The influence of histidine on cleavage C-terminal to acidic residues in doubly protonated tryptic peptides

Yingying Huang a, Vicki H. Wysocki a,*, David L. Tabb b, John R. Yates, III c

a Department of Chemistry, University of Arizona, Tucson, AZ 85721-0041, USA
b Department of Genome Sciences, University of Washington, Seattle, WA 98195, USA
c Department of Cell Biology, The Scripps Research Institute, La Jolla, CA 92037, USA

Received 20 November 2001; accepted 8 February 2002

Abstract

An ion-trap CID MS/MS spectral database of 505 doubly protonated tryptic peptides was used to investigate the influence of an internal basic residue on preferential fragmentation C-terminal to the acidic amino acid residues, aspartic acid (Asp) and glutamic acid (Glu). Because tryptic peptides, which contain C-terminal Lys or Arg, were selected for analysis, the majority of the peptides contain His as the internal basic residue. A comparison between spectra for peptides that do and do not contain an internal basic residue shows that cleavage is more prominent at Asp-Xxx bonds for peptides that do contain the internal basic residue. This result corroborates a previously published mechanism for Asp-Xxx cleavage that states that cleavage at Asp-Xxx is enhanced when protons are sequestered at basic sites, allowing the acidic hydrogen of the Asp side chain to initiate cleavage. The data suggest that in doubly-charged His-containing tryptic peptides, one proton is typically located at the C-terminal Arg or Lys while the mobility of the second proton is hindered by the His side chain’s relatively high basicity. The same investigation was performed for cleavage at Glu-Xxx amide bonds, but in this case there is only a marginal difference between peptides that do and do not contain the internal basic residue. (Int J Mass Spectrom 219 (2002) 233–244) © 2002 Elsevier Science B.V. All rights reserved.

Keywords: Statistical analysis; Tandem mass spectra; MS/MS spectral database; Tryptic peptides; Collision-induced dissociation; Fragmentation pattern; Histidine; Arginine; Lysine; Asp-Xxx cleavage; Glu-Xxx cleavage

1. Introduction

Peptide and protein structural characterization by tandem mass spectrometry is in widespread use [1–6]. However, greater accuracy in sequence determination requires improvements in current algorithms [7–15] for automated data analysis. Most algorithms use a simplistic method of predicting peptide fragmentation patterns that assumes that amide bond cleavage occurs uniformly throughout peptides, yielding fragment ions of equal abundance. While this approach is sufficient to identify many peptides, other peptides exhibit fragmentation patterns that significantly deviate from this model. A study by Reid and co-workers on the spectral identification of 319 peptides by SEQUEST [16], one of the most extensively used programs, shows that only 40% of the spectra were identified without any human intervention, while 24% of the spectra are classified as having "low information content" [16]. Results of this type underscore the need for more research into
gas phase peptide dissociation mechanisms, in order to make more effective algorithms possible. A general model has been established to explain how peptides fragment in the gas phase [17–20]. A series of publications [17–39] have been devoted to systematic mechanistic studies that reveal peptide fragmentation rules. The Wysocki laboratory is now attempting to identify characteristic fragmentation patterns and validate possible mechanisms by examining collections of spectra. The influence of the various side chains on fragmentation, individually and in combination, can be discerned by this approach. Computer programs to select spectra of interest help to narrow this search [40]. The analysis of acidic residue cleavage described below demonstrates this strategy.

Selective gas phase cleavage at the peptide bond C-terminal to an Asp residue (i.e., Asp-Xxx) in protonated or sodiated peptide ions has been investigated by several research groups [35,38,41–45]. Recent investigations show that when the ionizing protons are not “mobile,” i.e., when they are sequestered by Arg, or when a fixed-charge moiety is substituted for a protonated Arg, cleavage catalyzed by the acidic hydrogen of the Asp side chain, which requires a seven-member ring intermediate, becomes pronounced [38]. This leads to a five-membered ring succinic anhydride b ion structure [34] in contrast to the protonated oxazolone structure [46–48] typically assumed for b ions. This Asp-Xxx cleavage often results in a MS/MS spectrum in which this fragment ion is prominent and for which automated sequencing models may fail.

