. 1,014, respectively) is outweighed by a significant boost in proteoform characterization with AI-ETD. We also comment on the ability of AI-ETD to characterize proteoforms with post-translational modifications. In all, AI-ETD stands as a valuable dissociation method for LC-MS/MS top-down proteomics, offering improved sequence coverage and proteoform characterization without sacrificing sequencing depth.

MATERIALS AND METHODS

HCT 116 cells (ATCC CCL-247) were cultured via standard protocols, and protein concentration was measured using the Pierce BCA Protein Assay Kit (ThermoFisher Scientific) after cell lysis. Following the BCA assay, 375 µg protein was precipitated and pelleted by centrifugation, followed by decanting of the supernatant and washing of the pellet. Precipitated pellets were subsequently denatured and separated into 12 fractions of increasing molecular weight (MW) ranges on a 10% T gel-eluted liquid-fraction entrapment electrophoresis (GELFrEE) cartridge, following manufacturer's instructions (GELFrEE 8100 Fractionation System, Expedeon) and 10 µL sample was taken to visualize protein content and resolution by SDS-PAGE and subsequent silver nitrate stain.⁷³ Two fractions containing lowest MW (up to 20 kDa) range proteins were analyzed by LC-MS. A small mass range was chosen such that comparisons between activation methods could be more

reliably compared amongst each other. Immediately prior to LC–MS analysis, fractions were precipitated using the chloroform/methanol/water method.⁷⁴

Intact proteins were separated using ~15–20 cm reversed-phase columns packed in-house with 5 μ m particle size, 1000 Å pore size PLRP-S particles (Agilent Technologies). Online reversed-phase chromatography was performed using a 90 minute method with a Dionex Ultimate 3000 UPLC system (Thermo Fischer Scientific). Typical chromatograms are shown in Figure S1. Two microliters of sample were injected per analysis and were analyzed using a quadrupole-Orbitrap/quadrupole-linear ion trap hybrid MS system (Orbitrap Fusion Lumos Tribrid, Thermo Fisher Scientific) modified with a continuous wave CO₂ laser (10.6 μ m) as previously described.⁶⁹ Survey scans were collected from 500 – 2000 Th with an AGC target of 500,000, a resolution of 120,000 at 200 m/z, 4 μ scans averaged per spectrum, and the advanced precursor determination algorithm⁷⁵ enabled. MS/MS scans were collected with a resolution of 60,000 at 200 m/z, with 4 μ scans averaged per spectrum, an AGC target of 500,000, and a maximum injection time of 118 ms. All ETD reactions (including EThcD and AI-ETD analyses) were conducted for a static reaction time of 20 ms with a fluoranthene reagent anion AGC target of 700,000. Normalized collision energies (NCE) of 25 NCE and 20 NCE were used for HCD and EThcD, respectively, and 18W laser power was used for AI-ETD.

Raw data were uploaded to a custom version of the National Resource for Translational and Developmental Proteomics (NRTDP, Northwestern University, Evanston, IL) TDPortal²⁰ high performance computing environment (TDPortal 1.3 AI-ETD) for analysis of high-throughput top-down proteomics data. This service is available for academic collaborators at: <http://nrtdp.northwestern.edu/tdportal-request/>. A customized version of TDPortal v1.3 was required to handle the unique .RAW files generated by this project. Data was searched against a proteoform database based on the SwissProt 2016/04 release of the human proteome and scored as described previously.³⁴ Each hit was associated with: a P-score⁷⁶ and a multiple-test corrected E-value,¹⁸ to qualify protein identification; a C-score, to qualify proteoform characterization⁷⁷ (see discussion below). The confidence in each identification was determined by searching the data against a decoy, scrambled database and assuming all decoy hits as incorrect.¹⁷

Each forward hit with at least four matching fragment ions was compared to the decoy results from its search, and only hits scoring above the top 1% of decoys were accepted. Post-search results were visualized by TDViewer v. 1.0.4. Identified protein and proteoforms were exported in list form to Microsoft Excel 2013, where all statistical calculations were performed. The .raw files, .tdReport files, and other files with result metrics are available online at the Chorus Project (Project ID: 1444). See Supporting Materials and Methods for more detailed methods on sample preparation, LC-MS/MS methods, and data processing.

