

Published in final edited form as:

Int J Mass Spectrom. 2011 December 1; 308(2-3): 142–154. doi:10.1016/j.ijms.2011.04.008.

Investigation of VUV Photodissociation Propensities Using Peptide Libraries

Xiaohui Liu^{*}, Yong Fuji Li⁺, Brian C. Bohrer^{*}, Randy J. Arnold^{*}, Predrag Radivojac⁺, Haixu Tang⁺, and James P. Reilly^{*,#}

Department of Chemistry, Indiana University, Bloomington, Indiana 47405, and School of Informatics and Computing, Indiana University, Bloomington, Indiana 47408

Abstract

PSD does not usually generate a complete series of γ -type ions, particularly at high mass, and this is a limitation for *de novo* sequencing algorithms. It is demonstrated that b_2 and b_3 ions can be used to help assign high mass x_{N-2} and x_{N-3} fragments that are found in vacuum ultraviolet (VUV) photofragmentation experiments. In addition, v_N -type ion fragments with side chain loss from the N-terminal residue often enable confirmation of N-terminal amino acids. Libraries containing several thousand peptides were examined using photodissociation in a MALDI-TOF/TOF instrument. 1345 photodissociation spectra with a high S/N ratio were interpreted.

Keywords

De Novo Sequencing; Photodissociation; Mass Spectrometry; MALDI-TOF/TOF

Introduction

An important component of characterizing biological systems is protein identification.^{1–3} Peptides are typically generated by enzymatically digesting proteins and analyzed using a variety of chromatographic and mass spectrometric (MS) methods.^{4–7} Most often, the products of peptide ion fragmentation are identified by comparison with predicted fragments derived from protein sequence databases. In this approach, precursor masses are used to pre-filter peptide candidates. There are several database search engines available such as Mascot,⁸ Sequest,⁹ X!Tandem,¹⁰ Spectrum Mill,¹¹ Omssa¹² and others^{13–20}. One limitation to this method is that it is only applicable to organisms with sequenced genomes. Incomplete or incorrect databases can preclude identification or lead to errors.²¹ In addition, peptides with unexpected post-translational modifications (PTM) are a challenge since their precursor masses are shifted. The study of post-translational modifications is important for investigating the functions of proteins. It is generally recognized that *de novo* sequencing can alleviate some of these problems. In this approach, peptide sequences are directly derived from mass spectrometric data without the use of a protein database. Mass differences between appropriately chosen peaks in spectra are measured and interpreted.

© 2011 Elsevier B.V. All rights reserved.

[#]To whom correspondence should be addressed. reilly@indiana.edu.

^{*}Department of chemistry

⁺Department of Informatics and Computing

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Several *de novo* sequencing algorithms such as PEAKS, PepNovo, and MSNovo have been developed to interpret low-energy CID spectra.^{22–28} Since they rely on mass differences, a complete series of fragment ions is essential. Unfortunately, most CID spectra do not provide complete sequence information due to preferential bond cleavages and spectra are often dominated by a limited number of peaks.²⁹ This makes *de novo* sequencing difficult. Likewise, different types of ion fragments must be distinguished for this approach to be successful. Incorrect fragment assignments can lead to peptide sequence errors. Finally, isomeric leucine and isoleucine, which together typically represent more than 16% of the amino acids in a database, are not distinguished by low-energy fragmentation.³⁰

Various fragmentation techniques have been developed to generate more complete sequencing information.³¹ For example, electron capture dissociation (ECD) and electron transfer dissociation (ETD) are now widely used.^{32,33} These two radical-induced processes generate a series of c- and z ions.^{34,35} The fragments from ECD and ETD can be employed for *de novo* sequencing since they extend through entire sequences except N-terminal to proline.³⁶ C-type ions are 17Da heavier than corresponding b-type ions and y-type ions are 16Da heavier than z-type ions. Therefore, the combination of CID with ECD or ETD data yields b/c and y/z ion pairs that can confirm peak identifications and improve the confidence of sequencing results.^{37–39} However, ECD and ETD require multiply-charged ions formed by ESI and not by MALDI. High-energy CID fragmentation is an alternative that improves fragmentation yields and produces more fragment ions than low-energy CID. High-energy CID also produces side chain loss fragments including w-type ions that distinguish leucine and isoleucine.^{40,41} However, high-energy CID also generates an abundance of internal fragments particularly in the low-mass region and sequence information is still typically incomplete. In MALDI TOF-TOF experiments, the most common fragmentation process is post source decay (PSD) that involves precursor ion fragmentation during its flight in the mass spectrometer.^{42,43} PSD typically generates a-, b- and y-type ions. Because it is not very efficient, PSD rarely provides a complete series of peptide fragments.

We have shown that 157nm photodissociation generates abundant high-energy peptide ion fragments along the entire backbone.^{44–49} It produces a series of x-type ions for peptides with C-terminal arginine or C-terminal lysine, if the lysine is converted to homoarginine by guanidination.^{44,50,52} Abundant v- and w-type ions are also observed enabling leucine and isoleucine to be distinguished. Photodissociation has been implemented in a commercial MALDI-TOF/TOF instrument and high quality spectra are generated with this sensitive mass spectrometer.⁵⁰ Postsources decay (PSD) spectra can be recorded with the photodissociation light turned off. A *de novo* sequencing algorithm that combines photodissociation and PSD data has been developed.⁵¹ It exploits relationships such as the 26Da spacing between x and y-type ions to identify x/y ion pairs. Mass differences between adjacent x-type ions lead to peptide sequence information. The identification of corresponding v- and w-type ions increases the confidence of sequence assignments. The algorithm has been successfully applied to the identification of *Deinococcus radiodurans* ribosomal proteins.⁵² Five or more residues were sequenced in 298 ribosomal peptides using this sequencing algorithm. One shortcoming that was uncovered is that the N-terminal regions of tryptic peptides are often incompletely sequenced. This is a consequence of the limited number of high-mass y-type ions that are detected leading to a limited number of high-mass x/y pairs. In order to improve the *de novo* sequencing algorithm and further investigate photodissociation propensities, three libraries containing several thousand peptides were synthesized. The libraries contain peptides with 10 to 12 residues all with C-terminal arginine or lysine. Each of the three peptide library mixtures was chromatographically separated and analyzed in both photodissociation and PSD modes in our MALDI TOF-TOF instrument. Since the MALDI ionization efficiency of lysine-terminated peptides is not as high as it is for arginine-terminated peptides, very few of the

former were detected, and none of that data is included in this study. Although guanidination of lysine eliminates these problems,⁵² for the present mechanistic study, it was simpler to just analyze the arginine-terminated peptides. Because the library peptide sequences are known and there are no post-translational modifications, the number of false positive identifications should be limited and the effect that various residues have on photodissociation propensities can be directly probed. The observations of low-mass b-type ions and high-mass v- and w-type ions that are discussed below should lead to improvements in peptide *de novo* sequencing.

Experiment

Materials

Acetonitrile (ACN) and trifluoroacetic acid (TFA) were obtained from EMD chemicals (Gibbstown, NJ). α -cyano-4-hydroxycinnamic acid (CHCA) was purchased from Sigma (St. Louis, MO) and Protea Biosciences (Morgantown, WV). Phenylacetamidomethyl (PAM) resin beads were bought from Midwest Biotech (Indianapolis, IN). 3-(Diethoxyphosphoryloxy)-3*H*-benzo[*d*][1,2,3]triazin-4-one (DEPBT) and *N*-*tert*-butoxycarbonyl (Boc)-protected amino acids were obtained from National Biochemicals (Twinsburg, OH).

Peptide Synthesis

Three peptide libraries ranging from 10 to 12 residues in length were synthesized.⁵³ The sequences of components are listed in Table 1. A total of 13888 sequences were represented in the three libraries. The hydrophobicity distribution of each library was designed to imitate the peptides from human protein tryptic digests.⁵³ The peptides were synthesized following common solid-phase synthesis and split-and-mix techniques.⁵⁴ The synthesis started with PAM resin beads with arginine preloaded. 10-fold molar excess of Boc-protected amino acids and DEPBT was added to the vessel containing PAM resin beads to assure that all the resin was coupled with the appropriate amino acid. Then the beads were divided into equal portions and transferred to separate vessels in order to incorporate multiple residues at one position. The coupling of the next amino acid took place in individual vessels with the presence of Boc-protected residues and DEPBT at a 10-fold molar excess. When the synthesis was complete, peptides were cleaved from the resin and lyophilized.

LC separation and MALDI spotting

The three synthetic libraries were chromatographically separated and analyzed respectively. Synthetic peptides were separated by reverse phase chromatography (Eksigent nanoLC-2D system, Dublin, CA) and spotted onto MALDI plates using a robotic spotter (Eksigent, Dublin, CA). Mobile phase A was 100% water with 0.1% TFA and mobile phase B was 100% acetonitrile with 0.1% TFA. The spotter combined matrix solution containing 10g/L CHCA in 70% acetonitrile, 30% water and 0.1% TFA with the reverse phase effluent. The flowrates of the nanoLC and matrix solution were 600nL/min and 300nL/min respectively. 9 μ L of 0.2mg/mL sample was loaded into the column (~0.3 pmol of each peptide). 0.6 μ L spots were deposited every 40 seconds. Most peptides were deposited onto one or two spots, so each spot contain 0.1–0.3 pmol of each peptide. The reverse phase gradient lasted for 120 minutes.

