Journal of proteome research

pubs.acs.org/jpr

Article

Impact of Internal Fragments on Top-Down Analysis of Intact Proteins by 193 nm UVPD

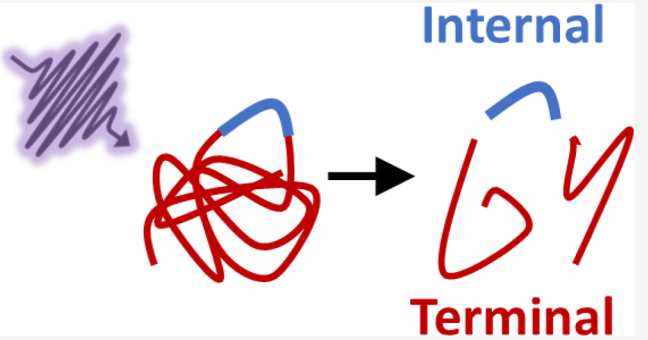
Sean D. Dunham, Benqian Wei, Carter Lantz, Joseph A. Loo, and Jennifer S. Brodbelt*

Cite This: *J. Proteome Res.* 2023, 22, 170−181

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: 193 nm ultraviolet photodissociation (UVPD) allows high sequence coverage to be obtained for intact proteins using terminal fragments alone. However, internal fragments, those that contain neither N- nor C- terminus, are typically ignored, neglecting their potential to bolster characterization of intact proteins. Here, we explore internal fragments generated by 193 nm UVPD for proteins ranging in size from 17−47 kDa and using the ClipsMS algorithm to facilitate searches for internal fragments. Internal fragments were only retained if identified in multiple replicates in order to reduce spurious assignments and to explore the reproducibility of internal fragments generated by UVPD. Inclusion of internal fragment improved sequence coverage by an average of 18% and 32% for UVPD and HCD, respectively, across all proteins and charge states studied. However, only an average of 18% of UVPD internal fragments were identified in two out of three replicates relative to the average number identified across all replicates for all proteins studied. Conversely, for HCD, an average of 63% of internal fragments were retained across replicates. These trends reflect an increased risk of false-positive identifications and a need for caution when considering internal fragments for UVPD. Additionally, proton-transfer charge reduction (PTCR) reactions were performed following UVPD or HCD to assess the impact on internal fragment identifications, allowing up to 20% more fragment ions to be retained across multiple replicates. At this time, it is difficult to recommend the inclusion of the internal fragment when searching UVPD spectra without further work to develop strategies for reducing the possibilities of false-positive identifications. All mass spectra are available in the public repository jPOST with the accession number JPST001885.



KEYWORDS: *top-down, mass spectrometry, ultraviolet photodissociation (UVPD), high-energy collisional dissociation (HCD), terminal fragment, internal fragment, proton-transfer charge reduction (PTCR) reaction*

INTRODUCTION

With the introduction of high-resolution mass spectrometers, the ability to apply top-down mass spectrometry to elucidate sequences of intact proteins *via* various MS/MS methods has advanced significantly,[1−5] and many impressive applications have been reported.[6−10] For example, 21 T FTICR mass spectrometers have achieved isotopic resolution for proteins larger than 100 kDa and have greatly facilitated the assignment of fragment ions in complicated MS/MS spectra.[6−11] Machine learning and other sophisticated algorithms have also improved the speed and accuracy of deconvolution of protein mass spectra.[12,13] Despite these advances, generating informative and interpretable MS/MS spectra for large proteins remains difficult owing to spectral congestion[14] and decreased fragment ion signal-to-noise[4] relative to smaller proteins. Many peaks are not assigned in top-down mass spectra, representing an untapped resource. Additionally, the performance of traditional collision-based methods is frequently lackluster for the sequence characterization of intact proteins. Other ion activation methods such as electron-based dissociation[3,4,15] and ultraviolet photodissociation (UVPD)[5,14,16−19] have been shown to achieve more efficient fragmentation, resulting in higher sequence coverages, but the large number of still unassigned peaks remains an issue. All of these ion activation methods generate internal fragment ions, products that do not contain either the N- or C-termini of the protein. Internal fragment ions are often ignored due to the computational difficulty to account for them and the potential high rate of false positives.[20]

The production of internal fragments was recognized even in some of the earlier MS studies of intact proteins and peptides.[21,22] Although terminal fragment ions are often sufficient for identification of proteins, inclusion of internal

Received: September 26, 2022
Published: December 12, 2022

fragment ions can help improve coverage in the middle section of protein sequences where assignable N- and C-terminal fragment ions are sparse. For example, inclusion of internal fragments and other non-canonical cleavages led to a 45% increase sequence coverage of intact proteins when performing collision-induced dissociation (CID).[23] The inclusion of internal fragment ions also resulted in an increase in the total percentage of matched fragment ions from 40 to 75% for beam-type CID of ubiquitin.[24] Statistical analysis of top-down MS data has revealed a bias for terminal fragment formation over internal fragments for data derived from UVPD and higher-energy collisional dissociation (HCD).[25] Interestingly, more of the ion current was shown to contribute to terminal fragment ion formation for UVPD compared to HCD.[25] In the same study, formation of internal fragment ions was shown to be less prominent for large proteins compared to small proteins, despite the increase in the number of unique internal fragment ions.[25] This non-intuitive outcome was rationalized by considering that the chance of generating a specific internal fragment ion decreases with increasing size of the protein, thus decreasing the ability to identify a given internal fragment ion.[25] Despite the potential gain in information when considering internal fragments, internal fragments are typically ignored in proteomic studies owing to the high computational demands to search for them, the lack of appropriate software, and the need for high-resolution mass spectrometers to confidently assign internal fragment ions. Additionally, inclusion of internal fragments often leads to ambiguous assignments, in that a single theoretical mass can potentially be mapped to multiple internal fragments.[20] The enormous number of possible assignments that can fit a molecular composition necessitates a high mass measurement accuracy (e.g., <2 ppm or even better) to reduce ambiguous assignments and false positives when searching for internal fragments.

Ambiguous internal fragment ions can be classified into three categories: arrangement ambiguity in which the same mass maps to different segments of the protein sequence, frameshift ambiguity in which the middle portion of the sequence is identical with different flanking amino acids possessing the same chemical compositions, or mass accuracy ambiguity, in which a mass is assignable with perfect mass accuracy but not with the mass error used.[20] The prevalent source of ambiguity, frameshift ambiguity, can be reduced by extracting the common sequence from the confounding internal fragment ions with all but the flanking amino acids uncertain. Implementing this optimization protocol resulted in a 17% sequence coverage improvement for CID of superoxide dismutase.[20]

Recently, curation of internal fragments has been given more attention and has been shown to improve sequence coverage of proteins when properly assigned.[15,20,26−28] These initial studies focused on the production of internal fragment ions by CID,[20,26,27] electron ionization dissociation (EID),[15,26,27] and 213 nm UVPD.[25,26] The latter method, UVPD, might be expected to be especially prone to extensive generation of internal fragment ions because it is a high-energy activation method that produces a variety of fragment ion types (a/x, b/y, and c/z and variations thereof).[5,29−31] The significant fragmentation of proteins promoted by UVPD comes at the cost of low S/N of product ions, increased spectral congestion, and the risk of secondary dissociation, the latter of which contributes to the prevalence of internal fragment ions.

Using longer laser irradiation times or more laser pulses during UVPD exacerbates the occurrence of secondary dissociation and generation of internal fragment ions. Auxiliary methods have been developed to counter secondary dissociation during UVPD and infrared multiphoton dissociation (IRMPD) when irradiating proteins in ion trap cells.[32−34] These strategies entail the application of low amplitude broadband dipolar excitation waveforms in which frequencies corresponding to the precursor protein ions are omitted, allowing for selective excitation of fragment ions in a manner that moves them away from the path of the laser. In one iteration of this methodology, termed fragment ion protection (FIP), the S/N of fragment ions was increased by up to a factor of 3.5, and sequence coverage of proteins smaller than 17 kDa was increased by up to 16%.[33] In another application of FIP that focused on intact proteins (29−55 kDa), the increase in the signal intensity enabled the identification of additional ions originating from backbone cleavages in the mid-sections, where sequence coverage is often lacking.[34] Even if internal fragments are accounted for, internal fragment generation also leads to increased spectral congestion, which is particularly problematic for UVPD, as a result of the plethora of fragment ion types generated compared to other methods of ion activation.

