Published in final edited form as:

J Am Soc Mass Spectrom. 2017 June ; 28(6): 1203–1215. doi:10.1007/s13361-017-1635-x.

Defining gas-phase fragmentation propensities of intact proteins during native top-down mass spectrometry

Nicole A. Haverland, **Owen S. Skinner**, **Ryan T. Fellers**, **Areeba A. Tariq**, **Bryan P. Early**, **Richard D. LeDuc**, **Luca Fornelli**, **Philip D. Compton**, and **Neil L. Kelleher***

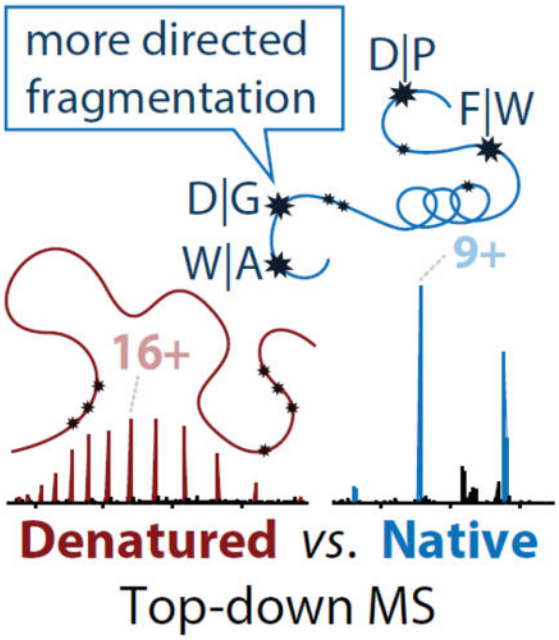
Departments of Chemistry and Molecular Biosciences, And the Proteomics Center of Excellence, Northwestern University, 2170 Campus Drive, Evanston, IL 60208

Abstract

Fragmentation of intact proteins in the gas phase is influenced by amino acid composition, the mass and charge of precursor ions, higher order structure, and the dissociation technique used. The likelihood of fragmentation occurring between a pair of residues is referred to as the fragmentation propensity and is calculated by dividing the total number of assigned fragmentation events by the total number of possible fragmentation events for each residue pair. Here, we describe general fragmentation propensities when performing top-down mass spectrometry (TDMS) using denaturing or native electrospray ionization. A total of 5,311 matched fragmentation sites were collected for 131 proteoforms that were analyzed over 165 experiments using native top-down mass spectrometry (nTDMS). These data were used to determine the fragmentation propensities for 399 residue pairs. In comparison to denatured top-down mass spectrometry (dTDMS), the fragmentation pathways occurring either N-terminal to proline or C-terminal to aspartic acid were even more enhanced in nTDMS as compared to other residues. More generally, 257/399 (64%) of the fragmentation propensities were significantly altered ($p \leq 0.05$) when using nTDMS as compared to dTDMS and of these, 123 were altered by 2-fold or greater. The most notable enhancements of fragmentation propensities for TDMS in native vs. denatured mode occurred (1) C-terminal to aspartic acid, (2) between phenylalanine and tryptophan (F|W), and (3) between tryptophan and alanine (W|A). The fragmentation propensities presented here will be of high value in the development of tailored scoring systems used in nTDMS of both intact proteins and protein complexes.

Graphical Abstract

*Address reprint requests to: Neil L. Kelleher, Northwestern University, 2170 Campus Drive, Evanston, IL 60208, n-kelleher@northwestern.edu, Phone: (847) 467-4362, *Fax: (847) 467-3276.



Introduction

The analysis of intact proteins via “top-down” proteomics [1, 2] has grown in popularity in recent years and has been used to show proteomic differences associated with senescence [3–5], a model of bacterial infection [6], and myocardial infarction [7]. A unique advantage of top-down proteomics is the accurate measurement of the mass of an intact proteoform. The term “proteoform” describes the biological variability for a protein by taking into account gene or protein processing events (*e.g.,* mutations, post-translational modifications) [8]. To fully characterize the proteoform and localize the modifications, it is necessary to controllably disassemble protein ions in the gas phase using tandem mass spectrometry (MS/MS).

Since the first correlation of gaseous protein fragment ions to primary sequence in 1990 [9], MS/MS has been widely used to enable mass spectrometry-based analysis of peptides [10, 11, 12] and proteins [13, 14]. The process requires the measurement of the intact mass of the precursor, which is then isolated in the gas phase and activated to break its interresidue bonds. Many methods for activation and fragmentation are available (reviewed in [15, 16]), but one of the most common is via energetic collisions with a neutral gas, which is termed collisionally-induced dissociation (CID). Variations of CID exist, with the two major types being (1) ion trap CID, which uses resonant excitation to excite ions and potentiate radial movement that slowly dissociates trapped ions after many collisions, or (2) beam-style CID, which is performed using higher-energy axial acceleration to rapidly activate and dissociate the ions. While these techniques both produce *b*- and *y*-type protein backbone cleavages [17], each can produce a different set of cleavage products [18, 19]. As opposed to CID, electron-based and ultraviolet photon-based fragmentation techniques rely on much faster processes, breaking different backbone bonds and producing potentially more than ten

different fragment types [20–22]. Regardless of the method used for dissociation, the fragment ions can be used to determine the residue fragmentation propensities, which are described here as the chance that a fragmentation event will occur between any given pair of residues.

The efforts that have been put forth to analytically describe the gas-phase residue fragmentation propensities for peptides using a variety of fragmentation techniques have contributed to the development of the mobile proton model [23] (reviewed in [24, 25]). This model describes fragmentation of peptides in the gas phase with considerations for residue composition [23, 26–28], gas-phase basicity [23, 28–30], secondary structure [30], relative location of the basic residue [31, 32], and the theoretical mobility of the proton (e.g., sequestered vs. mobile) [23, 32]. For instance, following ionization, a peptide that has a greater number of charges than the number of residues with high gas-phase basicity (*e.g.,* arginine [33]) will have protons readily mobilize to amide bonds to induce backbone fragmentation [23]. The energy requirements for fragmentation of peptides with mobile protons are relatively low [23, 28, 29]. In contrast, if the peptide has fewer charges than the number of basic residues, the proton(s) is sequestered to the basic residue(s) and the activation energy needed to mobilize the proton and induce backbone fragmentation is increased substantially [23, 28–30, 34]. The conventional fragmentation pathways that are best described by the mobile proton model are grouped into four major clusters, including fragmentation occurring (i) N-terminal to proline (X|P), (ii) C-terminal to isoleucine, leucine, or valine (I/L/V|X), (iii) C-terminal to aspartic acid or glutamic acid (D/E|X), and (iv) “b&y” fragmentation [31, 32]. This “b&y” fragmentation pathway generally includes ions with arginine near the N-terminus (*e.g.,* in the case of missed tryptic cleavages) or internal fragment ions (ions that form by re-fragmentation of a *b*- or *y*-type ion) [31, 32, 35]. Fragmentation through these four major channels has also been investigated for intact proteins analyzed under denaturing conditions [35]. It is important to note that development and application of this model primarily used peptides and proteins that were analyzed using acidified, denatured electrospray ionization (dESI) conditions. Although it is well established that the ESI solvent can affect peptide and protein conformation and protonation [36–38], the extension of the mobile proton model as applied to the analysis of intact proteins and complexes using native ESI (nESI) has had limited exploration.

Beyond the improved understanding of the general trends of fragmentation for whole protein cations, characterization of residue fragmentation propensities for different protein activation types can also be used as priors to enhance protein identification [39] and characterization of proteoforms [40] in top-down proteomics. However, since the vast majority of MS/MS-based top-down proteomics has been performed on denatured proteins, the scoring systems used for identification and characterization of proteoforms have naturally been developed and tested using datasets derived from denatured top-down mass spectrometry (dTDMS) experiments. Native top-down mass spectrometry (nTDMS) approaches are proliferating because nESI can preserve many native properties of proteins and protein complexes, including aspects of structure [41, 42], interactions [41, 43–47], and activity [41, 48] (these topics are also reviewed in [49–51]). Further, nTDMS has an additional benefit of increased signal-to-noise due to fewer overall charge states [52]. All in all, nTDMS is readily becoming a powerful technique for the analysis of intact proteins. In fact, the Kaltashov and

Ge groups have recently made advancements in on-line separations coupled with native mass spectrometry for the analysis of native proteins and complexes [53–55], which is an important step in establishing nTDMS as a technique capable of supporting discovery mode studies. However, a scoring system will need to be optimized for nTDMS as the fragmentation patterns for native proteins appear to be altered as compared to those observed for denatured proteins [33, 37, 54, 56–58]. An essential step for optimizing these scoring systems is to better define the differences in fragmentation propensity when analyzing proteins using nTDMS or dTDMS. Here, we provide the first statistically-backed analysis of the differences in residue fragmentation propensities between intact, endogenous proteins analyzed with nTDMS as compared to dTDMS.

Experimental

Sample preparation

Four human cell lines were obtained from the German Collection of Microorganisms and Cell Cultures GmbH (www.dsmz.de) or the American Type Culture Collection (www.atcc.org), including Hg-3 (DSMZ ACC 765), Ramos (ATCC CRL-1596), Jurkat (ATCC TIB-152), and HEK-293T (ATCC CRL-11268) cell lines. Cells were grown following manufacturer recommended culture methods using 150 mm^2 flasks to obtain >100 million cells per cell line. Cells were collected by centrifugation (300 × *g* for 5 min at 4 °C), washed using phosphate buffered saline (PBS), re-pelleted, and stored at −80 °C until processed.

