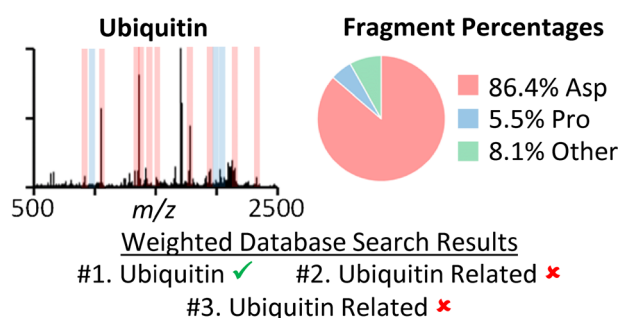CrossMark

Maximizing Selective Cleavages at Aspartic Acid and Proline Residues for the Identification of Intact Proteins

David J. Foreman,[1] Eric T. Dziekonski,[2] Scott A. McLuckey[1]

[1]Department of Chemistry, Purdue University, 560 Oval Drive, West Lafayette, IN 47907-2084, USA
[2]SCIEX, Concord, ON L4K 4V8, Canada

Abstract. A new approach for the identification of intact proteins has been developed that relies on the generation of relatively few abundant products from specific cleavage sites. This strategy is intended to complement standard approaches that seek to generate many fragments relatively non-selectively. Specifically, this strategy seeks to maximize selective cleavage at aspartic acid and proline residues via collisional activation of precursor ions formed via electrospray ionization (ESI) under denaturing conditions. A statistical analysis of the SWISS-PROT database was used to predict the number of arginine residues for a given intact protein mass and predict a m/z range where the protein carries a similar charge to the number of arginine residues thereby enhancing cleavage at aspartic acid residues by limiting proton mobility. Cleavage at aspartic acid residues is predicted to be most favorable in the m/z range of 1500–2500, a range higher than that normally generated by ESI at low pH. Gas-phase proton transfer ion/ion reactions are therefore used for precursor ion concentration from relatively high charge states followed by ion isolation and subsequent generation of precursor ions within the optimal m/z range via a second proton transfer reaction step. It is shown that the majority of product ion abundance is concentrated into cleavages C-terminal to aspartic acid residues and N-terminal to proline residues for ions generated by this process. Implementation of a scoring system that weights both ion fragment type and ion fragment area demonstrated identification of standard proteins, ranging in mass from 8.5 to 29.0 kDa.

Keywords: Ion/ion reactions, Intact protein identification, Top-down, Selective fragmentation

Introduction

The notion of using collision-induced dissociation (CID) to elucidate the primary structure of intact proteins was first presented in 1990 [1], and formed the basis for the development of "top-down" tandem mass spectrometry of proteins. Today, collisional activation remains one of the most common techniques for the identification of proteins in top-down

workflows [2–6]. Other dissociation approaches, such as electron capture dissociation [7], electron transfer dissociation [8], and UV photodissociation (UVPD) [9] have subsequently been applied to whole protein ions with a common objective of maximizing the number of sequence informative products. With this objective in mind, it is desirable that fragmentation be non-selective. However, as the number of different product ions increase, the overall product ion signal is dispersed among the many fragmentation channels [10]. There is merit, therefore, in having the option to maximize selective cleavages that tend to concentrate product ion signal into fewer channels. Scenarios that might benefit from this capability include, for example, the identification of low level proteins in cases when the dispersal of product ion signal among many channels limits

Correspondence to: Scott McLuckey; *e-mail:* mcluckey@purdue.edu

dynamic range and in single (or multiple) reaction monitoring workflows. We note here a distinction between "identification" of a protein or "proteoform" [11] and complete "characterization." Maximizing the number of cleavages is a priority for the latter objective in spite of the fact that this might limit detection limits and dynamic range.

The generation of almost exclusively b- and y-type fragment ions upon CID of multiply protonated proteins is commonly observed and is consistent with the mobile proton model [12–14]. Collisional activation of a peptide or protein leads to mobilization of the excess protons along the polypeptide backbone that can initiate a charge-directed fragmentation process to produce b- and y-fragment ions [15]. While a number of non-specific fragmentation pathways are observed when using collisional activation, particularly at intermediate charge states, several site-specific fragmentation pathways have been reported [16–20]. Specifically, under favorable conditions, cleavages at aspartic acid residues and proline residues have been shown to be the dominant pathways [16, 19, 21–24]. Despite both fragmentation pathways resulting in highly selective cleavages, the underlying mechanism regarding both cleavages differs. The higher basicity of the proline amide bond, relative to those of other amino acids, leads to preferential localization of a mobilized proton at the proline amide bond, initiating the charge-directed fragmentation N-terminal to the proline residue. The favorable cleavage at aspartic acid residues C-terminal to the acidic side chain does not require a mobile proton and can become prominent when the ionizing protons are sequestered by arginine residues [25].

Numerous studies have been performed to characterize the gas-phase fragmentation behavior of intact proteins [26–32], revealing that the information obtained in a CID experiment is highly dependent on the charge state subjected to activation as well as the number of basic residues present in the protein sequence [33, 34]. Fragmentation patterns can be rationalized on the basis of three major precursor ion types: high charge states, intermediate charge states, and low charge states. Activation of high charge states, where the large intramolecular Coulomb field tends to minimize proton mobilization, leads to fragmentation patterns that are difficult to predict a priori. Activation of intermediate charge states typically results in a high degree of sequence coverage and extensive non-specific fragmentation due to maximum proton mobility. Collisional activation of low charge state proteins, where protein mobility is hindered by solvation of a proton by multiple basic residues or where protons are sequestered onto mostly arginine residues, generates prominent small molecule loss and cleavage C-terminal to aspartic acid residues.

Recently, in recognition of the differences observed between the product ion spectra of protein ions generated under denaturing conditions versus those generated under native conditions [26–29, 35], the Kelleher group examined the gas-phase fragmentation propensities of native intact proteins [36]. In this study, a statistical analysis of 5311 matched fragments from 165 different experiments revealed striking differences in fragmentation propensities for native conditions when compared to denatured conditions. Most notably, under native conditions, fragmentation C-terminal to aspartic acid residues and N-terminal to proline residues showed significant enhancement in fragmentation tendency when compared to their denatured counterparts. This observation suggests that the CID of protein ions generated under native conditions could constitute a strategy for maximizing selective fragmentation at aspartic acid and proline residues. However, protein ion signals tend to be less intense using native conditions relative to denaturing conditions (e.g., low pH and, perhaps, some fraction of organic solvent) and the extent of sodium and potassium ion adduction tends to be greater due to differences in ionization mechanism [37].

Here, we present a methodology for the identification of intact proteins that maximizes the selective cleavages at aspartic acid and proline residues while retaining the advantages in signal levels and lower metal ion adduction afforded by generating ions under denaturing conditions [38–41]. A drawback associated with protein ionization under denaturing conditions is that the precursor ion signal is divided among many relatively high charge states. We therefore use a sequence of proton transfer ion/ion reactions within the mass spectrometer to, first, concentrate the precursor ion charge state distribution largely into one charge state prior to ion isolation and, second, to move the precursor ion charges into a range (i.e., m/z 1500–2500) that is most likely to lead to selective cleavages. This methodology is demonstrated here with four standard proteins often used by the National Resource for Translational and Developmental Proteomics in optimizing top-down workflows: ubiquitin (8.6 kDa), myoglobin (16.9 kDa), trypsinogen (24.0 kDa), and carbonic anhydrase (29.0 kDa) [42].