Spectral congestion is a major drawback of top-down analysis of large proteins, and ion−ion reactions may be used to manipulate charge states of ions and alleviate the spectra density.[6,14,35−43] Spectral congestion has been reduced following activation of intact proteins by using proton-transfer charge reduction (PTCR) reactions to lower the charge states of fragment ions and disperse them over a broad m/z range.[6,14,35,39,43] In recent studies, this process decreased spectral congestion and improved sequence coverage of proteins after UVPD[14] and ETD.[39] Ion charge states can be further manipulated through an auxiliary method known as ion parking, inhibiting the rate of ion−ion reactions by application of ion-selective waveforms.[37,38] In particular, ion parking can be performed during PTCR to consolidate a signal into one or fewer charge states to simultaneously improve signal and decrease spectral congestion.[11,40,41] Recently, a rapid parking strategy was implemented with PTCR to improve reaction rates by 5−10X without any compromise in parking efficiency.[11] This strategy was showcased in an LCMS analysis of cell lysate protein fractions derived from *H. sapiens* MCF7 cell lysates and improved the number of proteoform identifications from 1404 without the PTCR rapid park strategy to 2472 with the PTCR rapid park method.[11]

Despite the aforementioned methods to mitigate spectral congestion and secondary dissociation, there are numerous unassigned peaks in UVPD mass spectra, suggesting that a deeper inspection of internal fragment ions is warranted. Recently, the ClipsMS algorithm was developed and used to aid the search for internal fragments.[15,26−28] In these studies, internal fragments were shown to increase sequence coverage of all proteins studied when using an array of ion activation methods including EID, CID, electron capture dissociation (ECD), and 213 nm UVPD.[26,27] Notably, inclusion of internal fragments expanded sequence coverage in the center section of proteins where coverage was lacking when only considering terminal fragments.[26,27] In another application, inclusion of internal fragments enabled the identification of 7% more proteoform spectral matches in a MCF7[44] top-down proteomic data set.[45] The availability of a breakthrough

https://doi.org/10.1021/acs.jproteome.2c00583
J. Proteome Res. 2023, 22, 170−181

algorithm like ClipsMS makes it possible to evaluate internal fragment ions produced by other MS/MS methods, such as 193 nm UVPD. The 193 nm UVPD differs from 213 nm UVPD, in that protein amide bonds absorb more strongly around 193 nm.[46] Furthermore, 213 nm UVPD is most often undertaken using a lower power solid-state laser, requiring the use of longer activation periods (more laser pulses) than that of 193 nm and potentially increasing the risk of secondary dissociation. The aim of the present study was to further assess the contributions of internal fragments to top-down mass spectra of intact proteins generated by 193 nm UVPD.

METHODS

Materials

Lyophilized myoglobin (16 kDa, 153 amino acids) from equine skeletal muscle, bovine carbonic anhydrase II (29 kDa, 259 amino acids), aldolase from rabbit muscle (39 kDa, 363 amino acids), and enolase from baker's yeast (46 kDa, 436 amino acids) were purchased from Millipore Sigma and dissolved in 50% methanol with 0.1% formic acid at a concentration of 5−10 μM. Each solution was directly infused into the mass spectrometer using a heated electrospray ionization (HESI) source at a flow rate of 3 μl/min and using applied voltages ranging from 3.2−4.5 kV.

Mass Spectrometry

Experiments were performed on a Thermo Fisher Scientific Orbitrap Fusion Lumos Tribrid mass spectrometer (San Jose, CA) modified to implement 193 nm UVPD in the dual linear ion trap using a 500 Hz excimer laser (Coherent Excistar XS) as described previously.[47] The mass spectrometer was also modified to implement PTCR using the reagent perfluoroperhydrophenanthrene (PFPP), similar to that described previously.[36,41] For PTCR, reaction conditions were varied as needed to reduce the charge states of fragment ions and thus spread them up to a range of m/z 3500. The PTCR reaction time ranged from 5 to 100 ms. In some cases, up to three rounds of charge reduction were performed by repeated injections of the PFPP reagent ion (2×10^5 charges), a method described previously.[48] UVPD was performed primarily in the low-pressure cell (LPT) of the dual linear ion trap. The pressure in the ion routing multipole was reduced to 0.002 torr to optimize transmission and detection of fragment ions. The instrument was calibrated on each day of use to help maintain a high mass accuracy of <2 ppm. Furthermore, Easy-IC, an internal calibration procedure using fluoranthene as an internal calibrant, was used to assist in maintaining high mass accuracy. UVPD mass spectra were collected with a resolution of 240,000 at m/z 400, an automatic gain target (AGC) of 1×10^6 charges and using full profile mode. HCD spectra were collected with a 1×10^5 AGC target and NCE 20 unless stated otherwise. For UVPD and HCD experiments, 200 transients were averaged for myoglobin and carbonic anhydrase and 300 or 400 transients were averaged for aldolase and enolase, respectively, to maximize S/N of fragment ions.

Data Analysis

Xtract was used to deconvoluted mass spectra using a fit factor of 44% and S/N threshold of 10. ClipsMS was used to search for terminal and internal fragments using an error tolerance of 2 ppm and smallest internal fragment size of 5 amino acids.[27] Hydrogen gains and losses expected for UVPD ions were searched for all fragments. Other neutral losses were not considered. To accommodate hydrogen gains and losses in ClipsMS searches, unlocalized modifications were searched on all residues regardless of the residue at the cut site. Consequently, these modifications were applied without consideration of ion type, allowing improbable fragments, such as $b + 1$ among others, to be found. A script in R was used to filter out all improbable fragments and ambiguous assignments with the exception of frameshift ambiguous internal fragments, those where flanking amino acids of the fragment ion are uncertain. Additionally, frameshift internal fragments were only considered with a shift of up to two amino acids. Motivated by a previous study,[20] frameshift ambiguous fragments were redefined as their unambiguous sequence as a means to address them in a consistent manner. Furthermore, only fragments identified in two of three replicates were retained in order to reduce spurious assignments. Fragment ion searches were performed on a computer with an AMD Ryzen 7 2700 processor (8-core, 3200 MHz) with 16 GB of RAM. Fragment ion isotope distributions were simulated using the averaging model and a script in MATLAB. P-scores[49] were calculated by the software ProSight Lite and only considered terminal fragment ions. In general, higher p-scores indicate a higher potential chance of false positives and thus are less favorable than lower p-scores. Intensities of ions were normalized to the TIC. Internal fragment ions are labeled using the format $[\text{ion type}]_{[\text{start amino acid}]-[\text{end amino acid}]}^{[\text{charge state}]}$ with the ion types being *ax*, *ay*, *az*, *bx*, *bz*, *cx*, and *cy*, a notation adopted in a previous study.[28] For example, the ion type *ax* means *a*/*x* type cleavage at the two ends of the internal fragment, and thus, $ax_{10\text{-}90}$ signifies an ion containing 81 residues that runs from the 10th amino acid to the 90th amino acid.

Internal fragment searches necessitate a tight error tolerance of <2 ppm to mitigate ambiguous and false-positive assignments at the cost of total number of fragment ion assignments. For instance, UVPD spectra were collected for carbonic anhydrase (25+, m/z 1161.9) in full profile and with a laser energy of 1.5 mJ and 1 pulse. An average sequence coverage of 75 ± 1% was obtained when using a fragment ion mass tolerance of 10 ppm and considering just terminal fragment ions (based on 3 replicates), whereas a sequence coverage of 62 ± 3% was obtained when utilizing a mass tolerance of 2 ppm. Many of the fragments identified at error tolerances >2 ppm appeared to be valid assignments based on their isotopic profiles, suggesting the inability to maintain high mass accuracy for all fragment ions and the loss of potentially valid assignments. Additionally, previous work has shown a high degree of internal fragment ion ambiguity when using higher error tolerances in internal fragment ion searches,[20] necessitating a need for tight error tolerances of 2 ppm or lower. A relatively high AGC target of 1×10^6 was used for UVPD MS/MS data collection. The use of Easy-IC helped maintain a high degree of mass accuracy for assignment of UVPD fragment ions using the 1×10^6 AGC target, allowing for a greater number of fragment ion identifications than possible with a lower AGC target. Full profile spectral acquisition allowed for more fragment ion identifications compared to reduced profile acquisition at the cost of an increased risk of false-positive identifications. A higher S/N threshold was used during deconvolution to minimize false-positive identifications. To optimize the S/N threshold value, UVPD spectra of carbonic anhydrase (25+) were deconvoluted at multiple S/N values and plotted against −log(p-score), as shown in Figure S1. The

https://doi.org/10.1021/acs.jproteome.2c00583
J. Proteome Res. 2023, 22, 170−181

average −log (p-score) increased with S/N and plateaued around an S/N threshold of 10, motivating the use of a S/N threshold of 10 when deconvoluting spectra. This method of optimization the S/N threshold when deconvoluting was reported in a previous study and thus used for consistency here.[50]

RESULTS AND DISCUSSION

UVPD Internal Fragment Ion Types

UVPD generates a plethora of fragment ion types, including *a/x* (cleavage of C_α—C backbone bonds), *b/y* (cleavage of C−N amide backbone bonds), and *c/z* (cleavage of N—C_α bonds), some also accompanied by hydrogen migrations. Owing to the potential cleavage of many backbone bonds upon UVPD, a variety of internal fragment ion types may be generated, including *ax, ay, az, bx, by, bz, cx, cy,* and *cz*. If the backbone cleavages occur between the same pairs of residues, the resulting *ax, by,* and *cz* ions have the same mass, and therefore, all three products will simply be treated as *ax* fragments. Consideration of these product ion types results in a large search space (and considerable search time) for assigning internal fragments. UVPD spectra were collected for four proteins and searched for possible internal fragment ions. To confirm the assignment of internal fragment ions, theoretical isotope distributions generated from the averagine model were overlaid for isotope distributions of suspected internal fragment ions, as illustrated for several product ions of myoglobin (*m/z* 1130.9, 15 + charge state), as shown in Figure S2. Simulated isotope distributions aligned with the observed isotope distribution, and therefore, the following ion types were considered for the rest of the study: *a*, *a* + 1, *b*, *x*, *x* + 1, *y*, *y*−1, *y*−2, *z*, *ax*, *ax* + 1, *ax* + 2, *ax* + 3, *ay*, *ay* + 1, *ay* + 2, *az*, *az* + 1, *bx*, *bx* + 1, *bx*−1, *bx*−2,, *bz*, *bz*−1, *bz*−2, *cx*, *cx* + 1, and *cy*. The *ax* and *ay* ions assigned by ClipsMS searches were also 2H greater than expected based on the preliminary results, but the difference in mass is attributed to hydrogen migrations common during UVPD. Therefore, *ax*−2 and *ay*−2 ions identified by ClipsMS were redefined as *ax* and *ay*, respectively. Additional neutral losses were not included due to the already large computational load of performing internal fragment ion searches for UVPD data, which for larger proteins like enolase may take several hours.