All chemicals were obtained from Sigma, unless otherwise noted. Cells were resuspended in a hypotonic buffer composed of 15mM Tris-HCl (pH 7.5), 60 mM KCl, 15 mM NaCl, 5 mM $MgCl_2$, 1 mM $CaCl_2$, 250 mM sucrose, 10 mM sodium butyrate, and 1x HALT protease and phosphatase inhibitor (Thermo). Cells were allowed to swell for 45 – 60 min. Cells were then lysed using mild sonication with a probe sonicator set to 30% amplitude and 2 × 30 second pulse. Cellular debris, including the nuclei and other intact organelles, were pelleted using centrifugation (11,000 × *g* for 5 min at 4 °C). The soluble cytosolic proteins were then filtered using a Millex HV PVDF 0.45 μm syringe filter. Next, the samples were desalted using Amicon Ultra-0.5 centrifugal filter units in a stepwise manner: First, the lysate was concentrated using 100 kDa nominal molecular weight limit (NMWL) filter. The flow-through was retained and concentrated using a 3 kDa NMWL centrifugal filter, effectively creating a fractionated sample with proteins (and complexes) ranging from ~3–100 kDa. This 3–100 kDa sample was then buffer exchanged using a 10 mM ammonium acetate (Sigma) buffer prepared using Optima LC/MS water (Fisher) within the 3 kDa NMWL centrifugal filter; >5 rounds of desalting was performed to maximize downstream separation.

Ion-exchange chromatography (IEX)

Samples were fractionated by IEX using the Agilent series 1100 (degasser, pumps, column compartment, UV module) and 1200 (fraction collector, fraction chiller). The column was a mixed-bed IEX column composed of an equal proportion of PolyCAT A and PolyWAX LP (PolyLC; Columbia, MD). Two columns were used over the course of this project: 12 μm resin with a 1500 Å pore size and 5 μm resin with a 1000 Å pore size. Optimized methods

for separation were developed using the Agilent Chemstation control software using a gradient of Buffer A (10 mM ammonium acetate, pH 7) and Buffer B (1 mM ammonium acetate, pH 7). Fractions were collected into a 96 deep-well plate using 1.5 min time slices, which were not dependent on the elution profile. The separation was monitored using UV absorbance at 280 nm. The column compartment was maintained at a constant temperature of 21 °C for the duration of the run. Fractions were concentrated using a 3 kDa NMWL centrifugal filters and desalted using a 150 mM ammonium acetate buffer, pH 7.

Mass spectrometry

Electrospray ionization was performed using a custom nano electrospray source [59] applying between 0.8–1.6 kV voltage. All analyses were performed on a modified Q-Exactive HF mass spectrometer [60] using methods as previously described [57, 58]. Briefly, the intact mass spectrum (MS^1) was acquired using low (~5 V) source-induced dissociation to relieve salt adducts from the precursor ions. Next, a single charge state of a precursor was quadrupole-isolated and fragmented (MS^2) using higher-energy collisional dissociation (HCD), which is a type of beam-style CID [61]. For most analyses, a resolving power of >120,000 (at 200 *m/z*) was used for both MS^1 and MS^2. Deconvolution of isotopically resolved precursor and fragment ions was performed using the Xtract software (Thermo). When the precursor was not isotopically resolved, the average mass was manually calculated from an MS^1 collected using either a resolving power of 15,000 or 30,000 (at 200 *m/z*).

Informatics and statistics

The ProSightPC 4.0 software suite (Thermo) was used for identification and characterization of proteoforms using the precursor mass (monoisotopic or average) and monoisotopic fragment ion masses, which were searched using a database consisting of Swiss-Prot entries from UniProt (database retrieved on October 27, 2015). The precursor mass tolerance was set between 200–4000 Da to account for unexpected mass shifts resulting from ligand/co-factor binding, splice variants, post-translational modifications, and/or truncation events. The fragment ion mass tolerance was set to ≤20 ppm. Proteoform data were exported from ProSightPC to ProSight Lite [62] to generate graphical fragment maps, which were saved as .pcml files. These .pcml files were used to count the total number of assigned fragmentation events for each residue pair as well as the total number of possible fragmentation events, that latter of which was multiplied by 2 to account for bidirectional cleavages. Ratios were calculated as the total assigned divided by total possible fragmentation events for each residue pair; these ratios were used as the fragmentation propensities.

Sampling of the nTDMS dataset and dTDMS dataset were performed using (1) the entire nTDMS dataset and (2) HCD hits of 13,034 proteoforms that were identified with a 1% global false discovery rate (FDR) from the dTDMS dataset in reference [4]. For each round of sampling per dataset, 50% of the hits were randomly selected and the fragmentation propensity was calculated. A total of 10 passes were performed for each dataset and the average fragmentation propensity and standard error were calculated for each residue pair. Differences between the nTDMS dataset and the dTDMS dataset were determined using an independent samples Student’s t-test with Bonferroni correction to account for multiple

comparisons. The percentage point change is the absolute difference between two percentages and was calculated by subtracting the dTDMS fragmentation propensities from the nTDMS propensities. In contrast, the percent change is the relative change in residue fragmentation propensities and was calculated by dividing the percentage point difference of the nTDMS and dTDMS fragmentation propensities by the dTDMS fragmentation propensities (((nTDMS – dTDMS)/dTDMS) × 100%).

Results and Discussion

Acquisition of nTDMS data on protein monomers (4–70 kDa)

While conducting previous work in the analysis of protein complexes using nTDMS [58], we observed a trend in fragmentation patterns: matched fragment ions in nTDMS were seemingly even more prevalent for the C-terminal side of aspartic and glutamic acid residues and the N-terminal side of prolines than previously established for dTDMS [63]. In order to quantitatively examine this trend, we compiled a dataset containing fragmentation events collected from human proteins that were analyzed using nTDMS/MS with HCD. This dataset includes 85 human proteins with 131 unique proteoforms that were collected during 165 independent acquisitions of MS/MS data. To be considered an independent acquisition, the data must have been acquired from a unique charge state, from separate sample preparations, and/or from distinct cell lines. The molecular weight distribution of proteoforms ranged from 4 kDa to over 70 kDa. Our approach for the analysis of protein complexes [58, 60, 64] uses energy to eject monomers in the source region of the mass spectrometer. Often, these ejected monomers will be observed with a disproportionately high number of charges as compared to the intact precursor in a process known as asymmetric charge partitioning [46, 65], which can attributed to partial or complete unfolding of the monomer [46, 66, 67] or to heterolytic scission of ion pairs [68]. Because the preferred fragmentation pathways may be affected by this increased protonation, monomers that were ejected from complexes were *not* included in this study.

Overall, a total of 28,250 fragment ion masses were acquired and of these, 5,311 (18.8%) were assigned to the precursor sequence positions with high confidence. The number of fragmentation events were not evenly distributed across all residue pairs; rather, 60% of the ~5300 assigned fragmentation events occurred C-terminal to aspartic acid, glutamic acid, or lysine residues or N-terminal to proline or glycine (Figure 1A). When deconstructed to characterize the effect of unique residue pairs on fragmentation, large biases in residue fragmentation were observed (Figure 1B). Assuming an even distribution of fragmentation across all 400 residue pairs, ~13 fragmentation events would map to each pair on average. However, the number of fragmentation events ranged from 0 on the low end to over 100 events mapped to a single pair (8-fold greater than expected for an even distribution, Figure 1B).

The fragmentation propensity describes the likelihood of fragmentation occurring between a specific set of residues while normalizing for differences in residue pair frequency occurring for proteins within the dataset. The residue fragmentation propensities were calculated by dividing the total number of observed matching fragment ions for the residue pair by the total number of possible fragmentation events for that same residue pair. The average

fragmentation propensity across all pairs was 8%, and ranged from 0% to 64% (Figure 2A, B). The fragmentation pathways that occurred either C-terminal to aspartic acid or N-terminal to proline were well above the average with propensities of 38% and 24% (Figure 2A), respectively. These preferred cleavage pathways have been previously documented in nTDMS when assessing the relationship between the precursor's charge state and residue fragmentation propensities, but have been limited in scope to one or a few purified protein species [35, 37, 56, 69–73]. In addition to fragmentation pathways involving proline or aspartic acid, several residue pairs displayed notably higher or lower than expected fragmentation propensities defined as >20% or <5%, respectively (Table 1). Altogether, these results indicate that specific residues can impact - in a strong and local fashion - the residue fragmentation propensity during a nTDMS/MS experiment using HCD.

Fragmentation propensities are significantly different for nTDMS versus dTDMS

Next, the HCD fragmentation propensities for nTDMS were compared to those for dTDMS. First, the fragmentation propensities for all residue pairs were calculated using an extensive dTDMS dataset published previously (Figure 2A) [4]. This dataset was filtered to include 13,034 proteoforms, which were identified with a global FDR of <1% and fragmented using HCD. Few dramatic differences were observed when comparing the fragmentation pathways for each residue in nTDMS and dTDMS (Figure 2A); the notable exception was fragmentation occurring C-terminal to aspartic acid, which was increased 2-fold in the nTDMS dataset as compared to the dTDMS dataset. When deconstructed by residue pair, a trend for increased fragmentation propensity was observed within the dTDMS dataset among the more hydrophobic residue pairs (leucine, isoleucine, valine, phenylalanine, and methionine; Figure 2C), which is not as evident within the nTDMS dataset (Figure 2B). Overall, the residue fragmentation propensities for each pair appears to be more evenly distributed in dTDMS as compared to nTDMS with an average residue fragmentation propensity of 9% and a range of 0.02% to 37%.