Materials and Methods

Sample Preparation

Ubiquitin from bovine erythrocytes, myoglobin from equine skeletal muscle, trypsinogen from bovine pancreas, carbonic anhydrase from bovine erythrocytes, 2,2,3,3,4,4,5,5,6,6,7,7,8,8,8-pentadecafluoro-1-octanol (PFO), ammonium bicarbonate, dithiothreitol (DTT), iodoacetamide, and formic acid were purchased from Sigma-Aldrich (St. Louis, MO, USA). Optima-grade acetonitrile, HPLC-grade methanol, and Optima LC/MS-grade water were purchased from Fisher Scientific (Fair Lawn, NJ, USA). Chloroform was purchased from Mallinckrodt (Phillipsburg, NJ, USA) and urea was purchased from Pierce Chemical (Rockford, IL, USA). PFO was made at a concentration of 200 μM in a methanol solution containing 2% ammonium hydroxide.

Stock solutions of ubiquitin, myoglobin, and carbonic anhydrase were prepared at ~2 mg/mL in water. Fifty microliters of the ubiquitin stock, 80 μL of the myoglobin stock, and 80 μL of the carbonic anhydrase stock were combined and precipitated using 1:1:4:3 protein/chloroform/methanol/water [43]. The dried precipitated protein mixture was reconstituted in 100 μL of mobile phase A. The proteins were then separated

via reverse-phase HPLC using an Agilent 1200 series with manual injector (Palo Alto, CA, USA) using an Agilent Zorbax 300SB-C3 4.6 × 150 mm, 5 μm (Palo Alto, CA, USA) analytical HPLC column. The HPLC method consisted of a linear gradient from 5% mobile phase B to 50% mobile phase B over 60 min (mobile phase A: 95% water, 5% acetonitrile, and 0.2% formic acid, mobile phase B: 5% water, 95% acetonitrile, 0.2% formic acid) at flow rate of 1 mL/min and a detection wavelength of 215 nM.

A total of four 20 μL injections were made. For each injection, each protein was collected into a separate Eppendorf tube for a total of four fractions for each protein. The first three fractions were dried under vacuum and reconstituted in the fourth sample. The concentrations of the working samples are approximately 9, 8, and 4 μM for ubiquitin, myoglobin, and carbonic anhydrase, respectively.

Reduction and Alkylation of Trypsinogen

Approximately 2 mg of trypsinogen was dissolved in 1 mL of reduction buffer (100 mM ammonium bicarbonate, 7 M urea). DTT was added from a 500 mM stock to a final concentration of 5 mM and was incubated for 45 min at 55 °C. The solution was cooled to room temperature and centrifuged briefly. Iodoacetamide was added to a final concentration of 14 mM, from a 500 mM freshly prepared stock solution. The solution was incubated at room temperature in the dark for 30 min. Unreacted iodoacetamide was quenched with addition of another 5 mM of DTT and incubated at room temperature for 15 min in the dark. One hundred microliters of the reduced and alkylated trypsinogen was subjected to the precipitation and HPLC procedures as described above. The final trypsinogen concentration is estimated to be approximately 6 μM.

Mass Spectrometry

All data were collected using a modified TripleTOF 5600 quadrupole/time-of-flight mass spectrometer (SCIEX, Concord, ON, Canada), modified to perform ion/ion reactions [44]. Alternatively pulsed nano-electrospray (nESI) ionization allows for sequential injection of multiply protonated proteins and singly deprotonated pentadecafluoro-1-octanol dimer (PFO) [45]. Analyte cations were injected into the mass spectrometer and trapped in q0 followed immediately by injection of the reagent anions. Instrument parameters were tuned such that PFO dimer was injected and transmitted through q0 while still trapping protein cations, resulting in consecutive proton transfer ion/ion reactions [46, 47]. During the reaction, the period which anions are passing through q0, an excitation voltage was applied at a slightly higher frequency than the desired ion's fundamental frequency [48]. This process is referred to as ion parking. Application of the excitation voltage slows the ion/ion reaction rate of a specific mass-to-charge, allowing concentration of signal into a single, lower charge state. Ion parking was performed at the charge state immediately prior to m/z 1500. Reaction times varied from 100 to 300 ms.

Next, the charge reduced protein was isolated in Q1 and transferred to q2. PFO dimer was once again generated via negative nESI and transferred to q2. The ions were mutually stored for a reaction time optimized to generate a distribution of charge states between m/z 1500 and m/z 2500 (10–30 ms) [49]. Application of two high amplitude frequency sweeps were used to eject any low m/z ions (< m/z 1500) and any high m/z ions (> m/z 2500). All charge states between m/z 1500 and 2500 were isolated and subjected to dipolar direct current (DDC) collisional activation [50]. Mass analysis was performed using time of flight (TOF).

Database Search

Database searches were performed using a program written in MATLAB (MathWorks®, Natick, MA, USA). All entries of the SWISS-PROT database (556,197 entries) were processed to create separate entries for the cleavage of the N-terminal initiator methionine residue and cleavage into individual protein chains [51]. The processing of the SWISS-PROT entries resulted in a database of 624,084 entries across all species.

Uninterpreted MS/MS spectra were processed using a zero-charge THRASH deconvolution algorithm written in MATLAB, using a S/N threshold of three [52]. The output contained a list of monoisotopic zero-charge masses as well the total area under each fragment's isotopic distribution. Theoretical zero-charge b- and y-fragment masses for each protein, within a ± 100 Da window of the experimentally derived protein mass, were compared to the masses of the deconvoluted peak list with a fragmentation tolerance of ± 15 ppm. A McLuckey Score, with modified coefficients for native-like protein charge states, was calculated for each protein within the 200-Da mass window as shown below,

$$Score = (5 \times nD \times \sum D_A) + (3 \times nP \times \sum P_A)$$
$$+ (nX \times \sum X_A)$$

where n and Σ are, respectively, the number of matched fragments and summation of fragment area (where the smallest area is normalized to an area of 1) for either D (aspartic acid), P (proline), or X (non-specific) cleavages [53]. These coefficients are based qualitatively off our understanding of the well-known aspartic acid effect and the fragmentation propensities of native intact proteins as determined by Kelleher and coworkers [36]. This scoring system accounts for the preferential cleavage at aspartic acid and proline observed with native-like protein charge states.

Results and Discussion

Experimental Design and Rationale

The approach to intact protein analysis as described here exploits the highly selective cleavages C-terminal to aspartic acid

residues and N-terminal to proline residues upon collisional activation, resulting in fewer, more abundant fragment ions, when compared to conventional top-down protocols where complete sequence coverage is desirable. Using a 27-residue polypeptide represented schematically in Fig. 1 as an example, cleavage at every residue would result in a minimum of 26 fragments ions, inherently dispersing the precursor ion intensity across 26 different fragmentation channels. Realistically, the resulting tandem mass spectrum will likely contain more than 26 fragment ions as both complementary fragments from a single cleavage often appear in the spectrum, as do products from small molecule losses, amino acid side chain losses, and from multiple backbone cleavages (i.e., internal fragment ions). As an alternative to complete sequence coverage, one could manipulate the protein charge state distribution to produce abundant aspartic acid cleavages. As shown in Fig. 1, cleavage at aspartic acid residues, as represented by the red amino acids, generates six fragment ions. Consequently, the precursor ion intensity is divided into only six fragments of higher abundance. In principle, this approach can lead to an extension of the lower limit of the dynamic range for top-down protein analysis.