MALDI TOF-TOF mass spectrometry

Samples deposited on MALDI plates were analyzed by our modified 4700 MALDI TOF-TOF instrument (Applied Biosystems, Framingham, MA) that has previously been described.⁵⁰ Mass spectra were first collected using the reflectron-TOF. Photodissociation was performed by vacuum ultraviolet light (VUV) from an F₂ laser (Coherent Lambda Physik, Germany). Peptide PSD spectra were separately collected with the VUV light turned

off. In either photodissociation or PSD experiments, the 10 most intense precursor ions with S/N above 80 from each MALDI spot were selected and isolated by a timed ion gate. Precursor ions, photofragments and PSD fragments all contributed to photodissociation spectra. A total of 2000 shots were averaged for each photodissociation and PSD MS/MS spectrum. For *de novo* sequencing, peak lists were extracted from both sets of data. The threshold for the software's automated peak detection was set at a signal to noise ratio of 10.

De novo Sequencing

A *de novo* sequencing program was previously developed to derive peptide sequences from photodissociation and PSD data.^{51,52} The *de novo* sequencing algorithm combines peak lists from photodissociation and PSD data to identify x/y ion pairs. As noted above, for the lysine terminated peptides, the ionization yield in MALDI source is low and thus the precursors of these peptides are low in abundance. Therefore, no attempt was made to sequence lysine-terminated library peptides. In a previous investigation we photodissociated ribosomal peptides from *Deinococcus radiodurans* and these data provided access to some sequences not present in our peptide libraries.⁵² These peptides were guanidinated so they yielded intense MALDI signals. Two examples of these spectra are included in the present study. For the peptides with C-terminal arginine, the spectra contain numerous x- and y-type ions. Although y ions appear in both PD and PSD spectra, x ions only occur in the former. The program looks for x-type ions from photodissociation data that are 25.98Da heavier than y-type ions found in PSD spectra. The spacing between adjacent x-type ions identifies an amino acid. The presence of corresponding v-, w- and y-type ions enables the confidence of sequence assignments to be estimated. If an x/y ion pair is missing, a gap is left in the sequence. A homology search program was developed to identify peptides by matching partial *de novo* sequences against a database that included all peptides in the library. Precursor masses with a tolerance of 0.3Da are used to confirm homology matches. 80–90% of the peptide residues are typically derived with the *de novo* sequencing algorithm. Fragment ion mass accuracy was assumed to be 0.3Da.

Mascot Search

The database search engine Mascot was also used to match photodissociation data with the peptide library sequence database assuming tryptic digestion with no missed cleavages and the same 0.3Da mass error. Fragments including a-, b-, x-, y-, y-NH₃-, y-H₂O, w- and v-type ions were all considered by Mascot. Peptides with Mascot scores above 40 were considered to be confident identifications.

Candidate b-x ion pairs

There are two steps to identify candidate b₂/x_{N-2} and b₃/x_{N-3} pairs. First, the 20 amino acids lead to 400 or 8000 different possible b₂ or b₃ sequences. However, leucine and isoleucine are isomers and some b₂ or b₃ ion masses with different sequences differ by less than 0.0001Da and for all practical purposes their masses can be considered identical. The error tolerance to assign b₂ or b₃ ions was set as 0.1Da since low-mass b ions are usually measurable to about 0.05Da. Only 177 and 1484 b₂ and b₃ ion masses that differ by more than 0.0001Da exist. The observed peaks in a spectrum are compared with these 177 or 1484 precomputed theoretical mono-isotopic m/z values. When a peak mass matches one of these numbers to within 0.1Da, it is considered as a candidate b₂ or b₃ ion. The sum of the mass of a b_n ion and its complementary x_{N-n} ion is 26.99Da larger than the precursor mass. This relation is used to search for a tentative x_{N-n} ion fragment. An attempt was made to estimate how often false b-x ion pairs were detected using this approach. Spectra were divided into two groups based on Mascot peptide identifications. In the first group, Mascot identified b₂/x_{N-2} or b₃/x_{N-3} pairs and in the second group of spectra, none were found. Results of comparing the two groups to estimate the number of random matches are discussed below.

Results and Discussion

De novo sequencing

Each synthetic library was chromatographically separated and MALDI ionized respectively. Ion fragmentation occurred from post-source decay and was induced by VUV photodissociation. Figure 1 shows MS/MS spectra for a typical library peptide, EGADLSIVTR generated by these two processes. The PSD spectrum contains low-energy fragments including a-, b- and y-type ions. In addition to these, photodissociation produces x-, v- and w-type ions. (Note that with our apparatus, PSD products necessarily appear in photodissociation spectra, sometimes with higher and sometimes with lower intensities than in the PSD spectrum). Isoleucine and threonine each have two different β -substituents on their side chains and as a result can produce two types of ions- w_a and w_b . However, the latter is normally much lower in intensity than the former. This means that the ethyl group loss of isoleucine and the hydroxyl group loss of threonine are usually favored in photodissociation, and this is the case in Figure 1. PSD yields y_1 to y_6 as well as y_9 ions. However, y_7 and y_8 are missing. Consequently our *de novo* algorithm recognizes only six consecutive x/y ion pairs along with the x_9 , y_9 pair leading to a partial sequence of [372.230]LSIVTR. In this notation, [372.230] represents the mass gap between the sequenced portion of the peptide and the precursor mass. Observation of v_2 , v_3 and v_5 ions confirms threonine, valine and serine assignments. The w_4 and w_6 fragments also distinguish leucine and isoleucine in the peptide. Nevertheless, the program fails to identify the four N-terminal residues. Thus, although photofragmentation provides a complete series of x-type ions from x_1 to x_9 , the absence of high-mass y_7 and y_8 PSD ions that would confirm the identities of x_7 and x_8 leads to a sequencing gap. This is consistent with our previous observation that high-mass y-type ions are often missing.⁵² Other data interpretation strategies are therefore needed to confirm the identities of large x-type ions.

Synthetic peptide libraries

The three synthetic libraries of peptides were studied to extract photofragmentation rules that might improve *de novo* sequencing. The synthetic libraries contain equal numbers of peptides with C-terminal arginine and lysine. However, as noted above, data for only arginine-terminated peptides were interpreted in this study. Although it would have been helpful to have spectra with intense fragments, each synthetic library contained thousands of peptides with similar sequences and the same length (Table 1), so the peptides could not be completely separated. As a result, many peptides with similar masses were deposited on the same MALDI spot and ionization competition led to some low precursor ion intensities. Photodissociation spectra generated from these precursors contained correspondingly weak peaks. Because each MS/MS spectrum was recorded by averaging 2000 laser shots, repeating this process for the ten most intense precursor ions largely consumed a MALDI spot. Therefore, the last few photodissociation spectra from each spot tended to be mediocre. Moreover, the automated process of depositing MALDI spots could split some peptide samples onto adjacent spots, also weakening precursor ion signals. Thus, while synthetic peptide libraries are in principle attractive for studying peptide fragmentation, the experiment was still challenging. Reflecting this, when Mascot was applied to interpret photodissociation MS/MS spectra, only 1345 of the 6944 arginine-terminated peptides in the three libraries were identified with scores above 40. The 1345 spectra were used to derive some new insights about photofragmentation. These include a new appreciation for the ubiquity of low-mass b-type ions in VUV photodissociation spectra, and the surprising observation of v_N - and w_N -type ions (where N is the number of residues in the peptide). Similarly, the library data hint at an unexpected enhanced fragmentation of peptides next to threonine but only when it is close to the peptide N-terminus. Finally, Y-type ions (that are 2 Da lighter than their y-type counterparts) were found to be much more abundant in

photofragmentation data than had been previously reported. A complete discussion of these observations follows.

Assignment of high-mass x-type ions using b_2 and b_3 ions

b_2 ions are commonly formed by low-energy activation of peptide ions.^{55, 56} Figure 2 shows an example of PSD-generated b_2 and b_3 ions. In this case, their intensities increase with photofragmentation. In the 1345 photodissociation spectra of synthetic peptides identified by Mascot, numerous small b ions appeared. Table 2 indicates that 1337 and 1297 of these spectra contain b_2 and b_3 ion peaks, respectively. Over 96% of photodissociation spectra contain both. The b_2 and b_3 ion intensities depend somewhat on peptide sequence. In our synthetic libraries, 14 and 32 different sequences for b_2 and b_3 ions, respectively, were represented. In addition, the libraries contain a variety of different third and fourth residues that could affect b_2 and b_3 ion fragmentation. Intensities of b_2 and b_3 ions were extracted from our 1345 peptide spectra and normalized with respect to the sum of the intensities of all assigned peaks. Results are displayed in Figure 3. The numbers of identified b_2 or b_3 ions for each sequence are also noted in the Figure. The peptide detection exhibits some sequence preference probably due to ionization competition. Therefore, some N-terminal sequences are rarely detected. For example, very few peptides with N-terminal sequences NTM and GND were found resulting in very few b_2 ions. Likewise peptides with N-terminal sequences NTMI, NTMD, GNDS, ETMD, ATAH, AQIH and AQEH yielded almost no b_3 ions. However, fragmentation propensities can still be derived from other sequences. Figure 3a displays 32 different b_2 ions, their sequences and the number of different peptides from which they derived. It is obvious that peptides with glycine or threonine on the second residue from the N-terminus show relatively low b_2 ion intensities compared to glutamine, especially when isoleucine is next to glutamine. However, b_3 ion intensity seems to depend more on the fourth residue in the peptide (Figure 3b). No peptide with GLDS sequence at its N-terminus was found with a Mascot Search score above 40 even though 128 peptides in the libraries contain this N-terminal sequence. 75 identified b_3 ions are listed in Figure 3b. Bars for which standard deviations are not noted correspond to cases for which only a single peptide was detected. Peptides containing aspartate or histidine at the fourth position produce relatively weak or no b_3 ions.

