# Investigation of Neutral Loss during Collision-Induced Dissociation of Peptide Ions

**ARTICLE** *in* ANALYTICAL CHEMISTRY · SEPTEMBER 2005

Impact Factor: 5.64 · DOI: 10.1021/ac050701k · Source: PubMed



**5 AUTHORS**, INCLUDING:

Jimmy K Eng
University of Washington Seattle
**140** PUBLICATIONS **23,305** CITATIONS
SEE PROFILE

Alexey I Nesvizhskii
University of Michigan
**124** PUBLICATIONS **12,046** CITATIONS
SEE PROFILE

Published in final edited form as:
*Anal Chem*. 2005 August 1; 77(15): 4870–4882.

# Investigation of neutral loss during collision induced dissociation of peptide ions.

**Daniel B. Martin**[1,2,*], **Jimmy K. Eng**[1,3], **Alexey I. Nesvizhskii**[1], **Andrew Gemmill**[1], and **Ruedi Aebersold**[1,4]

1 *Institute for Systems Biology, Seattle, WA 98103*

2 *Divisions of Human Biology and Clinical Research, Fred Hutchinson Cancer Research Center Seattle WA 98109-1024*

3 *Public Health Sciences Division, Fred Hutchinson Cancer Research Center Seattle WA 98109-1024*

4 *Institute of Molecular Systems Biology,, ETH Zurich and Faculty of Natural Sciences, University of Zurich, Switzerland*

## Abstract

MS/MS fragmentation of peptides is dominated by overlapping b and y ion series. However, alternative fragmentation possibilities exist, including neutral loss. A database was generated containing 8400 MS/MS spectra of tryptic peptides assigned with high probability to an amino acid sequence (true positives), and a set of certified false (true negative) assignments for analysis of the amino terminus. A similar database was created for analysis of neutral loss at the carboxy termini using a dataset of chymotryptic peptides. The analysis demonstrated that the presence of an internal basic residue, limiting proton mobility, has a profound effect on neutral loss. Peptides with fully mobile protons demonstrated minimal neutral loss, with the exception of amide bonds with proline on the carboxy terminal side, which created an intense neutral loss peak. In contrast, peptides with *partial* proton mobility contained many amino acids on either side of the amide bond associated with a strong neutral loss peak. Most notable among these was proline on the carboxy terminal side of an amide bond and aspartic acid, on the amino terminal side of a bond. All results were found to be consistent for doubly and triply charged peptides and after adjustment for pairings across the amide bonds with particularly labile residues. The carboxy terminal of chymotryptic peptides also demonstrated significant neutral loss events associated with numerous amino acid residues. Clarification of the rules that govern neutral loss, when incorporated into analysis software, will improve our ability to correctly assign spectra to peptide sequences.

High throughput mass spectrometry has become an increasingly important tool in investigative biology. One of the most actively employed strategies for deriving sequence information for peptides utilizes soft ionization techniques such as electrospray ionization (ESI) or matrix assisted laser desorption/ionization (MALDI) combined with low-energy collision-induced dissociation (CID) to produce fragmentation. Tandem mass spectra generated by the fragmentation of peptide ions in the gas phase at low collision energy are dominated by fragment ions resulting from cleavage at the amide bonds, with very little side chain fragmentation observed [1]. While numerous fragmentation pathways are possible and have been extensively studied [reviewed in [2]], the predominant fragmentation pathway generates overlapping b and y ion series [3, 4]. A number of computer programs compare MS/MS

* Corresponding author: Institute for Systems Biology (phone) 206 732-1365 (fax) 206 732-1299 email dmartin@systemsbiology.org

Supporting Information Available.

Supporting information as stated in the text. This material is available free of charge via the Internet at http://pubs.acs.org.

fragmentation patterns generated *in silico* from protein and DNA databases with the b and y ion spectra determined experimentally to assign a peptide sequence to the spectrum[5, 6]. The performance of these search algorithms, therefore, depends on a thorough understanding of the process of fragmentation.

Electrospray ionization systems typically generate doubly and triply charged precursor ions that dissociate into fragments each of which typically carries at least one charge. However, alternative fragmentation possibilities exist, including those where all charge is retained on one of the fragments resulting in one fragment in the same charge state as the precursor and a neutral fragment. Such fragmentation events are termed neutral loss. Neutral loss is quite well known and has typically been described as the loss of small molecules such as water, ammonia, or carbon monoxide from peptides, or the loss of larger molecules such as $H_3PO_4$ or $CH_3SOH$ from covalently modified peptides. The neutral loss of amino acid residues from the N- and C-termini of doubly and triply protonated tryptic peptides is also well known [2, 7–12]; it was recently described to produce useful y and b type ion ladders that could be used as an additional source of information for the conclusive assignment of MS/MS spectra to peptide sequences [13]. These authors also noted that peptides undergoing neutral loss tended to contain either an internal basic residue or an internal proline, suggesting that the neutral loss phenomena relates to a restriction of proton mobility.

While not typically associated with neutral loss, proton mobility has been the subject of a number of research studies in the past few years that attempted to elucidate the rules governing gas-phase fragmentation [2, 14–16]. The model of proton mobility holds that fragmentation of amide bonds under low-energy activation is brought about by the migration of a proton from the initial site of protonation (N-terminal amino groups, basic side chains) along the peptide backbone to an amide bond, where the localization of the proton facilitates cleavage by neighboring group attack from an adjacent nucleophilic amide carbonyl group [2]. This is referred to as “charge-directed” cleavage.

Tryptic peptides ideally contain one basic group at the amino terminal and a second one at the side chain of the C-terminal amino acid (K or R). The presence of additional internal basic residue(s) can cause localization of a charge to their side chain(s). This results in substantial energy requirements for proton mobilization, with a net result that proton movement is less favorable during CID. In such cases alternative, “non-charge-directed” fragmentation mechanisms predominate. Kapp and co-workers recently further refined the mobility model by developing a “relative proton mobility scale” that incorporates peptide charge state with basic residue content [17]. These authors divided a large database of peptides into those possessing mobile, partially mobile, and non-mobile protons, and defined peptide bond cleavage propensities according to this categorization. Peptides were classified as non mobile if the arginine residue number was greater than or equal to the proton number. Peptides were classified as “mobile” if the number of basic residues totaled less than total proton number. All others were classified as partially mobile and could be sub-categorized according to composition. This analysis was specific for fragmentation of the amide bond along the b/y ion pathway, but did not address neutral loss. In this paper we conduct a statistical analysis of observed neutral loss events at both the amino and carboxy termini from two large datasets of tryptic and chymotryptic peptides, respectively, to determine the frequencies and statistical significance of this occurrence if the input data are controlled for proton mobility.

## Experimental section

### Preparation of peptides for analysis

All reagents were purchased from Sigma (St Louis, MO, USA) unless otherwise noted. Spectra used for the tryptic analysis were derived from a series of experiments testing strong cation

exchange (SCX) and hydrophilic interaction chromatography (HILIC). Except as below, all experiments were performed using Saccharomyces cerevisiae (strain BY4741) prepared as described in [18] using iodoacetaminde, and digested overnight with trypsin (Promega, Madison, WI, USA) and dried in a low pressure centrifuge. Cation exchange was performed as described [19] using a polysulfoethyl A 2.1 mm x 200 mm column (PolyLC, Columbia MD). HILIC was performed using a 2.1 mm x 10 mm Polyhydroxyethyl column (PolyLC, Columbia MD). For HILIC samples, yeast digest was dissolved in 90% acetonitrile and 10% water. HILIC gradients were run using a gradient of acetonitrile (buffer A) and water (buffer B) from 90 to 10 percent A over 50 minutes. Yeast chymotryptic peptides were prepared using a yeast lysate grown as above followed digestion with chymotrypsin (Roche, Basel, Switzerland). Chymotryptic digests were separated using strong cation exchange and hydrophilic exchange chromatography as above. A twenty-five ml E.coli culture (strain DH5α) was grown overnight in LB amp+ media and lysed by sonication in 0.1% SDS and the supernatant retained. Proteins were denatured reduced and alkylated with iodoacetaminde and digested with chymotrypsin as above. Peptides were separated into six fractions using strong cation exchange chromatography as above.

**Mass Spectrometry and database searching—**Each chromatograpy fraction was run on an automated mass spectrometry system as described [20]. 5 μl of each sample was loaded onto a 75 μm internal diameter fused silica capillary column packed to a bed length of 10 cm with C18 spherical silica 5 μm mean particle size resin (Magic C18Aq Michrom Bioresources). The loaded column was washed isocratically for 5 minutes with buffer A (0.1% formic acid). Elution of peptides was performed using a linear gradient 10–35% buffer B (0.1% formic acid, 100%ACN) over 75 minutes at a flow rate of 200 nl/min using a HP 1100 solvent delivery system (Agilent, Palo Alto, CA). Mass spectrometry including CID was performed in an automated fashion using the dynamic exclusion option on an LCQ Deca XP mass spectrometer (ThermoFinnigan, San Jose, CA) equipped with an in-house built microspray device [21]. A number of the chromatography fractions were run more than once. A total of 330,123 spectra were searched against a Yeast ORF database (orf_trans.fasta downloaded 12/27/2003) from SGD at Stanford University or the E. coli database (E. coli subset of a non-redundant protein database from NCI downloaded 04/27/2004) using SEQUEST (version 27). To increase search speed, searches were performed requiring potential peptides to have at least one tryptic terminus. Identifications were filtered based on a PeptideProphet [22] probability of 0.9 using the INTERACT program [23]. Identifications were sorted based on peptide sequence. Many peptides were represented by multiple spectra after filtering: in these cases, the first spectrum for each peptide in the sorted list was chosen as representative for that peptide. This resulted in a tryptic database containing 4917 unique doubly charged peptides, and 3501 unique triply charged peptides. The chymotryptic database contained 651 and 659 unique doubly charged and triply charged peptides, respectively. Given the large number of spectra passing selection criteria, no attempt was made to manually validate those used in the analysis.

