

Statistical Characterization of the Charge State and Residue Dependence of Low-Energy CID Peptide Dissociation Patterns

Yingying Huang,[†] Joseph M. Triscari,[‡] George C. Tseng,[§] Ljiljana Pasa-Tolic,^{||} Mary S. Lipton,^{||} Richard D. Smith,^{||} and Vicki H. Wysocki^{*†}

Department of Chemistry, University of Arizona, Tucson, Arizona 85721, Science Application International Corporation, Tucson, Arizona 85711, Department of Biostatistics and Department of Human Genetics, University of Pittsburgh, Pittsburgh, Pennsylvania 15261, and Biological Sciences Division, Pacific Northwest National Laboratory, Richland, Washington 99352

Data mining was performed on 28 330 unique peptide tandem mass spectra for which sequences were assigned with high confidence. By dividing the spectra into different sets based on structural features and charge states of the corresponding peptides, chemical interactions involved in promoting specific cleavage patterns in gas-phase peptides were characterized. Pairwise fragmentation maps describing cleavages at all Xxx–Zzz residue combinations for b and y ions reveal that the difference in basicity between Arg and Lys results in different dissociation patterns for singly charged Arg- and Lys-ending tryptic peptides. While one dominant protonation form (proton localized) exists for Arg-ending peptides, a heterogeneous population of different protonated forms or more facile interconversion of protonated forms (proton partially mobile) exists for Lys-ending peptides. Cleavage C-terminal to acidic residues dominates spectra from singly charged peptides that have a localized proton and cleavage N-terminal to Pro dominates those that have a mobile or partially mobile proton. When Pro is absent from peptides that have a mobile or partially mobile proton, cleavage at each peptide bond becomes much more prominent. Whether the above patterns can be found in b ions, y ions, or both depends on the location of the proton holder(s) in multiply protonated peptides. Enhanced cleavages C-terminal to branched aliphatic residues (Ile, Val, Leu) are observed in both b and y ions from peptides that have a mobile proton, as well as in y ions from peptides that have a partially mobile proton; enhanced cleavages N-terminal to these residues are observed in b ions from peptides that have a partially mobile proton. Statistical tools have been designed to visualize the fragmentation maps and measure the similarity between them. The pairwise cleavage patterns observed expand our knowledge of peptide gas-phase fragmentation behaviors and

may be useful in algorithm development that employs improved models to predict fragment ion intensities.

Low-energy collision-induced dissociation (CID) in mass spectrometry has been used extensively in peptide sequencing for protein identification and analysis of protein modifications.^{1,2} Computer algorithms are employed to automate the spectral analysis and assign peptide sequences to the fragmentation spectra.^{2–4} Numerous peptide fragmentation spectra are obtained everyday, but only a low percentage (e.g., 10–35%) of them are successfully identified by readily available algorithms.^{5–8} One explanation for the high failure rate of current algorithms is the very simple fragmentation model on which they are based, which assumes that cleavage will occur mainly, if not exclusively, at the amide bonds between consecutive amino acid residues to produce ions of identical abundance (or intensity), regardless of the identity or the chemical property of the residues.^{4,9} Although intensity patterns of the fragment ions from the same peptide under the same experimental settings are highly reproducible, current readily available algorithms only use the mass-to-charge (m/z) information from a tandem mass (MS/MS) spectrum and ignore the other information available—the relative intensity information. While this approach is sufficient to identify many peptides, many others yield fragmentation patterns that significantly deviate from a uniform fragmentation model.⁸ A more detailed and chemically meaningful fragmentation model might improve algorithm performance and might include using different models for different candidate sequences. In addition, the gas-phase unimolecular

* Corresponding author. E-mail: vwyssocki@email.arizona.edu, Fax: 520-621-8407.

[†] University of Arizona.

[‡] Science Application International Corp.

[§] University of Pittsburgh.

^{||} Pacific Northwest National Laboratory.

(1) Aebersold, R.; Goodlett, D. R. *Chem. Rev.* 2001, 101, 269–295.

(2) Aebersold, R.; Mann, M. *Nature* 2003, 422, 198–207.

(3) Fenyo, D. *Curr. Opin. Biotechnol.* 2000, 11, 391–395.

(4) Eng, J. K.; McCormack, A. L.; Yates, J. R. *J. Am. Soc. Mass Spectrom.* 1994, 5, 976–989.

(5) Resing, K. A.; Meyer-Arendt, K.; Mendoza, A. M.; Aveline-Wolf, L. D.; Jonscher, K. R.; Pierce, K. G.; Old, W. M.; Cheung, H. T.; Russell, S.; Wattawa, J. L.; Goehle, G. R.; Knight, R. D.; Ahn, N. G. *Anal. Chem.* 2004, 76, 3556–3568.

(6) Tabb, D. L.; Smith, L. L.; Breci, L. A.; Wysocki, V. H.; Lin, D.; Yates, J. R. *Anal. Chem.* 2003, 75, 1155–1163.

(7) Elias, J. E.; Gibbons, F. D.; King, O. D.; Roth, F. P.; Gygi, S. P. *Nat. Biotechnol.* 2004, 22, 214–219.

(8) Smith, L. L. Ph.D. Dissertation, University of Arizona, 2003.

(9) Perkins, D. N.; Pappin, D. J. C.; Creasy, D. M.; Cottrell, J. S. *Electrophoresis* 1999, 20, 3551–3567.

dissociation of protonated peptides is of fundamental interest because knowledge of the chemistry increases our understanding of molecular structure and intramolecular proton-transfer dynamics in the absence of solvent.

Previous studies have shown that peptide dissociation behavior in MS/MS spectra is a complex function of many factors including the residue content, charge state, and sequence of the peptide. Traditional studies using only a few model peptides^{10–13} often yield complicated results, and the data are insufficient to elucidate all possible mechanisms. The alternative approach, using data-mining methods to analyze a large set of MS/MS spectra,^{6,7,14–17} requires a large set of high-quality spectra whose sequences are assigned unambiguously. Assembling the required high-quality data set is a major bottleneck. Several earlier attempts used either a high cutoff score in available algorithms^{7,14,15,18} or manual validation combined with a cutoff score^{5,16} to filter the spectra in the databases. Varying numbers of spectra from ~500 to ~5000 were used,^{6,7,14–16,19} and these studies often focus on a single charge state (most often doubly charged). More recently, researchers generated validated data sets²⁰ by combining the analyses of two types of mass spectrometers together: The same peptides were analyzed by FT-ICR and ion trap spectrometers under the same chromatographic conditions. The accurate mass measurements of the precursor ions were obtained from the FT-ICR, while the fragmentation information in the MS/MS spectra from the same peptides were recorded by the ion trap. When the top hit of the sequences assigned by SEQUEST from the MS/MS spectrum correlated (within 1 ppm) with the accurate mass measurement under the same ($\pm 5\%$) retention time, a spectrum was denoted to have a sequence assigned with very high confidence.

The data generated from such an approach should have few incorrectly assigned spectra, because it takes advantage, at a given retention time, of two different mass spectrometry techniques and records different information from the same peptide, with both measurements contributing to the confidence of the sequence assignment. Although a commercial database-searching algorithm was still relied upon in generating peptide sequence assignments, this approach eliminates the requirement of high cutoff scores that results in tremendous loss of information. It also eliminates human errors, which are fairly common in manual validation.²¹

In this study, we use peptide spectra collected from such an approach to catalog fragmentation behavior of peptides that have different structural motifs. Unlike previous studies that focused only on peptides that were doubly charged, mainly because spectra from doubly charged peptides are the most abundant species that can pass through the filters, this study shows pairwise cleavage statistics among peptides with different structural motifs under different charge states. Special attention has been given to the sorting and filtering of the peptides, which allows factors contributing to the fragmentation process to be studied alone or in combination. Statistical tools are developed to visualize and validate the patterns observed in the pairwise fragmentation maps, revealing a detailed description of the fragmentation behavior of different peptides. The fragmentation statistics from this study may prove useful for improving the fragmentation model in sequencing algorithms, which could have an immediate impact on practical proteomics studies. At the same time, the correlation of fragmentation statistics with peptide structural features can help us understand chemical interactions or residue combinations involved in promoting specific cleavage pathways, guide the design of specific chemical mechanism studies using model systems, and thus have a longer term impact by increasing our fundamental knowledge of gas-phase unimolecular dissociation.

EXPERIMENTAL SECTION

Spectra Collecting and Filtering. The experiments to acquire the spectra were described in previous publications.^{17,20} Peptide MS/MS spectra from the proteomes of two organisms, *Shewanella oneidensis* and *Deinococcus radiodurans*, were collected using HPLC with ion trap instruments. The SEQUEST algorithm⁴ was used to assign peptide sequences to these spectra with minimum XCorr score 1.5 for peptides with MW <1000, 2.0 for all other peptides. Using the same chromatographic conditions, accurate masses of the precursor ions detected at the same retention time (\pm) by FT-ICR were used to confirm the assigned sequences. A total of 1137 composite spectra (spectra that show more than one peptide might have been selected for fragmentation simultaneously) were identified and purged from the *Deinococcus* data set. (A spectrum is composite if more than one peptide was assigned to it with high SEQUEST XCorr scores and both their precursor ions were also found in the FT-ICR measurement.) A total of 1974 spectra (730 from *Shewanella* and 1244 from *Deinococcus*) were also purged if their sequences contain ¹⁵N isotope or the mass difference is more than 3 Da between the theoretical mass of the singly charged precursor ion (MH⁺) and the recorded MH⁺ (provided regardless of the actual charge state, in the first line of the *.dta file from the LCQ ion trap MS). After all the filtering steps, 28 330 spectra of unique sequence and charge state (16 008 from *Shewanella* and 12 322 from *Deinococcus*) were used for the analyses in this paper. (A copy of this data set may be requested via e-mail to vwyssocki@email.arizona.edu.)

Computing the Fragmentation Statistics. Computer programs written in PERL (version 5.6.3, <http://www.perl.com>) and MatLab (6.0, release 12, student version, <http://www.mathworks.com>), as well as in SQL under Microsoft SQL-Server 2000 (<http://www.microsoft.com/sql>), were used to extract fragmentation statistics. The 28 330 spectra were sorted by the charge states and the num-

- (10) Tsaprailis, G.; Nair, H.; Somogyi, A.; Wysocki, V. H.; Zhong, W. Q.; Futrell, J. H.; Summerfield, S. G.; Gaskell, S. J. *J. Am. Chem. Soc.* **1999**, *121*, 5142–5154.
- (11) Burllet, O.; Yang, C. Y.; Gaskell, S. J. *J. Am. Soc. Mass Spectrom.* **1992**, *3*, 337–344.
- (12) Dongre, A. R.; Jones, J. L.; Somogyi, A.; Wysocki, V. H. *J. Am. Chem. Soc.* **1996**, *118*, 8365–8374.
- (13) Gu, C. G.; Tsaprailis, G.; Brecci, L.; Wysocki, V. H. *Anal. Chem.* **2000**, *72*, 5804–5813.
- (14) Huang, Y.; Wysocki, V. H.; Tabb, D. L.; Yates, J. R. *Int. J. Mass Spectrom.* **2002**, *219*, 233–244.
- (15) Tabb, D. L.; Huang, Y.; Wysocki, V. H.; Yates, J. R. *Anal. Chem.* **2004**, *76*, 1243–1248.
- (16) Kapp, E. A.; Schutz, F.; Reid, G. E.; Eddes, J. S.; Moritz, R. L.; O'Hair, R. A. J.; Speed, T. P.; Simpson, R. J. *Anal. Chem.* **2003**, *75*, 6251–6264.
- (17) Huang, Y.; Triscari, J. M.; Pasa-Tolic, L.; Anderson, G. A.; Lipton, M. S.; Smith, R. D.; Wysocki, V. H. *J. Am. Chem. Soc.* **2004**, *126*, 3034–3035.
- (18) Tabb, D. L.; Smith, L. L.; Brecci, L. A.; Wysocki, V. H.; Lin, D.; Yates, J. R. I. *Analytical Chemistry* **2003**, *75*, 1155–1163.
- (19) Brecci, L. A.; Tabb, D. L.; Yates, J. R.; Wysocki, V. H. *Anal. Chem.* **2003**, *75*, 1963–1971.
- (20) Smith, R. D.; Anderson, G. A.; Lipton, M. S.; Pasa-Tolic, L.; Shen, Y. F.; Conrads, T. P.; Veenstra, T. D.; Udseth, H. R. *Proteomics* **2002**, *2*, 513–523.

