Exploring N- and C-terminal Fragment Ion Biases in UV-Photodissociation Mass Spectrometry of Intact Proteins

Presented by Raymond Lam
Spring 2024

In partial fulfillment of the requirements for graduation with the Dean's Scholars Honors Degree in the Department of Chemistry.

Dr. Jennifer S. Brodbelt
Supervising Professor
Department of Chemistry

Dr. Carlos R. Baiz
Honors Advisor in Chemistry

Abstract

193 nm UV-photodissociation (UVPD) is a powerful ion activation method in tandem mass spectrometry (MS/MS) for analyzing complex biomolecules and proteins. Sample ions are isolated by their mass-to-charge m/z ratio and exposed to pulsed UV light, causing absorption of UV photons and cleavage along the amide backbone to generate informative fragment ions. The high energy deposition of UVPD and preservation of higher-order structure and modifications makes it particularly appealing for analysis of large, modified, heterogeneous, or multimeric protein states that prove challenging for conventional tandem mass spectrometry methods. It is expected that fragment ions containing the N- and C-terminal ends of each protein ion should be produced and detected equally upon amide backbone cleavage. Data among recent 193 nm UVPD-MS analyses and other MS/MS methods (such as collisionally-activated dissociation) show bias in the production of N- and C-terminal ions in different *m/z* mass-to-charge regions of mass spectra. Owing to the development of UVPD for top-down protein analysis, this method was examined in more detail for the N-term/C-term bias and was the focus of the present study. Among a set of six proteins prepared in denaturing solution conditions to generate standard and "supercharged" charge states, fragment ion identifications from proteins with greater numbers of basic residues were biased towards N-terminal ions in lower *m/z* regions, while proteins with acidic residues biased C-terminus-containing fragment ions in the same region. The backbone sites of generated fragment ions showed precursor charge-state dependence on the degree of ion current bias but remained intrinsically biased towards N/C-terminal fragment ions, while explorations of the charge of fragment ions proved insufficient to provide rationale for the migration of fragment ions into different m/z regions. This promotes further study of biases in top down MS/MS analysis of proteins, particularly as applied to MS techniques attempting to resolve information from spectrally noisy yet information-rich UVPD fragment ions, like proton-transfer charge reduction and internal fragment ion assignment.

Key Terms: Mass Spectrometry, UV-photodissociation, Protein Ions, Fragment Ions, N/C-termini

Introduction

Mass spectrometry, a technique analyzing volatilized ions based on their *m/z* mass-to-charge ratios, has permitted broad explorations of biomolecules in a rapid and facile manner. As driven by instrumentation and computational developments, mass spectrometry-based proteomics is used to both identify and characterize protein sequences, stoichiometries, binding interactions, and biological activities.

In the field of tandem mass spectrometry, selected precursor ions (termed the "MS^1" spectrum) undergo isolation and fragmentation into constituent product ions (the "MS/MS" or "MS^2" spectrum) used to obtain sequence and structural characteristics of the original precursor ion. In Tribrid systems such as the Orbitrap Fusion Lumos mass spectrometers, three primary portions of the instrument, the quadrupole, linear ion trap, and Orbitrap mass analyzer, work in tandem to accomplish this. The quadrupole applies and spectra through alternating rF voltages to allow an initial selection of ions by their *m/z* mass-to-charge ratio; the linear ion trap allows accumulation of ions for administered fragmentation events, either occurring in the linear ion trap or an associated collisional cell; and the Orbitrap mass analyzer monitors the oscillation of ions around a spindle which is converted by Fourier transformation into an output *m/z* signal.[10] Fragmentation of ions is most commonly accomplished through collisional activation methods, such as collision-induced dissociation (CID) or higher energy collisional activation dissociation (HCD). Collisional methods accelerate ions into neutral gas molecules to induce fragmentation, generating *b/y*-type fragment ions based on cleavages between the carbonyl and amino groups of the amide backbone.[1,3,4]

In the search for methods characterizing more challenging biomolecular targets, such as large proteins that don't fragment well or proteins with modifications, higher-energy activation methods have been created to allow for greater degrees of characterization, including electron-based methods and photodissociation methods. In particular, UV-photodissociation (UVPD) is capable of generating cleavages at a greater variety of positions along the amide backbone. It provides additional *a/x*-type ions between C_α-C bonds and *c/z*-type ions between N-C_α bonds in addition to the C-N bond cleavage available from *b/y*-type ions (Figure 1).[3,4] One key strength of these higher energy activation methods is through its mechanism of fragmentation: its single-step energy deposition allows for more stochasticity in the generation of protein ion fragmentation (Figure 2).

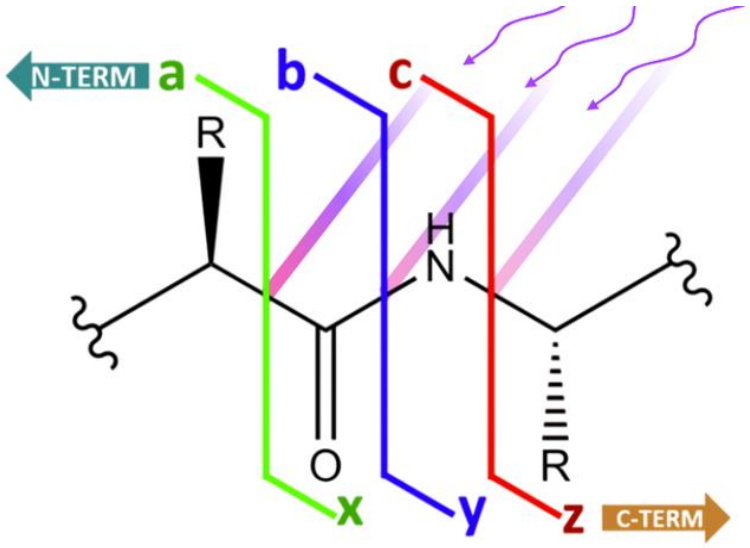

Figure 1. *a/x*, *b/y*, and *c/z*-type protein fragment ion nomenclature along the amide backbone, representing fragmentation pathways accessible to UV-Photodissociation.

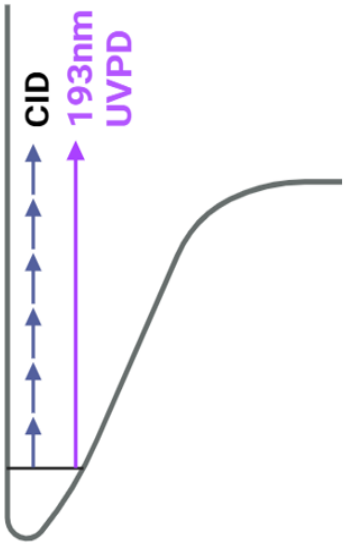

Figure 2. Proposed potential energy representation of collisionally-induced dissociation (CID) versus UV-photodissociation (UVPD) direct dissociation fragmentation pathway.

Though the mechanism of action is not fully resolved, most incident photons absorbed are thought to energize ions to excited states, and direct dissociation from the excited states results in a/x fragment ions. There may be internal conversion to the ground electronic state and heating, generating collisional-like *b/y*-type ions if there is sufficient energy for fragmentation.[5] The majority of ion cleavages from UVPD appear to derive from direct dissociation of the amide backbone from incident photons, causing the majority of observed fragmentation and the generation of *a/x*, *b/y*, and *c/z* fragment ions.

Different ion activation methods may favor fragmentation between specific residues; collisional-induced dissociation is known for producing cleavages N-terminal to proline, while 213 nm has been demonstrated to favor *a* and *y-2* ion generation N-terminal to proline residues compared to 193 nm UVPD.[6] In comparison, 193 nm UVPD appears to display few biases in preferential cleavages; while characteristic *a*-ions appear abundant, residue-biased fragmentation pathways rarely appear when considering all fragment ions generated from 193 nm UVPD spectra.

Most mass spectrometry applications have utilized proteolytic digests of proteins separated by liquid chromatography and further analyzed, termed bottom-up mass spectrometry, in order to generate smaller peptides for mass spectrometry analysis. These smaller peptides can be separated and fragmented more readily, permitting large-scale analyses of mixtures of proteins or cell lysates on rapid timescales.[1] This digestion, however, causes the loss of key structural and sequence information retained in the intact protein species. In contrast, top-down and native mass spectrometry analyzes undigested and intact proteins, allowing deeper exploration of proteins' structure and localization of modifications or binding interactions.[7] When aided by the use of high-energy tandem mass spectrometry like UVPD, its extensive fragmentation is tenable to uncovering these key features lost in bottom up mass spectrometry methods.

One way top-down analysis has been performed in recent years has been to vary the charge state of the proteins analyzed, allowing proteins to access different conformations and fragment at different residue sites. Proteins and their constituent fragment ions may be analyzed at lower charge states, wherein the decrease in charge decreases the degree of coulombic repulsion and adopts a more compact conformation, or they may be analyzed at higher charge states, where repulsion between residues causes them to become more extended. Prior work with UV-photodissociation and proteins that are "supercharged" by the addition of charging reagents in solution has shown concentration of fragment ion signal towards particular residue sites, often providing novel backbone cleavage sites unavailable in lower charge states to confer new sequence and structural information.[8]

Additional means of enhancing the information gained from the complexity of top-down mass spectral data, particularly for higher-energy methods like UV-photodissociation that generate information-rich yet spectrally congested spectra, has been through manipulation of the regions of *m/z* analysis performed by the mass spectrometer. One such method relies on "fractionation" of UV-photodissociation spectra, isolating narrower *m/z* ranges of ions in the gas phase and analyzing data from the varied spectral regions separately, in order to excise noise from unfragmented precursor ions and dedicate more signal to generated fragment ions at higher and lower *m/z* regions.[9] This strategy improved sequence coverage, or the proportion

of linkages between adjacent residues characterized by at least one fragment ion, by a factor of 1.3 to 2.3, garnering more instances for sequence and structural characterization to occur. For large proteins, where the broader range of fragment ion charge states and possible fragmentation sites creates a much higher degree of spectral overlap within the mass spectrometer's *m/z* spectrum, this fractionation strategy helps dedicate more of the fragment ion signal towards assigning fragment ions with greater confidence.

These fractionated regions of the *m/z* spectrum may be further analyzed through a combination of charge modulation and fractionation in the gas phase with gas-phase proton-transfer charge reduction. Here, product fragment ions from spectra are fractionated into different *m/z* regions higher and lower than the m/z of the precursor and charge-reduced in the gas phase with a reagent like nitrogen-adducted fluoranthene or other perfluoroalkyl substances, reducing spectral congestion as fragment ions in dense regions are separated in *m/z* space to better resolve fragment ion isotopic distributions for identification. In one study, this was applied to UVPD analysis of calf thymus histones, relatively basic proteins with site-specific chemical modifications responsible for epigenetic regulation of chromatin transcription; gas-phase proton transfer charge reduction helped attain high sequence coverage for N-terminal acetylation, as well as lysine and arginine methylation near the proteins' N-terminal residues.[10] In concert, this broad spectrum of techniques showcases the many means employed to better resolve the rich and complex spectral data generated by UV-photodissociation, particularly for the heterogeneous and complex targets it aims to best resolve.

Fractionation of UV-photodissociation spectra appears to generate certain biases in the types of ions detected in separate *m/z* regions. In the aforementioned proton transfer charge reduction-histone study, for example, analysis of one particular H4 histone proteoform, acH4R3me2K12ac, revealed a significant higher proportion of N-terminal ions in lower *m/z*-fractionated spectra, whereas higher *m/z*-regions were dominated by C-terminal ion identifications.[10] This was examined over a relatively narrow charge range from 12 to 15+ precursors, and the majority of these biases were largely attributed to *a* and *x*-type ions for the lower and higher *m/z* regions, respectively. Residues and their associated modifications must be isolated by "bracketed" by fragment ions on the residues directly adjacent to them in order to resolve the exact location of the modification/residue; the mass difference observed between an adjacent modification-containing and modification-lacking fragment ion within the spectra allows the relevant structural characteristic to be assigned.

This study then aims to examine N- and C-terminal ion biases in a broader set of proteins and charge states, seeking to determine probable causes for the preferential production of N- or C-terminal ions among a broad set of proteins ranging in size, residue composition, charge, and

structural features. In particular, the inclusion of supercharged states, wherein proteins adopt more unfolded confirmations and have previously shown increased affinity for fragmentation at particular residue sites independent of residue identity, was hypothesized to provide interesting lead for the examination of N and C-terminal ion biases across a broader range of protein charge states and confirmations.

Experimental Section

Spectral data of UVPD-fragmented peptides were collected as described previously.[8] Raw data are available in the jPOSTrepo with jPOST ID JPST002123.

Sample Preparation

Carbonic anhydrase II (bovine, 29.0 kDa), ubiquitin (bovine, 8.6 kDa), myoglobin (equine, 16.9 kDa), and cytochrome C (equine, 12.4 kDa) were purchased from Millipore Sigma, while staphylococcal nuclease (*Staphylococcus aureus*, 16.1 kDa) and dihydrofolate reductase (*Escherichia coli*, 19.0 kDa) were expressed as described previously.[11,12] Proteins were prepared at 5 µM concentration in 50% acetonitrile and 50% LC-MS Grade Water with 0.1% formic acid, all purchased from Fisher Scientific, while solutions of supercharged proteins were prepared with additional 1% 4-Vinyl-1,3-dioxolan-2-one purchased from Millipore Sigma.

Top-Down Mass Spectrometry

Mass spectrometry was performed using a Thermo Scientific Orbitrap Fusion Lumos mass spectrometer with addition of a 500 Hz 193 nm ArF Excimer Laser (Coherent Excistar XS) for fragmentation. Electrospray ionization with Pd/Au-coated static tips pulled in-house and loaded with protein solution were used to spray samples with 1.6 kV applied voltage. 3 *m/z*-wide windows were used to isolate precursor protein charge states, 2 each from supercharged and non-supercharged protein solutions, and 193 nm light delivered at 1 pulse and 1-2 mJ to ions accumulated in the linear ion trap of the Lumos system, optimized for fragmentation efficiency to fragment protein ions. Spectra were collected in triplicate in intact protein mode from 150-2000 *m/z* and 240000 resolution at an AGC target of 1E6, S-lens Radio Frequency of 60, and 200 averages.

Data Analysis

MS/2 spectra “lower” and “higher” *m/z* regions were demarcated as zones lower and higher in *m/z* than the precursor outside of a 3 *m/z* window surrounding the precursor ion. MS/2 Spectra were deconvoluted using data from full, lower, and higher *m/z* regions using Thermo Scientific’s FreeStyle and its Xtract algorithm at a S/N threshold of 3 and fit factor of 44% between the

theoretical and observed fragment ion isotope distributions. ProSight Lite was used to resolve the number of deconvoluted matched fragments and sequence coverage at a maximum PPM tolerance of 10; results from each activation parameter's three replicates were averaged.[13] MS-TAFI, a Python tool developed in-house for visualization of intact protein mass spectra as described previously, was used to generate weighted fragment ion intensities and maps of fragment ion signal along their backbone fragmentation site with maximum PPM tolerance at 10 using one of the three replicates.[14] Additionally, ProSight Native was used to resolve charge state and mass distributions of protein fragment ions from MS2 spectra with relevant ion intensities, employing THRASH deconvolution at a S/N threshold of 3 and a maximum PPM tolerance of 10 among one of the three replicates.[15] Hydrogen shifted variants of fragment ions (e.g. *a+1, x+1, y-1, y-2* ions) were consolidated in the statistics of their non-shifted forms (*a, x, y* ions) during all analysis.

Results and Discussion

In this study, top-down 193 nm UV-photodissociation spectra of proteins, including those in "supercharged" states allowing possible extended conformations and alternate fragmentation to occur, were analyzed to assess possible biases in the population of ions generated through its high-energy ion activation method. Multiple avenues were explored in order to both characterize and rationalize the dominance of N and C-terminal ions detected in regions lower, higher, and around the *m/z* of detected fragment ions.