**Data Analysis—**For each high probability spectral assignment, a matched list of false positive assignments was created using the *second* highest scoring peptide from the SEQUEST search (excluding those answers indistinguishable from the highest scoring peptide due to isobaric residues). For each spectral assignment on each list, a table was generated containing the peptide sequence, charge state, proton mobility, and amino acid identity for the first five N-terminal residues. An entry was also created for the intensity of the fragment resulting from the loss of uncharged N-terminal residues for each of the first four amide bonds. This intensity was normalized to the base (most intense) peak in the spectrum and therefore has a value between 0 and 1. For chymotryptic peptides, for the first four residues counting from the C-terminal were entered and the intensity of the fragment resulting from the loss of uncharged carboxy terminal residues for each of the first four amide bonds was also normalized to the

base peak. Tables are available online. Non-mobile peptides were defined as those with a number of arginine residues greater to or equal to the proton number, mobile peptides were those with a number of basic residue totaling less than total proton number. All others were classified as partially mobile. Data was analyzed using the statistical software package, SPSS (SPSS Inc, Chicago, IL). Error bar plots were created using a 95% confidence interval for the mean.

## Results and Discussion

### Visualizing neutral losses

Collision induced dissociation of peptides in the mass spectrometer generates informative b and y type ion peaks used for peptide assignment. However, other informative peaks also exist. Among those, some of the most potentially useful are generated by neutral loss. Noting the frequency of peaks associated with neutral loss of fragments from the N-terminal of tryptic peptides in an ion trap instrument, we modified a software tool, COMET Spectrum View, to permit easy visualization of the phenomena to improve confidence in peptide assignment. The result was a spectral viewer (available http://cvs.sourceforge.net/viewcvs.py/sashimi/trans_proteomic_pipeline/) in which neutral loss peaks in MS/MS spectra of peptides are labeled (Figure 1). This spectral viewer has been previously used to visualize the quality of the fit of the assigned peptide to the spectra by marking the matched b and y type ions. The changes allow the user to select the ion series and charge state of the ions to be included in the display, including charge states equal to the precursor ions (fragments generated by the loss of a neutral).

With this viewer, we confirmed that neutral losses frequently occurred at the amino terminal of tryptic peptides, with proline frequently giving a very intense peak when located within the first four residues of the amino terminus. We also observed the presence of informative sequence ladders indicating that neutral losses occurred at sequential amide bonds of residues other than proline. These ladders occurred with strong intensity at the first two to or three amide bonds, and with much lower intensity at the forth amide bond and beyond. We also observed a parallel phenomena occurring at the carboxy terminus that was specifically associated with peptides lacking the carboxy terminal basic amino acid that is typical in a tryptic digest.

Our modified viewer allows assignment of intense unmatched peaks to neutral loss fragments, which provides rapid support to a peptide assignment. Being of obvious utility the changes in the viewer have been implemented as the default view. Our observations suggested the common occurrence of neutral loss during peptide fragmentation. It was most obvious with proline present near the amino terminal, but was also observed with most other residues. It also prompted a more thorough study of the phenomena.

### Generation of a database consisting of true positive and true negative assignments

Because it was clear that neutral loss peaks were quite common and specifically associated with a particular residue, we sought to quantify the frequency, relative intensity and residue specific behavior of the phenomena.

To perform this analysis we generated two spectral datasets for comparison of the properties of neutral loss. The first consisted of a set of spectra for which high confidence peptide assignments have been made i.e. “true positives“. The second set consisted of assignments for which there is a high confidence of *incorrect* peptide assignment, or “true negatives”. The first set included 4917 doubly and 3501 triply charged peptide spectra derived by filtering the 330,123 ion trap search results with PeptideProphet [22], and keeping only those with a probability of being a correctly assigned of greater than 0.9. When multiple spectra were

assigned to the same peptide, the first spectrum in the sorted list of peptides was arbitrarily chosen to represent that peptide.

The second spectral dataset was derived from the same spectra as the true positive set. However in this case, each spectrum was matched to the *second* best peptide assignment from the search results. Because the second best answer was taken from the list of assignments where the top assignment was presumed to be correct, we considered the second set to be true negatives.

These spectral datasets were used to generate a database with which we could perform analysis of neutral loss. For each spectrum in the true positive and negative spectral datasets, an entry in our database was created noting the intensity of the neutral loss peaks at each of the first four amide bonds moving inward from amino and carboxy termini. Specifically, the spectrum was examined for peaks within a window of width +/− 0.95 Da around the expected m/z value for the charged peptide resulting from a neutral loss at each of the first four amide bonds both termini. The matched intensity was normalized to the base peak of the spectrum and therefore takes values between zero and one. Also entered in the database was the identity of the residues on both sides of each amide bond as well as the proton mobility of the assigned peptide, using the criteria described by Kapp et al (see above).

A similar database was made for neutral losses occurring at the carboxy terminal using peptides generated by chymotryptic digestion. The spectral dataset consisted of 651 doubly charged and 659 triply charged chymotryptic peptides. From this spectral dataset, a chymotryptic database was created as above for the first three amide bonds moving inward from the carboxy termini. The database is available for download. While there is a standardized nomenclature for describing fragment ions and amide bonds counting inward from the amino terminus (aa(n)---aa(n+1)—etc) [3, 4, 24], our analysis required moving inward from either terminal. Thus, for purposes of communicating and analyzing data in this specific application, our database employs a simplification of the above nomenclature describing both the residue and bond location counting inward from the amino and carboxy termini, respectively (Figure 2).

Analysis of the dataset itself produced some interesting findings. In line with the original report by Kapp [17], those peptides classified as partially mobile make up a substantial fraction of the entire population. Table 1b gives the frequency of counts for arginine and lysine for peptides without proline (to avoid counting lysines expected as part of the KP sequence not digested by trypsin) for the same 2+ data. Peptides with two or more lysines are significantly more common than those with two or more arginines (Table 1b). This is not due to missed cleavages (equally likely with KP excluded from the dataset) or residue content, which is very similar for arginine and lysine [25], but rather to the fact that non-mobile protons (arginines equal or greater than charge state) fragment poorly in low energy CID [14, 17] and those peptides with partially mobile peptides (lysines equal to charge state) are more likely to generate interpretable spectra.

Our neutral loss database is large enough to facilitate an in depth analysis of neutral loss at each amide bond. The database is sufficiently large enough to approximate the behavior expected with a random assignment of a peptide to a spectra. The division of peptide matches according to mobility database permits further characterization of the neutral loss phenomena, which is relevant to the analysis as discussed above.

## Identifying the effect of proton mobility on neutral loss

Review of the literature revealed that neutral loss is generally associated with the presence of proline or internal basic residues. This suggested that fragmentation resulting in a neutral loss may be enhanced by the sequestration of the ionizing protons. We employed our database to test whether this observation is generalizable to both amino and carboxy termini, and the extent

to which the neutral losses occur at amide bonds moving inward from either end of the molecule, potentially generating informative sequence ladders.

We generated a summary table comparing the mean neutral loss intensity (normalized to the base peak) between the true positive and true negative populations at the first four amino and carboxy amide bonds, (Table 2 a-d).

Our results indicate that there are dramatic differences between the behaviors of the peptides according to proton mobility. As shown for doubly charged data in the first row of Table 2a, there is very little difference between mean neutral loss intensity between the true positive and the true negative populations when the peptide assigned is classified as non-mobile. This finding is true for the first four amide bonds moving inward from the amino terminus. In other words, those correctly matched spectra had statistically indistinguishable neutral loss behavior from that of the incorrectly matched spectra, and therefore in this class of peptide neutral losses are not likely to generate informative peaks.

In contrast the partially mobile peptides demonstrate highly significant class differences between the populations, a difference that was seen at all four of the first amide bonds. The magnitude of neutral loss peak is quite substantial; at the second amide bond (N2-3) the *average* neutral loss peak was 28% of the base peak of the spectrum. We also observed that the neutral loss intensity peaked at N2-3 and decreased moving inward.

The peptides with full proton mobility showed similar behavior to the partially mobile peptides. This group also had statistically significantly differences at the second through forth amide bond. However the neutral loss peaks had a substantially lower intensity compared to the partially mobile class.