(21) Aebersold, R. IPAM Workshop on Proteomics: Sequence, Structure, Function: UCLA, Los Angeles, CA, Keynote Talk, March 23, 2004.

ber of missed cleavages. Multiple analyses from two categories were then performed. Different sets of spectra and different schemes for peak intensity normalization were used in each category.

In the first category of analyses, summarized in Table 3, the frequencies and intensities of different ion types from peptides of different charge states and basic residue content are characterized and compared. One m/z entry in the *.dta file was considered one peak in the spectrum. Because the monoisotopic peak of the peptide was selected to obtain the MS/MS spectra, isotopic fragment peaks were not commonly observed in our data. Ion types investigated here include several ion types that may be observed in low-energy CID spectra. These include the following: y , $y - H_2O$, $y - NH_3$, b , $b - H_2O$, $b - NH_3$, $b + H_2O$, a , $a - H_2O$, $a - NH_3$, internal, internal- H_2O , internal- NH_3 , internal-CO, $M - H_2O$, $M - NH_3$, and immonium ions. When the charge state of the precursor ion is greater than 1, both singly and doubly charged fragment ions were considered. Two additional ion types, x and c ions, which are not considered to exist typically in low-energy CID spectra, were used as the control to determine the probability of random matching and the noise level. All ions from the above ion types were identified from each spectrum according to the assigned sequence using a mass window of ± 0.5 around the theoretical m/z . Then they were normalized to the total ion abundance in the spectrum. (Due to the low mass cutoff problem intrinsic to ion trap instruments, as well as the instrument detection limit at 2000 m/z , the ions whose theoretical m/z fall into the above mass cutoff ranges were excluded from our analyses.) This normalization scheme allows direct comparisons of the abundances from different ion types and enables other fragmentation statistics be computed, e.g., the percentage of total identifiable ion abundance. With each spectrum, the normalized ion intensities from the same ion type are summed together as the sum of normalized intensities for that ion type in that spectrum. The frequency is calculated as the number of spectra that contain a nonzero sum of normalized intensities for that specific ion type divided by the number of total spectra in that data set. For example, if half the spectra of a set contain at least one b^+ , the frequency for b^+ would be 0.5. A median value is then calculated from those spectra showing a nonzero sum of normalized intensities for that ion type.

In the second category of analyses, which make up the majority of the results discussed, the analyses focus on ion types that are most prominent in low-energy CID spectra and useful in sequencing—cleavage of amide bonds with the retention of charge on either the N-terminal fragment (b ion) or the C-terminal fragment (y ion). A different normalization scheme was used: singly charged b ions (b^+) and y ions (y^+) were normalized to the most abundant peak among all b^+ and y^+ . This normalization separates the intensities of b^+ and y^+ from distortions caused by other fragmentation pathways, e.g., neutral loss or internal fragmentation, as well as eliminating the undesirable disruption from spectra with different signal-to-noise ratios. This normalization also allows direct comparison between b^+ and y^+ relative abundance and quantitative comparison of the lability of a certain amide bond to the rest. The normalized abundances of b^+ and y^+ were cataloged by the pair of the AA residues at the cleavage site as the relative abundances for such a pairwise cleavage. From each spectral set, two fragmentation maps showing the median

relative abundances of bond cleavages at each residue combination, one for each ion type (b^+ and y^+), were then obtained.

Visualization of the Fragmentation Maps. Computer programs written in C# were used to generate the fragmentation maps. A color gradient scheme (shown on the right side of each fragmentation map below) is used to represent the wide variance in cleavage likelihood among all residue combinations, with the darker color corresponding to the more abundant cleavages. (Although the color gradient has been chosen to accommodate black and white printing, the authors recommend viewing these maps in color to avoid missing information.) Pairwise fragmentation maps are obtained for b^+ and y^+ from different sets of spectra. The color of each circle or ellipse in these maps shows the median of a certain $X_{xx}-Z_{zz}$ cleavage. Single letter codes of AA residues listed in the left-most column correspond to the N-terminal residue (X_{xx}) in an $X_{xx}-Z_{zz}$ pair, while those listed along the topmost row correspond to the C-terminal residue (Z_{zz}) in an $X_{xx}-Z_{zz}$ pair. Cysteine (Cys or C) and tryptophan (Trp or W) are the two residues that have the lowest occurrences among all residues because Cys residues were not reduced and alkylated, while Trp residues are prone to oxidation. Therefore, the statistics for these two residues are often missing in the fragmentation maps. The horizontal dimension of each ellipse is proportional to the count (number of the particular $X_{xx}-Z_{zz}$ residue combination in that data set) so the visualization impact of a lower-count AA pair is reduced. The minimum count and the count at 90% quantile are illustrated under the color bar.

Fisher Information (FI) Threshold. A statistical threshold called the FI threshold is employed to determine whether the cleavage abundance information for a particular residue combination should be shown when the count of such a residue combination is less than 10. FI describes the amount of information about an unobserved parameter that the observed data carry.²² The FI threshold was designed to validate the distribution of the cleavage abundance when the occurrence is low. Denote data $X = (X_1, \dots, X_n)$ as the fragmentation intensities of a residue combination. FI is defined as

$$FI(X) = \frac{n}{\sigma^2} = \frac{n \times (n - 1)}{\sum (X_i - \bar{X})^2}$$

where n is the count and σ^2 is the sample variance. A larger n and a smaller σ^2 correspond to larger FI. We further denote $X^{(i)} = (X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_n)$ the leave-one-out data and $F\tilde{I}(X) = \max(FI(X^{(1)}), \dots, FI(X^{(n)}))$ the Fisher information when the greatest outlying observation is treated as an outlier and detected. Therefore, $F\tilde{I}(X) > > FI(X)$ implies that the distribution has a significant outlier. Finally we include the residue combinations in the fragmentation map if they satisfy *all* of the following: (1) $n \geq 3$. (2) When $3 \leq n \leq 9$, $FI(X)$ is larger than or equal to 500. (3) When $6 \leq n \leq 9$ and $FI(X) < 500$, $F\tilde{I}(X) > 1000$.

Note that rule 3 is introduced as a conservative procedure to recover distributions that should have large Fisher information but were deteriorated by an extreme outlier.

Statistical Comparison (Similarity Test) of Two Fragmentation Maps. After the pairwise fragmentation maps are generated for different spectral sets, statistical comparisons are performed

(22) Bickel, P. J.; Doksum, K. A. *Mathematical Statistics: Basic Ideas and Selected Topics*, 2nd ed.; Prentice Hall: Upper Saddle River, NJ, 2001; Vol I.

between them to validate the observed patterns using computer programs written in C++. To compare the overall difference between two maps, each corresponding cleavage site needs to be compared first. Denote X_{ijk} , $1 \leq k \leq n_{ij}$, the normalized intensities at cleavage site i followed by j in map α and denote Y_{ijk} , $1 \leq k \leq m_{ij}$, the corresponding intensities in map β . Here n_{ij} and m_{ij} are the total cleavage occurrences at site i followed by j in maps α and β , respectively. The detection of differences at individual cleavage site i followed by j of two maps can be performed by the Wilcoxon (or Mann–Whitney) rank sum test:²³

$$Z_{ij} = \frac{\sum_{k=1}^{n_{ij}} R(X_{ijk}) - \frac{n_{ij}(n_{ij} + m_{ij} + 1)}{2}}{\sqrt{\frac{n_{ij}m_{ij}(n_{ij} + m_{ij} + 1)}{12}}}$$

where $R(X_{ijk})$ is the rank of X_{ijk} among the combined set of $\{X_{ij1}, \dots, X_{ijm_{ij}}, Y_{ij1}, \dots, Y_{ijm_{ij}}\}$. The Wilcoxon test detects the difference of two distributions without distribution assumptions, and it is known that under null hypothesis (no difference) the test statistic is approximately $N(0,1)$ when $n_{ij} \geq 10$ and $m_{ij} \geq 10$. For each cleavage site with $n_{ij} \geq 10$ and $m_{ij} \geq 10$, a Z score is obtained. A larger absolute Z score represents a more significant difference and thus results in a smaller p value. Cleavage sites with $n_{ij} < 10$ or $m_{ij} < 10$ are treated as information missing. Denote λ as the total number of cleavage sites satisfying $n_{ij} \geq 10$ and $m_{ij} \geq 10$. To account for the fact that λ independent hypothesis tests are simultaneously performed, a p value rejection threshold of 0.001 instead of the usual 0.05 is used to avoid increased false positives.

After performing the above test for each corresponding cleavage site between two maps, to compare the overall differences of two maps, the following test statistic is considered:

$$W = \sum_{(i,j): n_{ij} > 10 \text{ and } m_{ij} > 10} |Z_{ij}|^d$$

where d is a tuning parameter and the hypothesis testing is more sensitive to differences of individual cleavage sites for larger d . In a normal situation, $d = 2$ is chosen and statistical theory has shown that in this case W follows a χ^2 distribution of degree of freedom λ under the null hypothesis. To demonstrate the magnitude of the differences between multiple fragmentation maps, the statistic $V = W/\lambda$ is used to take into account different λ in each comparison, in which V is a measurement of the degree of difference between different fragmentation maps.

RESULTS AND DISCUSSION

In the text below, the composition of the 28 330 spectra is briefly summarized and statistics are provided on the ion types that make up the spectra. The remainder of the text then focuses on a presentation and discussion of the chemically meaningful amide bond cleavage patterns for peptides with different structural motifs. When interpreting the data, we typically assume that the backbone amide bond cleavages that occur are charge-directed; i.e., that

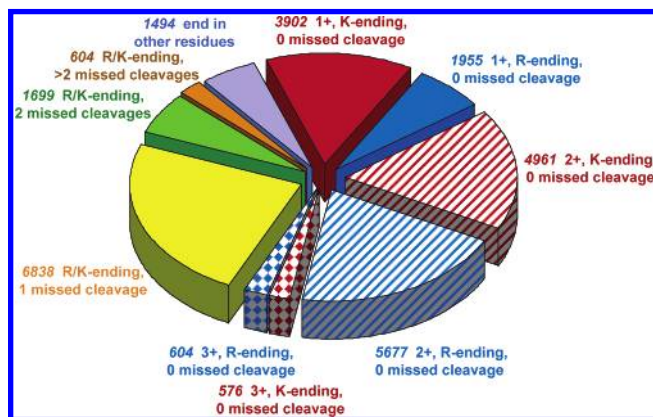


Figure 1. Composition of the 28 330 peptide MS/MS spectral database.

they occur via involvement of a proton with the exception of cleavage at Asp–Xxx.²⁴ This assumption is most likely to break down for peptides in which the number of protons is less than or equal to the number of Arg.