To further quantify the differences in residue fragmentation propensities for nTDMS and dTDMS, bootstrapping was performed for each dataset to obtain summary statistics (standard error and mean fragmentation propensity). Using the Student's t-test with the Bonferroni correction for multiple comparisons (n=399), the fragmentation propensities for a total of 257 residue pairs were significantly different ($p \leq 0.05$) between the two datasets and of these, 123 residue pairs were altered by 2-fold or greater (Figure 3). The most extreme relative change was proline|histidine with an increase of >12-fold in nTDMS as compared to dTDMS. In contrast, the largest relative decrease was -2-fold, which was the case when no matched fragmentation events were observed for a residue pair. Additionally, several trends were observed in nTDMS as compared to dTDMS, including enhanced fragmentation occurring C-terminal to aspartic acid and diminished fragmentation between the more bulky and hydrophobic residues (leucine, isoleucine, valine, phenylalanine, methionine, and tryptophan; Figure 3). The complete dataset containing residue fragmentation propensities from both nTDMS and dTDMS is provided in Online Resource 1. These results indicate that even while using the same type of dissociation, ionization of the protein in either native or denatured mode will significantly influence the fragmentation propensities for backbone amide bonds flanked by the majority of amino acid pairings.

Correlating fragmentation with charge state, sources of mobile protons, and tertiary structures

The number of protonation events for a protein can influenced in part by the intrinsic properties of a protein and the ionization technique that is used [74, 75] (reviewed in [76]). In fact, one of the more profound observable differences between dESI and nESI the effect of ESI solvent on the charge state distribution of a protein [37]. For example, a protein that has been denatured is able to carry additional protons, which is indicative of an elongated structure with a higher surface area and increased accessibility to ionizable residues. In contrast, a globular protein that retains a native-like fold during the transfer from solution to the gas phase will likely have a decreased surface area. In the case of globular proteins, the theoretical maximum number of charges that can be imparted to the protein during electrospray is described by the Rayleigh charge (z_R) limit theory, which has been defined and discussed elsewhere [74, 75, 77, 78] and is largely influenced by the surface tension of the solvent and the radius of the ion. Thus, because a denatured protein will have a much larger surface area than its native (compact) form, the denatured protein will not necessarily conform to the Rayleigh charge limit theory and may be observed with charge states greater than the Rayleigh charge ($z > z_R$). (For a review on protein charging and supercharging with a focus on ESI, see [79]). Based on the results presented here and elsewhere [35, 51, 63, 73, 80], we posit that the residue fragmentation propensities for intact proteins are primarily influenced by the thermalized structures, the net charge, dissociation technique, and the residue composition of precursor ions.

A given charge state for a protein is typically designated as a high, intermediate, or low charge state, which is a historic reference to the McLuckey group's designation for ubiquitin, with high charge states defined as 13+ to 10+, intermediate as 9+ to 7+, and low as 6+ and below [56]. It is possible to extrapolate these designations to the proteins analyzed here by nTDMS by determining the ratio of observed charge to Rayleigh charge (z/z_R); the latter of which can be calculated as $z_R = 0.0778m^{½}$, but has the assumption that the structure of the ion is a compacted globular sphere in an aqueous environment and, as such, the ion density is $\rho = 1$ g/cm^3 and the surface tension is that of water, $\gamma = 0.072$ N/m [75]. Therefore, given that the mass of ubiquitin is 8.5 kDa, the z_R is calculated as 7 (rounded to the nearest whole integer) and the z/z_R for the high charge states is >1.43 and the z/z_R for low charge states is <0.86, with the intermediate charge states falling in between these values. In calculating the z/z_R ratio for each of the precursor ions used in the nTDMS dataset reported here, we found 102 had low charge state precursor ions, 57 had intermediate charge state precursor ions, and 6 had high charge state precursor ions (Figure 4A). These results indicate that the majority of precursor ions used in this analysis conform to the Rayleigh charge limit theory predictions. However, a nontrivial number of precursors have $z > z_R$, especially among precursors with a mass of <30 kDa (Figure 4A). We anticipated that these high and intermediate charge state precursor ions may be an outcome of non-sphericity due to the intrinsic fold of the protein (*e.g.*, inclusion of fibrous and disordered protein precursor ions). In fact, among the precursor ions with the high charge state designation was STMN1 (P16949), which is a founding member for the class of intrinsically disordered/unstructured proteins [81–83]. Additionally, COF1 (P23528), which has an unstructured N-terminus when unmodified [84], had precursor ions that spanned all three charge state designations

and were included in the nTDMS dataset. The fragmentation patterns for COF1 are distinct for each charge state designation (Figure 4B, C, D), which can be attributed to the differences in net charge. However, one should also consider that starting precursor intensity can be a confounding factor and influence the ability to detect fragment ions, which may explain the relatively few fragment ions that were observed in the high charge state designation of COF1 (Figure 4D). Additionally, the fragmentation patterns observed for COF1 during dTDMS are distinct from even the high charge state designation in nTDMS (Figure 4E, F), which suggests that aspects of tertiary structure could correlate strongly to the observed gas-phase fragmentation patterns.

Although one cannot assume that the gas-phase structure of a protein will mimic its crystal structure, we investigated the possible role of tertiary structure on fragmentation by mapping the fragmentation events to the crystal structure for a variety of proteins. The largest protein that had multiple charge state designations was GDIR1 (P52565; 23 kDa), which has a globular core with an extended, flexible arm and a disordered region (Figure 5A). The two charge states that were included in the nTDMS portion of this study represent a low and intermediate charge state with z/z_R ratios of 0.76 (9+) and 1.18 (14+). The fragmentation patterns for these charge states are similar with several notable differences, including an increased number of cleavages occurring at aspartic acid in the 9+ precursor as compared to the 14+ precursor (Figure 5B, C). The fragmentation patterns for both of the nTDMS precursors sharply contrast the patterns observed by dTDMS (Figure 5D). Interestingly, mapping the fragmentation events observed for the 9+ precursor ion of GDIR1 to its crystal structure reveals a trend for fragmentation occurring at or near the solvent accessible surface area of the protein and proximal to several arginine residues (Figure 5A). This trend was also observed in other precursor ions representing both low and intermediate charge states across a wide range of masses (Figure 6).

With some appreciation for the relationships between charge state, tertiary structure of the analogous crystal structure, and preferred fragmentation patterns in nTDMS, we revisit the mobile proton model. In contrast to dESI, which imparts charge to a variety of residues with largest proton affinities (arginine, lysine, histidine, and the N-terminal α-amino group [37]), nESI results in protonation primarily at arginine residues [85]. The difference in protonation of the precursor is important when considering proton mobility and sequestration. A proton is considered mobile if the total number of charges is greater than the number of basic residues present in the ion, which is often the case with dTDMS experiments. When a low level of activation energy is applied, the proton readily mobilizes from the side chain of the residue to the backbone amide bond to induce charge-directed fragmentation. The specificity or preference for specific fragmentation pathways is diminished with mobile protons, which can result in fragmentation that occurs between a greater variety of residue pairs [31, 35, 56] and was consistent with the dTDMS dataset reported here (Figure 2C) and elsewhere [35]. In contrast, a proton is sequestered if the total number of charges is equal to or less than the number of basic residues present in the ion, which is more reflective of the conditions in nESI. In this case, fragmentation will occur via pathways that do not require intramolecular proton transfer (e.g., charge-remote fragmentation) or if the activation energy is increased to promote mobilization of the sequestered proton [23, 28–30, 34]. However, if the proton is sequestered and an acidic residue is present, the acidic hydrogen in the side chain of the

residue can act as the proton source needed to induce fragmentation [28, 30, 34]. This is observed as an increased preference for fragmentation occurring C-terminal to glutamic acid and aspartic acid, and is consistent with the nTDMS data presented here (Figure 2B). In fact, of the 20 possible residue pairs with fragmentation occurring C-terminal to aspartic acid, the propensities for 17 of these pairs were increased by over 50% in nTDMS as compared to dTDMS (Figure 3). Based on these observations, the canonical view of the mobile proton model as it relates to sequestered protons corresponds well with the observations and trends reported here for both nTDMS and dTDMS.

The “proline effect” is the observation that upon collisional ion activation there is an increased fragmentation propensity occurring N-terminal to proline residues [86], which is likely influenced by the rigid structure and substituted amine in the *y*-ion’s leaving group of this cyclized residue. Loo and colleagues report that as the mass of a precursor ion increases (*e.g.*, peptide to protein to large protein), the likelihood of producing an interior cleavage event decreases; however, this trend is reversed when an interior proline residue(s) is present [86]. The Wysocki group expanded on this observation, noting that the energy demands for fragmentation is increased for precursor ions containing internal prolines, likely a consequence of the unique structure of the residue [23]. Further, an overall increase in fragmentation occurs N-terminal to proline when the residue pairs include valine, histidine, aspartic acid, isoleucine, and leucine; however, fragmentation is less likely when the residues N-terminal to the cleavage site are proline and glycine [26, 32]. Additionally, there is a relationship between increased proton mobility and increased fragmentation occurring N-terminal to proline residues in peptides [26, 31, 32] and proteins [35], despite the proton not necessarily being localized to the proline residue itself [87]. Given that the “proline effect” is heavily reliant on the sequence of the precursor ion, it would be expected that the residue fragmentation propensity occurring N-terminal to proline in both nTDMS and dTDMS would be relatively equivalent. However, the fragmentation propensities for several residue pairs were altered by over 2-fold when using nTDMS as compared to dTDMS, including alanine|proline, tyrosine|proline, and arginine|proline (Figure 3).

