Journal of proteome research

pubs.acs.org/jpr

Technical Note

MS-TAFI: A Tool for the Analysis of Fragment Ions Generated from Intact Proteins

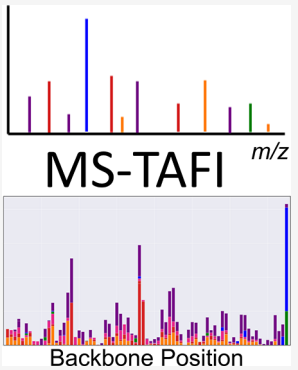
Kyle J. Juetten and Jennifer S. Brodbelt*

Cite This: *J. Proteome Res.* 2023, 22, 546−550

Read Online

ACCESS | Metrics & More | Article Recommendations

ABSTRACT: Tandem mass spectrometry (MS/MS) spectra of intact proteins can be difficult to interpret owing to the variety of fragment ion types and abundances. This information is crucial for maximizing the information derived from top-down mass spectrometry of proteins and protein complexes. MS-TAFI (Mass Spectrometry Tool for the Analysis of Fragment Ions) is a free Python-based program which offers a streamlined approach to the data analysis and visualization of deconvoluted MS/MS data of intact proteins. The application also contains tools for native mass spectrometry experiments with the ability to search for fragment ions that retain ligands (holo ions) as well as visualize the location of charge sites obtained from 193 nm ultraviolet photodissociation data. The source code and complete application for MS-TAFI is available for download at https://github.com/kylejuetten.



KEYWORDS: *MS/MS analysis, proteomics software, protein fragmentation*

INTRODUCTION

Top-down mass spectrometry (TD-MS) is an increasingly popular strategy at the forefront of proteomics that focuses on the identification and characterization of intact proteins and protein complexes.[1,2] This technique has emerged as an appealing choice in recent years because it obviates the need for protein digestion unlike the well-established peptide-based bottom-up approach.[3,4] Much of the interest in TD-MS is aimed at the characterization of post-translationally modified proteins (or proteoforms) for which key contextual insight may be lost by analyzing incomplete mixtures of small proteolytic peptides.[5−7] Robust activation methods are needed to maximize the information obtained from tandem mass spectrometry (MS/MS) spectra in order to determine the identity and location of post-translational modifications (PTMs), uncover the breadth of sequence variations, and map sites of preferential cleavage, all features which can be crucial for characterizing proteoforms.[8−10]

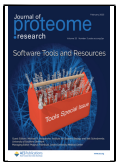
There is an ongoing need for improved tools related to top-down data analysis.[11−13] Several popular programs exist in this space such as ProSight Lite which allows users to generate publication-quality sequence coverage maps, or TDValidator which serves as a one-stop shop for proteoform analysis.[12,14] MASH Explorer offers another comprehensive software suite for proteomics with additional focus on native MS analysis.[15] TDFragMapper is another resource which aids the optimization of experimental parameters for top-down experiments.[16] However, these programs have limitations pertaining to the types of information provided (such as sorting ions by type) and/or their accessibility. Herein we present the Mass Spectrometry Tool for the Analysis of Fragment Ions, or MS-TAFI, as a free python-based software program for the analysis and visualization of TD-MS data. By matching the fragments of deconvoluted MS/MS spectra to a proposed sequence, a user can determine the sequence coverage, both the identities and locations of post-translational modifications, and visualize the fragment intensities as a function of backbone cleavage position. The application supports data from a variety of activation methods (CID, ETD, ETciD, UVPD) and contains additional utilities for native mass spectrometry data such as the analysis of ligand binding and charge site localization.