Composition of the MS/MS Database. Among the 28 330 spectra from peptides of unique sequences and charge states, the lengths of the peptides range from 5 to 55 AA residues, with the median and average both at 16 residues. A summary of the spectral database composition by the charge states, the terminating residue, and the number of missed cleavages is shown in Figure 1. Among the 28 330 spectra, 7181 (25.3%) are from singly charged (1+) peptides, with a median length of 11 residues; 17 660 (62.3%) are from doubly charged (2+) peptides, with a median length of 16 residues; 3489 (12.3%) are from triply charged (3+) peptides, with a median length of 26 residues. Separating the spectra by the terminating residue of the peptides, 12 583 (44.4%) are from peptides that end in Arg and 14 253 (50.3%) are from peptides that end in Lys. There is no significant difference in peptide length between the Arg- and the Lys-ending peptides.

Ion Statistics. To analyze the spectral content of the 28 330 peptide MS/MS spectra and to glean chemical information from the analyses, it is first desirable to characterize the spectra in terms of the number and types of product ions typically produced in the ion trap. Five sets of spectra from tryptic peptides with zero missed cleavages and peptides without basic residues are characterized (Table 1). (Only two spectra exist for triply protonated peptides without basic residues, therefore that data set is not included.) Two sets of analyses were performed on these five data sets. The first set of analyses focuses on the impact of internal series ions on peak assignment (See Table 1.). The second set of analyses compare the frequencies and intensities of different noninternal ions (See Table 2.). Conclusions drawn from these analyses are summarized briefly as follows: (More detailed discussions are provided in the Supporting Information.)

- (1) The peptides are longer in higher charge state data sets.
- (2) A significant portion of the spectra from triply charged tryptic peptides and doubly charged peptides without basic residues, which are the longest, is cut off due to the experimental cutoff at m/z 2000.
- (3) There is a considerable fraction of the total peaks and total ion intensity in ion trap spectra that cannot be labeled. This fraction is most significant in singly charged tryptic peptides.

(23) Rice, J. A. *Mathematical Statistics and Data Analysis*, 2nd ed.; Duxbury Press: Belmont, CA, 1995.

Table 1. Characterizing Spectral Sets of Different Charge States and Basic Residue Content^a

| | 3+ tryptic | 2+ tryptic | 1+ tryptic | 2+ no RKH | 1+ no RKH |
|------------------------|------------|------------|------------|------------|------------|
| av length ± SD | 25.7 ± 7.8 | 16.4 ± 5.8 | 11.5 ± 3.6 | 20.2 ± 6.4 | 11.8 ± 4.8 |
| median length | 25 | 17 | 16 | 19 | 17 |
| % peptide mass >2000 | 82.9% | 32.4% | 0 | 48.9% | 0 |
| av no. of peaks per AA | 12.0 | 17.3 | 23.2 | 14.6 | 19.1 |
| no. of spectra | 1180 | 10638 | 5857 | 131 | 103 |
| % intensity identified | | | | | |
| w/o internals | 48.0 | 48.2 | 38.0 | 36.9 | 44.3 |
| with internals | 80.4 | 65.1 | 48.1 | 60.0 | 54.6 |
| % peaks identified | | | | | |
| w/o internals | 32.2 | 29.0 | 19.6 | 26.9 | 20.9 |
| with internals | 74.7 | 52.3 | 31.1 | 52.1 | 34.3 |
| % multiple assignments | | | | | |
| w/o internals | 22.2 | 16.1 | 3.8 | 17.5 | 3.6 |
| with internals | 63.5 | 42.5 | 19.2 | 47.5 | 21.3 |

^a 3+, 2+, and 1+: refer to the charge states. Tryptic: Arg or Lys only occurs at the C-terminus of the peptide. No RKH: stands for peptides without Arg, Lys, or His. Median values are shown unless specified otherwise. w/o internals: uses y, y - H₂O, y - NH₃, b, b - H₂O, b - NH₃, b + H₂O, a, a - H₂O, a - NH₃, M - H₂O, M - NH₃, and immonium ions to label the spectra. With internals: uses the same ion types plus the internal series (including internal, internal-H₂O, internal-NH₃, internal-CO). % peptide mass >2000: the percentage of spectra with a corresponding peptide MW greater than 2000 Da.

Table 2. Frequencies (Freq) and Sum of Normalized Intensities (Int) of Different Ion Types among Different Spectral Sets^a

| | 3+ tryptic | | 2+ tryptic | | 1+ tryptic | | 2+ no RKH | | 1+ no RKH ion type | |
|--------------------------------------|------------|-------------|------------|-------------|-----------------|------------|-----------|------------|--------------------|-------------|
| | Freq | % Int | Freq | % Int | Freq | % Int | Freq | % Int | Freq | % Int |
| y ⁺ | 1.000 | 7.2 | 1.000 | 19.2 | 1.000 | 9.5 | 1.000 | 7.5 | 1.000 | 7.9 |
| y ²⁺ | 0.997 | 10.0 | 0.973 | 2.1 | na ^b | na | 0.924 | 0.8 | na | na |
| [y - H ₂ O] ⁺ | 0.945 | 1.0 | 0.976 | 1.4 | 0.983 | 1.6 | 0.977 | 1.7 | 0.971 | 1.7 |
| [y - H ₂ O] ²⁺ | 0.985 | 2.2 | 0.955 | 1.1 | na | na | 0.954 | 0.9 | na | na |
| [y - NH ₃] ⁺ | 0.956 | 1.0 | 0.972 | 1.3 | 0.980 | 1.8 | 0.962 | 1.4 | 0.903 | 1.4 |
| [y-NH ₃] ²⁺ | 0.986 | 2.7 | 0.956 | 1.2 | na | na | 0.969 | 0.9 | na | na |
| total Int of y series | na | 24.1 | na | 26.3 | na | 12.8 | na | 13.2 | na | 11.0 |
| b ⁺ | 0.997 | 3.8 | 1.000 | 6.5 | 1.000 | 5.3 | 1.000 | 6.9 | 1.000 | 11.4 |
| b ²⁺ | 0.997 | 4.6 | 0.921 | 0.8 | na | na | 0.969 | 1.3 | na | na |
| [b - H ₂ O] ⁺ | 0.967 | 1.6 | 0.988 | 2.6 | 0.995 | 3.0 | 0.992 | 1.8 | 0.971 | 4.7 |
| [b - H ₂ O] ²⁺ | 0.980 | 2.2 | 0.902 | 0.7 | na | na | 0.947 | 0.9 | na | na |
| [b - NH ₃] ⁺ | 0.963 | 1.2 | 0.981 | 2.0 | 0.981 | 2.0 | 0.985 | 1.6 | 0.932 | 3.8 |
| [b - NH ₃] ²⁺ | 0.981 | 2.4 | 0.902 | 0.7 | na | na | 0.947 | 1.0 | na | na |
| [b + H ₂ O] ⁺ | 0.891 | 0.7 | 0.868 | 0.5 | 0.911 | 0.7 | 0.855 | 0.5 | 0.796 | 0.5 |
| [b + H ₂ O] ²⁺ | 0.968 | 1.2 | 0.917 | 0.7 | na | na | 0.931 | 0.9 | na | na |
| a ⁺ | 0.945 | 0.8 | 0.970 | 1.1 | 0.969 | 1.2 | 0.947 | 0.8 | 0.971 | 1.2 |
| a ²⁺ | 0.969 | 1.5 | 0.901 | 0.7 | na | na | 0.939 | 0.8 | na | na |
| [a - H ₂ O] ⁺ | 0.922 | 0.7 | 0.922 | 0.7 | 0.924 | 0.7 | 0.924 | 0.6 | 0.893 | 0.7 |
| [a - H ₂ O] ²⁺ | 0.969 | 1.2 | 0.887 | 0.6 | na | na | 0.954 | 0.5 | na | na |
| [a - NH ₃] ⁺ | 0.914 | 0.8 | 0.944 | 0.9 | 0.961 | 1.0 | 0.908 | 0.8 | 0.932 | 1.1 |
| [a - NH ₃] ²⁺ | 0.969 | 1.5 | 0.901 | 0.7 | na | na | 0.939 | 0.8 | na | na |
| total Int of b series | na | 24.1 | na | 19.2 | na | 13.9 | na | 19.1 | na | 23.5 |
| [M - H ₂ O] ⁺ | 0 | na | 0.002 | 0.1 | 0.448 | 0.3 | 0 | na | 0.398 | 0.3 |
| [M - H ₂ O] ²⁺ | 0.114 | 0.2 | 0.314 | 0.3 | na | na | 0.435 | 0.8 | na | na |
| [M - NH ₃] ⁺ | 0 | na | 0.002 | 0.1 | 0.953 | 2.5 | 0 | na | 0.893 | 1.2 |
| [M - NH ₃] ²⁺ | 0.120 | 0.1 | 0.507 | 0.5 | na | na | 0.595 | 0.8 | na | na |
| immonium | 0.028 | 0.3 | 0.057 | 0.3 | 0.004 | 0.2 | 0.008 | 6.0 | 0.029 | 0.1 |
| x ⁺ | 0.891 | 0.6 | 0.865 | 0.5 | 0.841 | 0.5 | 0.908 | 0.6 | 0.786 | 0.5 |
| c ⁺ | 0.904 | 0.6 | 0.865 | 0.5 | 0.845 | 0.5 | 0.931 | 0.8 | 0.816 | 0.5 |

^a The intensities of those ions that are significant are highlighted in boldface type. (See text and Supporting Information for details.) ^b na, not available.

(4) For a given number of basic residues in a peptide, cleavages are less selective (or more uniform) in peptides of higher charge states.

(5) Internal ions, formed by two cleavages of the peptide backbone, encompass a significant portion of the total ion intensity in ion trap spectra. Their abundances are higher in spectra with higher charge states.

(6) A significant portion of the spectra from triply charged tryptic peptides are cut off and the presence of abundant internal ions in these spectra are the main reasons for the lower identification rate of triply charged tryptic peptides.

(7) The low resolution of ion trap causes ambiguity in peak labeling, which leads to multiple assignments to the same peak. This can create problems for sequencing algorithms.

