Dissociation channel dependence on peptide size observed in electron capture dissociation of tryptic peptides

Guillaume van der Rest[*]*, Renjie Hui, Gilles Frison, Julia Chamot-Rooke*

Laboratoire des Mécanismes Réactionnels, Departement of Chemistry, Ecole Polytechnique and CNRS, 91128 Palaiseau Cedex, France

TITLE RUNNING HEAD: Peptide size effects in ECD.

Address reprint requests to Guillaume van der Rest, Laboratoire des Mécanismes Réactionnels, Ecole Polytechnique, 91128 Palaiaseau Cedex, France, Phone : +33 1 69 33 48 43, Fax : +33 1 69 33 48 03, e-mail gvdr@dcmr.polytechnique.fr.

Abstract

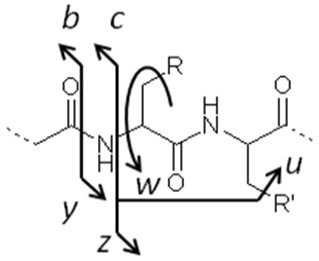
Electron capture dissociation (ECD) of a series of five residue peptides led to the observation that these small peptides did not lead to the formation of the usual *c*/*z* ECD fragments, but to *a*, *b*, *y* and *w* fragments. In order to determine how general this behavior is for small sized peptides, the effect of peptide size on ECD fragments using a complete set of ECD spectra from the SwedECD spectra database was examined. Analysis of the database shows that *b* and *w* fragments are favored for small peptide sizes and that average fragment size shows a linear relationship to parent peptide size for most fragment types. From these data, it appears that most of the *w* fragments are not secondary fragments of the major *z* ions, in sharp contrast with the proposed mechanism leading to these ions. These data also show that *c* fragment distributions depend strongly on the nature of C-terminal residue basic site: arginine leads to loss of short neutral fragments, whereas lysine leads to loss of longer neutral fragments. It also appears that *b* ions might be produced by two different mechanisms depending on the parent peptide size. A model for the fragmentation pathways in competition is proposed. These relationships between average fragment size and parent peptide size could be further exploited also for CID fragment spectra and could be included in fragmentation prediction algorithms.

Introduction

Mass spectrometry based proteomics relies on two types of strategies: direct measurement of the masses of ionized peptides (MS strategies), or fragmentation of these peptides in the mass spectrometer leading to fragmentations which are related to peptide sequence (MS/MS strategies) [1]. In general, available genomic data, completed with transcription, splicing and translation rules for deriving proteins (and peptides) from these data, can provide a set of peptides to which the experimental data can be related. For the MS based strategies, this comparison is straightforward, as experimental masses are directly compared to database masses. Most of the difficulty is related to efficient construction of the set of database mass values to which the experimental masses should be compared, as to statistical validation of the resulting comparison. For MS/MS based strategies, an additional bottleneck is that there are currently few predictive models for peptide fragmentation. Therefore, comparison with database values is generally based on very simple assumptions on the general rules followed by fragmentation of peptides (such as the type of fragments which can be expected for a given ion activation method), and does not take into account fragment intensities. In recent years, thanks to the availability of large collections of MS/MS spectra of peptides, efforts have been devoted to build empirical predictive models both for low energy collision induced dissociation (CID) [2, 3] and electron based activation methods [4]. Although empirical, these predictive models rely on the selection of key parameters (such as hydrogen or proton affinities of amide carbonyls and of the backbone atoms, complex dissociation Arrhenius parameters, …) delineating fragmentation pathways.

In this context, when we observed, a couple of years ago a peptide (AGWLK) which did not follow the expected fragmentation behavior upon electron capture dissociation [5], we sought further investigation in the generality of this observation, as in the parameters leading to this fragmentation behavior. Electron capture dissociation has been introduced in 1998 by Zubarev et al [6] as a method to fragment a multiply and positively charged peptide by direct capture of an electron. The electron – positive charge recombination provides energy to the system but also leads to the formation of an odd-electron species from the initially even-electron species. The presence of radical based fragmentation

pathways allows breaking the N-Cα peptide bond, leading to *c*/*z* fragment types, which are usually not observed in low energy CID conditions [7]. Although usually dominant in ECD fragmentation of peptides, *c* and *z* fragments are far from being the only type of fragments present in ECD spectra: the initial work had reported the presence of *a* and *y* type ions [8], and further studies have shown many other fragmentation pathways stemming from a direct break of the peptide backbone such as *b* ions [9, 10, 11, 12], from side chain [13, 14, 15] or hydrogen losses [16, 17], or from a combination of both leading to *w* or *u* ions [18, 19, 20]. The picture would not be complete without mentioning that the radical process leads to intramolecular hydrogen transfers [21, 22, 23, 24], as well as possible hydrogen losses [23], leading to a shift in masses of *c* and *z* ions by up to two hydrogen masses.