RESULTS AND DISCUSSION

AI-ETD has outperformed several dissociation methods in previous studies that relied on direct infusion of intact protein standards,^{68,70,71,78} so a major goal of this study was to compare AI-ETD performance in high-throughput LC-MS/MS analyses of complex

mixtures intact proteins relative to other commonly used dissociation methods, *i.e.*, HCD and ETD. Furthermore, we wanted to benchmark how AI-ETD performs as a supplemental activation strategy for top-down experiments compared to EThcD. Figure 1 displays the overlap in protein identifications (*i.e.*, unique UniProt accession numbers) and the number of proteoforms characterized by each dissociation method. Figure S4 presents overlap information for proteoforms. AI-ETD provided the most proteoform identifications (935) of the ETD-based methods (915 and 871 proteoforms for ETD and EThcD, respectively), which was also 92% of the total proteoforms identified via HCD (1,014 proteoforms). This is somewhat surprising, as the scan sequences required for HCD fragmentation can be executed more rapidly, which often enables more identifications due to more acquired MS/MS spectra, regardless of overall quality. In top-down experiments, however, the use of the Orbitrap for high resolution MS/MS scans and acquisition of microscans (used to improve spectral quality, 4 microscans per spectrum in this study) slow down the overall duty cycle of the instrument, somewhat leveling the playing field between HCD and ETD-based fragmentation. In these experiments, HCD methods accounted for approximately 20% more MS/MS scans (~500 scans) per 90 minute analysis than ETD-based methods, but translated to only modest gains in identifications overall.

This disparity points to the overall improvement in spectral quality in ETD-based methods, especially AI-ETD. Figure 2a shows the MS/MS success rate of the four dissociation methods, which is a percentage of the number of total MS/MS spectra that were successfully identified as a proteoform. AI-ETD converts MS/MS scans to identifications at the highest rate, and each of the ETD-methods performs better than HCD. Unexpectedly, ETD slightly outperforms EThcD in this metric. Figure 2b and 2c plot the distribution of characterization scores (C-scores) and expectation values (E-values) of proteoforms identified with each method, respectively. The C-score is a metric that represents a nonlinear transformation of posterior probability, which was ultimately designed as a scoring system for proteoform characterization scoring to better capture the information content in high-resolution top down proteomics.⁷⁷ E-values, on the other hand, are a nonlinear transformation of the number of matching fragment ions in a spectrum, without regard to the extent of chemical characterization at the proteoform level. Although less sophisticated, E-values are still valuable for characterizing quality of spectra via the number of sequence-informative fragment ions generated. For both metrics, AI-ETD shifts the distribution to more favorable scores, highlighting the spectral quality it affords. The data in Figure 2 highlight how AI-ETD can provide a similar number of proteoform identifications to HCD even with fewer MS/MS scans, and it indicates that the identifications achieved are of a significantly higher quality. The E-value also provides a surrogate for measuring the proportion total fragments present in a spectrum that match to an expected sequence-informative fragment. If a dissociation method were producing a multitude of fragments ions that did not match to the expected product ion types, E-values would decrease. Thus, AI-ETD is not producing significantly more fragments that are not matched product ion types as compared to the other fragmentation regimes. Figure S5 provides an example of this for the top proteoform hits for each of the dissociation methods, where ~65–70% of fragments are matched to expected product ions in AI-ETD spectra compared to ~35%, ~55%, and ~50% in HCD, ETD, and EThcD spectra, respectively.

Figure 3 provides a more in-depth look at sequence-informative product ions generated by each dissociation method, providing another metric to evaluate their performance. Figure 3a shows the distribution of different fragment types detected for proteoforms identified with each fragmentation method, reported as a percentage of total matching fragments (*i.e.*, relative contribution of each fragment type for a given identified spectrum). This normalized distribution of fragment types allows direct comparison of relative amounts of fragments, especially between ETD-based methods, and Figure 3b reports an average number of total matching fragments per proteoform to give context to the relative distributions shown in panel a. Note, HCD and ETD data includes only b/y-type or c/z[•]-type product ions, respectively. AI-ETD retains similar contributions of c/z[•]-type fragments to ETD, with a small contribution of y-type fragments and minimal amounts of b-type fragments. EThcD, on the other hand, has more significant contributions of b- and y-type fragments, but this comes at a loss of z[•]-type fragments, which can be rationalized by the fragility of radical z[•]-type products during the HCD supplemental activation process. Ultimately, AI-ETD produces similar product ion types to ETD, while providing a higher total number of matching fragments on average than both ETD and EThcD.

