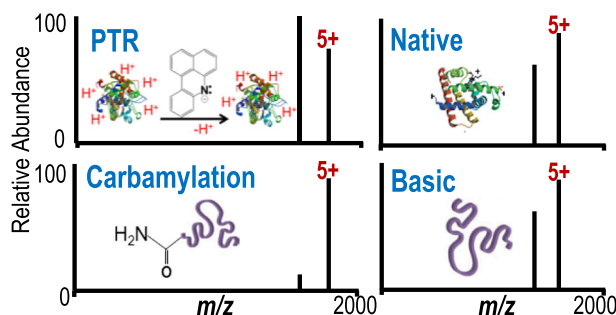


# Top-Down Analysis of Proteins in Low Charge States

Aarti Bashyal, James D. Sanders, Dustin D. Holden, Jennifer S. Brodbelt 

Department of Chemistry, The University of Texas at Austin, Austin, TX 78712, USA



**Abstract.** The impact of charging methods on the dissociation behavior of intact proteins in low charge states is investigated using HCD and 193 nm UVPD. Low charge states are produced for seven different proteins using the following four different methods: (1) proton transfer reactions of ions in high charge states generated from conventional denaturing solutions; (2) ESI of proteins in solutions of high ionic strength to enhance retention of folded native-like conformations; (3)

ESI of proteins in high pH solutions to limit protonation; and (4) ESI of carbamylated proteins. Comparison of sequence coverages, degree of preferential cleavages, and types and distribution of fragment ions reveals a number of differences in the fragmentation patterns depending on the method used to generate the ions. More notable differences in these metrics are observed upon HCD than upon UVPD. The fragmentation caused by HCD is influenced more significantly by the presence/absence of mobile protons, a factor that modulates the degree of preferential cleavages and net sequence coverages. Carbamylation of the lysines and the N-terminus of the proteins alters the proton mobility by reducing the number of proton-sequestering, highly basic sites as evidenced by decreased preferential fragmentation C-terminal to Asp or N-terminal to Pro upon HCD. UVPD is less dependent on the method used to generate the low charge states and favors non-specific fragmentation, an outcome which is important for obtaining high sequence coverage of intact proteins.

**Keywords:** Top-down, Ultraviolet photodissociation, Protein, Low charge state

Received: 18 November 2018/Revised: 30 January 2019/Accepted: 30 January 2019/Published Online: 22 February 2019

## Introduction

Top-down proteomics has emerged as a powerful tool for the analysis of intact proteins as it allows both accurate mass determination and the potential for unambiguous identification of a proteoform based on the production of sufficient diagnostic fragment ions from MS/MS to localize the number and types of modifications [1–3]. The term “proteoform” refers to different variations of a protein that arise owing to mutations, alternative splicing at the RNA level, and post-translational modifications (PTMs) at specific positions within the sequence [3]. Physiologically, proteins often exist as numerous proteoforms, each potentially of distinct function. For example, phosphorylation of kinases often activates their catalytic

activity. Also, methylation and acetylation of histone tails modulate gene transcription. Top-down proteomics has been most successful for characterization of proteoforms below 30 kDa and has yet fully untapped capabilities for characterization of even larger proteins [4]. The development of high-performance mass analyzers and sophisticated data analysis software has contributed to the advancement of top-down proteomics as a desirable tool for proteoform characterization [2], but there also remains significant room for improvement in the understanding of the factors that influence protein fragmentation. These factors include amino acid composition, charge state, charge sites, secondary structure, and conformation, among others, all of which modulate the fragmentation patterns generated upon activation of proteins in MS/MS analysis [5–11].

Considerable effort has been dedicated to optimizing, improving, and developing new activation methods for characterization of proteins [1]. Collisional activation methods like collision-induced dissociation (CID) and higher energy

**Electronic supplementary material** The online version of this article (<https://doi.org/10.1007/s13361-019-02146-1>) contains supplementary material, which is available to authorized users.

Correspondence to: Jennifer Brodbelt; e-mail: jbrodbelt@cm.utexas.edu

collisional dissociation (HCD or beam-type CID) [12] are the most widely used and well understood activation methods for proteomics [13]. Electron-based methods, such as electron capture dissociation (ECD) and electron transfer dissociation (ETD), have also evolved as methods complementary to collisional activation owing to their notable ability to map post-translational modifications [14–16]. However, the success of all of these methods relies on the number of protons and/or charge density of the protein ions, factors which influence the sequence information that can be obtained [5, 6]. More recently, ultraviolet photodissociation (UVPD) using 193 nm photons has demonstrated the most extensive fragmentation of intact proteins and does not exhibit a significant dependence on charge state [17–21]. This finding is attributed to the absorption of high-energy photons (6.4 eV per photon) which results in excited electronic states, thus accessing high-energy fragmentation pathways that lead to the formation of *a/x*- and *c/z*-type fragment ions in addition to *b/y* ions commonly produced upon collisional activation [22].

Combining the use of complementary activation methods, like CID, HCD, ECD, ETD, and UVPD, affords one strategy for increasing the information content in top-down proteomics, and further gains can be obtained by modulating the charge sites and charge states of proteins [1, 13]. In general, proteins sprayed from denaturing solutions produce a wide range of higher charge states, and this approach has been the conventional domain of high-throughput top-down proteomics strategies that use organic solvents for chromatographic separations [5, 6, 9]. In contrast, proteins sprayed from aqueous solutions with high salt content (typically ammonium acetate) generate native-like folded proteins in low charge states [23, 24]. Charge state distributions of proteins can also be lowered by proton transfer reactions (PTR) in the gas phase (ion/ion reaction with an anion) [5, 25] or by carbamylation reactions [18] in solution to convert the primary amines of N-termini and lysine residues to less basic carbamates. Likewise, the addition of volatile bases, like ammonium hydroxide, to solutions results in deprotonation of proteins and, thus, production of lower charge states [26]. In contrast, charge states of proteins can be increased by chemical derivatization of primary amines to more basic groups [27] or by adding supercharging reagents [28, 29] to the solutions prior to ESI. This array of methods for altering charge states raises many fundamental questions about the impact of charge state on protein fragmentation.

It is well-established that the outcome of collisional activation of peptides and proteins depends on the presence of mobile protons [30]. According to the mobile proton model, protons are initially located at the most basic sites, and the addition of internal energy to the ions mobilizes the protons to promote different fragmentation pathways [30]. However, when fewer charges are present than the number of basic sites, these charges remain sequestered at the most basic sites (e.g., side chains of Lys and Arg), and fragmentation occurs via charge-remote pathways [7, 30]. Several systematic studies have focused on modulating the charge states of proteins to understand the role of the charge state on the fragmentation pathways from

collisional activation. For example, Reid et al. used PTR to generate low charge states ( $1^+$  to  $5^+$ ) of ubiquitin and demonstrated that low proton mobility enhanced cleavages of C-terminal to aspartic acid residues and caused loss of ammonia or water from the precursor ion, whereas higher proton mobility (precursor charge states of  $7^+$  to  $9^+$  of ubiquitin) resulted in non-specific backbone cleavage throughout the protein [5]. Chanthamontri et al. also demonstrated that the level of sequence information obtained from both CID and HCD methods correlated with the proton mobility and charge state of  $\alpha$ -synuclein [6]. Overall, these studies established that proteins in low charge states do not fragment well by collisional activation and are dominated by fragmentation C-terminal to acidic residues. Alternatively, Iavarone et al. reported that collisional activation of highly charged proteins generated by adding *m*-nitrobenzyl alcohol to solution prior to ESI resulted in a few predominant backbone cleavage sites along with limited backbone fragmentation at nearby sites, outcomes governed by the location of the charges [28]. Similarly, Pitteri et al. reported that after the guanidination of lysines, the fragmentation behavior of ubiquitin ( $10^+$ ) was similar to that obtained for lower charge states (i.e., increased fragmentation of C-terminal to acidic residues) owing to decreased proton mobility [31].

More recently, Haverland et al. examined the overall fragmentation patterns of native-like and denatured proteins and mapped the probability of backbone bond cleavages between hundreds of different residue pairs based on HCD [9]. Cleavages of C-terminal to aspartic acid, between phenylalanine and tryptophan, between tryptophan and alanine, and N-terminal to proline were more enhanced for proteins in low charge states compared to those in high charge states [9]. Similarly, Greer et al. reported that the fragmentation patterns of carbamylated proteins obtained by HCD were significantly altered compared to the patterns obtained for non-carbamylated proteins in identical charge states, and overall sequence coverage was substantially decreased for the carbamylated proteins [18]. However, little to no changes were observed for fragmentation caused by UVPD of the same proteins prior to and after carbamylation [18]. It was rationalized that carbamylation had an impact on protein fragmentation by HCD because the primary amines of lysines and the N-terminus were converted to less basic carbamates, thus changing proton mobility.

Harnessing the strategic use of charge state modulation for intact protein analysis requires a better understanding of the impact of both charge state and charge site on protein fragmentation. Here, we focus on examining the dissociation behavior of intact proteins in low charge states using HCD and 193 nm UVPD. Low charge states are generated for seven proteins using the following four different methods: (1) proton transfer reactions in the gas phase of proteins sprayed from conventional denaturing solutions; (2) ESI of proteins in solutions of high ionic strength to enhance retention of folded “native-like” conformations; (3) ESI of proteins in high pH solutions to limit protonation; and (4) ESI of carbamylated proteins. Comparison of the sequence coverage, the degree of preferential cleavages, the types and distribution of fragment ions, and the charge site

assignment is undertaken for both HCD and UVPD to demonstrate the effect of charging methods on the fragmentation of proteins.

## Experimental

### Materials

Ubiquitin (bovine), cytochrome c (equine), transthyretin (human), myoglobin (equine), and urea were obtained from Sigma-Aldrich (St. Louis, MO, USA). Alpha synuclein (human, expressed in *Escherichia coli*) was obtained from rPeptide. Calmodulin (human) and staphylococcal nuclease (*Staphylococcus aureus*) were expressed as described previously [32]. Solvents were purchased from Thermo Fisher Scientific (Pittsburgh, PA). Proteins were cleaned up by using Bio-Spin 6 columns from Bio-Rad Laboratories (Hercules, CA).