It has been noted above that missing high-mass x/y pairs lead to *de novo* sequencing failures even when photodissociation produces complete x-type ion series. Table 2 indicates that 1076 out of 1345 confidently detected peptides yielded x_{N-2} photofragments, while only 318 yielded corresponding y_{N-2} ions (where N is the length of the peptide). Analogously, x_{N-3} ions were observed for 1125 out of 1345 peptides but y_{N-3} ions were detected in only 485 cases. Because of their high abundances, b_2 and b_3 ions can help identify x-type ions. It is well known that the sum of a b_n and a complementary y_{N-n} ion is the precursor mass plus 1.01 Da. Likewise, the sum of a b_n -type ion and its complementary x_{N-n} -type ion is 26.99Da larger than the precursor mass. This relation can therefore be applied to identify corresponding high-mass x-type ions. As indicated in Table 2, in these 1345 photodissociation spectra, 1053 b_2/x_{N-2} and 1056 b_3/x_{N-3} ion pairs can be recognized, while only 247 x_{N-2}/y_{N-2} and 368 x_{N-3}/y_{N-3} ion pairs are found. This suggests that the identification of b_2/x_{N-2} and b_3/x_{N-3} ion pairs can be much more effective for high-mass x-type ion confirmation. However, to exploit these ion pairs for sequencing, they must first be correctly identified. Random combinations of peaks can possibly satisfy the above mass relations. Internal fragment ions of a given sequence have the same mass as b_2 and b_3 ions of that sequence. In order to determine how often random ion pairs satisfy the relation $b_n + x_{N-n} = M + 26.99$ (where M is the precursor mass), a program was developed to search for the b_2/x_{N-2} and b_3/x_{N-3} ion pairs in photodissociation data. The 1345 spectra with confident Mascot identifications were examined. All low-mass fragment ions were compared to a list

of the 177 unique b_2 and 1484 b_3 ion masses mentioned above. All matches were included in a list of potential b_2/b_3 candidates. The program then tested each one of these ions by looking for complementary high-mass peaks that could be summed with the candidate b_2/b_3 ions to yield the precursor mass plus 26.99Da. The numbers of *false* b_2/x_{N-2} and b_3/x_{N-3} ion pairs uncovered by this procedure are plotted in Figures 4a and 4b respectively. In each case, the abscissa is the number of false matches and the ordinate shows the number of peptides (i.e. spectra) that contained this number of false matches. It can be concluded that more than 90% of the 1345 spectra have four or fewer false b_2/x_{N-2} matches and over 700 spectra have zero or one false identification. Likewise, about 90% of the spectra yield five or fewer false b_3/x_{N-3} matches and about half have two false matches or fewer. One ameliorating factor is that in this peptide library study, more than one precursor may be contributing to some spectra due to the wide ion isolation window. In that case, some spectra may have more than one *legitimate* b_2/x_{N-2} or b_3/x_{N-3} ion pair. Although this procedure has found multiple false b/x ion pairs, one other check can be applied to the data: the x_{N-1}/x_{N-2} , x_{N-2}/x_{N-3} and x_{N-3}/x_{N-4} mass spacings should match the mass of some amino acid. Application of these constraints will increase the number of valid x ion identifications and reduce the number of false positives.

It has been reported that accurate masses of dipeptides can be used to extract sequence information for peptide identification.⁵⁷ In the present case, the residues in potential b_2 ion candidates can be identified if the masses of these fragment ions are measured accurately enough. All the peptides in our synthetic libraries are arginine-terminated and all of the spectra contain a 175.117Da y_1 ion. Therefore, this mass can be used as an internal calibrant. The Data Explorer software from ABSciex was capable of recalibrating the spectra using a peak assigned with an accurate mass. Since the calibrant mass is 175.117Da, the recalibration particularly improves the mass accuracy of the low-mass region. We investigated the accuracy with which b_2 and b_3 ions could be measured before and after spectral recalibration using y_1 as internal standard. The results for 100 typical spectra are displayed in Figure 5. Before recalibration, only about half of the b_2 and b_3 measured masses were within 0.02Da of their correct values. After y_1 recalibration, about 95% of b_2 and b_3 ions were within 0.02Da and almost all of these ions were within 0.03Da mass error. This recalibration increases the mass accuracy of MS/MS ion masses and thereby improves the possibility of predicting peptide N-terminal residues.

Identification of x_{N-1} , w_N , v_N and v_{N+1} ions

Photodissociation generates a series of x -type ions in the spectra. The mass spacing between x_{N-1} and the precursor is unique for every N-terminal residue. Therefore, the presence of x_{N-1} provides an opportunity to identify the peptide N-terminus. Over 75% of the 1345 library spectra contained x_{N-1} ions, although as shown in Table 3, the probability was quite residue-dependent. The mass spacings between each x_{N-1} and the precursor are listed in Table 4. Some masses spacings in this table are not unique for different N-terminal residues. However, the peaks in the high-mass region of photodissociation spectra are as redundant as in the low-mass region. Therefore, these mass spacings can be used to tentatively assign x_{N-1} ions and the N-terminal residues.

Photodissociation of various peptides in our library yielded w_N , v_N or v_{N+1} ions (where N is the number of residues) that are associated with side chain loss from the N-terminal residue. Some examples are shown in Figure 6. Observation of these ions was somewhat surprising since it has been proposed that v - and w -type ions are formed from x ion precursors (Schemes 1a and 1b)⁴⁵ and it is not possible to form x_N ions. Since v_N and w_N ions are generated from unique side chain losses that are residue-dependent, the mass spacings between these v_N or w_N type ions and the precursor should facilitate confirmation of peptide N-terminal residues. To investigate this, our library peptides with six different N-termini,

Asn, Glu, Ala, Tyr, Phe and Gly were examined. The top part of Table 3 indicates the percentage of the 1345 identified spectra that contain five different ion fragments. N-terminal glycine can produce an x_{N-1} , but not a v_N or w_N ion. It is apparent that different N-terminal residues produce different types of v_N - or w_N -type ions. In our spectra, N-terminal phenylalanine did not lead to a v_{N+1} radical, and N-terminal asparagine was the only residue that yielded a w_N ion. 285 of our confidently identified peptides have N-terminal glutamic acid. 90% of these generate v_{N+1} radicals after photodissociation and about 42% yield v_N ions. Although only about 40% of these peptides produce both v_{N+1} and v_N ions, the v_N ion intensity is typically three times weaker suggesting that the former is the dominant product. Figure 6a presents the high-mass region of a PD spectrum displaying an intense v_{N+1} radical ion and relatively weak v_N ion. The peaks marked with asterisks predominantly result from ^{13}C isotope contributions and they are not treated as different fragments by the instrument software. 69% of peptides with N-terminal asparagines produce w_N ions and 46% of them generate v_N ions but only 33% of them yield both w_N and v_N fragments. Figure 6d displays one example with N-terminal asparagine. Table 3 indicates that 15% of these peptides yield v_{N+1} radical ions. However, the ^{13}C isotope peak associated with the adjacent v_N ion occurs at this same mass, so asparagine's v_{N+1} ions may well be artifacts that will not be helpful for peptide sequencing. Interestingly, tyrosine and phenylalanine have similar aromatic side chains but peptides with N-terminal phenylalanine only generate v_N ions and peptides with N-terminal tyrosine produce both v_N and v_{N+1} ions by photodissociation.

Whether or not a v_{N+1} ion can be used to identify an N-terminal residue depends on the uniqueness of its mass (as listed in Table 4) and whether it can be distinguished from side chain loss m-type ions that would have the same mass. The latter can result from side chain loss from any residue in the peptide.⁴⁴ For example, peptides with N-terminal glutamic acid produce strong v_{N+1} radical ions. The m ion formed from non-terminal glutamic acid side chain loss would have the same mass, and this would confuse v_{N+1} radical ion identification. In our synthetic library study, only the peptides containing glutamic acid and aspartic acid are found to have m ions from photodissociation. We compared the intensities of m ions from interior glutamic acid side chain loss with those of v_{N+1} ions from peptides with N-terminal glutamic acid. Over half of the peptides with non-terminal glutamic acid yielded m ions in photodissociation but their intensities were 10–20% of those associated with v_{N+1} radical ions. Because m ions from non-terminal glutamic acid side chain loss are much weaker than the v_{N+1} ions from the N-terminal glutamic acid side chain loss, the latter should be recognizable by sequencing programs. N-terminal aspartic acid was not included in our libraries, and this residue requires further study.