Improved spectral quality and higher numbers of sequence-informative product ions translate to improvements in sequence coverage, and ultimately, proteoform characterization. Figure 4 gives several comparisons of sequence coverage metrics between the different dissociation methods. Figure 4a plots sequence coverage as a function of precursor molecular weight, showing the AI-ETD consistently provides the highest sequence coverage across the molecular weight range. EThcD also improves coverage over ETD in most cases, although not to the extent that AI-ETD does, and ETD outperforms HCD for larger precursors. On average, AI-ETD provides a boost of approximately 10% sequence coverage over EThcD, 12% sequence coverage over ETD, and 18% sequence coverage over HCD (Figure 4b). When comparing overlapping proteoform identifications between AI-ETD and the other dissociation methods, AI-ETD provides better sequence coverage for ~85% of the shared identifications, as shown by the difference plots in Figure 4c–e. Similar plots are available in Figure S6 to compare the other dissociation methods. Figure S7 displays boxplots to provide distributions of percent sequence coverage values for the four dissociation methods.

Post-translational Modifications with AI-ETD

Identifying sites of post-translational modification is a major component to proteoform identification. A key benefit of the increases in fragmentation quality and sequence coverage offered by AI-ETD is the ability to localize PTMs to improve confidence in proteoform characterization. Figure 5 summarizes the performance of AI-ETD for characterizing a proteoform of human mitochondrial acyl carrier protein (O14561; gene name, NDUFAB1) that has a PTM, O-phosphopantetheine, not often characterized in proteomic experiments. This protein, beyond functioning as a subunit in NADH dehydrogenase (*i.e.*, complex I) in the mitochondrial electron transport chain, is a key component of fatty acid biosynthesis. Its 4'-phosphopantetheine prosthetic group (Figure 5c) covalently attaches to growing fatty acid chains and carries them to active sites of various enzymes during the synthesis process.^{79–81} All three of the ETD-based methods identified this acyl carrier protein with a O-

phosphopantetheine modification, but ETD and EThcD provided relatively poor fragmentation (14% and 17% sequence coverage, respectively) that did not allow confident localization of the modification to the correct serine. AI-ETD, on the other hand, generated 54% sequence coverage to provide confident localization (Figure 5b). This improvement in spectral quality is reflected in the E-values for these spectra: 6.2×10^{-45} E-value for AI-ETD compared to an E-value of 3.5×10^{-5} for ETD and an E-value of 1.5×10^{-6} for EThcD. Note, this proteoform was not identified with HCD (Figure S8). Although collisional activation can retain modifications on intact protein ions,^{76,82} elimination of phosphopantetheinyl arms from gas phase ions during vibrational activation is a known phenomenon,⁸³ which is likely the reason HCD was not suitable for characterizing the proteoform in this case. It appears that although AI-ETD leverages vibrational activation to unfold precursor ions, labile modifications can still be retained as expected in electron-driven dissociation (also observed for phosphorylation⁷⁰). Interestingly, the region of the protein that had the least sequence coverage with AI-ETD is in a known hydrophobic transmembrane region of the protein. These regions can challenge ETD-based fragmentation.³⁹ Future studies could investigate how AI-ETD performs for intact proteins with hydrophobic transmembrane proteins specifically, especially because previous AI-ECD approaches have been beneficial.⁵⁵ Furthermore, the identification of NDUFB1 presents an interesting case because its precursor ion has relatively low charge density ($\sim 1502 \text{ m/z}$, $z = 7$), where ETD efficiency is low. EThcD does not offer a substantial boost in fragmentation quality in this example, but AI-ETD significantly improves the number of fragment ions and spectral quality, even for a low charge-density precursor with a relatively large modification.

CONCLUSION

The addition of AI-ETD to the top-down proteomics workflow improves proteoform identification and characterization in complex mixtures, including the ability to localize labile PTMs with confidence. The added spectral quality and sequence coverage that AI-ETD offers through improved precursor-to-product ion conversion makes it an attractive option for a number of applications in top-down proteomics, both in discovery and targeted experiments. While this study focuses on relatively small proteins from complex mixtures, expanding the molecular weight range that top-down proteomics can effectively sample is a current goal of the field.^{20,21} AI-ETD has shown promise for improving fragmentation of larger proteins (30–70 kDa),⁷¹ although the amount of spectral averaging required to sufficiently characterize most larger proteins limits the ability to analyze complex mixtures of them on the LC-MS/MS timescale. We note that HCD can be used to identify larger proteins with a lesser degree of scan averaging, but sequence coverage and thorough proteoform characterization are limited in such a regime. We expect a combination of more expedient HCD scans for identification and targeted AI-ETD scans for improved sequence coverage to be a valuable complementary approach, especially for analyses of larger proteins. One potential strategy may be to integrate AI-ETD into intelligent acquisition schemes, e.g., Autopilot,³⁸ which would be able to select optimal fragmentation conditions in real time and determine what precursors may need multiple MS/MS scans to get sufficient sequence coverage.⁸⁴ Additionally, combining AI-ETD with multiple fill analyses, ion parking strategies to protect highly charged fragment ions from subsequent ETD-reactions,