### Carbamylation

The carbamylation reactions were carried out by incubating ~20 µg protein in 100 mM ammonium bicarbonate at 80 °C for 4–5 h in the presence of 8 M urea, as reported previously [33]. Carbamylated proteins were cleaned up using Amicon Ultra 3 kDa MWCO filters from EMD Millipore (Billerica, MA) or Bio-Spin 6 from Bio-Rad and buffer exchanged into water. The efficiency of carbamylation was estimated to be above 90% except for transthyretin that showed more heterogeneous distribution of carbamylated species.

### Mass Spectrometry

Proteins were diluted to a concentration of 10 µM. For experiments involving carbamylated proteins or PTR, proteins were prepared from solutions of water, acetonitrile, and formic acid (49.5/49.5/1). Proteins in basic solutions were prepared in solutions of water, acetonitrile, and ammonium hydroxide (49.5/49.5/1), and proteins for native MS experiments were prepared in 100 mM ammonium acetate solution (pH 6.9). Samples were analyzed on a Thermo Scientific Orbitrap Elite mass spectrometer (Thermo Fisher Scientific, Bremen, Germany). UVPD was performed in the HCD cell using a single 5 ns pulse of 193 nm photons from a Coherent ExciStar excimer laser (Santa Clara, CA) as described previously [19]. All spectra were collected at a resolving power of 120,000 at  $m/z$  400 averaged over 300 scans. Samples were infused with a static emitter, and the spray voltage was varied between 0.9 and 1.2 kV to obtain the optimal spray stability. The automated gain control (AGC) target for MS1 was set to 1E6 and 5E5 for MS2, and the ion isolation width was set to 20  $m/z$ . This isolation width is commonly used for top-down analysis, and it means that co-isolation and co-activation of neighboring adducts are possible. PTR was performed in the high-pressure cell of the linear ion trap as previously described [25], and the automatic gain control target of the reagent anion (nitrogen-adducted fluoranthene anions of  $m/z$  216) was set to 2E5. The

normalized collision energy (NCE) for HCD was optimized between 20 and 35%, and the laser energy for UVPD was optimized between 2.5 and 4 mJ with a single 5 ns pulse to obtain the most sequence information. Collision cross-sections were measured in the Orbitrap mass analyzer as described previously [34].

### Data Analysis

MS2 spectra were deconvoluted using the Xtract algorithm (Xcalibur Qual Browser, Thermo Fisher Scientific) with a S/N threshold of 2. ProSight Lite (<http://prosiglitem.northwestern.edu/>) was used to generate sequence coverage maps using an error tolerance of 10 ppm [35]. Fragment intensity maps were plotted in Microsoft Excel using the output from MS Product (<http://prospector.ucsf.edu/prospector/cgi-bin/msform.cgi?form=msproduct>) which was further processed to obtain the abundances and backbone cleavage positions of the ions. Charge site localization plots were generated in Microsoft Excel using the output generated by the Thrash algorithm from ProSight PC and UV-POSIT [36]. Fragmentation efficiency was estimated by comparing the abundances of matched fragment ions to the total ion current.

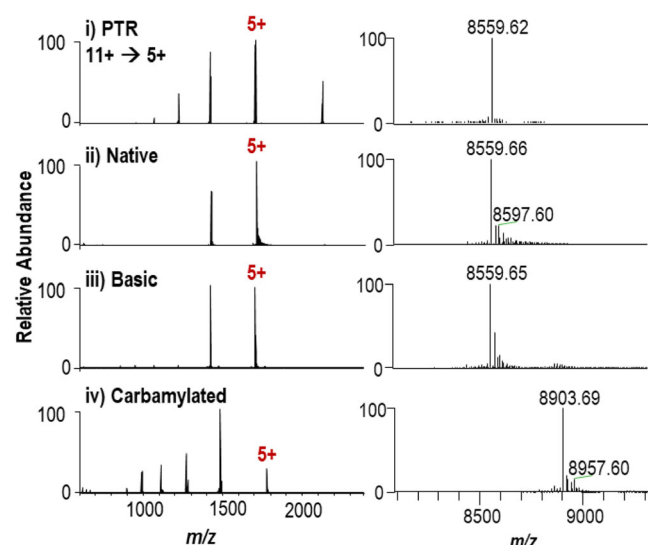
## Results and Discussions

In this study, the fragmentation behavior of the seven proteins in low charge states is evaluated. The proteins range in size from 8.5 to 16.9 kDa and with a varying number of residues that modulate charge states and/or preferential backbone cleavages (arginine, lysine, histidine, proline, aspartic acid, and glutamic acid). The sequences of the seven proteins are listed in Fig. S1 along with their molecular weights and summaries of their compositions. Representative ESI mass spectra are shown in Fig. 1 and Figs. S2–S8. For this systematic study, low charge states of the proteins are produced by the following four different methods: (1) utilizing proton transfer reactions of the proteins in the gas phase, a process which strips protons from more highly charged (denatured) proteins, to decrease their charge states; (2) spraying the proteins in aqueous solutions containing 100 mM ammonium acetate to generate native-like folded structures; (3) spraying the proteins in basic solutions (pH 9 via addition of ammonium hydroxide) to limit the extent of protonation; and (4) undertaking carbamylation reactions to convert primary amines of lysines and the N-terminus into non-basic carbamate groups prior to ESI. Among these methods, three of them (1, 3, 4) are expected to result in denatured protein structures, likely partially unfolded and/or collapsed structures. The second method (ESI of proteins in high ionic strength solutions) is anticipated to yield more native-like conformations. At least one common low charge state is generated for each protein using each of the four methods, and the remainder of the study focuses on MS/MS analysis of this common charge state as follows: ubiquitin ( $5^+$ ) (Fig. 1), cytochrome c ( $7^+$ ) (Fig. S2), transthyretin (TTR) monomer ( $7^+$ )

generated by in-source dissociation of native tetrameric TTR complexes (Fig. S3),  $\alpha$ -synuclein ( $7^+$ ) (Fig. S4), staph nuclease ( $8^+$ ) (Fig. S5), calmodulin ( $7^+$ ) (Fig. S6), and myoglobin ( $8^+$ ) (Fig. S7). The deconvoluted MS1 spectra obtained for each protein by each method are shown in Fig. S8. For two of the proteins ( $\alpha$ -synuclein and staph nuclease), the carbamylation reactions were unsuccessful, an outcome attributed to the precipitation of the proteins or inefficient clean-up which impeded the detection of carbamylated products, and, thus, the carbamylated versions of these two proteins could not be studied.

### Sequence Coverage

One benchmark metric for evaluating the impact of the method used to generate the low charge states of the proteins is sequence coverage. The sequence coverage represents the percentage of the inter-residue bonds that are cleaved in a protein, and this value would not be expected to vary if each of the four methods used to generate the low charge states yielded identical populations of protein ions. For this assessment of sequence coverage, one common low charge state of each protein is isolated and activated by UVPD and HCD, and the results are summarized in Table 1. Ubiquitin is a small protein (8.5 kDa) that is efficiently fragmented by both HCD and UVPD, irrespective of charge state [5], resulting in rich fragmentation patterns for both activation methods (Fig. S9). UVPD results in 96–98% sequence coverage for the  $5^+$  charge state of ubiquitin regardless of the method used to produce the ions; HCD results in 74 to 89% coverage for the four different charging methods (Table 1). The lowest coverage is obtained for the carbamylated form of ubiquitin, an outcome attributed to the restricted mobility of protons when localized at the highly basic



**Figure 1.** MS1 spectra of ubiquitin generated by the following four methods: (a) PTR ( $11^+ \rightarrow 5^+$ , 30 MS), (b) native solution (100 mM ammonium acetate), (c) basic solution (1% ammonium hydroxide), and (d) carbamylated. Corresponding deconvoluted MS1 spectra are shown on the right

arginines. The variation in sequence coverage as a function of the ion production method is more pronounced for larger proteins (Table 1, Figs. S11–S17). For example, the sequence coverage of the  $7^+$  charge state of cytochrome c varies from 33 to 82% for HCD depending on the charging method; the variation for UVPD is less extreme (72 to 93%). Similarly, sequence coverage of transthyretin ( $7^+$ ) ranges from 31 to 67% for HCD and from 58 to 80% for UVPD. Likewise, a strong dependence on the charging method is observed for the sequence coverages of the other four proteins ( $\alpha$ -synuclein, staph nuclease, calmodulin, and myoglobin), particularly for HCD and to a lesser extent for UVPD.

In general, carbamylation consistently results in the most negative impact on sequence coverage obtained by HCD for all of the proteins. For example, the sequence coverage of myoglobin ( $8^+$ ) and cytochrome c ( $7^+$ ) upon carbamylation decreases by more than half upon HCD when compared to the corresponding unmodified proteins, an outcome also reported in a recent study that examined fragmentation of high charge states of proteins [18]. Both myoglobin and cytochrome c contain the greatest number of lysine residues (19 for each) among the proteins included in the present study, and, thus, these two proteins experience a significant shift in the localization of protons after carbamylation. Transthyretin and calmodulin, with fewer lysine residues (eight for each), show smaller overall changes in the sequence coverage obtained by HCD after carbamylation when compared to myoglobin and cytochrome c. There are several factors that may contribute to the decrease in sequence coverage obtained by HCD of the carbamylated proteins. The fragmentation efficiency upon HCD is adversely affected upon carbamylation of the proteins, resulting in lower abundances of fragment ions (Fig. S10). Since the mechanism of HCD relies on the presence of mobile protons [1], the conversion of basic lysines to non-basic sites alters proton mobility in a way that may be particularly detrimental for backbone cleavages. For example, backbone cleavages adjacent to acidic residues are particularly suppressed for the carbamylated proteins upon HCD (as described in more detail later). In addition, the types of intramolecular interactions of the proteins are expected to vary considerably upon carbamylation in which the ionizable lysine residues are converted to non-basic ones. Even if the proteins do not adopt well-defined structures in the gas phase, the proteins in the low charge states targeted in the present study are not anticipated to be elongated but rather more compact, somewhat folded, and/or collapsed structures. The structural differences between carbamylated and non-carbamylated protein may contribute to variations in the fragmentation pathways and dissociation efficiencies. For all proteins analyzed, the reduction in sequence coverage for the carbamylated proteins is less notable for UVPD compared to HCD, and this result reflects the lack of dependence of the mechanism of UVPD on proton mobility.

