

Characterization of the Fragmentation Pattern of Peptide from Tandem Mass Spectra

Sangeetha Ramachandran and Tessamma Thomas*

Department of Electronics, Cochin University of Science and technology, Cochin-682022, Kerala, India

Received March 27, 2019; Revised April 26, 2019; Accepted May 08, 2019

First published on the web June 30, 2019; DOI: 10.5478/MSL.2019.10.2.50

Abstract : The fragmentation statistics of ion trap CID (Collision-Induced Dissociation) spectra using 87,661 tandem mass spectra of doubly charged tryptic peptides are analyzed here. In contrast to the usual method of using intensity information, the frequency of occurrence of fragment ions, with respect to the position of the cleavage site and the residues at these sites is studied in this paper. The analysis shows that the frequency of occurrence of fragment ion peaks is more towards the middle of the peptide than its ends. It was noted that amino acid with an aromatic and basic side chain at N- & C- terminal end of the peptide stimulates more peaks at the lower end of the spectrum. The residue pair effect was shown when the amide bond occurs between acidic and basic residues. The fragmentation at these sites (D/E-H/R/K) stimulates the generation of the y-ion peak. Also, the cleavage site H-H/R/K stimulates the generation of b-ions. K-P environment in the peptide sequence has more tendency to generate y-ions than b-ions. Statistical analysis helps in the visualization of the CID fragmentation pattern. Cleavage pattern along the length of the peptide and the residue pair effects, enhance the knowledge of fragmentation behavior, which is useful for the better interpretation of tandem mass spectra.

Keywords : Proteomics, MS/MS, CID, fragmentation pattern, residue dependence

Introduction

Tandem mass spectrometry is emerging as the most important tool for the analysis of complex protein samples.¹ Collision-induced dissociation (CID) is the commonly used activation technique that energizes the mass selected precursor peptide ion and induces the fragmentation at the amide bond. Sequencing of peptide relies on the knowledge of fragment ion produced. CID fragmentation causes cleavage at the peptide backbone leading to b-, y- and a- type ion that generates a set of (*m/z*) peaks in the mass spectrum. A better understanding of the fragmentation process can improve the interpretation of CID spectra which in turn serve to identify peptide from mass spectra.

Numbers of software are available to analyze the fragmentation pattern produced by CID, which either

matches the pattern to the peptide in sequence database or directly derives denovo peptide sequence from the mass spectrum.^{2,3} The efficiency of these algorithms highly depends on the quality of the product ion spectra. A number of research groups have studied the fragmentation mechanism relied on model peptides.⁴ The mobile proton model proposed by Wyoski et al.⁵ elucidates the fundamental theory behind the gas phase fragmentation. According to this model, charge from the initial site of protonation is transferred along the peptide backbone to an amide bond and subsequently fragmented.^{5,6} Kapp et al.⁷ further expanded the model into 'Relative Mobile Proton' (RMP) revealed that basic residue within the peptide hinders the mobility of the proton. Studies have proved that the amide bond fragmentation may enhance due to the presence of particular amino acid within the peptide.⁷⁻⁹ The comprehensive fragmentation pattern has been studied with statistical analysis of the relatively large data set, providing more insight into peptide fragmentation behavior.¹⁰ Data mining procedures were used to investigate a large set of data, predicted the intensity patterns of the spectra and discovered the influence of specific residues in the fragmentation process.^{4,5,7,9,10} Cleavage at the N- terminal to proline residue is found to be the most preferred event.^{4,7,9,11-13} Researchers have also observed an enhanced or suppressed cleavage near acidic and basic residues.^{4,7-10} In addition, enhanced cleavage at the C- terminal to the histidine triggers b- ion formation.⁷

Open Access

*Reprint requests to Tessamma Thomas
E-mail: tessamma1@gmail.com

All MS Letters content is Open Access, meaning it is accessible online to everyone, without fee and authors' permission. All MS Letters content is published and distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/3.0/>). Under this license, authors reserve the copyright for their content; however, they permit anyone to unrestrictedly use, distribute, and reproduce the content in any medium as far as the original authors and source are cited. For any reuse, redistribution, or reproduction of a work, users must clarify the license terms under which the work was produced.