The influence of precursor ion charge state on the gas-phase fragmentation behavior of multiply charged proteins is illustrated by ubiquitin (Fig. 2). The ion trap CID product ion spectra of $[M + 8H]^{8+}$, $[M + 7H]^{7+}$, $[M + 6H]^{6+}$, and $[M + 5H]^{5+}$ precursor ions show drastic differences in both fragment ion identities and abundances. At intermediate charge states, as shown in Fig. 2a, b, extensive non-specific cleavage was observed. For lower charge states, fragmentation at aspartic acid residues becomes more prominent. Activation of the $[M + 6H]^{6+}$ ion, the spectrum of which is shown in Fig. 2c, shows a base peak corresponding to the doubly charged y_{18} fragment (aspartic acid

channel), and activation of the $[M + 5H]^{5+}$ ion (Fig. 2d) produces almost exclusively cleavages at aspartic acid residues. These results are consistent with fragmentation characterization studies of ubiquitin and are consistent with the mobile proton model for peptide and protein dissociation [26, 32]. For ubiquitin, a small protein containing four arginine residues, the number of ionizing protons for the $[M + 6H]^{6+}$ and $[M + 5H]^{5+}$ ions begins to approach the number of arginine residues. The arginine residues, which have the highest gas-phase basicity of any amino acid [54, 55], sequester the protons allowing for processes that do not require proton mobilization (e.g., cleavage at aspartic acid residues and small molecule loss) to dominate the spectrum. These processes are also observed for the $[M + 4H]^{4+}$ precursor ions, where the number of ionizing protons is equal to the number of arginine residues. The ion trap CID spectrum, as shown in supplemental Fig. S1, shows a dominant water loss as well as cleavages C-terminal to aspartic acid residues.

Based on the assumption that proton mobility in protein ions is inhibited when the number of excess protons is roughly equal to the number of arginine residues in the protein, we sought to determine the m/z range within which most protein ions with limited proton mobility are likely to fall. A program written in MATLAB was used to count the number of arginine residues and calculate the mass of each protein ≤ 250,000 Da in the SWISS-PROT database. Linear regression of mass versus arginine count was used to predict the number of arginine residues for each entry (Fig. 3a). Next, the true count, the actual number of arginine residues present in the protein sequence, was compared to the predicted count, the number of arginine residues predicted by the linear model. For any sequence ± 3 arginine residues to the true count, seven mass-to-charge ratios were generated
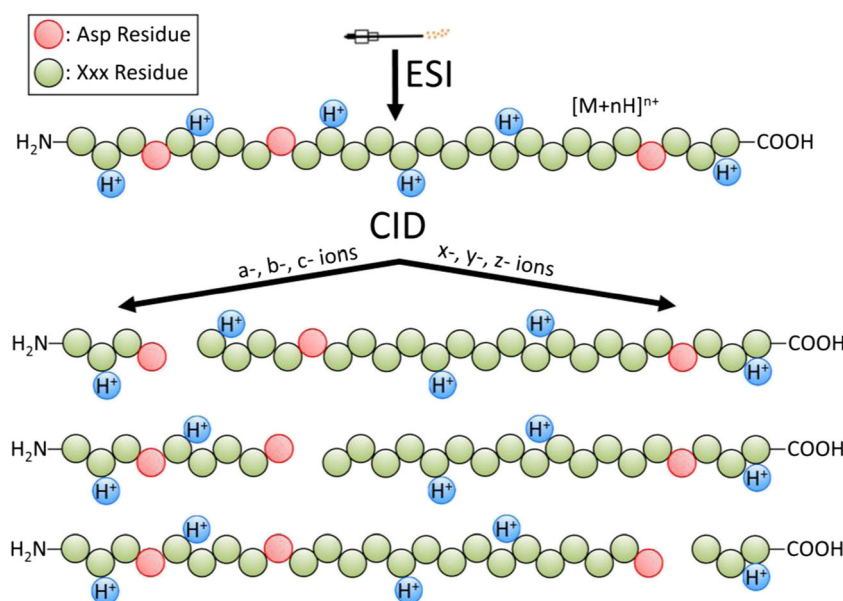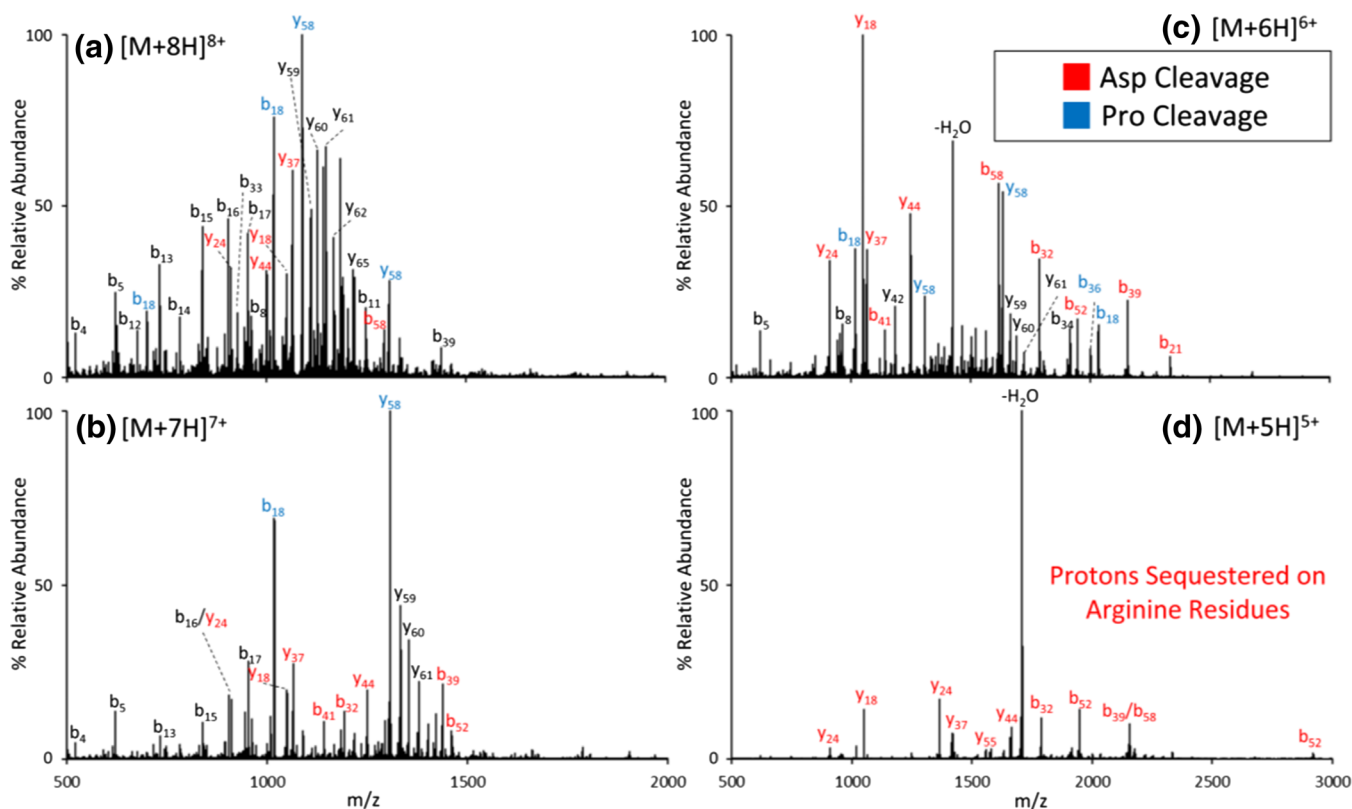