A somewhat notable decrease in the sequence coverage is also observed upon HCD for several of the proteins in charge states generated by PTR, particularly staph nuclease, calmodulin, cytochrome c, and transthyretin, compared to the same

**Table 1.** Sequence Coverage Summary for HCD and UVPD

Protein	MW (kDa)	Charge State	HCD				UVPD			
			PTR (%)	Native (%)	Basic (%)	Carbam (%)	PTR (%)	Native (%)	Basic (%)	Carbam (%)
Ubiquitin	8.5	5 <sup>+</sup>	89 ± 2	84 ± 4	89 ± 4	74 ± 2	96 ± 1	98 ± 1	96 ± 1	97 ± 1
Cytochrome c <sup>a</sup>	12.3	7 <sup>+</sup>	56 ± 1	80 ± 1	82 ± 1	33 ± 3	80 ± 2	88 ± 1	93 ± 1	72 ± 3
Transthyretin	13.8	7 <sup>+</sup>	45 ± 5	67 ± 4	55 ± 1	31 ± 2	58 ± 3	78 ± 1	80 ± 3	59 ± 1
α-Synuclein	14.5	7 <sup>+</sup>	76 ± 3	75 ± 1	88 ± 1	<sup>b</sup>	76 ± 2	82 ± 1	82 ± 2	<sup>b</sup>
Staph nuclease	16.1	8 <sup>+</sup>	34 ± 1	53 ± 3	59 ± 1	<sup>b</sup>	52 ± 1	79 ± 4	81 ± 1	<sup>b</sup>
Calmodulin	16.6	7 <sup>+</sup>	19 ± 1	43 ± 4	33 ± 1	22 ± 2	50 ± 1	66 ± 2	63 ± 2	46 ± 2
Myoglobin	16.9	8 <sup>+</sup>	43 ± 2	60 ± 1	63 ± 1	19 ± 2	67 ± 3	68 ± 2	62 ± 2	64 ± 2

Results are averaged over three replicates

<sup>a</sup>Molecular mass of cytochrome c contains a covalently bound heme and N-terminal acetylation

<sup>b</sup>Carbamylation of α-synuclein and staph nuclease was not successful

charge states generated from the native or basic solutions. During PTR in the gas phase, multiple protons are removed from a denatured, highly charged protein, and the remaining protons may be located in different positions than a folded protein or one in which basic residues are deprotonated in solution prior to transfer to the gas phase. Depending on where the protons are located, the backbone cleavages and extent of fragmentation obtained by HCD may be altered significantly. In addition, as protons are stripped from a protein during the PTR process, the degree of electrostatic repulsions is decreased, and an “unfolded” denatured protein may adopt a different structure (or population of structures), including ones that are far more compact or collapsed as charge-induced repulsions are alleviated and new intramolecular interactions are formed. The new compact or collapsed structures may differ significantly from folded proteins generated from native-like solutions, thus leading to different fragmentation behavior even if the charge state is the same.

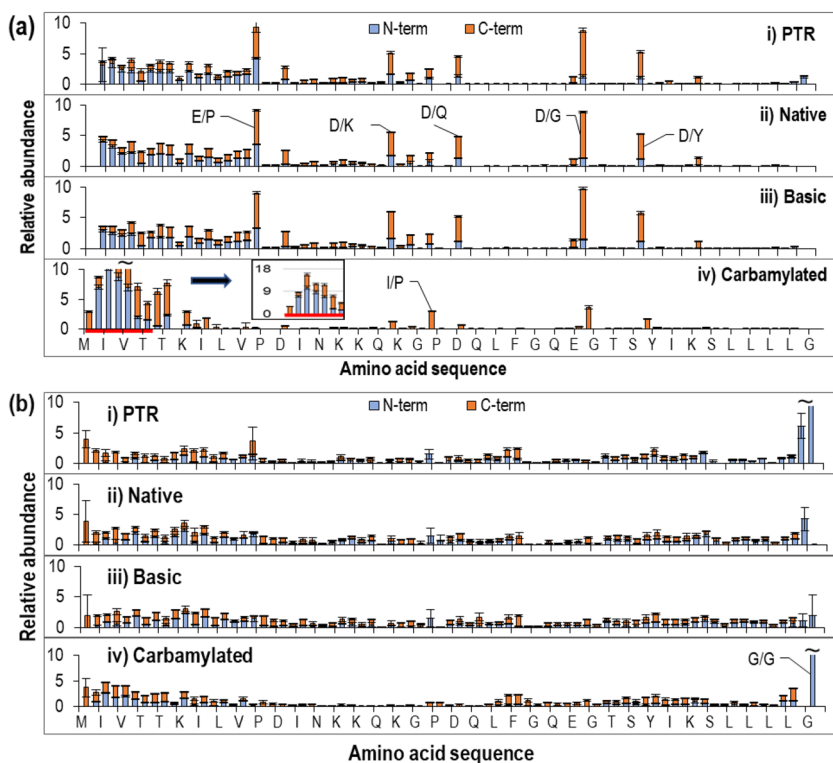
### Distribution of Backbone Fragmentation

The sequence coverage illustrates the total extent of fragmentation, but it does not indicate how the cleavages are distributed along the backbone of the protein. More specific nuances in fragmentation patterns are obtained by examination of the sequence maps that show the distributions of fragment ions as a function of the backbone cleavage sites, as illustrated in Fig. 2 for ubiquitin (5<sup>+</sup>) and Figs. S18 to S23 for the other six proteins.

**Ubiquitin** HCD results in enhanced fragmentation near the N-terminus of ubiquitin (5<sup>+</sup>) generated by all four methods (Fig. 2a) and yields several dominant cleavages of C-terminal to aspartic (D) and glutamic (E) acid residues (Glu18, Asp32, Asp39, Asp52, and Asp58) for the 5<sup>+</sup> charge states generated by PTR, the native (high ionic strength) solution, and the basic (high pH) solution, but not for the carbamylated protein. Ubiquitin has four Arg and seven Lys, all of which are basic sites that sequester protons. For the 5<sup>+</sup> charge states generated by PTR, the native solution, and the basic solution, ubiquitin is not expected to have any mobile protons. It is known that backbone

cleavage C-terminal to Asp (D) occurs by the attack of the carboxyl oxygen of Asp on the adjacent carbonyl carbon in the absence of mobile protons during collisional activation [31, 37]. The suppression of these Asp (D) and Glu (E) cleavages for carbamylated ubiquitin reflects the greater proton mobility relative to the other 5<sup>+</sup> charge states (PTR, native, basic) [37]. For carbamylated ubiquitin, four protons are expected to be sequestered at the very basic arginine residues, and the fifth proton should be mobile and not localized at a lysine residue nor the N-terminus (all of which are carbamylated). Very similar distributions of backbone cleavages are observed for the 5<sup>+</sup> charge state of ubiquitin produced by the other three methods (PTR, native solution, and basic solution) suggesting that these populations of ions adopt similar protonation sites with similar overall proton mobilities. Unlike HCD, extensive non-specific backbone fragmentation is observed upon UVPD of all four versions of the 5<sup>+</sup> charge state of ubiquitin (Fig. 2b) with minimal preferential cleavages.

**Cytochrome c** Enhanced fragmentation is observed near both of the N- and C-termini for each of the 7<sup>+</sup> charge states of cytochrome c upon HCD (Fig. S18a), with particular enhancement observed for the carbamylated version at the expense of extensive fragmentation in the mid-section of the protein (as also reflected in the low sequence coverage seen in Table 1). Unlike ubiquitin, enhanced fragmentation C-terminal to acidic residues is not the dominant pathway for cytochrome c. Fragmentation next to Pro76 is enhanced for each of the four 7<sup>+</sup> charge states of cytochrome c, especially for the carbamylated protein. Although the fragmentation pattern of cytochrome c obtained by HCD does not exhibit a significant number of preferential cleavages, there are a few enhanced cleavages adjacent to Asp50, Lys53, Lys55, Ile57, Asn70, and Ile75. Generally, the fragmentation is patchier across the backbone for the 7<sup>+</sup> charge states created by PTR and for the carbamylated protein. Fragmentation of the carbamylated protein is particularly suppressed in the mid-section of the protein as evidenced by a long stretch spanning Lys13 to Tyr74 with minimal backbone cleavages, thus explaining the low sequence coverage of only 33%.



**Figure 2.** Backbone cleavage histograms of ubiquitin ( $5^+$ ) from HCD (a) and UVPD (b): (i) PTR ( $11^+ \rightarrow 5^+$ , 30 ms), (ii) native solution (100 mM ammonium acetate), (iii) basic solution (1%  $\text{NH}_4\text{OH}$ ), and (iv) carbamylated. Some of the more prominent fragmentation sites are labeled. The sequence of ubiquitin is shown along the x-axis (every other residue is shown). N-terminal ions are shown in blue bars, and C-terminal ions are shown in orange bars

Like ubiquitin, UVPD of cytochrome *c* results in non-specific fragmentation along the protein backbone along with a few more prominent cleavages (Gly1/Asp2, Ile9/Phe10, and Ile75/Pro76) for the  $7^+$  charge states produced by all four methods (Fig. S18b). However, there are no backbone cleavages observed for the carbamylated protein for one stretch between Lys55 and Leu68, thus accounting for the decrease in sequence coverage (72%) when compared to the  $\sim 90\%$  coverages obtained for the unmodified protein. The region between Cys14 to His18 remained unfragmented by either HCD or UVPD owing to the stabilization of this region by the heme-binding domain [18].