Researchers have explored the intensity information from the mass spectra in order to investigate the fragmentation characteristics of mass spectra. The information obtained from the mass spectrum is not fully exploited in the previous studies and many new features can be obtained by analyzing large sets of data, this motivated the present study.

In this work, a comprehensive analysis of ion trap CID fragmentation mechanism of doubly charged tryptic peptides was performed. Relative frequencies of occurrence of fragmentation ion peaks were analyzed with respect to the position and the residue at the cleavage sites. Effect of position of cleavage site along the peptide backbone and influence of residues was investigated in each of the fragment ion types such as b-, y-, a- ions. Results from this study also confirm the previous observations of dependency of residue at the cleavage site on fragmentation.

Method

Data sets

Data sets were downloaded from the NIST peptide spectral library (NIST-chemdata.nist.Gov). It consists of spectrum segment pairs, where segments were identified by searching in the protein sequence database with accepted search programs.¹⁴ The proteins are digested by trypsin and the protonated tryptic peptide ions are fragmented using CID. Mass spectra are generated by electrospray ionization in LC-MS/MS experiments. From human spectral library with 340,357 ion trap spectra, doubly charged precursors with no peptide modifications and no missed cleavages comprising of peptide length of 6 to 21 residues were extracted for the study. Thus the data set used for this study contains 87,661 peptide mass spectra assigned with the corresponding peptide sequence and peak annotation details.

Methodology for analyzing fragmentation characteristics

A novel method has been developed for analyzing fragmentation characteristics of ion trap CID spectra using fragmentation statistics. The statistical approach was used and programs were written in MATLAB to analyze the effect of the residue cleavage site and its position along the peptide length on CID fragmentation. It calculates the effect that the residue/residue pair (n-Xaa-Xaa-c) at cleavage site on N- and C- terminal side along the peptide length, has on bond breakage and product ion observation. Initially, the entire possible residue pairs at the cleavage site together with its position along the peptide length were extracted from the peptide sequences in the data set. Twenty naturally occurring amino acids will create 400 residue pairs at the cleavage site. The data set has tryptic peptides, according to Keil rule,¹⁵ trypsin cleaves next to arginine or lysine and not before proline. Therefore, proline

is the only possible residue following arginine and lysine. Thus, 362 possible amide bonds of residue pair (Xaa-Xaa) exist at the fragmentation site in the data set. The maximum length of the peptide used for the study was 21, so there were 20 cleavage positions possible. Cleavage position from N- terminal to C- terminal extracted for b- ion and a- ion analysis and that from C- terminal to N-terminal extracted for y- ion analysis. These values were extracted from the entire set of 87,661 peptides in the data set. The number of times each residue pair at cleavage site was found at each position was calculated from the peptide sequences and was denoted as $N_{rp, p}^t$, where 'rp' is the residue pair at the cleavage site, and 'p' denotes its position along the peptide length or the number of amino acids present in the fragment.

The mass spectrum data have peaks assigned to the fragment ion types (b-, y- and a-ions) along with the number of residues present in corresponding product ions. From spectrum data, the fragmentation ion type, position or number of residues in the product ion was extracted. And derived corresponding residue pair at the cleavage site from the assigned peptide. The number of times a particular ion type observed for a given amide bond cleavage in the mass spectrum was calculated along with the position of amide bond and is denoted as $C_{rp, p}^t$. The relative frequency of occurrence of fragmentation at the given amide bond with respect to each position along the peptide length is given by

$$F_t(rp, p) = \sum C_{rp, p}^t / \sum N_{rp, p}^t \quad (1)$$

where 't' denotes the fragmentation ion type; b-, y- & a-ions. The frequency information provides the extent of fragmentation occurring both N- and C- terminal side of the residue in each position along the peptide backbone.