Impact of Laser Pulses and Energy on Internal Fragmentation

Top-down 193 nm UVPD experiments are frequently performed using a single laser pulse with the expectation that multiple laser pulses may in some cases convert large sequence ions into smaller (less informative) ions or cause excessive internal fragmentation. Therefore, the impact of the number of laser pulses on the degree of internal fragmentation was explored. UVPD spectra were collected for myoglobin (*m/z* 848.5, 20 + charge state) using 1 to 5 laser pulses (1.5 mJ per pulse), and those fragment ions identified in two out of three replicates were classified as terminal (containing either the N- or C-terminus) or internal (containing neither the N- or C-terminus). The sequence coverages obtained without and with the inclusion of internal fragments are plotted in Figure 1A. Sequence coverages when considering only terminal fragment ions decreased as the number of laser pulses increased, but the sequence coverage slightly increased when including internal fragments. The abundances of terminal and internal fragment ions were summed, and the proportions of the terminal versus

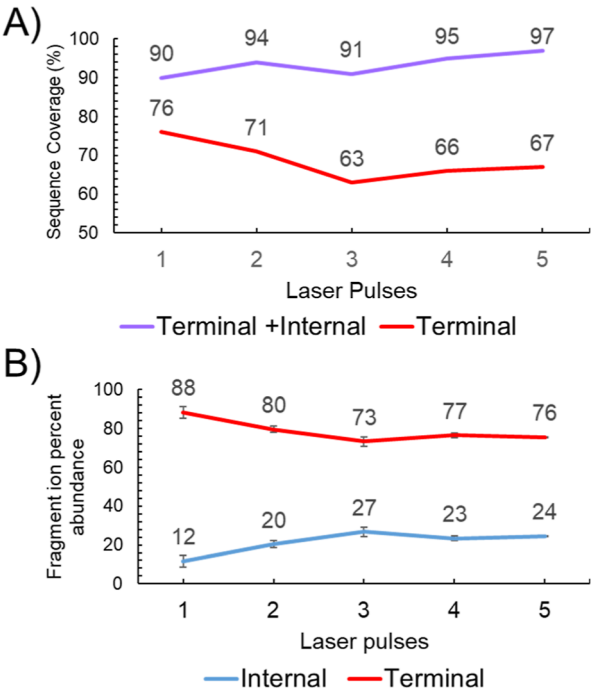


Figure 1. (A) Sequence coverage obtained without and with the inclusion of internal fragments. (B) Proportion of terminal and internal fragment ions based on summed ion abundances. UVPD spectra were collected for myoglobin (20+) at a laser energy of 1.5 mJ and using 1−5 laser pulses. Only fragment ions identified in at least two of three replicates were retained.

internal fragment ion abundance are plotted in Figure 1B. The abundance and overall proportion of internal fragment ions increased with the number of laser pulses up to 3 pulses, then plateaued around 25% abundance of internal fragment ions for 3, 4, or 5 laser pulses. Violin plots were also generated to illustrate the distribution of fragment ion sizes (mass) produced using 1, 3, or 5 laser pulses, as shown in Figure 2, based on the number of fragment ions. As the number of laser pulses increased, the distribution of fragment ion masses shifted to lower masses. The collection of results in Figures 1 and 2 confirms that the internal fragmentation plays an increasing role when utilizing multiple laser pulses.

The impact of the laser energy on the prevalence of internal fragmentation was also evaluated. The overall flux of photons increases with laser energy, thus increasing the chances for multiphoton absorption and the possibility of internal fragmentation. UVPD spectra were collected for carbonic anhydrase (*m/z* 1161.9, 25 + charge state) with one laser pulse, while the laser energy was varied from 0.5 to 2.5 mJ per pulse. Both the internal fragment ion sequence coverage (Figure S3A) and the proportion of internal fragment ion abundance (Figure S3B) increased as a function of laser energy, consistent with increased secondary dissociation of terminal fragment ions and their conversion into internal fragment ions. Additionally, violin plots for the fragment ion size distribution were generated for UVPD spectra of carbonic anhydrase (*m/z* 1161.9, 25 + charge state) at 0.5, 1.5, and 2.5 mJ, as displayed in Figure S4. The mass distribution exhibited a trend to smaller fragment ions as the laser energy increased, consistent with the expectation that higher photon fluxes would lead to more substantial internal fragmentation.

https://doi.org/10.1021/acs.jproteome.2c00583
J. Proteome Res. 2023, 22, 170−181

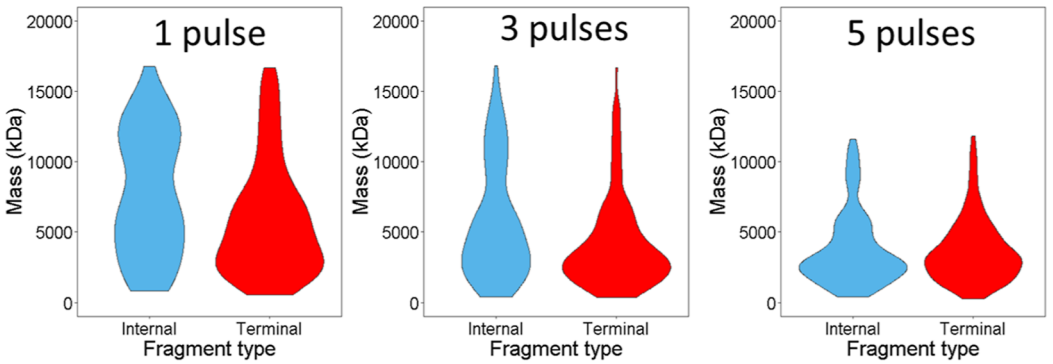


Figure 2. UVPD spectra were collected for myoglobin (20+, m/z 848.5) at a laser energy of 1.5 mJ and using 1, 3, or 5 laser pulses. Only fragment ions identified in at least two of three replicates were retained. Violin plots were generated from the number of terminal and internal fragments as a function of their masses.

Impact of Internal Fragment Ions on Sequence Coverage

Sequence maps were constructed for four proteins based on the UVPD mass spectra without and with internal fragment ions (Figure S5). Including internal fragments allowed for the identification of more UVPD fragment ions than when considering terminal fragment ions alone. For instance, the average number of fragment ions identified for UVPD of carbonic anhydrase (m/z 1161.9, 25 + charge state) using 1.5 mJ and one pulse increased from 237 ± 29 to 692 ± 11 upon the consideration of non-ambiguous and frameshift ambiguous internal fragments. However, the total number of identified fragment ions decreased to 284 when retaining only those ions found in at least two of three replicates. Specifically, the total number of internal fragments decreased from an average of 455 ± 21 to 74, and the total number of terminal fragment ions decreased from an average of 237 ± 30 to 210. Therefore, many internal fragment identifications are not consistently reproduced in UVPD spectra, and deploying more rigorous constraints, like requiring duplication of fragment ions in multiple data sets, increases confidence in assignments.

Internal fragment ions are expected to have lower S/N than terminal fragment ions because they require two cleavage events, reducing their probability of formation. To ascertain whether internal fragments were not being reliably identified due to low S/N, the S/N values of all fragment ions in a UVPD spectrum of carbonic anhydrase (25+, 1.5 mJ, 1 pulse, 1 replicate) were tabulated based on the S/N of the most abundant m/z value in each isotopic profile as determined by Xtract for all fragment ions identified by ClipsMS. Box plots overlaid with dot plots were generated to illustrate the distribution of S/N values of fragment ions in Figure S6. The average S/N of terminal and internal fragment ions was 74 and 42, respectively, based on a total of 542 terminal ions and 563 internal ions. Fragment ions that were not identified in multiple replicates were removed, and the average S/N of internal fragments retained (104 ions) and removed (459 ions) was determined to be 60 and 38, respectively. These trends suggest that some internal fragments are not being reliably identified as a consequence of low S/N. Furthermore, the fit factor, that is the quality of an isotope distribution as reported by Xtract, was determined for all of the same fragment ions. On average, the fit factor was 88 and 78 for terminal and internal fragment ions, respectively. Thus, internal fragment identifications were of lower quality and potentially have a higher risk of mis-assignment. For this aspect of the analysis, deconvolution was performed with consideration of overlap of isotope distributions, allowing for more generous scoring of isotope clusters. However, none of these comparisons were of statistical significance (i.e., averages were within one standard deviation of one another), and we are unable to draw confident conclusions regarding S/N and fit factor as a contributor to the inability to reliably identify the majority of internal fragment ions. Using stricter isotope scoring and S/N thresholds may be valuable in future internal ion fragment studies. The complexity of UVPD spectra is likely a contributing factor, in that it may lead to a higher risk of mis-assignment, especially when considering the increased size of the theoretical mass list size for fragment ions when including internal fragments.

Furthermore, 12% of internal fragments exhibited frameshift ambiguity, leading to additional uncertainty in the exact cleavage positions of some internal fragments. However, since the middle section of these fragments is known, we believe that these fragments are still valuable and can assist in characterization of proteins when considered in conjunction with other fragment ions. Nevertheless, inclusion of internal fragments in 193 nm UVPD improved sequence coverage for all proteins studied, as summarized in Figure 3A, with the gain in sequence coverage averaging 15% for the four proteins analyzed.