and proton transfer reactions to more evenly distribution product ion signal across the m/z range would all further improve AI-ETD characterization in LC-MS/MS top-down experiments,^{85–89} enabling the faster interrogation of larger proteins sacrificing neither speed nor depth of proteoform characterization. Even with these potential improvements, this study shows that AI-ETD is well-suited for high-throughput top-down proteomic experiments using LC-MS/MS, and we expect it will provide a direct benefit to most applications where ETD is used for intact protein characterization.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors gratefully acknowledge support from Thermo Fisher Scientific and NIH Grants P41 GM108538 (JJC) and P41 GM108569 (NLK). NMR was funded through an NIH Predoctoral to Postdoctoral Transition Award (F99 CA212454).

References

1. Zhang Y, Fonslow BR, Shan B, Baek MC, Yates JR. *Chem Rev.* 2013; 113:2343–2394. [PubMed: 23438204]
2. Aebersold R, Mann M. *Nature.* 2016; 537:347–355. [PubMed: 27629641]
3. Riley NM, Hebert AS, Coon JJ. *Proteomics Moves into the Fast Lane. Cell Systems.* 2016; 2:142–143. [PubMed: 27135360]
4. Pirmoradian M, Budamgunta H, Chingin K, Zhang B, Astorga-Wells J, Zubarev RA. *Mol Cell Proteomics.* 2013; 12:3330–3338. [PubMed: 23878402]
5. Hebert AS, Richards AL, Bailey DJ, Ulbrich A, Coughlin EE, Westphall MS, Coon JJ. *Mol Cell Proteomics.* 2014; 13:339–347. [PubMed: 24143002]
6. Richards AL, Hebert AS, Ulbrich A, Bailey DJ, Coughlin EE, Westphall MS, Coon JJ. *Nat Protoc.* 2015; 10:701–714. [PubMed: 25855955]
7. Scheltema RA, Hauschild JP, Lange O, Hornburg D, Denisov E, Damoc E, Kuehn A, Makarov A, Mann M. *Mol Cell Proteomics.* 2014; 13:3698–3708. [PubMed: 25360005]
8. Kelstrup CD, Jersie-Christensen RR, Batth TS, Arrey TN, Kuehn A, Kellmann M, Olsen JV. *J Proteome Res.* 2014; 13:6187–6195. [PubMed: 25349961]
9. Bekker-Jensen DB, Kelstrup CD, Batth TS, Larsen SC, Haldrup C, Bramsen JB, Sorensen KD, Hoyer S, Orntoft TF, Andersen CL, Nielsen ML, Olsen JV. *Cell Syst.* 2017; 4:587–599. [PubMed: 28601559]
10. Smith LM, Kelleher NL. *Nat Methods.* 2013; 10:186–187. [PubMed: 23443629]
11. Catherman AD, Skinner OS, Kelleher NL. *Biochem Biophys Res Commun.* 2014; 445:683–693. [PubMed: 24556311]
12. Toby TK, Fornelli L, Kelleher NL. *Annu Rev Anal Chem.* 2016; 9:499–519.
13. Patrie SM. *Adv Exp Med Biol.* 2016; 919:171–200. [PubMed: 27975217]
14. Chen B, Brown KA, Lin Z, Ge Y. *Top-Down Proteomics: Ready for Prime Time? Analytical Chemistry.* 2018; 90:110–127. [PubMed: 29161012]
15. Aebersold R, Agar JN, Amster IJ, Baker MS, Bertozzi CR, Boja ES, Costello CE, Cravatt BF, Fenselau C, Garcia BA, Ge Y, Gunawardena J, Hendrickson RC, Hergenrother PJ, Huber CG, Ivanov AR, Jensen ON, Jewett MC, Kelleher NL, Kiessling LL, Krogan NJ, Larsen MR, Loo JA, Ogorzalek Loo RR, Lundberg E, MacCoss MJ, Mallick P, Mootha VK, Mrksich M, Muir TW, Patrie SM, Pesavento JJ, Pitteri SJ, Rodriguez H, Saghatelian A, Sandoval W, Schluter H, Sechi S, Slavoff SA, Smith LM, Snyder MP, Thomas PM, Uhlen M, Van Eyk JE, Vidal M, Walt DR, White