Results and Discussion

Investigation of large spectral dataset elucidates the peptide fragmentation behavior of CID spectra. Product ions produced by CID have a strong influence on the sequence of peptides analyzed. In order to visualize the dependencies of residue/residue pair at the cleavage site and the position of cleavage sites, different plots were drawn using relative frequency information.

Fragmentation characteristics with respect to the position of cleavage site along the peptide backbone

Initially, analysis primarily focused to learn the significance of the position of the cleavage site along the peptide backbone. Relative frequency of occurrence of fragment ion types; b-, y- and a- ions were calculated with respect to each cleavage position along the peptide backbone. The plot of the relative frequency of occurrence of fragmentation with respect to the position of cleavage

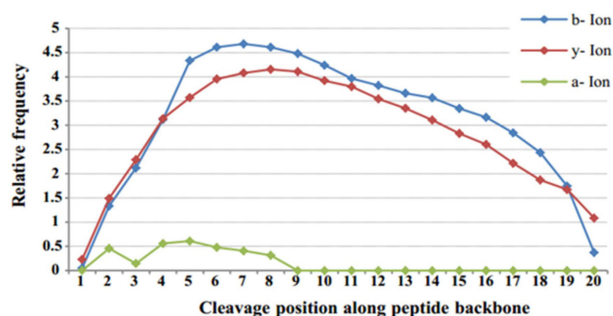


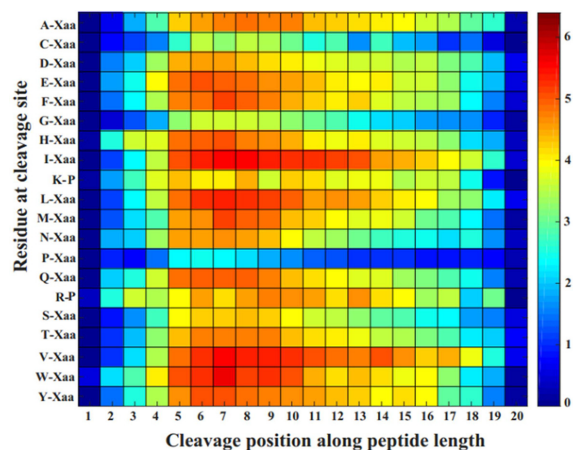
Figure 1. Relative frequency versus cleavage position. The x-axis represents the position of the cleavage site. The y-axis represents the relative frequency of occurrence of fragment ion peaks.

site for ion types; b-, y- and a- ions are shown in Figure 1. Here, b- & y- type fragment label also encompasses with ion type, isotope peak, and neutral loss peaks.

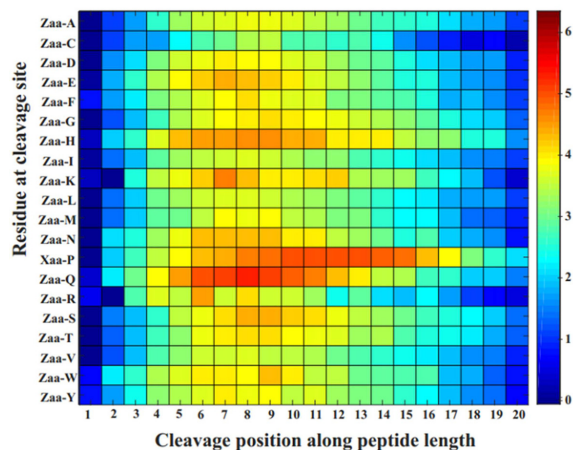
From the plot in Figure 1, it is clear that the fragmentation of peptide is more likely to occur towards the middle of the peptide rather than ends. The frequency of occurrence of peaks in the spectrum corresponding to b-ions are more than y-ions as the length of peptide increases or the number of amino acid present in the fragment ion increases. This is because b-ions are structurally less stable and have the tendency to fragment further by losing neutral molecules and causes more peaks in the mass spectrum.¹⁶ a- type ions are formed by the loss of mass of 28 mass units (corresponding to the mass of CO) from b- type ions. 'a-b' pairs are often observed in the tandem mass spectrum, which is used as an indicator for b-ions. a- ion peaks are less frequent than b- and y-ions. 'a-' ions are observed up to the position eight residues from the N- terminal end of the peptide. The scanty occurrence of a3- ion peak is also spotted, as the a3- ion tends to immediately fragment to b2- ion.¹⁷