**Transthyretin** HCD of the  $7^+$  charge states of transthyretin consistently results in relatively lower sequence coverage compared to UVPD. Backbone cleavages of C-terminal to Asp38 and Asp99 are enhanced for the  $7^+$  charge states produced by PTR or the native or high pH solutions (Fig. S19a). There is virtually no fragmentation of the region between Asp18 and Asp38, a stretch that contains four basic residues, including two Arg, one His, and one Lys. In addition, fragmentation, next to Pro125, is particularly favored for transthyretin after carbamylation or when sprayed from a high pH solution. The number of charges (seven) is lower than the number of basic residues (4R/8K/4H) of transthyretin which causes the charges to be sequestered at the basic sites and, hence, enhances fragmentation via charge remote pathways. No fragmentation is

observed for a relatively long stretch of carbamylated transthyretin from Pro11 to Lys76 upon HCD, thus accounting for the low 31% sequence coverage obtained for the carbamylated protein. For UVPD of transthyretin, enhanced fragmentation of the N- and C-terminal regions and greater sequence coverage are obtained irrespective of the manner used to generate the  $7^+$  charge states (Fig. S19b).

**$\alpha$ -Synuclein** Activation of the  $7^+$  charge states of  $\alpha$ -synuclein generates predominantly fragment ions containing the N-terminus (*a*-, *b*-, *c*-type) for both HCD and UVPD (Fig. S20a,b), a result consistent with the greater frequency of basic residues in the first half of the sequence and the concomitant higher frequency of acidic residues in the second half of the sequence. Unlike other proteins examined, HCD of  $\alpha$ -synuclein results in mostly non-specific fragmentation along the backbone with only a few notable enhanced backbone cleavages—Val66/Gly67, Asp98/Gln99, Asp119/Pro120, Met127/Pro128 (Fig. S20a). Moreover, high sequence coverages ranging from 75-88% are obtained regardless of the method used to generate the low  $7^+$  charge states. UVPD of  $\alpha$ -synuclein also results in extensive cleavages across nearly the entire backbone with only a small stretch between Asp98 and Asp115 that remains unfragmented regardless of the method used to generate the  $7^+$  charge states (Fig. S20b). As a result, slightly lower coverage (82% on average) is obtained for UVPD compared to HCD (88%). This stretch between residues

98–115 (DQLGKNEEGAPQEGILED) contains 7 residues (two D, one P, four E) for which charge-remote fragmentation pathways are known to be enhanced for low charge states upon collisional activation, and this high frequency of preferential cleavage sites rescues the performance of collisional activation but serves no benefit for UVPD. Owing to the fact that  $\alpha$ -synuclein lacks arginines and contains only one histidine, carbamylation of all lysines leaves the protein nearly devoid of protonation sites and results in the unsuccessful observation of ions by ESI (presumably due to its high hydrophobicity and precipitation).

**Staph Nuclease** HCD of the  $8^+$  charge states of staph nuclease generates mainly *b* ions (blue bars) covering the first half of the sequence and only low abundance *y* ions (orange bars) in the second half of the protein (Fig. S21). As expected, preferential fragmentation C-terminal to Asp19, Asp40, and Asp77 and N-terminal to Pro10 is observed for the  $8^+$  charge states. This finding is consistent with the mobile proton model as the presence of twenty Lys and five Arg can efficiently sequester eight protons, thus enhancing the predominance of charge-remote pathways and suppressing non-specific backbone cleavages. Backbone cleavages are particularly suppressed across a long stretch of residues (Asp89 to Glu123), and only charge-remote pathways occur which limits the overall sequence coverage obtained by HCD. Interestingly, the number of backbone cleavages in the C-terminal end of the  $8^+$  charge state generated by PTR is significantly reduced, resulting in only 34% sequence coverage when compared to 59% for the  $8^+$  charge state produced by spraying the protein directly from a basic solution or 53% for the  $8^+$  charge state of the native-like protein. This finding suggests that PTR removes protons from sites that differ from those typically occupied when the protein emerges from the basic or high ionic strength solutions.

UVPD of the  $8^+$  charge state of staph nuclease also predominantly generates N-terminal ions (*b* ions; blue bars), and fragmentation is less efficient in the C-terminal end of the sequence (Fig. S21b). The fragmentation of the first half of the protein sequence is much more uniform for UVPD compared to HCD, reflecting a higher degree of non-specific fragmentation and fewer enhanced charge-remote pathways. A decrease in sequence coverage is observed upon UVPD of the  $8^+$  charge state generated by PTR, similar to the one noted for HCD. This parallel outcome suggests that the low abundance of sequence ions covering the C-terminal region might arise from a general lack of charges in this region (despite the ample number of basic residues), impeding detection of the resulting products. Like  $\alpha$ -synuclein, staph nuclease could not be successfully detected after carbamylation (likely owing to precipitation).

**Calmodulin** HCD of the  $7^+$  charge states of calmodulin results in the limited formation of *b* ions (blue bars) in the N-terminal region and *y* ions (orange bars) in the C-terminal region (Fig. S22a). Sequence coverages returned by HCD are

generally low for calmodulin and particularly low for the  $7^+$  charge states created by PTR (19%) and for the carbamylated protein (23%). The types of backbone cleavages that contribute to these low net coverages are very different. For the  $7^+$  charge state generated by PTR, nine charge-remote pathways, all preferential backbone cleavages of C-terminal to acidic residues, dominate the HCD spectrum. In contrast, for the  $7^+$  charge state of the carbamylated protein, these same charge-remote pathways are not active and instead only non-specific cleavages close to the N- and C-terminus occur, resulting in very short sequence ions. The  $7^+$  charge states obtained for the protein sprayed from solutions of high ionic strength or high pH yield similar fragmentation patterns upon HCD, each exhibiting a few preferential cleavages at acidic residues, more extensive fragmentation near the N- and C-termini, and a new but limited series of products from non-specific backbone cleavages from Ala46 to Gly61.

Unlike HCD, UVPD results in no preferential charge-remote backbone cleavages. Instead, UVPD promotes a broad array of non-specific cleavages, yielding primarily N-terminal ions (blue bars) in the first half of the sequence and C-terminal ions (orange bars) in the second half (Fig. S22b). Low abundance fragment ions cover the mid-section of the sequence, with the exception of the stretch from residues 104 to 123 that contains seven acidic residues and for which few backbone cleavages occur (as also observed for HCD). Total sequence coverage obtained by UVPD ranges from 46% for the carbamylated protein to 66% for the native-like protein, again significantly higher than the coverage obtained by HCD for the  $7^+$  charge states.

**Myoglobin** Like calmodulin, fragmentation of the  $8^+$  charge state of myoglobin by HCD results in clusters of *b* ions (blue bars) in the N-terminal section of the sequence and *y* ions (orange bars) in the C-terminal region (Fig. S23a). Several preferential cleavages consistent with charge-remote pathways are observed for the  $8^+$  charge states created by PTR or for the native-like or basic solutions. For example, backbone cleavages C-terminal to acidic residues (Asp4/Gly5, Asp20/Ile21, Asp44/Lys45, Asp60/Leu61, Asp122/Phe123, Asp126/Ala127, Asp141/Ile142), C-terminal to lysine and N-terminal to proline (Lys63/His64, Lys79/Gly80, His119/Pro120) are predominant. In contrast, HCD of the carbamylated protein only results in fragmentation close to the N- and C-terminus but no preferential cleavages. Moreover, a large portion of the sequence between Trp14 to Asp122 remains uncharacterized with very few backbone cleavages and resulting in only 20% sequence coverage for the carbamylated protein compared to over 60% coverage obtained by HCD of the  $8^+$  charge states generated by analysis of the native or basic solutions.

Fragmentation of the  $8^+$  charge states by UVPD yields backbone cleavage histograms (Fig. S23b) that are strikingly different from the ones obtained by HCD. Total sequence coverages are somewhat higher for UVPD than HCD, but more notable is the near absence of the types of preferential

cleavages that are highly enhanced for HCD. UVPD of all the  $8^+$  charge states results in abundant fragment ions in the region spanning Phe33 to Phe46 along with extensive lower abundance fragment ions through the mid-section of the protein. UVPD of carbamylated myoglobin also results in similar fragmentation patterns and similar coverage except for a few observed differences in cleavages of N-terminal to proline residues (K87/P88, Ile99/Pro100, and His119/Pro120). These results reiterate the ability of UVPD to fragment the protein backbone more uniformly with fewer preferential cleavages than HCD, regardless of the method used to generate the low charge states.