Figure 1. Illustration of selective cleavages upon collisional activation

Figure 2. Ion trap CID spectrum of ubiquitin for (**a**) $[M+8H]^{8+}$, (**b**) $[M+7H]^{7+}$, (**c**) $[M+6H]^{6+}$, and (**d**) $[M+5H]^{5+}$ precursor ions. Cleavages C-terminal to aspartic acid residues are labeled in red, cleavages N-terminal to proline residues are labeled in blue, and non-specific cleavages are labeled in black

from $[M+n_{(PR-3)}H]^{n+}$ to $[M+n_{(PR+3)}H]^{n+}$, where PR is the number of predicted arginine residues. Inspection of the histogram plot of these mass-to-charge ratios provided in Fig. 3b reveals that greater than 75% of the charge states fall between *m/z* 1500 and *m/z* 2500.

Protocol Demonstration Using Reduced and Alkylated Trypsinogen

nESI of the reduced and alkylated trypsinogen sample under acidic denaturing conditions generates a broad and bi-modal distribution of charge states (Fig. 4a) with the highest and lowest
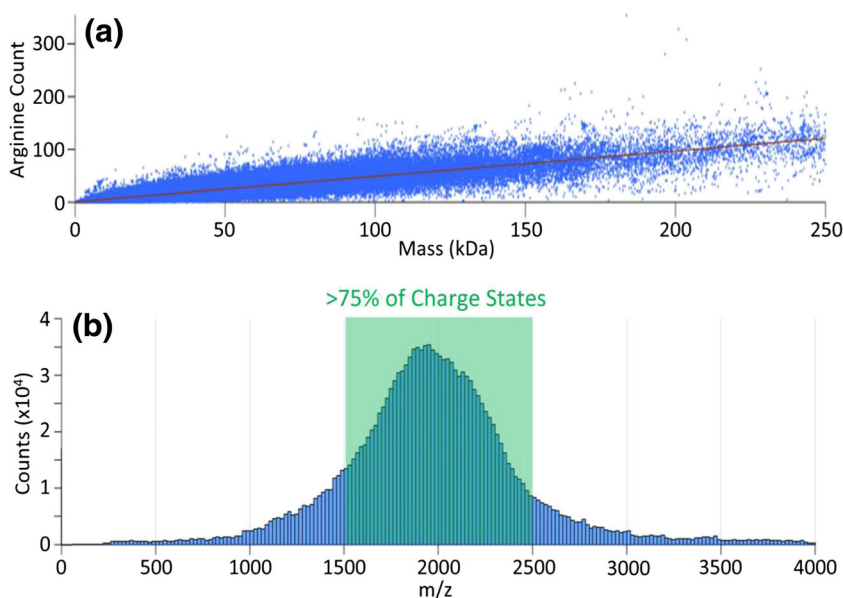
Figure 3. (**a**) Scatter plot of arginine count versus protein mass. (**b**) Histogram of mass-to-charge ratios calculated using ± 3 the predicted number of arginine residues
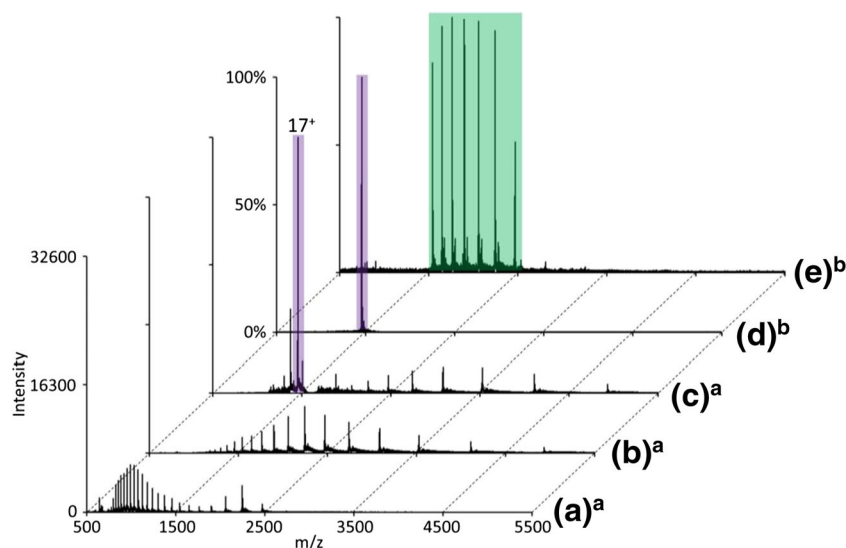
Figure 4. (**a**) Nano-electrospray mass spectrum of reduced and alkylated trypsinogen. (**b**) Post ion/ion spectrum from reaction of (**a**) with PFO dimer. (**c**) q0 ion parking of the $[M + 17H]^{17+}$ ion under the same reaction conditions as (**b**). (**d**) Q1 isolation of (**c**). (**e**) Post ion/ion spectrum and isolation of m/z 1500 to m/z 2500 from the reaction of (d) with PFO dimer. [a](a), (b), and (c) are plotted on the same absolute intensity scale while [b](d) and (e) are plotted on a scale of percent relative abundance

observable charge states being the $[M + 33H]^{33+}$ and $[M + 9H]^{9+}$ species, respectively. Much of the protein signal at the highest charge states was concentrated into a single lower charge state, $[M + 17H]^{17+}$, via a series of q0 transmission mode proton transfer ion/ion reactions and ion parking (Fig. 4c). Signal enhancement is evidenced by the ion abundance of the $[M + 17H]^{17+}$ species (compare Fig. 4c to Fig. 4a, b). Figure 4b demonstrates that there is no signal enhancement in the absence of ion parking, as no auxiliary excitation voltage is applied to inhibit the ion/ion reaction of the $[M + 17H]^{17+}$ species in that experiment.

The concentrated $[M + 17H]^{17+}$ ion was isolated in Q1 (Fig. 4d) and transferred to q2 where it was subjected to a second set of proton transfer ion/ion reactions. Reaction conditions were tuned such that the reduced charge states fell largely into the

m/z 1500–2500 window. This region is highlighted in green in Fig. 4e. Two high amplitude frequency sweeps were applied to q2, one using a frequency to eject ions lower in m/z than 1500, and one using a frequency to eject ions higher in m/z than 2500. The resulting spectrum is shown in Fig. 4e.