Identical behavior with respect to neutral loss was seen in the triply charged data (Table 2b). In this group, the same classes of peptides (those with partially and fully mobile protons) gave statistically significant neutral losses with an intensity peak again at N2-3. The average intensities of the triply charged ions was approximately 60% that seen in the doubly charged ions.

Analysis of neutral loss at the carboxy termini in peptides produced by chymotryptic digest showed that the phenomena also occur at this end of peptides with dependence on proton mobility. Table 2c and 2d shows that in this group, neutral loss peaks are generated for the partially mobile peptides as well as the fully mobile peptides, with the doubly charged spectra having more intense neutral loss peaks than triply charged spectra. Sample size was inadequate to analyze non-mobile peptides. In contrast to the amino terminal, the peak of neutral loss intensity occurs at the first amide bond in chymotryptic peptides.

These findings show that neutral loss is a common phenomenon occurring at both termini, with substantial spectral intensity. Also apparent is that the neutral loss occurrence depends on proton mobility. Trends that emerge are a greater intensity seen in doubly charged verses triply charged peptides as well as a predictable pattern of intensity with respect to distance from the carboxy and amino termini.

## Evaluating residue specific effects on neutral loss using a graphical display

Having established the strong dependence of neutral loss on proton mobility, we sought to analyze the relationship between the residues present on either side of an amide bond and the occurrence of neutral loss at both amino and carboxy termini.

We employed a graphical visualization tool to rapidly assess the effects of the bounding residues for each amide bond moving inward from both ends of the peptide. As a result of our findings above, we limited this analysis to only those peptides with partially and fully mobile protons. The effects of individual residues at a given bond was studied by creating a box plot of the mean normalized intensity and 95% confidence interval (CI) of the neutral loss for each amino acid for both the true positive and true negative datasets. The plots were then visually compared to determine whether the populations differed in the datasets. This graphical approach was taken to simultaneously examine all residues for each bond on both the amino or carboxy sides for both doubly and triply charged peptides. An example of this visualization is shown in Figure 3 for at the N1-2 bond for doubly charged peptides with partial mobility. In this case, a doubly charged peptide has lost an uncharged N1 residue resulting in a doubly charged fragment ion. Figure 3a examines neutral loss according to residue on the amino terminal side of the amide bond (N1) and 3b according to residue on the carboxy side (N2). In these and all other plots, the incorrect population is plotted in red and the correct population in green.

These graphs clearly demonstrate the dramatic positional effect of neighboring amino acids on neutral loss. Not surprisingly, the effect of amino acids is quite dramatically different on opposite sides of the amide bond. Most obvious is the effect of proline when on the carboxy terminal side of an amide bond and aspartic acid on the amino terminal side. However, neutral loss with intensity significantly greater than that expected by chance appears to occur in the setting of many other amino acids surrounding the amide bond, and the graph illustrates how ubiquitous the phenomena is among the amino acids.

## Detailed analysis of the effects of local residues on neutral losses at the amino terminal

We sought to extend our study to characterize the effects of all residues on both sides of the first four amino terminal amide bonds using the graphical format. This analysis was carried out by generating box plots like the one above for the full and partial proton mobility groups for each of the amide bonds. These are sub grouped by proton mobility below.

**Results for peptides with full proton mobility—**Peptides with full proton mobility showed neutral loss that was restricted to proline. Figure 4 shows that dramatic effect of proline on the carboxy terminal side of the amide bond in doubly charged ions in all positions relative to the amino terminal. No other residues were consistently associated with neutral loss with the exceptions of mild effects of serine, threonine, leucine, and valine on the amino terminal side of the second amide bond. Identical results with reduced intensity were seen for triply charged ions (supplemental Figure 1). While there may be other residues associated with neutral loss, this study does not have the statistical power to identify these smaller effects. It is also notable that the statistical differences between the correct and incorrect full proton mobility datasets as a whole (Table 2a) above are due almost entirely to the effect of proline. Without proline, nearly all statistically significant difference seen in the Table 2a between the true positive and true negative populations disappears (data not shown).

These results of neutral loss occurring in association with proline are consistent with literature reports; however the restriction of the phenomenon to only proline containing full mobility peptides is of significance. Salik noted neutral loss peaks and sequence ladders associated with proline [13]. The analysis by Kapp [17], showed fragmentation at the Xaa-Pro bond to be most favored event in b/y type fragmentation. Applying this information to the use of our peptide viewer suggests an observed neutral peak associated with a carboxyl proline supports the peptide assignment. Neutral loss associated with other residues is considerably less informative and should be used with caution. The lack of neutral loss with residues other than proline in full mobility peptides has other serious implications. Because this class of peptides makes up

50% of correct search results and approximately 80% of these peptides do not have a proline bounding any of the first four amide bonds, applications to employ neutral loss information as part of a generalized strategy to adjust confidence of peptide assignment in this group will not be relevant to a large fraction of peptide spectra.

**Results for peptides with partial proton mobility—**Unlike peptides with full proton mobility, peptides with partial proton mobility show extensive neutral loss that occurs with significant intensity for most amino acid residues on both sides of the amide bond, and for all four amide bonds studies (Figure 5). The most dramatic effects are those illustrated above for proline and aspartic acid. The presence of proline on the carboxy terminal side of an amide bond gives extremely strong neutral loss peaks. The *mean* intensity of peaks of this type is approximately 75% of the spectral base peak at N2-3. The trend of peak strength peaking at N2-3 and decreasing thereafter is seen here for proline also.

Likewise, aspartic acid on the amino terminal side of a bond results in a consistent neutral loss event at the amino terminal. This effect is generalizable for residues with acidic side chains, as glutamic acid has a similar effect when positioned on the amino terminal side of the amide bond, however with reduced peak intensity when compared to aspartic acid. Hydrophobic residues are also associated with neutral loss in our analysis. On the amino terminal side of amide bonds at all positions, isoleucine, leucine, and valine were associated with strong neutral loss peaks. Hydrophobic residues were also seen to be associated with neutral loss on the carboxy terminal side of the amide bond. In this position though, the effect is less dramatic and less consistent.

The hydrophilic-neutral residues serine, threonine, and tyrosine behave similarly to the hydrophobic residues, with moderate strength neutral loss peaks seen on both sides of the amide bond. The effect also peaks at N2-3 and declines thereafter. The basic charged residues, histidine and lysine are notable for a significant neutral loss when these residues are located on the carboxy terminal side of an amide bond. Histidine, in particular, was associated with dramatic neutral losses. Glutamine and asparagine were less closely associated with neutral loss, though particular examples were seen where either was. Cysteine (alkylated in this experiment) was not associated in any case. Triply charged peptides with a partially mobile proton shared residue specific features with the doubly charged data. However, in general, the intensities of the neutral loss peaks were smaller. As a consequence, fewer residues on both sides of each bond were associated with a neutral loss (supplementary Figure 2).

The extensive nature of neutral loss with in peptides with partial mobility suggests significant opportunities to extract sequence specific information from spectra. The finding of enhanced neutral loss with proline on the carboxy side of a bond in partially mobile peptides is consistent with results from fully mobile data discussed above. The effect of aspartic and glutamic acid residues on the amino terminal side of the amide bonds is consistent with previous reports showing facile cleavage C-terminal to Asp in peptides with non-mobile protons [26, 27]. It has been suggested that this type of cleavage does not involve the ionizing proton and is therefore “charge remote”, occurring through the COOH group of the side chain [28, 29].

The analysis by Kapp et al. [17] calculated a ratio of cleavage intensity (CIR) for all residues on both sides of the amide bond demonstrating that in partially mobile peptides fragmentation at the Xaa-Pro bond was the most favored event. They also noted that there was enhanced cleavage at bonds containing either an aspartic acid or glutamic acid residue on the N-terminal side, with the Glu-Xaa bond approximately ~50% less labile than Asp. While this analysis was directed at cleavage within peptides through b and y ion pathway, not neutral loss, the phenomenon may be applicable to bond breakage leading to neutral loss. Our analysis confirms

both the easy lability of the amide bond when an acidic residue is located on the amino terminal side and that neutral loss associated with Glu is substantially less intense than that of Asp.

The above analysis of CIR data also demonstrated that Leu-Xaa, Val-Xaa, and Ile-Xaa were also associated with preferential cleavage, a finding that we also observed to be applicable in the setting of neutral loss.

### Adjustment for the bias due to particularly labile residues

Careful examination of nearly any of the neutral loss graphs from peptides with partially mobile protons reveal that there tend to be a few residues on both sides of the amide bond associated with very strong neutral loss peaks, i.e. proline and histidine on the carboxy terminal, and aspartic acid and hydrophobic residues on the amino terminal side. Because all residues on the amino terminal side could be paired with a proline and likewise any carboxy terminal residue could be paired with an aspartic acid, we sought to determine whether observed neutral loss intensity resulted from particularly labile residues such as proline confounding the data on the opposite side of the bond.

We reanalyzed our dataset to remove the effects of specific residues. While examination across the matrix of all bond pairs would be the best solution, it was not possible due to the limited size of our database. Instead, the studies described above were repeated for each bond with the most labile residues removed from the dataset. In this way we could see whether the mean neutral loss associated with some residues was in fact being elevated through pairings with “active” residues across the amide bond. These data are presented in Figure 6 for doubly charged data.