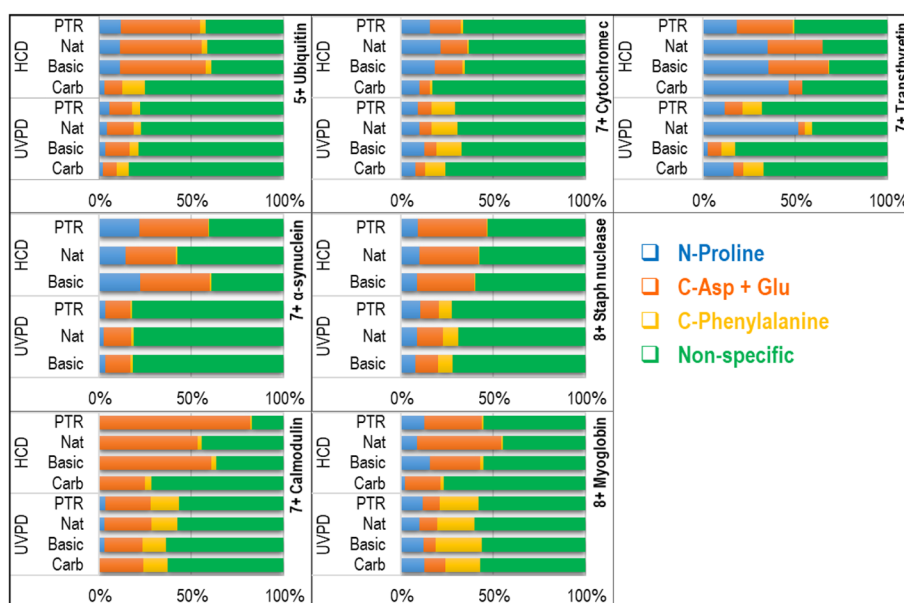
### Preferential Cleavages

The fragmentation of proteins by collisional activation methods relies on the availability of mobile protons which directs the prevalence of preferential fragmentation pathways [6]. The two most common preferential cleavages occur next to proline for proteins in high charge states or adjacent to glutamic acid and aspartic acid for proteins in low charge states (limited proton mobility) [5]. Obtaining high sequence coverage of a protein may be impeded if the products from these preferential cleavages dominate the spectra. To assess the degree of preferential cleavages for the low charge states of proteins showcased in this study, the total abundances of product ions originating from the following three types of preferential cleavage sites were evaluated: N-terminal to proline (N-Pro), C-terminal to phenylalanine (C-Phe), and C-terminal to glutamic and aspartic acid (C-Glu + Asp). Cleavages of C-terminal to phenylalanine were previously reported to be enhanced upon UVPD of intact proteins [18]. The distribution of these three types of preferential cleavages compared to all other non-specific pathways is

represented in a bar graph for each of the seven proteins studied (Fig. 3).

For all of the proteins, two types of preferential cleavages (N-Pro, C-Glu + Asp) are particularly prominent for HCD for the low charge states generated by PTR or from ESI of the high ionic strength (native-like) or basic solutions. The prevalence of these preferential cleavages generally decreases for the carbamylated proteins upon HCD, a result consistent with the change in proton mobility upon conversion of highly basic lysines to non-basic sites. For example, backbone cleavage of C-terminal to acidic residues is highly enhanced for the  $5^+$  charge states of ubiquitin generated by PTR (43%) or from the native (44%) or basic (46%) solutions. However, this percentage plummets to 10% for carbamylated ubiquitin. Preferential cleavages next to acidic residues account for over 50% of the product ion abundance for calmodulin, a protein containing an unusually high number of acidic residues (17 Asp and 21 Glu). The range of C-Glu + Asp cleavages for calmodulin varies from 82% for the ions created by PTR to 53 or 60% for ions produced from native-like or basic solutions, indicating that the structure or charge-site locations vary for these different populations of  $7^+$  charge states. After carbamylation, the portion of products originating from the cleavages next to acidic residues decreases to less than 25%, and instead non-specific charge-directed cleavages dominate. The portion of products from N-Pro cleavages appears to be less sensitive to the method used to generate the protein ions and instead more closely track the total number of proline residues in each protein.

Unlike HCD, UVPD favors non-specific cleavages (60 to 80%) for all proteins studied. The portions of products from backbone fragmentation of C-terminal to acidic residues or N-terminal to proline are significantly lower upon UVPD than HCD and are relatively insensitive to carbamylation. However,



**Figure 3.** Distribution of fragment ions generated by HCD and UVPD categorized as preferential cleavages (N-terminal to proline, C-terminal to glutamic and aspartic acid residues, C-terminal to phenylalanine) and all non-specific N- and C-terminal cleavages



fragmentation of C-terminal to Phe is favored upon UVPD for all proteins analyzed except  $\alpha$ -synuclein (which only has two Phe). Cleavages of C-terminal to Phe are presumably enhanced owing to the site-specific absorption of UV photons by the aromatic ring of Phe. The abundances of the Phe-directed cleavage products are consistent with the number of Phe residues in the protein sequence, with the Phe cleavage products accounting for up to 26% of the product ion current for myoglobin which contains seven Phe residues. These results support the lack of dependence of UVPD fragmentation on many of the characteristics that influence collisional activation.

### Distribution of Fragment Ion Types

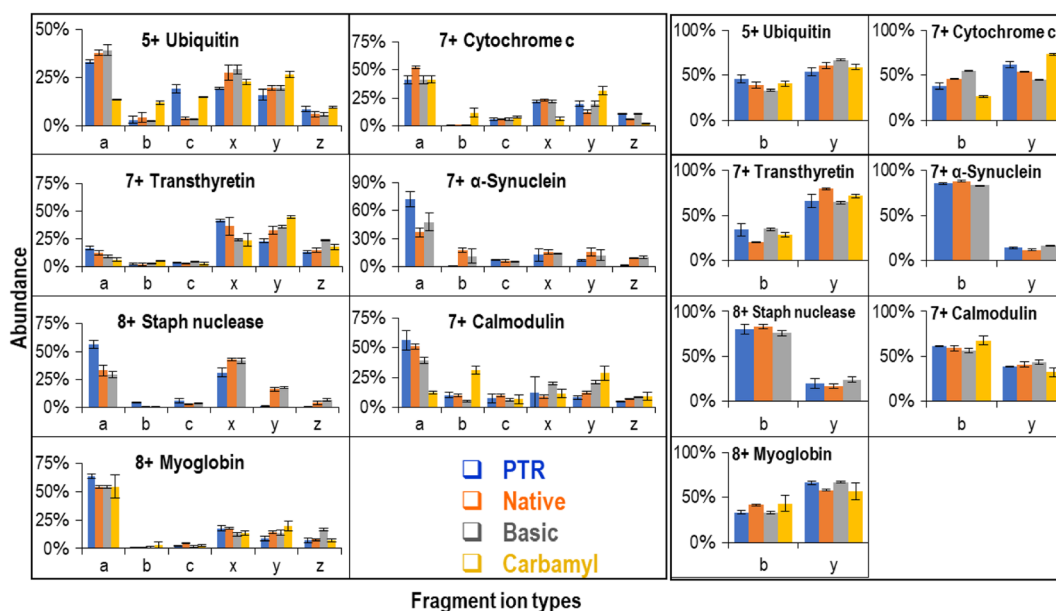
UVPD generates  $a/x$ -,  $b/y$ -, and  $c/z$ -type fragment ions, typically with  $a$ -ions most prominent as well as ample abundances of  $x/y$  ions and lower abundances of  $b/c/z$  ions. The distributions of the N-terminal ions ( $a$ ,  $b$ ,  $c$ ) and C-terminal ions ( $x$ ,  $y$ ,  $z$ ) generated by UVPD and  $b/y$  ions generated by HCD for each protein are summarized in Fig. 4. Upon UVPD, production of N-terminal fragment ions is favored or is fairly balanced with C-terminal fragment ions for all proteins except for transthyretin, which significantly favors C-terminal ions. Upon HCD, ubiquitin, cytochrome c, and transthyretin favor the formation of C-terminal ( $y$ ) ions, whereas  $\alpha$ -synuclein, staph nuclease, and calmodulin favor the formation of N-terminal ( $b$ ) ions. The formation of C-terminal ( $x$ ,  $y$ ,  $z$ ) ions is enhanced for transthyretin for both UVPD and HCD. Transthyretin has a slightly higher ratio of basic to acidic residues in the second half of the sequence which may partially account for its imbalanced distribution of C-terminal versus N-terminal ions.

An interesting observation is that the decreased abundances of  $a$ -ions and enhancement of  $b$ -ions are consistently observed

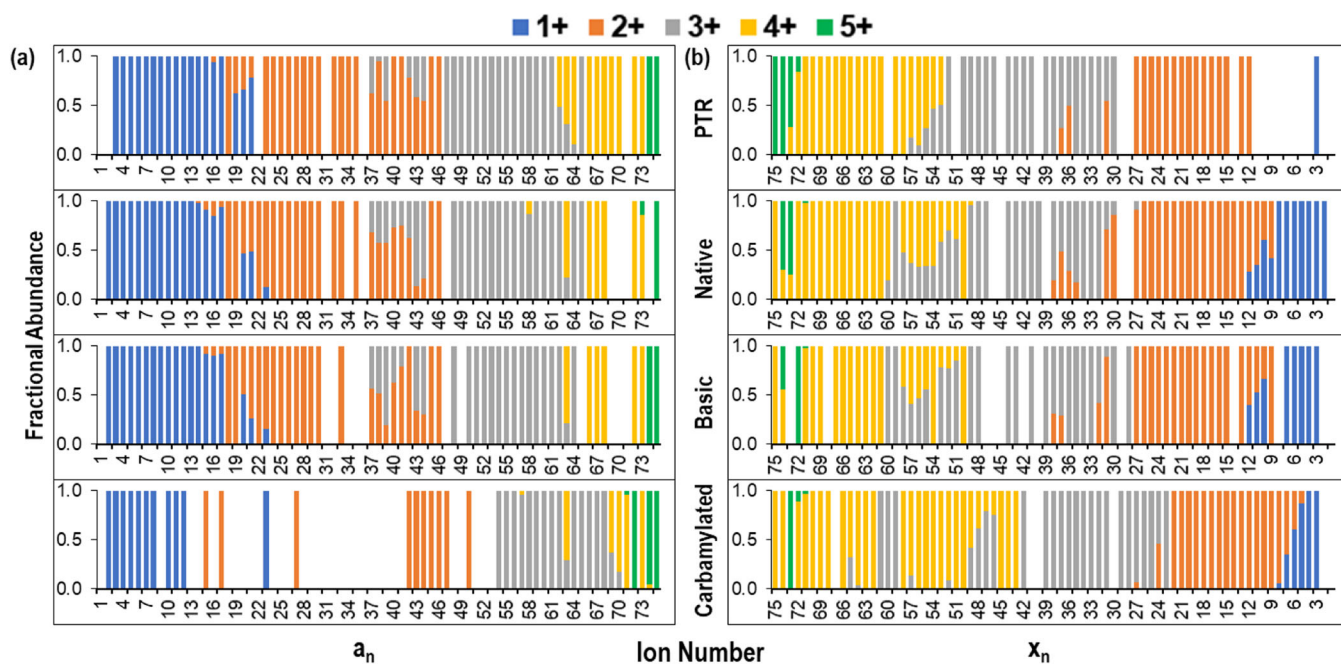
upon UVPD of the carbamylated proteins compared to the unmodified proteins (except for myoglobin). This outcome suggests that a pathway in which  $b$ -ions decay to  $a$ -ions by the loss of carbon monoxide is suppressed for the carbamylated proteins or that the pathway which creates  $a$ -ions directly from excited states is less favorable for the carbamylated proteins. These trends merit further evaluation based on larger sets of proteins which may yield additional clues about the mechanism of UVPD and the impact of charge sites and protein structure on UVPD.