Initially, the analyzed proteins were sorted based on their basicity, pursuing a possible link to a residue-based rationale for non-stochastic fragmentation, a possible link to the mobile proton model for *b/y*-type ion formation (Table 1).

***Species* and Protein**	**Uniprot Accession**	**Monoisotopic Mass (Da)**	**Structural Feature/Modifications**	**Charge Range (standard, supercharged)**	**% Basic residues (K, R)**	**pI**
E. coli Dihydrofolate Reductase	P0ABQ4	18954.31	Disulfide Link (C85-152)	13, 19, 22, 24+	9.0	5.46
Bos taurus Carbonic Anhydrase II	P00921	28964.67	N-terminal acetylation	22, 26, 34, 38+	10.4	6.40
Bos taurus Ubiquitin	P0CG53	8559.62	-	8, 10, 12, 14+	14.5	6.56
Equus caballus Myoglobin	P68082	16940.56	Non Covalent heme	15, 17, 22, 24+	13.6	7.36
Staphylococcus aureus Staphylococcal Nuclease	Q8NXI6	16106.39	-	13, 18, 25, 27+	17.5	9.24
Equus caballus Cytochrome C	P00004	12351.52	N-terminal acetylation, Disulfide Link (C14-17), Covalent Heme (C14-17)	13, 15, 20, 22+	20.2	9.59

Table 1. Species, mass, structural features, charge range, and basicity of analyzed proteins.

Overall matched fragment ions from information garnered from the full *m/z* spectrum, representing fragment ion signal masses lower, high, and inclusive of the precursor ion's *m/z* appear to be evenly matched for most proteins between N and C-terminal ion identifications (Figure 3, S1-S5). The majority of N-terminal fragment identifications appear to derive from *a*-type ions as documented among 213 and 193 nm UV-photodissociation described previously.[6] More massive proteins within the set analyzed, such as carbonic anhydrase at 29 kDa mass, appear to have slightly more bias in matched fragment ions, with higher proportions of *a*-type ions contributing to a higher proportion of N-terminal fragment identifications (Figure S1). Interestingly, carbonic anhydrase appeared to display the lowest sequence coverage of all identified proteins, defined as the percent of linkages between adjacent residues characterized

by at least one fragment ion, between 55 to 68% sequence coverage (Figure 4). Sequence coverage was also observed to vary between lower and higher *m/z* spectral regions, with lower *m/z* regions displaying sequence coverage similar or equal to the sequence coverage obtained from full *m/z* analysis for most proteins (Figure S6-S10). This may be due to greater spectral congestion observed in higher *m/z* spectra, wherein fragment ions display overlapping isotopic distributions hindering fragment identifications.

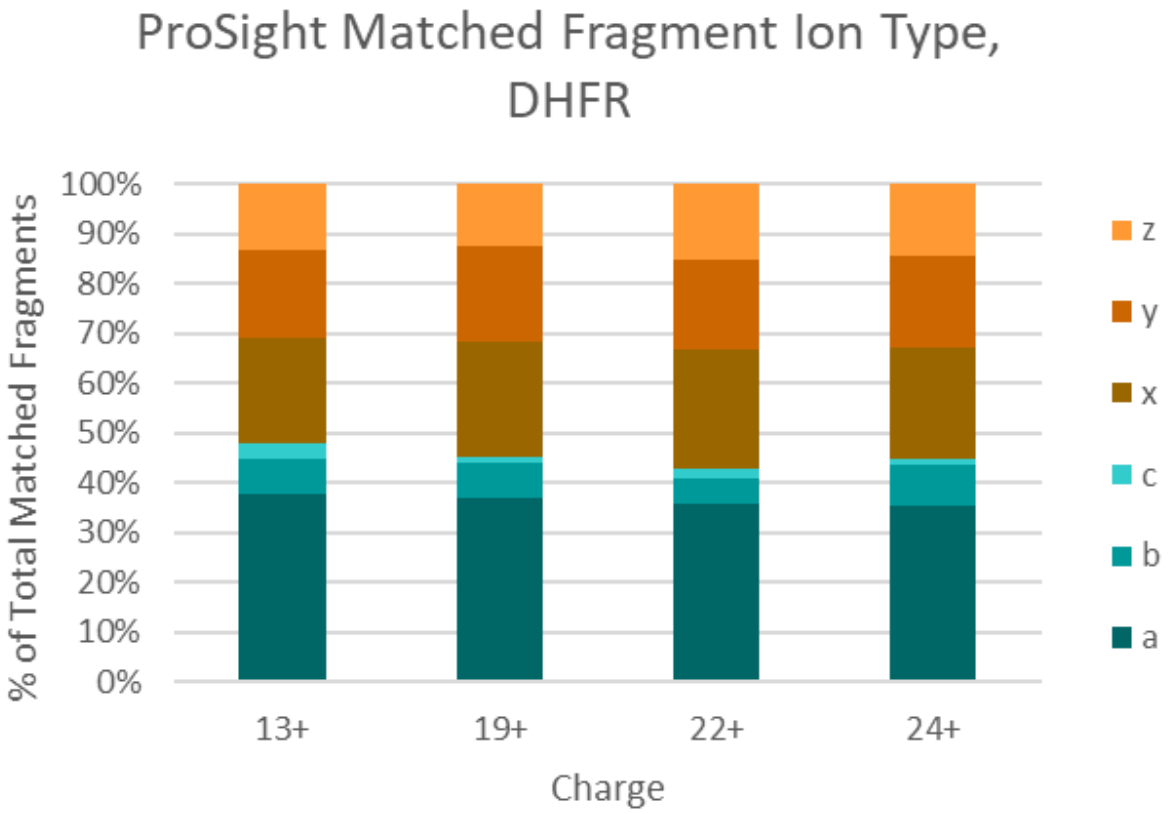

Figure 3. Representative matched fragment identifications of dihydrofolate reductase obtained from full m/z.

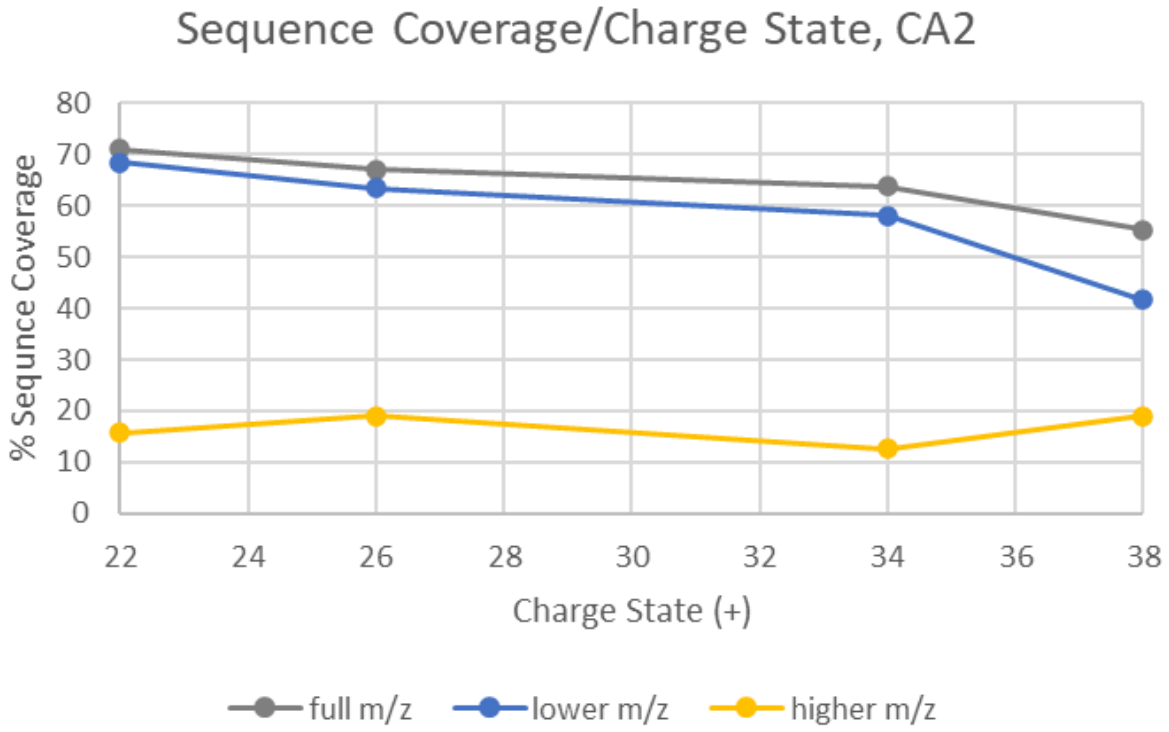

Figure 4. Sequence coverage garnered from lower, higher, and full *m/z* spectral data of carbonic anhydrase from UVPD spectra of 22, 26, 34, and 38+ precursors.

Lower and Higher m/z Fragment Ion Bias

Alternatively, fragment ion analysis of spectral data fractionated into lower and higher *m/z* regions reveals biases in N and C-terminal ion detection that remain constant across precursor charge states.[9] One key distinction from previous studies is the demarcation of lower and higher *m/z* during data processing rather than to exclude the noise of dominant precursor ion signal during data collection; this still provides insight into N and C-terminal ion detection.

For the more acidic subset of proteins analyzed with isoelectric points at or below pI=7.5, representing dihydrofolate reductase, ubiquitin, and myoglobin, C-terminal ions are more often detected in lower *m/z* regions across all charge states (Figure 5-7). This counters the earlier trend observed by Walker and colleagues of predominantly N-terminal ions in the lower *m/z* fractionated region of histone H4 proteoforms.[10] The set of proteins similarly displays a preference for *a*-ions as the majority of the N-terminal ion identifications in both lower and higher *m/z* regions, particularly for dihydrofolate reductase and myoglobin. The pronounced nature of *a* ions may have a link to the ease of fragmentation of each protein; dihydrofolate reductase and myoglobin are much larger proteins than ubiquitin, with the former two at 19 and 16.9 kDa and ubiquitin at 8.7 kDa, and this disparity is reflected in lower sequences

compared to the near-100% sequence coverage for all charge states and fractionation windows of ubiquitin (Table 1; Figure S6-8). Interestingly, this intrinsic preference for N and C-terminal ions in each *m/z* region remains across the four charge states examined within each protein despite fluctuations in sequence coverage, a similar finding to the previous study by Walker and colleagues.

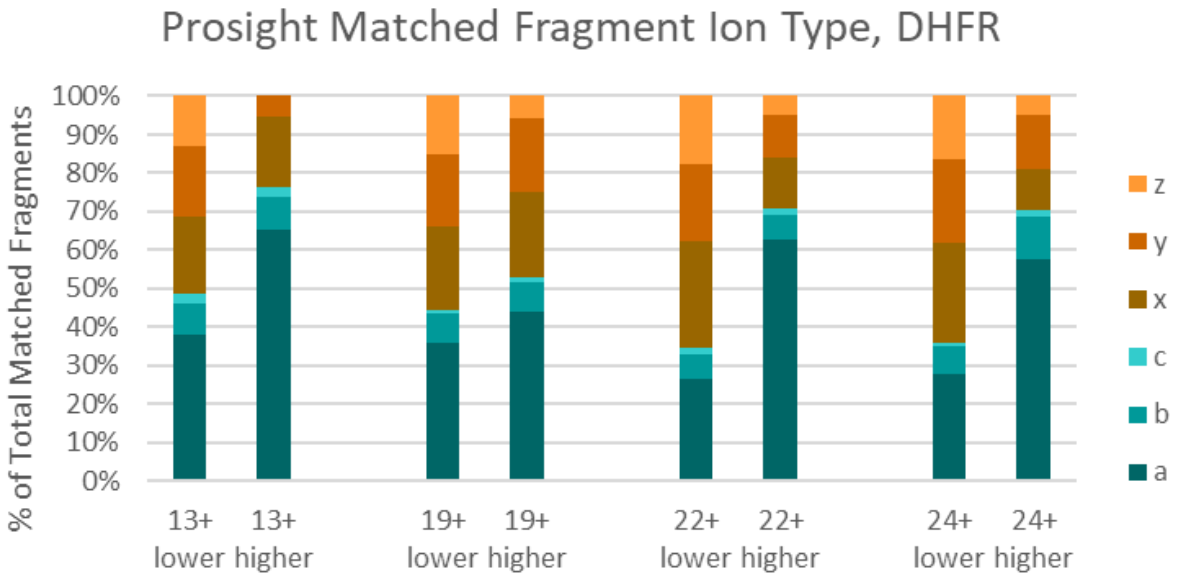

Figure 5. Matched fragment ion identifications of dihydrofolate reductase obtained from lower and higher *m/z* of UVPD spectra of 13, 19, 22, and 24+ precursors.

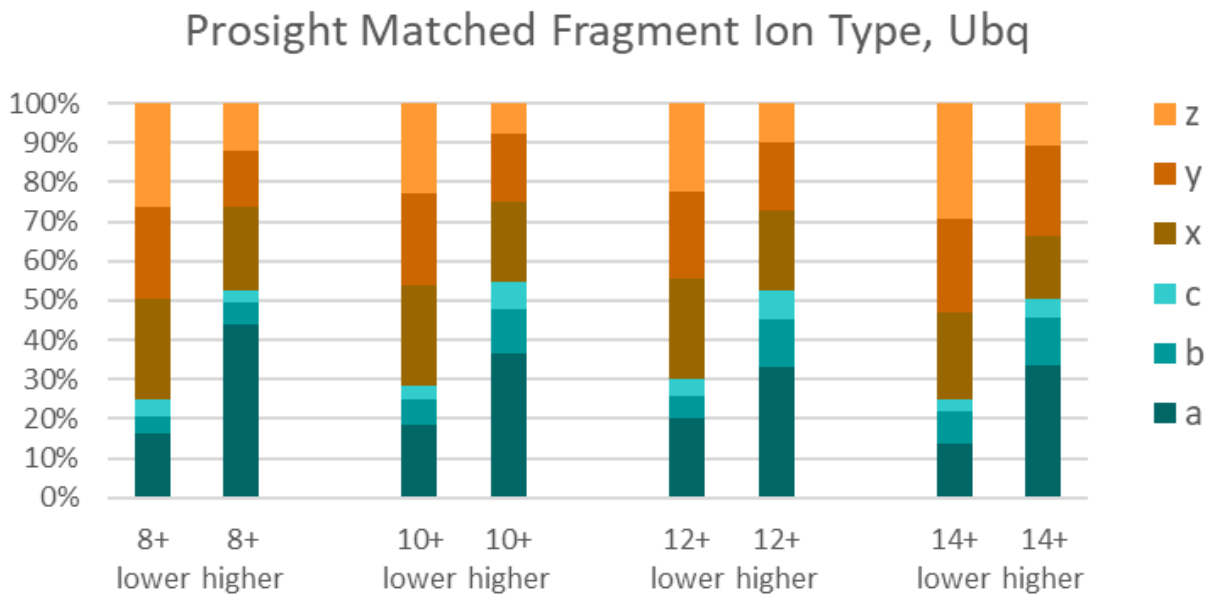

Figure 6. Matched fragment ion identifications of ubiquitin obtained from lower and higher *m/z* fractionation analysis of UVPD spectra of 8, 10, 12, and 14+ precursors.