The adjusted results for doubly charged peptides with partially mobile protons by amino and carboxy termini did not differ dramatically from the unadjusted results; most of the residues associated with a neutral loss were still quite easily differentiated from the incorrect set. There was a general downward movement of the means for all of the residues, which would be expected when data with the most intense neutral loss peaks is removed. In a few cases, such as serine and threonine on the amino terminal side of N4-5, the mean percent base peak became less easily distinguished between correct and incorrect populations. It is possible that with a larger sample size, this distinction would be more robust.

The adjusted results for triply charged peptides with partially mobile protons also did not differ dramatically from the unadjusted results though for some residues, the correction made a clear distinction between true and false populations more difficult (supplementary Figure 3).

These results indicated that the association of neutral loss with particular residues is not due to the effects of residues across the amide bond. Correction for this effect produced similar results as the uncorrected; many residues are associated with neutral loss. This information can be utilized in analyzing the peptide assignment to a given spectra.

### Neutral loss at the carboxy terminal

A separate analysis was performed to examine whether particular residues could be linked to a neutral loss event at the carboxy terminal of a peptide. Loss at the carboxy terminal generates b type ions with a charge unchanged from that of the parent. The database used for this analysis was derived from 651 doubly charged and 659 triply charged peptides generated by chymotryptic digest.

In the C-terminal analysis the partial and full mobility data were combined. This was done for two reasons. First the database size was considerably smaller than that for the investigation of N-terminal neutral losses. Second, chymotryptic peptides lack a carboxy terminal basic residue

to localize charge. This means that internal charge localization is possible in chymotryptic peptides even with only a single basic residue present (full mobility). Separate analysis of these states of proton mobility demonstrated similar behavior between the classes (data not shown).

Neutral losses occur at the carboxy terminal of chymotryptic peptides with strong intensities and are associated with a large number of amino acids on either side of the amide bond. Figure 7 shows that for doubly charged peptides the effect of the residue on the amino side is far more influential and is similar to that seen earlier in the tryptic dataset. In this setting acidic residues and hydrophobic residues generally are associated with strong peaks, with aspartic acid again having the strongest effect. On the carboxy terminal side of the amide bond, only leucine was consistently associated with neutral loss. Neutral loss from triply charged chymotryptic peptides was only appreciably observed at the amino terminal side of C1-2, with mean intensities significantly lower than those seen in the doubly charged population (supplementary Figure 4).

Neutral loss occurs at either end of a protonated peptide, with the amino acid sequence determining the intensity. In the case of the tryptic peptides, the protonated carboxy terminal basic residue makes a loss at this end unlikely as all fragments retain a proton and are therefore charged. In the case of tryptic mis-cleavage or cleavage with other proteases such as chymotrypsin that do not generate peptides with basic C-termini, neutral loss occurs with significant intensity. Like the effect in partially mobile tryptic peptides, this effect can add useful information in the analysis of a spectrum. In this case however, the information could potentially be incorporated into a search strategy because it is independent of the proton mobility and can occur in all peptides.

## Implications for proposed mechanisms of amino terminal neutral loss

The mechanisms of neutral loss occurring at the amino terminal have been investigated by a number of groups using reionization experiments [7, 9–12]. We applied our neutral loss database to determine whether they supported any of the proposed mechanisms. Neutral loss at the N1, N2, and N3 bonds and beyond are thought to occur through different mechanisms (Figure 9). The loss of the N-terminal residue occurs through nucleophilic attack of the N-terminal nitrogen at the first carbonyl carbon of the backbone, generating a neutral three-member aziridinone ring (9A). The loss of a *dipeptide* neutral unit results from a nucleophilic attack of the N-terminal nitrogen at the second carbonyl carbon of the peptide backbone forming a six-member ring diketopiperazine. In both these cases, because the proton affinities of cyclic peptides are lower than those of linear peptides, the extra proton transfers to the C-terminal fragment and the complex dissociates to form the y ion and the cyclic peptide as its neutral counterpart.

For losses occurring beyond the second carbonyl carbon three or more residues are most likely eliminated as cyclic peptides, however it is not settled as to whether these peptides are macrocycles or substituted smaller cyclic peptides. One model proposed holds that for species with three and four glycyl residues calculations predict cyclic dipeptides (diketopiperazines) to be more stable [11, 13], with larger rings becoming comparably favorable with increasing size. A second model proposes that the corresponding neutrals are cyclotri-, cyclo-tetra, and cyclo-penta-peptides, respectively [2, 12]. We sought to use our database to determine whether support could be found for the macrocyclic or substituted smaller cyclic peptide model. Our hypothesis was that a macrocyclic model places the N1 residue adjacent to the amide bond while in the substituted small cyclic model the N1 is be further from the bond. Strong effects of N1 individual residues on the N3-4 and N4-5 bonds would imply a closer proximity and support the macrocyclic model.

We constructed graphs for the neutral loss at the N2-3, N3-4, and N4-5 bonds sorted by the N1 residue. In stark contrast to the effects of the amino and carboxy terminal amino acids of these bonds, the identity of the N1 residue seemed to have little effect of the intensity of a neutral loss peak. All residue specific peaks were essentially indistinguishable from one another. The exception was that the basic residues at N1 were not associated with any neutral loss.

These findings suggest that there is little effect of the N1 amino acid on neutral loss at bonds that it does not bound. The uniform distribution of peak intensity occurs because while there is neutral loss occurring, it is divided among the N1 residues randomly, implying that that the side chain properties are not influential. The absence of neutral loss with basic residues at N1 results because in this position they sequester a proton and ions lost are therefore charged. While not mechanistic evidence, the impressive effects of the resides bounding the amide bond on the amino side, and the lack of any effect of the N1 residues on N2-3 and inward support the substituted smaller cyclic peptide over a macrocyclic model

## Conclusions

Our current utilization of the information available from neutral loss occurring during CID has great utility for checking the peptide assignment made by SEQUEST for a spectrum. We frequently find peptides of borderline significance after SEQUEST search are often confirmed through the matches of the neutral loss peaks. It is our intention to take greater advantage of this data in order to incorporate it into either a search algorithm such as Sequest or a secondary application that reviews search results, such as PeptideProphet or LOD based scores [22, 30], and applies a probability to each one. It may be that interrogation of neutral loss data, because it is spectral, could allow a secondary evaluation of the top search results to identify any that contain highly suggestive features that could serve to alert the user that an assignment is good, or that a peptide scoring lower deserves closer inspection. We have shown that neutral loss is a common occurrence in CID performed on an ion trap instrument. It takes place at both termini, depending on peptide characteristics. For the analysis of tryptic peptides, which is currently a method of choice, it is frequently one of the defining characteristics of the spectra. We have begun to clarify the rules that govern this phenomena, which, when incorporated into analysis software, will improve our ability to assign spectra with confidence.