### Charge Site Location

As described previously, the fragmentation patterns of each protein vary depending on the method used to generate the same low charge states. The variations are evidenced in the changes in sequence coverages, the differences in the types of fragment ions produced, and even in the distributions of backbone cleavages. Although in some cases the variations are relatively subtle, they suggest several possibilities that could rationalize the changes in fragmentation patterns based on potential differences in the structures and/or the locations of charge sites adopted by the proteins generated by each of the four methods. To illustrate the impact of the charge site locations, the charge locations are assigned by mapping the charge states of  $a$ - and  $x$ -type fragment ions for each protein using a strategy previously described [11]. In essence, the charge states of all  $a$ - and  $x$ -ions produced upon UVPD are plotted in a histogram as a function of the number of residues (length) of each fragment ion for each protein (see Fig. 5 for ubiquitin and S24–S29 for the other six proteins). When a larger fragment ion appears in a higher charge state compared to a smaller fragment ion, it indicates the addition of another proton to the fragment



**Figure 4.** Fragment ion distributions for ubiquitin (5<sup>+</sup>), cytochrome c (7<sup>+</sup>), transthyretin (7<sup>+</sup>),  $\alpha$ -synuclein (7<sup>+</sup>), staph nuclease (8<sup>+</sup>), calmodulin (7<sup>+</sup>), and myoglobin (8<sup>+</sup>) generated by four different methods upon UVPD (left) showing six different ion types ( $a/x$ ,  $b/y$ ,  $c/z$ ) and that by HCD (right) showing  $b$ - and  $y$ -ions



**Figure 5.** Fractional abundances of charge states of (a)  $a$ -ions and (b)  $x$ -ions of ubiquitin ( $5^+$ ) generated by four methods upon UVPD activation

ion, and the additional charge is presumed to be localized within the stretch where the charge state shifts from one fragment ion to another up until the point where the charge state shifts again. Results are described in detail for two of the proteins—ubiquitin ( $5^+$ ) or cytochrome  $c$  ( $7^+$ ).

For the  $a$ -ion series produced upon UVPD of the  $5^+$  charge state of ubiquitin generated by proton transfer reactions (Fig. 5a), the smallest fragment ion is  $a_2$  indicating that the first proton is located in the first two residues. Fragments  $a_2$  to  $a_{17}$  are all singly charged, indicating that no additional protons are localized in this region. The fragment ions  $a_{18}$  to  $a_{36}$  are doubly charged, and the well-demarcated shift corresponds to the second proton located at the amide of Glu18 or potentially migrating further along this stretch of the protein between residues 18 and 36 [38]. Additional protons are sequentially localized based on the shifts in the charge states observed in the fragment ion histograms. In some cases, fragment ions are produced in two charge states, thus indicating alternative schemes for proton locations. This latter phenomenon is observed for the region of the histogram covering  $a_{19}$  to  $a_{21}$  (Fig. 5a) which exhibits  $a$ -ions in both  $1^+$  and  $2^+$  charge states as well as the region covering  $a_{38}$ ,  $a_{39}$ , and  $a_{40}$  which indicates both  $2^+$  and  $3^+$  charge states. The histograms for the complementary  $x$ -ions (Fig. 5b) generally mirror the trends observed for the  $a$ -ions. The  $x$ -ion series can be used to infer additional charge site information when the  $a$ -ion series is incomplete or vice versa.

The  $a/x$  ion charge state plots for the  $5^+$  charge states of ubiquitin produced by each of the four methods (i.e., PTR, native solution, basic solution, and carbamylated) show similar transition points as well as some notable differences in the production of product ions in multiple charge states. The  $a_n$ -ion charge plot for carbamylated ubiquitin ( $5^+$ ) is significantly

patchier than the ones for the  $5^+$  charge states generated by PTR or ESI of the native solution or the basic solutions, and this patchiness is mirrored in the portion of  $a$ -ion coverage observed for ubiquitin in Fig. 4. Upon carbamylation of ubiquitin, C-terminal  $x$ ,  $y$ , and  $z$  ions are significantly enhanced relative to  $a$ -type ions, making it difficult to pinpoint charge site locations based on the  $a$ -ions. However, the production of  $x$ -ions is robust upon UVPD, as illustrated by the  $x_n$  ion distributions in Fig. 5, and the shifts in the charge states of the fragment ions are more readily visualized. Although there are similarities in the charge site patterns, the plots are not identical for ubiquitin ( $5^+$ ) produced by the four different methods. These results suggest that some of the variations in the fragmentation patterns of the proteins produced in the same charge state by different methods arise from differences in the localization of charges.

For the  $a$ -ion series produced upon UVPD of the  $7^+$  charge state of cytochrome  $c$  generated by PTR or by native ESI, the  $a_7$  to  $a_{13}$  ions were all singly charged, suggesting localization of the first proton on the first seven residues for PTR and native ESI (Fig. S24). For all four versions of cytochrome  $c$ , no fragmentation is observed in the heme-binding region (Cys14 to Cys17) and is mirrored by the absence of  $a_{14}$  to  $a_{17}$  ions. The ions  $a_{18}$  to  $a_{24}$  are all doubly charged, indicating that the second proton is localized in this stretch. The locations of the additional protons are more difficult to assign owing to a greater diversity of fragment ions in multiple charge states. There are several transitions observed for the triply charged  $a$ -ions, indicating that the third proton can be located in a long stretch of residues spanning His26 to Ile57 (a stretch that contains eight basic residues). Similarly, the fourth proton is predicted to be located between Arg38 and Ile75 (with six basic residues in this

region), the fifth proton in the stretch between Tyr67 and Lys86 (with four basic residues in this region), the sixth proton at Arg91, and the seventh proton at Lys100. The absence of sharp transitions among the more highly charged  $a$ -ions suggests that there are several populations of ions with protons localized at different residues. This hypothesis is also supported by the presence of several  $a$ - $x$ -ions in the mid-section of the protein that are found in two different charge states. For carbamylated cytochrome c, the second proton may be located from Gln12 to Gly24 as evidenced by doubly charged  $a_{12}$  through  $a_{24}$  ions, and the third proton is localized from His26 to Arg38 owing to the presence of the triply charged  $a_{24}$  to  $a_{38}$  ions. The absence of sharp transitions and the lack of a complete  $a$ -ion series make it more difficult to pinpoint the locations of the additional protons for carbamylated cytochrome c. In general, variations in the charge state distributions of  $a$ - and  $x$ -ions are more prominent for the  $7^+$  charge states of cytochrome c generated by the four different methods than noted for ubiquitin.

Similar charge site analysis was performed for the other five proteins and is summarized in Figs. S25–S29. The estimated location of charges of all seven proteins is displayed in Table S2. Owing to less complete sequence coverages for these proteins, particularly less extensive series of  $a/x$  ions, the assignment of charge sites is not as comprehensive. In general, the location of the first charge seems to be shifted more towards the N-terminus for the proteins sprayed from the basic solutions and for the carbamylated proteins. Also, the locations of the charge sites vary slightly for the carbamylated proteins relative to the charge states created by the other methods.

These general observations about the charge states of the  $a/x$  product ions further support the premise that changes in the fragmentation patterns of the proteins originate in part from variations in the positions of charges. However, there are also other considerations. This method of assigning charge sites derives locations based only on examination of  $a/x$  ions generated by UVPD, not other types of ions generated by UVPD nor by HCD. The  $a/x$  ions are surmised to originate from fast dissociation of ions in excited electronic states [13, 22], representing one population of ions accessed by UVPD. Additional proton migration may occur for ions that undergo vibrational energy re-distribution, whether after relaxation to ground states (for UVPD) or during collisional activation.

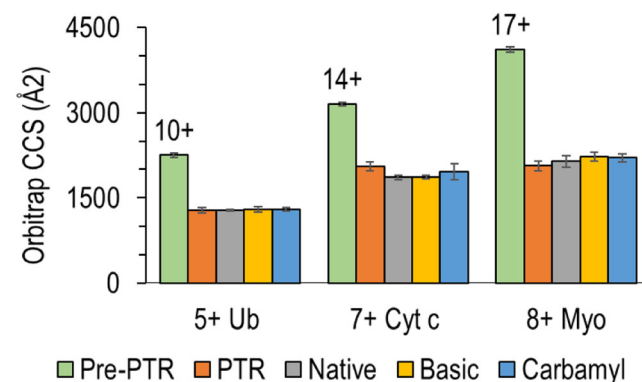
### Collision Cross-Sections

In addition to variations in the locations of charges, it is also possible that the proteins generated in low charge states by the four methods adopt different conformations (or populations of conformations). To explore this possibility, collision cross-sections (CCSs) were measured for three of the proteins in an Orbitrap mass analyzer. The collision cross-section of a protein is one indicator of its molecular shape/size in the gas phase. CCSs of proteins in the gas phase are traditionally measured by ion mobility (IM) methods [39]. Several studies have explored the effect of ESI solvent or PTR on the CCSs of protein ions [40–43], demonstrating that conformations varied as a function

of solution conditions and that charge reduction via PTR did not lead to significant structural collapse. Although IM mass spectrometry is a robust method for measuring CCS of protein ions, estimation of CCS values using an Orbitrap analyzer allows interrogation of the ions on the same platform as used for the HCD and UVPD experiments reported in this study. For this method, the decay rate of the transient signal in an Orbitrap mass analyzer [34] is used to estimate the CCS values. This method relies on the fact that the transient decay rates for proteins scale with their conformational size; in essence, unfolded protein ions with larger CCSs have higher probabilities of ion-neutral collisions than globular or folded proteins ions with smaller CCSs, thus leading to decay rates proportional to protein size [34].