Terminal and internal fragment ions were counted based on the backbone positions cleaved to generate the fragment ions for each protein and are displayed in Venn diagrams, as shown in Figure 3B–E. Interestingly, some of the backbone cleavage sites were uniquely observed for internal fragment ions, indicating that internal fragments provide both new and shared cleavage sites. Across all four proteins, an average of 33% of the backbone cleavage sites were common to both terminal and internal fragment ions. One may expect more overlap between the cleavage sites that lead to terminal and internal ions in UVPD, especially given that significant overlap between terminal and internal fragments has been observed upon CID.[28] This overlap may be tempered in UVPD by the inability to assign all generated fragment ions as a consequence of low S/N stemming from multiple fragmentation pathways and spectral congestion. Furthermore, these mitigating issues are exacerbated as protein size increases and the risk of false assignments increases.

Terminal and internal fragments provided both unique and shared coverage across all proteins. In particular, the inclusion of internal fragments expanded sequence coverage in the middle sequence section of large proteins, exemplified by the sequence coverage plots shown in Figure S5, where there are few terminal fragment ions originating from backbone

https://doi.org/10.1021/acs.jproteome.2c00583
J. Proteome Res. 2023, 22, 170−181

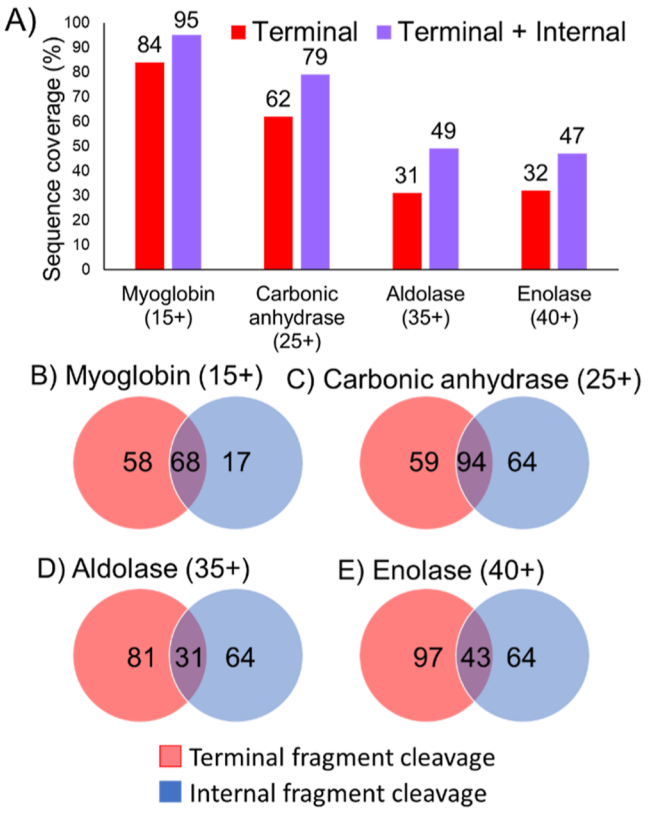


Figure 3. (A) Sequence coverage obtained with and without internal fragments. (B,C,D,E) Backbone cleavage locations corresponding to the fragment ions produced were counted and binned as terminal fragments, internal fragments, or shared by both. UVPD spectra were collected for myoglobin (15+), carbonic anhydrase (25+), aldolase (35+), and enolase (40+) using 1.5 mJ and one pulse. Fragment ions were only retained if identified in two out of three replicates.

cleavages in the central region. The fragment ions may be categorized based on whether they provide coverage of the first third of the protein sequence (N-terminal region), middle third section, or last third of the sequence (C-terminal region), as summarized in Table S1. When including internal fragment ions, gains in sequence coverage were evident in all three sections of the proteins, but the increases were most evident in the middle third and last third regions for the two larger proteins. These gains were particularly significant for aldolase (m/z 1121.3, 35 + charge state), for which sequence coverage in the middle region increased from 8 to 33% upon inclusion of internal fragment ions and from 9 to 35% for the last region.

UVPD spectra were collected for multiple charge states of each protein, and the sequence coverages and proportion of terminal fragment ions versus internal fragment ions were determined, as summarized in Figure S7. Whether considering only terminal fragment ions or including both terminal and internal fragment ions, the sequence coverage decreased modestly or remained nearly the same for each protein as charge state increased (Figure S7A,C,E,G). The abundances of all terminal and internal fragments were summed, and the proportion of terminal versus internal fragment ions was calculated for each protein in each of three charge states (Figure S7B,D,F,H). As the size of the protein increased, the proportion of internal fragment ions increased, averaging 12% for myoglobin (17 kDa, 15+) up to 22% for enolase (47 kDa, 40+) or 42% (50+). The latter trend suggests an increasing propensity for internal fragmentation with protein size and likely reflects the greater number of possible backbone sites that can be cleaved to create internal fragment ions for larger proteins. A previous study suggested that larger proteins should have additional internal fragmentation pathways, resulting in dilution of the internal fragment ion signal across more pathways.[25]

HCD and Internal Fragment Ions

The prevalence of internal fragment ions in HCD mass spectra was also explored to benchmark the impact of internal fragmentation for collisional activation versus UVPD of proteins on the same Orbitrap mass spectrometer. HCD spectra for carbonic anhydrase (25+) were collected using collision energies of 10, 20, and 30 NCE using an AGC target of 1×10^5. The sequence coverages and proportions of terminal and internal fragment ion abundances are summarized in Figure S8. Sequence coverage based on terminal fragment ions was greatest (33%) when applying 20 NCE for HCD. However, the inclusion of internal fragments resulted in a maximum sequence coverage (73%) when using a higher collision energy (30 NCE) (Figure S8A). At a low collision energy (10 NCE), for which the precursor ion was not fully depleted, internal fragment ions constituted only 7% of the identified fragment ion population (Figure S8B). However, upon increasing the NCE to 30 for which the precursor is nearly entirely depleted, the proportion of internal fragments increased substantially to 54% of the identified fragment ion abundance. Such a high degree of internal fragmentation contrasts the findings of a previous FTICR study in which internal fragments accounted for less than 20% of the fragment ion signal when performing resonance CID on a range of charge states of cytochrome *c*, myoglobin, and carbonic anhydrase and using a collisional energy sufficient to deplete the precursor to an abundance of 5% relative to the most abundant fragment ion in the MS/MS spectrum.[26] However, when performing resonance CID, an excitation waveform is utilized to excite only the precursor ion, and any resulting fragment ions that do not have the same resonance frequency are not activated. In contrast, when performing HCD (beam-type non-resonance CID), fragment ions may undergo activation and experience secondary dissociation as they travel through the HCD collision cell. This difference in the activation process would account for the difference in prevalence of internal fragment ions generated by collision-based dissociation between the FTICR and Orbitrap platforms. In another study which examined the production of internal fragment ions of ubiquitin (7+, 10+, and 13+) by HCD in an Orbitrap mass spectrometer, terminal fragments constituted the majority of the signal at typical collision energies (25 NCE).[24] However, internal fragments represented a larger proportion of signal when using higher collision energies,[24] consistent with the results presented here. Therefore, HCD conditions can be biased to favor the terminal fragment ion, internal fragment ions, or a mixture of both.

The propensity for internal fragmentation during HCD was briefly explored as a function of the collision energy based on the number of identified fragment ions. HCD spectra were collected for myoglobin (15+), carbonic anhydrase (25+), aldolase (35+), and enolase (40+) using NCE values of 10, 20, and 30, and the number of *b*, *y*, and *by* fragment ions was counted, as displayed in Figure S9. For three of the proteins, the number of internal by ions increased significantly with