Table 3. Summary of the Nine Different Spectral Sets for Which Fragmentation Maps Are Analyzed in This Paper

| roman numeral | charge state | no. of P | basic residue content | structural motif representation | no. of spectra | most abundant cleavages |
|---------------|------------------|----------|-----------------------|--|----------------|-------------------------|
| I | 1 H ⁺ | ≥1 | 1 R, 0 K, 0 H | [...P...noH...R] ⁺ | 674 | D-X, E-X in y |
| II | 1 H ⁺ | ≥1 | 0 R, 1 K, 0 H | [...P...noH...K] ⁺ | 1,216 | D-X, X-P in y |
| III | 2 H ⁺ | ≥1 | 1 R, 0 K, 0 H | [...P...noH...R] ²⁺ | 2,182 | X-P in y |
| IV | 2 H ⁺ | ≥1 | 0 R, 1 K, 0 H | [...P...noH...K] ²⁺ | 1,834 | X-P in y |
| V | 2 H ⁺ | ≥1 | 2 R, 0 K, 0 H | [...R...P...noH...D/E...R] ²⁺ | 201 | D-X, E-X in b and y |
| VI | 2 H ⁺ | ≥1 | 0 R, 2 K, 0 H | [...K...P...noH...D/E...K] ²⁺ | 230 | X-P in y |
| VII | 1 H ⁺ | 0 | 0 R, 1 K, 0 H | [...noP...noH...K] ⁺ | 1,978 | D-X, E-X in y |
| VIII | 2 H ⁺ | 0 | 1 R or 1 K, 0 H | [...noP...noH...R/K] ²⁺ | 2755 | I-X, V-X, L-X in y |
| IX | 2 H ⁺ | varies | 1 R or 1 K, varies H | [...R/K] ²⁺ | 10,638 | X-P in y |

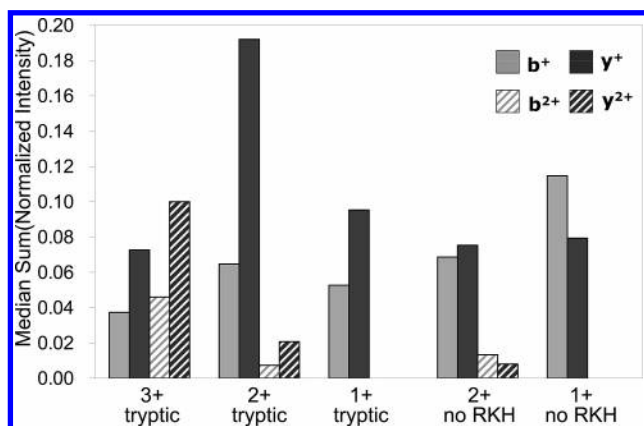


Figure 2. Relative abundances of different ion types in different spectral sets. Ion intensities are normalized to the total ion abundances in the spectrum and then summed together according to the ion types. Ion types listed from the left to right for each data set: b⁺, y⁺, b²⁺, y²⁺. 3+ tryptic, 2+ tryptic, and 1+ tryptic stand for triply charged, doubly charged, and singly charged Arg- or Lys-ending peptides with 0 missed cleavages, respectively. 2+ no RKH and 1+ no RKH stand for doubly charged and singly charged peptides that do not contain any basic residues, respectively.

(8) Longer peptides have a larger number of possible fragment ions. Therefore, spectra of higher charge states have higher probabilities of multiple assignments to a single peak. The possible internal series ions, in particular, increase exponentially with the peptide length. Multiple assignments are most significant among triply charged tryptic peptides and doubly charged peptides with no basic residues when internals are included in the labeling scheme.

(9) Different ion types have different frequencies and abundances among different data sets.

(10) The probability of randomly matching a peak is high in low-energy CID spectra. This probability increases with the charge state of the precursor ion. However, the intensity from such random matching is low.

(11) Singly charged fragments are more abundant than doubly charged fragments when the precursor ions are doubly charged. Doubly charged fragments become more abundant than singly charged fragments when the precursor ions are triply charged.

(12) b⁺ and y⁺ (or b²⁺ and y²⁺ in triply charged peptides) are the most abundant fragment ion types when internal ions are not included (Figure 2). Tryptic peptides have a preference of forming y series ions over b series ions, while among peptides without basic residues, abundances of b series ions are comparable to, if not more intense than, y series ions. The largest preference for

formation of y series over b series ions is found in doubly charged tryptic peptides.

Fragmentation Maps of Different Spectral Sets. (a) Nine Sets of Peptides with Specific Structural Features and Charge States. The spectra were sorted based on the structural features and charge states of the corresponding peptides. To understand the roles of Lys and Arg protonation, peptides with internal His were not included in most spectral sets. Table 3 shows a summary of nine different spectral sets for which fragmentation maps are analyzed in this paper. These are (I) 674 spectra from singly charged tryptic peptides that contain one or more Pro, no His (indicated by noH), and C-terminal Arg ([...P...noH...R]⁺); (II) 1216 spectra from singly charged tryptic peptides that contain one or more Pro, no His, and C-terminal Lys ([...P...noH...K]⁺); (III) 2182 spectra from doubly charged tryptic peptides that contain one or more Pro, no His, and C-terminal Arg ([...P...noH...R]²⁺); (IV) 1834 spectra from doubly charged tryptic peptides that contain one or more Pro, no His, and C-terminal Lys ([...P...noH...K]²⁺); (V) 201 spectra from doubly charged peptides that contain one or more Pro, one or more acidic residues (Asp or Glu), no His or Lys, C-terminal Arg, and only one internal Arg N-terminal to the acidic residue ([...R...P...noH...D/E...R]²⁺); (VI) 230 spectra from doubly charged peptides with the same structural motif as in (V) but substitute Arg by Lys ([...K...P...noH...D/E...K]²⁺); (VII) 1978 spectra from singly charged tryptic peptides that contain no Pro or His, and C-terminal Lys ([...noP...noH...K]⁺); (VIII) 2755 spectra from doubly charged tryptic peptides that contain no Pro or His ([...noP...noH...R/K]²⁺); (IX) 10 638 spectra from doubly charged tryptic peptides that contain any combination of residues ([...R/K]²⁺). In the subsequent text, these spectral sets will be referred to by the Roman numeral preceding them. Note that data set IX corresponds directly to the data set previously reported in a communication¹⁷ but includes almost twice as many spectra (10 638 vs 5654) and is presented in the new visualization scheme with the current statistical methods.

(b) Frequency of Selective Cleavage in 1+ R- versus K-Ending Peptides. Selective cleavage has been noted for years in the literature, especially in peptides where the number of protons is less than or equal to the number of Arg.^{10,13,24–30} In

(24) Wysocki, V. H.; Tsaprailis, G.; Smith, L. L.; Breci, L. A. *J. Mass Spectrom.* **2000**, *35*, 1399–1406.

(25) Tsaprailis, G.; Somogyi, A.; Nikolaev, E. N.; Wysocki, V. H. *Int. J. Mass Spectrom.* **2000**, *196*, 467–479.

(26) Sullivan, A. G.; Brancia, F. L.; Tyldesley, R.; Bateman, R.; Sidhu, K.; Hubbard, S. J.; Oliver, S. G.; Gaskell, S. J. *Int. J. Mass Spectrom.* **2001**, *210*, 665–676.

(27) Papayannopoulos, I. A. *Mass Spectrom. Rev.* **1995**, *14*, 49–73.

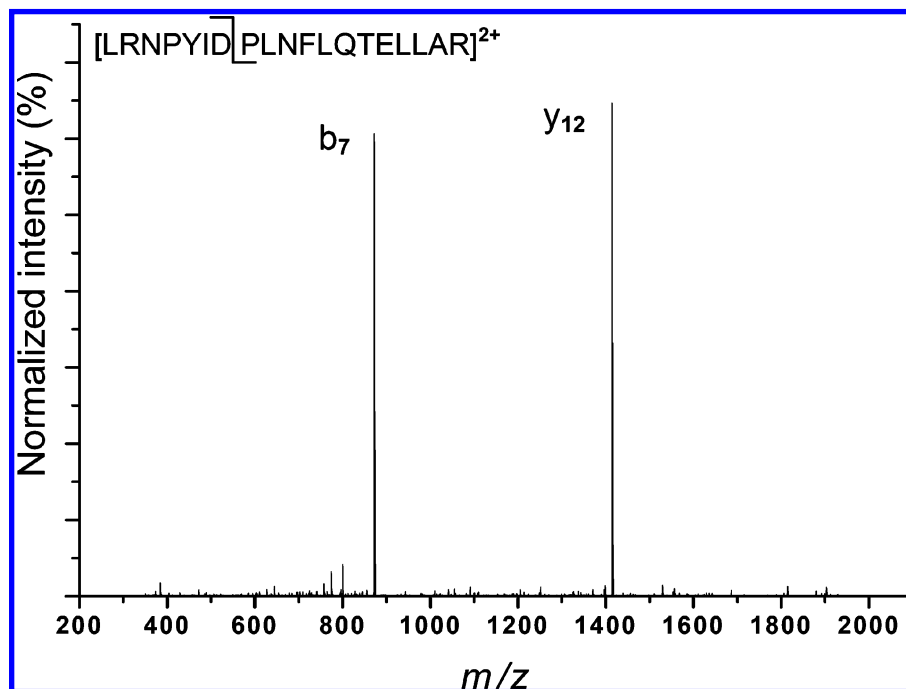


Figure 3. Example of a spectrum showing selective cleavage from doubly charged peptide LRNPYIDPLNFLQTELLAR.

spectra that show selective cleavages, certain cleavage sites become the preferred pathways for dissociation, while cleavages at other sites are diminished. Figure 3 is an example of a spectrum showing selective cleavage at the Asp–Pro peptide bond from a doubly charged peptide LRNPYIDPLNFLQTELLAR. Product ions b_7 and y_{12} from such a selective cleavage dominate the spectrum and encompass 65% of the total ion intensity in the spectrum. It is important to study spectra of these kind because they deviate the most from the uniform fragmentation model and are usually classified as “low-information” content³¹ or sometimes “low quality” because the contiguous ion series are missing.

To study how the identity of different basic residues influences the fragmentation statistics, spectral data set I ([...P...noH...R]⁺) is compared to set II ([...P...noH...K]⁺). The histidine residue is excluded because His is also a basic residue whose side chain can serve as a “proton holder” and allow selective cleavage.^{14,32} By excluding His, the main structural element that leads to the difference in basicity between these data sets is the C-terminal residue. An analysis was performed to see whether these spectra fall into specific categories that involve nonselective versus selective cleavage. The result shows that 53% of the spectra from Arg-ending peptides (data set I) have 25% or more of the total fragment ion abundance contributed by the two most abundant fragment ion peaks; on average, these two most abundant peaks account for $52 \pm 30\%$ of the total abundance. However, only 26% of the spectra from peptides terminating in Lys (data set II) have 25% or more of the total abundance contributed by the two most

abundant peaks; on average, these two most abundant peaks account for $45 \pm 24\%$ of the total abundance.

Since the only difference in sequence motif between (I) and (II) is the terminal basic residue, one can conclude that the difference in the basicity between Arg and Lys results in different degrees of proton localization, thus leading to different degrees of selective cleavage. To find out which cleavages among the 400 residue combinations are selective, the fragmentation maps for these two data sets are generated.

(c) Fragmentation Maps for y Ions from [...P...noH...R]⁺ versus [...P...noH...K]⁺. Panels I_y and II_y in Figure 4 show the maps for y ions from (I) ([...P...noH...R]⁺) and (II) ([...P...noH...K]⁺), respectively. For Arg-terminating peptides I, cleavage is very strong and concentrated C-terminal to Asp and Glu (rows labeled D and E), with the Asp cleavage stronger than Glu cleavage. Weak cleavage was observed N-terminal to Pro (column labeled P) and C-terminal to asparagine (Asn or N). Cleavages at other sites are minimal. For Lys-ending peptides II, the dominant cleavage is N-terminal to Pro, followed by C-terminal to Asp. Only weak cleavage is observed C-terminal to Glu. Weak cleavage is also observed N-terminal to Gly. While minimal cleavages are observed C-terminal to Gly and Pro, most residue combinations show more abundant cleavage in Figure 4II_y than in Figure 4I_y. Rows for Arg and Lys are missing because Arg and Lys only occur at the C-terminus in the peptides of these data sets; i.e., they are not located N-terminal to any other residues. Most of the column information for Arg and Lys is missing because y_1 product ions are often below the low-mass cutoff inherent to ion trap instruments.