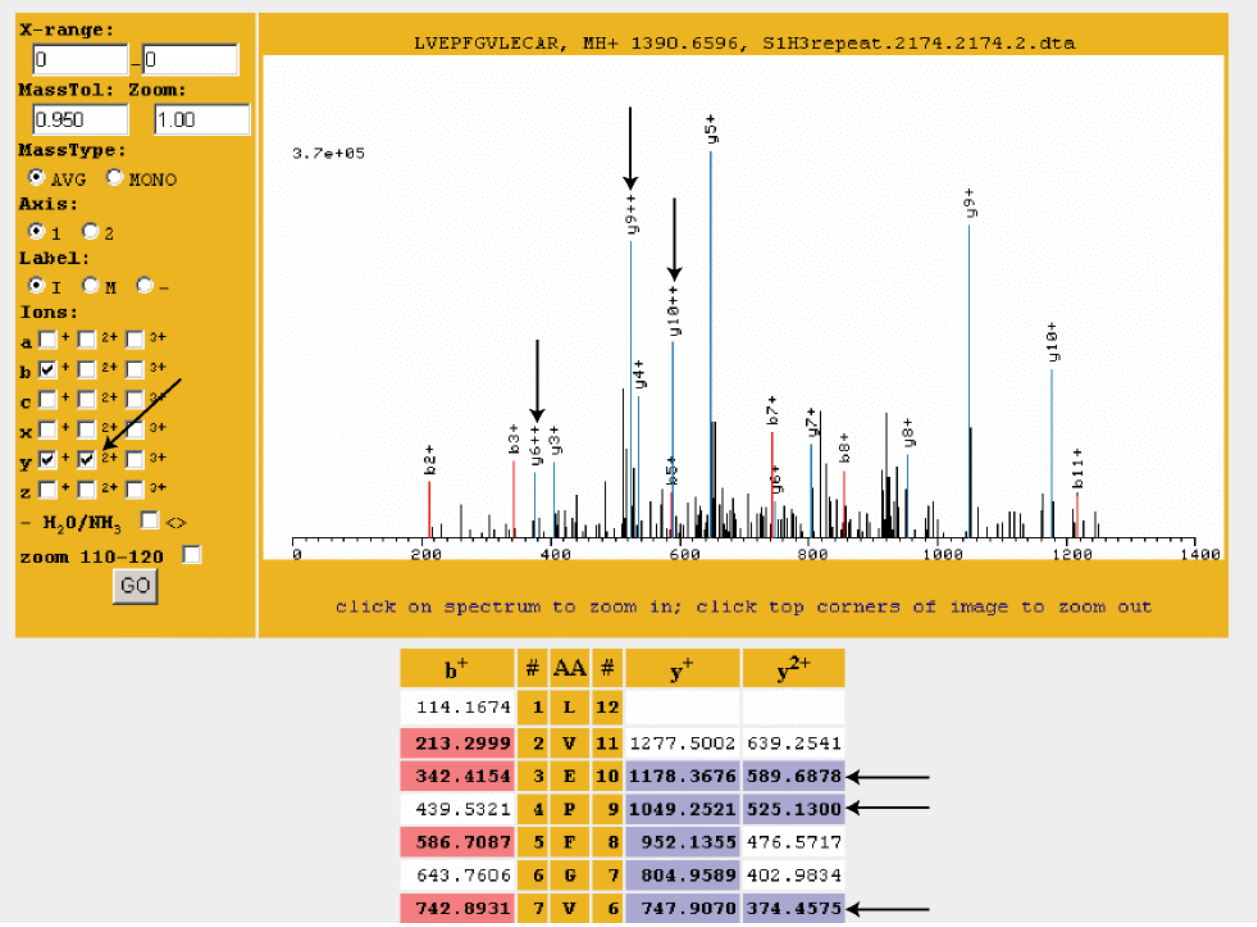
#### Acknowledgements

This work was supported by National Cancer Institute grant K08 CA097282 to D. Martin, contract N01-HV-28179 from the National Heart, Lung, and Blood Institute to R. Aebersold

## References

1. Hunt DF, Yates JR 3rd, Shabanowitz J, Winston S, Hauer CR. Proc Natl Acad Sci U S A 1986;83:6233–6237. [PubMed: 3462691]
2. Paizs, B.; Suhai, S. Mass Spectrom Rev 2004.
3. Roepstorff P, Fohlman J. Biomed Mass Spectrom 1984;11:601. [PubMed: 6525415]
4. Biemann K. Methods Enzymol 1990;193:886–887. [PubMed: 2074849]
5. Eng JK, McCormack AL, Yates JRr III. J Am Soc Mass Spectrom 1994:976–989.
6. Perkins DN, Pappin DJ, Creasy DM, Cottrell JS. Electrophoresis 1999;20:3551–3567. [PubMed: 10612281]
7. Cordero MM, Houser JJ, Wesdemiotis C. Anal Chem 1993;65:1594–1601. [PubMed: 8328675]
8. Yalcin, T.; Khouw 1995.
9. Nold MJ, Wesdemiotis C, Yalcin T, Harrison AG. International Journal of Mass Spectrometry and Ion Processes 1997;164:137–153.
10. Nold MJ, Cerda BA, Wesdemiotis C. J Am Soc Mass Spectrom 1999;10:1–8. [PubMed: 9888180]

11. Polce MJ, Ren D, Wesdemiotis C. Journal of Mass Spectrometry 2000;35:1391–1398. [PubMed: 11180629]
12. Paizs B, Suhai S. Journal of the American Society for Mass Spectrometry 2004;15:103–113. [PubMed: 14698560]
13. Salek M, Lehmann WD. J Mass Spectrom 2003;38:1143–1149. [PubMed: 14648821]
14. Wysocki VH, Tsaprailis G, Smith LL, Breci LA. J Mass Spectrom 2000;35:1399–1406. [PubMed: 11180630]
15. Cox KA, Gaskell SJ, Morris M, Whiting A. Journal of the American Society for Mass Spectrometry 1996;7:522–531.
16. Jones J, Dongre AR, Somogyi A, Wysocki VH. J Am Chem Soc 1994;116:8368 – 8369.
17. Kapp EA, Schutz F, Reid GE, Eddes JS, Moritz RL, O'Hair RA, Speed TP, Simpson RJ. Anal Chem 2003;75:6251–6264. [PubMed: 14616009]
18. Smolka M, Zhou H, Aebersold R. Mol Cell Proteomics 2002;1:19–29. [PubMed: 12096137]
19. Martin DB, Gifford DR, Wright ME, Keller A, Yi E, Goodlett DR, Aebersold R, Nelson PS. Cancer Res 2004;64:347–355. [PubMed: 14729644]
20. Yi EC, Lee H, Aebersold R, Goodlett DR. Rapid Commun Mass Spectrom 2003;17:2093–2098. [PubMed: 12955739]
21. Gygi SP, Han DK, Gingras AC, Sonenberg N, Aebersold R. Electrophoresis 1999;20:310–319. [PubMed: 10197438]
22. Nesvizhskii AI, Keller A, Kolker E, Aebersold R. Anal Chem 2003;75:4646–4658. [PubMed: 14632076]
23. Han DK, Eng J, Zhou H, Aebersold R. Nat Biotechnol 2001;19:946–951. [PubMed: 11581660]
24. Biemann K. Biomed Environ Mass Spectrom 1988;16:99–111. [PubMed: 3072035]
25. Creighton, T. E. Proteins-Structures and Molecular Properties; W.H Freeman and Company: New York, 1993.
26. Yu W, Vath JE, Huberty MC, Martin SA. Anal Chem 1993;65:3015–3023. [PubMed: 8256865]
27. Qin J, Chait B. Journal of the American Chemical Society 1995;119:5411–5412.
28. Lee SW, Kim H, Beauchamp JL. J Am Chem Soc 1998;120:3188–3195.
29. Gu C, Tsaprailis G, Breci L, Wysocki VH. Anal Chem 2000;72:5804–5813. [PubMed: 11128940]
30. Elias JE, Gibbons FD, King OD, Roth FP, Gygi SP. Nat Biotechnol 2004;22:214–219. [PubMed: 14730315]




| $b^+$ | # | AA | # | $y^+$ | $y^{2+}$ |
|---|---|---|---|---|---|
| 114.1674 | 1 | L | 12 | | |
| 213.2999 | 2 | V | 11 | 1277.5002 | 639.2541 |
| 342.4154 | 3 | E | 10 | 1178.3676 | 589.6878 |
| 439.5321 | 4 | P | 9 | 1049.2521 | 525.1300 |
| 586.7087 | 5 | F | 8 | 952.1355 | 476.5717 |
| 643.7606 | 6 | G | 7 | 804.9589 | 402.9834 |
| 742.8931 | 7 | V | 6 | 747.9070 | 374.4575 |

**Figure 1.**

A modified spectral viewer was designed that automatically assigned matches to neutral loss peaks from the amino or carboxy termini. The acquired spectrum of a doubly charged precursor is shown. The viewer shows matches to doubly charged neutral loss peaks (arrows on spectra), and can be adjusted through an interface (arrow at left) to label peaks of any ion series. While labeled with this viewer, the neutral loss ions were not used by Sequest to assign a sequence to this spectrum. In this example the box for doubly charged y ions is checked, which is our default view for doubly charged precursors. The matched peaks are listed by ion series in tabular form below. In addition the viewer can be used to observe neutral losses of water/ammonium, to ion matches (I), or peak mass (M).


**Figure 2.**
Peptides numbered from the amino and carboxy termini. Amide bonds are labeled according to the end of the peptide and residues on both sides. A peptide bond between the second and third residues from the amino terminus is labeled N2-3.

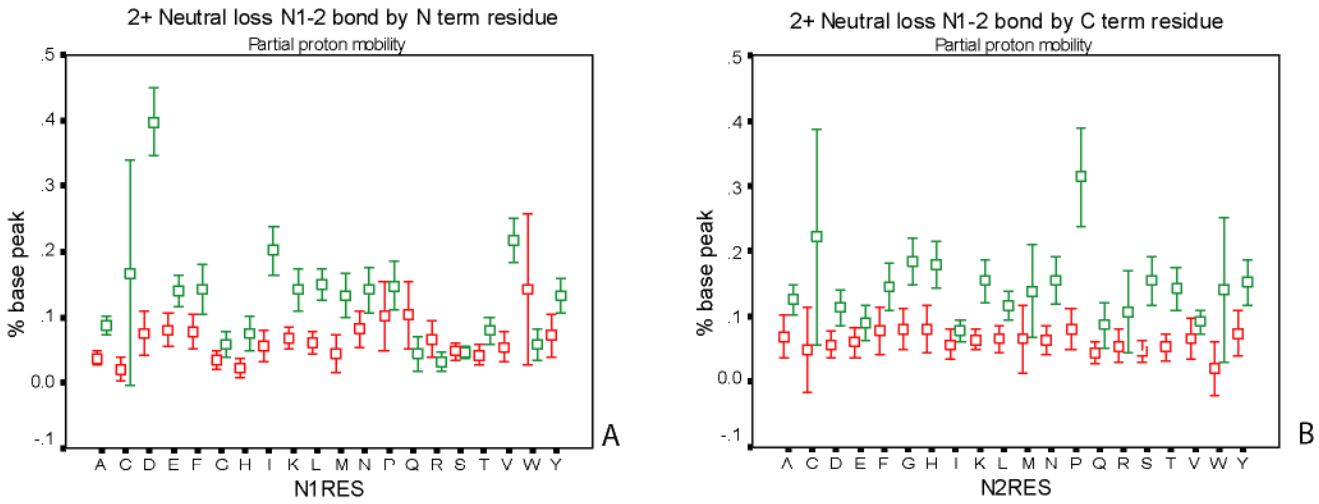



**Figure 3.**

Neutral loss viewer. Loss of neutral fragments at the N1-2 bond for doubly charged peptides with partial proton mobility is graphed according to amino acid. The mean normalized neutral loss intensity is indicated by the box and the 95% confidence interval by the bars. The true negative data is plotted in red and true positive data in green. Results are plotted for all amino acids on both the amino (A) and carboxy (B) side of the N1-2 amide bond.