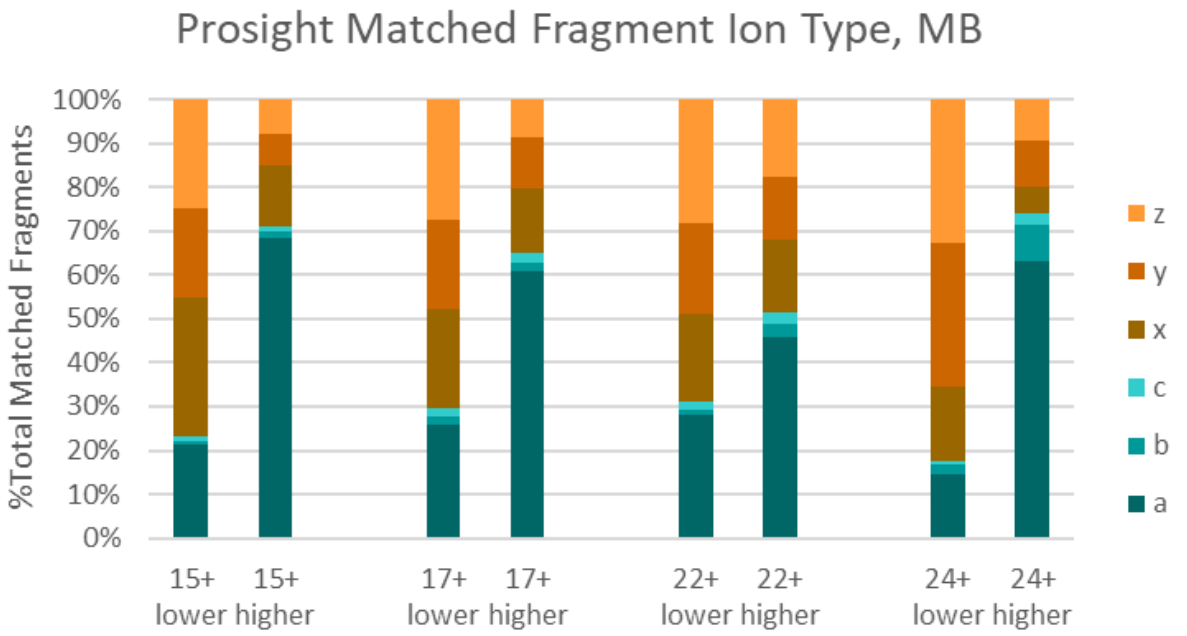

Figure 7. Matched fragment ion identifications of myoglobin obtained from lower and higher *m/z* fractionation analysis of UVPD spectra of 15, 17, 22, and 24+ precursors.

UVPD fragmentation of staphylococcal nuclease and cytochrome C, both with isoelectric points 9.24 and 9.59, respectively, display an inverted trend to the acidic proteins examined, with N-

terminal ions biased in lower *m/z* regions (Figure 8-9). This was also observed across supercharged and non-supercharged charge states, though the degree of bias similarly varies between each protein. Staphylococcal nuclease appeared to display >50% favorability for N-terminal ions at all lower *m/z* regions, with the majority of matched fragments identified as *a*-type ions, whereas >50% favorability for N-terminal ions in lower *m/z* regions of cytochrome C were only apparent at supercharged precursor charge states 20 and 22+.

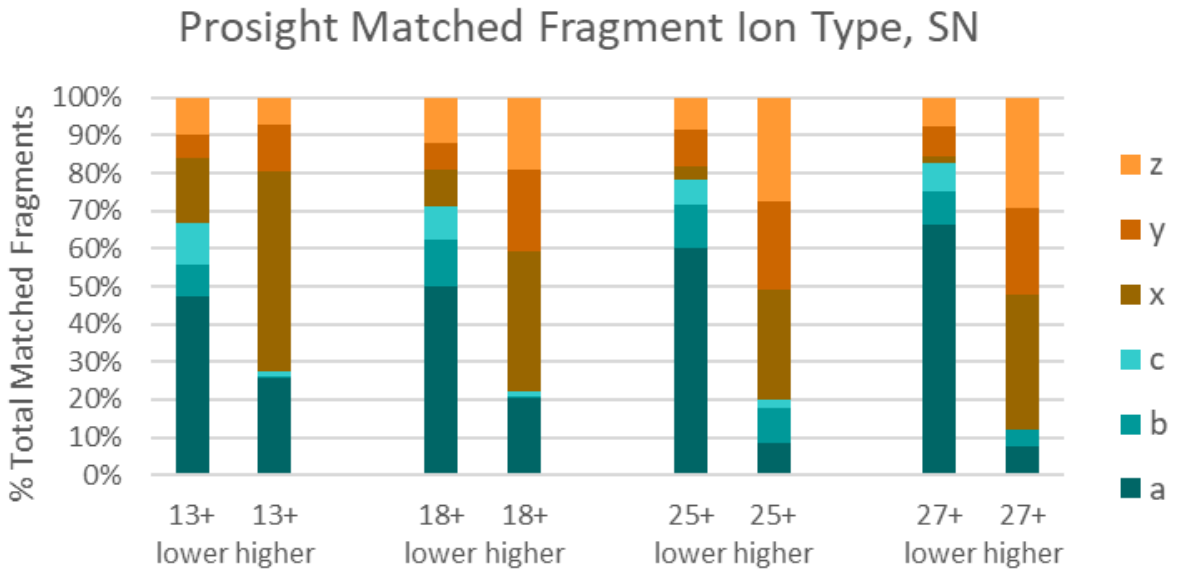

Figure 8. Matched fragment ion identifications of staphylococcal nuclease obtained from lower and higher *m/z* fractionation analysis of UVPD spectra of 13, 18, 25, and 27+ precursors.

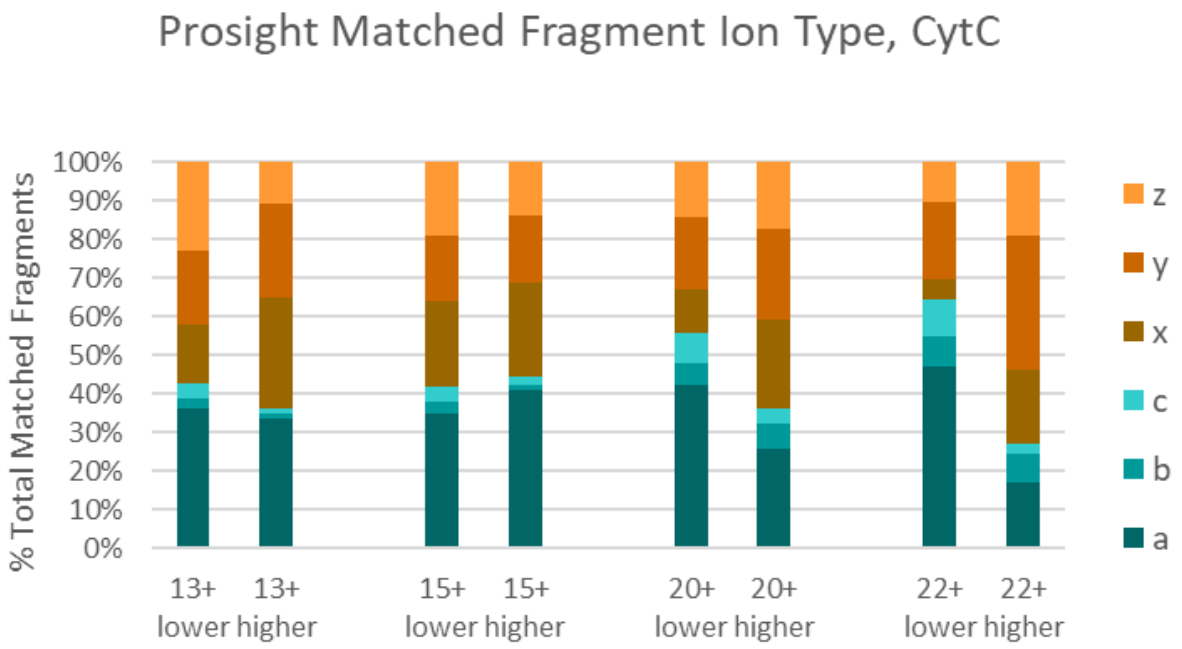

Figure 9. Matched fragment ion identifications of cytochrome C obtained from lower and higher *m/z* fractionation analysis of UVPD spectra of 13, 15, 20, and 22+ precursors.

Carbonic anhydrase appeared to disrupt this tentative trend of basicity dictating N and C-terminal bias. In spite of its more acidic composition at an isoelectric point of 6.40, it displays behavior similar to staph nuclease and the previously observed basic H4 histone proteins, with N-terminal ions comprising the majority of identifications in all observed lower charge states (Figure 10). The persistence of the degree of bias did display fluctuation, with *a*-ions varying between 59 and 74% of matched fragment ions in lower *m/z* between the 22 and 34+ precursor, respectively.

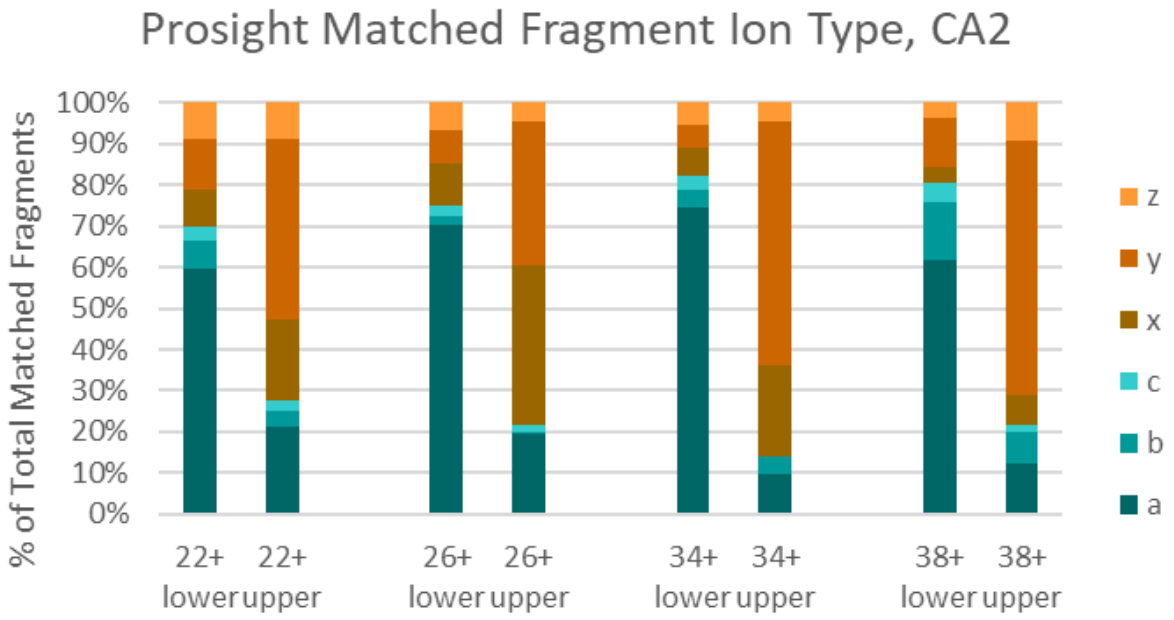

Figure 10. Matched fragment ion identifications of carbonic anhydrase obtained from lower and higher *m/z* fractionation analysis of UVPD spectra of 22, 26, 34, and 38+ precursors.

The high degree of *a* ion generation may be due in part to its more massive structure at 29 kDa, but does not fully provide a rationale for the strength of bias towards N-terminal ions in lower *m/z* at all charge states. Countering trends of acidic proteins analyzed through UV-photodissociation, this precluded assignment of a purely residue-based model of fragmentation; fragmentation and N and C-terminal bias is then likely occurring through an alternate mechanism not fully explained by the presence of basic and acidic residues alone.

Backbone Fragmentation Maps

Another means of determining the generation of fragment ions is to visualize the relative fragment ion signal from fragment ions generated at each specific residue; these are referred to as "backbone fragmentation maps" throughout the remainder of the study. Programs like MS-TAFI allow visualization of the shifts in fragment ion abundance from different sites upon precursor charge state modulation or structure modification, gaining structural insights into which residues of the protein are accessible and facile to fragmentation.[14]

In previous studies, increases in the precursor charge state resulted in 193 nm UVPD-generated fragment ions concentrating towards fragmentation at fewer residues, with *b/y*-type ions occurring adjacent to proline residues.[8] For proteins like dihydrofolate reductase, this resulted in concentration of fragment ion signal to a narrower set of residue sites; however, the specific location of the preferential cleavage tends to vary by charge state, with the most prominent

fragment ion fluctuating between an A/M, F/P, E/P, and E/S cleavage between the 13 and 24+ charge state (Figure 11).

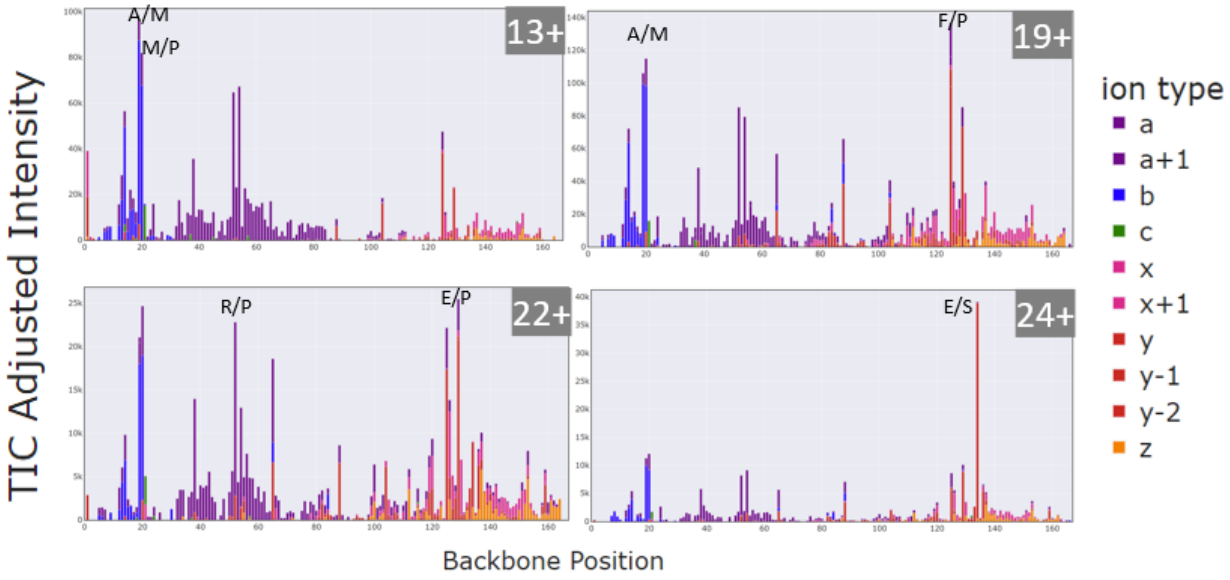

Figure 11. Backbone fragmentation map of dihydrofolate reductase garnered from full *m/z* spectral information of UVPD spectra of 13, 19, 22, and 24+ precursors.

Segregating this analysis into spectral data obtained from lower and higher *m/z* regions further defines this trend, with cleavages adjacent to proline similarly occurring and HCD-type *b/y* ions becoming increasingly favored at higher precursor charge states. However, the type and location of proline cleavages appears to display some degree of orthogonality between the higher and lower *m/z* spectra, with each lower / higher *m/z* fractionation pair displaying an alternate site with the most fragment ion signal. For example, in dihydrofolate reductase, the most prominent cleavages in the 24+ charge state occur between M/P and E/P sites in the lower *m/z* region but A/M and K/P in the higher *m/z* (Figure 12).

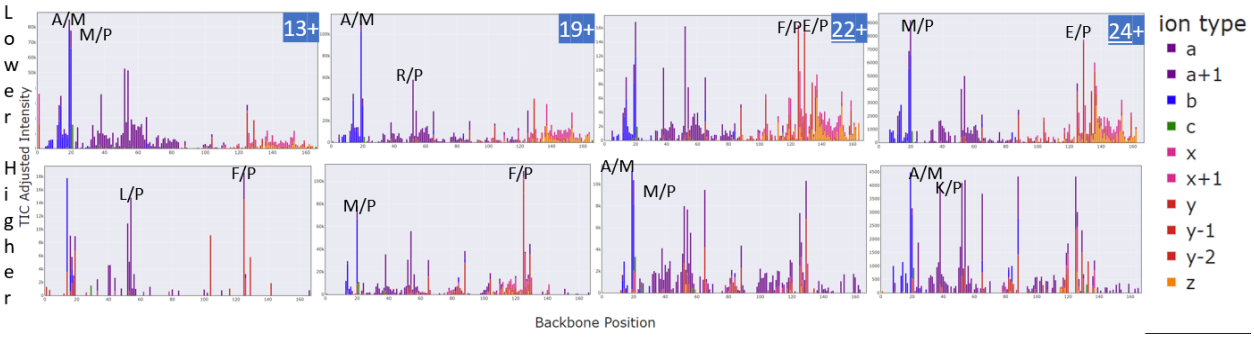

Figure 12. Backbone fragmentation map of dihydrofolate reductase garnered from lower and higher *m/z* spectral information of UVPD spectra of 13, 19, 22, and 24+ precursors.