Based on previous studies, enhanced cleavage at acidic residues demands the absence of a proton at Asp–Xxx or Glu–Xxx, i.e., the acid groups are neutral COOH.¹³ The observation that cleavages at Asp–Xxx and Glu–Xxx are stronger in Arg-ending than Lys-ending peptides corroborates the argument that

(28) Loo, J. A.; Edmonds, C. G.; Smith, R. D. *Anal. Chem.* **1993**, *65*, 425–438.

(29) Reid, G. E.; Wu, J.; Chrisman, P. A.; Wells, J. M.; McLuckey, S. A. *Anal. Chem.* **2001**, *73*, 3274–3281.

(30) Reid, G. E.; Stephenson, J. L.; McLuckey, S. A. *Anal. Chem.* **2002**, *74*, 577–583.

(31) Simpson, R. J.; Connolly, L. M.; Eddes, J. S.; Pereira, J. J.; Moritz, R. L.; Reid, G. E. *Electrophoresis* **2000**, *21*, 1707–1732.

(32) Tsaprailis, G.; Nair, H.; Zhong, W.; Kuppanan, K.; Futrell, J. H.; Wysocki, V. H. *Anal. Chem.* **2004**, *76*, 2083–2094.

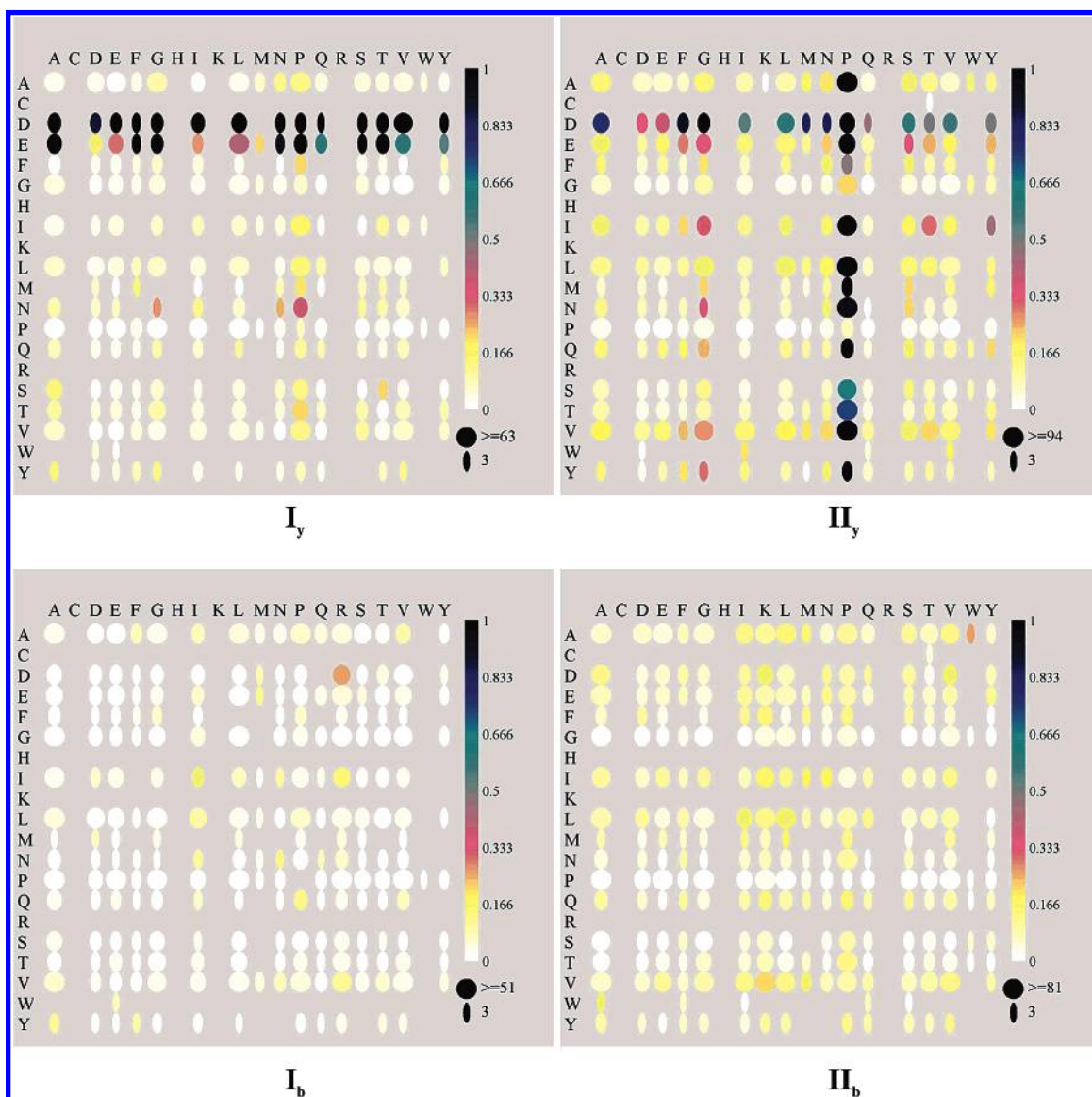


Figure 4. Pairwise fragmentation maps showing median bond cleavage intensities at specific Xxx–Zzz residue combinations from **I** [...P...noH...R]⁺ and **II** [...P...noH...K]⁺. The ion intensities are normalized to the most abundant peak in b and y ions combined. Subscripts b and y correspond to the ion types. The single letter codes of AA residues listed in the leftmost column correspond to the N-terminal residue (Xxx) in an Xxx–Zzz pair while those listed along the topmost row correspond to the C-terminal residue (Zzz). The color scheme is shown on the right side of each map. The horizontal dimension of each ellipse is proportional to the count of such pairwise cleavages. Illustrations of such proportion between the size and the count are shown under the color bar at the lower right corner. See Table 3 for the number of peptides in each set.

the Arg side chain is more effective in sequestering the ionizing proton than the side chain of Lys.^{12,14,26} The fact that both data sets contain at least one Pro in their sequence, but cleavage at Xxx–Pro is enhanced only in Lys-ending peptides, indicates that differences in proton localization causes the difference in Xxx–Pro cleavage patterns—cleavage at Xxx–Pro is charge dependent, i.e., requires a proton at the Xxx–Pro amide bond (see also results for data sets III and IV).

For singly charged Arg-ending peptides, the result suggests that only one dominant protonation form exists in which a proton is completely “sequestered” by the guanidino side chain and not accessible to the peptide backbone. This is consistent with the observation that (1) no enhanced cleavage occurs at Xxx–Pro, even though Pro is present in all peptides and (2) cleavages at Asp–Xxx and Glu–Xxx are the dominant cleavage pathways, because these cleavages are “charge-remote”,^{13,24} i.e., they do not

involve the added proton. Lys-ending singly charged peptides fragment at both Asp–Xxx and Xxx–Pro, cleavages that were previously suggested as being “charge-remote” and “charge-directed”,²⁴ respectively. The strong cleavage at Xxx–Pro is detected in all cases where there is at least one proton in excess of the number of basic residues (See Table 3, data sets III and IV). Thus, for Lys-ending peptides, the data suggest that the side chain of Lys is not basic enough to completely sequester the ionizing proton. Therefore, when energy is added during the low-energy CID process, the proton is able to leave the Lys side chain and reach the peptide backbone. A heterogeneous population of two or more protonated forms exists in such peptides. The first protonation form is similar to that of Arg-ending peptides where the proton is localized by the basic side chain allowing charge-remote selective cleavage at Asp–Xxx. A second form leads to charge-directed selective cleavage at Xxx–Pro. In the subsequent

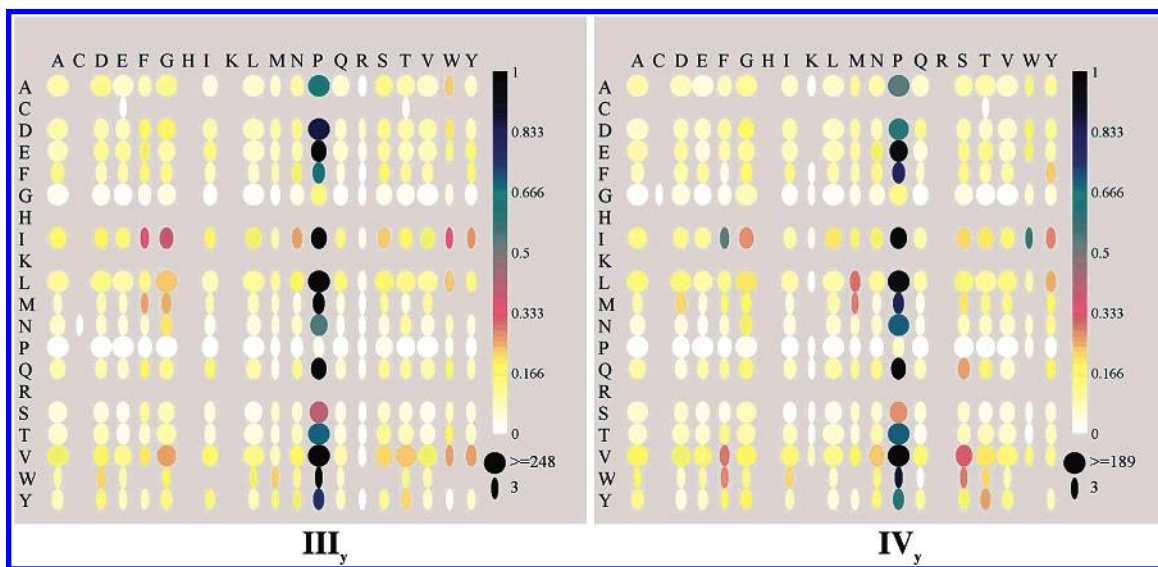


Figure 5. Pairwise fragmentation maps from III [...P...noH...R]²⁺ and IV [...P...noH...K]²⁺. See caption for Figure 4.

text, “a partially mobile proton” will be used to denote the existence of multiple protonation forms, i.e., more facile intramolecular protein transfer. This terminology is a related but perhaps less rigid definition of “partially mobile” than used previously¹⁶ by Kapp et al. Rather than counting the number of charges versus the number of basic residues, we base our definition on fragmentation maps presented in this paper. The difference in the definition of “a partially mobile proton” between ours and that used by Kapp et. al. will be further contrasted in section f for data set VI.

(d) Fragmentation Maps for b Ions from [...P...noH...R]⁺ versus [...P...noH...K]⁺. Fragmentation maps for b ions from data sets I and II are shown in Figure 4I_b and II_b. Comparing the overall intensities in these b ion maps to those in the y ion maps from the same data sets (Figure 4I_y and II_y) shows that b ions are significantly less abundant than y ions, since each peak is normalized to the most abundant peak among all b and y ions. This correlates well with the statistics in Figure 2. Cleavages overall are more abundant in Lys-ending peptides than Arg-ending peptides. No enhanced cleavages C-terminal to acidic residues or N-terminal to Pro are observed in these b ion maps.