Additional investigations will be required to better define the interplay between gas-phase ion structure, protonation, and preferred fragmentation pathways as they relate to the analysis of intact proteins with native-like conformations. Likewise, defining the extent of internal fragmentation that occurs during nTDMS would also be beneficial, as it could provide a unique insight into how the structure of a precursor ions in the gas phase can influence fragmentation pathways, especially in comparison to its denatured counterpart [35, 73].

Conclusions

Here we have provided the first extended analysis of residue fragmentation propensities for proteoforms analyzed using nTDMS with beam-type HCD fragmentation. Additionally, we established that a great number of fragmentation pathways are significantly altered when using nTDMS as compared to dTDMS. Overall, we report a data-driven assertion that nTDMS funnels fragmentation into a fewer number of more highly preferred sites than dTDMS. Additionally, it was unexpected that the already high propensities for

fragmentation “hot spots” in dTDMS (*e.g.*, occurring C-terminal to aspartic acid) would be doubled when using nTDMS. It also seems that the mobile proton model and the proline effect can be used to explain many of the fragmentation pathways that are enhanced during nTDMS. The nTDMS results also show a strong correlation to surface exposed residues as viewed on the tertiary structures of selected examples.

Thus, it stands to reason that as nTDMS becomes more routine for the analysis of proteoforms and their complexes, expert scoring systems should be updated and informed with these trends. The residue fragmentation propensities presented here will be expanded and become an essential piece of information when adopting and adapting Bayesian scoring metrics, such as the C-score for proteoforms [40], the MPC-score for multi-proteoform complexes [58], and other scores generated by the growing community of developers and practitioners of TDMS.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors would like to acknowledge the W. M. Keck foundation for generous support and funding (DT061512). In addition, this work was partially supported by R01GM067193 (NIGMS). O.S.S. was supported by an NSF graduate research fellowship (2014171659).

References

1. Catherman AD, Skinner OS, Kelleher NL. Top Down proteomics: facts and perspectives. Biochem Biophys Res Commun. 2014; 445:683–693. [PubMed: 24556311]
2. Toby TK, Fornelli L, Kelleher NL. Progress in Top-Down Proteomics and the Analysis of Proteoforms. Annu Rev Anal Chem. 2016; 9:499–519.
3. Tran JC, Zamdborg L, Ahlf DR, Lee JE, Catherman AD, Durbin KR, Tipton JD, Vellaichamy A, Kellie JF, Li M, Wu C, Sweet SM, Early BP, Siuti N, LeDuc RD, Compton PD, Thomas PM, Kelleher NL. Mapping intact protein isoforms in discovery mode using top-down proteomics. Nature. 2011; 480:254–258. [PubMed: 22037311]
4. Catherman AD, Durbin KR, Ahlf DR, Early BP, Fellers RT, Tran JC, Thomas PM, Kelleher NL. Large-scale top-down proteomics of the human proteome: membrane proteins, mitochondria, and senescence. Mol Cell Proteomics. 2013; 12:3465–3473. [PubMed: 24023390]
5. Durbin KR, Fornelli L, Fellers RT, Doubleday PF, Narita M, Kelleher NL. Quantitation and Identification of Thousands of Human Proteoforms below 30 kDa. J Proteome Res. 2016; 15:976–982. [PubMed: 26795204]
6. Ansong C, Wu S, Meng D, Liu X, Brewer HM, Deatherage Kaiser BL, Nakayasu ES, Cort JR, Pevzner P, Smith RD, Heffron F, Adkins JN, Pasa-Tolic L. Top-down proteomics reveals a unique protein S-thiolation switch in Salmonella Typhimurium in response to infection-like conditions. Proc Natl Acad Sci USA. 2013; 110:10153–10158. [PubMed: 23720318]
7. Peng Y, Gregorich ZR, Valeja SG, Zhang H, Cai W, Chen YC, Guner H, Chen AJ, Schwahn DJ, Hacker TA, Liu X, Ge Y. Top-down proteomics reveals concerted reductions in myofilament and Z-disc protein phosphorylation after acute myocardial infarction. Mol Cell Proteomics. 2014; 13:2752–2764. [PubMed: 24969035]
8. Smith LM, Kelleher NL. Consortium for Top Down P. Proteoform: a single term describing protein complexity. Nat Methods. 2013; 10:186–187. [PubMed: 23443629]

9. Loo JA, Edmonds CG, Smith RD. Primary sequence information from intact proteins by electrospray ionization tandem mass spectrometry. Science. 1990; 248:201–204. [PubMed: 2326633]
10. Sickmann A, Reinders J, Wagner Y, Joppich C, Zahedi R, Meyer HE, Schonfisch B, Perschil I, Chacinska A, Guiard B, Rehling P, Pfanner N, Meisinger C. The proteome of Saccharomyces cerevisiae mitochondria. Proc Natl Acad Sci USA. 2003; 100:13207–13212. [PubMed: 14576278]
11. Gygi SP, Rist B, Gerber SA, Turecek F, Gelb MH, Aebersold R. Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. Nat Biotechnol. 1999; 17:994–999. [PubMed: 10504701]
12. Haverland NA, Fox HS, Ciborowski P. Quantitative proteomics by SWATH-MS reveals altered expression of nucleic acid binding and regulatory proteins in HIV-1-infected macrophages. J Proteome Res. 2014; 13:2109–2119. [PubMed: 24564501]
13. Xie Y, Zhang J, Yin S, Loo JA. Top-down ESI-ECD-FT-ICR mass spectrometry localizes noncovalent protein-ligand binding sites. J Am Chem Soc. 2006; 128:14432–14433. [PubMed: 17090006]
14. Schennach M, Schneeberger EM, Breuker K. Unfolding and Folding of the Three-Helix Bundle Protein KIX in the Absence of Solvent. J Am Soc Mass Spectrom. 2016; 27:1079–1088. [PubMed: 26936183]
15. Sleno L, Volmer DA. Ion activation methods for tandem mass spectrometry. J Mass Spectrom. 2004; 39:1091–1112. [PubMed: 15481084]
16. Brodbelt JS. Ion Activation Methods for Peptides and Proteins. Anal Chem. 2016; 88:30–51. [PubMed: 26630359]
17. Roepstorff P, Fohlman J. Proposal for a common nomenclature for sequence ions in mass spectra of peptides. Biomed Mass Spectrom. 1984; 11:601. [PubMed: 6525415]
18. Frese CK, Altelaar AF, Hennrich ML, Nolting D, Zeller M, Griep-Raming J, Heck AJ, Mohammed S. Improved peptide identification by targeted fragmentation using CID, HCD and ETD on an LTQ-Orbitrap Velos. J Proteome Res. 2011; 10:2377–2388. [PubMed: 21413819]
19. Xia Y, Liang X, McLuckey SA. Ion trap versus low-energy beam-type collision-induced dissociation of protonated ubiquitin ions. Anal Chem. 2006; 78:1218–1227. [PubMed: 16478115]
20. Zubarev RA, Kelleher NL, McLafferty FW. Electron Capture Dissociation of Multiply Charged Protein Cations. A Nonergodic Process J Am Chem Soc. 1998; 120:3265–3266.
21. Syka JE, Coon JJ, Schroeder MJ, Shabanowitz J, Hunt DF. Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. Proc Natl Acad Sci USA. 2004; 101:9528–9533. [PubMed: 15210983]
22. Shaw JB, Li W, Holden DD, Zhang Y, Griep-Raming J, Fellers RT, Early BP, Thomas PM, Kelleher NL, Brodbelt JS. Complete protein characterization using top-down mass spectrometry and ultraviolet photodissociation. J Am Chem Soc. 2013; 135:12646–12651. [PubMed: 23697802]
23. Dongré AR, Jones JL, Somogyi Á, Wysocki VH. Influence of Peptide Composition, Gas- Phase Basicity, and Chemical Modification on Fragmentation Efficiency: Evidence for the Mobile Proton Model. J Am Chem Soc. 1996; 118:8365–8374.
24. Wysocki VH, Tsaprailis G, Smith LL, Breci LA. Mobile and localized protons: a framework for understanding peptide dissociation. J Mass Spectrom. 2000; 35:1399–1406. [PubMed: 11180630]
25. Paizs B, Suhai S. Fragmentation pathways of protonated peptides. Mass Spectrom Rev. 2005; 24:508–548. [PubMed: 15389847]
26. Breci LA, Tabb DL, Yates JR, Wysocki VH. Cleavage N-Terminal to Proline: Analysis of a Database of Peptide Tandem Mass Spectra. Anal Chem. 2003; 75:1963–1971. [PubMed: 12720328]
27. Gu C, Tsaprailis G, Breci L, Wysocki VH. Selective Gas-Phase Cleavage at the Peptide Bond C-Terminal to Aspartic Acid in Fixed-Charge Derivatives of Asp-Containing Peptides. Anal Chem. 2000; 72:5804–5813. [PubMed: 11128940]
28. Gu C, Somogyi Á, Wysocki VH, Medzihradszky KF. Fragmentation of protonated oligopeptides XLDVLQ (X=L, H, K or R) by surface induced dissociation: additional evidence for the ‘mobile proton’ model. Anal Chim Acta. 1999; 397:247–256.