Histidine, an amino acid residue often found in active sites of proteins, has a relatively basic side chain whose gas phase basicity is similar to that of Lys. The gas phase proton affinities of the Lys and His side chains are both lower than that of Arg but higher than those of any other amino acid side chain. Although the proton affinities of the amino acids Lys and His have been determined ([49] and reference cited therein), there is disagreement about their exact values and even the relative order of proton affinities of Lys and His. The values of the proton affinities and gas phase basicities of their side chains in the gas phase have not been established. Morgan and Bursey studied low-energy CID spectra of a set of singly protonated G-G-X (X = G, A, V, L, I, F, Y, W, P, M, E, H, R, K) model peptides and proposed that the site of protonation changes from the peptide backbone to the side chains for His, Arg and Lys [50]. In a statistical analysis, Dongen et al. showed that His has the highest calculated mean charge retention factors among residues in 138 model peptides containing 3–9 residues and activated by high-energy CID conditions [51]. Because tryptic digestion is more or less a standard procedure in proteomics studies employing tandem mass spectrometry, histidine is usually the only basic residue that is not located at the carboxyl terminus in a peptide sequence. It is informative, and useful for future sequencing algorithm development, to determine how the presence of an internal His influences peptide fragmentation. In this study, peptides that have at least one internal arginine or lysine, i.e., formed by an incomplete trypic digest, are also included as comparisons with those that contain histidine.

To investigate how the presence of a basic residue like His influences the preferential fragmentation C-terminal to acidic residues, the Wysocki Laboratory examined a dataset of 505 peptides for comparison with previous conclusions [35,38] drawn from studies of relatively small numbers of peptides. The spectra of two types of doubly protonated tryptic peptides were interrogated: (1) those with at least one acidic residue (Asp or Glu) and at least one basic residue (Arg, Lys, or His) in addition to the Arg or Lys at the C-terminus and (2) those without the internal basic residue and with at least one acidic residue (Asp or Glu). The cleavages of these peptides C-terminal to the acidic residue were compared. The results for peptides with internal His were also compared to results for peptides with internal Arg or Lys. In tryptic peptides containing at least one internal basic residue, one proton is expected to be localized at the C-terminal Arg or Lys while the other may be localized at the internal basic residue. If this holds true, then cleavage at Asp-Xxx (and possibly at Glu-Xxx) should be more dominant for the peptides with internal basic residue(s) than for peptides without the internal basic residue, because...
the Asp-Xxx mechanism becomes favorable when charges are “localized” [38].

2. Experimental

The spectra used were obtained in the Yates Lab from an upgraded Finnigan LCQ™ Classic Ion-Trap mass spectrometer with an ESI nano-spray source. The samples were a collection of tryptic peptides from a digestion of yeast total cell lysate analyzed by MudPIT. MudPIT stands for multidimensional protein identification technology, and involves rapid and large-scale analysis by multidimensional liquid chromatography, tandem mass spectrometry, and database searching by the SEQUEST algorithm [52,53]. CID spectra of the peptides were acquired under 35% “Normalized Collision Energy” [54] with helium as the collision gas.

The acquired spectra were analyzed by SEQUEST on the Yates Lab computer cluster. DTASelect, software developed by the Yates Lab [40], was used to filter the spectra by SEQUEST cross-correlation score, admitting only doubly-charged tryptic peptides and filtering for sequences containing different combinations of basic and (or) acidic residues. The compiled sequence database only included peptides with a minimum cross-correlation (Xcorr) score of 2.5 and the existence of 50% expected ions to ensure that the correct sequences were assigned by SEQUEST. Manual interpretation was also performed on a random selection of the spectra to check the correct assignment of the sequence. Another program developed in the Yates Lab, El CID, generated a database of identified fragment ions (not including doubly-charged fragment ions or immonium ions) corresponding to the identified database sequences.