Because our synthetic peptide libraries contain only six different N-terminal residues, photodissociation data of tryptic peptides from *D. radiodurans* ribosomal proteins were also briefly examined. Peptides with 16 different N-termini were confidently identified using Mascot. In this sample set, only one to ten peptides for each N-terminal residue were found, and so accurate appearance percentages for various fragment ions were not derived. The lower section of Table 3 simply indicates whether or not v_N , w_N or v_{N+1} ions were observed in the examples that were studied. No peptides with N-terminal arginine, cysteine, lysine and proline were in this sample set. v_N ions were generally observed except for peptides with N-terminal methionine. It can be seen that some residues yield v_{N+1} radical ions but others do not. Figures 6b and 6c display examples of PD spectra of peptides with N-terminal aspartic acid and glutamine. Glutamic acid and glutamine terminated peptides yield the same weak v_N ions and intense v_{N+1} ions. By comparison, aspartic acid and asparagine terminated peptides produce strong w_N and v_N ions (Figures 6b and 6d). The side chains of glutamic acid and glutamine are only one CH_2 longer than those of aspartic acid and asparagines and the w_N ions at the latter two are apparently stabilized by formation of a

stable CO₂ side product. It is less obvious what affects the v_{N+1}/v_N peak ratios but the possibility of conjugation in the glutamic acid side product may be involved. Scheme 1c displays how v_N or v_{N+1} fragment ions form.

Scheme 2 proposes a mechanism for producing w_N and v_N ions from N-terminal aspartic acid and asparagine. A hydrogen bond between the N-terminal amine and the side chain enables the formation of a six-member ring. If the hydrogen atom from the amine is bridged to the oxygen or nitrogen of the adjacent side chain, a v_N ion is produced following elimination of CH₂CO and either H₂O or NH₃ as shown in Scheme 2b, 2d. If the hydrogen from the side chain is bridged to the nitrogen on the N-terminus, then w_N ions form (Scheme 2a, 2c). Because glutamic acid and glutamine contain one more CH₂ and a seven-member ring is less stable, w_N ions are not formed and v_N ions are not the primary product. Instead, light activation breaks the N-terminal side chain C_α-C_β bond, the side chain group is eliminated as a radical (that probably loses a hydrogen to form a conjugated species) and the peptide is left as a v_{N+1} radical ion. Presumably N-terminal residues other than aspartic acid and asparagine lose their side chains to form either v_N and v_{N+1} ions as shown in Scheme 1c. Interestingly, asparagine and aspartic acid are the only two N-terminal residues that yield w_N ions (Table 3). Even though the mass difference between these w_N ions and precursor is not unique (as shown in Table 4), based on other information derived from the *de novo* program, the presence of w_N ions for N-terminal aspartic acid or asparagine should help to identify these N-terminal residues.

Peptides with N-terminal threonine and serine produce strong v_{N+2} ion peaks as seen, for example, in Figures 6e and 6f. Proposed mechanisms of v_{N+2} ion formation for threonine and serine are shown in Schemes 3a and 3b. The hydrogen bond between hydroxyl hydrogen and carbonyl oxygen facilitates a six-member ring formation, which helps hydrogen transfer and side chain elimination. The v_{N+2} ion is produced after the elimination of CH₃CHO and CH₂O groups for threonine and serine respectively and hydrogen is rearranged to form a carbonyl group. As shown in Table 4, the mass of a v_{N+2} ion for N-terminal threonine happens to match the mass of the precursor with CO₂ loss. However, the unique mass of v_{N+2} ion for N-terminal serine is indicative of its presence. Threonine and serine are the only two N-terminal residues that yield all three v_N , v_{N+1} and v_{N+2} ions (Table 3) and this should be useful for confirming their presence. Since the *D. radiodurans* ribosomal protein sample provided only a few examples for each N-terminal residue, further study of these peptides is warranted. Table 4 lists the mass spacing between the precursor ion and the largest x, v and w ions for all possible N-terminal residues. Some of these mass spacings are unique. In other cases, they can be used as constraints to limit the number of N-terminal residue candidates.

Immonium ions and internal fragments

Immonium ions and internal fragments are commonly observed in all photodissociation spectra of synthetic peptide libraries, and their observation has been discussed previously.^{44,50,62} Not every residue yields an immonium ion. Figures 7A and 7B display spectra that include immonium ions of histidine, arginine, glutamine, isoleucine and tyrosine. Histidine and tyrosine immonium ions are dominant peaks. Photodissociation also produces immonium ions for proline, phenylalanine and tryptophan that are not shown in these examples. Such immonium ions should help to confirm peptide *de novo* sequences. Numerous internal fragments also appear in Figure 7. Typical of our synthetic library spectra, most of these intense internal fragments are either dipeptides or tripeptides that contain proline or histidine. These internal ions provide 50% sequence coverage of ELAYHEIQLR and 73% sequence coverage of YQEVGPNSER. Therefore, their recognition should be useful for confirming *de novo* sequences and identifying peptides.⁵⁷

Glutamic acid, asparagine and threonine effects

It is well known that fragmentation is enhanced C-terminal to glutamic acid.^{58,59} Figure 8 displays a typical example of photodissociation spectra of peptides with glutamic acid and asparagine at the peptide N-terminus. The peptides with N-terminal glutamic acid provide intense x_9 and y_9 fragments (Figure 8a and 8b), but y_9 is missing for the peptide with N-terminal asparagine (Figure 8c). The enhanced cleavage next to glutamic acid facilitates recognition of its x/y ion pair. In our 1345 peptides whose spectra were confidently identified by Mascot, there were 956 glutamic acid residues, 284 of them at the N-terminus. 97% of the glutamic acid-containing peptides cleave next to this residue to generate y -type ions. 80% of peptides with N-terminal glutamic acid produce y_{N-1} fragments. By comparison, of 618 asparagine-containing peptides, about 42% cleave next to asparagine to form y ions, but only 18 out of 272 peptides with N-terminal asparagine produce y_{N-1} fragments. These observations directly impact *de novo* sequencing that requires the observation of y -type ions.

Surprisingly, threonine was also found to enhance peptide fragmentation when it is close to the peptide N-terminus. In synthetic libraries, 330 out of 1345 identified peptides contain threonine next to the N-terminus and 268 of them produce y_{N-2} fragments. By contrast, the other 1015 peptides generate only 50 y_{N-2} ions. The y_{N-2} intensity normalized by the total intensity of all assigned peaks was extracted and compared for different residues. The average of normalized y_{N-2} intensity for threonine at the second N-terminal residue is 0.008431 and it is only 0.00038 for all other second residues. However, when threonine is the third residue from the C-terminus, the averaged y_2 intensity is 0.066731 and it is 0.053687 for all other residues on the same position. Therefore, the fragmentation enhancement C-terminal to threonine is not obvious if it is near the C-terminus rather than the N-terminus. Figure 8 displays an example of the threonine effect. The peptides involved have threonine or leucine as the second residue from the N-terminus. In these rather typical spectra, cleavage next to threonine appears to be enhanced. Although the mechanism for this is not clear, it may involve the side chain OH interacting with the adjacent amide group through hydrogen bonding just as the OH of aspartic acid does when facilitating fragmentation next to that residue.⁵⁹ The dissociation of peptide ions formed by electrospray ionization was studied by Wysocki and coworkers and no fragmentation enhancement next to threonine was observed.^{60,61} Further work on singly-charged ions with threonine and serine near the N-terminus is warranted.

Identification of Y-type ions

Photodissociation produces intense Y-type ions terminated by proline.^{51,52} The observation of Y/ y pairs in photodissociation and PSD data can be used to identify proline residues.^{41,42} However, examination of over one thousand library spectra led to the conclusion that cleavage next to glycine, glutamine, alanine, proline, histidine and tyrosine can also yield Y-type ions by photodissociation. Figure 9 shows some typical examples. Only y - and Y-type ions are marked in these spectra. The Y ions formed by cleavage next to proline are certainly the most intense in Figure 9. The Y ions from glutamine and alanine terminated fragments are also significant but less intense in Figure 9. Y ions terminated with histidine are typically relatively weak (e.g. Y_2 in Figure 9a and Y_8 in Figure 9b). However, Y-type ions terminated by the residues glycine, glutamic acid, alanine, histidine and tyrosine do not appear consistently in photodissociation spectra. For example, glycine N-terminated fragments have both Y_8 and y_8 ions in Figure 9a, but only have y_5 and y_9 ions in Figure 9c. In photodissociation of synthetic peptides, Y-type ions terminated by glutamine are generated for only 65% of 232 spectra containing this residues; Likewise, tyrosine-terminated Y ions appear in only 47% of 176 peptide spectra that include this residue. Therefore, the presence of Y-type ions in photodissociation spectra cannot be relied upon for sequencing efforts.

However, since the library data suggest that Y-type ions terminated by just certain residues tend to form, this can be used as corroborating information.