29. Summerfield SG, Gaskell SJ. Fragmentation efficiencies of peptide ions following low energy collisional activation. Int J Mass Spectrom Ion Processes. 1997; 165–166:509–521.
30. Tsaprailis G, Nair H, Somogyi Á, Wysocki VH, Zhong W, Futrell JH, Summerfield SG, Gaskell SJ. Influence of Secondary Structure on the Fragmentation of Protonated Peptides. J Am Chem Soc. 1999; 121:5142–5154.
31. Huang Y, Triscari JM, Tseng GC, Pasa-Tolic L, Lipton MS, Smith RD, Wysocki VH. Statistical Characterization of the Charge State and Residue Dependence of Low-Energy CID Peptide Dissociation Patterns. Anal Chem. 2005; 77:5800–5813. [PubMed: 16159109]
32. Huang Y, Tseng GC, Yuan S, Pasa-Tolic L, Lipton MS, Smith RD, Wysocki VH. A Data-Mining Scheme for Identifying Peptide Structural Motifs Responsible for Different MS/MS Fragmentation Intensity Patterns. J Proteome Res. 2008; 7:70–79. [PubMed: 18052120]
33. Bojesen G. The order of proton affinities of the 20 common L-alpha.-amino acids. J Am Chem Soc. 1987; 109:5557–5558.
34. Tsaprailis G, Somogyi Á, Nikolaev EN, Wysocki VH. Refining the model for selective cleavage at acidic residues in arginine-containing protonated peptides2. Int J Mass Spectrom. 2000; 195–196:467–479.
35. Cobb JS, Easterling ML, Agar JN. Structural Characterization of Intact Proteins Is Enhanced by Prevalent Fragmentation Pathways Rarely Observed for Peptides. J Am Soc Mass Spectrom. 2010; 21:949–959. [PubMed: 20303285]
36. Chowdhury SK, Katta V, Chait BT. Probing conformational changes in proteins by mass spectrometry. J Am Chem Soc. 1990; 112:9012–9013.
37. Loo JA, Loo RRO, Udseth HR, Edmonds CG, Smith RD. Solvent-induced conformational changes of polypeptides probed by electrospray-ionization mass spectrometry. Rapid Commun Mass Spectrom. 1991; 5:101–105. [PubMed: 1666527]
38. Katta V, Chait BT, Carr S. Conformational changes in proteins probed by hydrogen-exchange electrospray-ionization mass spectrometry. Rapid Commun Mass Spectrom. 1991; 5:214–217. [PubMed: 1666528]
39. Meng F, Cargile BJ, Miller LM, Forbes AJ, Johnson JR, Kelleher NL. Informatics and multiplexing of intact protein identification in bacteria and the archaea. Nat Biotechnol. 2001; 19:952–957. [PubMed: 11581661]
40. LeDuc RD, Fellers RT, Early BP, Greer JB, Thomas PM, Kelleher NL. The C-score: a Bayesian framework to sharply improve proteoform scoring in high-throughput top down proteomics. J Proteome Res. 2014; 13:3231–3240. [PubMed: 24922115]
41. Siuzdak G, Bothner B, Yeager M, Brugidou C, Fauquet CM, Hoey K, Change CM. Mass spectrometry and viral analysis. Chem Biol. 1996; 3:45–48. [PubMed: 8807827]
42. Robinson CV, Grosz M, Eyles SJ, Ewbank JJ, Mayhew M, Hartl FU, Dobson CM, Radford SE. Conformation of GroEL-bound [alpha]-lactalbumin probed by mass spectrometry. Nature. 1994; 372:646–651. [PubMed: 7990955]
43. Rostom AA, Sunde M, Richardson SJ, Schreiber G, Jarvis S, Bateman R, Dobson CM, Robinson CV. Dissection of multi-protein complexes using mass spectrometry: Subunit interactions in transthyretin and retinol-binding protein complexes. Proteins. 1998; 33:3–11.
44. Gervasoni P, Staudenmann W, James P, Gehrig P, Plückthun A. beta-Lactamase binds to GroEL in a conformation highly protected against hydrogen/deuterium exchange. Proc Natl Acad Sci USA. 1996; 93:12189–12194. [PubMed: 8901555]
45. Ganem B, Li YT, Henion JD. Detection of noncovalent receptor-ligand complexes by mass spectrometry. J Am Chem Soc. 1991; 113:6294–6296.
46. Light-Wahl KJ, Schwartz BL, Smith RD. Observation of the Noncovalent Quaternary Associations of Proteins by Electrospray Ionization Mass Spectrometry. J Am Chem Soc. 1994; 116:5271–5278.
47. Laganowsky A, Reading E, Allison TM, Ulmschneider MB, Degiacomi MT, Baldwin AJ, Robinson CV. Membrane proteins bind lipids selectively to modulate their structure and function. Nature. 2014; 510:172–175. [PubMed: 24899312]

48. Gologan B, Takáts Z, Alvarez J, Wiseman JM, Talaty N, Ouyang Z, Cooks RG. Ion soft-landing into liquids: Protein identification, separation, and purification with retention of biological activity. J Am Soc Mass Spectrom. 2004; 15:1874–1884. [PubMed: 15589764]
49. Winston RL, Fitzgerald MC. Mass spectrometry as a readout of protein structure and function. Mass Spectrom Rev. 1997; 16:165–179. [PubMed: 9449832]
50. Sharon M, Robinson CV. The Role of Mass Spectrometry in Structure Elucidation of Dynamic Protein Complexes. Annu Rev Biochem. 2007; 76:167–193. [PubMed: 17328674]
51. Breuker K, McLafferty FW. Stepwise evolution of protein native structure with electrospray into the gas phase, 10–12 to 102 s. Proc Natl Acad Sci USA. 2008; 105:18145–18152. [PubMed: 19033474]
52. Compton PD, Zamdborg L, Thomas PM, Kelleher NL. On the scalability and requirements of whole protein mass spectrometry. Anal Chem. 2011; 83:6868–6874. [PubMed: 21744800]
53. Muneeruddin K, Thomas JJ, Salinas PA, Kaltashov IA. Characterization of small protein aggregates and oligomers using size exclusion chromatography with online detection by native electrospray ionization mass spectrometry. Anal Chem. 2014; 86:10692–10699. [PubMed: 25310183]
54. Muneeruddin K, Nazzaro M, Kaltashov IA. Characterization of intact protein conjugates and biopharmaceuticals using ion-exchange chromatography with online detection by native electrospray ionization mass spectrometry and top-down tandem mass spectrometry. Anal Chem. 2015; 87:10138–10145. [PubMed: 26360183]
55. Chen B, Peng Y, Valeja SG, Xiu L, Alpert AJ, Ge Y. Online Hydrophobic Interaction Chromatography-Mass Spectrometry for Top-Down Proteomics. Anal Chem. 2016; 88:1885–1891. [PubMed: 26729044]
56. Reid GE, Wu J, Chrisman PA, Wells JM, McLuckey SA. Charge-State-Dependent Sequence Analysis of Protonated Ubiquitin Ions via Ion Trap Tandem Mass Spectrometry. Anal Chem. 2001; 73:3274–3281. [PubMed: 11476225]
57. Skinner OS, Do Vale LH, Catherman AD, Havugimana PC, de Sousa MV, Compton PD, Kelleher NL. Native GELFrEE: a new separation technique for biomolecular assemblies. Anal Chem. 2015; 87:3032–3038. [PubMed: 25664979]
58. Skinner OS, Havugimana PC, Haverland NA, Fornelli L, Early BP, Greer JB, Fellers RT, Durbin KR, Do Vale LH, Melani RD, Seckler HS, Nelp MT, Belov ME, Horning SR, Makarov AA, LeDuc RD, Bandarian V, Compton PD, Kelleher NL. An informatic framework for decoding protein complexes by top-down mass spectrometry. Nat Methods. 2016; 13:237–240. [PubMed: 26780093]
59. Wojcik R, Dada OO, Sadilek M, Dovichi NJ. Simplified capillary electrophoresis nanospray sheath-flow interface for high efficiency and sensitive peptide analysis. Rapid Commun Mass Spectrom. 2010; 24:2554–2560. [PubMed: 20740530]
60. Belov ME, Damoc E, Denisov E, Compton PD, Horning S, Makarov AA, Kelleher NL. From protein complexes to subunit backbone fragments: a multi-stage approach to native mass spectrometry. Anal Chem. 2013; 85:11163–11173. [PubMed: 24237199]
61. Olsen JV, Macek B, Lange O, Makarov A, Horning S, Mann M. Higher-energy C-trap dissociation for peptide modification analysis. Nat Methods. 2007; 4:709–712. [PubMed: 17721543]
62. Fellers RT, Greer JB, Early BP, Yu X, LeDuc RD, Kelleher NL, Thomas PM. ProSight Lite: graphical software to analyze top-down mass spectrometry data. Proteomics. 2015; 15:1235–1238. [PubMed: 25828799]
63. Schaaff TG, Cargile BJ, Stephenson JL, McLuckey SA. Ion Trap Collisional Activation of the (M + 2H)2+ (M + 17H)17+ Ions of Human Hemoglobin β-Chain. Anal Chem. 2000; 72:899–907. [PubMed: 10739190]
64. Compton PD, Fornelli L, Kelleher NL, Skinner OS. Probing asymmetric charge partitioning of protein oligomers during tandem mass spectrometry. Int J Mass Spectrom. 2015; 390:132–136. [PubMed: 26692813]
65. Schwartz BL, Bruce JE, Anderson GA, Hofstadler SA, Rockwood AL, Smith RD, Chilkoti A, Stayton PS. Dissociation of tetrameric ions of noncovalent streptavidin complexes formed by electrospray ionization. J Am Soc Mass Spectrom. 1995; 6:459–465. [PubMed: 24214298]