A deficit of peaks at the lower end of the spectrum causes difficulty in complete sequencing of peptides. Hence, for extracting the features at these positions, the analysis of the dependency of residue at each fragmentation locations along the peptide backbone was studied and it is shown in the color map in Figure 2. The frequency of occurrence peaks of fragment ions are calculated with respect to the residue at the N- and C-terminal side of the cleavage site. The plot of b- ion and a- ion with respect to residue on the N- terminal side of the cleavage site is shown in Figure 2(a) and 2(c) respectively. The plot of y-ion with respect to residue on the C- terminal side of the cleavage site is shown in Figure 2(b).

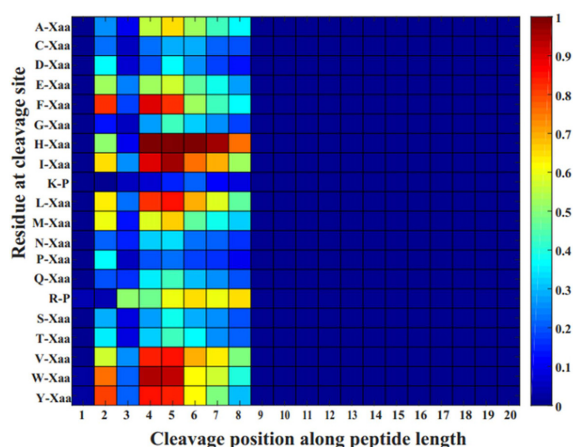
The results obtained from Figure 2 are discussed as follows. b1- ions are seldom observed and being tryptic peptides, y1 peaks are appearing when residues with basic side chain (R, K) on the C- terminal end of the spectrum



(a) b- ion; N- terminal residue



(b) y- ion; C- terminal residue



(c) a- ion; N- terminal residue

Figure 2. The color map of frequency of occurrence of residue specific cleavages at each position along the peptide backbone. The x-axis represents the position of cleavage site from the N- & C- terminal end of the peptide for b-ion and y-ion respectively. The y-axis represents the residue at the cleavage site. Xaa represents the 20 amino acids and Zaa represent the residues except K and R.

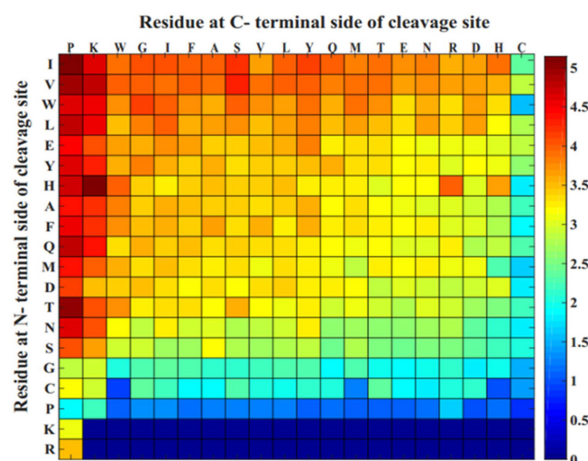
confirm the previous results.¹⁸ Also, cysteine shows a reduced tendency to produce fragment ions.¹⁹ Additionally, here the b1- ion appears in case of tryptophan and arginine at the N-terminal end of the peptide. From Figure 2(b), it is noted that a significant number of y1- ions appeared in case of residues with aromatic (F, W, Y) side chain on the C-terminal end of the peptide. Figure 2(c) shows aromatic residues on the N- terminal side of the peptide produce a-ions (a2-) on the lower part of the spectrum. Basic residue histidine and hydrophobic amino acids (F, W, Y, I, L, V, M) show a relatively significant number of fragment ion for a-ions. Thus, it is inferred that the residue with an aromatic or a basic side chain at the N- and C- terminal ends of the peptide stimulates more peaks at the lower end of the spectrum.