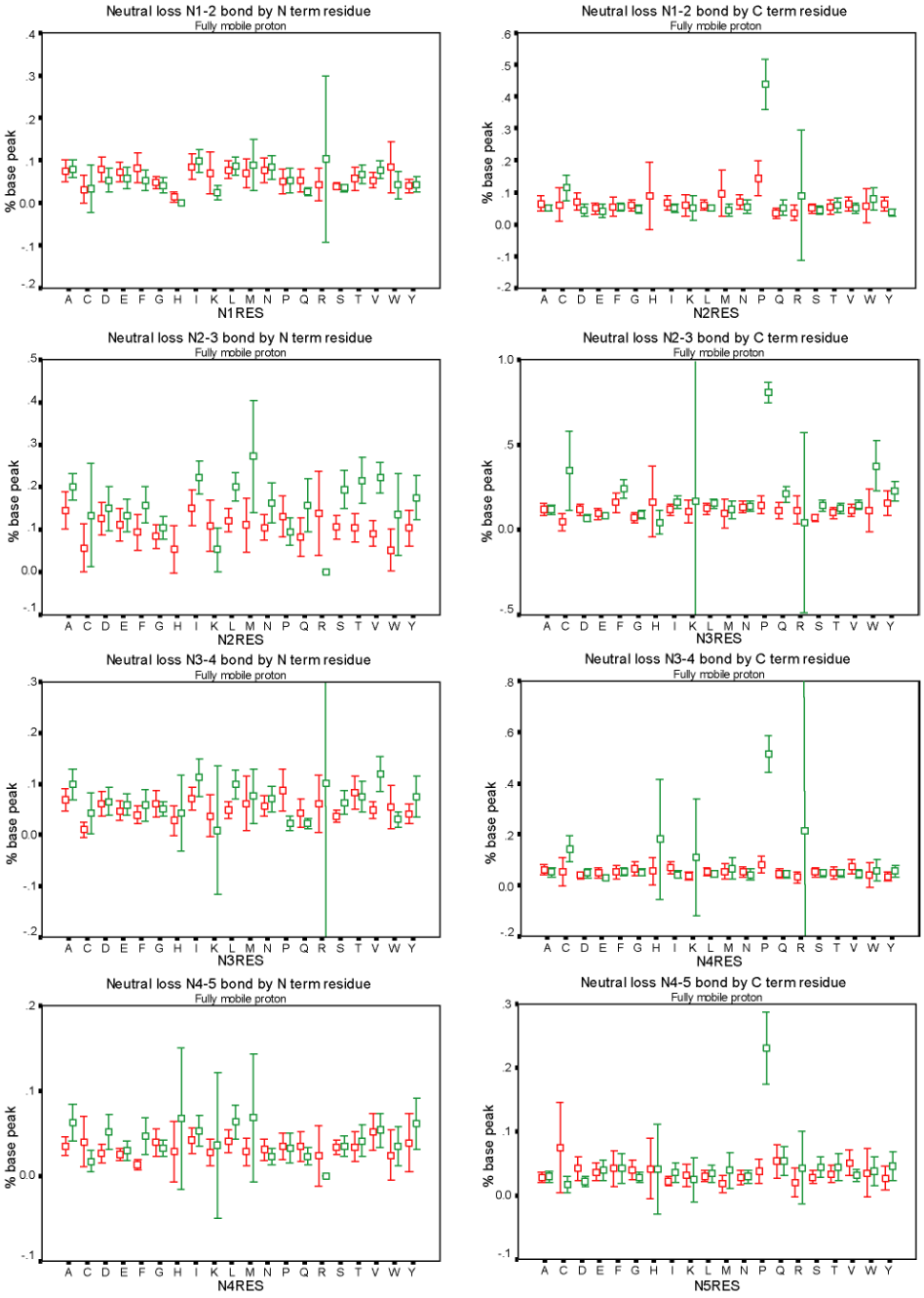



**Figure 4.**

Loss of neutral fragments for doubly charged tryptic ions with full proton mobility at the first four amino terminal amide bonds. The normalized intensity for incorrect (red) and correct (green) populations was plotted for all amino acids on the amino (left column) and carboxy side (right column) of the first four amide bonds. Ninety-five percent confidence intervals are given by the error bars. Truncated error bars are the result of a small sample size resulting in a large confidence interval.

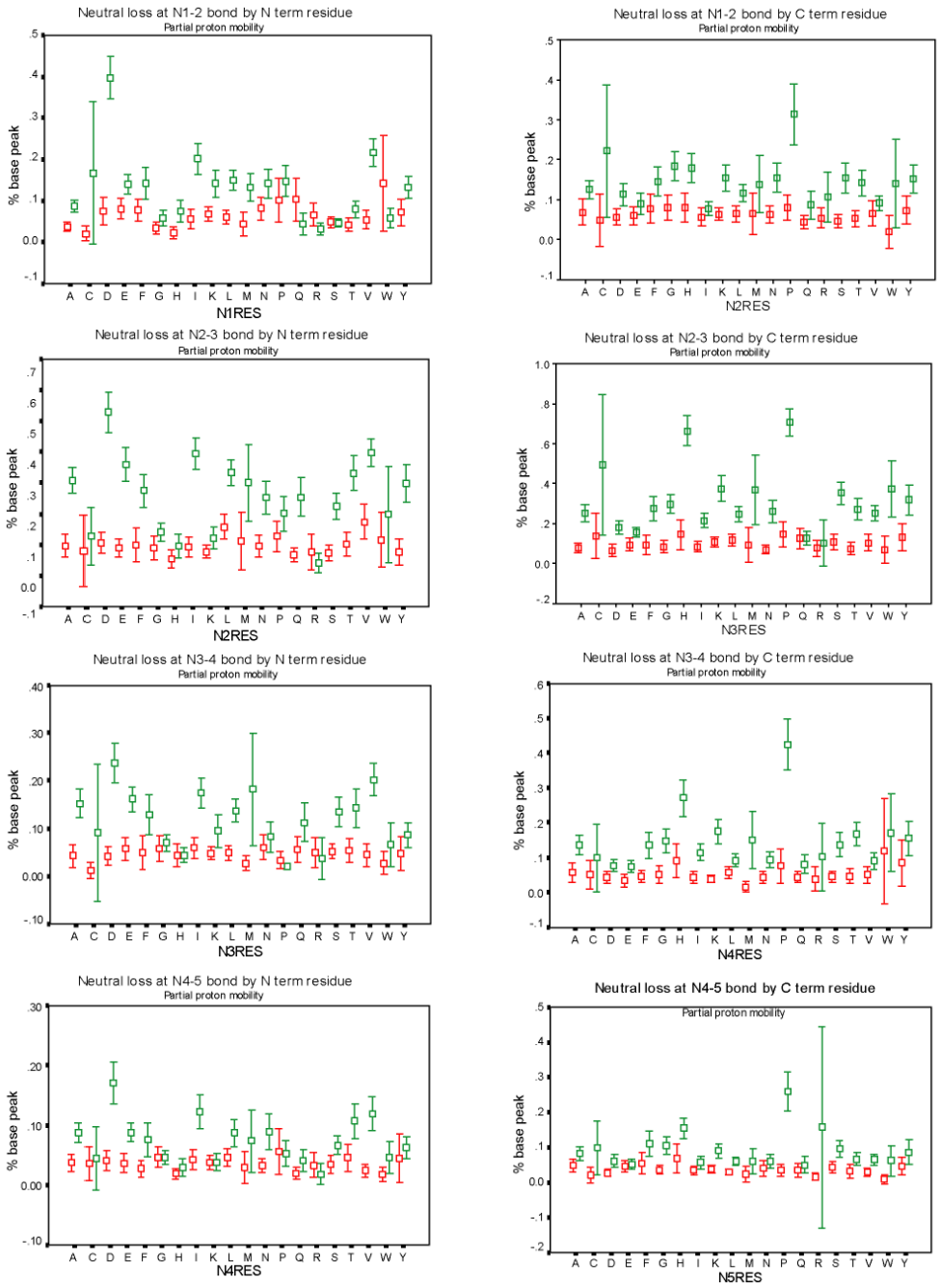



**Figure 5.**

Loss of neutral fragments for doubly charged tryptic ions with partial proton mobility at the first four amino terminal amide bonds. The normalized intensity for incorrect (red) and correct (green) populations was plotted for all amino acids on the amino (left column) and carboxy side (right column) of the first four amide bonds.