66. Jurchen JC, Williams ER. Origin of Asymmetric Charge Partitioning in the Dissociation of Gas-Phase Protein Homodimers. J Am Chem Soc. 2003; 125:2817–2826. [PubMed: 12603172]

67. Jurchen JC, Garcia DE, Williams ER. Further studies on the origins of asymmetric charge partitioning in protein homodimers. J Am Soc Mass Spectrom. 2004; 15:1408–1415. [PubMed: 15465353]

68. Loo RRO, Loo JA. Salt Bridge Rearrangement (SaBRe) Explains the Dissociation Behavior of Noncovalent Complexes. J Am Soc Mass Spectrom. 2016; 27:975–990. [PubMed: 27052739]

69. Newton KA, Pitteri SJ, Laskowski M, McLuckey SA. Effects of Single Amino Acid Substitution on the Collision-Induced Dissociation of Intact Protein Ions: Turkey Ovomucoid Third Domain. J Proteome Res. 2004; 3:1033–1041. [PubMed: 15473693]

70. Engel BJ, Pan P, Reid GE, Wells JM, McLuckey SA. Charge state dependent fragmentation of gaseous protein ions in a quadrupole ion trap: bovine ferri-, ferro-, and apo-cytochrome c. Int J Mass Spectrom. 2002; 219:171–187.

71. Hogan JM, McLuckey SA. Charge state dependent collision-induced dissociation of native and reduced porcine elastase. J Mass Spectrom. 2003; 38:245–256. [PubMed: 12644985]

72. Pitteri SJ, Reid GE, McLuckey SA. Affecting Proton Mobility in Activated Peptide and Whole Protein Ions via Lysine Guanidination. J Proteome Res. 2004; 3:46–54. [PubMed: 14998162]

73. Durbin KR, Skinner OS, Fellers RT, Kelleher NL. Analyzing Internal Fragmentation of Electrosprayed Ubiquitin Ions During Beam-Type Collisional Dissociation. J Am Soc Mass Spectrom. 2015; 26:782–787. [PubMed: 25716753]

74. Wilm M. Principles of Electrospray Ionization. Mol Cell Proteomics. 2011; 10

75. Fernandez de la Mora J. Electrospray ionization of large multiply charged species proceeds via Dole's charged residue mechanism. Anal Chim Acta. 2000; 406:93–104.

76. Smith RD, Loo JA, Loo RRO, Busman M, Udseth HR. Principles and practice of electrospray ionization—mass spectrometry for large polypeptides and proteins. Mass Spectrom Rev. 1991; 10:359–452.

77. Loo JA, Udseth HR, Smith RD. Peptide and protein analysis by electrospray ionization-mass spectrometry and capillary electrophoresis-mass spectrometry. Anal Biochem. 1989; 179:404–412. [PubMed: 2774189]

78. Tolid LP, Anderson GA, Smith RD, Brothers HM Ii, Spindler R, Tomalia DA. Electrospray ionization Fourier transform ion cyclotron resonance mass spectrometric characterization of high molecular mass Starburst dendrimers. Int J Mass Spectrom Ion Processes. 1997; 165–166:405–418.

79. Loo RRO, Lakshmanan R, Loo JA. What Protein Charging (and Supercharging) Reveal about the Mechanism of Electrospray Ionization. J Am Soc Mass Spectrom. 2014; 25:1675–1693. [PubMed: 25135609]

80. Kruger NA, Zubarev RA, Carpenter BK, Kelleher NL, Horn DM, McLafferty FW. Electron capture versus energetic dissociation of protein ions. Int J Mass Spectrom. 1999; 182–183:1–5.

81. Uversky VN, Gillespie JR, Fink AL. Why are "natively unfolded" proteins unstructured under physiologic conditions? Proteins. 2000; 41:415–427. [PubMed: 11025552]

82. Tompa P. Intrinsically unstructured proteins. Trends Biochem Sci. 2002; 27:527–533. [PubMed: 12368089]

83. Uversky VN. Natively unfolded proteins: A point where biology waits for physics. Protein Sci. 2002; 11:739–756. [PubMed: 11910019]

84. Frantz C, Barreiro G, Dominguez L, Chen X, Eddy R, Condeelis J, Kelly MJS, Jacobson MP, Barber DL. Cofilin is a pH sensor for actin free barbed end formation: role of phosphoinositide binding. J Cell Biol. 2008; 183:865. [PubMed: 19029335]

85. Carbeck JD, Severs JC, Gao J, Wu Q, Smith RD, Whitesides GM. Correlation between the Charge of Proteins in Solution and in the Gas Phase Investigated by Protein Charge Ladders, Capillary Electrophoresis, and Electrospray Ionization Mass Spectrometry. J Phys Chem B. 1998; 102:10596–10601.

86. Loo JA, Edmonds CG, Smith RD. Tandem mass spectrometry of very large molecules. 2 Dissociation of multiply charged proline-containing proteins from electrospray ionization. Anal Chem. 1993; 65:425–438. [PubMed: 8382455]

87. Vaisar T, Urban J. Probing Proline Effect in CID of Protonated Peptides. J Mass Spectrom. 1996; 31:1185–1187. [PubMed: 8916427]

88. Grizot S, Fauré J, Fieschi F, Vignais PV, Dagher MC, Pebay-Peyroula E. Crystal Structure of the Rac1 RhoGDI Complex Involved in NADPH Oxidase Activation. Biochemistry. 2001; 40:10007–10013. [PubMed: 11513578]

89. Ruiz Carrillo D, Chandrasekaran R, Nilsson M, Cornvik T, Liew CW, Tan SM, Lescar J. Structure of human Rack1 protein at a resolution of 2.45 Å. Acta Crystallogr Sect F: Struct Biol Cryst Commun. 2012; 68:867–872.

90. Chen L, Yang S, Jakoncic J, Zhang JJ, Huang XY. Migrastatin analogues target fascin to block tumour metastasis. Nature. 2010; 464:1062–1066. [PubMed: 20393565]

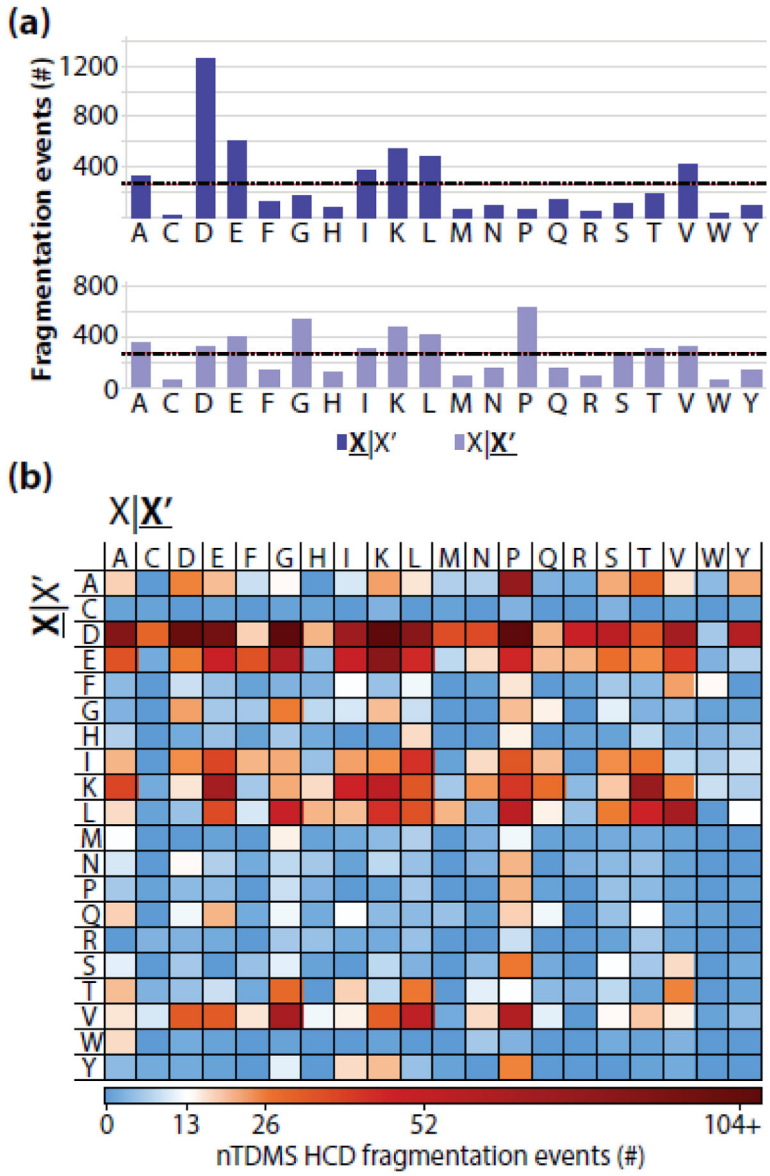


Figure 1.