This trend of exclusivity between the higher and lower *m/z* is observed for most proteins, with the change in the site of preferential cleavages between charge states possibly hinting at some degree of conformational change as each proteins' confirmation varies upon charging (Figure S11-15). While *a* ions appear in all versions of the protein, proteins such as carbonic anhydrase and staphylococcal nuclease appear to show the most separation of *a* fragment ion signal into their lower and higher *m/z* regions; in particular, staphylococcal nuclease appears to display most *a* ion signal from the in the N-terminal half of the protein in lower *m/z* and *y* ion signal from the C-terminal half of the protein in higher *m/z* (Figure S11, S14).

While likely in unstructured or extended conformations, groups have previously attempted to characterize charge state grouping confirmers or "conformational families" of proteins via their thermal unfolding transitions during ion mobility analysis.[16] In supercharging each protein, the observed residue sites, such as those adjacent to proline, may become more accessible while producing non-reproducible extended conformations or conformational ensembles. Work by the Clemmer group has unveiled distinct "conformational families" of myoglobin in ammonium acetate solution upon heating, including one *apo*-myoglobin transition, lacking a heme group, they refer to as G3 encompassing the 17 to 21+ charge states.[17] The fluctuation between proline cleavages at alternate backbone sites despite continued bias for N-terminal ions in lower *m/z* may imply the presence of multiple structures within one of these conformational families, allowing varying proline residues to be exposed and preferably fragmented (Figure S13). Additionally, the inclusion of charge states beyond this family, such as the 15, 22, and 24+ charge states, may further provide rationale for the variability between each charge state with common characteristics of N-terminal bias in lower *m/z*.

Charge and Mass of Fragment Ions

Fragment ion data was also analyzed for the mass and charge of the N and C-terminal ions collected from the full *m/z* range each spectrum, aiming to observe to see if either factor influenced a migration of fragment ions to lower and higher *m/z* regions relative to the precursor ion. This data was weighted in terms of the fragment ion intensity, where the signal generated by each fragment ion was considered to generate a "weighted" perspective on charge and mass data of N and C-terminal ions (Equation 1)

$$\frac{\sum_{n=1}^{k}[Fragment\ Ion\ TIC\ Intensity]\ x\ [Fragment\ Ion\ Charge\ z\ \ or\ Mass\ m]}{\sum_{n=1}^{k}[Fragment\ Ion\ TIC\ Intensity]}$$

Equation 1. function for averaging charge and mass of identified fragments by weighting total ion current (TIC).

It was initially hypothesized that the type of fragment ions biased present in lower *m/z* regions, such as the N-terminal ions abundant in cytochrome C's lower *m/z* region from its supercharged 20 and 22+ precursors, would overall have a lower mass *m* and higher charge *z* (Figure 9). However, these trends often appear contradictorily in many of the proteins examined, with mass and charge data often conflicting and failing to rationalize migration to *m/z*.

N-terminal ions collected from the full *m/z* scan of cytochrome C's 20 and 22+ precursors have a weighted average charge from 3 to 4+ compared to 6 to 8+ for C-terminal ions, yet the fragment ions also display an average mass from 1413 to 2344 daltons, much lower than 2093 to 4022 dalton mass range of C-terminal fragment ions (Figure 13-14). In addition, the relative difference in charge and mass appears to reverse as a function of the precursor charge state; in the non-supercharged 13 and 15+ charge states, N-terminal ions are simultaneously more massive and more charged. Among all charge states, the differences in mass and charge between N and C-terminal ions are only supported by either mass *or* charge in describing their migration to different *m/z* regions.

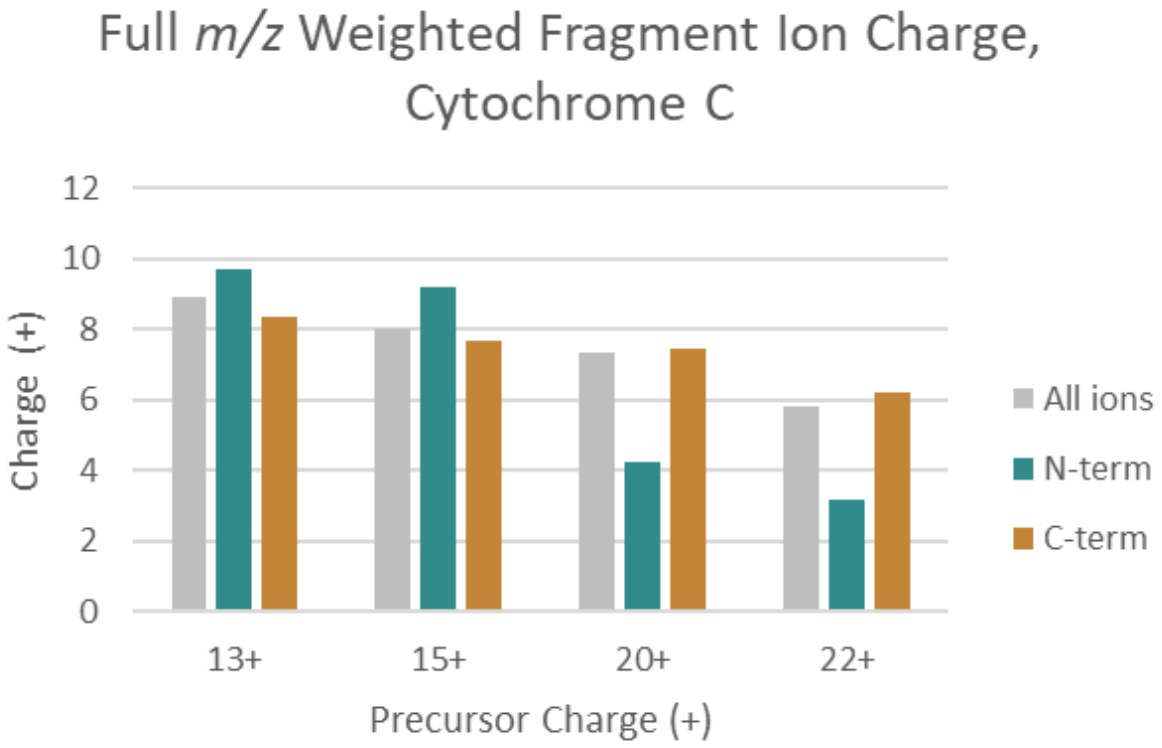

Figure 13. Weighted fragment ion charge of cytochrome C from UVPD spectra of 13, 15, 20, and 22+ precursors.

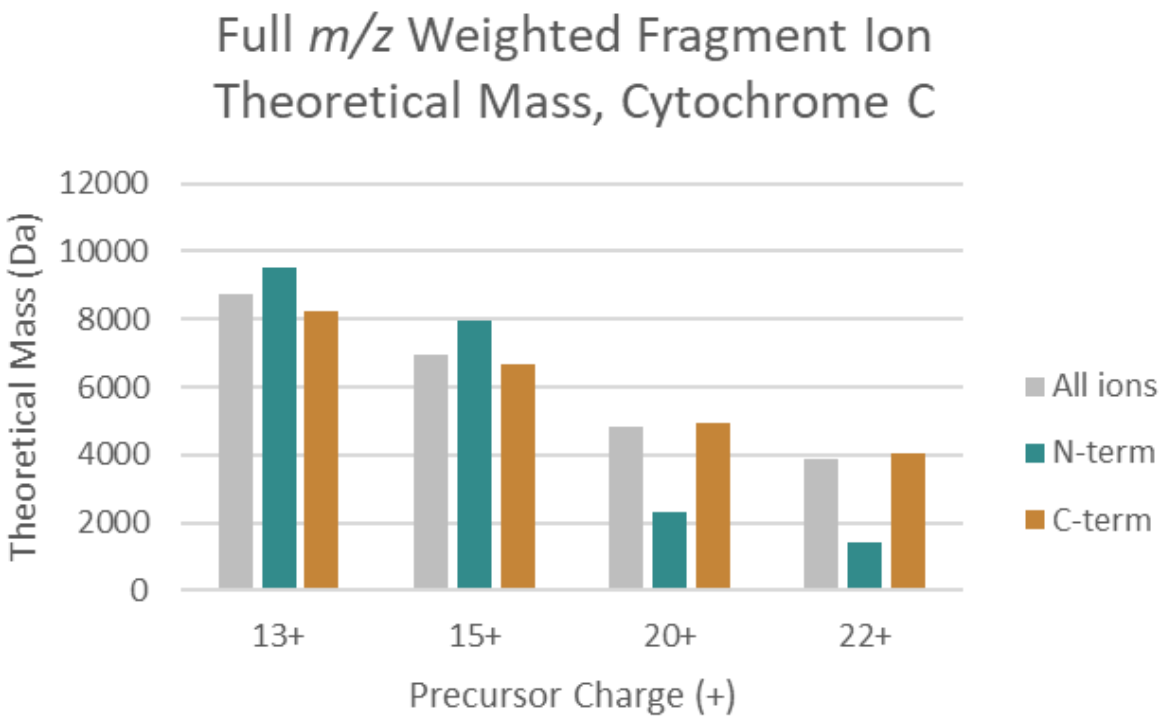

Figure 14. Weighted fragment ion mass of cytochrome C from UVPD spectra of 13, 15, 20, and 22+ precursors.

For most other proteins examined, the differences in charge state and mass appear marginal between N and C-terminal ions or don't fully support mass or charge-based migration to different *m/z* (Figure S16-S19; Figure S20-S23). For ubiquitin, its bias for C-terminal ion identifications in lower *m/z* is maintained throughout all charge states, yet the largest charge disparity between N and C terminal ions exists for products of the 14+ precursor, where they have a weighted average charge of 7 and 9+, respectively (Figure 15). The relative mass difference between N and C terminal ions varies as a function of the precursor charge state, with C-terminal ions more charged from the 12 and 14+ precursor ions and less massive from 8 and 10+ precursors supportive of their bias in lower *m/z* regions across charge states (Figure 16). A similar contradictory behavior exists with only charge or mass data, not both, supporting a specific precursor's fragment ion biases into different mass-to-charge *m/z* regions. The differences between these parameters appear marginal for 8, 10, and 12+ precursors, however, and C-terminal ions still comprise 70 to 75% of matched fragments for all lower *m/z* fractionated spectra (Figure 6).

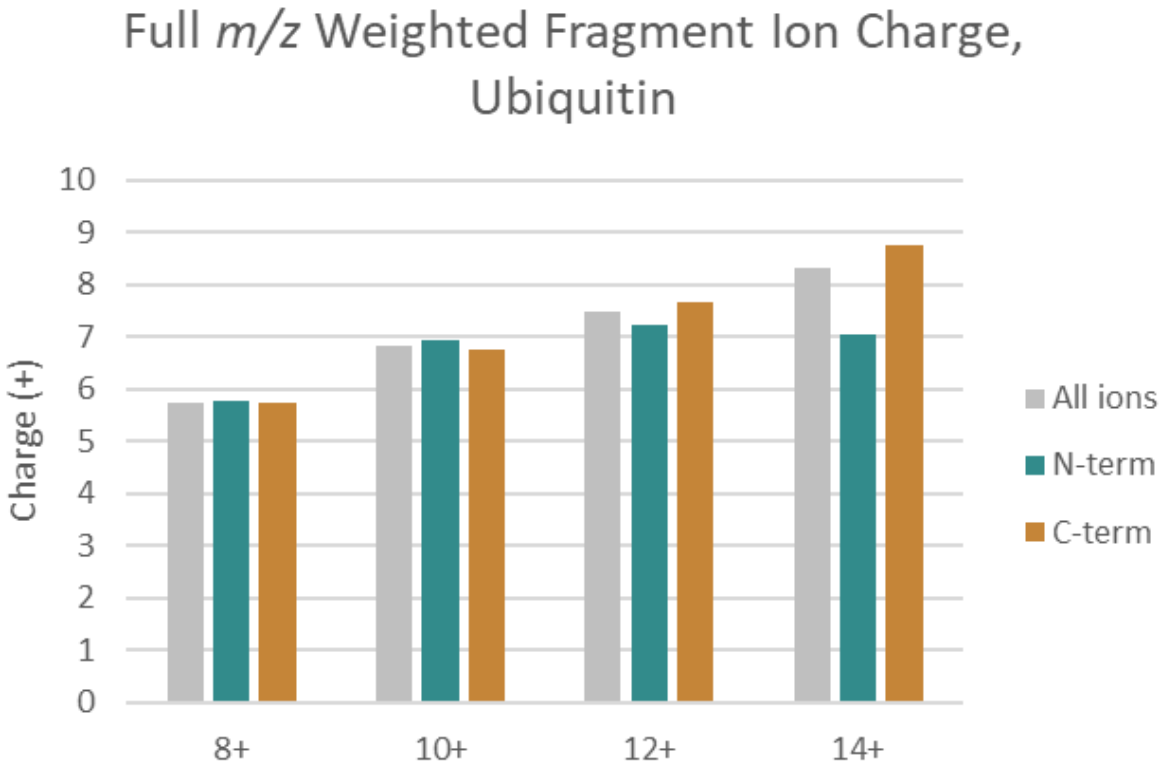

Figure 15. Weighted fragment ion charge of ubiquitin from UVPD spectra of 8, 10, 12, and 14+ precursors.

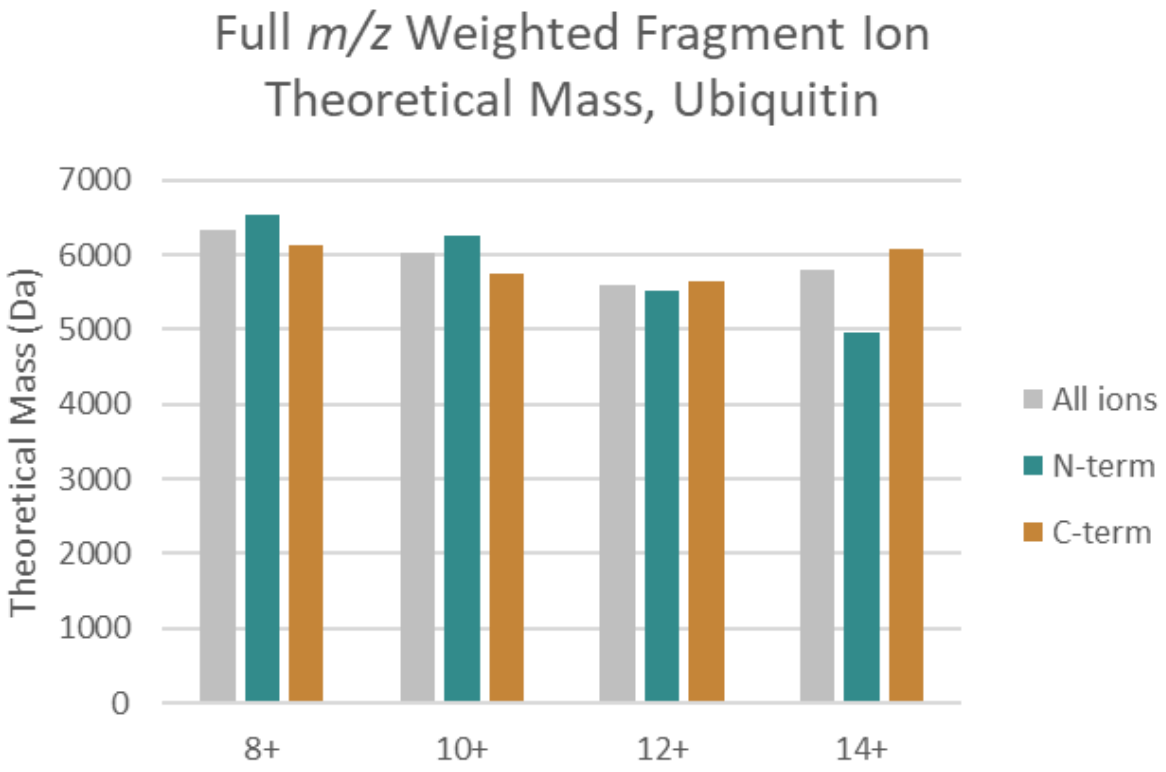

Figure 16. Weighted fragment ion mass of ubiquitin from UVPD spectra of 8, 10, 12, and 14+ precursors.