FEATURES

The general understanding of protein fragmentation established by the mobile proton model can be supplemented by probing the abundance of fragments as this can help reveal the preferential nature of fragments according a protein's sequence and the activation method used.[17−19] Here, an easy-to-use tool is presented to visualize the preferential sites of backbone

Special Issue: Software Tools and Resources 2023

Received: September 28, 2022
Published: December 14, 2022



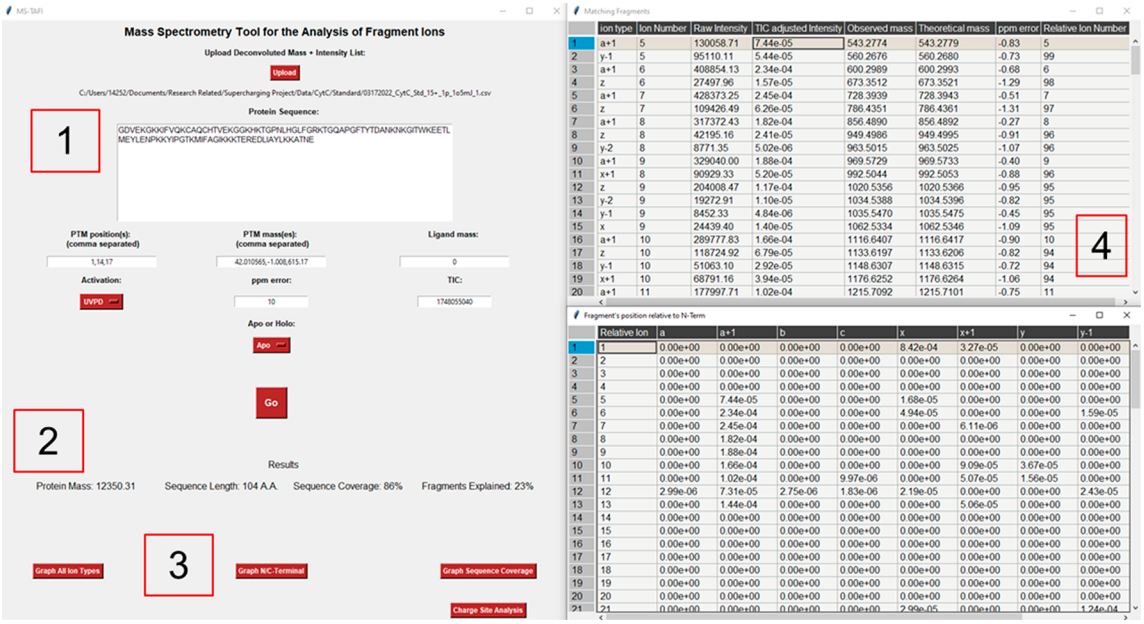


Figure 1. Main window of MS-TAFI contains sections for (1) data entry, (2) monoisotopic mass and sequence coverage, (3) graphing utilities, and (4) fragment information including type of fragment ion (*a*, *b*, *c*, ...), intensity and mass value, and ppm error. In the upper table, ion number indicates the number of amino acids in the fragment ion, and relative ion number describes the backbone position relative to the N-terminus. Raw intensities are intensity values extracted directly from the deconvoluted mass list. The raw intensity values are divided by the total ion current (TIC) to obtain the TIC adjusted intensity values. Data shown is from cytochrome c (15+) activated using 193 nm UVPD (1 pulse at 1.5 mJ).

cleavages and categorize the types of fragment ions (*a*, *b*, *c*, *x*, *y*, *z*) that originated from those backbone cleavage sites. The fragment intensity tool reads in deconvoluted data in a TAB delimited format. Any deconvolution algorithm may be used, given that the output contains information related to fragment mass and intensity. Once the deconvoluted data is imported, the user can enter the relevant sequence information, the locations and masses of PTMs, the desired error tolerance (in ppm) for fragment identification, the activation method used, and the total ion current for normalizing fragment intensities. MS-TAFI is compatible with MS/MS data generated by CID (*b*/*y*), ETD (*c*/*z*), ETciD (*b*/*c*/*y*/*z*), and UVPD (*a*/*a* + 1/*b*/*c*/*x*/*x* + 1/*y*/*y* + 1/*y* + 2/*z*). Executing the program produces a monoisotopic mass for the input sequence (including the masses of PTMs if present) along with the sequence length, sequence coverage, and percent of fragments explained. These are standard metrics in TD-MS for evaluating the extent of characterization. Two tables are generated which display additional fragment information. The upper table displays all the fragment ion output with respect to assignments, raw and TIC adjusted intensities, and ppm errors. By default, the top table lists fragments in order of increasing numerical value in the position column (i.e., a_3 ions come before b_5 ions); however, this table can easily be reordered by simply clicking on the desired column. This allows the user to easily restructure the table to evaluate a parameter of interest (ion intensity, ppm error, etc.). The lower table only contains the ion type, position, and TIC adjusted intensity, making it easier for users to transfer the data content into a spreadsheet for further processing or graphing (Figure 1). Fragment ions are assigned based on the backbone position cleaved between pairs of residues and are not "assigned" to one of the amino acids connected by the cleavage backbone bond, as done with one of our former tools.[20]