The measured CCS values are shown in Fig. 6 for ubiquitin ( $5^+$ ), cytochrome c ( $7^+$ ), and myoglobin ( $8^+$ ) (and specific values are summarized in Table S1). Some increase in CCS after carbamylation is expected owing to the greater mass of the protein (e.g., net mass increase of 4.0, 6.6, and 5.1% for fully carbamylated ubiquitin, cytochrome c, and myoglobin, respectively). The differences in the CCS values for the  $5^+$  charge state of ubiquitin created by the four methods proved to be insignificant based on a  $t$  test (differing by less than 2%, ranging from  $1283 \pm 45$  to  $1305 \pm 47$  Å<sup>2</sup>) and in reasonable agreement with the CCS values reported in the literature for ubiquitin ( $5^+$ ) (Table S1) [34, 44–46]. As expected, all of these CCS values are much smaller than those of denatured ubiquitin ( $10^+$ ,  $2253 \pm 36$  Å<sup>2</sup>). The CCS after PTR of ubiquitin ( $10^+ \rightarrow 5^+$ ) is almost identical to the CCS of native version of the protein ( $1284 \pm 12$  Å<sup>2</sup>), suggesting that PTR caused the denatured protein ions to refold or collapse significantly. Results based on ion mobility measurements by Laszlo et al. indicated that native-like ions and ions in very low charge states created by PTR exhibited similar sizes and, hence, may share similar structural motifs [47].

The CCS value for the  $7^+$  charge state of cytochrome c generated by PTR ( $14^+ \rightarrow 7^+$ ) ( $2056 \pm 83$  Å<sup>2</sup>) is larger than the CCS values obtained for the  $7^+$  charge states generated



**Figure 6.** Orbitrap CCS values of  $5^+$  ubiquitin,  $7^+$  cytochrome c, and  $8^+$  myoglobin generated by four different methods. CCS values of denatured  $10^+$  ubiquitin,  $14^+$  cytochrome c, and  $17^+$  myoglobin are also included for comparison. All the experiments were performed in triplicate

from the native ( $1857 \pm 39 \text{ \AA}^2$ ) and basic solutions ( $1862 \pm 35 \text{ \AA}^2$ ) and much smaller than the CCS of the original  $14^+$  charge state ( $3154 \pm 35 \text{ \AA}^2$ ). The observation of a larger CCS of the  $7^+$  charge state of cytochrome c after PTR than the CCS of the  $7^+$  charge state of the native-like protein generated from the solution of high ionic strength agrees with a similar finding that reported somewhat larger CCSs for cytochrome c after PTR compared to the CCS values of the native-like protein in the same charge states [48]. Jhingree et al. also reported similar observations after electron transfer to the  $8^+$  charge state of native cytochrome c (charge reduction without dissociation, ETnoD). They reported that the  $7^+$  charge state formed after electron transfer had a CCS of  $2120 \pm 73 \text{ \AA}^2$  ( $\text{He} \rightarrow \text{N}_2$ ), larger than the  $7^+$  charge state of native-like cytochrome c ( $2043 \pm 34 \text{ \AA}^2$ ) [34, 49]. The present results suggest that after PTR, the denatured/unfolded  $14^+$  ions of cytochrome c refold or collapse to form a structure that is about 10% larger than the native-like version, and complete re-folding to a native-like state is not achieved. After carbamylation of 19 lysines, the CCS of cytochrome c ( $7^+$ ) was 5% larger than the native version which is consistent with the net mass increase arising from carbamylation of 20 sites.

The CCS value of the  $8^+$  charge state of myoglobin generated from the basic solution ( $2224 \pm \text{ \AA}^2 77$ ) was similar to that of the native-like protein ( $2143 \pm \text{ \AA}^2 101$ ), but the CCS decreased by 4% for the  $8^+$  charge state generated by PTR ( $17^+ \rightarrow 8^+$ ) ( $2064 \pm \text{ \AA}^2 89$ ). After carbamylation (19 lysines and the N-terminus), the CCS of myoglobin was 3% larger than the native version of the protein ( $2206 \pm \text{ \AA}^2 73$ ), readily accounted for by the increase in mass owing to carbamylation [50]. Overall, the CCS measured for the three proteins are in reasonable agreement with those reported in the literature (Table S1).

The modest variations in CCS values observed for the different low charge states of the proteins in Fig. 6 reflect subtle differences in the sizes of the proteins but do not reveal specific changes in the conformations that could account for the variations in the fragmentation patterns.

## Conclusion

Analysis of a panel of seven proteins in low charge states revealed differences in the fragmentation patterns depending on the method used to generate low charge state ions. In general, such metrics as sequence coverage, degree of preferential cleavages, and the types ( $a$ ,  $b$ ,  $c$ ,  $x$ ,  $y$ ,  $z$ ) and distributions of fragment ions varied more dramatically for HCD than UVPD, underscoring differences in the mechanisms of the two activation methods. Carbamylation led to a significant loss of sequence coverage of each protein upon HCD, particularly through the mid-section of the protein. The impact of carbamylation was especially interesting because the fully carbamylated proteins should have higher proton mobility owing to the conversion of highly basic proton-sequestering lysine residues to non-basic residues. However, the apparent

lower fragmentation efficiency of the carbamylated proteins upon HCD may contribute to the loss of sequence coverage. For the most part, the CCS values did not change significantly based on the method used to generate the ions in low charge states; all of the low charge states adopted structures far more compact than the denatured proteins in high charge states. Variation in the locations of the charge sites provides one rationale to account for changes in the fragmentation patterns of the ions in low charge states created by the four different methods. Other factors that may influence the fragmentation include variations in both the arrangements of intramolecular interactions and structures of the ions, irrespective of whether they have defined conformations or collapsed structures; these types of differences might not be revealed by measurements of average CCS values. Moreover, the mechanisms of the prevalent fragmentation pathways may be modulated by intramolecular interactions of the ions and the possibility of proton migration. It has been previously proposed that collisional activation of multimeric noncovalent protein complexes may promote the reorganization of ion pairing interactions, such as salt bridges, without concomitant changes in global conformation or charge state [51], and these concepts could extend to individual proteins in low charge states like the ones in the present study. The robust performance of UVPD for fragmentation of proteins in low charge states offers a promising avenue for enhancing the characterization of larger proteins. The option of generating and analyzing low charge states affords special utility for situations where overlapping charge states of fragment ions create spectral congestion that impedes assignment of fragment ions of more highly charged proteins.

## Acknowledgements

We appreciate the donation of proteins from Dr. Kevin Dalby (calmodulin) and Dr. Lauren Webb (Staph nuclease). Funding from the NIH (R01GM121714) and the Robert A. Welch Foundation (F-1155) is acknowledged.

## References

1. Toby, T.K., Fomelli, L., Kelleher, N.L.: Progress in top-down proteomics and the analysis of proteoforms. *Annu. Rev. Anal. Chem.* **9**, 499–519 (2016)
2. Catherman, A.D., Skinner, O.S., Kelleher, N.L.: Top down proteomics: facts and perspectives. *Biochem. Biophys. Res. Commun.* **445**, 683–693 (2014)
3. Smith, L.M., Kelleher, N.L., Proteomics, T.C. for T.D., Linial, M., Goodlett, D., Langridge-Smith, P., Goo, Y.A., Safford, G., Bonilla\*, L., Kruppa, G., Zubarev, R., Rontree, J., Chamot-Rooke, J., Garavelli, J., Heck, A., Loo, J., Penque, D., Hornshaw, M., Hendrickson, C., Pasatolic, L., Borchers, C., Chan, D., Young\*, N., Agar, J., Masselon, C., Gross\*, M., McLafferty, F., Tsybin, Y., Ge, Y., Sanders\*, I., Langridge, J., Whitelegge\*, J., Marshall, A.: Proteoform: a single term describing protein complexity. *Nat. Methods.* **10**, 186–187 (2013)
4. Durbin, K.R., Fornelli, L., Fellers, R.T., Doubleday, P.F., Narita, M., Kelleher, N.L.: Quantitation and identification of thousands of human proteoforms below 30 kDa. *J. Proteome Res.* **15**, 976–982 (2016)
5. Reid, G.E., Wu, J., Chrisman, P.A., Wells, J.M., McLuckey, S.A.: Charge-state-dependent sequence analysis of protonated ubiquitin ions