Fragmentation characteristics with respect to residue/residue pair at the cleavage site

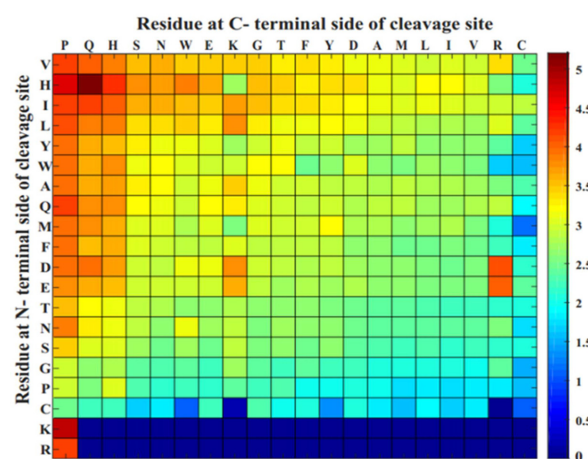
The degree of fragmentation at a given amide bond is found to have a dependency on the amino acid content in the peptide.⁷ In this study, the frequency of occurrence of fragmentation, at the N- and C- terminal side of each of the residue within the peptide sequence has been elucidated using equation 1. The frequency of occurrence peaks of b-, y- and a- ions are calculated with respect to the residue pair at cleavage site are plotted in colormap in Figure 3.

Similar to the previous results, here also observed a substantial enhancement in the frequency of occurrence of fragment ions when there is proline on the C- terminal side of the cleavage bond.¹⁰ Glycine at the N- terminal side of the cleavage site shows a characteristic reduction in the frequency of occurrence of b- and y- type ion. Also, enhanced cleavage preference is shown by branched aliphatic residues (I, L, V) and histidine at the N- terminal side of cleavage site.^{7,9} In addition to previous results, this analysis also shows that residues, tryptophan and tyrosine at the N- terminal side of cleavage site show enhanced frequency for b-, y- and a- ion. In addition, lysine at the C- terminal side of the cleavage site shows an enhancement in frequency for b- ion and a-ion. Histidine or glutamine at the C- terminal side of the cleavage site leads to a significant increment in the frequency of y-ion formation. The residue pair effect is shown by histidine when it has an amide bond with basic residues. The number of peaks for b- ion appears in mass spectra if the amide bond is between the basic residues: H-H, H-R, H-K. In the case of y- ion, H-H shows a higher frequency of occurrence while H-R and H-K show considerably lesser frequency. Amide bond of acidic and basic residue (D/E- H/R/K) create a number of y- ion peaks in the mass spectrum.

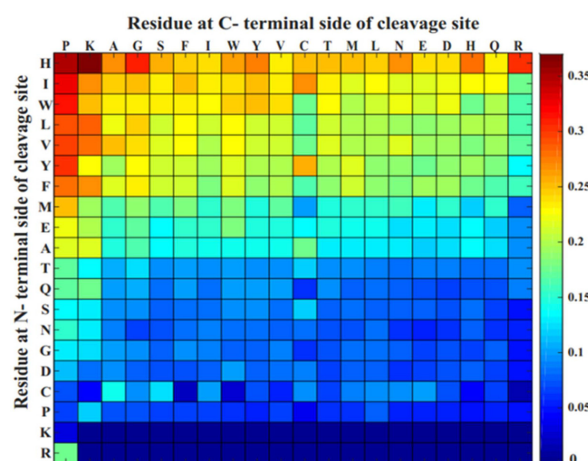
Figure 4 shows the bar graph of relative frequency information for b- and y- fragments with respect to residues at both N- and C- terminal side of the cleavage site. Four bar chart is drawn for each residue at the cleavage site in the order of N- & C-terminal residue for b-