- FM, Williams ER, Wohlschlager T, Wysocki VH, Yates NA, Young NL, Zhang B. *Nat Chem Biol*. 2018; 14:206–214. [PubMed: 29443976]
16. Smith LM, Kelleher NL. *Science* (80-). 2018; 359:1106–1107.
17. Tran JC, Zamdborg L, Ahlf DR, Lee JE, Catherman AD, Durbin KR, Tipton JD, Vellaichamy A, Kellie JF, Li M, Wu C, Sweet SMM, Early BP, Siuti N, LeDuc RD, Compton PD, Thomas PM, Kelleher NL. *Nature*. 2011; 480:254–258. [PubMed: 22037311]
18. Catherman AD, Durbin KR, Ahlf DR, Early BP, Fellers RT, Tran JC, Thomas PM, Kelleher NL. *Mol Cell Proteomics*. 2013; 12:3465–3473. [PubMed: 24023390]
19. Durbin KR, Fornelli L, Fellers RT, Doubleday PF, Narita M, Kelleher NL. *J Proteome Res*. 2016; 15:976–982. [PubMed: 26795204]
20. Fornelli L, Durbin KR, Fellers RT, Early BP, Greer JB, LeDuc RD, Compton PD, Kelleher NL. *J Proteome Res*. 2017; 16:609–618. [PubMed: 28152595]
21. Anderson LC, Dehart CJ, Kaiser NK, Fellers RT, Smith DF, Greer JB, LeDuc RD, Blakney GT, Thomas PM, Kelleher NL, Hendrickson CL. *J Proteome Res*. 2017; 16:1087–1096. [PubMed: 27936753]
22. Cai W, Tucholski T, Chen B, Alpert AJ, McIlwain S, Kohmoto T, Jin S, Ge Y. *Anal Chem*. 2017; 89:5467–5475. [PubMed: 28406609]
23. Huang Y, Triscari JM, Tseng GC, Pasa-Tolic L, Lipton MS, Smith RD, Wysocki VH. *Anal Chem*. 2005; 77:5800–5813. [PubMed: 16159109]
24. Catherman AD, Li M, Tran JC, Durbin KR, Compton PD, Early BP, Thomas PM, Kelleher NL. *Anal Chem*. 2013; 85:1880–1888. [PubMed: 23305238]
25. Haverland NA, Skinner OS, Fellers RT, Tariq AA, Early BP, LeDuc RD, Fornelli L, Compton PD, Kelleher NL. *J Am Soc Mass Spectrom*. 2017; 28:1203–1215. [PubMed: 28374312]
26. Riley NM, Westphall MS, Coon JJ. *J Proteome Res*. 2017; 16:2653–2659. [PubMed: 28608681]
27. Shaw JB, Li W, Holden DD, Zhang Y, Griep-Raming J, Fellers RT, Early BP, Thomas PM, Kelleher NL, Brodbelt JS. *J Am Chem Soc*. 2013; 135:12646–12651. [PubMed: 23697802]
28. Zubarev R, Kelleher NL, McLafferty FW. *J Am Chem Soc*. 1998; 120:3265–3266.
29. Syka JEP, Coon JJ, Schroeder MJ, Shabanowitz J, Hunt DF. *Proc Natl Acad Sci U S A*. 2004; 101:9528–9533. [PubMed: 15210983]
30. Coon JJ, Ueberheide B, Syka JEP, Dryhurst DD, Ausio J, Shabanowitz J, Hunt DF. *Proc Natl Acad Sci U S A*. 2005; 102:9463–9468. [PubMed: 15983376]
31. Shaw JB, Robinson EW, Paša-Toli L. *Anal Chem*. 2016; 88:3019–3023. [PubMed: 26882021]
32. Julian R. *J Am Soc Mass Spectrom*. 2017; 28:1823–1826. [PubMed: 28702929]
33. Cannon JR, Cammarata MB, Robotham SA, Cotham VC, Shaw JB, Fellers RT, Early BP, Thomas PM, Kelleher NL, Brodbelt JS. *Anal Chem*. 2014; 86:2185–2192. [PubMed: 24447299]
34. Cleland TP, Dehart CJ, Fellers RT, Vannispén AJ, Greer JB, LeDuc RD, Parker WR, Thomas PM, Kelleher NL, Brodbelt JS. *J Proteome Res*. 2017; 16:2072–2079. [PubMed: 28412815]
35. Brodbelt JS. Ion Activation Methods for Peptides and Proteins. *Analytical Chemistry*. 2016; 88:30–51. [PubMed: 26630359]
36. Greer SM, Brodbelt JS. *J Proteome Res*. 2018; 17:1138–1145. [PubMed: 29343059]
37. Riley NM, Coon JJ. *Anal Chem*. 2018; 90:40–64. [PubMed: 29172454]
38. Durbin KR, Fellers RT, Ntai I, Kelleher NL, Compton PD. *Anal Chem*. 2014; 86:1485–1492. [PubMed: 24400813]
39. Skinner OS, Catherman AD, Early BP, Thomas PM, Compton PD, Kelleher NL. *Anal Chem*. 2014; 86:4627–4634. [PubMed: 24689519]
40. Rožman M, Gaskell SJ. *Rapid Commun Mass Spectrom*. 2012; 26:282–286. [PubMed: 22223314]
41. Ahlf DR, Compton PD, Tran JC, Early BP, Thomas PM, Kelleher NL. *J Proteome Res*. 2012; 11:4308–4314. [PubMed: 22746247]
42. Coon JJ. *Anal Chem*. 2009; 81:3208–3215. [PubMed: 19364119]
43. Zubarev Ra, Zubarev AR, , Savitski MM. *J Am Soc Mass Spectrom*. 2008; 19:753–761. [PubMed: 18499036]