A key feature of MS-TAFI is the graphing capabilities which allow visualization of the relative abundances of backbone cleavages while also retaining information related to the fragment type. The added capability to normalize abundances according to total ion current allows abundances to be compared across different samples. This type of analysis is particularly beneficial when deducing preferential backbone cleavage sites. MS-TAFI allows data to be graphed in three different formats; the first displaying all ion types, which retains individual fragment identities (*a*/*b*/*c*/*x*/*y*/*z*, etc.), the second grouping fragment ions based on whether they contain the N-terminal (*a*, *b*, *c*) or C-terminal (*x*, *y*, *z*), and the third providing a sequence coverage map that delineates the backbone cleavage sites and types of fragment ions associated with each cleavage site. Amino acids with PTMs are highlighted in this latter view. Examples of these three formats are shown in Figure 2 for cytochrome c (15+) activated using 193 nm UVPD. Graphs are generated in a Web browser and are fully interactive, allowing a user to parse through data in a customized manner. The interactive view allows the user to omit or combine specific ion types, zoom or pan across the data set, and download graphs as a. png file for use in figures.

The fragment intensity tool also supports searches for apo (no ligand) and holo (containing ligand) fragment ions when interested in examining features related to ligand binding.[21] By entering a ligand mass into the designated box and selecting Apo or Holo from the drop down menu, the user can easily search for ions which contain (or lack) the bound ligand. Analyzing the intensities of these fragments allows inquiry into the preferred binding sites of a particular ligand. An example of this function is displayed in Figure 3 as shown for the myoglobin−heme complex (8+) activated using 193 nm UVPD.

The second feature of MS-TAFI is a tool for deriving and visualizing potential charge sites along the protein backbone