https://doi.org/10.1021/acs.jproteome.2c00583
J. Proteome Res. 2023, 22, 170−181

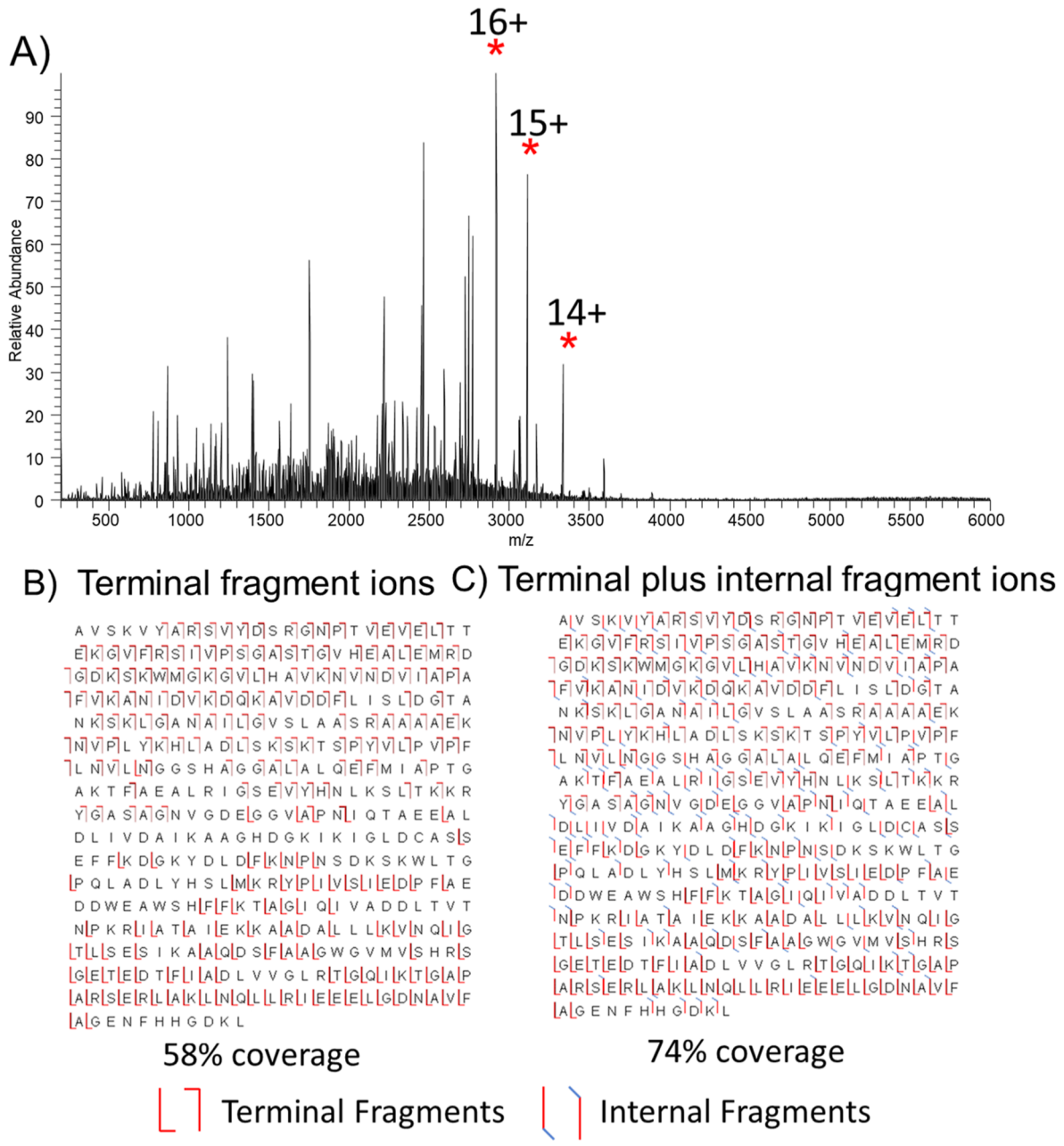


Figure 4. UVPD spectra were collected for enolase (40+, *m/z* 1167.7) using 1.5 mJ and one pulse. Following UVPD, three rounds of PTCR reactions were conducted for 50 ms each and using a reagent target of 2×10^5. Prominent charge-reduced precursor ions are labeled with asterisks. (A) Representative UVPD-PTCR spectrum. (B) Sequence coverage map obtained based on terminal fragment ions. (C) Sequence coverage map obtained based on including both terminal and internal fragment ions. Fragment ions were retained if present in two of three replicates.

increased NCE. A more in-depth study with a larger number of proteins, collision energies, and charge states would be of value to fully evaluate these trends.

HCD was performed on multiple charge states of each of the four proteins: myoglobin (15+, 20+, 25+), carbonic anhydrase (25+, 30+, 35+), aldolase (35+, 40+, 45+), and enolase (40+, 45+, 50+), and the resulting sequence coverages and portion of terminal versus internal fragment ions are displayed in Figure S10. For HCD, terminal fragment ion sequence coverage was at its lowest level for the highest charge state of each protein (Figure S10A,C,E,G), similar to the UVPD results. The proportion of internal fragment abundance generally increased with both protein size and charge state for HCD, reaching an average high of 62% for enolase (50+) (Figure S10B,D,F,H). The inclusion of internal fragments generally provided a more substantial increase in sequence coverage compared to 193 nm UVPD. For instance, inclusion of internal fragments when performing UVPD improved the sequence coverage of enolase by a factor of 1.5−1.8 depending on the charge state, whereas sequence coverage improved by a factor of 2.3−2.8 for HCD. Furthermore, a higher proportion of internal fragment ion abundance was observed for HCD (averaging 47% across all proteins and charge states) compared to UVPD (20%), based on the results shown in Figures S7 and S10. This outcome is rationalized by considering that UVPD creates a greater array of fragment ions, decreasing the probability of observing any individual internal fragment ion. Moreover, many internal fragment ions potentially generated by UVPD are not

https://doi.org/10.1021/acs.jproteome.2c00583
J. Proteome Res. 2023, 22, 170−181

reproducibly identified across multiple replicates owing to low abundance or errors in deconvolution, a less significant outcome for HCD spectra which are less congested than UVPD spectra.

The average number of terminal and internal fragment ions identified per spectrum and the number of fragments ions retained after applying a constraint that they must be observed in two of three replicates are tabulated for all proteins for both HCD and UVPD, as shown in Table S2. On average, 98% of the terminal fragment ions appear in two out of three replicates for HCD and 90% for UVPD, meaning excellent run-to-run consistency for production and identification of terminal fragment ions. However, for internal fragment ions, on average, 63% appear in two out of three replicates for HCD and only 18% for UVPD, indicating a significant drop in the reproducible generation of the internal fragment ions. The impact is magnified as a function of protein size for UVPD. For the two largest proteins (aldolase and enolase), the number of internal fragment ions per spectrum increases to nearly 700−800, yet the numbers that are reproducibly identified plummet to only 60−70 when applying the replication constraint. HCD produces a fewer number of theoretical internal fragment ions. Thus, one may expect a higher probability of generating and observing a given internal fragment ion compared to UVPD, for which the theoretical number of internal fragment ions is larger. Moreover, HCD internal fragments may exhibit a higher S/N than that of UVPD as a result of less signal dilution across multiple fragment pathways. The inability to reliably identify the majority of UVPD internal fragments in multiple replicates reinforces the concern about false positives when including internal fragment ions in searches of UVPD mass spectra.

Proton-Transfer Charge Reduction (PTCR) Reactions and UVPD Internal Fragments

PTCR reactions can be performed to decrease the charge states of fragment ions, thus dispersing them over a broader m/z range and mitigating spectral congestion.[11,14,39] Integration of UVPD and PTCR has allowed for identification of additional fragment ions following UVPD.[14] To explore the impact of PTCR on the identification of additional internal fragment ions, UVPD spectra were collected for carbonic anhydrase (m/z 1161.9, 25+), aldolase (m/z 1121.2, 35+), and enolase (m/z 1167.7, 40+) using 1.5 mJ and one pulse without and with PTCR after UVPD. To facilitate the fairest comparisons of performance, UVPD spectra were re-collected without or with PTCR on the same day, and thus, some of the sequence coverages are not identical to the previously described results. An example of a UVPD-PTCR mass spectrum for enolase (40+) is displayed in Figure 4A, and the sequence maps generated without and with inclusion of internal fragment ions are shown in Figure 4B,C. The inclusion of internal fragment ions improved the sequence coverage of enolase from 58 to 74%, substantially filling out the sequence coverage in the center of the protein. For comparison, a sequence coverage of 53% was obtained when including both terminal and internal fragments without PTCR.

Additionally, the inclusion of PTCR has a considerable impact on the mass lists used for the searches. For UVPD of enolase (40+, m/z 1167.7), the size of the deconvoluted fragment ion mass prior to any fragment ion searches decreased from 2297 (without PTCR) to 1865 (with PTCR), the resulting average sequence coverage increased from 39% (without PTCR) to 57% (with PTCR), and the average p-scores (using only terminal fragment ions) decreased from 3.0×10^{-77} to 4.0×10^{-215}. When restricting the fragment ion pool to only those found in at least two of three replicates, 247 fragment ions were identified without PTCR, but 435 were identified with PTCR (Figure S11A), revealing that PTCR allowed a much higher proportion of fragment ions to be reproducibly identified. Similar trends were also observed for carbonic anhydrase and aldolase when utilizing UVPD-PTCR, as tabulated in Figure S11A. The results indicate a decrease in the rate of false-positive identifications when utilizing UVPD-PTCR, highlighting the potential benefits of employing PTCR to uncover both terminal and internal fragments. Furthermore, UVPD-PTCR improved the sequence coverage of all three proteins when including internal fragments, as displayed in Figure S11B,C, with coverages increasing an average of 15% for UVPD-PTCR.

Interestingly, the relative proportion of abundances of terminal and internal fragment ions generally did not change significantly between the UVPD and UVPD-PTCR spectra for aldolase and enolase. For example, for UVPD and UVPD-PTCR of enolase (40+), the proportion of terminal fragment ion abundance averaged 86% and 85%, respectively, meaning that the contribution from internal fragment ions remained around 14−15% abundance.

Despite the potential dividends gained by decongesting the dense regions of UVPD mass spectra using PTCR, PTCR also offers the risk of diminishing the abundance of larger, highly charged fragment ions to the point that they may no longer be identified (insufficient S/N). Therefore, reaction time should be carefully optimized to mitigate the loss of fragment ion identifications.