- via ion trap tandem mass spectrometry. *Anal. Chem.* **73**, 3274–3281 (2001)
6. Chanthamontri, C., Liu, J., McLuckey, S.A.: Charge state dependent fragmentation of gaseous  $\alpha$ -synuclein cations via ion trap and beam-type collisional activation. *Int. J. Mass Spectrom.* **283**, 9–16 (2009)
  7. Dongré, A.R., Jones, J.L., Somogyi, Á., Wysocki, V.H.: Influence of peptide composition, gas-phase basicity, and chemical modification on fragmentation efficiency: evidence for the mobile proton model. *J. Am. Chem. Soc.* **118**, 8365–8374 (1996)
  8. Tsaprailis, G., Nair, H., Somogyi, Á., Wysocki, V.H., Zhong, W., Futrell, J.H., Summerfield, S.G., Gaskell, S.J.: Influence of secondary structure on the fragmentation of protonated peptides. *J. Am. Chem. Soc.* **121**, 5142–5154 (1999)
  9. Haverland, N.A., Skinner, O.S., Fellers, R.T., Tariq, A.A., Early, B.P., LeDuc, R.D., Fornelli, L., Compton, P.D., Kelleher, N.L.: Defining gas-phase fragmentation propensities of intact proteins during native top-down mass spectrometry. *J. Am. Soc. Mass Spectrom.* **28**, 1203–1215 (2017)
  10. 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.* **21**, 949–959 (2010)
  11. Morrison, L.J., Brodbelt, J.S.: Charge site assignment in native proteins by ultraviolet photodissociation (UVPD) mass spectrometry. *Analyst.* **141**, 166–176 (2015)
  12. Olsen, J.V., Macek, B., Lange, O., Makarov, A., Horning, S., Mann, M.: Higher-energy C-trap dissociation for peptide modification analysis. *Nat. Methods.* **4**, 709–712 (2007)
  13. Brodbelt, J.S.: Ion activation methods for peptides and proteins. *Anal. Chem.* **88**, 30–51 (2016)
  14. Min-Sik, K., Akhilesh, P.: Electron transfer dissociation mass spectrometry in proteomics. *Proteomics.* **12**, 530–542 (2012)
  15. Syka, J.E.P., Coon, J.J., Schroeder, M.J., Shabanowitz, J., Hunt, D.F.: Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. *Proc. Natl. Acad. Sci.* **101**, 9528–9533 (2004)
  16. Riley, N.M., Coon, J.J.: The role of electron transfer dissociation in modern proteomics. *Anal. Chem.* **90**, 40–64 (2018)
  17. Cannon, J.R., Martinez-Fontes, K., Robotham, S.A., Matouschek, A., Brodbelt, J.S.: Top-down 193-nm ultraviolet photodissociation mass spectrometry for simultaneous determination of polyubiquitin chain length and topology. *Anal. Chem.* **87**, 1812–1820 (2015)
  18. Greer, S.M., Holden, D.D., Fellers, R., Kelleher, N.L., Brodbelt, J.S.: Modulation of protein fragmentation through carbamylation of primary amines. *J. Am. Soc. Mass Spectrom.* **28**, 1587–1599 (2017)
  19. 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.* **135**, 12646–12651 (2013)
  20. 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.* **86**, 2185–2192 (2014)
  21. Greer, S.M., Brodbelt, J.S.: Top-down characterization of heavily modified histones using 193 nm ultraviolet photodissociation mass spectrometry. *J. Proteome Res.* **17**, 1138–1145 (2018)
  22. Julian, R.R.: The mechanism behind top-down UVPD experiments: making sense of apparent contradictions. *J. Am. Soc. Mass Spectrom.* **28**, 1823–1826 (2017)
  23. Susa, A.C., Xia, Z., Tang, H.Y.H., Tainer, J.A., Williams, E.R.: Charging of proteins in native mass spectrometry. *J. Am. Soc. Mass Spectrom.* **28**, 332–340 (2017)
  24. Konermann, L.: Addressing a common misconception: ammonium acetate as neutral pH “buffer” for native electrospray mass spectrometry. *J. Am. Soc. Mass Spectrom.* **28**, 1827–1835 (2017)
  25. 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.* **88**, 1008–1016 (2016)
  26. Iavarone, A.T., Jurchen, J.C., Williams, E.R.: Effects of solvent on the maximum charge state and charge state distribution of protein ions produced by electrospray ionization. *J. Am. Soc. Mass Spectrom.* **11**, 976–985 (2000)
  27. Krusemark, C.J., Frey, B.L., Belshaw, P.J., Smith, L.M.: Modifying the charge state distribution of proteins in electrospray ionization mass spectrometry by chemical derivatization. *J. Am. Soc. Mass Spectrom.* **20**, 1617–1625 (2009)
  28. Iavarone, A.T., Williams, E.R.: Collisionally activated dissociation of supercharged proteins formed by electrospray ionization. *Anal. Chem.* **75**, 4525–4533 (2003)
  29. Teo, C.A., Donald, W.A.: Solution additives for supercharging proteins beyond the theoretical maximum proton-transfer limit in electrospray ionization mass spectrometry. *Anal. Chem.* **86**, 4455–4462 (2014)
  30. Wysocki, V.H., George, T., Smith Lori, L., Brechi, L.A.: Mobile and localized protons: a framework for understanding peptide dissociation. *J. Mass Spectrom.* **35**, 1399–1406 (2001)
  31. Pitteri, S.J., Reid, G.E., McLuckey, S.A.: Affecting proton mobility in activated peptide and whole protein ions via lysine guanidination. *J. Proteome Res.* **3**, 46–54 (2004)
  32. Lee, K., Alphonse, S., Piserchio, A., Tavares, C.D.J., Giles, D.H., Wellmann, R.M., Dalby, K.N., Ghose, R.: Structural basis for the recognition of eukaryotic elongation factor 2 kinase by calmodulin. *Struct. Lond. Engl.* **1993**(24), 1441–1451 (2016)
  33. Greer, S.M., Cannon, J.R., Brodbelt, J.S.: Improvement of shotgun proteomics in the negative mode by carbamylation of peptides and ultraviolet photodissociation mass spectrometry. *Anal. Chem.* **86**, 12285–12290 (2014)
  34. Sanders, J.D., Grinfeld, D., Aizikov, K., Makarov, A., Holden, D.D., Brodbelt, J.S.: Determination of collision cross-sections of protein ions in an orbitrap mass analyzer. *Anal. Chem.* **90**, 5896–5902 (2018)
  35. 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.* **15**, 1235–1238 (2015)
  36. Rosenberg, J., Parker, W.R., Cammarata, M.B., Brodbelt, J.S.: UV-POSIT: Web-based tools for rapid and facile structural interpretation of ultraviolet photodissociation (UVPD) mass spectra. *J. Am. Soc. Mass Spectrom.* **29**, 1323–1326 (2018)
  37. Gu, C., Tsaprailis, G., Brechi, L., Wysocki, V.H.: Selective gas-phase cleavage at the peptide bond C-terminal to aspartic acid in fixed-charge derivatives of Asp-containing peptides. *Anal. Chem.* **72**, 5804–5813 (2000)
  38. Morrison, L.J., Rosenberg, J.A., Singleton, J.P., Brodbelt, J.S.: Statistical examination of the a and a+1 fragment ions from 193 nm UVPD reveals local hydrogen bonding interactions. *J. Am. Soc. Mass Spectrom.* **27**, 1443–1453 (2016)
  39. Harper, B., Neumann, E.K., Stow, S.M., May, J.C., McLean, J.A., Solouki, T.: Determination of ion mobility collision cross sections for unresolved isomeric mixtures using tandem mass spectrometry and chemometric deconvolution. *Anal. Chim. Acta.* **939**, 64–72 (2016)
  40. Shi, H., Atlasevich, N., Merenbloom, S.I., Clemmer, D.E.: Solution dependence of the collisional activation of ubiquitin  $[M+7H]^+$  ions. *J. Am. Soc. Mass Spectrom.* **25**, 2000–2008 (2014)
  41. Shi, H., Pierson, N.A., Valentine, S.J., Clemmer, D.E.: Conformation types of ubiquitin  $[M+8H]^+$  ions from water:methanol solutions: evidence for the N and A states in aqueous solution. *J. Phys. Chem. B.* **116**, 3344–3352 (2012)
  42. Valentine, S.J., Counterman, A.E., Clemmer, D.E.: Conformer-dependent proton-transfer reactions of ubiquitin ions. *J. Am. Soc. Mass Spectrom.* **8**, 954–961 (1997)
  43. Laszlo, K.J., Bush, M.F.: Interpreting the collision cross sections of native-like protein ions: insights from cation-to-anion proton-transfer reactions. *Anal. Chem.* **89**, 7607–7614 (2017)
  44. 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.* **82**, 9557–9565 (2010)
  45. Salbo, R., Bush, M.F., Naver, H., Campuzano, I., Robinson, C.V., Pettersson, I., Jørgensen, T.J.D., Haselmann, K.F.: Traveling-wave ion mobility mass spectrometry of protein complexes: accurate calibrated collision cross-sections of human insulin oligomers. *Rapid Commun. Mass Spectrom.* **26**, 1181–1193 (2012)
  46. Lermyte, F., Łacki, M.K., Valkenborg, D., Gambin, A., Sobott, F.: Conformational space and stability of ETD charge reduction products of ubiquitin. *J. Am. Soc. Mass Spectrom.* **28**, 69–76 (2017)
  47. Laszlo, K.J., Munger, E.B., Bush, M.F.: Folding of protein ions in the gas phase after cation-to-anion proton-transfer reactions. *J. Am. Chem. Soc.* **138**, 9581–9588 (2016)
  48. Laszlo, K.J., Buckner, J.H., Munger, E.B., Bush, M.F.: Native-like and denatured cytochrome c ions yield cation-to-anion proton transfer

- reaction products with similar collision cross-sections. *J. Am. Soc. Mass Spectrom.* **28**, 1382–1391 (2017)
49. Jhingree, J.R., Beveridge, R., Dickinson, E.R., Williams, J.P., Brown, J.M., Bellina, B., Barran, P.E.: Electron transfer with no dissociation ion mobility–mass spectrometry (ETnoD IM-MS). The effect of charge reduction on protein conformation. *Int. J. Mass Spectrom.* **413**, 43–51 (2017)
  50. Mosier, P.D., Counterman, A.E., Jurs, P.C., Clemmer, D.E.: Prediction of peptide ion collision cross sections from topological molecular structure and amino acid parameters. *Anal. Chem.* **74**, 1360–1370 (2002)
  51. Loo, R.R.O., Loo, J.A.: Salt bridge rearrangement (SaBRe) explains the dissociation behavior of noncovalent complexes. *J. Am. Soc. Mass Spectrom.* **27**, 975–990 (2016)