The absence of enhanced cleavages at Asp–Xxx and Xxx–Pro in these b ion maps can be explained by the location of the proton during and after the fragmentation process. In tryptic peptides, when selective cleavage occurs C-terminal to acidic residues, the proton is at the basic side chain of the C-terminal Arg or Lys. This proton retention is most likely to continue after the Asp–Xxx bond breaks, which results in a y ion and a neutral N-terminal anhydride fragment. For selective cleavage N-terminal to Pro, the proton is likely at the carbonyl oxygen, or perhaps amide nitrogen, of the Xxx–Pro bond to initiate such cleavage. When the Xxx–Pro bond breaks, the C-terminal piece is most likely to retain the single charge. The N-terminal fragment, which would form a b ion if the charge were retained, cannot compete with the C-terminal fragment because of the large basicity difference: (1) Pro, with a secondary amine group, is at the N-terminus and (2) Lys is at the C-terminus of the forming C-terminal fragment. To see the corresponding b ions from these cleavages, an additional proton and an additional basic residue as the proton holder on the N-terminal side of the Asp–Xxx or

Xxx–Pro bond may be needed. (See section j for characterization of spectra from doubly charged tryptic peptides with one internal basic residue on the N-terminal side of the Asp–Xxx or Xxx–Pro bond.)

(e) Fragmentation Maps for y Ions from 2+ R-Ending and K-Ending Peptides. Figure 5III_y and IV_y are the y ion fragmentation maps generated for III ([...P...noH...R]²⁺) and IV ([...P...noH...K]²⁺). In contrast to their singly charged counterparts (Figure 4I_y and II_y), which differ for Arg- versus Lys-ending peptides, these maps appear very similar. Our statistical similarity test gives a *p* value of 0.0025 between these two maps, which is greater than the threshold 0.001. Therefore, these two maps do not display significant difference in their fragmentation patterns. Cleavage N-terminal to Pro is the dominant cleavage in both maps, but not as abundant as that in Figure 4II_y. Minimal cleavages are observed C-terminal to Gly and Pro, while weak cleavages are observed C-terminal to Ile, Val, and Leu and N-terminal to Phe, Gly, and Ser.

These two maps corroborate well with our earlier analysis¹⁷ and are expansions of it. In the earlier analysis published in a communication, there is no separation based on the terminal basic residue, His is not purged from the sequence and that leads to greater variation of the basicity among the data set. Also, the statistical tools employed in the current study make the new plots richer in information and statistically more meaningful (a gradient is used for the color scheme instead of a step function, count information is imbedded in the current plots, and the FI threshold is used to validate the distribution for AA pairs that have lower occurrence). Since only one basic residue exists in these peptides, but two protons are available, these data represent the cases when one proton is “localized” at the basic side chain, and the other proton is mobile, i.e., being transferred intramolecularly along the peptide backbone, initiating cleavage at any one of a variety of sites for individual protonated molecules of the whole activated population. Clearly shown from these two maps is that, even when a mobile proton is available, there are favored and unfavored cleavage sites. Chemical interactions between the side chains and the backbone, as well as secondary structures of the peptides, are the proposed main factors behind these patterns.

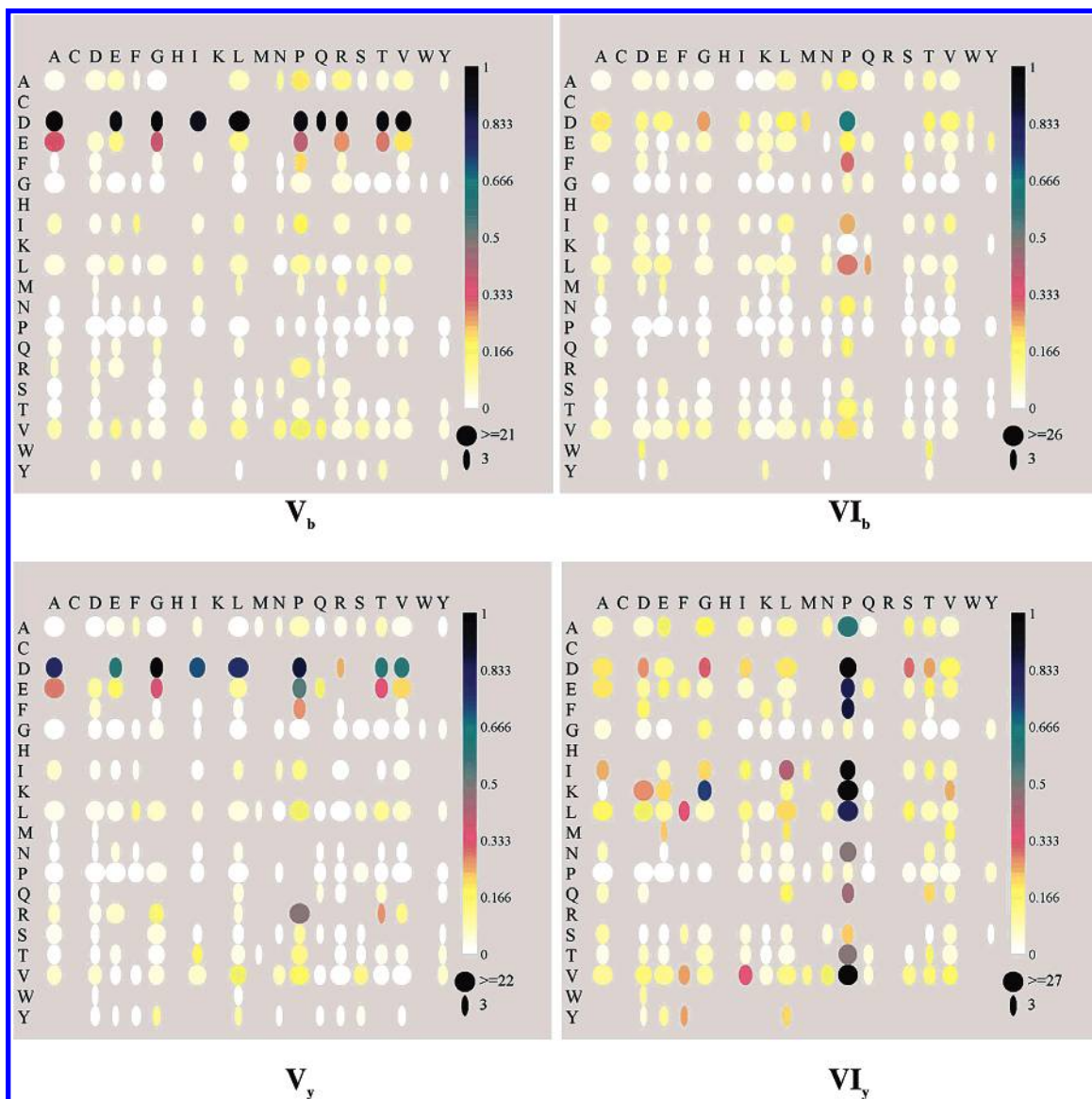


Figure 6. Pairwise fragmentation maps from V [...R...P...noH...D/E...R]²⁺ and VI [...K...P...noH...D/E...K]²⁺. See caption for Figure 4.

The b ion fragmentation maps for III and IV are also generated (see Supporting Information). Preferential cleavage N-terminal to Pro is observed in both maps, but its abundance is far less in b ion maps than those in the y ion maps. When these b ions are normalized to themselves, i.e., to the most abundant b ion, their fragmentation maps show patterns similar to those of y ions¹⁷ (See also section i, Figure 8IX_b and IX_y). The overall cleavage intensities observed in these b ion maps are less than those of the b ion maps from singly protonated Lys-ending peptides (Figure 4II_b).

The fact that preferential cleavage occurs at Pro when a second charge is added to Arg- or Lys-ending tryptic peptides is consistent with the involvement of an added proton in the cleavage at Xxx-Pro. The preferential cleavage at Pro does not indicate localization of the added proton at Pro because fragmentation is a kinetic process (a Pro-protonated form may fragment faster even if other protonated forms exist). Other cleavages are also detected in the presence of Pro. A recent publication shows that, for *N*-acetyl-*O*-methylproline, a proton can be transferred from carbonyl oxygen to amide nitrogen with assistance of an adjacent carbonyl.³³ When there is only one proton available, preferential cleavage at Xxx-

Pro was seen only for Lys-ending (Figure 4II_y) but not for Arg-ending peptides (Figure 4I_y). This is consistent with the lower basicity of Lys allowing the proton to migrate away from the Lys side chain.

(f) Fragmentation Maps for b Ions and y Ions from [...R...P...noH...D/E...R]²⁺ versus [...K...P...noH...D/E...K]²⁺. The y ions are dominant for the spectral sets presented above. This is consistent with the fact that these peptides all have only one basic residue, and that basic residue is at the C-terminus. However, in practical proteomics studies, tryptic peptides with missed cleavages are also common (Figure 1). For those peptides, basic residues are present in more than one location, which may change the b versus y intensity patterns.¹⁵ Fragmentation maps of b and y ions from V ([...R...P...noH...D/E...R]²⁺) and VI ([...K...P...noH...D/E...K]²⁺) are generated (Figure 6V_b, V_y, VI_b, and VI_y). Since these two sets of spectra are from peptides with very specific structural motifs, the numbers of spectra that fit the criteria are limited. Hence, for these two data sets, the occurrences

(33) Komaromi, I.; Somogyi, A.; Wysocki, V. H. *Int. J. Mass Spectrom.* **2005**, *76*, 315–323.

Table 4. Degree of Difference between Selected Pairs of Fragmentation Maps^a

| Degree of Difference | I _y | II _y | I _b | II _b | III _y | IV _y | V _b | VI _b | V _y | VI _y | VII _y | VIII _y | VII _b |
|----------------------|----------------|-----------------|----------------|-----------------|------------------|-----------------|----------------|-----------------|----------------|-----------------|------------------|-------------------|------------------|
| I _y | - | 7.40 | - | - | 11.04 | 10.49 | 3.32 | 6.21 | 3.55 | 6.11 | 14.74 | 16.01 | - |
| II _y | 7.40 | - | - | - | 4.97 | 4.97 | 6.64 | 8.05 | 6.52 | 2.36 | 9.25 | 8.53 | - |
| I _b | - | - | - | 3.41 | - | - | 4.37 | 3.41 | - | - | - | - | 12.47 |
| II _b | - | - | 3.41 | - | - | - | 4.18 | 2.75 | - | - | - | - | 7.17 |
| III _y | 11.04 | 4.97 | - | - | - | 1.27 | - | - | 7.18 | 0.55 | 21.53 | 9.02 | - |
| IV _y | 10.49 | 4.97 | - | - | 1.27 | - | - | - | 7.31 | 1.54 | 19.48 | 8.05 | - |
| V _b | 3.32 | 6.64 | 4.37 | 4.18 | - | - | - | 2.68 | 2.07 | 4.23 | - | - | 8.55 |
| VI _b | 6.21 | 8.05 | 3.41 | 2.75 | - | - | 2.68 | - | 2.90 | 3.51 | - | - | 9.51 |
| V _y | 3.55 | 6.52 | - | - | 7.18 | 7.31 | 2.07 | 2.90 | - | 4.36 | - | - | - |
| VI _y | 6.11 | 2.36 | - | - | 0.55 | 1.54 | 4.23 | 3.51 | 4.36 | - | - | - | - |
| VII _y | 14.74 | 9.25 | - | - | 21.53 | 19.48 | - | - | - | - | - | 10.81 (2.46) | 12.04 |
| VIII _y | 16.01 | 8.53 | - | - | 9.02 | 8.05 | - | - | - | - | 10.81 (2.46)* | - | - |
| VII _b | - | - | 12.47 | 7.17 | - | - | 8.55 | 9.51 | - | - | - | - | - |

^a A smaller number means the two sets compared are more similar. The Roman numerals correspond to spectral sets described in Table 3. The subscripts b and y correspond to the ion types. Some pairs are not selected because they have very few or no common elements in their fragmentation maps shown in Figures 4–7. The numbers in parentheses correspond to the comparison between VII_y and VIII_y with the rows of D and E purged.

are low (<10) for quite a few pairwise cleavages. The FI threshold is particularly useful here to overcome the lower number of spectra and show statistically valid trends.