Conclusion

Three synthetic peptide libraries with C-terminal arginine were analyzed by a modified 4700 MALDI TOF-TOF instrument to study 157nm photodissociation. Some fragmentation rules were extracted with the eventual goal of improving *de novo* sequencing. Sequencing peptide N-terminal regions often fails because high-mass y-type ions are missing from spectra. However, this work demonstrated that b_2 and b_3 ions appear in most photodissociation spectra. Their ubiquity enables b_2/x_{N-2} and b_3/x_{N-3} ion pairs to be identified which should help confirm the identities of high-mass x-type ions and low-mass b_2 and b_3 ions. Accurate masses of b_2 ions should also facilitate assigning the first two N-terminal residues. The spacings between precursor mass and x_{N-1} ion fragments can often determine the peptide N-terminal amino acid. In addition, v_{N-} , w_{N-} and v_{N+1} -type ions can help to confirm N-terminal residue assignments. Immonium ions and internal fragments can be utilized to confirm *de novo* derived sequences. The photofragmentation propensities observed in this work should improve the photofragmentation-based sequencing of peptides. Based on these improvements, we anticipate being able to identify more of the peptides in these libraries.

Acknowledgments

The authors thank Youyou Yang and Dr. Liangyi Zhang for providing photodissociation spectra from *D. radiodurans* ribosomal proteins. This work is supported by grants from the National Institutes of Health (R01 RR024236-01A1) and the National Science Foundation (CHE-1012855).

References

1. Aebersold R, Goodlett DR. Mass spectrometry in proteomics. *Chem Rev.* 2001; 101(2):269–295. [PubMed: 11712248]
2. Aebersold R, Mann M. Mass spectrometry-based proteomics. *Nature.* 2003; 422(6928):198–207. [PubMed: 12634793]
3. Yates JR. Mass spectrometry and the age of the proteome. *J Mass Spectrom.* 1998; 33(1):1–19. [PubMed: 9449829]
4. Hunt DF, Yates JR, Shabanowitz J, Winston S, Hauer CR. Protein Sequencing by Tandem Mass Spectrometry. *Proc Natl Acad Sci USA.* 1986; 83(17):6233–6237. [PubMed: 3462691]
5. Biemann K, Scoble HA. Characterization by tandem mass spectrometry of structural modifications in proteins. *Science.* 1987; 237(4818):992–998. [PubMed: 3303336]
6. Henzel WJ, Billeci TM, Stults JT, Wong SC, Grimley C, Watanabe C. Identifying proteins from two-dimensional gels by molecular mass searching of peptide fragments in protein sequence databases. *Proc Natl Acad Sci USA.* 1993; 90(11):5011–5015. [PubMed: 8506346]
7. Yates JR, Speicher S, Griffin PR, Hunkapiller T. Peptide mass maps: a highly informative approach to protein identification. *Anal Biochem.* 1993; 214(2):397–408. [PubMed: 8109726]
8. Perkins DN, Pappin DJC, Creasy DM, Cottrell JS. Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis.* 1999; 20(18):3551–3567. [PubMed: 10612281]
9. Yates JR, Eng JK, Clauser KR, Burlingame AL. Search of sequence databases with uninterpreted high-energy collision-induced dissociation spectra of peptides. *J Am Soc Mass Spectrom.* 1996; 7(11):1089–1098.
10. Craig R, Beavis RC. TANDEM: matching proteins with tandem mass spectra. *BIOINFORMATICS.* 2004; 20(9):1466–1467. [PubMed: 14976030]
11. Yi D, Perkins PD. Identification of ubiquitin nitration and oxidation using a liquid chromatography/mass selective detector system. *J Biomol Tech.* 2005; 16(4):364–370. [PubMed: 16522858]

12. Geer LY, Markey SP, Kowalak JA, Wagner L, Xu M, Maynard DM, Yang XY, Shi WY, Bryant SH. Open mass spectrometry search algorithm. *J Proteome Res.* 2004; 3(5):958–964. [PubMed: 15473683]
13. Xu CX, Ma B. Software for computational peptide identification from MS-MS data. *Drug Discov Today.* 2006; 11(13–14):595–600. [PubMed: 16793527]
14. Zhang N, Li XJ, Ye ML, Pan S, Schwikowski B, Aebersold R. ProbiDtree: An automated software program capable of identifying multiple peptides from a single collision-induced dissociation spectrum collected by a tandem mass spectrometer. *Proteomics.* 2005; 5(16):4096–4106. [PubMed: 16196091]
15. Bafna V, Edwards N. SCOPE: a probabilistic model for scoring tandem mass spectra against a peptide database. *Bioinformatics.* 2001; 17S:S13–21. [PubMed: 11472988]
16. Zhang N, Aebersold R, Schwikowski B. ProbiD: A probabilistic algorithm to identify peptides through sequence database searching using tandem mass spectral data. *Proteomics.* 2002; 2(10):1406–1412. [PubMed: 12422357]
17. Colinge J, Masselot A, Giron M, Dessingy T, Magnin J. OLAV: Towards high-throughput tandem mass spectrometry data identification. *Proteomics.* 2003; 3(8):1454–1463. [PubMed: 12923771]
18. Fu Y, Yang Q, Sun RX, Li DQ, Zeng R, Ling CX, Gao W. Exploiting the kernel trick to correlate fragment ions for peptide identification via tandem mass spectrometry. *Bioinformatics.* 2004; 20(12):1948–1954. [PubMed: 15044235]
19. Wan YH, Yang A, Chen T. PepHMM: a hidden Markov model based scoring function for mass spectrometry database search. *Anal Chem.* 2006; 78(2):432–437. [PubMed: 16408924]
20. Tabb DL, Narasimhan C, Strader MB, Hettich RL. DBDigger: Reorganized proteomic database identification that improves flexibility and speed. *Anal Chem.* 2005; 77(8):2464–2474. [PubMed: 15828782]
21. Keller A, Nesvizhskii AI, Kolker E, Aebersold R. Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. *Anal Chem.* 2002; 74(20):5383–5392. [PubMed: 12403597]
22. Taylor JA, Johnson RS. Implementation and uses of automated *de novo* peptide sequencing by tandem mass spectrometry. *Anal Chem.* 2001; 73(11):2594–2604. [PubMed: 11403305]
23. Zheng ZQ. *De novo* peptide sequencing based on a divide-and-conquer algorithm and peptide tandem spectrum simulation. *Anal Chem.* 2004; 76(21):6374–6383. [PubMed: 15516130]
24. Bern M, Goldberg D. *De novo* analysis of peptide tandem mass spectra by spectral graph partitioning. *J Comput Biol.* 2006; 13(2):364–378. [PubMed: 16597246]
25. Frank A, Pevzner P. PepNovo: *De novo* peptide sequencing via probabilistic network modeling. *Anal Chem.* 2005; 77(4):964–973. [PubMed: 15858974]
26. Fischer B, Roth V, Roos F, Grossmann J, Baginsky S, Widmayer P, Gruissem W, Buhmann JM. NovoHMM: A hidden Markov model for *de novo* peptide sequencing. *Anal Chem.* 2005; 77(22):7265–7273. [PubMed: 16285674]
27. Ma B, Zhang KZ, Hendrie C, Liang CZ, Li M, Doherty-Kirby A, Lajoie G. PEAKS: powerful software for peptide *de novo* sequencing by tandem mass spectrometry. *Rapid Commun Mass Spectrom.* 2003; 17(20):2337–2342. [PubMed: 14558135]
28. Mo LJ, Dutta D, Wan YH, Chen T. MSNovo: A dynamic programming algorithm for *de novo* peptide sequencing via tandem mass spectrometry. *Anal Chem.* 2007; 79(13):4870–4878. [PubMed: 17550227]
29. Huang YY, Triscari JM, Tseng GC, Pasa-Tolic L, Lipton MS, Smith RD, Wysocki VH. Statistical characterization of the charge state and residue dependence of low-energy CID peptide dissociation patterns. *Anal Chem.* 2005; 77(18):5800–5813. [PubMed: 16159109]
30. Kjeldsen F, Haselmann KF, Sorensen ES, Zubarev RA. Distinguishing of Ile/Leu amino acid residues in the PP3 protein by (hot) electron capture dissociation in Fourier transform ion cyclotron resonance mass spectrometry. *Anal Chem.* 2003; 75(6):1267–1274. [PubMed: 12659185]
31. Sleno L, Volmer DA. Ion activation methods for tandem mass spectrometry. *J Mass Spectrom.* 2004; 39(10):1091–1112. [PubMed: 15481084]