https://doi.org/10.1021/acs.jproteome.2c00594
J. Proteome Res. 2023, 22, 546−550

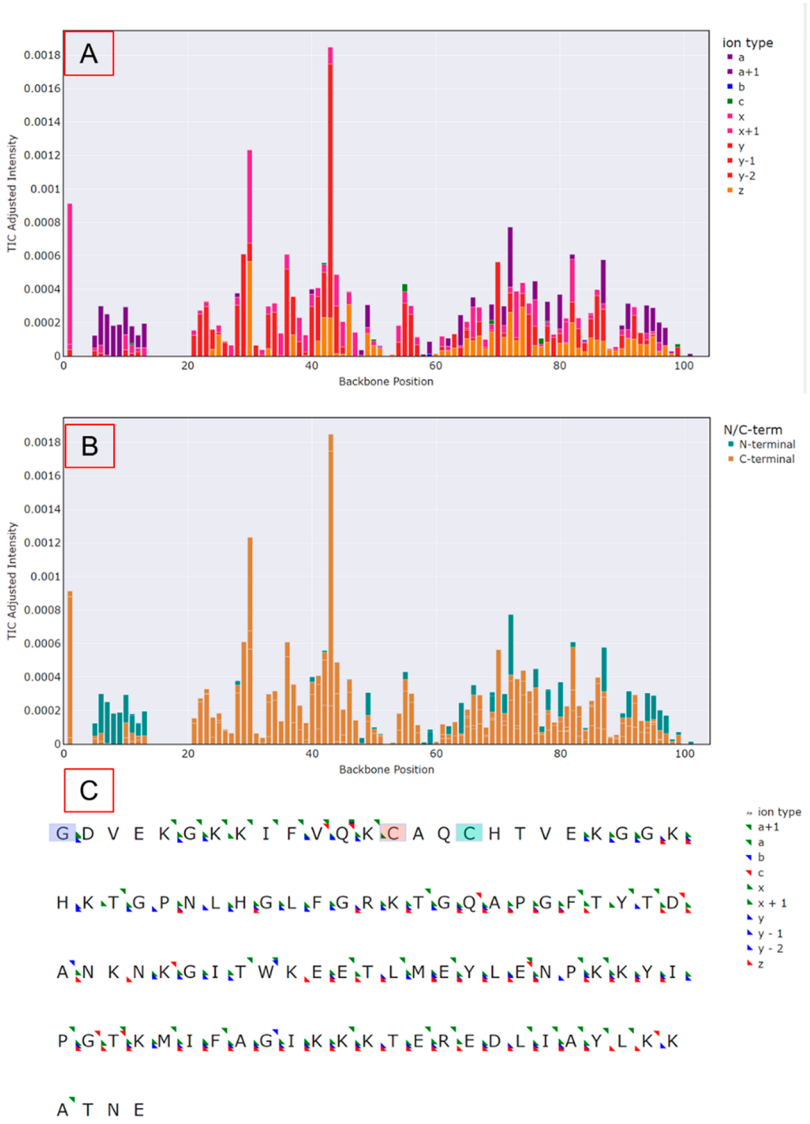


Figure 2. MS-TAFI displays fragment information in three different formats, as illustrated for cytochrome c (15+) activated using 193 nm UVPD (1 pulse, 1.5 mJ). Fragments intensities are displayed as TIC adjusted values. (A) graph of all fragment ion types based on the backbone position cleaved, (B) graph of fragment ion types according to whether they include the N-terminus or C-terminus based on the backbone position cleaved, and (3) sequence coverage map. Specific ion types can be selected, combined, or omitted from the graphs. Ion types from the same family (i.e., y, $y-1$, $y-2$) share the same color. Locations and types of PTMs are color-coded in the sequence coverage map according to their unique mass. Here acetylation is shaded blue, a hydrogen loss is shaded red (to account for the attachment of heme), and the attachment of heme is shaded green.

based on examination of the charge sites of specific fragment ions produced by UVPD.[22] This analysis utilizes a strategy established previously which maps the *a*- and *x*-type ions generated from UVPD to assign the charge site locations within a sequence.[22] The abundances of each *a*- and *x*-type ion of each charge state are tabulated, and the relative abundances of each charge state of each *a*- or *x*-type ion are calculated as a function of the total *a*- or *x*-type ion population. The results display the percent abundances of fragments in each charge state according to the backbone site that was cleaved to generate the fragment ions. Comparing the results generated using the two different ion types (*a* versus *x*) allows additional confirmation of the assigned charge locations. The charge state analysis tool requires deconvoluted mass, intensity, and fragment charge information and has been designed around data generated by Xtract within Thermo Fisher Scientific's Freestyle software suite. The graphs generated from this analysis allow the user to easily visualize predicted proton locations (Figure 4).

CONCLUSIONS AND FUTURE DIRECTIONS

A free python-based software program, MS-TAFI, was developed to aid the analysis and visualization of fragmentation data generated from MS/MS of intact proteins. MS-TAFI utilizes deconvoluted fragment ion data to visualize fragment abundances based on the backbone cleavage sites from which they originate, probe ligand binding through apo/holo fragment ion analysis, and determine the charge sites of proteins. Future development is aimed toward expanding applications of MS-TAFI for other classes of biological molecules such as oligonucleotides and lipids.