Proton-Transfer Charge Reduction Reactions and Internal Fragments from HCD

HCD spectra were also collected without or with PTCR for carbonic anhydrase (m/z 1161.9, 25+), aldolase (m/z 1121.2 35+), and enolase (m/z 1167.7, 40+) using 20 NCE and a PTCR reaction time of 50 ms. Unlike with UVPD-PTCR, the average number of deconvoluted fragment ions did not change significantly for HCD-PTCR relative to HCD and the number of terminal fragments identified decreased with PTCR in all cases (Figure S12), resulting in higher p-scores for each protein when considering only terminal fragment ions. HCD spectra for proteins are less complex and less congested than UVPD spectra, and consequently, PTCR yields less notable gains or may even result in dilution of fragment ions across multiple charge states. For instance, for HCD of enolase (40+), the number of terminal fragment ions identified across multiple replicates decreased from 104 (no PTCR) to 100 (with PTCR), as shown in Figure S12A. The terminal fragment ion sequence coverage decreased for all proteins for HCD-PTCR relative to HCD alone, as displayed in Figure S12B,C. Nevertheless, PTCR did increase the number of internal fragments identified in addition to the proportion of fragment ions retained across replicates, yielding an increase in sequence coverage when considering both terminal and internal fragments. For example, the sequence coverage obtained for enolase (40+) increased from 57% for HCD to 67% for HCD-PTCR when both terminal and internal fragment ions were included. Furthermore, post-HCD-PTCR results were combined with post-UVPD-PTCR results for enolase (40+), achieving sequence coverages of 65% and 90%, without and

https://doi.org/10.1021/acs.jproteome.2c00583
J. Proteome Res. 2023, 22, 170−181

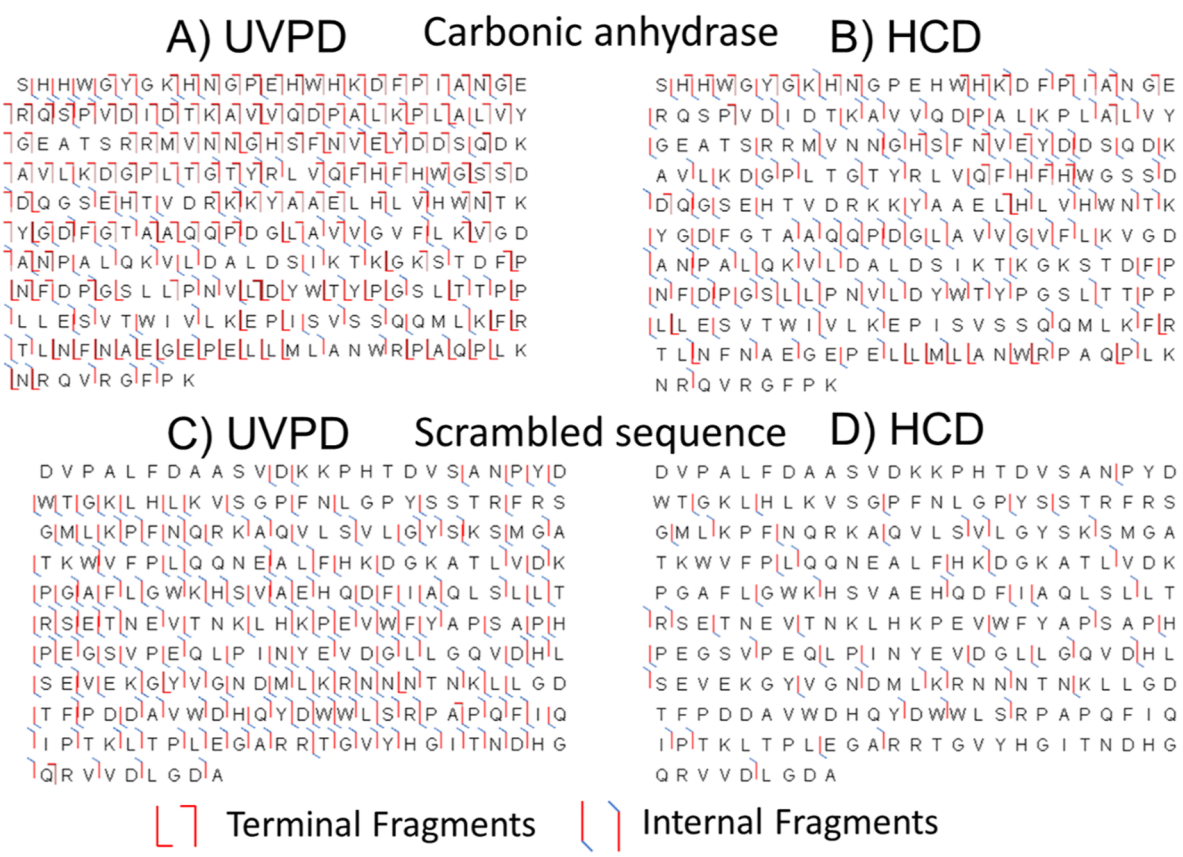


Figure 5. Sequence coverage maps for carbonic anhydrase and a scrambled sequence based on (A,C) UVPD and B,D) HCD spectra. The experimental data were collected for carbonic anhydrase (25+) using 1.5 mJ with one pulse for UVPD or 20 NCE for HCD. Fragment ions were searched using both the correct and scrambled sequences. Fragment ions were only retained if detected in two out of three replicates.

with internal fragments, respectively, showcasing the complementary nature of different activation techniques.

The reduced degree of spectral congestion in HCD spectra compared to UVPD spectra may also be reflected by the decreased average charge state of fragment ions. For MS/MS spectra acquired without and with PTCR, the average charge states of all fragment ions were determined based on fragment ions identified in multiple replicates, as summarized in Table S3 for carbonic anhydrase (25+), aldolase (35+), and enolase (40+). The average charge state of HCD fragments ions was lower than that of UVPD for all proteins.

Assessing False-Positive Identifications of Internal Fragments

Including internal fragments offers an avenue of expanding sequence coverage but comes at the increased risk of false-positive identifications. To explore this risk, internal fragments were searched in the MS/MS spectra of carbonic anhydrase (25+) using a scrambled sequence. Sequence coverage plots obtained for carbonic anhydrase and the scrambled sequence are displayed in Figure 5 based on UVPD and HCD. Primarily internal fragments and few terminal fragments were identified for the scrambled sequence, providing a sequence coverage of 65% for the scrambled sequence compared to 79% for the correct sequence for UVPD. For HCD, a sequence coverage of 21% was obtained for the scrambled sequence, while a sequence coverage of 57% was obtained for the correct sequence. Additionally, terminal fragment ion identifications were sparse for scrambled sequences where only three and zero terminal fragments were identified for UVPD and HCD, respectively. When searching for internal fragments, the library of internal fragment theoretical masses for UVPD is much larger than that for HCD, thus increasing the probability of any random mass matching an internal fragment. These comparisons suggest that internal fragments should not be considered for the purpose of protein identification, even when relatively high mass measurement accuracy is used (≤2 ppm, in this case). Currently, we propose that when performing internal fragment ions searches, more weight should be given to terminal fragment ions when developing a scoring system that reflects the confidence of a protein or proteoform assignment. In general, internal fragments could be used to bolster characterization when the identity and primary structure (i.e., sequence) of a protein are known, although caution should be used, especially for UVPD for which terminal fragments are required for confident assignments of PTMs.

Neutral Loss Considerations

Neutral losses were not considered in fragment ion searches due to the complexity of such searches. However, it is important to acknowledge that there is a risk of mapping internal fragment ions to fragment ions that originate from conventional neutral loss pathways. To assess this possibility, theoretical mass lists were generated for proteins using ProteinProspector v 6.4.0.[51] Internal fragments identified in mass spectra generated with UVPD-PTCR and HCD-PTCR of carbonic anhydrase (25+, m/z 1161.9) and enolase (40+, m/z 1167.7) were compared against ProteinProspector mass lists to determine the number of ions that could have been mapped to neutral loss fragments with higher mass accuracy. For HCD,

https://doi.org/10.1021/acs.jproteome.2c00583
J. Proteome Res. 2023, 22, 170−181

the following neutral losses were considered: loss of H_2O from S, T, E, and D residues and loss of NH_3 from R, K, Q, and N residues. For UVPD, the following neutral losses were considered: loss of H_2O, loss of NH_3 and side-chain losses resulting in *d* and *w* ions. For UVPD, no internal fragment masses were re-mapped to neutral loss products, an outcome that might have been tempered because many internal fragments were removed due to not being identified in multiple replicates. For HCD, only one and four ions were re-mapped to neutral losses for carbonic anhydrase and enolase, respectively. For example, one fragment ion (10124.2530 Da) generated from HCD of enolase could be assigned as internal fragment $by_{151\text{-}250}$ with a mass accuracy of 1.2 ppm. Alternatively, this same species could be assigned as y_{94}-NH_3 with a mass accuracy of 0.5 ppm. Overall, the overlap of internal fragment ions and neutral loss products seemed to have little impact on the trends uncovered in the present study.

CONCLUSIONS

Internal fragments are prevalent in UVPD spectra and may offer value to inform fragmentation conditions. However, the incorporation of internal fragment ions results in an increased risk of false-positive identifications. This is particularly problematic for UVPD, where the majority of internal fragment ion identifications are not reliably identified across multiple replicates, and an abundance of caution must be used when considering them. Additionally, deconvolution is known to produce spurious mass assignments, and thus, when internal fragments are considered, the probability of matching any mass to a fragment ion increases. Moreover, for UVPD, there is a larger library of theoretical masses against which fragment ions may be matched to. These concerns appear to have a less significant impact on searches that consider internal fragments generated by HCD owing to a smaller potential library of fragments, simpler mass spectra, and consequently more reliable deconvolution results, as inferred by the ability to retain more internal fragments across replicates compared to UVPD. Even so, with large proteins, as much as 50% of internal fragment identifications are not confirmed in multiple replicates. However, these concerns can be partially mitigated by utilizing PTCR after HCD or UVPD, allowing a higher percentage of internal fragment ions to be reliably identified across multiple replicates. These benefits were particularly notable for UVPD in which the percentage of all fragment ions retained in multiple replicates increased by 7−20% depending on the protein. The ability to reliably identify internal fragments across replicates may be used as another metric to inform the reliability of deconvolution algorithms.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jproteome.2c00583.