(a) b- ion



(b) y- ion



(c) a- ion

Figure 3. The color map of frequency of occurrence of cleavages at each residue pair. X-axis represents the C- terminal residue of the cleavage site. The y-axis represents the N- terminal residue of the cleavage site. Figure 3(a), 3(b), 3(c). represents b- ion, y- ion and a-ion frequency information respectively.

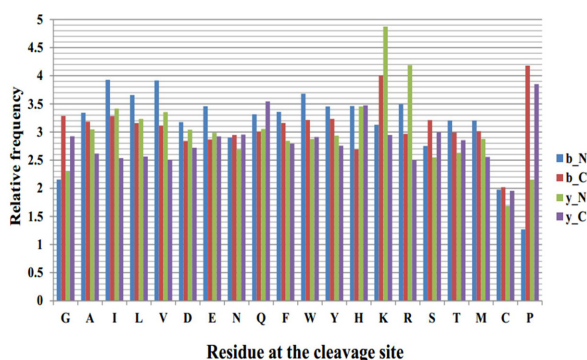


Figure 4. Relative frequency of b- & y- ion peaks when the residue is at the N- and C- terminal side of the cleavage sites. The x-axis represents the residue at the cleavage site. The y-axis represents the relative frequency occurrence of fragment ion peaks.

ion, N- & C- terminal residue for y- ion. Analyzing the influence of residues at N- & C- terminal ends of cleavage site for b- and y- ion, residues- P, G, S have increased bias on N-terminal cleavage confirms the previous result.²⁰ Furthermore, in case of lysine when it is placed on the N-terminal side of the cleavage site (K-Xaa), the frequency of y- ion is remarkably higher than that of b- ion. Since tryptic peptide has only K-P environment, and proline has increased the tendency to cleave at the N- terminus, K-P environment has less tendency to generate b- ions and more likely to produce y- ions peaks.

Conclusion

This study performed the statistical characterization of the fragment ion peaks in CID spectra of doubly charged tryptic peptides. A large set of data of 87,661 spectra from the peptide spectral library was used to identify the features influencing the CID fragmentation pattern. The study was based on spectral data from spectral library which contain high quality and well annotated spectra only (NIST *.msp spectral file). The study provides the relative frequency of occurrence of product ion peaks in the tandem mass spectra, which point out the existing and the novel statistical features of the CID spectra like residue specific and position specific cleavage preferences of the CID fragmentation pattern. The analysis shows that the frequencies of fragment ion peaks are higher towards the middle of the peptide length than the ends. The novel observations obtained from this study are as follows. An amino acid with an aromatic or basic side chain at either the N- or C- terminal end of the peptide stimulates more peaks of b₁-, y₁- & a₂- ions at the lower end of the spectrum. Lysine (K) at the C- terminal side of the cleavage site (Xaa-K) shows an increment in the occurrence of b- ion and a-ion frequency. Histidine (H) or

glutamine (Q) at the C- terminal side of the cleavage site (Xaa-Q) promotes a significant increment in the occurrence of y-ion. The amide bond occurred between acidic and basic residue (D/E-H/R/K) shows a residue pair effect in CID fragmentation. The fragmentation at these sites stimulates the generation of the y-ion peak. Residue pair- H-H at the cleavage site generates both b- & y- ions and cleavage at H-R/K generate more b- ions, and very less y-ion peaks. K-P environment has more tendency to generate y-ions than b-ions. This study provides the feature set consisting of the frequency of occurrence of product ion peaks in the mass spectrum, which provides residue specific and position specific cleavage preferences of the CID fragmentation pattern. This feature set gives more hints to distinguish the fragment ion peaks in the mass spectrum, which helps in predicting the fragmentation pattern of a particular peptide sequence.