In the case of reduced and alkylated trypsinogen, m/z 1500 to m/z 2500 contains seven charge states, ranging from the $[M + 16H]^{16+}$ ion to the $[M + 10H]^{10+}$ ion. These charge states were subjected to broadband DDC collisional activation, maximizing the chances of observing the preferential cleavage at aspartic acid and proline residues. The MS/MS spectrum is provided in Fig. 5. Only b- and y-type fragment ions, matched within a ± 15 ppm tolerance, are labeled. As summarized in Table 1, a total of 12 fragments were identified, 6 coinciding
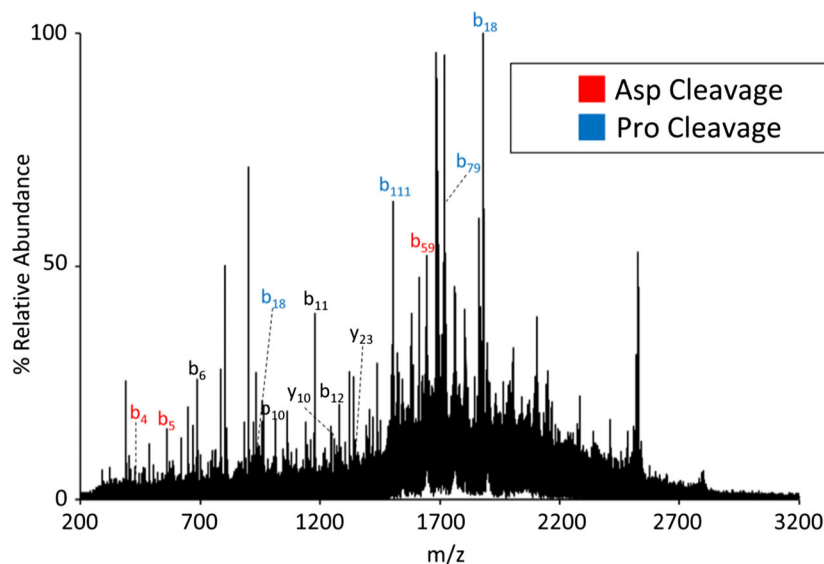
Figure 5. Simultaneous activation of the $[M + 16H]^{16+}$ to $[M + 10H]^{10+}$ charge states of reduced and alkylated trypsinogen via DDC collisional activation. Only peaks matched to ± 15 ppm of theoretical b- and y-fragment ions are labeled

Table 1. Summary of Database Search Results

Observed mass (monoisotopic)	8559.72
Theoretical mass (monoisotopic)	8559.62
Mass difference (ppm)	12.08
Identity	Ubiquitin
Species	Various
Sequence	MQIFVKTLTGKTITLEVEPSDTIENVKAKIQDKEGIPPDQQRLIFAGKQLEDGR TLSDYNIQKESTLHLVLRLRGG
Number of basic residues	Arg: 4, His: 1, Lys: 7, Total: 12
Matches	22 of 100 (22 of 100)[a]
McLuckey Score	100,719.97 (76,750.99)[b]
P-Score	3.8e−29 (3.8e−29)[c]
Matched fragments	b_7, b_8, b_9, b_{11}, b_{13}, b_{16}, b_{18}, b_{21}, b_{32}, b_{36}, b_{39}, b_{52}, y_9, y_{12}, y_{13}, y_{16}, y_{18}, y_{24}, y_{25}, y_{37}, y_{55}
Observed mass (monoisotopic)	16,941.02
Theoretical mass (monoisotopic)	16,940.96
Mass difference (ppm)	3.27
Identity	Myoglobin
Species	*Equus caballus* (horse)
Sequence	GLSDGEWQQVLNVWGKVEADIAGHGQEVLIRLFTGHPETLEKFDKFKHLKTEA EMKASEDLKKHGTVVLTALGGILKKKGHHEAELKPLAQSHATKHKIPIKYLEF ISDAIIHVLHSKHPGDFGADAQGAMTKALELFRND IAAKYKELGFQG
Number of basic residues	Arg: 2, His: 11, Lys: 19, Total: 32
Matches	24 of 45 (11 of 45)[a]
McLuckey Score	36,787.9 (5729.11)[b]
P-Score	6.1e−34 (9.9e−13)[c]
Matched fragments	b_{20}, b_{41}, b_{44}, b_{47}, b_{63}, b_{78}, b_{79}, b_{96}, b_{118}, b_{122}, y_7, y_8, y_9, y_{10}, y_{11}, y_{12}, y_{13}, y_{27}, y_{31}, y_{34}, y_{44}, y_{74}, y_{90}
Observed mass (monoisotopic)	24,662.18
Theoretical mass (monoisotopic)	24,661.82
Mass difference (ppm)	14.43
Identity	Trypsinogen
Species	*Bos taurus* (bovine)
Sequence	VDDDDKIVGGYTCGANTVPYQVSLNSGYHFCGGSLINSQWVVSAAHCYKSGIQVRLGEDNINV VEGNEQFISASKSIVHPSYNSNTLNNDIMLIKLKSAASLNSRVASISLPTSCASAGTQCLISGWGN TKSSGTSYPDVLKCLKAPILSDSSCKSAYPGQITSNMFCAGYLEGGKDSCQGDSGGPVVCSGKL QGIVSWGSGCAQKNKPGVYTKVCNYVSWIKQTIASN
Number of basic residues	Arg: 2, His: 3, Lys: 15, Total: 20
Matches	12 of 143 (2 of 143)[a]
McLuckey Score	129,384.11 (12,828.84)[b]
P-Score	3.7e−10 (0.24)[c]
Matched fragments	b_4, b_5, b_6, b_{10}, b_{11}, b_{12}, b_{18}, b_{59}, b_{79}, b_{111}, y_{10}, y_{23}
Observed mass (average)	29,024.55
Theoretical mass (average)	29,024.63
Mass difference (ppm)	− 2.83
Identity	Carbonic anhydrase 2
Species	*Bos taurus* (bovine)
Sequence	Ac-SHHWGYGKHNGPEHWHKDFPIANGERQSPVD IDTKAVVQDPALKPLALVYGEATSRRMVNNGHSFN VEYDDSQDKAVLKDGPLTGTYRLVQFHFHWGSSDDQGSEHTVDRKKYAAELHLVHWNTKYGDFGTA AQQPDGLAVVGVFLKVGDANPALQKVLDALDSIKTKGKSTDFPNFDPGSLLPNVLDYWTYPGSLTTPP LLESVTWIVLKEPISVSSQQMLKFRTLNFNAEGEPELLMLANWRPAQPLKNRQVRGFPK
Number of basic residues	Arg: 9, His: 11, Lys: 18, Total: 38
Matches	21 of 69 (2 of 69)[a]
McLuckey Score	183,239.01 (12,439.74)[b]
P-Score	1.6e−28 (0.047)[c]
Matched fragments	b_3, b_6, b_8, b_{17}, b_{18}, b_{31}, b_{33}, b_{38}, b_{40}, b_{40}, b_{70}, y_{17}, y_{27}, y_{47}, y_{48}, y_{61}, y_{62}, y_{63}, y_{64}, y_{67}, y_{68}

[a]The number of matched fragments of the second ranked protein is provided in parenthesis
[b]The McLuckey Score of the second ranked protein is provided in parenthesis
[c]The P-Score of the second ranked protein is provided in parenthesis

with selective fragmentation, 3 at aspartic acid, and 3 at proline. Visually, it is evident proline fragment ions are among the most abundant peaks in the spectrum.