32. Zubarev RA, Kelleher NL, McLafferty FW. Electron capture dissociation of multiply charged protein cations. A nonergodic process. *J Am Chem Soc.* 1998; 120(13):3265–3266.
33. Syka JEP, Coon JJ, Schroeder MJ, Shabanowitz J, Hunt DF. Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. *Proc Natl Acad Sci USA.* 2004; 101(26): 9528–9533. [PubMed: 15210983]
34. Stensballe A, Jensen ON, Olsen JV, Haselmann KF, Zubarev RA. Electron capture dissociation of singly and multiply phosphorylated peptides. *Rapid Commun Mass Spectrom.* 2000; 14(19):1793–1800. [PubMed: 11006587]
35. Shi SDH, Hemling ME, Carr SA, Horn DM, Lindh I, McLafferty FW. Phosphopeptide/phosphoprotein mapping by electron capture dissociation mass spectrometry. *Anal Chem.* 2001; 73(1):19–22. [PubMed: 11195502]
36. Zubarev RA, Horn DM, Fridriksson EK, Kelleher NL, Kruger NA, Lewis MA, Carpenter BK, McLafferty FW. Electron capture dissociation for structural characterization of multiply charged protein cations. *Anal Chem.* 2000; 72(3):563–573. [PubMed: 10695143]
37. Nielsen ML, Savitski MM, Zubarev RA. Improving protein identification using complementary fragmentation techniques in Fourier transform mass spectrometry. *Mol Cell Proteomics.* 2005; 4(6):835–845. [PubMed: 15772112]
38. Zubarev RA, Zubarev AR, Savitski MM. Electron capture/transfer versus collisionally activated/induced dissociations: Solo or duet? *J Am Soc Mass Spectrom.* 2008; 19(6):753–761. [PubMed: 18499036]
39. Savitski MM, Nielsen ML, Kjeldsen F, Zubarev RA. Proteomics-grade *de novo* sequencing approach. *J Proteome Res.* 2005; 4(6):2348–2354. [PubMed: 16335984]
40. Johnson RS, Martin SA, Biemann K, Stults JT, Watson JT. Novel fragmentation process of peptides by collision- induced. decomposition in a tandem mass spectrometer: Differentiation of leucine and isoleucine. *Anal Chem.* 1987; 59(21):2621–2625. [PubMed: 3688448]
41. Hines WM, Falick AM, Burlingame AL, Gibson BW. Pattern-based algorithm for peptide sequencing from tandem high-energy collision-induced dissociation mass spectra. *J Am Soc Mass Spectrom.* 1992; 3(4):326–336.
42. Kaufmann R, Kirsch D, Spengler B. Sequencing of peptides in a time-of-flight mass spectrometer - Evaluation of post-source decay following matrix-assisted laser desorption ionisation (MALDI). *Int J Mass Spectrom Ion Processes.* 1994; 131:355–385.
43. Kaufmann R, Spengler B, Lutzenkirchen F. Mass spectrometric sequencing of linear peptides by product-ion analysis in a reflectron time-of-flight mass spectrometer using matrix-assisted laser desorption ionization. *Rapid Commun Mass Spectrom.* 1993; 7(10):902–910. [PubMed: 8219321]
44. Thompson MS, Cui WD, Reilly JP. Fragmentation of singly charged peptide ions by photodissociation at $\lambda=157$ nm. *Angew Chem Int Ed.* 2004; 43(36):4791–4794.
45. Cui WD, Thompson MS, Reilly JP. Pathways of peptide ion fragmentation induced by vacuum ultraviolet light. *J Am Soc Mass Spectrom.* 2005; 16(8):1384–1398. [PubMed: 15979330]
46. Devakumar A, O'Dell DK, Walker JM, Reilly JP. Structural analysis of leukotriene C-4 isomers using collisional activation and 157 nm photodissociation. *J Am Soc Mass Spectrom.* 2008; 19(1): 14–26. [PubMed: 18024058]
47. Devakumar A, Mechref Y, Kang P, Novotny MV, Reilly JP. Laser-induced photofragmentation of neutral and acidic glycans inside an ion-trap mass spectrometer. *J Am Soc Mass Spectrom.* 2007; 21(8):1452–1460.
48. Zhang L, Reilly JP. Use of 157-nm photodissociation to probe structures of γ - and b -type ions produced in collision-induced dissociation of peptide ions. *J Am Soc Mass Spectrom.* 2008; 19(5): 695–702. [PubMed: 18325783]
49. Zhang L, Reilly JP. Extracting both peptide sequence and glycan structural information by 157 nm photodissociation of N-linked glycopeptides. *J Proteome Res.* 2008; 8(2):734–742. [PubMed: 19113943]
50. Zhang LY, Reilly JP. Peptide photodissociation with 157 nm light in a commercial tandem time-of-flight mass spectrometer. *Anal Chem.* 2009; 81(18):7829–7838. [PubMed: 19702244]
51. Zhang LY, Reilly JP. Peptide *de novo* sequencing using 157 nm photodissociation in a tandem time-of-flight mass spectrometer. *Anal Chem.* 2010; 82(3):898–908. [PubMed: 20058881]

52. Zhang L, Reilly JP. *De novo* sequencing of tryptic peptides derived from *deinococcus radiodurans* ribosomal proteins using 157 nm photodissociation MALDI TOF/TOF mass spectrometry. *J Proteome Res.* 2010; 9(6):3025–3034. [PubMed: 20377247]
53. Bohrer BC, Li YF, Reilly JP, Clemmer DE, DiMarchi RD, Radivojac P, Tang H, Arnold RJ. Combinatorial libraries of synthetic peptides as a model for shotgun proteomics. *Anal Chem.* 2010; 82(15):6559–6568. [PubMed: 20669997]
54. Lebl M, Krchnak V. Synthetic peptide libraries. *SOLID-PHASE PEPTIDE SYNTHESIS.* 1997; 289:336–392.
55. Savitski MM, Fälth M, Fung YME, Adams CM, Zubarev RA. Bifurcating fragmentation behavior of gas-phase tryptic peptide dications in collisional activation. *J Am Soc Mass Spectrom.* 2008; 19(12):1755–1763. [PubMed: 18799320]
56. Godugu B, Neta P, Simón-Manso Y, Stein SE. Effect of N-terminal glutamic acid and glutamine on fragmentation of peptide ions. *J Am Soc Mass Spectrom.* 2010; 21(7):1169–1176. [PubMed: 20413325]
57. Schlosser A, Lehmann WD. Patchwork peptide sequencing: Extraction of sequence information from accurate mass data of peptide tandem mass spectra recorded at high resolution. *Proteomics.* 2002; 2(5):524–533. [PubMed: 11987126]
58. Qin J, Chait BT. Preferential fragmentation of protonated gas-phase peptide ions adjacent to acidic aminoacid residues. *J Am Chem Soc.* 1995; 117(19):5411–5412.
59. Tsaprailis G, Somogyi A, Nikolaev EN, Wysocki VH. Refining the model for selective cleavage at acidic residues in arginine-containing protonated peptides. *int J Mass Spectrom.* 2000; 156–157:467–479.
60. Huang YY, Triscari JM, Tseng GC, Pasa-Tolic L, Lipton MS, Smith RD, Wysocki VH. Statistical characterization of the charge state and residue dependence of low-energy CID peptide dissociation patterns. *Anal Chem.* 2005; 77(18):5800–5813. [PubMed: 16159109]
61. Huang YY, Triscari JM, Pasa-Tolic L, Anderson GA, Lipton MS, Smith RD, Wysocki VH. Dissociation behavior of doubly-charged tryptic peptides: Correlation of gas-phase cleavage abundance with Ramachandran plots. *J Am Chem Soc.* 2004; 126(10):3034–3035. [PubMed: 15012117]
62. Reilly JP. Ultraviolet photofragmentation of biomolecular ions. *Mass Spec Rev.* 2009; 28(3):425–447.

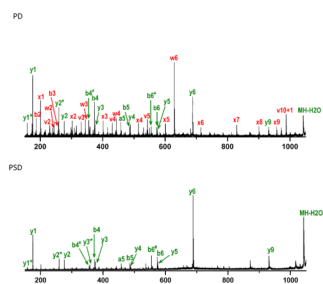
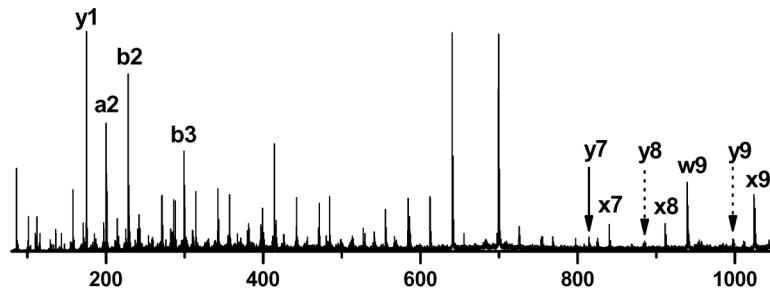


Figure 1.
Photodissociation (PD) and postsource decay (PSD) spectra of EGADLSIVTR.

PD



PSD

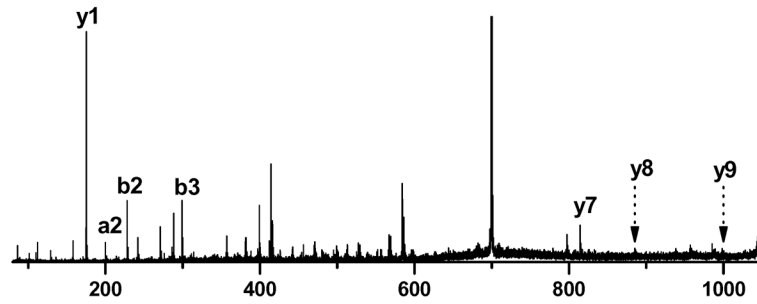


Figure 2. b_2 , b_3 ions and complementary x- and y-type ions in the photodissociation (PD) and postsource decay (PSD) spectrum of NLADLGIQLR.