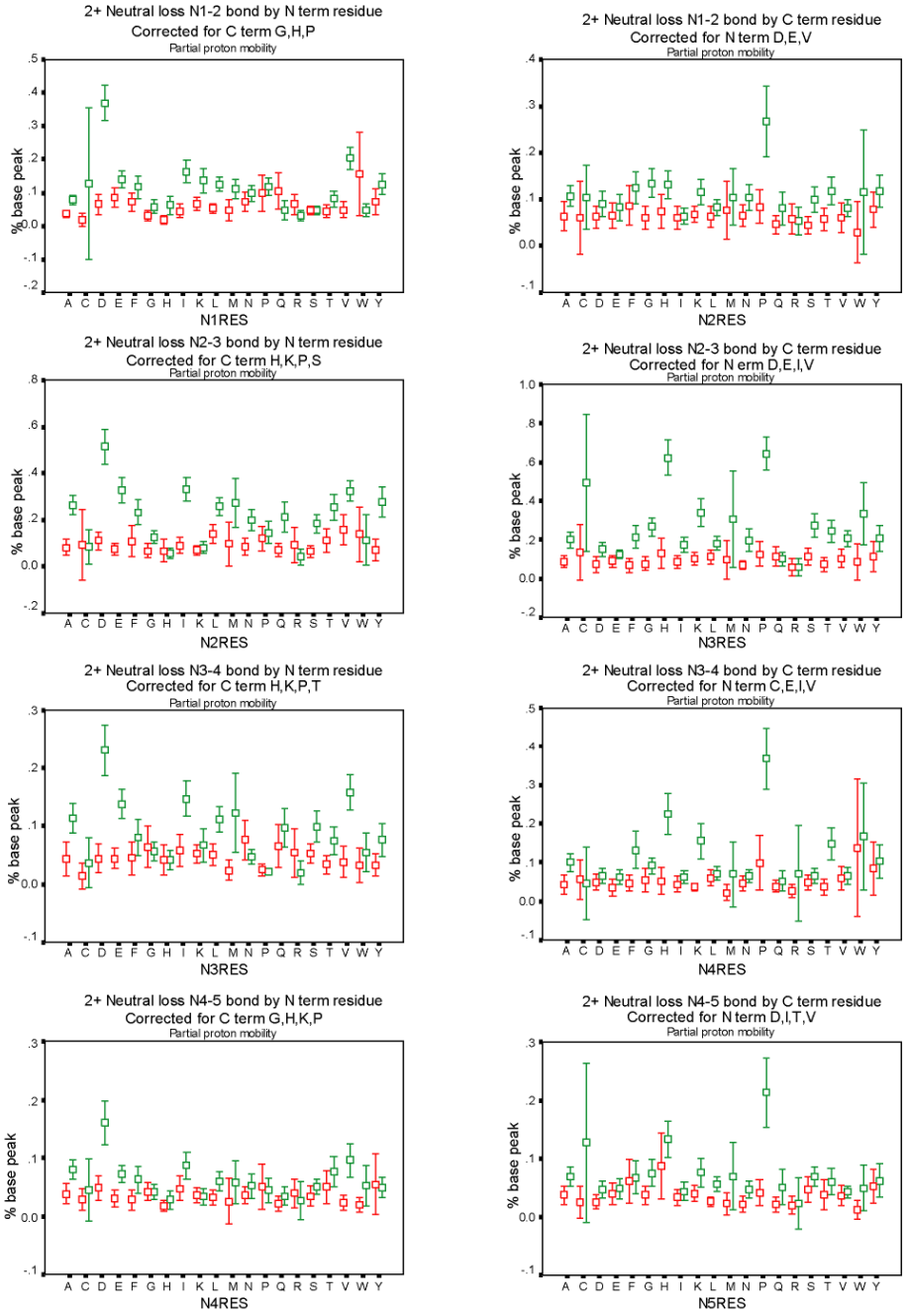



**Figure 6.**

Loss of neutral fragments for doubly charged tryptic ions with partial mobility at the first four amino terminal amide bonds with labile residues removed. At each position, for residues on both sides of the amide bond, data was reanalyzed by removing those spectra that contained residues with strong neutral loss peaks on the opposite side of the bond. For example, in 7A, in the neutral loss at N1-2 by amino terminal residue, those spectra with labile proline, histidine, or glycine residues on the *carboxy* terminal side were removed from the analysis.

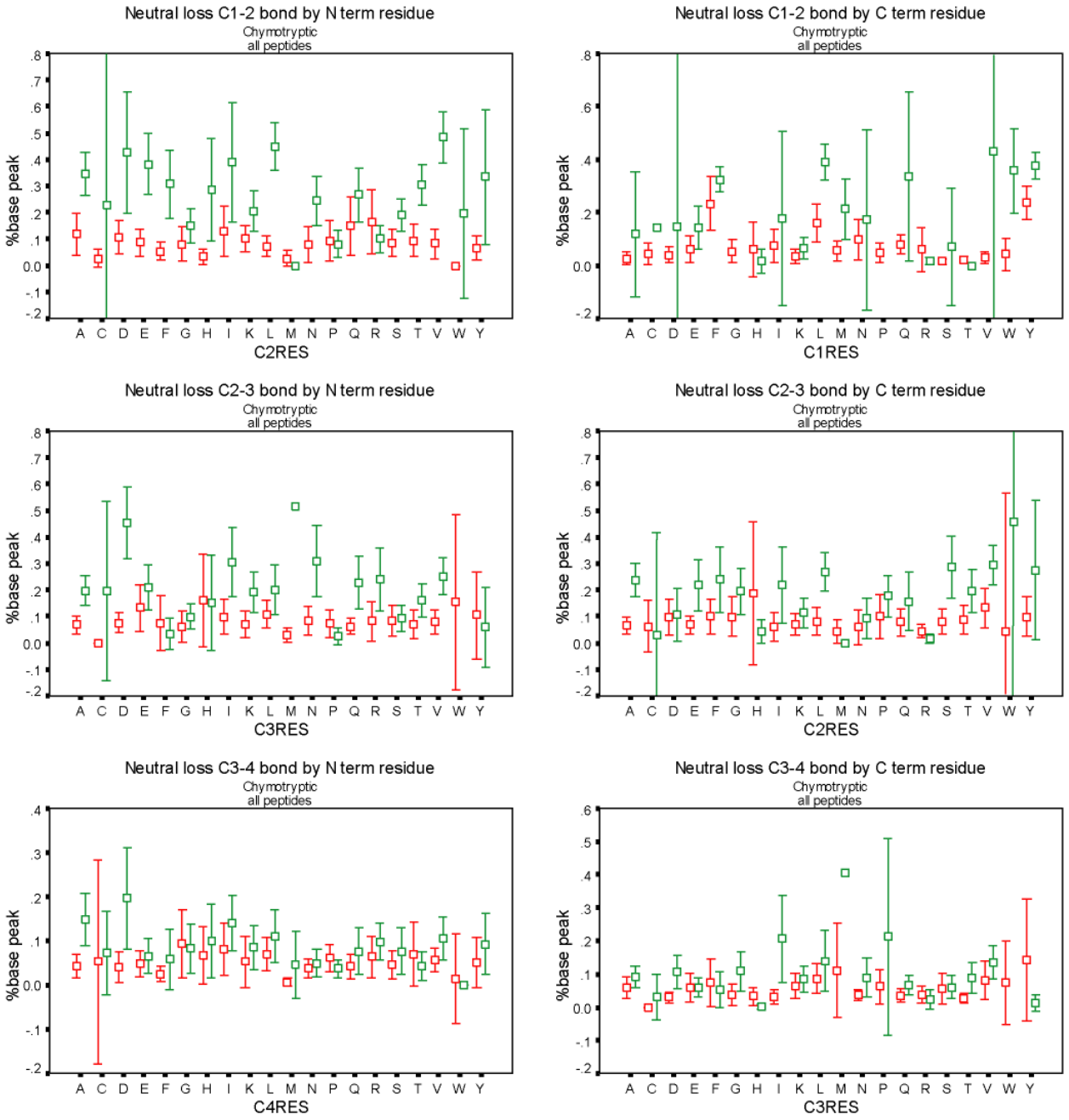



**Figure 7.**

Loss of neutral fragments for doubly charged chymotryptic ions without separation by proton mobility at the first three carboxy terminal amide bonds. The normalized intensity for incorrect (red) and correct (green) populations was plotted for all amino acids on the amino and carboxy side of the first four amide bonds moving inward from the C-terminus