The b and y ion maps (Figure 6V_b and V_y) from V ([...R...P...noH...D/E...R]²⁺) are both similar to the y ion map (Figure 4I_y) from I ([...P...noH...R]⁺). Enhanced cleavage is observed C-terminal to Asp and Glu; some weak cleavage was observed N-terminal to Pro; while cleavages at other AA pairs are suppressed. These results support our hypothesis that the reason enhanced cleavages at Asp–Xxx and Xxx–Pro were not observed in b ion maps of singly charged tryptic peptides I and II is because there is only one charge and one proton holder at the C-terminal side of the cleavage bond, but no proton holder available at the N-terminal side. Once an additional proton is added and a basic residue exists at the N-terminal side of Asp or Pro, the b ions from cleavages at Asp–Xxx or Xxx–Pro show up. In addition, the b ion map (Figure 6V_b) shows more abundant cleavage at acidic residues than the y ion map (Figure 6V_y) and is more similar to Figure 4I_y. Similarity tests show that among all the maps present in this study, V_b has the shortest similarity distance to I_y (Table 4). This slight predominance of b ions over y ions correlates with our observation on cleavage from peptides without basic residues (Figure 2). When the basicity of the residues from either side of the cleavage is the same, the N-terminal amine is giving b ions preference over y ions. However, a similar b versus y intensity pattern was not found in Lys-containing data set VI.

The fragmentation maps from VI ([...K...P...noH...D/E...K]²⁺) are shown in Figure 6VI_b and VI_y. According to the “relative proton mobility scale”¹⁶ by Kapp et al., “peptide ions” in VI should contain “partially mobile” protons. However, the fragmentation maps show that these peptides behave more similarly to peptides that have a mobile proton. Despite the fact that there is one Lys on the N-terminal side of the peptides, the y ions (Figure 6VI_y) are significantly more abundant than the b ions (Figure 6VI_b). The y ion map VI_y shows patterns more similar to those from III_y ([...P...noH...R]²⁺) and IV_y ([...P...noH...K]²⁺), rather than II_y ([...P...noH...K]⁺). In fact, our statistical similarity tests show a *p* value of 0.0027 between VI_y and III_y and a *p* value of 0.000 13

between VI_y and IV_y (as will be shown in Table 4 below, these two pairs of sets have two of the smallest degrees of difference between sets). The b ion map VI_b does show some degree of similarity to II_y ([...P...noH...K]⁺): cleavages at Asp–Xxx and Xxx–Pro are more abundant than cleavages at other places; however, their intensities are rather weak. The dominance of y ions in VI suggests that when Lys is present at the middle of the sequence rather than at the C-terminus, its ability to sequester the added proton is significantly hindered. In addition, cleavage at Asp–Xxx is significantly weaker in VI versus V. Such a difference was not observed between the singly charged peptides II versus I (Figure 4II_y and 4I_y). These observations suggest a possible interaction between the side chain of the internal Lys and the side chain of Asp among the majority of the peptide molecules in VI. A bridge structure may exist in these molecules that prevents the Lys side chain from sequestering the proton, allowing a mobile proton to induce fragmentation similar to that in doubly charged tryptic peptides without internal basic residues (III and IV). Whether or not such an interaction can form may depend on the distance between Lys and Asp, and the secondary structure of the peptide. When such interaction does not happen for a small population of the peptides, Lys then has a behavior in VI similar to that of Arg in V, initiating selective cleavage at Asp.

(g) Fragmentation Maps for y Ions and b Ions from Tryptic Peptides That Do Not Contain Pro. Cleavage N-terminal to Pro dominates the y ion fragmentation maps from peptides that have a mobile or partially mobile proton (Table 3). When their no-Pro counterpart peptides VII ([...noP...noH...K]⁺) and VIII ([...noP...noH...R/K]²⁺) are selected, more abundant cleavages are observed in y ions at almost all residue combinations (Figure 7VII_y and VIII_y). Despite the difference in their charge states, VII_y and VIII_y show very similar patterns, except for the enhanced cleavage C-terminal to acidic residues. The similarity tests confirmed our observation: by excluding the cleavages C-terminal to acidic residues in the calculation, the similarity distance decreases from 10.81 to 2.46 between these two maps (Table 4). Enhanced cleavage C-terminal to branched aliphatic residues (Val, Ile, Lue) and suppression of cleavage C-terminal to Gly are observed in

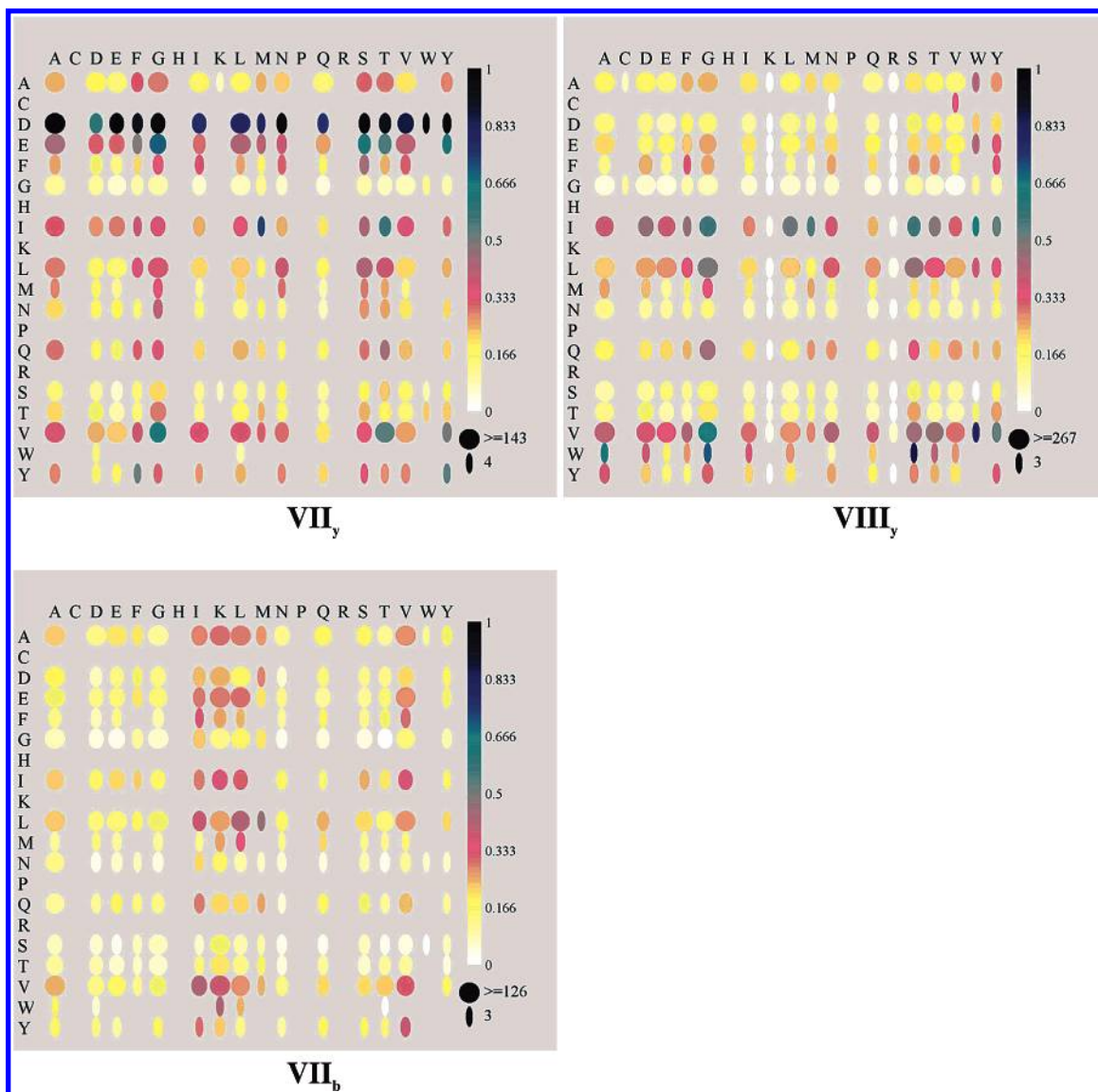


Figure 7. Pairwise fragmentation maps from VII [...noP...noH...K]⁺ and VIII [...noP...noH...R/K]²⁺. See caption for Figure 4.

both maps. Relatively strong cleavages C-terminal or N-terminal to Tyr are also observed in both maps. The preference of Gly, Ser, and Thr to cleave at their N-terminal sides rather than their C-terminal sides is more prominent in the map for singly charged peptides (Figure 7VII_y). Data set VIII has enough counts to show that Trp (W) can have relatively strong cleavage occurring at either its N- or its C-terminal side (Figure 7VIII_y).

The dramatic change in fragmentation maps (Figure 4II_y vs 7VII_y, Figure 5III_y and 5IV_y vs 7VIII_y) shows that when Xxx-Pro is not available, other possible pathways take over. Figure 7VII_y also supports our conclusion that multiple protonation forms exist in singly charged Lys-ending peptides. While enhanced cleavages at acidic residues account for one protonation form where the proton is at the basic side chain, similar to those observed for singly charged Arg-ending peptides in Figure 4I_y, strong and weak cleavage patterns at other amino acid residues are similar to those observed for doubly charged tryptic peptides (Figure 7VIII_y), in which one proton is transferred along the peptide backbone.

The b ion fragmentation maps for VII (Figure 7VII_b) and VIII (see Supporting Information) are also prepared. While the b ions

from VIII ([...noP...noH...R/K]²⁺), similar to the b ions from their Pro-containing counterparts III and IV, are much less abundant than the y ions from the same data sets, b ions from VII ([...noP...noH...K]⁺) show interesting patterns that have not been observed in other data sets. Similarity tests show VII_b is very different from any other b ion maps (Table 4). Enhanced cleavage N-terminal to Lys and branched aliphatic residues (Ile, Val, Leu) are observed in Figure 7VII_b. The degree of enhancement is higher at cleavage sites where the C-terminal residues are also aliphatic. Most patterns observed in the y ions from the same data set are not observed in the b ions, e.g., no enhanced cleavage C-terminal to acidic residues or N-terminal to Gly, Ser, and Thr.