Identification of Model Proteins

For each protein, database searching was initiated by retrieval of all proteins within the user-defined precursor tolerance. In all cases, this tolerance was set to ± 100 Da of the experimentally derived precursor mass, obtained from deconvolution based on Bayesian statistics within the PeakView Software (SCIEX, Concord, ON, Canada). Theoretical b- and y-fragment zero-charge masses were calculated for each protein in the retrieved list and compared to the experimental zero-charge masses. Proteins were ranked according to their McLuckey Score.
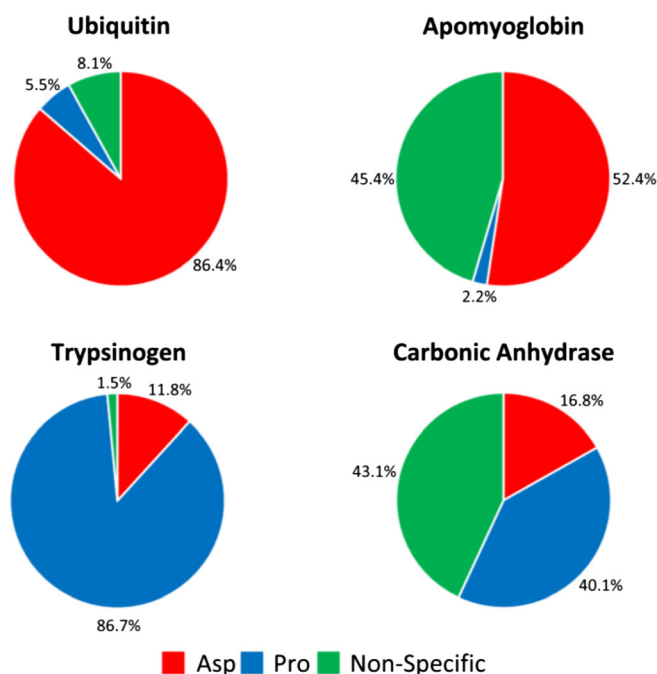
Figure 6. Percentage of the total matched fragment area for (red) cleavages C-terminal to aspartic acid residues, (blue) cleavages N-terminal to proline residues, or (green) non-specific cleavages

The peak lists used in the database searches are provided in the supplemental information.

In the case of ubiquitin, 2350 proteins were retrieved and scored in the database search. The top 153 scores were identical and, upon inspection, all 153 were correctly identified as ubiquitin. This is an artifact of the database creation as there are several parent proteins, across all species, in the SWISS-PROT database containing ubiquitin chains. Of the 100 fragment masses obtained by the THRASH deconvolution algorithm, 22 corresponded to b- and y-fragments of ubiquitin. The results of the top scoring protein are summarized in Table 1. By concentrating the precursor signal via proton transfer ion/ion reaction in conjunction with ion parking, and moving the charge state distribution to a region which maximizes selective cleavages, we observed greater than 90% of the matched fragment ion signal corresponding to cleavage at aspartic acid and proline residues. Specifically, 86.4% of the total matched fragment area was contained in seven unique aspartic acid fragments and 5.5% was contained in two proline fragment ions (Fig. 6).

ProSight Lite was used to calculate a P-Score for each of the top five ranked proteins [56]. While the output of the database search ranked according to the McLuckey Score does not necessarily coincide with ranking based on P-Score, in the case of ubiquitin, the top five unique protein sequences based on P-Score were identical to the top five unique protein sequences ranked according to their McLuckey Score. As shown in Table 1 and Table 2, the top two ranked proteins each matched 22 fragment masses of the 100 experimental fragment masses, resulting in equivalent P-Scores [57]. In our approach, three specific aspartic acid fragments, b_{52}, y_{37}, and y_{55}, provided sufficient information to distinguish between ubiquitin and ubiquitin-related proteins (supplementary Table S1). These results demonstrate the utility of using a scoring system that is weighted both by abundance and cleavage site as the conventional approach was unable to distinguish between ubiquitin and the ubiquitin-related protein.

For apomyoglobin, the database search returned 2446 results. The highest scoring protein was correctly identified as myoglobin from *Equus caballus*, while the second and third scoring proteins, as shown in supplementary Table S2, were found to be myoglobin from different species. Here, greater than 50% of the matched fragment area is comprised of seven aspartic acid fragments (Fig. 6). While equine myoglobin contains only two arginine residues, there are 19 lysine residues and 11 histidine residues in addition to the N-terminus, for a total of 33 nominal basic sites. When the charge state distribution is moved to the m/z 1500 to m/z 2500 region, $[M + 11H]^{11+}$ to $[M + 7H]^{7+}$, the ionizing protons may well be solvated by multiple basic residues, thereby reducing the proton mobility and thereby allowing for prominent cleavage at the aspartic acid residues.

Like equine myoglobin, trypsinogen is another protein containing only two arginine residues. The database search, which included carbamidomethylation (+ 57.0198 Da) of cysteine residues, successfully identified trypsinogen, matching only 12 fragment masses. Despite the lowest number of matched fragments, there remains a factor 10 difference in the score between trypsinogen and the second ranking protein (Table 1). Unlike apomyoglobin, where the majority of information was comprised of aspartic acid fragment ions, activation of trypsinogen resulted in cleavage at three proline residues, accounting for 86.7% of the matched fragment area (Fig. 6). The differences between the two systems arise from the total number of basic residues. Trypsinogen contains only 20 basic residues compared to myoglobin's 32. As previously stated, the m/z region subjected to DDC excitation for trypsinogen contains the $[M + 16H]^{16+}$ to $[M + 10H]^{10+}$ charge states. While some protons may be solvated by two basic residues, a sufficient

Table 2. Summary of the Top Five Proteins Ranked According to P-Score for Ubiquitin MS/MS Data

	Identity	Species	P-Score	Matches
1	Ubiquitin	Various	3.8e−29	22 of 100
2	Ubiquitin-related	Various	3.8e−29	22 of 100
3	Ubiquitin-related	*Equus caballus* (horse)	3.3e−24	19 of 100
4	Ubiquitin-related 1	*Cricetulus griseus* (Chinese hamster)	1.3e−22	18 of 100
5	Ubiquitin-related 1	*Dictyostelium discoideum* (slime mold)	5.2e−21	17 of 100

number of mobile protons remain to initiate cleavage at proline residues.

Lastly, two database searches for carbonic anhydrase were performed. The first search was performed with no N-terminal modifications while the second accounted for N-terminal acetylation (+ 42.0095 Da). The results of the former are summarized in supplementary Table S4. As shown in Table S4, the top two proteins exhibit the same score and the same number of matched fragments. The two sequences were identified as bovine carbonic anhydrase II, where the sequences differ only by the cleavage of the initiator methionine residue. In both cases, the 10 matched fragments were identical (y-type fragments). It was observed that the experimentally derived average mass of the precursor was 89 Da lower in mass than carbonic anhydrase with the N-terminal methionine, whereas the average mass of the precursor was 42 Da greater in mass than the sequence lacking the initiator methionine.

While it is possible to deduce the correct sequence based on mass, a second search was performed with N-terminal acetylation. Here, the top scoring protein was, again, correctly identified as carbonic anhydrase II lacking the initiator methionine with N-terminal acetylation. Note that the results of the top five proteins, as shown in Table S5, do not include carbonic anhydrase II with the initiator methionine since addition of the N-terminal modification puts this sequence outside of the ± 100-Da retrieval window. When comparing the two searches, N-terminal acetylation matched 11 additional fragments (b-type fragments), for a total of 21 matched fragments (Table 1). Despite being the largest protein examined in this study, selective cleavages are still observed, with 56.9% of the total matched fragment area held in aspartic acid and proline cleavages (Fig. 6).