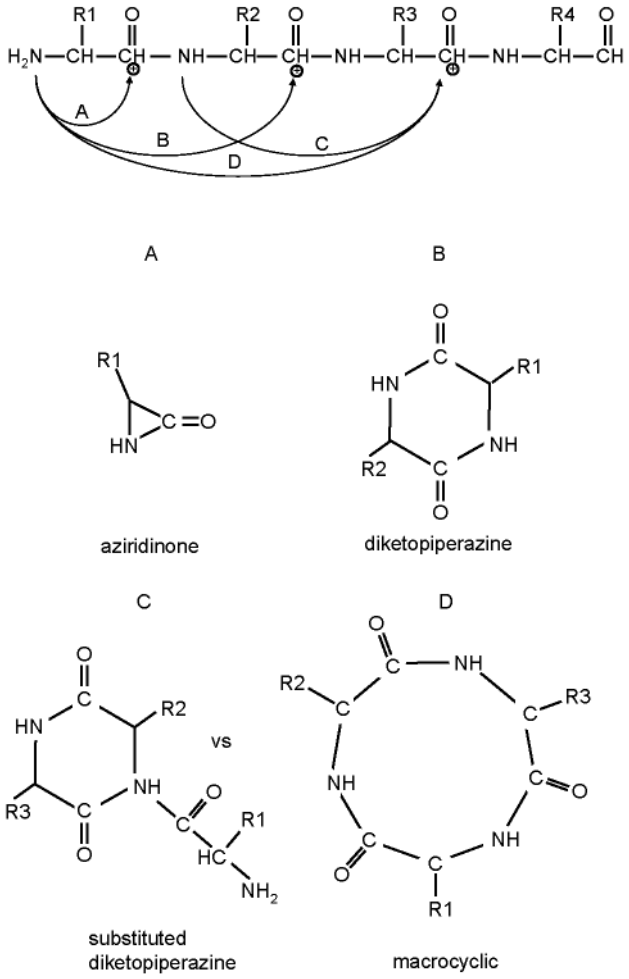



**Figure 8.**

Mechanisms of neutral loss. The loss of the N-terminal residue occurs through nucleophilic attack of the N-terminal nitrogen at the first carbonyl carbon of the backbone generating a three-member aziridinone ring structure (A). Loss of a *dipeptide* neutral unit (B) results from a nucleophilic attack of the N-terminal nitrogen at the second carbonyl carbon of the peptide backbone forming a six-member ring diketopiperazine. Losses occurring distal to the second carbonyl carbon three or more residues are most likely eliminated as cyclic peptides either as substituted cyclic dipeptides (C) or as macrocyclics (D).

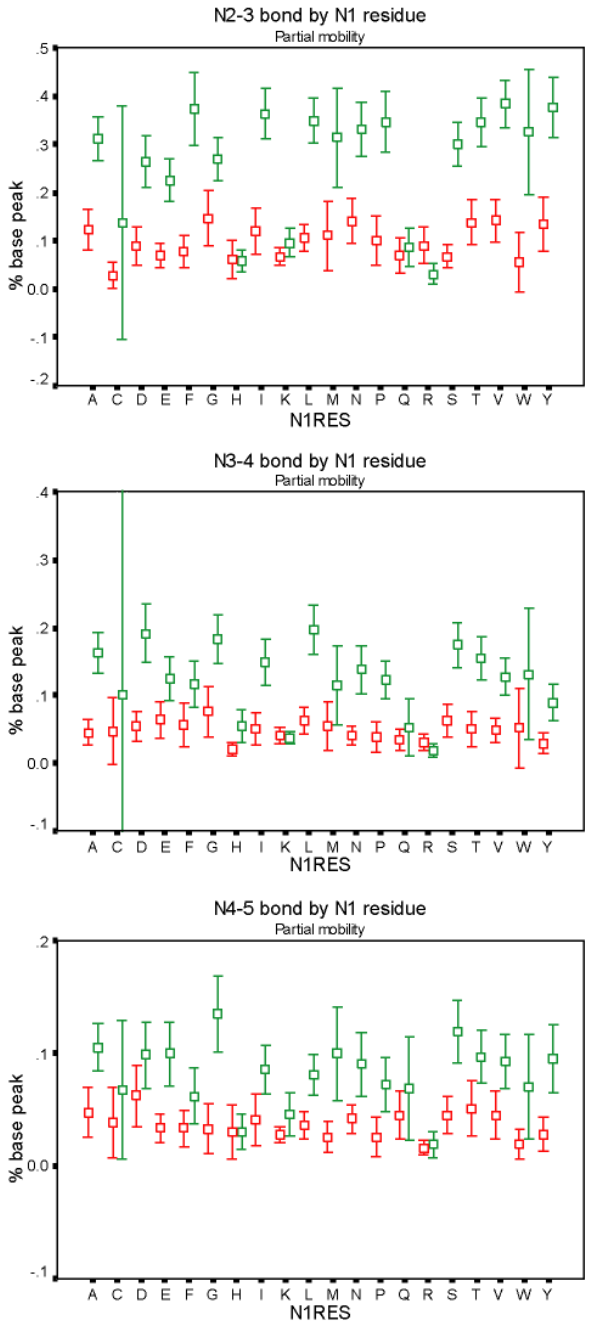



**Figure 9.**
Neutral loss at the amino terminal bonds N2-3, N3-4, and N4-5 according to the N1 residue.

## Table 1

Mobility states and Arg/Lys numbers compared. A: Distribution of proton mobility states for doubly and triply charged peptides. B: Distribution of arginine and lysine residue number in a subset of correct doubly charged peptides that excludes proline residues. The exclusion prevents bias due to missed cleavages at the sequence KP.

**a**

| | 2+ Correct | | 3+ Correct | |
|---|---|---|---|---|
| | Count | % | Count | % |
| **Non-mobile** | 110 | 2.20% | 13 | 0.40% |
| **Partially mobile** | 2520 | 51.30% | 1291 | 36.90% |
| **Mobile** | 2287 | 46.50% | 2197 | 62.80% |

**b**

| | R | K |
|---|---|---|
| **0** | 1340 | 836 |
| **1** | 1002 | 1122 |
| **2** | 50 | 370 |
| **3** | 2 | 63 |
| **4** | 0 | 3 |

## Table 2

Average intensity of neutral loss. Neutral loss intensity normalized to spectral base peak is given for each amide bond moving inward from the amino and carboxy termini, comparing the results for the true negative (TN) and true positive (TP) data. Tables are subdivided by row for proton mobility. A: results for doubly charged tryptic peptides. B: triply charged peptides. C: Chymotryptic data, doubly charged. D: Chymotryptic data, triply charged. C and D include entries only for partially and fully mobile data because of limited data size. Statistical significance for differences between means is indicated for each comparison.

**A Tryptic doubly charged**

| | | Bond N1-2 mean % base peak | *p* | Bond N2-3 mean % base peak | *p* | Bond N3-4 mean % base peak | *p* | Bond N4-5 mean % base peak | *p* |
|---|---|---|---|---|---|---|---|---|---|
| Non-mobile proton | TN | 0.063 | 0.143 | 0.086 | 0.831 | 0.053 | 0.071 | 0.035 | 0.526 |
| | TP | 0.094 | | 0.082 | | 0.090 | | 0.044 | |
| Partialy mobile | TN | 0.063 | <.001 | 0.100 | <.001 | 0.049 | <.001 | 0.038 | <.001 |
| | TP | 0.135 | | 0.287 | | 0.135 | | 0.087 | |
| Fully mobile | TN | 0.064 | 0.489 | 0.113 | <.001 | 0.056 | <.001 | 0.035 | 0.001 |
| | TP | 0.067 | | 0.180 | | 0.074 | | 0.045 | |

**B Tryptic triply charged**

| | | Bond N1-2 mean % base peak | *p* | Bond N2-3 mean % base peak | *p* | Bond N3-4 mean % base peak | *p* | Bond N4-5 mean % base peak | *p* |
|---|---|---|---|---|---|---|---|---|---|
| Non-mobile proton | TN | 0.054 | 0.267 | 0.050 | 0.220 | 0.047 | 0.902 | 0.032 | 0.574 |
| | TP | 0.039 | | 0.113 | | 0.044 | | 0.045 | |
| Partialy mobile | TN | 0.053 | <.001 | 0.068 | <.001 | 0.043 | <.001 | 0.029 | <.001 |
| | TP | 0.107 | | 0.177 | | 0.071 | | 0.055 | |
| Fully mobile | TN | 0.052 | 0.657 | 0.054 | <.001 | 0.037 | 0.006 | 0.027 | 0.007 |
| | TP | 0.054 | | 0.089 | | | | 0.033 | |

**C Chymotryptic doubly charged**

| | | Bond C1-2 mean % base peak | *p* | Bond C2-3 mean % base peak | *p* | Bond C3-4 mean % base peak | *p* |
|---|---|---|---|---|---|---|---|
| Partialy mobile | TN | 0.084 | <.001 | 0.077 | <.001 | 0.058 | <.001 |
| | TP | 0.371 | | 0.272 | | 0.125 | |
| Fully mobile | TN | 0.096 | <.001 | 0.092 | 0.003 | 0.050 | 0.140 |
| | TP | 0.240 | | 0.143 | | 0.067 | |

**D Chymotryptic triply charged**

| | | Bond C1-2 mean % base peak | *p* | Bond C2-3 mean % base peak | *p* | Bond C3-4 mean % base peak | *p* |
|---|---|---|---|---|---|---|---|
| Partialy mobile | TN | 0.071 | <.001 | 0.060 | <.001 | 0.037 | <.001 |
| | TP | 0.238 | | 0.138 | | 0.079 | |
| Fully mobile | TN | 0.069 | 0.131 | 0.076 | 0.167 | 0.042 | 0.480 |
| | TP | 0.088 | | 0.058 | | 0.049 | |