Acknowledgments

The authors would like to acknowledge the financial support from the University Grant Commission, Government of India.

References

1. Aebersold, R.; Mann, M. *Nature* **2003**, 422, 198.
2. Diamant, B.; Noble, W. S. *J. Proteomics* **2011**, 10, 3871.
3. Dancik, V.; Dancik, V.; Addona, T. A.; Clauser, K. R.; Vath, J. E.; Pevzner, P. A. *J. Comput. Biol.* **1999**, 6, 327.
4. Huang, Y.; Tseng, G. C.; Yuan, S.; Pasa-Tolic, L.; Lipton, M. S.; Smith R. D.; Wysocki, V. H. *J. Proteome Res.* **2008**, 7, 70.
5. Wysocki, V. H.; Tsaprailis, G.; Smith L. L.; Brei, L. A. *J. Mass Spectrom.* **2000**, 35, 1399.
6. Boyd, R.; Somogyi, A. *J. Am. Soc. Mass Spectrom.* **2010**, 21, 1275.
7. Kapp, E. A.; 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.
8. Elias, J. E.; Gibbons, F. D.; King, O. D.; Roth F. P.; Gygi, S. P. *Nat. Biotechnol.* **2004**, 22, 214.
9. Huang, Y.; Triscari, J. M.; Tseng, G. C.; Pasa-Tolic, L.; Lipton, M. S.; Smith R. D.; Wysocki, V. H. *Anal. Chem.* **2005**, 77, 5800.
10. Tabb, D. L.; Smith, L. L.; Brei, L. A.; Wysocki, V. H.; Lin, D.; Yates, J. R. *Anal. Chem.*, **2003**, 75, 1155.
11. Martin, D. B.; Eng, J. K.; Nesvizhskii, A. I.; Gemmill, A.; Aerosol, R. *Anal. Chem.* **2005**, 77, 4870.
12. Schütz, F.; Kapp, E. A.; Simpson R. J.; Speed, T. P.; *Biochem. Soc. Trans.* **2003**, 31, 1479.
13. Raulfs, M. D. M.; Brei, L.; Bernier, M.; Hamdy, O. M.; Janiga, A.; Wysocki V.; Poutsma, J. C. *J. Am. Soc. Mass Spectrom.* **2014**, 25, 1705.
14. Dataset: NIST Library of Peptide Ion Fragmentation

- Spectra, **2008**, NIST/EPA/NIH Mass Spectral Library: <http://chemdata.nist.gov/>.
15. Rodriguez, J.; Gupta, N.; Smith R. D.; Pevzner, P. A. *J. Proteome Res.* **2008**, 7, 300.
 16. Lau, K. W.; Hart, S. R.; Lynch, J. A.; Wong, S. C. C.; Hubbard, S. J.; Gaskell, S. J. *Rapid Commun. Mass Spectrom.* **2009**, 23, 1508.
 17. Allen, J. M.; Racine, A. H.; Berman, A. M.; Johnson, J. S.; Bythell, B. J.; Paizs B.; Glish, G. L. *J. Am. Soc. Mass Spectrom.* **2008**, 19, 1764.
 18. Medzihradszky, K. F.; Chalkley, R. J. *Mass Spectrom. Rev.* **2015**, 34, 43.
 19. Khatun, J.; Ramkissoon, K.; Giddings, M. C. *Anal. Chem.* **2007**, 79, 3032.
 20. Loo, J. A.; Edmonds C. G.; Smith, R. D. *Anal. Chem.* **1993**, 65, 425.