Conclusion

While UVPD still displays a high degree of fragmentation with a lower degree of preferential cleavages when considering full *m/z* spectra, fractionation strategies aimed at resolving additional information from high-noise spectra appear to confer a degree of bias in the types of ions detected, particularly in the detection of N and C-terminal ions in lower and higher *m/z* fractionated regions. The above analysis methods, dissecting fragment ion information obtained from full and lower and higher-fractionated *m/z*, provide some insight into the behavior of fragment ions in the gas phase but do not fully resolve a model for rationalizing the persistence of N and C-terminal bias, a seemingly intrinsic property to each protein subjected to UVPD-MS conditions.

Mass and charge data collected from full *m/z* spectra often displayed contradictory or negligible differences between N and C-terminal spectra, but the presence and variability of backbone fragment ions across non-supercharged and supercharged precursor charge states proved particularly interesting, possibly showcasing the presence of variable structures within "conformational families" of proteins to allow fragmentation of different residues sites while displaying biases towards residue identity and N/C-terminal biases. Future examinations may

opt for analysis of broader set of proteins in order to more fully resolve the trends and understand outliers to the initial basicity-driven trends for N/C-terminal biases, such as the C-terminal preferentialism for carbonic anhydrase fragment ions in lower *m/z*, as well as examination of lower "native-like" charge states and protein confirmations to assess the universality of N/C-terminal biases for fractionation strategies of UV-photodissociation of intact proteins.

References

(1) Peters-Clarke, T.; Coon, J.; Riley, N. Instrumentation at the Leading Edge of Proteomics. ChemRxiv April 22, 2024. https://doi.org/10.26434/chemrxiv-2023-8l72m.

(2) 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* (3), 520–531. https://doi.org/10.1007/s13361-015-1306-8.

(3) Brodbelt, J. S. Ion Activation Methods for Peptides and Proteins. *Anal. Chem.* **2016**, *88* (1), 30–51. https://doi.org/10.1021/acs.analchem.5b04563.

(4) Roepstorff, P.; Fohlman, J. Letter to the Editors. *Biomed. Mass Spectrom.* **1984**, *11* (11), 601–601. https://doi.org/10.1002/bms.1200111109.

(5) R. Julian, R. The Mechanism Behind Top-Down UVPD Experiments: Making Sense of Apparent Contradictions. *J. Am. Soc. Mass Spectrom.* **2017**, *28* (9), 1823–1826. https://doi.org/10.1007/s13361-017-1721-0.

(6) Lanzillotti, M.; Brodbelt, J. S. Comparison of Top-Down Protein Fragmentation Induced by 213 and 193 Nm UVPD. *J. Am. Soc. Mass Spectrom.* **2023**, *34* (2), 279–285. https://doi.org/10.1021/jasms.2c00288.

(7) Lermyte, F.; Tsybin, Y. O.; O'Connor, P. B.; Loo, J. A. Top or Middle? Up or Down? Toward a Standard Lexicon for Protein Top-Down and Allied Mass Spectrometry Approaches. *J. Am. Soc. Mass Spectrom.* **2019**, *30* (7), 1149–1157. https://doi.org/10.1007/s13361-019-02201-x.

(8) Juetten, K. J.; Brodbelt, J. S. Top-Down Analysis of Supercharged Proteins Using Collision-, Electron-, and Photon-Based Activation Methods. *J. Am. Soc. Mass Spectrom.* **2023**, *34* (7), 1467–1476. https://doi.org/10.1021/jasms.3c00138.

(9) Dunham, S. D.; Brodbelt, J. S. Enhancing Top-Down Analysis of Proteins by Combining Ultraviolet Photodissociation (UVPD), Proton-Transfer Charge Reduction (PTCR), and Gas-Phase Fractionation to Alleviate the Impact of Nondissociated Precursor Ions. *J. Am. Soc. Mass Spectrom.* **2024**, *35* (2), 255–265. https://doi.org/10.1021/jasms.3c00351.

(10) Walker, J. N.; Lam, R.; Brodbelt, J. S. Enhanced Characterization of Histones Using 193 Nm Ultraviolet Photodissociation and Proton Transfer Charge Reduction. *Anal. Chem.* **2023**, *95* (14), 5985–5993. https://doi.org/10.1021/acs.analchem.2c05765.

(11) Crittenden, C. M.; Novelli, E. T.; Mehaffey, M. R.; Xu, G. N.; Giles, D. H.; Fies, W. A.; Dalby, K. N.; Webb, L. J.; Brodbelt, J. S. Structural Evaluation of Protein/Metal Complexes via Native Electrospray Ultraviolet Photodissociation Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* **2020**, *31* (5), 1140–1150. https://doi.org/10.1021/jasms.0c00066.

(12) 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* (28), 9128–9135. https://doi.org/10.1021/jacs.5b04628.

(13) 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. https://doi.org/10.1002/pmic.201570050.

(14) Juetten, K. J.; Brodbelt, J. S. MS-TAFI: A Tool for the Analysis of Fragment Ions Generated from Intact Proteins. *J. Proteome Res.* **2023**, *22* (2), 546–550. https://doi.org/10.1021/acs.jproteome.2c00594.

(15) Durbin, K. R.; Robey, M. T.; Voong, L. N.; Fellers, R. T.; Lutomski, C. A.; El-Baba, T. J.; Robinson, C. V.; Kelleher, N. L. ProSight Native: Defining Protein Complex Composition from Native Top-Down Mass Spectrometry Data. *J. Proteome Res.* **2023**, *22* (8), 2660–2668. https://doi.org/10.1021/acs.jproteome.3c00171.

(16) Bush, M. F.; Hall, Z.; Giles, K.; Hoyes, J.; Robinson, C. V.; Ruotolo, B. T. Collision Cross Sections of Proteins and Their Complexes: A Calibration Framework and Database for Gas-Phase Structural Biology. *Anal. Chem.* **2010**, *82* (22), 9557–9565. https://doi.org/10.1021/ac1022953.

(17) Woodall, D. W.; Henderson, L. W.; Raab, S. A.; Honma, K.; Clemmer, D. E. Understanding the Thermal Denaturation of Myoglobin with IMS-MS: Evidence for Multiple Stable Structures and Trapped Pre-Equilibrium States. *J. Am. Soc. Mass Spectrom.* **2021**, *32* (1), 64–72. https://doi.org/10.1021/jasms.0c00075.

Supporting Information

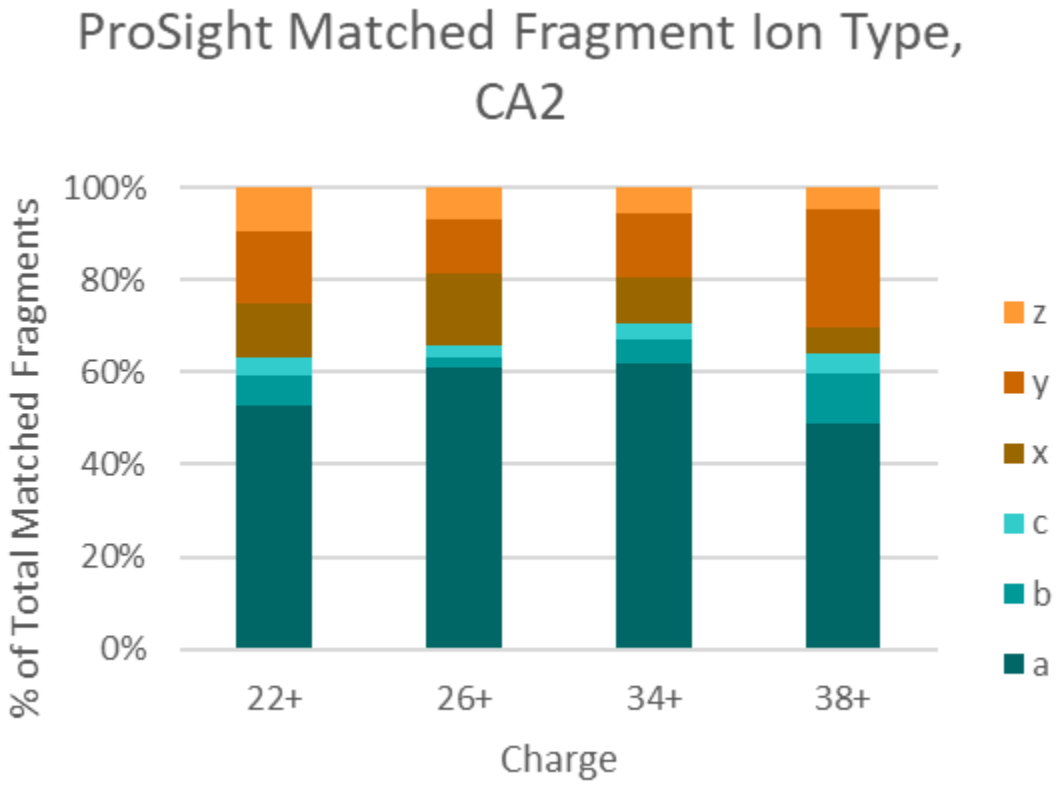

Figure S1. Matched fragment identifications of carbonic anhydrase from full *m/z* UVPD spectra of 22, 26, 34, and 38+ precursors.

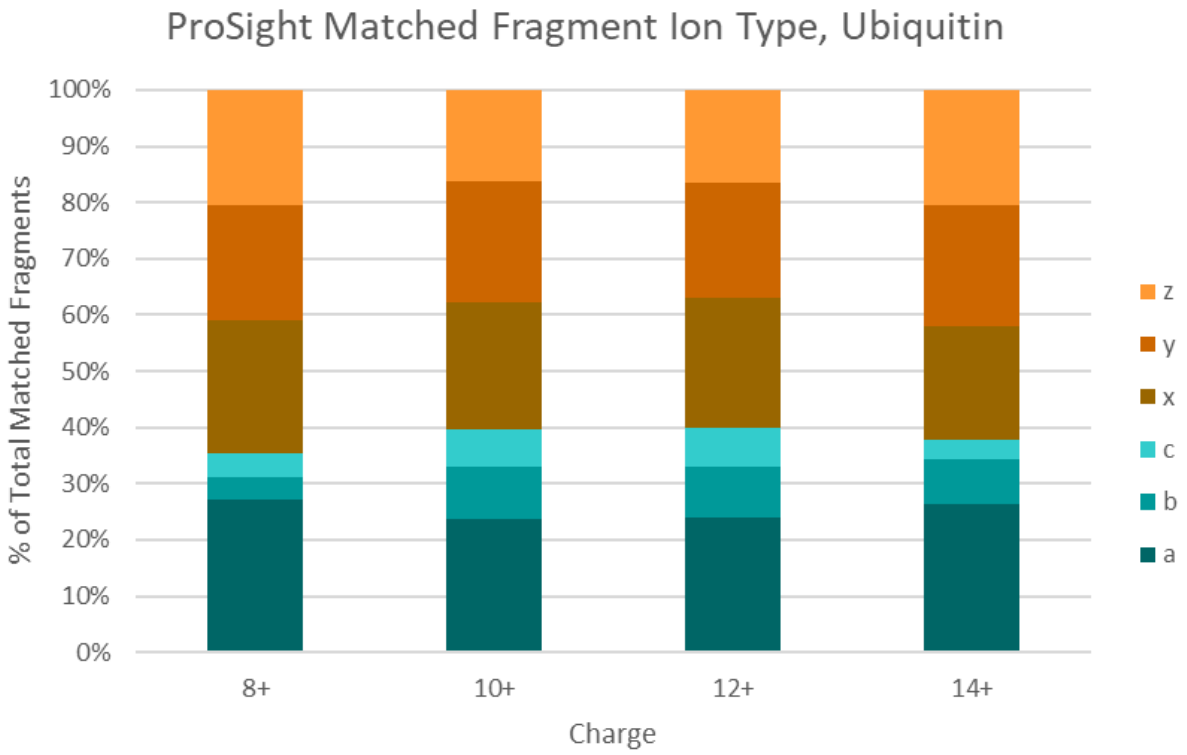

Figure S2. Matched fragment identifications of ubiquitin from full *m/z* UVPD spectra of 8, 10, 12, and 14+ precursors.

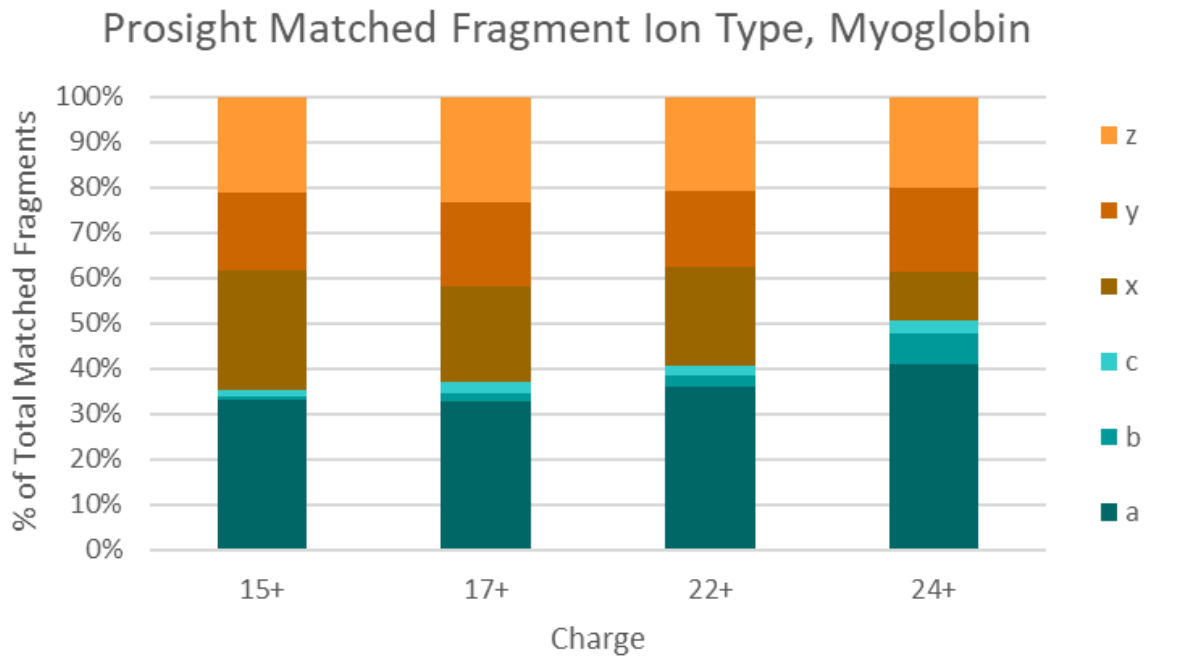

Figure S3. Matched fragment identifications of myoglobin from full *m/z* UVPD spectra of 15, 17, 22, and 24+ precursors.