44. Good DM, Wirtala M, McAlister GC, Coon JJ. *Mol Cell Proteomics*. 2007; 6:1942–1951. [PubMed: 17673454]
45. Pitteri SJ, Chrisman PA, McLuckey SA. *Anal Chem*. 2005; 77:5662–5669. [PubMed: 16131079]
46. Horn DM, Ge Y, McLafferty FW. *Anal Chem*. 2000; 72:4778–4784. [PubMed: 11055690]
47. Han H, Xia Y, McLuckey S. *Rapid Commun mass* 2007; 21:1567–1573.
48. Swaney DL, McAlister GC, Wirtala M, Schwartz JC, Syka JEP, Coon JJ. *Anal Chem*. 2007; 79:477–485. [PubMed: 17222010]
49. Frese CK, Altelaar AFM, van den Toorn H, Nolting D, Griep-Raming J, Heck AJR, Mohammed S. *Anal Chem*. 2012; 84:9668–9673. [PubMed: 23106539]
50. Ledvina AR, McAlister GC, Gardner MW, Smith SI, Madsen JA, Schwartz JC, Stafford GC Jr, Syka JE, Brodbelt JS, Coon JJ. *Angew Chem Int Ed Engl*. 2009; 48:8526–8528. [PubMed: 19795429]
51. Ledvina AR, Beauchene Na, McAlister GC, Syka JEP, Schwartz JC, Griep-Raming J, Westphall MS, Coon JJ. *Anal Chem*. 2010; 82:10068–10074. [PubMed: 21062032]
52. Ledvina AR, Rose CM, McAlister GC, Syka JEP, Westphall MS, Griep-Raming J, Schwartz JC, Coon JJ. *J Am Soc Mass Spectrom*. 2013; 24:1623–1633. [PubMed: 23677544]
53. Cannon JR, Holden DD, Brodbelt JS. *Anal Chem*. 2014; 86:10970–10977. [PubMed: 25270663]
54. Tsybin YO, He H, Emmett MR, Hendrickson CL, Marshall AG. *Anal Chem*. 2007; 79:7596–7602. [PubMed: 17874851]
55. Zabrouskov V, Whitelegge JP. *J Proteome Res*. 2007; 6:2205–2210. [PubMed: 17441748]
56. Lermyte F, Sobott F. *Proteomics*. 2015; 15:2813–2822. [PubMed: 26081219]
57. Sze SK, Ge Y, Oh H, McLafferty FW. *Proc Natl Acad Sci U S A*. 2002; 99:1774–1779. [PubMed: 11842225]
58. Lermyte F, Konijnenberg A, Williams JP, Brown JM, Valkenborg D, Sobott F. *J Am Soc Mass Spectrom*. 2014; 25:343–350. [PubMed: 24408179]
59. Lermyte F, Łcki MK, Valkenborg D, Gambin A, Sobott F. *J Am Soc Mass Spectrom*. 2017; 28:69–76. [PubMed: 27495285]
60. Sze SK, Ge Y, McLafferty FW. *Anal Chem*. 2003; 75:1599–1603. [PubMed: 12705591]
61. Hakansson K, Chalmers MJ, Quinn JP, McFarland MA, Hendrickson CL, Marshall AG. *Anal Chem*. 2003; 75:3256–3262. [PubMed: 12964777]
62. Tsybin YO, Witt M, Baykut G, Kjeldsen F, Hakansson P. *Rapid Commun mass Spectrom*. 2003; 17:1759–1768. [PubMed: 12872281]
63. Ge Y, Lawhorn BG, ElNaggar M, Strauss E, Park J-H, Begley TP, McLafferty FW. *J Am Chem Soc*. 2002; 124:672–678. [PubMed: 11804498]
64. Tsybin YO, Witt M, Baykut G, Hakansson P. *Rapid Commun Mass Spectrom*. 2004; 18:1607–1613. [PubMed: 15282786]
65. Mikhailov VA, Cooper HJ. *J Am Soc Mass Spectrom*. 2009; 20:763–771. [PubMed: 19200749]
66. Xia Y, Han H, McLuckey SA. *Anal Chem*. 2008; 80:1111–1117. [PubMed: 18198896]
67. Brunner AM, Lossel P, Liu F, Huguet R, Mullen C, Yamashita M, Zabrouskov V, Makarov A, Altelaar AFM, Heck AJR. *Anal Chem*. 2015; 87:4152–4158. [PubMed: 25803405]
68. Riley NM, Westphall MS, Coon JJ. *Anal Chem*. 2015; 87:7109–7116. [PubMed: 26067513]
69. Riley NM, Westphall MS, Hebert AS, Coon JJ. *Anal Chem*. 2017; 89:6358–6366. [PubMed: 28383247]
70. Riley NM, Hebert AS, Durnberger G, Stanek F, Mechtler K, Westphall MS, Coon JJ. *Anal Chem*. 2017; 89:6367–6376. [PubMed: 28383256]
71. Riley NM, Westphall MS, Coon JJ. *J Am Soc Mass Spectrom*. 2018; 29:140–149. [PubMed: 29027149]
72. Zhao Y, Riley NM, Sun L, Hebert AS, Yan X, Westphall MS, Rush MJP, Zhu G, Champion MM, Medie FM, Champion PAD, Coon JJ, Dovichi NJ. *Anal Chem*. 2015; 87:5422–5429. [PubMed: 25893372]
73. Shevchenko A, Wilm M, Vorm O, Mann M. *Anal Chem*. 1996; 68:850–858. [PubMed: 8779443]
74. Wessel D, Flugge UI. *Anal Biochem*. 1984; 138:141–143. [PubMed: 6731838]