https://doi.org/10.1021/acs.jproteome.2c00594
J. Proteome Res. 2023, 22, 546−550

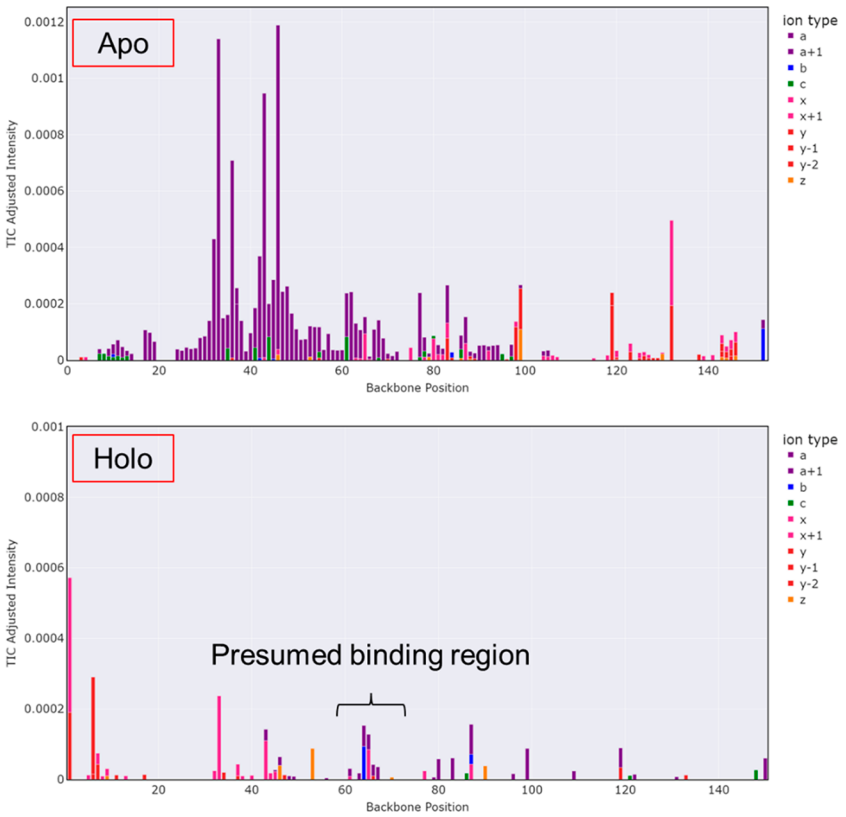


Figure 3. Output for apo/holo fragment ion search of myoglobin−heme complex (8+) activated using 193 nm UVPD (1 pulse, 3 mJ). Tracking the distribution of N/C terminal fragments allows insight into the presumed region of binding.