Java™ programs were written in the Wysocki Lab to perform automatic analysis of the ion intensity patterns. For each ion type (e.g., b or y ions), two different average intensities (or average cleavage abundance) are calculated for expected peaks that are within the instrument scan range: one for those arising from Asp-Xxx cleavage, one for the rest of the peaks, i.e., non-Asp-Xxx cleavage. The cleavage enhancement ratio (CER) at Asp-Xxx for each ion type (b or y) is calculated as the average Asp-Xxx cleavage abundance divided by the average non-Asp-Xxx cleavage abundance (see Eq. (1)). For a set of spectra, the overall cleavage enhancement ratio is the average value obtained after all the spectra in the database are processed. According to this definition, if cleavage at Asp-Xxx is comparable to other cleavage sites, i.e., not enhanced or suppressed, then its cleavage enhancement ratio will be 1.0. If, however, selective or enhanced cleavage occurs at Asp-Xxx, then it is expected that on average, its cleavage enhancement ratio will be above 1.0. In this way, the cleavage enhancement ratio quantifies the degree of enhanced cleavage. In a similar fashion, cleavage at any other sites (e.g., Glu-Xxx) can be calculated. The distributions of b and y ion CER for different datasets are plotted as histograms in which the first pair of bars corresponds to CER of 0, the second pair of bars is for $0 < \text{CER} \leq 1$, the third pair is for $1 < \text{CER} \leq 2$ and so on.

$$\text{CER} = \frac{\sum_{i=1}^{n} \text{DXpeak}_i / n}{\sum_{i=1}^{m} \text{non-DXpeak}_i / m}$$

where CER is the cleavage enhancement ratio for one ion type and a selected cleavage site; DXpeak, is the ion intensity at Asp-Xxx cleavage site in the $i$th instance; non-DXpeak, is the ion intensity at non-Asp-Xxx cleavage site in the $i$th instance; $n$ is the count of Asp-Xxx cleavage of this ion type in a spectrum; and $m$ is the count of non-Asp-Xxx cleavage of this ion type in a spectrum.

3. Results and discussion

The database investigated in the present work contains 462 spectra of doubly protonated tryptic peptides with at least one acidic residue and at least one internal basic residue in addition to the C-terminal Arg or Lys, and 43 spectra without the internal basic residue and with at least one acidic residue (see Tables 1 and 2). Peptides may contain more than one internal basic residue. The length of the peptides ranges from 6 to 25 amino acid residues.
### Table 1
Average CER at Asp-Xxx for peptides in different categories

<table>
<thead>
<tr>
<th>CER for peptides containing</th>
<th>No internal basic residue</th>
<th>No internal basic residue and no Pro</th>
<th>Internal basic residue(s)</th>
<th>Internal basic residue(s) but no Pro</th>
<th>Internal His only</th>
<th>Internal Lys only</th>
<th>Internal Arg only</th>
</tr>
</thead>
<tbody>
<tr>
<td>Asp and Glu</td>
<td>1.0</td>
<td>0.78</td>
<td>0.74</td>
<td>0.79</td>
<td>4.1</td>
<td>3.5</td>
<td>3.8</td>
</tr>
<tr>
<td>Asp, no Glu</td>
<td>0.98</td>
<td>0.62</td>
<td>N.A.</td>
<td>N.A.</td>
<td>3.5</td>
<td>4.5</td>
<td>N.A.</td>
</tr>
</tbody>
</table>

N.A.: not applicable or not performed.
Table 2
Summary of peptide database composition