Leaving Pro out from the singly charged Arg-ending peptides without His did not change the fragmentation patterns. The b and y ion fragmentation maps from 829 singly charged Arg-ending peptides without His or Pro ([...noP...noH...R]⁺; see Supporting Information) are very similar to those from their Pro-containing counterparts (Figure 4I_y and 4I_b). This corroborates our earlier argument for a single protonation form among such peptides in which the proton is sequestered by the basic Arg side chain.

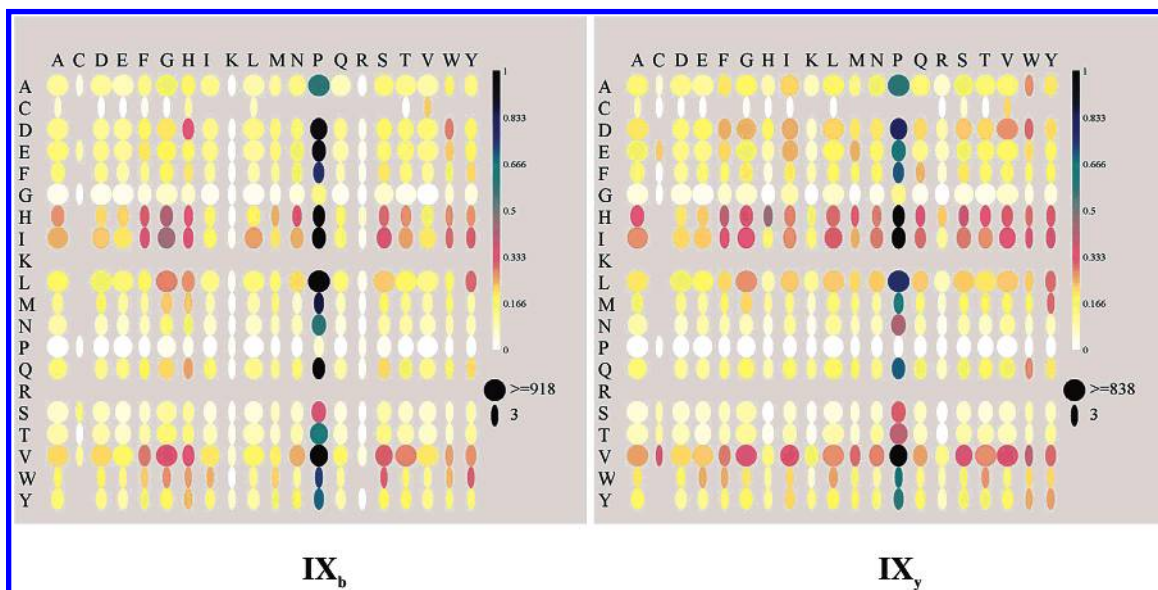


Figure 8. Pairwise fragmentation maps from IX [...R/K]²⁺. The intensity of each ion is normalized to the most abundant peak of that ion type.

(h) Similarity between Fragmentation Maps. Statistical similarity tests between the fragmentation maps from data sets I–VIII presented above were performed. Each map is compared to the rest to determine whether they exhibit similar fragmentation patterns. Except for pairs III_y and IV_y, III_y and VI_y, and IV_y and VI_y, all other maps display unique fragmentation patterns that are significantly different from the rest (*p* value < 1 × 10⁻¹⁴). Further assessment of the magnitude of difference was designed to determine whether some maps are more similar to a specific map than to the rest. The degree of difference between selected pairs of fragmentation maps was calculated and the results are summarized in Table 4. The selection of the pairs of the maps to compare is based on whether the two maps display any common fragmentation patterns. Some pairs are not selected because they have very few or no common elements in their fragmentation patterns observed in Figures 4–7.

(i) Fragmentation Maps for Doubly Charged Tryptic Peptides. We previously reported the fragmentation map for 5654 doubly charged tryptic peptides using a 10-scale color scheme and without the imbedded count information.¹⁷ Here, the fragmentation behavior of 10 638 doubly charged tryptic peptides is shown in Figure 8IX_b and IX_y using the new visualization tools presented in this paper and twice as many unique peptides. A normalization scheme of the peak intensities similar to that in the communication,¹⁷ but different from those previously used in this paper, was applied for these two maps. Each ion is normalized to the most abundant peak of that ion type in that spectrum. These two maps (1) corroborate our previous results by using twice as many peptides and by using the new statistical visualization tools, and (2) show the behavior of His in our data set. The rest of the fragmentation maps presented in this paper do not contain the information for His, because His is excluded from the sequences to simplify interpretation of the possible chemical pathways.

The fragmentation patterns obtained in Figure 8IX_b and IX_y are almost identical to those previously reported,¹⁷ but more information is now provided. While b and y ions show very similar fragmentation patterns, the dominance of enhanced cleavage N-terminal to Pro indicates that one ionizing proton is at least

partially mobile, if not fully mobile. His shows enhanced cleavages at its C-terminal side (His–Xxx) in both ion types. Compared to peptides without His (III and IV), IX shows slight enhancement of cleavage C-terminal to Asp in y ions, which can be attributed to those peptides that contain a His, and agrees with a previous study.¹⁴ In addition, cleavages C-terminal to aliphatic residues, especially β-branched Ile and Val, are more enhanced in both IX_b and IX_y than III_y or IV_y.

CONCLUSIONS

Peptide dissociation patterns in MS/MS spectra are a complex function of the charge state, residue identities, sequence of the residues, and gas-phase structure. The results from this study elaborate how a change in the number of protons or AA content can lead to dramatically different fragmentation patterns in low-energy CID MS/MS of protonated peptides. In summary, analyses of a database of 28 330 spectra of unique sequence and charge states show that the most important factors influencing the fragmentation process are the mobility of the proton, the position and the basicity of the most basic residue, and the presence of Pro. When the proton is localized, cleavage C-terminal to acidic residues dominates. When the proton is mobile or partially mobile, cleavage N-terminal to Pro dominates. Specific fragmentation patterns found in b ions, y ions, or both are shown to depend on the location of the proton(s) or, more specifically, the availability and the relative position of the basic residue(s). The pairwise fragmentation map proves itself as a very useful tool in describing the different fragmentation behaviors from different sets of peptides. The overall statistical approach presented in this paper, including the computation of pairwise fragmentation statistics, the visualization tool, the FI threshold, the similarity test, and the distance measurement, can be applied to any ion type from any given set of spectra, as long as the number of unique spectra is large enough to give statistically valid results.

The y ion fragmentation maps from singly and doubly charged tryptic peptides (Figure 4I_y, 4II_y, 5III_y, and 5IV_y) explain why doubly charged tryptic peptides are more readily identifiable by current sequencing algorithms than singly charged tryptic pep-

tides (Figure 1). The behavior of doubly charged tryptic peptides resembles more closely the random cleavage model upon which the sequencing algorithms are based (At least modest cleavage happens at a variety of AA pairs.). Singly charged tryptic peptides deviate much more from such a model, with Arg-terminating peptides deviating more than Lys-ending peptides. This also explains why, in singly charged peptides, the number of identified spectra from Lys-ending peptides outnumber Arg-ending peptides (Figure 1). The Lys-ending peptides have more overall cleavage, producing more information for the search algorithm to use in comparison with candidate sequences. In doubly charged peptides, this difference in number of identified spectra diminishes (Figure 1) because Arg-ending peptides have behavior almost identical to Lys-ending peptides.

The great variability in relative cleavage efficiencies between different AA residue combinations in almost all our fragmentation maps shows that the chemical properties of the side chain play a significant role in determining how gas-phase peptides fragment. As we noted previously for peptides that have a mobile proton, the variability is dominated by enhanced or suppressed cleavage at those residues (Gly, Pro, pre-Pro) that have Ramachandran φ , ψ plots that differ from the general case, which suggests a steric influence on fragmentation.¹⁷ When Pro is not available in the sequence, those residues (Ile, Val) that occupy slightly lower percentages of total ψ , ϕ space than the other non-Pro residues³⁴ dominate fragmentation. High-order computational molecular modeling studies are desirable to further elaborate the possible "cleaving" structures for these residue combinations.

Note that the sorting of the data sets presented in this paper is heavily based on assumptions derived from prior chemical knowledge of peptide dissociation. Spectra that do not fall into the categories specified were not included in the analyses. This problem has been overcome in another study³⁵ using different data-mining techniques, e.g., unsupervised clustering and decision tree, to present the overall fragmentation behavior of the 28 330 spectra.

It would be interesting to look at the fragmentation maps of b^{2+} , y^{2+} from these data sets in a similar fashion. However, these doubly charged fragment ions are not abundant from doubly protonated precursor ions, and spectra from triply protonated peptides where these ions are abundant are limited in number. Extensive sorting of the spectra to study b^{2+} , y^{2+} fragmentation patterns often results in counts of pairwise residue combinations too low to be statistically meaningful. Studies of other ion types

(34) Kleywegt, G. J.; Jones, T. A. *Structure* **1996**, *4*, 1395–1400.

(35) Huang, Y.; Tseng, G. C.; Yuan, S.; Pasa-Tolic, L.; Lipton, M. S.; Smith, R. D.; Wysocki, V. H. Manuscript in preparation.

would also be desirable in future, e.g., characterizing water loss and ammonia loss peaks from peptides containing specific structural motifs.

Results from this study expand our knowledge of peptide fragmentation chemistry and can guide experimental fragmentation studies of model peptides, e.g., mechanistic studies on Ser-containing and Gly-containing peptides will be pursued based on the observation that cleavage at Ser and Gly occurs preferentially at their N-terminal amide bond but is suppressed at their C-terminal amide bond, especially for peptides where the proton is partially mobile. The fact that peptides with different chemical motifs display different fragmentation patterns should be considered in the fragmentation models for sequencing algorithms to improve the success rate of the algorithms; this work is in progress in our laboratory. Pairwise cleavage patterns will be utilized to allow the algorithms to predict fragment ion intensities given a candidate sequence.

ACKNOWLEDGMENT

We thank Li Ji in assisting spectra filtering. This work was financially supported by NIH Grant R01GM 51387 to V.H.W. and the 2003 Pfizer Graduate Research Fellowship in Analytical Chemistry to Y.H. Research at PNNL was partially supported by the U.S. DOE, Office of Biological and Environmental Research and the NIH National Center for Research Resources (RR018522). PNNL is operated by Battelle Memorial Institute for the U.S. DOE under Contract DE-AC06-76RLO 1830. The collaboration between Y.H. and G.C.T. was made possible by the UCLA Institute of Pure and Applied Mathematics and NSF, when Y.H. and G.C.T. participated in the IPAM spring 2004 Program "Proteomics: Sequence, Structure, Function", from March 8 to June 11, 2004.

SUPPORTING INFORMATION AVAILABLE

Detailed discussion of the Ion Statistics results shown in Tables 1 and 2, as well as Figure 2. Pairwise fragmentation maps for b ions from **III**, **IV**, **VIII**, as well as for y ions and b ions from 829 singly charged Arg-ending peptides without His or Pro ([...noP-...noH...R]⁺). This material is available free of charge via the Internet at <http://pubs.acs.org>.

Received for review December 24, 2004. Accepted June 3, 2005.

AC0480949