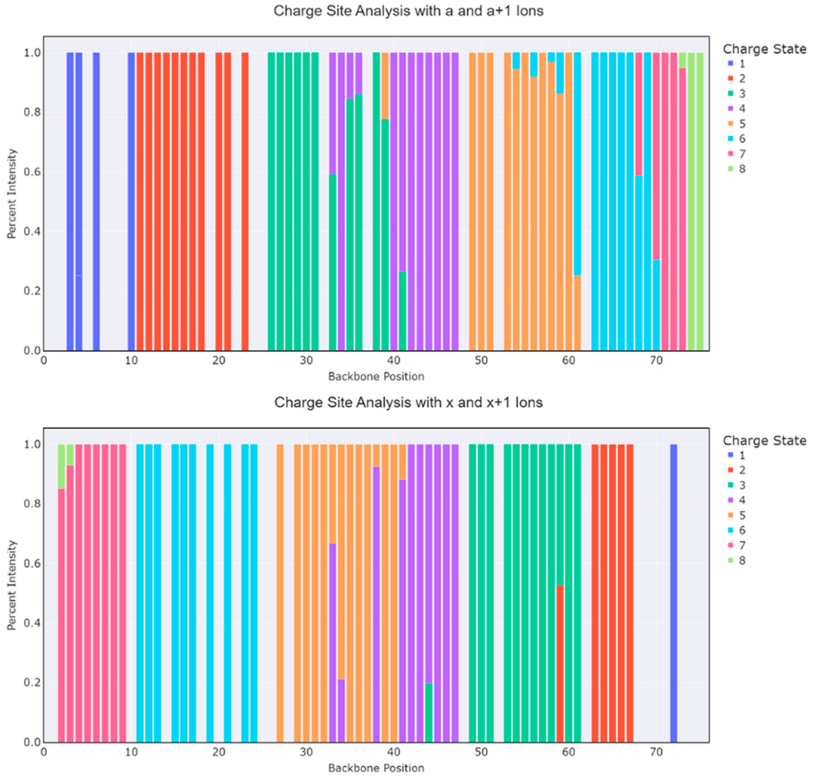


Figure 4. Charge site analysis feature displays the percentage of fragments in each charge state according to the backbone position cleaved. This analysis is performed using the $a/a + 1$ ions (top) or $x/x + 1$ ions (bottom). Data is shown for ubiquitin (8+) activated using 193 nm UVPD (1 pulse, 2 mJ).

https://doi.org/10.1021/acs.jproteome.2c00594
J. Proteome Res. 2023, 22, 546−550

AUTHOR INFORMATION

Corresponding Author

Jennifer S. Brodbelt − *Department of Chemistry, University of Texas at Austin, Austin, Texas 78712, United States;* orcid.org/0000-0003-3207-0217; Email: jbrodbelt@cm.utexas.edu

Author

Kyle J. Juetten − *Department of Chemistry, University of Texas at Austin, Austin, Texas 78712, United States*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jproteome.2c00594

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The author acknowledge support from the NSF (CHE-2203602), and the Robert A. Welch Foundation (F-1155).

REFERENCES

(1) Cui, W.; Rohrs, H. W.; Gross, M. L. Top-Down Mass Spectrometry: Recent Developments, Applications and Perspectives. *Analyst* **2011**, *136* (19), 3854−3864.

(2) Kelleher, N. L.; Lin, H. Y.; Valaskovic, G. A.; Aaserud, D. J.; Fridriksson, E. K.; McLafferty, F. W. Top Down versus Bottom Up Protein Characterization by Tandem High-Resolution Mass Spectrometry. *J. Am. Chem. Soc.* **1999**, *121* (4), 806−812.

(3) Tyers, M.; Mann, M. From Genomics to Proteomics. *Nature* **2003**, *422* (6928), 193−197.

(4) Catherman, A. D.; Skinner, O. S.; Kelleher, N. L. Top Down Proteomics: Facts and Perspectives. *Biochem. Biophys. Res. Commun.* **2014**, *445* (4), 683−693.

(5) Durbin, K. R.; Fornelli, L.; Fellers, R. T.; Doubleday, P. F.; Narita, M.; Kelleher, N. L. Quantitation and Identification of Thousands of Human Proteoforms below 30 KDa. *J. Proteome Res.* **2016**, *15* (3), 976−982.

(6) Smith, L. M.; Kelleher, N. L. Proteoform: A Single Term Describing Protein Complexity. *Nat. Methods* **2013**, *10* (3), 186−187.

(7) Siuti, N.; Kelleher, N. L. Decoding Protein Modifications Using Top-down Mass Spectrometry. *Nat. Methods* **2007**, *4* (10), 817−821.

(8) Toby, T. K.; Fornelli, L.; Kelleher, N. L. Progress in Top-Down Proteomics and the Analysis of Proteoforms. *Annu. Rev. Anal. Chem.* **2016**, *9* (1), 499−519.

(9) Riley, N. M.; Westphall, M. S.; Coon, J. J. Activated Ion Electron Transfer Dissociation for Improved Fragmentation of Intact Proteins. *Anal. Chem.* **2015**, *87* (14), 7109−7116.

(10) Brodbelt, J. S. Photodissociation Mass Spectrometry: New Tools for Characterization of Biological Molecules. *Chem. Soc. Rev.* **2014**, *43* (8), 2757−2783.