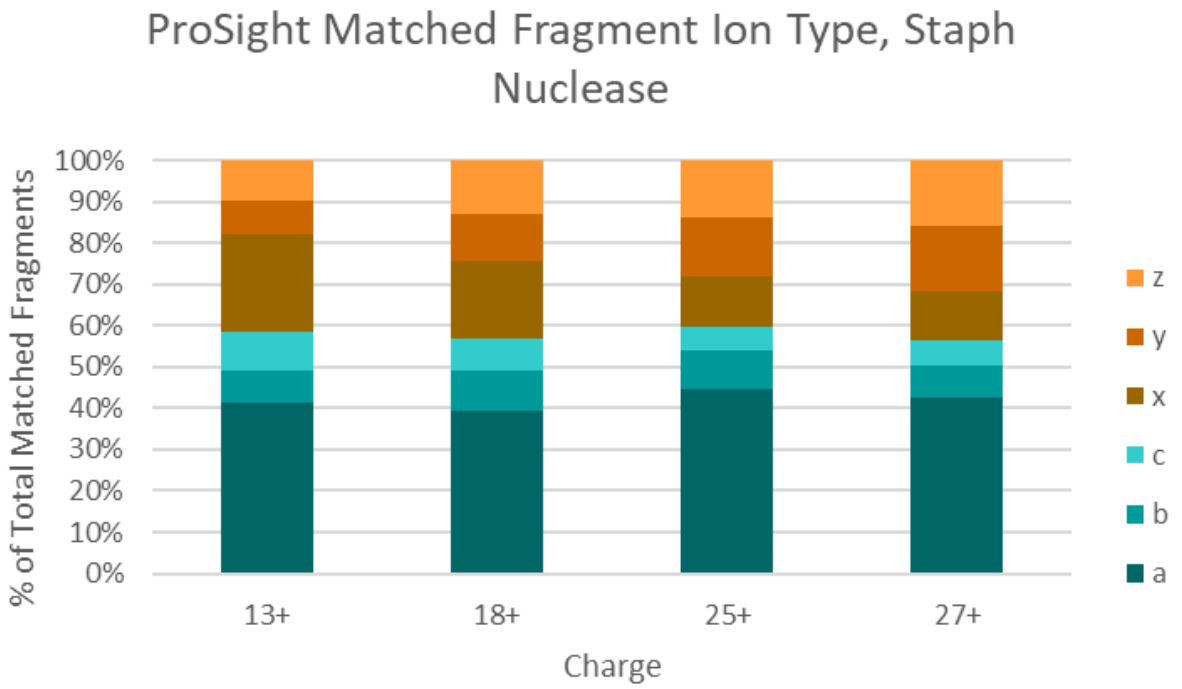

Figure S4. Matched fragment Identification of staph nuclease from full *m/z* UVPD spectra of 13, 18, 25, and 27+ precursors.

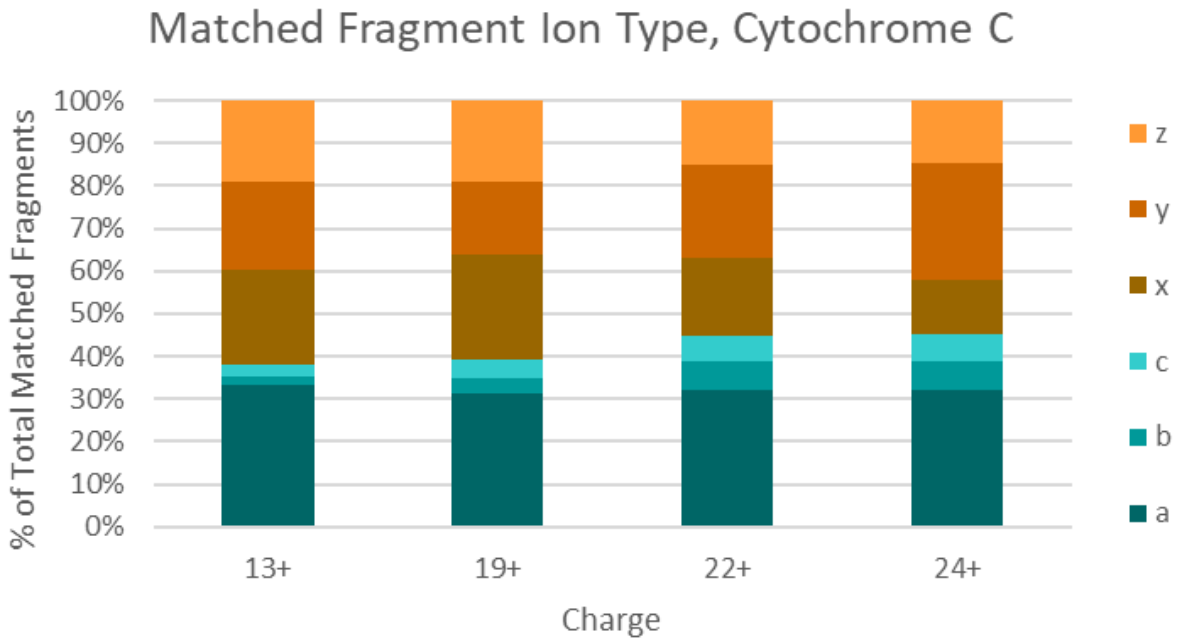

Figure S5. Matched fragment identifications of cytochrome C from full *m/z* UVPD spectra of 13, 19, 22, and 24+ precursors.

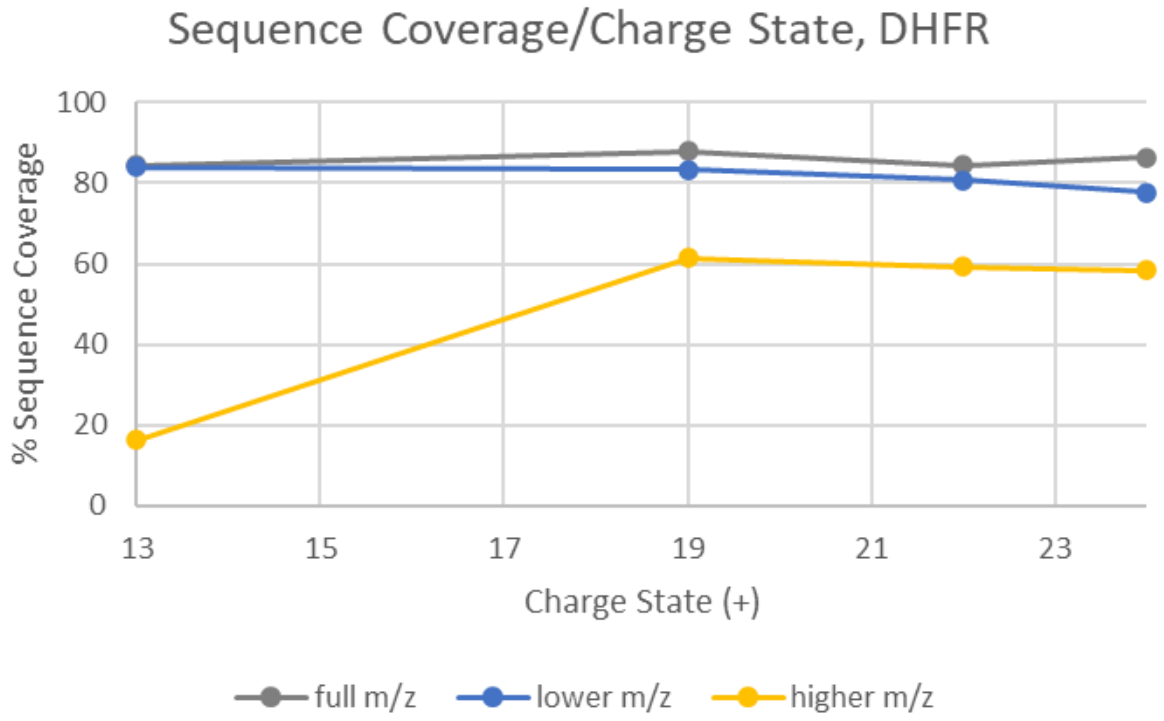

Figure S6. Sequence coverage of dihydrofolate reductase from UVPD spectra of 13, 19, 22, and 24+ precursors.

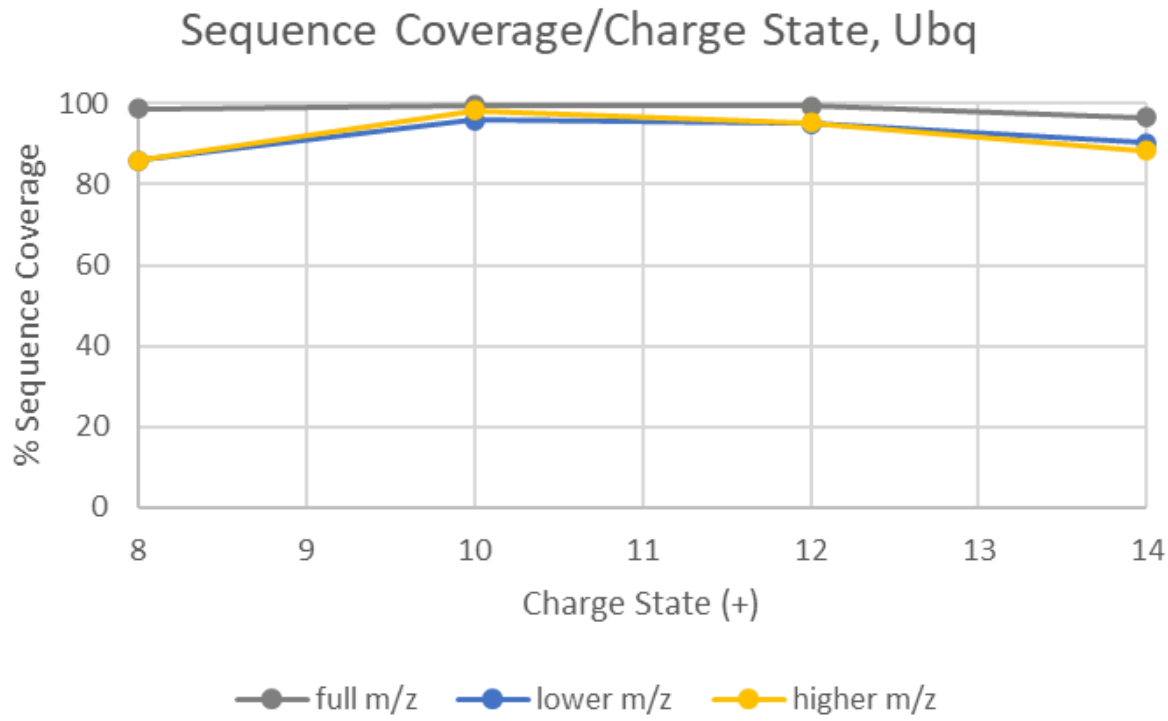

Figure S7. Sequence coverage of ubiquitin from UVPD spectra of 8, 10, 12, and 14+ precursors.

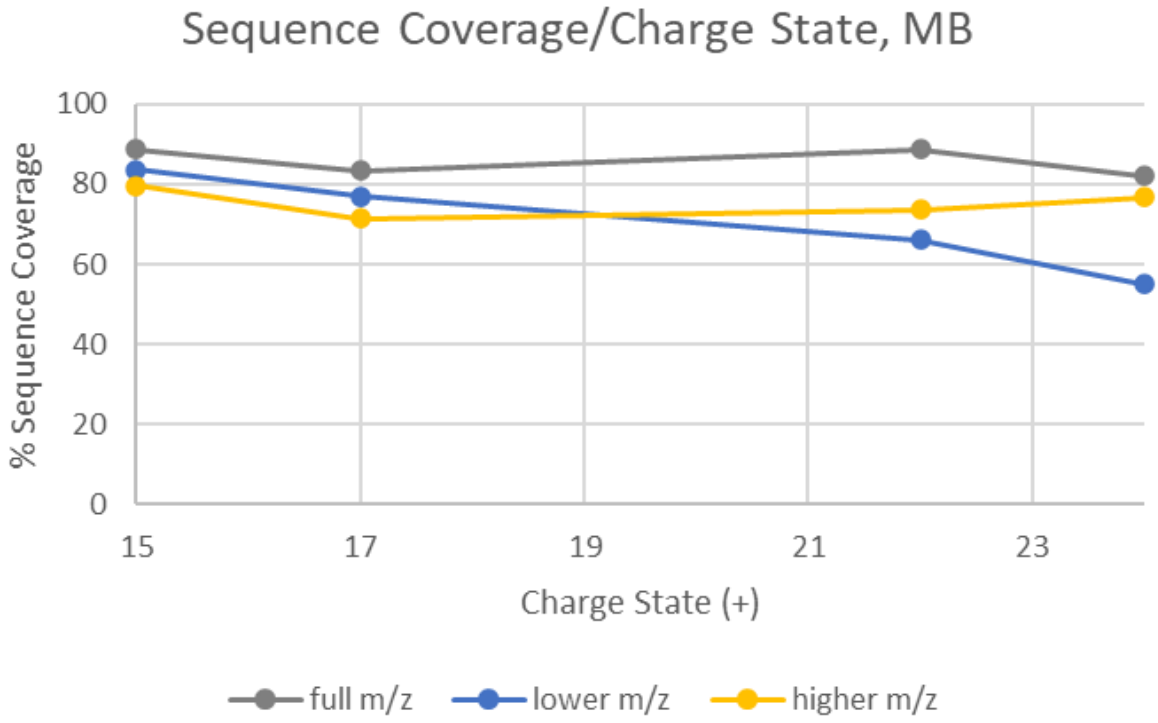

Figure S8. Sequence coverage of myoglobin from UVPD spectra of 15, 17, 22, and 24+ precursors.

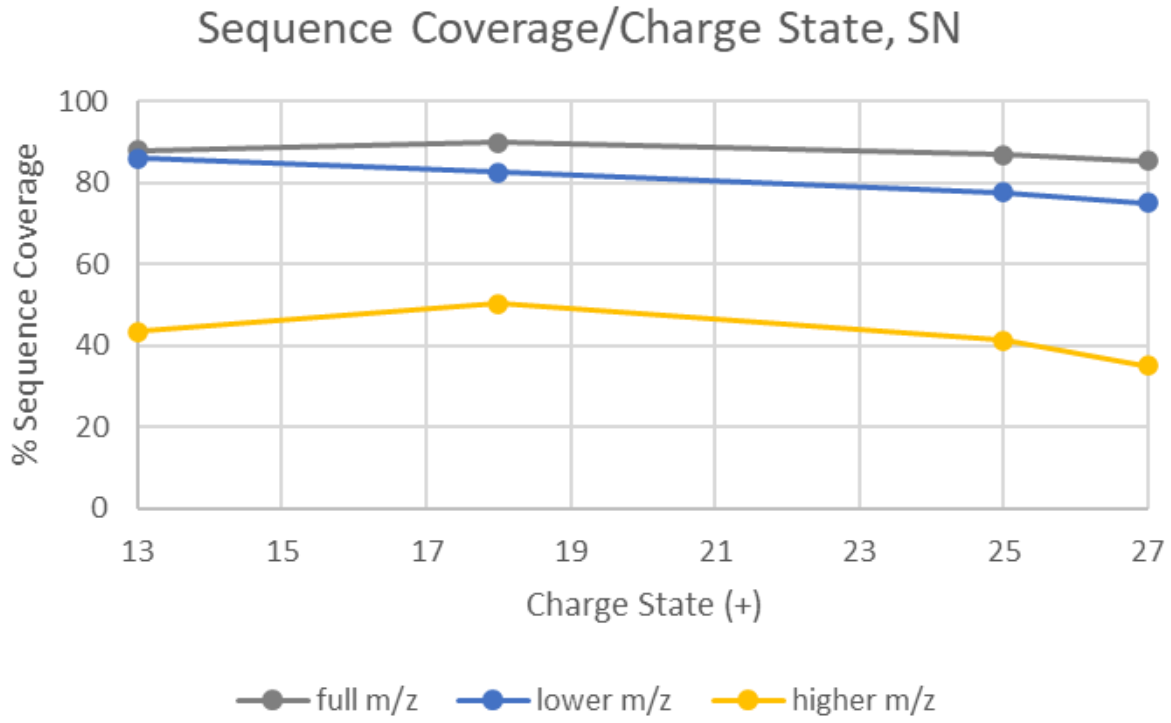

Figure S9. Sequence coverage of staphylococcal nuclease from UVPD spectra of 13, 18, 25, and 27+ precursors.