(a) The distribution of assigned residue fragmentation events in nTDMS. The total number of events that were assigned for each residue N-terminal to fragmentation (top) or C-terminal to fragmentation (bottom). The expected number of assigned fragmentation events assuming a uniform fragmentation propensity across all 20 residues is shown as a dashed black line (n = 266 events). **(b)** Assigned fragmentation events deconstructed by residue pair. The expected number of fragmentation events assuming uniform propensities across all 400 possible pairs is shown in white (n = 13 events). For all panels, **X**|X' refers to fragmentation occurring C-terminal to the amino acid residue whereas X|**X'** refers to fragmentation occurring N-terminal to the amino acid residue.

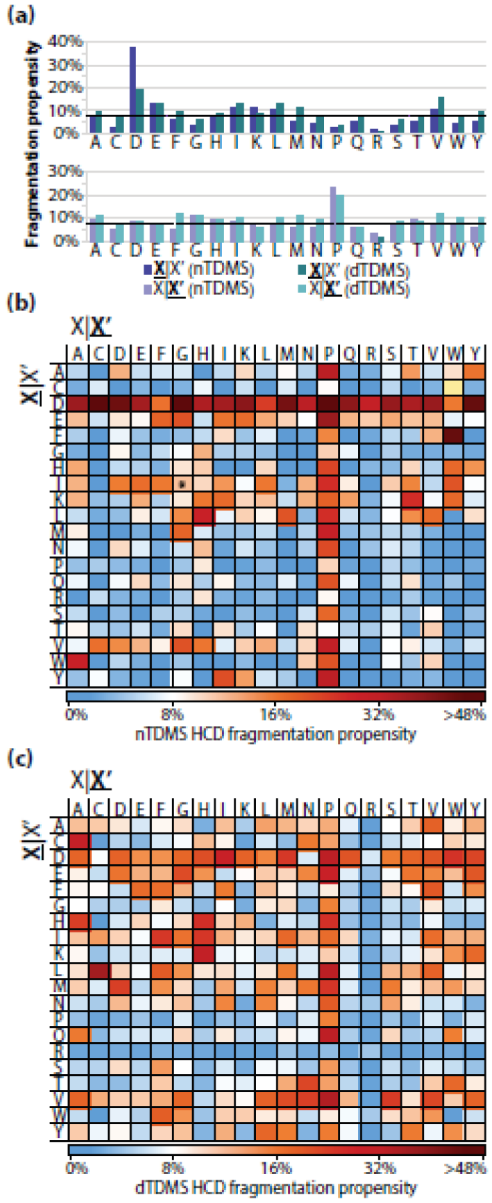


Figure 2.

(a) The HCD fragmentation propensities observed for the nTDMS dataset (purple) and dTDMS (turquoise) from Catherman, *et al.* [4]. The dashed black line at 8% represents the average fragmentation propensity across all residues from the nTDMS dataset. The top graph includes the fragmentation propensities for events occurring C-terminal to a given residue, whereas the bottom graph includes the fragment propensities for events occurring N-terminal to a given residue. **(b)** Residue fragmentation propensities for the nTDMS dataset. The asterisk on the cysteine|tryptophan pair indicates that no possible fragmentation event existed for that pair within the dataset. **(c)** Residue fragmentation propensities for the dTDMS dataset. For all panels, **X**|X' refers to fragmentation occurring C-terminal to the amino acid residue whereas X|**X'** refers to fragmentation occurring N-terminal to the amino acid residue.

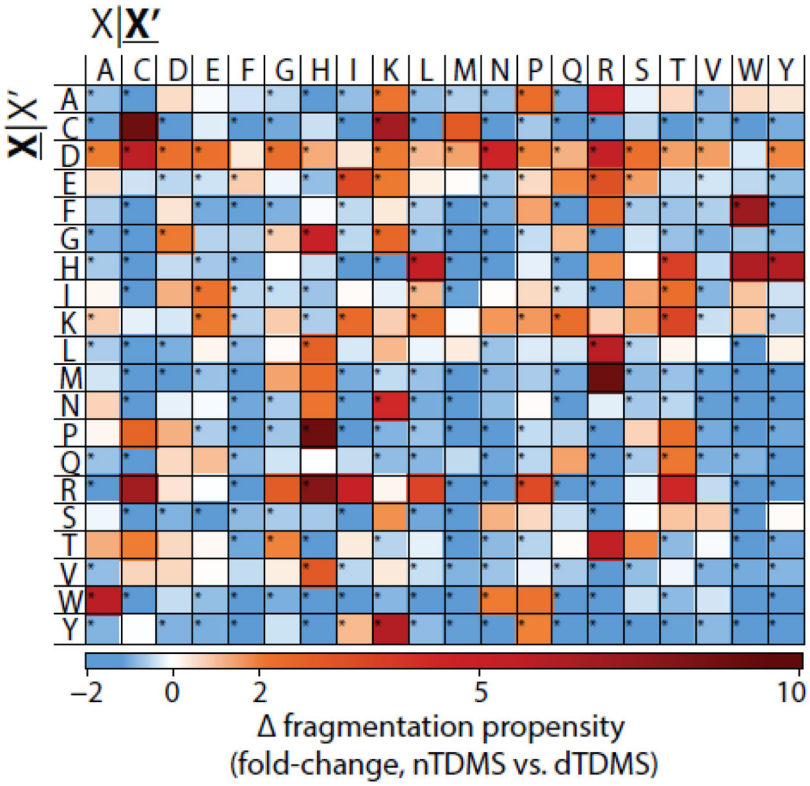


Figure 3.

nTDMS has a fewer number of highly preferred fragmentation pathways as compared to dTDMS. The fold-change in residue fragmentation propensity for nTDMS as compared to dTDMS. Blue indicates a decrease in fragmentation propensity in nTDMS as compared to dTDMS whereas red indicates an increase for nTDMS. Significant differences (p 0.05) are indicated with an asterisk. The yellow box for the cysteine|tryptophan pair indicates that no possible fragmentation event existed for the residue pair within the nTDMS dataset and, as such, a comparison cannot be made. **X**|X' denotes fragmentation occurring C-terminal to the amino acid residue whereas X|**X'** refers to fragmentation occurring N-terminal to the amino acid residue.

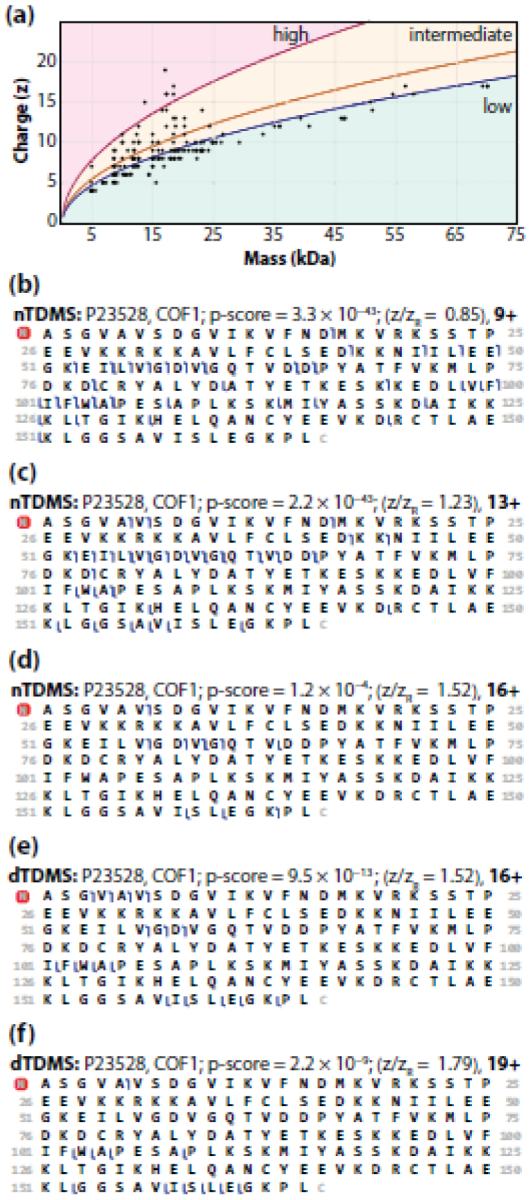


Figure 4.

(a) A relationship between charge and precursor mass. Precursor ions were selected for fragmentation without consideration for charge state designation, which resulted in the inclusion of fragmentation data from precursors with high (pink) and intermediate (yellow) charge state designations. The yellow curve ($y = 0.0778m^{½}$) represents the Rayleigh charge limit for each precursor, z_R (see text and [75]). The pink curve ($y = 0.1081m^{½}$) is the lower bound for the high charge state designation [56], which was determined by calculating the theoretical z_{High} ($z_{High} = z_R \times 1.43$) for each protein in the dataset. The blue curve ($y = 0.0667m^{½}$) is the upper bound for the low charge state designation [56], which was determined by calculating the theoretical z_{Low} ($z_{Low} = z_R \times 0.86$) for each protein in the dataset. The graphical fragmentation maps for COF1 obtained by nTDMS of a “low” **(b)**, “intermediate” **(c),** and “high” **(d)** charge state precursor ion as compared to dTDMS [4] of

an “intermediate” **(e)** and “high” **(f)** charge state precursor ion. For panels b–f, blue flags represent matched fragment ions with mass tolerance of 15 ppm.