75. Hebert AS, Thoing C, Riley NM, Kwiecien NW, Shiskova E, Huguet R, Cardasis HL, Kuehn A, Eliuk S, Zabrouskov V, Westphall MS, McAlister GC, Coon JJ. *Anal Chem.* 2018; 90:2333–2340. [PubMed: 29272103]
76. Meng F, Cargile BJ, Miller LM, Forbes AJ, Johnson JR, Kelleher NL. *Nat Biotechnol.* 2001; 19:952–957. [PubMed: 11581661]
77. Leduc RD, Fellers RT, Early BP, Greer JB, Thomas PM, Kelleher NL. *J Proteome Res.* 2014; 13:3231–3240. [PubMed: 24922115]
78. Riley NM, Westphall MS, Coon JJ. *J Proteome Res.* 2017; 16:2653–2659. [PubMed: 28608681]
79. Triepels R, Smeitink J, Loeffen J, Smeets R, Buskens C, Trijbels F, Van Den Heuvel L. *J Inherit Metab Dis.* 1999; 22:163–173. [PubMed: 10234612]
80. Feng D, Witkowski A, Smith S. *J Biol Chem.* 2009; 284:11436–11445. [PubMed: 19221180]
81. Elurbe DM, Huynen MA. *Biochim Biophys Acta - Bioenerg.* 2016; 1857:971–979.
82. Siuti N, Kelleher NL. *Nat Methods.* 2007; 4:817–821. [PubMed: 17901871]
83. Dorrestein PC, Bumpus SB, Calderone CT, Garneau-Tsodikova S, Aron ZD, Straight PD, Kolter R, Walsh CT, Kelleher NL. *Biochemistry.* 2006; 45:12756–12766. [PubMed: 17042494]
84. Bailey DJ, Rose CM, McAlister GC, Brumbaugh J, Yu P, Wenger CD, Westphall MS, Thomson JA, Coon JJ. *Proc Natl Acad Sci U S A.* 2012; 109:8411–8416. [PubMed: 22586074]
85. Senko MW, Remes PM, Canterbury JD, Mathur R, Song Q, Eliuk SM, Mullen C, Earley L, Hardman M, Blethrow JD, Bui H, Specht A, Lange O, Denisov E, Makarov A, Horning S, Zabrouskov V. *Anal Chem.* 2013; 85:11710–11714. [PubMed: 24251866]
86. Weisbrod CR, Kaiser NK, Syka JEP, Early L, Mullen C, Dunyach JJ, English AM, Anderson LC, Blakney GT, Shabanowitz J, Hendrickson CL, Marshall AG, Hunt DF. *J Am Soc Mass Spectrom.* 2017; 28:1787–1795. [PubMed: 28721671]
87. Anderson LC, English AM, Wang WH, Bai DL, Shabanowitz J, Hunt DF. *Int J Mass Spectrom.* 2015; 377:617–624. [PubMed: 25844056]
88. Anderson LC, Karch KR, Ugrin SA, Coradin M, English AM, Sidoli S, Shabanowitz J, Garcia BA, Hunt DF. *Mol Cell Proteomics.* 2016; 15:975–988. [PubMed: 26785730]
89. Zhang L, English AM, Bai DL, Ugrin SA, Shabanowitz J, Ross MM, Hunt DF, Wang WH. *Mol Cell Proteomics.* 2016; 15:1479–1488. [PubMed: 26621848]