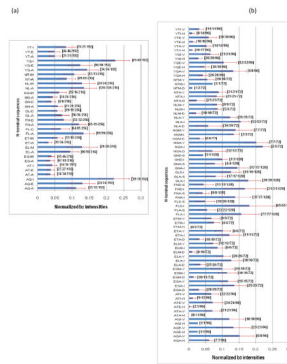


Figure 3. Normalized intensities of (a) b_2 ions; (b) b_3 ions with various sequences and different residues to provide amine groups for cleavage bonds. In square brackets, the numbers specify the number of [identified b_2 ions/identified peptides/peptides in libraries] respectively.

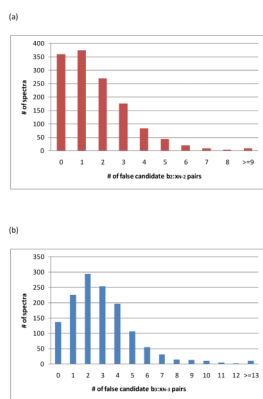


Figure 4. The number of spectra containing various of false b/x ion pairs. (a) b_2/x_{N-2} ion pairs; (b) b_3/x_{N-3} ion pairs

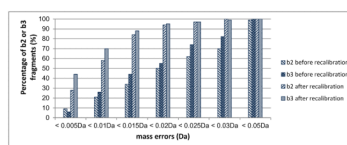


Figure 5. Comparison of the accuracy of identifying b_2 and b_3 fragments before and after recalibration of spectra using y_1 ions.

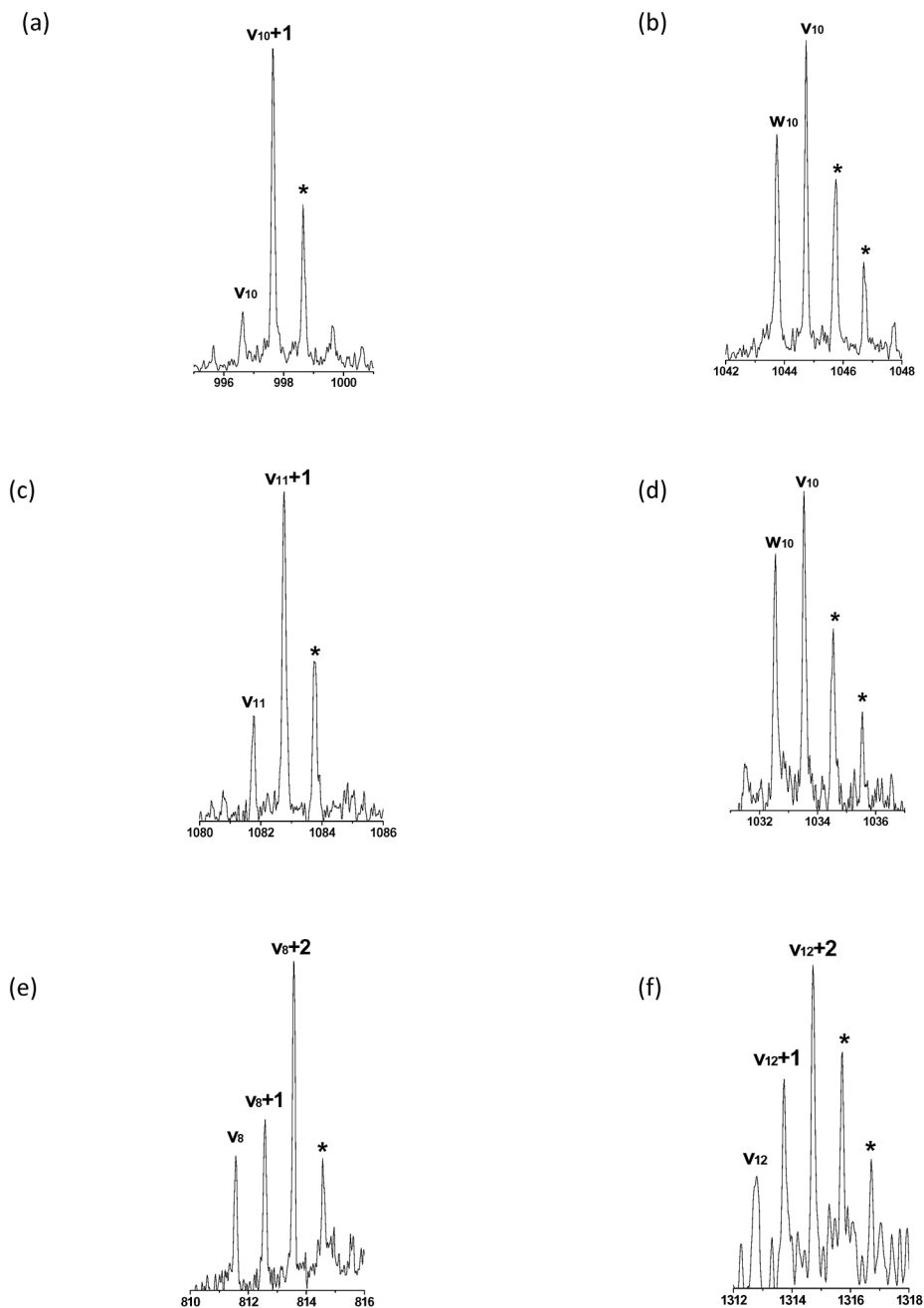


Figure 6. Assignments of (a) v_n+1 radical ion for EGAILSIVLR; (b) v_n , w_n ions of DAIQQAFGAR; (c) v_n+1 radical ion of QLVGQVAANVR; (d) v_n , w_n ions of NGAYLGIQTR; (e) v_n , w_n and v_n+1 radical ions of TVQALGLR; (f) v_n , w_n and v_n+1 radical ions of SLDLDSIIAEVK^g. * shows the isotopic peaks and ^g means that K is guanidinated.

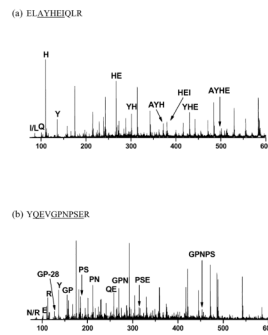


Figure 7. Immonium ions and internal fragments from the low-mass region of photodissociation spectra of (a) ELAYHEIQLR; (b) YQEVGNPSEK. Highlighted residues correspond to immonium ions and underlined sequences label internal fragments.

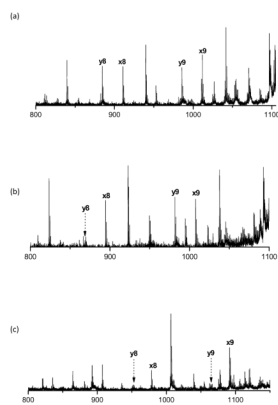


Figure 8. High-mass regions of photodissociation spectra of (a) ETAILSIVLR; (b) ELAILSPVLR; (c) NLADHEIVLR.