Total and regional sequence coverage of proteins following UVPD with the inclusion of internal fragment ions; average number of terminal and internal fragment ions identified and the number of terminal and internal fragment ions identified in multiple replicates with either UVPD or HCD; average charge state of HCD and UVPD spectra of carbonic anhydrase (25+), aldolase (35+), and enolase (40+) with or without PTCR; −log(p-scores) plotted as a function of S/N for UVPD of carbonic anhydrase (25+, *m/z* 1161.9); isotope distributions of selected fragment ions generated by UVPD; sequence coverages obtained with or without the inclusion of internal fragments and the percentage of terminal and internal fragment ions based on their summed abundances as a function of laser energy for UVPD of carbonic anhydrase (25+, *m/z* 1161.9); Violin plots for the distribution of fragment ion masses based off ion count following UVPD of carbonic anhydrase (25+, *m/z* 1161.9) at laser energies of 0.5, 1.5, and 2.5; sequence coverage plots with or without the inclusion of internal fragment ions for myoglobin (15+), carbonic anhydrase (25+), aldolase (35+), and enolase (40+); Box plots overlaid with dot plots for the distribution of S/N of fragment ions of a UVPD spectrum of carbonic anhydrase (25+, *m/z* 1161.9); summary of UVPD sequence coverage without or with the inclusion of internal fragment ions and the proportion of terminal versus internal fragment ions based on summed abundances as a function of charge state for all proteins; sequence coverages obtained with or without the inclusion of internal fragments and the percentage of terminal and internal fragment ions based on their summed abundances as a function of laser energy for HCD of carbonic anhydrase (25+, *m/z* 1161.9) at NCE of 10, 20, and 30 NCE; fragment ion counts for HCD spectra at NCE of 10, 20, and 30 for myoglobin 15+, carbonic anhydrase 25+, aldolase 35+, and enolase 40+; summary of HCD sequence coverage without or with the inclusion of internal fragment ions and the proportion of terminal versus internal fragment ions based on summed abundances as a function of charge state for all proteins; summary and comparison of UVPD-no PTCR and UVPD-PTCR terminal and internal fragment ion results; and summary and comparison of HCD-no PTCR and HCD-PTCR terminal and internal fragment ion results (PDF)

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

Authors

Sean D. Dunham − *Department of Chemistry, University of Texas at Austin, Austin, Texas 78712, United States;* orcid.org/0000-0002-6214-7468

Benqian Wei − *Department of Chemistry and Biochemistry, University of California Los Angeles, Los Angeles, California 90095, United States;* orcid.org/0000-0003-4853-4848

Carter Lantz − *Department of Chemistry and Biochemistry, University of California Los Angeles, Los Angeles, California 90095, United States*

Joseph A. Loo − *Department of Chemistry and Biochemistry, University of California Los Angeles, Los Angeles, California 90095, United States;* orcid.org/0000-0001-9989-1437

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jproteome.2c00583

https://doi.org/10.1021/acs.jproteome.2c00583
J. Proteome Res. 2023, 22, 170−181

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The funding from NSF (CHE-2203602), the Robert A. Welch Foundation (F-1155), and the UT System for support of the UT System Proteomics Core Facility Network is gratefully acknowledged. J.A.L. acknowledges the support from the US National Institutes of Health (R01GM103479 and R35GM145286), the US National Science Foundation (NSF) (CHE1808492), and the US Department of Energy (DE-FC02-02ER63421). C.L. acknowledges the support from the Ruth L. Kirschstein National Research Service Award program (GM007185). We thank Chad R. Weisbrod (National High Magnetic Field Laboratory), Christopher Mullen (Thermo Fisher Scientific), and John Syka (Thermo Fisher Scientific) for the advice about implementation of PFPP as a PTCR reagent. We acknowledge the repository jPost.[52]

■ REFERENCES

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

(2) Toby, T. K.; Fornelli, L.; Kelleher, N. L. Progress in Top-Down Proteomics and the Analysis of Proteoforms. *Annual Review of Analytical Chemistry* **2016**, *9*, 499−519.

(3) Riley, N. M.; Westphall, M. S.; Coon, J. J. Sequencing Larger Intact Proteins (30-70 KDa) with Activated Ion Electron Transfer Dissociation. *J. Am. Soc. Mass Spectrom.* **2018**, *29*, 140−149.

(4) Riley, N. M.; Mullen, C.; Weisbrod, C. R.; Sharma, S.; Senko, M. W.; Zabrouskov, V.; Westphall, M. S.; Syka, J. E. P.; Coon, J. J. Enhanced Dissociation of Intact Proteins with High Capacity Electron Transfer Dissociation. *J. Am. Soc. Mass Spectrom.* **2016**, *27*, 520−531.

(5) 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. Complete Protein Characterization Using Top-Down Mass Spectrometry and Ultraviolet Photodissociation. *J. Am. Chem. Soc.* **2013**, *135*, 12646−12651.

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

(7) 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. Identification and Characterization of Human Proteoforms by Top-Down LC-21 Tesla FT-ICR Mass Spectrometry. *J. Proteome Res.* **2017**, *16*, 1087−1096.

(8) Hendrickson, C. L.; Quinn, J. P.; Kaiser, N. K.; Smith, D. F.; Blakney, G. T.; Chen, T.; Marshall, A. G.; Weisbrod, C. R.; Beu, S. C. 21 Tesla Fourier Transform Ion Cyclotron Resonance Mass Spectrometer: A National Resource for Ultrahigh Resolution Mass Analysis. *J. Am. Soc. Mass Spectrom.* **2015**, *26*, 1626−1632.

(9) Shaw, J. B.; Lin, T.-Y.; Leach, F. E.; Tolmachev, A. V.; Tolić, N.; Robinson, E. W.; Koppenaal, D. W.; Paša-Tolić, L. 21 Tesla Fourier Transform Ion Cyclotron Resonance Mass Spectrometer Greatly Expands Mass Spectrometry Toolbox. *J. Am. Soc. Mass Spectrom.* **2016**, *27*, 1929−1936.

(10) He, L.; Anderson, L. C.; Barnidge, D. R.; Murray, D. L.; Dasari, S.; Dispenzieri, A.; Hendrickson, C. L.; Marshall, A. G. Classification of Plasma Cell Disorders by 21 Tesla Fourier Transform Ion Cyclotron Resonance Top-Down and Middle-Down MS/MS Analysis of Monoclonal Immunoglobulin Light Chains in Human Serum. *Anal. Chem.* **2019**, *91*, 3263−3269.

(11) Weisbrod, C. R.; Anderson, L. C.; Hendrickson, C. L.; Schaffer, L. V.; Shortreed, M. R.; Smith, L. M.; Shabanowitz, J.; Hunt, D. F. Advanced Strategies for Proton-Transfer Reactions Coupled with Parallel Ion Parking on a 21 T FT-ICR MS for Intact Protein Analysis. *Anal. Chem.* **2021**, *93*, 9119−9128.

(12) McIlwain, S. J.; Wu, Z.; Wetzel, M.; Belongia, D.; Jin, Y.; Wenger, K.; Ong, I. M.; Ge, Y. Enhancing Top-Down Proteomics Data Analysis by Combining Deconvolution Results through a Machine Learning Strategy. *J. Am. Soc. Mass Spectrom.* **2020**, *31*, 1104−1113.

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

(14) Sanders, J. D.; Mullen, C.; Watts, E.; Holden, D. D.; Syka, J. E. P.; Schwartz, J. C.; Brodbelt, J. S. Enhanced Sequence Coverage of Large Proteins by Combining Ultraviolet Photodissociation with Proton Transfer Reactions. *Anal. Chem.* **2020**, *92*, 1041−1049.

(15) Zenaidee, M. A.; Lantz, C.; Perkins, T.; Jung, W.; Loo, R. R. O.; Loo, J. A. Internal Fragments Generated by Electron Ionization Dissociation Enhances Protein Top-down Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* **2020**, *31*, 1896−1902.

(16) Fornelli, L.; Srzentić, K.; Toby, T. K.; Doubleday, P. F.; Huguet, R.; Mullen, C.; Melani, R. D.; dos Santos Seckler, H. S.; DeHart, C. J.; Weisbrod, C. R.; Durbin, K. R.; Greer, J. B.; Early, B. P.; Fellers, R. T.; Zabrouskov, V.; Thomas, P.; Compton, P. D.; Kelleher, N. L. Thorough Performance Evaluation of 213 Nm Ultraviolet Photodissociation for Top-down Proteomics. *Molecular & Cellular Proteomics* **2020**, *19*, 405−420.

(17) Gomes, F. P.; Diedrich, J. K.; Saviola, A. J.; Memili, E.; Moura, A. A.; Yates, J. R. EThcD and 213 Nm UVPD for Top-Down Analysis of Bovine Seminal Plasma Proteoforms on Electrophoretic and Chromatographic Time Frames. *Anal. Chem.* **2020**, *92*, 2979−2987.

(18) Cannon, J. R.; Cammarata, M. B.; Robotham, S. A.; Cotham, V. C.; Shaw, J. B.; Fellers, R. T.; Early, B. P.; Thomas, P. M.; Kelleher, N. L.; Brodbelt, J. S. Ultraviolet Photodissociation for Characterization of Whole Proteins on a Chromatographic Time Scale. *Anal. Chem.* **2014**, *86*, 2185−2192.

(19) Holden, D. D.; Schwartz, J. C. Enhancing Ultraviolet Photodissociation Performance on a Thermo Scientific Orbitrap Fusion Lumos Tribrid Mass Spectrometer for Small Molecule and Protein Analysis *66th ASMS Conference on Mass Spectrometry and Allied Topics*: San Diego, CA, 2018Jun 3− 7.

(20) Schmitt, N. D.; Berger, J. M.; Conway, J. B.; Agar, J. N. Increasing Top-Down Mass Spectrometry Sequence Coverage by an Order of Magnitude through Optimized Internal Fragment Generation and Assignment. *Anal. Chem.* **2021**, *93*, 6355−6362.

(21) Little, D. P.; Speir, J.; Senko, M. W.; O'Connor, P. B.; McLafferty, F. W. Infrared Multiphoton Dissociation of Large Multiply Charged Ions for Biomolecule Sequencing. *Anal. Chem.* **1994**, *66*, 2809−2815.