Conclusions

Proof-of-concept is presented here for a protein top-down tandem mass spectrometry methodology that exploits the advantages of ionization under denaturing conditions while also maximizing the likelihood for selective cleavages at aspartic acid and proline residues. The generation of precursor ions under denaturing conditions and the relatively short reaction times are amenable to an online LC-MS workflow, though demonstration of an online LC-MS workflow was outside the scope of the present work. Ion/ion reactions are used both to concentrate ions of interest into a single charge state for ion isolation and to generate subsequently precursor ions within a mass-to-charge range most likely to yield selective cleavages under collisional activation conditions. The objective is to generate sufficient fragmentation for identification while concentrating product ion signals into a relatively limited number of channels. This approach is not intended to compete with strategies that maximize the number of backbone cleavages for characterization purposes. Rather, it is a complementary approach that is effective in protein identification that offers advantages in dynamic range and quantitation via single- or

multiple-reaction monitoring due to the fact that product ion signal is not distributed among a large number of product channels. Broadband collisional activation using DDC of precursor ions within the targeted *m/z* window for all proteins studied here resulted in greater than 50% of the total matched fragment area arising from aspartic acid and proline cleavages (up to 91.9% for ubiquitin). While the conventional scoring system based on matching predicted and observed product masses can be used with this approach yielding respectable P-Scores, a scoring system that weights most heavily C-terminal aspartic acid and N-terminal proline cleavages can add discriminatory power to an approach like this that is intended to maximize such cleavages. The combination of the ion processing workflow described here with an abundance-weighted scoring system lead to the successful identification of four model proteins, ranging in mass from 8.5 to 29.0 kDa. In the case of trypsinogen, successful identification was achieved with only 12 matched fragments (3 aspartic acid, 3 proline, and 6 non-specific cleavages), demonstrating the utility of maximizing selective fragmentation pathways. Interestingly, two of the selected model proteins contain an unusually low number of arginine residues. Nevertheless, selective cleavage at either proline or aspartic acid residues was noted. In addition to the application of intact protein identification, it stands to reason this workflow may be useful in multiple-reaction monitoring, particularly of low level proteins.

References

1. Loo, J., Edmonds, C., Smith, R.: Primary sequence information from intact proteins by electrospray ionization tandem mass spectrometry. Science. **248**, 201–204 (1990)
2. Schaffer, L.V., Shortreed, M.R., Cesnik, A.J., Frey, B.L., Solntsev, S.K., Scalf, M., Smith, L.M.: Expanding proteoform identifications in top-down proteomic analyses by constructing proteoform families. Anal. Chem. **90**, 1325–1333 (2018)
3. Khan, A., Eikani, C.K., Khan, H., Iavarone, A.T., Pesavento, J.J.: Characterization of *Chlamydomonas reinhardtii* core histones by top-down mass spectrometry reveals unique algae-specific variants and post-translational modifications. J. Proteome Res. **17**, 23–32 (2018)
4. Lubeckyj, R.A., McCool, E.N., Shen, X., Kou, Q., Liu, X., Sun, L.: Single-shot top-down proteomics with capillary zone electrophoresis-electrospray ionization-tandem mass spectrometry for identification of nearly 600 *Escherichia coli* proteoforms. Anal. Chem. **89**, 12059–12067 (2017)
5. Cai, W., Tucholski, T., Chen, B., Alpert, A.J., McIlwain, S., Kohmoto, T., Jin, S., Ge, Y.: Top-down proteomics of large proteins up to 223 kDa

enabled by serial size exclusion chromatography strategy. Anal. Chem. **89**, 5467–5475 (2017)

6. Kilpatrick, L.E., Kilpatrick, E.L.: Optimizing high-resolution mass spectrometry for the identification of low-abundance post-translational modifications of intact proteins. J. Proteome Res. **16**, 3255–3265 (2017)

7. Zubarev, R.A., Kelleher, N.L., McLafferty, F.W.: Electron capture dissociation of multiply charged protein cations. A nonergodic process. J. Am. Chem. Soc. **120**, 3265–3266 (1998)

8. 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. U. S. A. **101**, 9528–9533 (2004)

9. Brodbelt, J.S.: Photodissociation mass spectrometry: new tools for characterization of biological molecules. Chem. Soc. Rev. **43**, 2757–2783 (2014)

10. Compton, P.D., Zamdborg, L., Thomas, P.M., Kelleher, N.L.: On the scalability and requirements of whole protein mass spectrometry. Anal. Chem. **83**, 6868–6874 (2011)

11. Smith, L.M., Kelleher, N.L.: The consortium for top down, P.: proteoform: a single term describing protein complexity. Nat. Methods. **10**, 186 (2013)

12. McCormack, A.L., Somogyi, A., Dongre, A.R., Wysocki, V.H.: Fragmentation of protonated peptides: surface-induced dissociation in conjunction with a quantum mechanical approach. Anal. Chem. **65**, 2859–2872 (1993)

13. Dongré, A.R., Somogyi, Á., Wysocki, V.H.: Surface-induced dissociation: an effective tool to probe structure, energetics and fragmentation mechanisms of protonated peptides. J. Mass Spectrom. **31**, 339–350 (1996)

14. Paizs, B., Suhai, S.: Fragmentation pathways of protonated peptides. Mass Spectrom. Rev. **24**, 508–548 (2005)

15. Paizs, B., Suhai, S.: Towards understanding the tandem mass spectra of protonated oligopeptides. 1: mechanism of amide bond cleavage. J. Am. Soc. Mass Spectrom. **15**, 103–113 (2004)

16. Yu, W., Vath, J.E., Huberty, M.C., Martin, S.A.: Identification of the facile gas-phase cleavage of the Asp-Pro and Asp-Xxx peptide bonds in matrix-assisted laser desorption time-of-flight mass spectrometry. Anal. Chem. **65**, 3015–3023 (1993)

17. Kish, M.M., Wesdemiotis, C.: Selective cleavage at internal lysine residues in protonated vs. metalated peptides. Int. J. Mass Spectrom. **227**, 191–203 (2003)

18. Gehrig, P.M., Roschitzki, B., Rutishauser, D., Reiland, S., Schlapbach, R.: Phosphorylated serine and threonine residues promote site-specific fragmentation of singly charged, arginine-containing peptide ions. Rapid Commun. Mass Spectrom. **23**, 1435–1445 (2009)

19. Bleiholder, C., Suhai, S., Harrison, A.G., Paizs, B.: Towards understanding the tandem mass spectra of protonated oligopeptides. 2: the Proline effect in collision-induced dissociation of protonated Ala-Ala-Xxx-Pro-Ala (Xxx = Ala, Ser, Leu, Val, Phe, and Trp). J. Am. Soc. Mass Spectrom. **22**, 1032–1039 (2011)

20. McGee, W.M., McLuckey, S.A.: The ornithine effect in peptide cation dissociation. J. Mass Spectrom. **48**, 856–861 (2013)

21. Schwartz, B.L., Bursey, M.M.: Some proline substituent effects in the tandem mass spectrum of protonated pentaalanine. Biol. Mass Spectrom. **21**, 92–96 (1992)

22. Vaisar, T., Urban, J.: Probing proline effect in CID of protonated peptides. J. Mass Spectrom. **31**, 1185–1187 (1996)