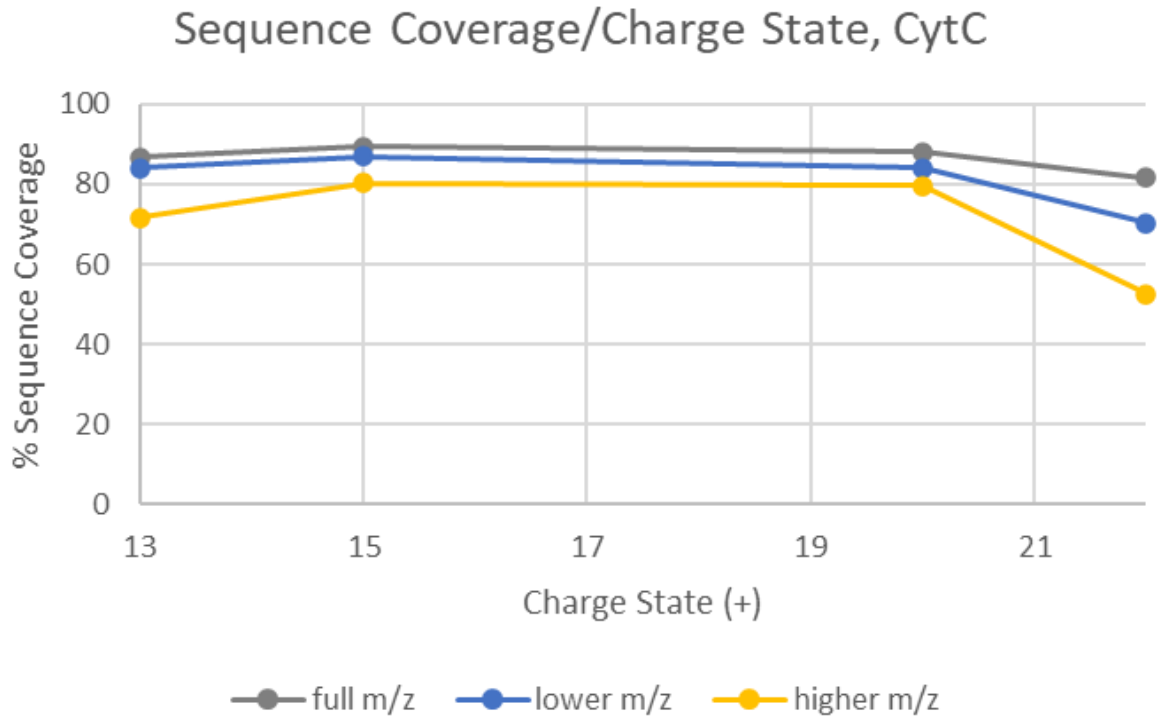

Figure S10. Sequence coverage of cytochrome C from UVPD spectra of 13, 15, 20, and 22+ precursors.

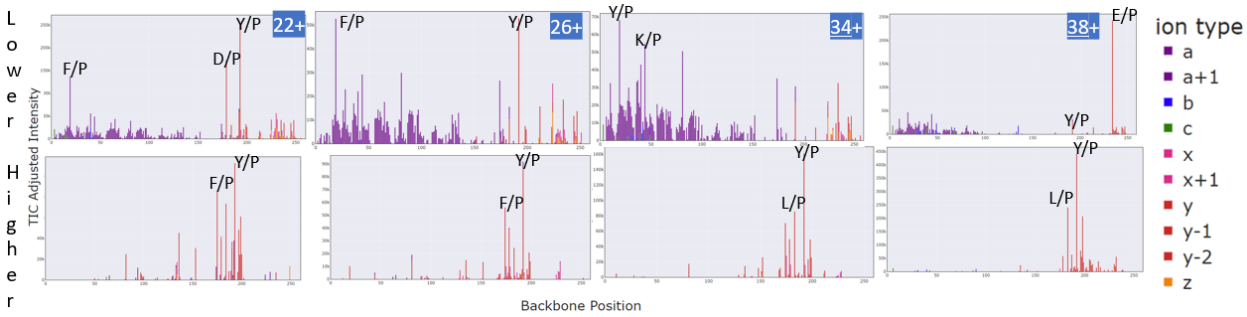

Figure S11. Backbone fragmentation map of carbonic anhydrase garnered from lower and higher *m/z* spectral information from UVPD spectra of 22, 26, 34, and 38+ precursors.

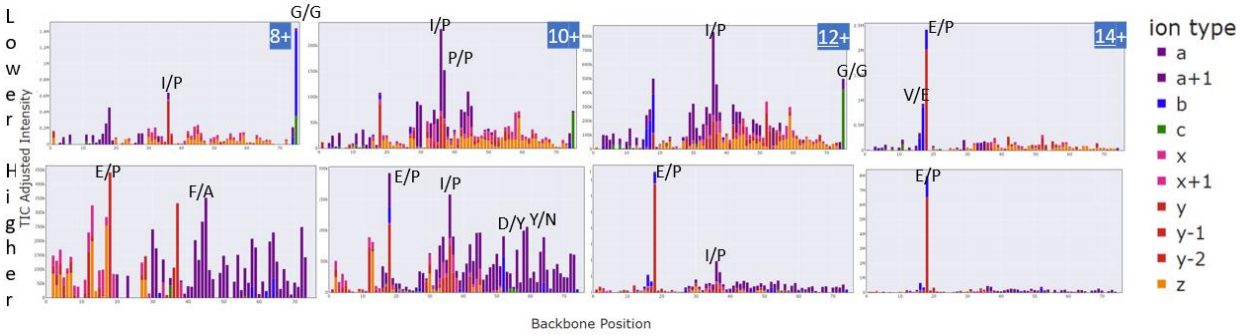

Figure S12. Backbone fragmentation map of ubiquitin garnered from lower and higher *m/z* spectral information of UVPD spectra of 8, 10, 12, and 14+ precursors.

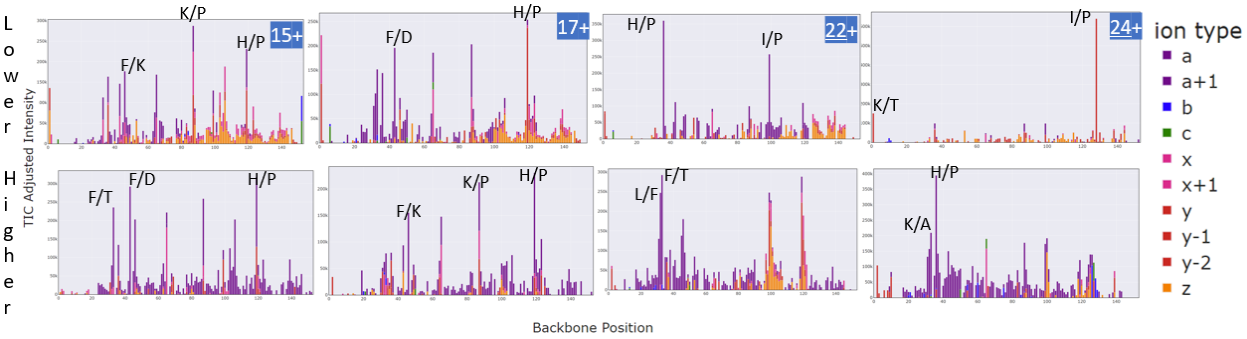

Figure S13. Backbone fragmentation map of myoglobin garnered from lower and higher *m/z* spectral information of UVPD spectra of 15, 17, 22, and 24+ precursors.

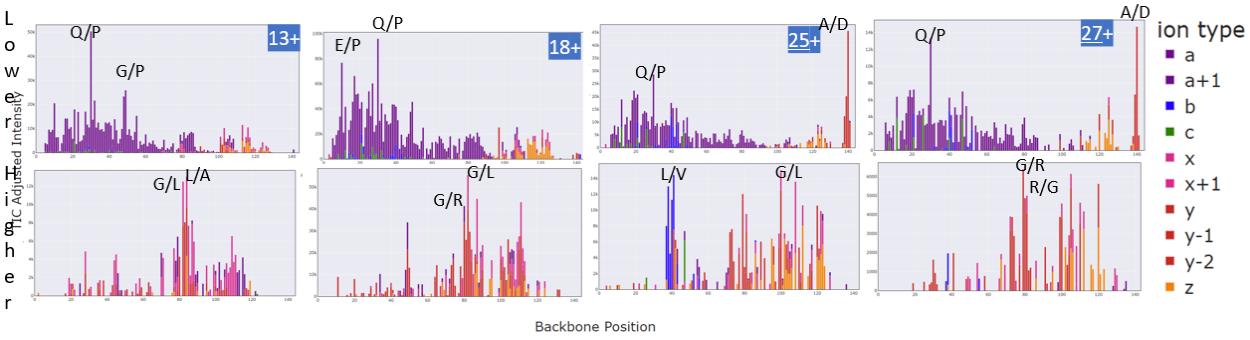

Figure S14. Backbone fragmentation map of staphylococcal nuclease garnered from lower and higher *m/z* spectral information of UVPD spectra of 13, 18, 25, and 27+ precursors.

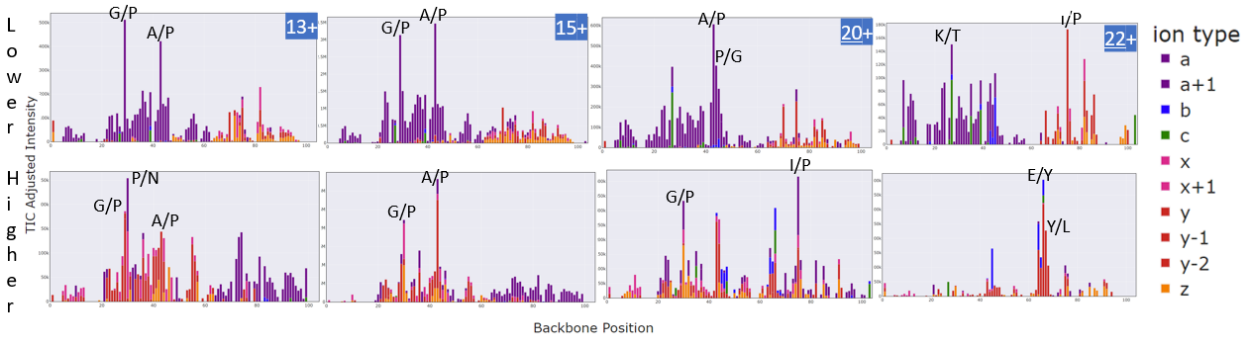

Figure S15. Backbone fragmentation map of cytochrome c garnered from lower and higher *m/z* spectral information of UVPD spectra of 13, 15, 20, and 20+ precursors.

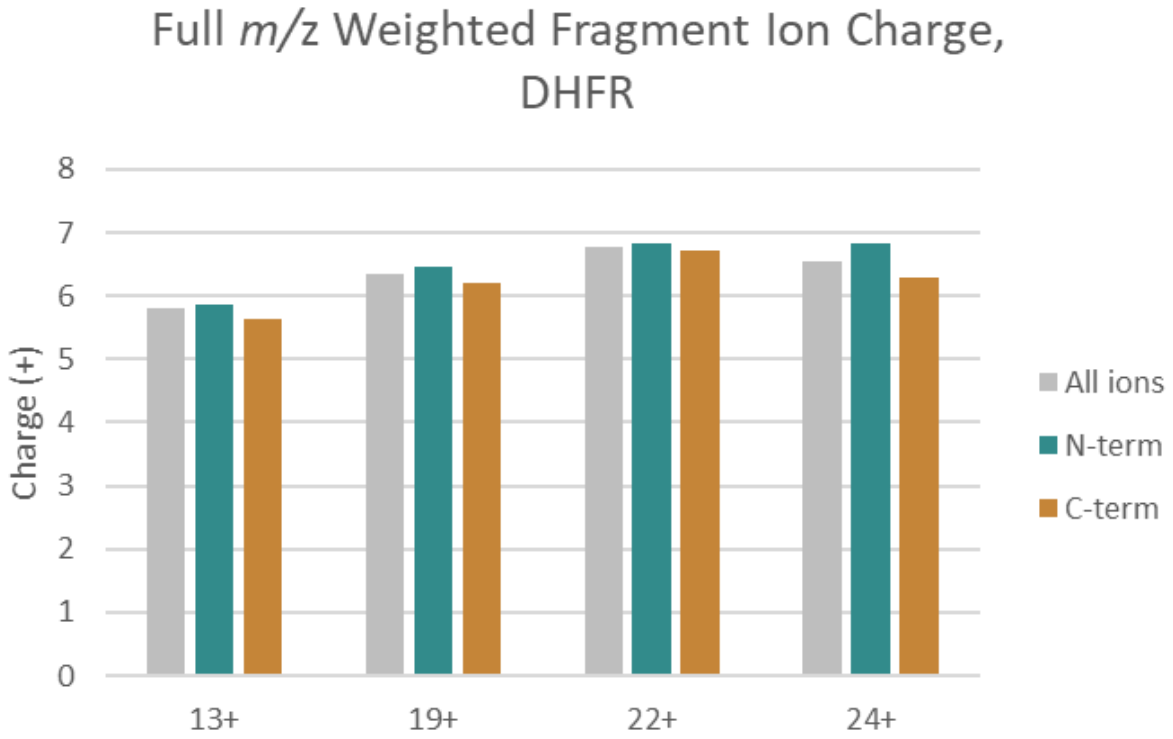

Figure S16. Weighted fragment ion charge of dihydrofolate reductase of UVPD spectra of 13, 19, 22, and 24+ precursors.

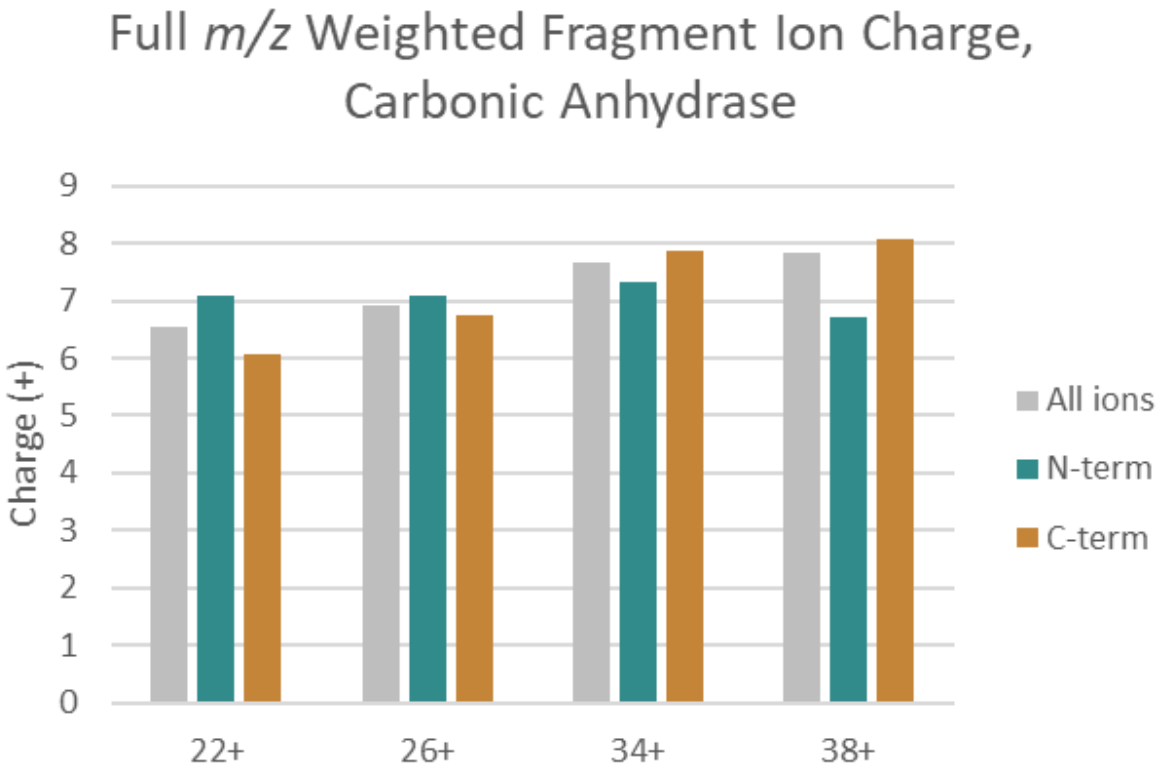

Figure S17 Weighted fragment ion charge of carbonic anhydrase from UVPD spectra of 22, 26, 34, and 38+ precursors.