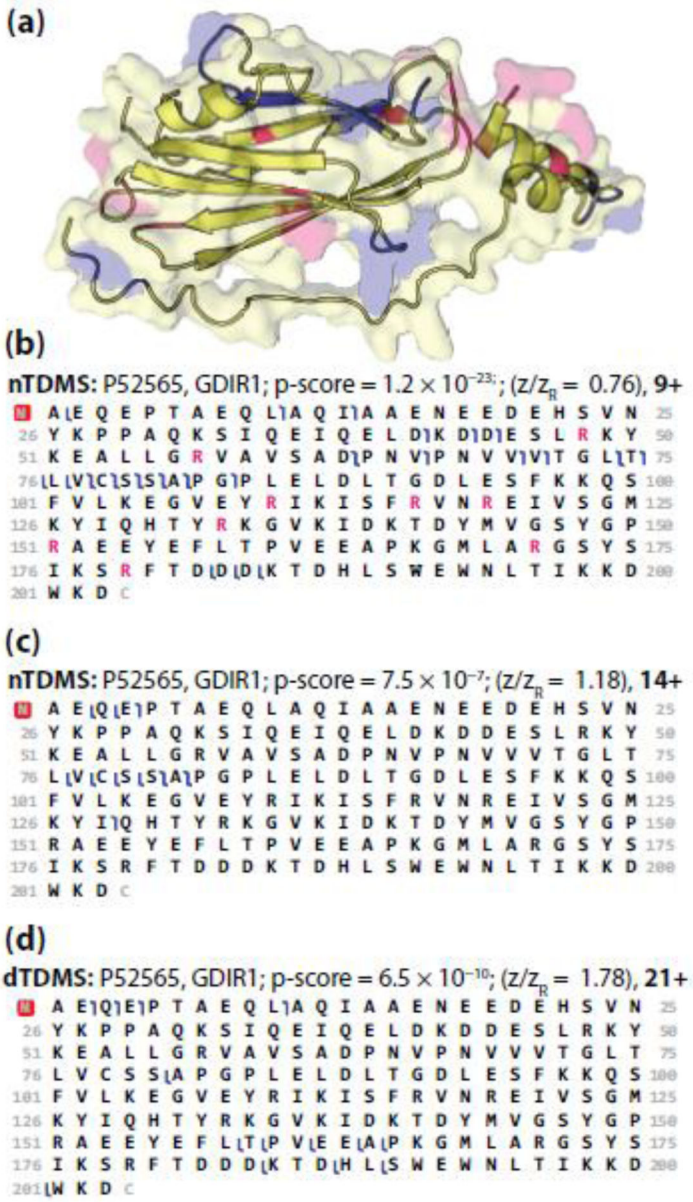


Figure 5.

(a) The surface and cartoon models of GDIR1 (PDB ID: 1hh4 [88]). Pink represents arginine residues and blue highlights the residues involved in fragmentation of the 9+ precursor ion (panel b). The two disordered regions, one at the N-terminus [A2–A7] and the other near the flexible arm [V59–P65], are not included in the crystal structure, and thus not shown. The fragmentation maps of GDIR1 obtained using nTDMS for a low **(b)** and intermediate **(c)** charge state precursor ion as compared to fragmentation of a high **(d)** charge state precursor ion obtained using dTDMS [4]. For panels b–d, blue flags represent matched fragment ions with mass tolerance of 15 ppm. In panel b, arginine residues are colored pink to match the arginine residues highlighted in panel a.

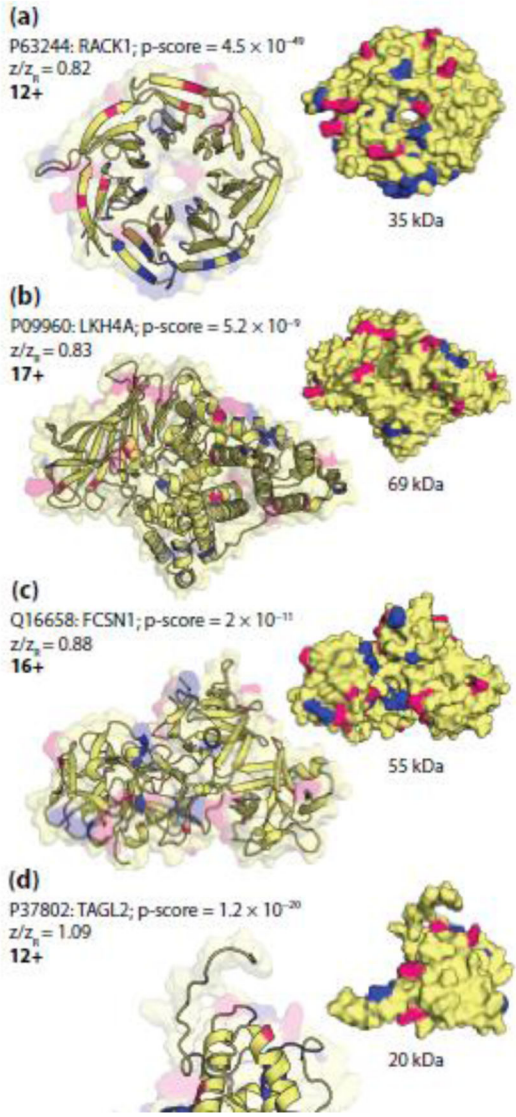


Figure 6.

The cartoon (left) and surface (right) models of **(a)** RACK1 (PDB ID: 4aow [89]), **(b)** LKH4A (PDB ID: 4rvb), **(c)** TAGL2 (PDB ID: 1wym), and **(d)** FCSN1 (PDB ID: 3llp [90]). Pink represents arginine residues and blue highlights the residues involved in fragmentation. Panels **(a)** and **(b)** represent examples of precursor ions with low charge state designations (<0.86) whereas panels **(c)** and **(d)** represent examples of precursor ions with intermediate charge state designations ($0.86 < z/z_R < 1.43$).

Table 1

The nTDMS fragmentation propensities ordered from the most frequently cleaved residue pairs (top) to the least frequently cleaved residues pairs (bottom). Residue pairs with fragmentation propensities >20% and number of assigned fragmentation events >13 are shown above the gray bar. In contrast, residue pairs with lower than expected fragmentation propensities (<5%) and >275 possible fragmentation events are shown below the gray bar.

Residue pair	Residue N-terminal to fragmentation	Residue C-terminal to fragmentation	Assigned fragmentation events	Possible fragmentation events	Fragmentation propensity
D|P	Aspartic Acid	Proline	121	190	64%
D|C	Aspartic Acid	Cysteine	32	66	49%
D|G	Aspartic Acid	Glycine	171	354	48%
D|Y	Aspartic Acid	Tyrosine	59	126	47%
F|W	Phenylalanine	Tryptophan	14	30	47%
D|D	Aspartic Acid	Aspartic Acid	100	220	46%
D|M	Aspartic Acid	Methionine	37	84	44%
D|K	Aspartic Acid	Lysine	124	304	41%
D|Q	Aspartic Acid	Glutamine	21	52	40%
D|S	Aspartic Acid	Serine	56	144	39%
D|V	Aspartic Acid	Valine	68	176	39%
E|P	Glutamic Acid	Proline	46	120	38%
D|A	Aspartic Acid	Alanine	85	224	38%
D|I	Aspartic Acid	Isoleucine	71	194	37%
D|E	Aspartic Acid	Glutamic Acid	95	260	37%
D|T	Aspartic Acid	Threonine	34	96	35%
D|H	Aspartic Acid	Histidine	21	60	35%
A|P	Alanine	Proline	77	234	33%
D|N	Aspartic Acid	Asparagine	38	116	33%
Y|P	Tyrosine	Proline	26	80	33%
V|P	Valine	Proline	60	192	31%
D|R	Aspartic Acid	Arginine	48	156	31%
W|A	Tryptophan	Alanine	17	58	29%
L|H	Leucine	Histidine	21	72	29%

J Am Soc Mass Spectrom. Author manuscript; available in PMC 2018 June 01.

Residue pair	Residue N-terminal to fragmentation	Residue C-terminal to fragmentation	Assigned fragmentation events	Possible fragmentation events	Fragmentation propensity
K|T	Lysine	Threonine	75	282	27%
I|P	Isoleucine	Proline	35	132	27%
L|P	Leucine	Proline	56	212	26%
D|L	Aspartic Acid	Leucine	83	336	25%
H|P	Histidine	Proline	15	64	23%
N|P	Asparagine	Proline	21	92	23%
Y|I	Tyrosine	Isoleucine	17	76	22%
V|G	Valine	Glycine	66	302	22%
E|G	Glutamic Acid	Glycine	63	290	22%
Q|P	Glutamine	Proline	18	84	21%
L|M	Leucine	Methionine	21	98	21%
A|A	Alanine	Alanine	18	392	5%
A|L	Alanine	Leucine	16	358	5%
K|D	Lysine	Aspartic Acid	16	388	4%
G|K	Glycine	Lysine	20	498	4%
G|L	Glycine	Leucine	9	332	3%
S|K	Serine	Lysine	8	304	3%