23. Tsaprailis, G., Somogyi, Á., Nikolaev, E.N., Wysocki, V.H.: Refining the model for selective cleavage at acidic residues in arginine-containing protonated peptides22 dedicated to Bob Squires for his many seminal contributions to mass spectrometry and ion chemistry. Int. J. Mass Spectrom. **195-196**, 467–479 (2000)

24. Sullivan, A.G., Brancia, F.L., Tyldesley, R., Bateman, R., Sidhu, K., Hubbard, S.J., Oliver, S.G., Gaskell, S.J.: The exploitation of selective cleavage of singly protonated peptide ions adjacent to aspartic acid residues using a quadrupole orthogonal time-of-flight mass spectrometer equipped with a matrix-assisted laser desorption/ionization source. Int. J. Mass Spectrom. **210-211**, 665–676 (2001)

25. Gu, C., Tsaprailis, G., Breci, 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)

26. 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)

27. Newton, K.A., Chrisman, P.A., Reid, G.E., Wells, J.M., McLuckey, S.A.: Gaseous apomyoglobin ion dissociation in a quadrupole ion trap: [M + 2H]$^{2+}$?-[M + 21H]$^{21+}$.? Int. J. Mass Spectrom. **212**, 359–376 (2001)

28. Engel, B.J., Pan, P., Reid, G.E., Wells, J.M., McLuckey, S.A.: Charge state dependent fragmentation of gaseous protein ions in a quadrupole ion trap: bovine ferri-, ferro-, and apo-cytochrome *c*. Int. J. Mass Spectrom. **219**, 171–187 (2002)

29. Hogan, J.M., McLuckey, S.A.: Charge state dependent collision-induced dissociation of native and reduced porcine elastase. J. Mass Spectrom. **38**, 245–256 (2003)

30. Huang, Y., Triscari, J.M., Tseng, G.C., Pasa-Tolic, L., Lipton, M.S., Smith, R.D., Wysocki, V.H.: Statistical characterization of the charge state and residue dependence of low-energy CID peptide dissociation patterns. Anal. Chem. **77**, 5800–5813 (2005)

31. Huang, Y., Tseng, G.C., Yuan, S., Pasa-Tolic, L., Lipton, M.S., Smith, R.D., Wysocki, V.H.: A data-mining scheme for identifying peptide structural motifs responsible for different MS/MS fragmentation intensity patterns. J. Proteome Res. **7**, 70–79 (2008)

32. 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)

33. Newton, K.A., Pitteri, S.J., Laskowski, M., McLuckey, S.A.: Effects of single amino acid substitution on the collision-induced dissociation of intact protein ions: turkey ovomucoid third domain. J. Proteome Res. **3**, 1033–1041 (2004)

34. 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)

35. Loo, J.A., Loo, R.R.O., Udseth, H.R., Edmonds, C.G., Smith, R.D.: Solvent-induced conformational changes of polypeptides probed by electrospray-ionization mass spectrometry. Rapid Commun. Mass Spectrom. **5**, 101–105 (1991)

36. 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)

37. Ahadi, E., Konermann, L.: Modeling the behavior of coarse-grained polymer chains in charged water droplets: implications for the mechanism of electrospray ionization. J. Phys. Chem. B. **116**, 104–112 (2012)

38. Wang, G., Cole, R.B.: Effect of solution ionic strength on analyte charge state distributions in positive and negative ion electrospray mass spectrometry. Anal. Chem. **66**, 3702–3708 (1994)

39. Pan, P., McLuckey, S.A.: Electrospray ionization of protein mixtures at low pH. Anal. Chem. **75**, 1491–1499 (2003)

40. Pan, P., Gunawardena, H.P., Xia, Y., McLuckey, S.A.: Nanoelectrospray ionization of protein mixtures: solution pH and protein pI. Anal. Chem. **76**, 1165–1174 (2004)

41. Yue, X., Vahidi, S., Konermann, L.: Insights into the mechanism of protein electrospray ionization from salt adduction measurements. J. Am. Soc. Mass Spectrom. **25**, 1322–1331 (2014)

42. National Resource for Translational and Developmental Proteomics Home Page. http://nrtdp.northwestern.edu/. Accessed 3 Mar 2018

43. Wessel, D., Flügge, U.I.: A method for the quantitative recovery of protein in dilute solution in the presence of detergents and lipids. Anal. Biochem. **138**, 141–143 (1984)

44. Xia, Y., Chrisman, P.A., Erickson, D.E., Liu, J., Liang, X., Londry, F.A., Yang, M.J., McLuckey, S.A.: Implementation of ion/ion reactions in a quadrupole/time-of-flight tandem mass spectrometer. Anal. Chem. **78**, 4146–4154 (2006)

45. Liang, X., Xia, Y., McLuckey, S.A.: Alternately pulsed nanoelectrospray ionization/atmospheric pressure chemical ionization for ion/ion reactions in an electrodynamic ion trap. Anal. Chem. **78**, 3208–3212 (2006)

46. Liang, X., McLuckey, S.A.: Transmission mode ion/ion proton transfer reactions in a linear ion trap. J. Am. Soc. Mass Spectrom. **18**, 882–890 (2007)

47. Emory, J.F., Hassell, K.H., Londry, F.A., McLuckey, S.A.: Transmission mode ion/ion reactions in the radiofrequency-only ion guide of hybrid tandem mass spectrometers. Rapid Commun. Mass Spectrom. **23**, 409–418 (2009)

48. McLuckey, S.A., Reid, G.E., Wells, J.M.: Ion parking during ion/ion reactions in electrodynamic ion traps. Anal. Chem. **74**, 336–346 (2002)

49. Xia, Y., Wu, J., McLuckey, S.A., Londry, F.A., Hager, J.W.: Mutual storage mode ion/ion reactions in a hybrid linear ion trap. J. Am. Soc. Mass Spectrom. **16**, 71–81 (2005)

50. Webb, I.K., Londry, F.A., McLuckey, S.A.: Implementation of dipolar direct current (DDC) collision-induced dissociation in storage and transmission modes on a quadrupole/time-of-flight tandem mass spectrometer. Rapid Commun. Mass Spectrom. **25**, 2500–2510 (2011)

51. Giglione, C., Boularot, A., Meinnel, T.: Protein N-terminal methionine excision. Cell. Mol. Life Sci. **61**, 1455–1474 (2004)

52. Horn, D.M., Zubarev, R.A., McLafferty, F.W.: Automated reduction and interpretation of high resolution electrospray mass spectra of large molecules. J. Am. Soc. Mass Spectrom. **11**, 320–332 (2000)

53. Reid, G.E., Shang, H., Hogan, J.M., Lee, G.U., McLuckey, S.A.: Gas-phase concentration, purification, and identification of whole proteins from complex mixtures. J. Am. Chem. Soc. **124**, 7353–7362 (2002)

54. Gorman, G.S., Speir, J.P., Turner, C.A., Amster, I.J.: Proton affinities of the 20 common alpha.-amino acids. J. Am. Chem. Soc. **114**, 3986–3988 (1992)

55. Harrison, A.G.: The gas-phase basicities and proton affinities of amino acids and peptides. Mass Spectrom. Rev. **16**, 201–217 (1997)

56. 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)

57. Meng, F., Cargile, B.J., Miller, L.M., Forbes, A.J., Johnson, J.R., Kelleher, N.L.: Informatics and multiplexing of intact protein identification in bacteria and the archaea. Nat. Biotechnol. **19**, 952 (2001)