(22) Ballard, K. D.; Gaskell, S. J. Sequential Mass Spectrometry Applied to the Study of the Formation of "Internal" Fragment Ions of Protonated Peptides. *Int. J. Mass Spectrom. Ion Processes* **1991**, *111*, 173−189.

(23) Cobb, J. S.; Easterling, M. L.; Agar, J. N. Structural Characterization of Intact Proteins Is Enhanced by Prevalent Fragmentation Pathways Rarely Observed for Peptides. *J. Am. Soc. Mass Spectrom.* **2010**, *21*, 949−959.

(24) Durbin, K. R.; Skinner, O. S.; Fellers, R. T.; Kelleher, N. L. Analyzing Internal Fragmentation of Electrosprayed Ubiquitin Ions during Beam-Type Collisional Dissociation. *J. Am. Soc. Mass Spectrom.* **2015**, *26*, 782−787.

(25) Lyon, Y. A.; Riggs, D.; Fornelli, L.; Compton, P. D.; Julian, R. R. The Ups and Downs of Repeated Cleavage and Internal Fragment Production in Top-Down Proteomics. *J. Am. Soc. Mass Spectrom.* **2018**, *29*, 150−157.

(26) Zenaidee, M. A.; Wei, B.; Lantz, C.; Wu, H. T.; Lambeth, T. R.; Diedrich, J. K.; Ogorzalek Loo, R. R.; Julian, R. R.; Loo, J. A. Internal

https://doi.org/10.1021/acs.jproteome.2c00583
J. Proteome Res. 2023, 22, 170−181

Fragments Generated from Different Top-Down Mass Spectrometry Fragmentation Methods Extend Protein Sequence Coverage. *J. Am. Soc. Mass Spectrom.* **2021**, *32*, 1752−1758.
(27) 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*, 1928−1935.
(28) Wei, B.; Zenaidee, M. A.; Lantz, C.; Ogorzalek Loo, R. R.; Loo, J. A. Towards Understanding the Formation of Internal Fragments Generated by Collisionally Activated Dissociation for Top-down Mass Spectrometry. *Anal. Chim. Acta* **2022**, *1194*, 339400.
(29) Greer, S. M.; Sidoli, S.; Coradin, M.; Schack Jespersen, M.; Schwämmle, V.; Jensen, O. N.; Garcia, B. A.; Brodbelt, J. S. Extensive Characterization of Heavily Modified Histone Tails by 193 Nm Ultraviolet Photodissociation Mass Spectrometry via a Middle-Down Strategy. *Anal. Chem.* **2018**, *90*, 10425−10433.
(30) Cammarata, M. B.; Thyer, R.; Rosenberg, J.; Ellington, A.; Brodbelt, J. S. Structural Characterization of Dihydrofolate Reductase Complexes by Top-Down Ultraviolet Photodissociation Mass Spectrometry. *J. Am. Chem. Soc.* **2015**, *137*, 9128−9135.
(31) Mehaffey, M. R.; Cammarata, M. B.; Brodbelt, J. S. Tracking the Catalytic Cycle of Adenylate Kinase by Ultraviolet Photodissociation Mass Spectrometry. *Anal. Chem.* **2018**, *90*, 839−846.
(32) Payne, A. H.; Glish, G. L. Thermally Assisted Infrared Multiphoton Photodissociation in a Quadrupole Ion Trap. *Anal. Chem.* **2001**, *73*, 3542−3548.
(33) Holden, D. D.; Sanders, J. D.; Weisbrod, C. R.; Mullen, C.; Schwartz, J. C.; Brodbelt, J. S. Implementation of Fragment Ion Protection (FIP) during Ultraviolet Photodissociation (UVPD) Mass Spectrometry. *Anal. Chem.* **2018**, *90*, 8583−8591.
(34) Dunham, S. D.; Sanders, J. D.; Holden, D. D.; Brodbelt, J. S. Improving the Center Section Sequence Coverage of Large Proteins Using Stepped-Fragment Ion Protection Ultraviolet Photodissociation. *J. Am. Soc. Mass Spectrom.* **2022**, *33*, 446−456.
(35) Earley, L.; Anderson, L. C.; Bai, D. L.; Mullen, C.; Syka, J. E. P.; English, A. M.; Dunyach, J.-J.; Stafford, G. C.; Shabanowitz, J.; Hunt, D. F.; Compton, P. D. Front-End Electron Transfer Dissociation: A New Ionization Source. *Anal. Chem.* **2013**, *85*, 8385−8390.
(36) Huguet, R.; Mullen, C.; Srzentić, K.; Greer, J. B.; Fellers, R. T.; Zabrouskov, V.; Syka, J. E. P.; Kelleher, N. L.; Fornelli, L. Proton Transfer Charge Reduction Enables High-Throughput Top-Down Analysis of Large Proteoforms. *Anal. Chem.* **2019**, *91*, 15732−15739.
(37) McLuckey, S. A.; Reid, G. E.; Wells, J. M. Ion Parking during Ion/Ion Reactions in Electrodynamic Ion Traps. *Anal. Chem.* **2002**, *74*, 336−346.
(38) Xia, Y.; Wu, J.; McLuckey, S. A.; Londry, F. A.; Hager, J. W. Mutual Storage Mode Ion/Ion Reactions in a Hybrid Linear Ion Trap. *J. Am. Soc. Mass Spectrom.* **2005**, *16*, 71−81.
(39) Kline, J. T.; Mullen, C.; Durbin, K. R.; Oates, R. N.; Huguet, R.; Syka, J. E. P.; Fornelli, L. Sequential Ion−Ion Reactions for Enhanced Gas-Phase Sequencing of Large Intact Proteins in a Tribrid Orbitrap Mass Spectrometer. *J. Am. Soc. Mass Spectrom.* **2021**, *32*, 2334−2345.
(40) Holden, D. D.; McGee, W. M.; Brodbelt, J. S. Integration of Ultraviolet Photodissociation with Proton Transfer Reactions and Ion Parking for Analysis of Intact Proteins. *Anal. Chem.* **2016**, *88*, 1008−1016.
(41) Ugrin, S. A.; English, A. M.; Syka, J. E. P.; Bai, D. L.; Anderson, L. C.; Shabanowitz, J.; Hunt, D. F. Ion-Ion Proton Transfer and Parallel Ion Parking for the Analysis of Mixtures of Intact Proteins on a Modified Orbitrap Mass Analyzer. *J. Am. Soc. Mass Spectrom.* **2019**, *30*, 2163−2173.
(42) Huang, T.-Y.; McLuckey, S. A. Top-down Protein Characterization Facilitated by Ion/Ion Reactions on a Quadrupole/Time of Flight Platform. *Proteomics* **2010**, *10*, 3577−3588.
(43) Duselis, E. M.; Panepinto, M. C.; Syka, J. E. P.; Mullen, C.; D'Ippolito, R. A.; English, A. M.; Ugrin, S. A.; Shabanowitz, J.; Hunt, D. F. Improved Sequence Analysis of Intact Proteins by Parallel Ion Parking during Electron Transfer Dissociation. *Anal. Chem.* **2021**, *93*, 15728−15735.
(44) Schaffer, L. V.; Anderson, L. C.; Butcher, D. S.; Shortreed, M. R.; Miller, R. M.; Pavelec, C.; Smith, L. M. Construction of Human Proteoform Families from 21 Tesla Fourier Transform Ion Cyclotron Resonance Mass Spectrometry Top-Down Proteomic Data. *J. Proteome Res.* **2021**, *20*, 317−325.
(45) Rolfs, Z.; Smith, L. M. Internal Fragment Ions Disambiguate and Increase Identifications in Top-Down Proteomics. *J. Proteome Res.* **2021**, *20*, 5412−5418.
(46) Song, S.; Asher, S. A. UV Resonance Raman Studies of Peptide Conformation in Poly(L-Lysine), Poly(L-Glutamic Acid), and Model Complexes: The Basis for Protein Secondary Structure Determinations. *J. Am. Chem. Soc.* **1989**, *111*, 4295−4305.
(47) Klein, D. R.; Holden, D. D.; Brodbelt, J. S. Shotgun Analysis of Rough-Type Lipopolysaccharides Using Ultraviolet Photodissociation Mass Spectrometry. *Anal. Chem.* **2016**, *88*, 1044−1051.
(48) Mullen, C.; Weisbrod, C. R.; Syka, J. E. P. Methods of Performing Ion-Ion Reactions in Mass Spectrometry. *US Patent* **2018**, *9107*, 991.
(49) Meng, F.; Cargile, B. J.; Miller, L. M.; Forbes, A. J.; Johnson, J. R.; Kelleher, N. L. Informatics and Multiplexing of Intact Protein Identification in Bacteria and the Archaea. *Nat. Biotechnol.* **2001**, *19*, 952−957.
(50) 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*, 8421−8429.
(51) Jacob, R. J.; Baker, P. R.; Baldwin, M. A.; Burlingame, A. L. Maximizing proteomic information from MS data: Enhancements to Protein Prospector, a suite of programs for mining genomic databases *48th ASMS Conference on Mass Spectrometry and Allied Topics*: Long Beach, CA, 2000June 11-15.
(52) Okuda, S.; Watanabe, Y.; Moriya, Y.; Kawano, S.; Yamamoto, T.; Matsumoto, M.; Takami, T.; Kobayashi, D.; Araki, N.; Yoshizawa, A. C.; Tabata, T.; Sugiyama, N.; Goto, S.; Ishihama, Y. jPOSTrepo: an international standard data repository for proteomes. *Nucl. Acids Res.* **2017**, *45*, D1107−D1111.

https://doi.org/10.1021/acs.jproteome.2c00583
J. Proteome Res. 2023, 22, 170−181