Figure 1. Overlap in protein identifications (*i.e.*, unique UniProt accession numbers) (**a**) and proteoforms identified (**b**) for HCD, ETD, EThcD, and AI-ETD analyses.

Figure 2.

Panel (a) shows MS/MS success rates for each of the four dissociation methods, where success rate is defined as the percentage of MS/MS scans that converted to a proteoform identification. Panels (b) and (c) compare C-score and $-\log(\text{E-value})$ distributions, respectively, for HCD, ETD, EThcD, and AI-ETD analyses.

Figure 3.

a) Distributions of relative amounts of b-, c-, y-, and z-product ion types, plotted as percentage of the total number of matched product ions, are shown for proteoform identifications from each dissociation method. **b)** The average number of total product ions matched per proteoform identification for each dissociation method.

Figure 4. Four-way and binary comparisons of proteoform fragmentation methods

a) Percent sequence coverage is plotted as a function of precursor molecular weight, where the percent sequence coverage is calculated as a rolling average of the 50 subsequent sequence coverage values. **b)** The average percent sequence coverage for all proteoform identifications based on dissociation method. Differences in sequence coverage for proteoforms seen with AI-ETD and **(c)** HCD, **(d)** ETD, and **(e)** ETHcD are plotted in rank order, where a positive value indicates a higher sequence coverage seen with AI-ETD (colored in red). In all cases, AI-ETD offered superior sequence coverage for approximately 85% of the proteoforms. Note, the value along the x-axes in panels **(c–e)** indicate the number of unique proteoforms identifications overlapping between AI-ETD and the given dissociation method.

Figure 5. Performance for a human mitochondrial acyl carrier protein proteoform (UniProt accession O14561) with a phosphopantetheine modification on serine 44

a) MS/MS spectra of the $z = 7$ precursor ion using ETD, EThcD, and AI-ETD, with the number of matched fragments and percent sequence coverage provided. All spectra are on the same intensity scale, and the grey regions show unreacted and charge reduced precursor ions at 10% their actual intensity. E-values are 3.5×10^{-5} , 1.5×10^{-6} , and 6.2×10^{-45} for the ETD, EThcD, and AI-ETD spectra, respectively. Each spectrum here represents a single scan number (i.e., 4 microscans). Note, this proteoform was not identified with HCD. **b)** A sequence coverage map for the AI-ETD spectrum showing matched product ions and location of the O-phospho- pantetheine-modified serine (light blue box). Backbone fragmentation markers shown in red and blue represent $c/z^{\bullet}$ -type and b/y -type product ions, respectively. **c)** Structure of O-phospho-pantetheine-modified serine, with grey background indicating the serine residue.