<table>
<thead>
<tr>
<th>Peptides containing</th>
<th>Count without internal basic residue</th>
<th>Count without internal basic residue and without Pro</th>
<th>Count with internal basic residue(s) but without Pro</th>
<th>Count with internal His only</th>
<th>Count with internal Lys only</th>
<th>Count with internal Arg only</th>
</tr>
</thead>
<tbody>
<tr>
<td>Asp and Glu</td>
<td>20</td>
<td>9</td>
<td>160</td>
<td>66</td>
<td>65</td>
<td>37</td>
</tr>
<tr>
<td>Asp, no Glu</td>
<td>15</td>
<td>N.A.</td>
<td>137</td>
<td>N.A.</td>
<td>73</td>
<td>19</td>
</tr>
<tr>
<td>Glu, no Asp</td>
<td>8</td>
<td>N.A.</td>
<td>165</td>
<td>N.A.</td>
<td>92</td>
<td>30</td>
</tr>
</tbody>
</table>

N.A.: not applicable or not performed.
Fig. 1. The distribution of b and y ion cleavage enhancement ratios at Asp-Xxx for peptides that contain both Asp and Glu, and (a) at least one internal basic residue, (b) no internal basic residue, and (c) at least one internal basic residue but no Pro.
Fig. 2. The distribution of b and y ion cleavage enhancement ratios at Glu-Xxx for peptides that contain both Asp and Glu, and (a) at least one internal basic residue, (b) no internal basic residue, and (c) at least one internal basic residue but no Pro.
The overall average cleavage enhancement ratio (CER) at Asp-Xxx for peptides containing internal basic residue(s) is more than four times higher than that for peptides without an internal basic residue in both b and y ions (4.25 vs. 1.01 for b ions; 3.49 vs. 0.71 for y ions). The distributions of individual CERs at Asp-Xxx are plotted in Fig. 1a and b. It is clear from the graphs that the distribution of the CERs are considerably broadened when one (or more) internal basic residue(s) are present. While about 60% b ions and 70% y ions have CERs less than 1 (no enhancement) when there is no internal basic residue, more than 70% of the b ions and y ions have CERs greater than 1 when one (or more) internal basic residue(s) are present. When a similar analysis is done for cleavage at Glu-Xxx, the trend is similar but not as significant as that for Asp-Xxx (see Fig. 2a and b). Leaving out peptides that contain proline (Pro), because Pro is known to be a residue that may cause enhanced cleavage (at Xxx-Pro) [55–57], does not lead to enhanced cleavage at Asp-Xxx and Glu-Xxx (see Figs. 1c and 2c). This means that Xxx-Pro cleavage is not competing with Asp-Xxx or Glu-Xxx cleavage. Leaving out peptides that contain Glu also does not change the

Fig. 3. The distribution of b and y ion cleavage enhancement ratios at Asp-Xxx for peptides that contain only Asp, not Glu, and (a) at least one internal basic residue, (b) no internal basic residue.
cleavage pattern of Asp-Xxx (see Fig. 3a and b). All of these data are consistent with a previous observation [35,38] that Asp-Xxx cleavage is enhanced when charge is sequestered at basic sites, which implies that one proton is localized at the C-terminal basic residue while the internal Arg, Lys or His residue retains the other. The reason for the broad distribution of CERs among the peptides containing one (or more) internal basic residue(s) is that the degree to which cleavage is enhanced depends on the identity of the residue immediately following Asp (Glu), and the number of basic residues in the peptides and their relative basicity.