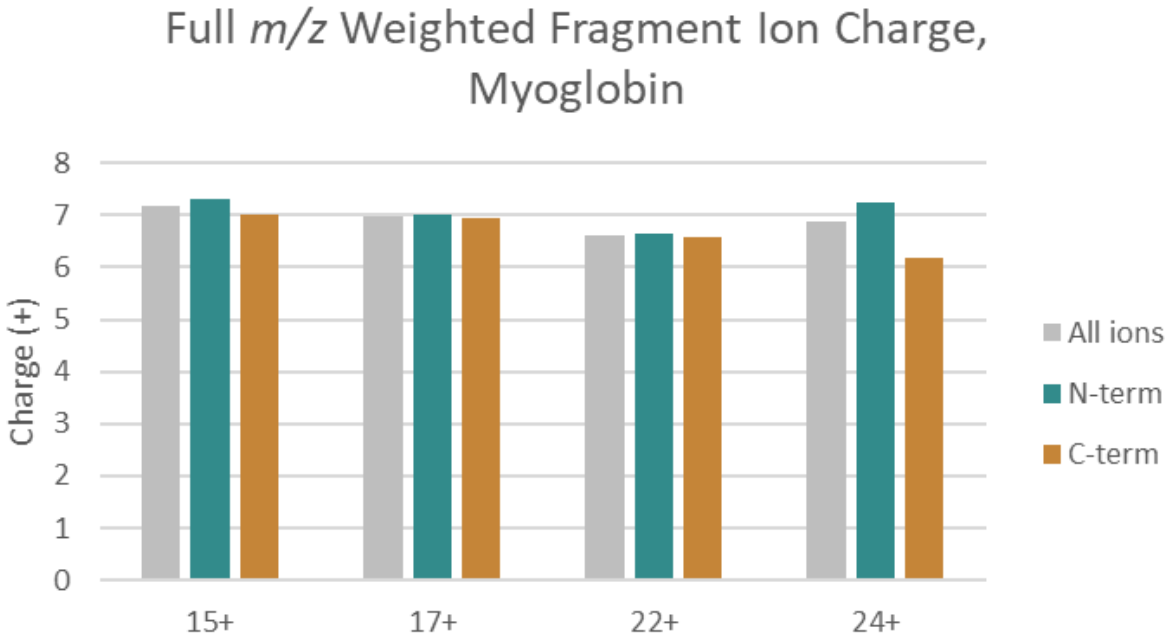

Figure S18. Weighted fragment ion charge of myoglobin from UVPD spectra of 15, 17, 22, and 24+ precursors.

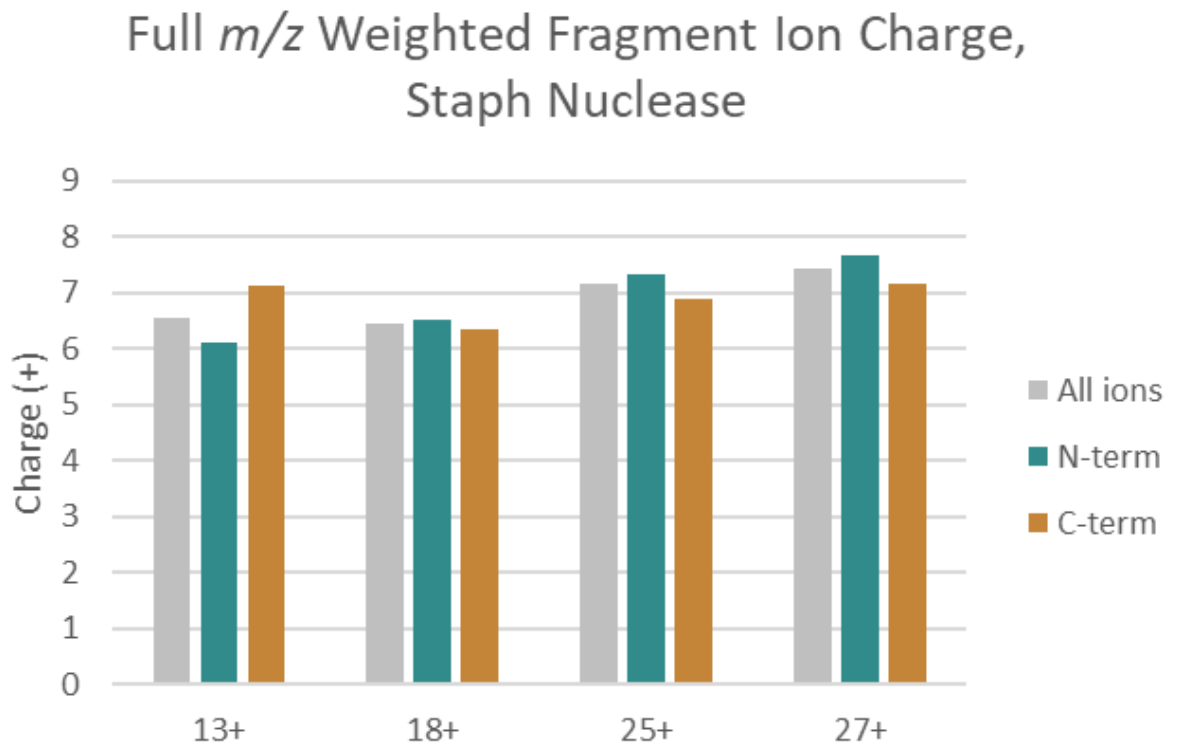

Figure S19. Weighted fragment ion charge of staphylococcal nuclease from UVPD spectra of 13, 18, 25, and 27+ precursors.

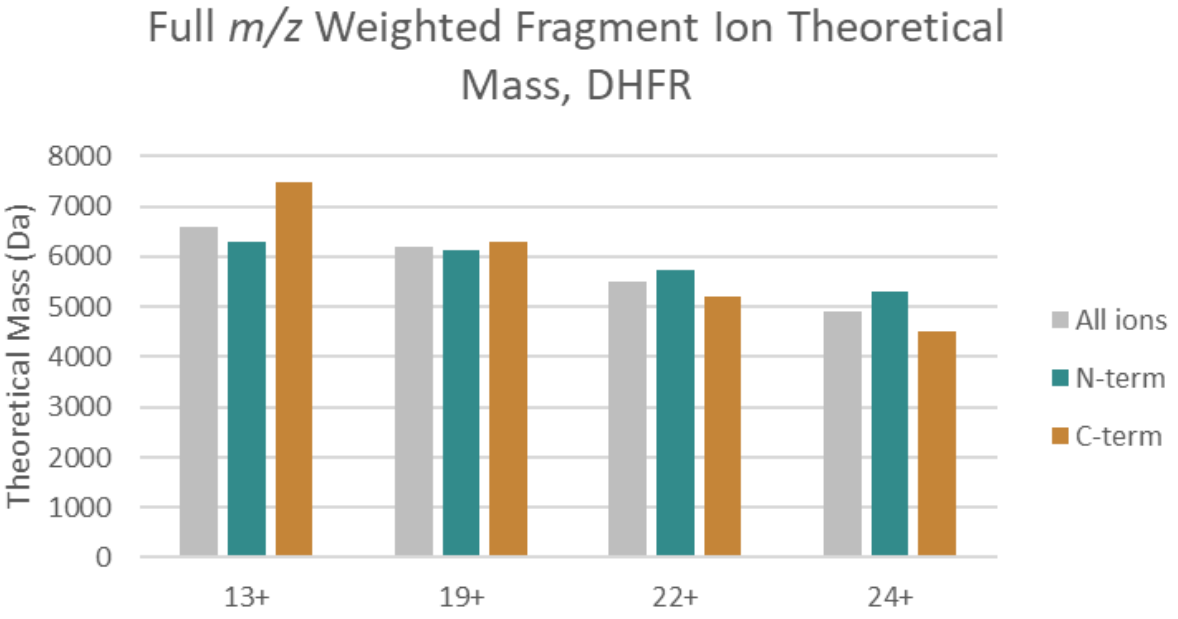

Figure S20. Weighted fragment ion mass of dihydrofolate reductase of UVPD spectra of 13, 19, 22, and 24+ precursors.

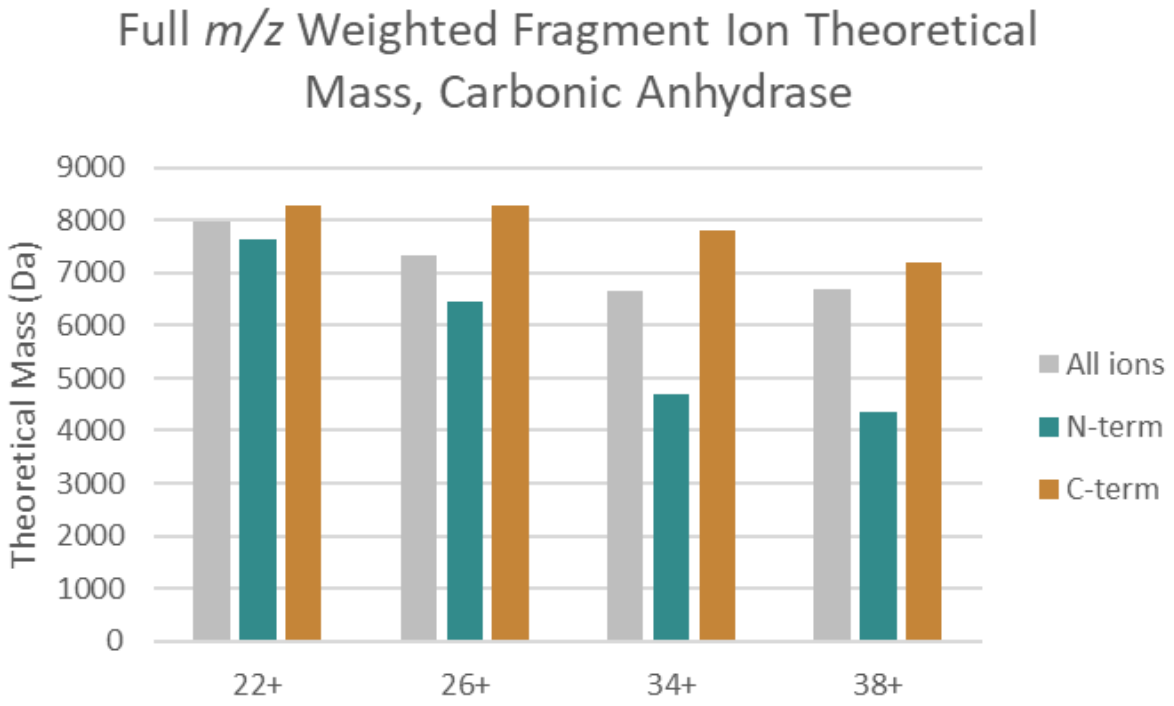

Figure S21. Weighted fragment ion mass of carbonic anhydrase from UVPD spectra of 22, 26, 34, and 38+ precursors.

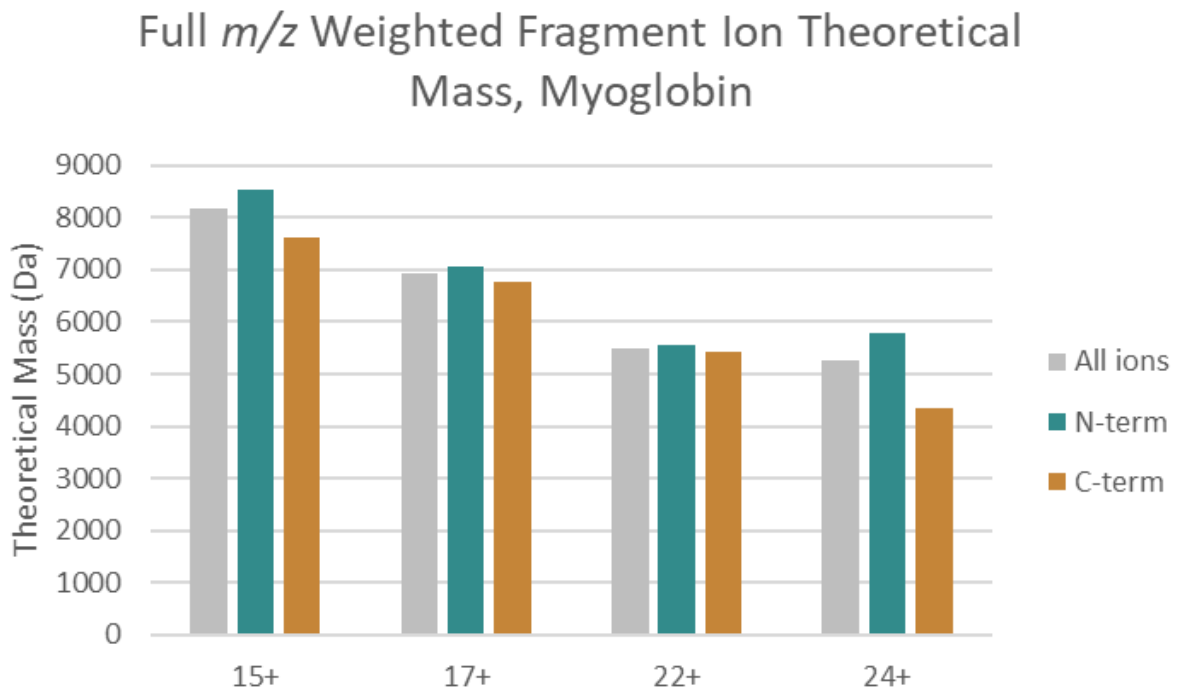

Figure S22. Weighted fragment ion mass of myoglobin from UVPD spectra of 15, 17, 22, and 24+ precursors.

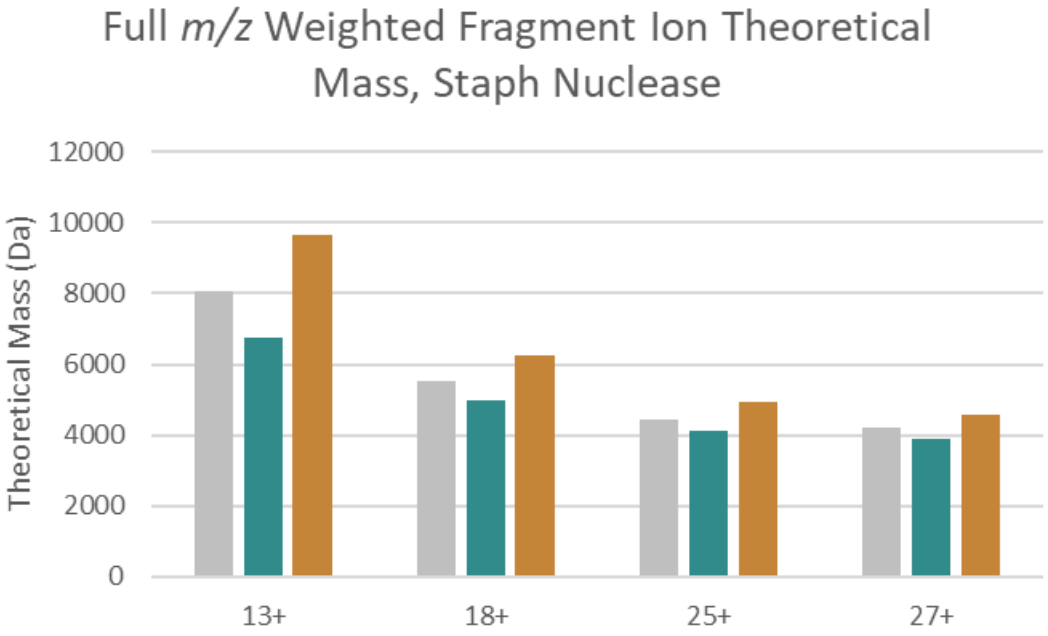

Figure S23. Weighted fragment ion mass of staphylococcal nuclease from UVPD spectra of 13, 18, 25, and 27+ precursors.