(11) Lantz, C.; Zenaidee, M. A.; Wei, B.; Hemminger, Z.; Ogorzalek Loo, R. R.; Loo, J. A. ClipsMS: An Algorithm for Analyzing Internal Fragments Resulting from Top-Down Mass Spectrometry. *J. Proteome Res.* **2021**, *20* (4), 1928−1935.

(12) Fellers, R. T.; Greer, J. B.; Early, B. P.; Yu, X.; LeDuc, R. D.; Kelleher, N. L.; Thomas, P. M. ProSight Lite: Graphical Software to Analyze Top-Down Mass Spectrometry Data. *Proteomics* **2015**, *15* (7), 1235−1238.

(13) Zamdborg, L.; LeDuc, R. D.; Glowacz, K. J.; Kim, Y.-B.; Viswanathan, V.; Spaulding, I. T.; Early, B. P.; Bluhm, E. J.; Babai, S.; Kelleher, N. L. ProSight PTM 2.0: Improved Protein Identification and Characterization for Top down Mass Spectrometry. *Nucleic Acids Res.* **2007**, *35* (Web Server), W701−W706.

(14) Fornelli, L.; Srzentić, K.; Huguet, R.; Mullen, C.; Sharma, S.; Zabrouskov, V.; Fellers, R. T.; Durbin, K. R.; Compton, P. D.; Kelleher, N. L. Accurate Sequence Analysis of a Monoclonal Antibody by Top-Down and Middle-Down Orbitrap Mass Spectrometry Applying Multiple Ion Activation Techniques. *Anal. Chem.* **2018**, *90* (14), 8421−8429.

(15) Wu, Z.; Roberts, D. S.; Melby, J. A.; Wenger, K.; Wetzel, M.; Gu, Y.; Ramanathan, S. G.; Bayne, E. F.; Liu, X.; Sun, R.; Ong, I. M.; McIlwain, S. J.; Ge, Y. MASH Explorer: A Universal Software Environment for Top-Down Proteomics. *J. Proteome Res.* **2020**, *19* (9), 3867−3876.

(16) Dhenin, J.; Borges Lima, D.; Dupré, M.; Chamot-Rooke, J. TDFragMapper: A Visualization Tool for Evaluating Experimental Parameters in Top-down Proteomics. *Bioinformatics* **2022**, *38* (4), 1136−1138.

(17) Wysocki, V. H.; Tsaprailis, G.; Smith, L. L.; Breci, L. A. Mobile and Localized Protons: A Framework for Understanding Peptide Dissociation. *J. Mass Spectrom* **2000**, *35* (12), 1399−1406.

(18) Dongré, A. R.; Jones, J. L.; Somogyi, Á.; Wysocki, V. H. Influence of Peptide Composition, Gas-Phase Basicity, and Chemical Modification on Fragmentation Efficiency: Evidence for the Mobile Proton Model. *J. Am. Chem. Soc.* **1996**, *118* (35), 8365−8374.

(19) Qin, J.; Chait, B. T. Preferential Fragmentation of Protonated Gas-Phase Peptide Ions Adjacent to Acidic Amino Acid Residues. *J. Am. Chem. Soc.* **1995**, *117* (19), 5411−5412.

(20) Rosenberg, J.; Parker, W. R.; Cammarata, M. B.; Brodbelt, J. S. UV-POSIT: Web-Based Tools for Rapid and Facile Structural Interpretation of Ultraviolet Photodissociation (UVPD) Mass Spectra. *J. Am. Soc. Mass Spectrom.* **2018**, *29* (6), 1323−1326.

(21) O'Brien, J. P.; Li, W.; Zhang, Y.; Brodbelt, J. S. Characterization of Native Protein Complexes Using Ultraviolet Photodissociation Mass Spectrometry. *J. Am. Chem. Soc.* **2014**, *136* (37), 12920−12928.

(22) Morrison, L. J.; Brodbelt, J. S. Charge Site Assignment in Native Proteins by Ultraviolet Photodissociation (UVPD) Mass Spectrometry. *Analyst* **2016**, *141* (1), 166−176.