To determine if Asp-Xxx cleavage is competing with Glu-Xxx cleavage, an analysis of peptides that contain only Glu, not Asp was also performed (see Fig. 4a and b). The results showed that in this dataset, the average CER for Glu-Xxx cleavage is almost the same for peptides with and without the internal basic residue(s) (1.81 vs. 1.77 for b ions, 1.63 vs. 1.14 for y ions). However, due to the small sample size for peptides without the internal basic residue (only eight peptides), these results need to be further confirmed. A slight shift in the distribution of CERs at Glu-Xxx was found when comparing this dataset to peptides that

---

**Fig. 4.** The distribution of b and y ion cleavage enhancement ratios at Glu-Xxx for peptides that contain only Glu, not Asp, and (a) at least one internal basic residue, (b) no internal basic residue.
Fig. 5. The distribution of b and y ion cleavage enhancement ratios at Asp-Xxx for peptides that contain both Asp and Glu, and (a) His as the internal basic residue, (b) Lys as the internal basic residue, (c) Arg as the internal basic residue.
contain both Asp and Glu. The populations of b and y ions at Glu-Xxx having CER greater than 1 increase when Asp is eliminated from the sequence (from 40 to 60% for b ions and 35 to 45% for y ions) for peptides with internal basic residue(s) (see Figs. 2a and 4a).

One possible reason that the analysis of Glu-Xxx cleavage does not show a pronounced difference for peptides with and without internal basic residues is that the increase of the Glu side chain by one methylene (compared with Asp) slows the reaction for entropic reasons; an eight-membered ring intermediate structure vs. a seven-membered ring is entropically less favorable. It may also be significant that the acidities of the Asp and Glu side chains differ: in solution, the carboxyl groups of Asp and Glu side chains ionize with intrinsic pK\text{a} values of 3.9 and 4.3, respectively [58].

To determine if the identity of the basic residue among Arg, Lys and His is significant in enhancement of amide bond cleavage at Asp-Xxx, the CERs for b and y ions were calculated for each residue and compared to the CERs for peptides without any internal basic residue (see Fig. 5a–c). These graphs show that with the presence of any internal basic residue, the cleavage at Asp-Xxx is enhanced, i.e., a significant portion of the b and y ions from peptides with internal His (>70% for b ions, >60% for y ions), internal Lys (~80% for b ions, >70% for y ions), or internal Arg (~85% for b ions, 100% for y ions) have CERs above 1.0. Peptides that do not have any internal basic residue show the opposite behavior (more than 60% b and y ions have CERs less than 1.0). For peptides with internal basic residue(s), the degree to which cleavage is enhanced varies, however, according to the ability of the basic side chain to “hold” the proton. Overall, it can be concluded that the degree to which cleavage at Asp-Xxx is enhanced increases with the gas phase basicity of the internal residue with Arg > Lys > His.

In the above analyses, plots for “a” ions are not shown because the “a” ions formed at Asp-Xxx are formed by a different mechanism from b and y ions. The anhydride b ions expected from side chain induced cleavage at Asp-Xxx do not readily lose CO to form “a” ions [35].

4. Conclusions

Analysis of the cleavage at Asp-Xxx showed that the cleavage is enhanced in doubly protonated tryptic peptides that contain one (or more) internal basic residue(s). These results are consistent with locating the second proton at a basic side chain while the first proton is sequestered at the C-terminal Arg (or Lys), allowing the acidic H on the Asp side chain to initiate cleavage. Although this general model for cleavage at acidic residues has been suggested before [35,38], earlier studies involved a relatively small number of related peptide sequences while the data reported here are based on several hundred peptides from a complex protein digest. Our analysis shows that the degree of enhancement of cleavage at Asp-Xxx varies with the internal basic residue in the order of Arg > Lys > His.

Because the database of spectra used for this study was produced by use of the SEQUEST algorithm and was filtered aggressively on its cross-correlation score, these spectra are examples of relatively “well behaved” peptides. Other spectra may plausibly fail to be correctly identified by SEQUEST because they are dominated by Asp-Xxx cleavage products. The use of a larger database of peptide spectra in the future, especially increasing the number of peptides without internal basic residues, will allow a more comprehensive study seeking residue-specific patterns. The fragmentation rule will also be incorporated into a sequencing algorithm and tested for its ability to improve identification of proteins.

Acknowledgements

This work was financially supported by a grant from the National Institutes of Health (GM RO151387 to V.H.W.)

References