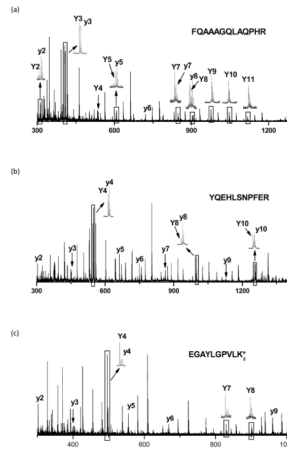
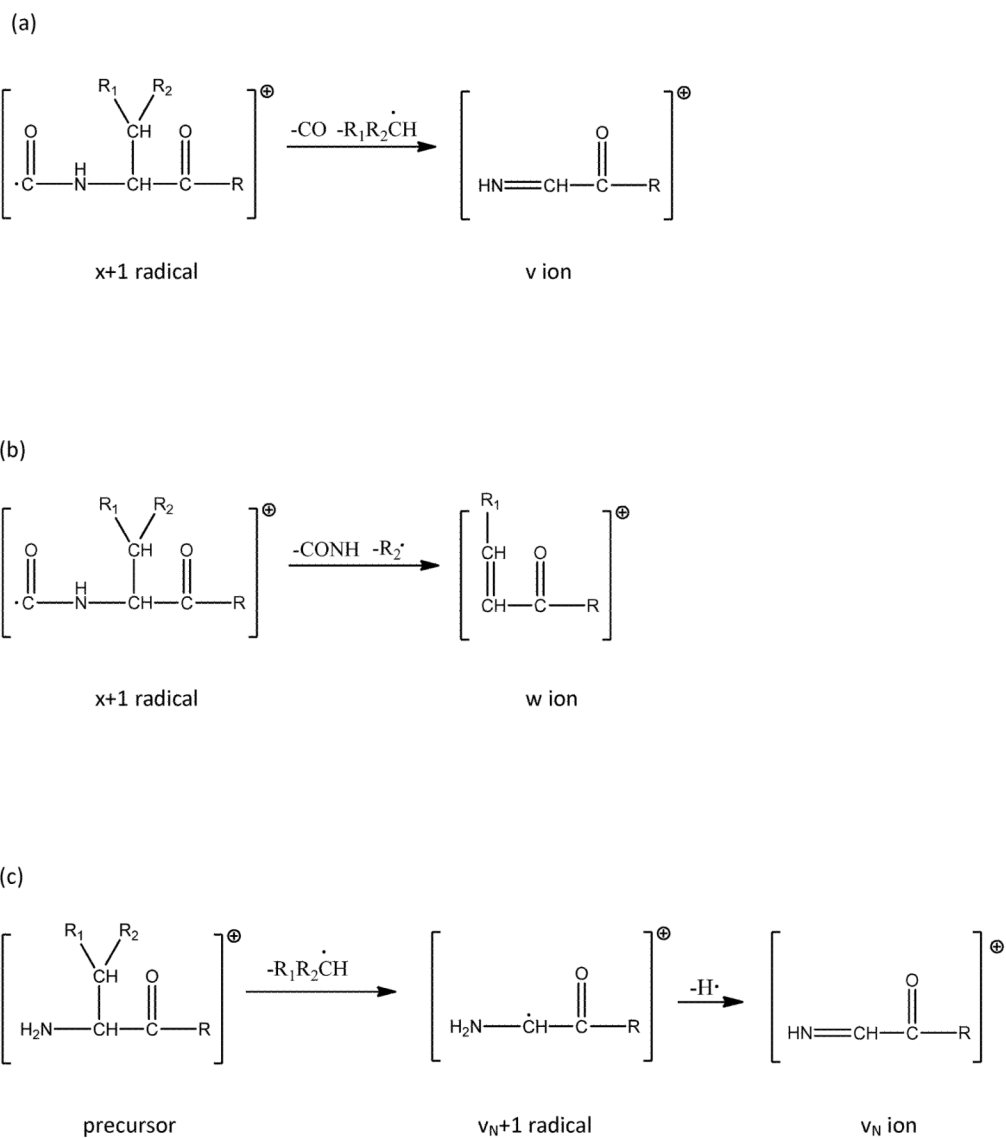
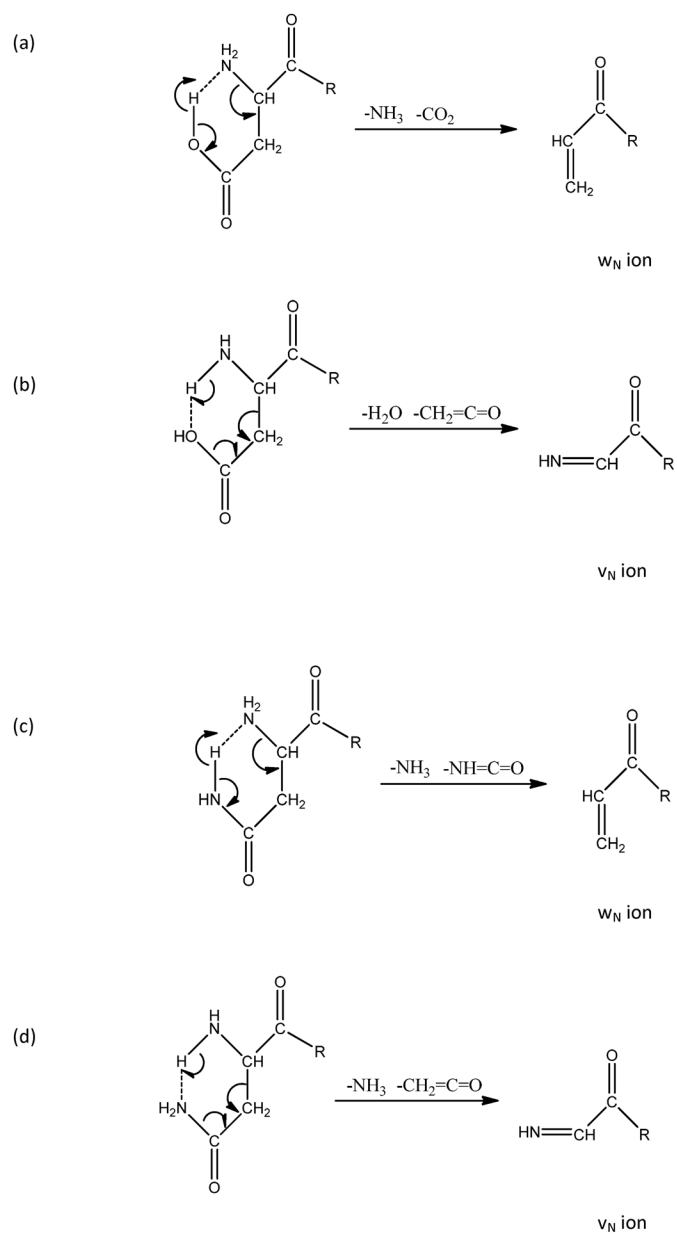


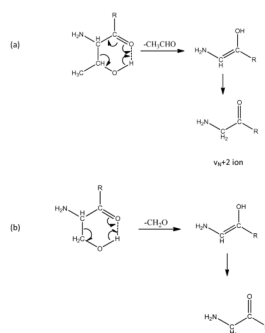
Figure 9. Y-type ions in photodissociation spectra of (a) FQAAAGQLAQPHER; (b) YQEHLSPFER; (c) EGAYLGPVLK^g. ^g means that K is guanidinated.

**Scheme 1.**

Pathways to form various fragment ions: (a) v- ions; (b) w ions; (c) v_N+0.1 and v_N ions.



Scheme 2.
Pathways to form w_N and v_N from (a)–(b) aspartic acid and (c)–(d) asparagine

**Scheme 3.**

Proposed pathways to form v_{N+2} ions from peptides with N-terminal (a) threonine; (b) serine.

Table 1

Composition of Synthetic Peptide Libraries

	1	2	3	4	5	6	7	8	9	10	11	12	# of peptides	
BB10A	N	T	M	I	L	S	I	C	T	R				
	E	L	A	D	H	E	P	V	L	K			5184	
	G		Y		G		Q							
BB11A	Y	Q	A	V	G	P	N	P	G	I	R			
	A	T	E	H	L	S	T	L	S	E	K		4608	
			I					F						
BB12A	G	N	D	I	G	H	P	E	M	Q	A	R		
	F	L	A	S	V	S	T	V	L	T	S	K	4096	

Table 2

Number of spectra for which various fragment ions were observed.

	b_2	x_{n-2}	y_{n-2}	b_3	x_{n-3}	y_{n-3}	b_2/x_{n-2}	y_{n-2}/x_{n-2}	b_3/x_{n-3}	y_{n-3}/x_{n-3}
# of peptides	1337	1076	318	1297	1125	485	1053	247	1056	368

Table 3

Percentage of w_N , v_N , v_{N+1} and x_{N-1} ions for various N-terminal amino acids.

N-terminal residue	w_N	v_N	v_{N+1}	v_{N+2}	x_{N-1}
Asn ^a	69%	46%	15%		93%
Glu ^a		42%	90%		89%
Ala ^a		28%	20%		86%
Tyr ^a		51%	85%		44%
Phe ^a		72%			76%
Gly ^a	-	-	-	-	51%
Asp ^b	x	x			
Leu ^b		x			
Ile ^b		x			
Val ^b		x			
His ^b		x	x		
Trp ^b		x	x		
Gln ^b		x	x		
Ser ^b		x	x	x	
Thr ^b		x	x	x	
Met ^b			x		

^aPeptides derived from synthetic libraries.

^bPeptides derived from *D. radiodurans* ribosomal proteins. X indicates that the fragment is present. - indicates ions that cannot be formed for N-terminal glycine.

Table 4Mass differences between precursor and x_{N-1} , v_N and w_N ions for different N-terminal amino acids.

	precursor- x_{N-1}	precursor- v_N	precursor- (v_N+1)	Precursor- $w_N/(v_N+2)$
Ala	45.058(Da)	16.031 (Da)	15.024(Da)	
Asn	88.064(Da)	59.037 (Da)	58.030(Da)	60.0324(Da) ^a
Asp	89.048(Da)	60.021 (Da)	59.014(Da)	61.0164(Da) ^a
Gln	102.080(Da)	73.053(Da)	72.046(Da)	
Glu	103.063 (Da)	74.037(Da)	73.030(Da)	
Gly	31.042 (Da)	-	-	
His	111.080(Da)	82.053 (Da)	81.046(Da)	
Ile	87.105(Da)	58.078 (Da)	57.071(Da)	
Leu	87.105(Da)	58.078 (Da)	57.071(Da)	
Met	105.061 (Da)	76.035(Da)	75.028(Da)	
Phe	121.089 (Da)	92.063(Da)	91.056(Da)	
Ser	61.053(Da)	32.026 (Da)	31.019(Da)	30.012(Da) ^b
Thr	75.068 (Da)	46.042(Da)	45.035(Da)	44.028(Da) ^b
Trp	160.100 (Da)	131.074(Da)	130.067(Da)	
Tyr	137.084 (Da)	108.058(Da)	107.051(Da)	
Val	73.089 (Da)	44.063(Da)	43.056(Da)	
Arg	130.122(Da)	101.095 (Da)	100.088(Da)	
Cys	77.030(Da)	48.003 (Da)	46.996(Da)	
Lys	102.116(Da)	73.089 (Da)	72.082(Da)	
Pro	71.074(Da)	-	-	

^aPrecursor- w_N ;^bPrecursor- (v_N+2) .