Scheme 1. Depiction of the major peptide fragment ion nomenclature used in this work.

A number of aspects governing the ECD mechanism has to be taken into account to delineate at which point the difference between fragmentation pathways arises: the electron capture process in itself is still a matter of debate as competing pathways have been proposed [25] (direct capture at the charge bearing ammonium [8] or capture in an excited carbonyl π* orbital stabilized by neighboring positively charged groups [26, 27, 28, 29]. Furthermore, the capture site will be strongly influenced by the nature and electron affinity of the charge bearing groups (ammonium vs guanidinium vs fixed charge group [30, 31, 32]), by the hydrogen bonding network involved in charge solvation as by the 3-dimensional structure of the peptide. Once the electron has been captured (in a ground or excited state), the system has to evolve, either through a direct hydrogen atom transfer [8], a concerted through bond electron transfer – proton transfer or through direct cleavage of the N-Cα bond prior to a proton migration on the anionic amidate [26] superbase. None of these schemes currently include the steps at which competitive

fragmentations leading to CID-like fragments could be observed. Considering that radical *b* or *y* ions are rarely observed, the simplest pathway would be loss of $H^{\bullet}$ following electron capture on an ammonium group, after which part of the energy deposited in the molecule upon electron capture is redistributed among the molecule vibrational modes leading to standard CID-like fragments. On the other hand, one cannot rule out that *b* and *y* fragments could arise from secondary fragmentation of respectively *c* and *z* fragments as such fragmentations (as a minor pathway for $z^{\bullet}$ fragments) were observed in collisional activation of ETD fragments [33, 34].

In these conditions, the observation of *a*, *b*, *y* and *w* fragments in the fragmentation of the AGWLK pentapeptide does not seem exceptional in itself. The absence of any *c* or *z* fragments is more intriguing: literature has constantly described *w* ions as originating from *z* ions [4, 18] and our spectra did not show any detectable amount of the associated *z* ion. Furthermore, no systematic study on the parameters controlling the evolution between "standard" (*c*/*z*) and "atypical" (other than *c*/*z*) fragmentation behaviors is present in the literature. For such a study, having at our disposal a peptide which exclusively fragments along "atypical" pathways is extremely useful, since it allows synthesis of variants of this peptide to test various hypothesis on the parameters disfavoring the standard fragmentation pathways. Among the parameters that could influence differing pathways, the size of the peptide could be a discriminating factor, since increasing size means additional internal degrees of freedom on which the energy deposited following electron capture can be redistributed, but also an increasing number of internal hydrogen bonds and an increasingly structured peptide. The nature of the charge carrying group could also be a major parameter as it will affect the electron capture step itself.

This article will focus on the relation between size and type of fragments observed. Using both a series of peptides derived from the AGWLK pentapeptide with varying lengths and compositions and the SwedECD database of spectra compiled by Roman Zubarev's group [18], it will be shown that (i) fragmentation of the AGWLK pentapeptide is not at all "atypical" for tryptic peptides of this size, (ii) both *b/y* and *w* fragmentation pathways are influenced by peptide size, but do not follow a similar trend,

(iii) fragment size distributions differ markedly between *c*/*z* and *w* fragments, suggesting that they do not originate from the same precursors.

Materials and Methods

ECD of synthetic peptides

All peptides used in this work are synthetic peptides purchased from GeneCust Europe (Dudelange, Luxemburg) with purities above 90% and used without further purification. Peptides were dissolved in water, and diluted in water / acetonitrile / formic acid 50 : 50 : 2 (v/v/v) to a working concentration of 10 µM. ECD experiments were run on a Bruker Apex III 7 T FT-ICR mass spectrometer upgraded with a indirectly heated hollow cathode [35]. The appropriate charge state of the parent peptide ion produced in the ESI source of the instrument was isolated by rf ejection of all unwanted ions, and fragmented by a pulse of electrons (cathode bias : 0.7 V, pulse duration 100 ms). Assignment of fragments in the mass spectrum was done using the Bruker BioTools software suite, followed by manual inspection of the data.

SwedECD database analysis

This work made use of the complete ECD spectra collection contained in the SwedECD [14] database downloaded from the project web page (http://www.bmms.uu.se/CAD/indexECD.html). Details on the spectra contained in the database are given in reference [14]. In summary, this 11 491 peptide spectra database is built from the proteomic analyses of trypsin digests of whole cell lysates. Sequence assigment of the peptides was based on database searches, and the authors claim a reliability on assigned sequences above 95%. For each assigned peptide, the database contains a list of peaks (*m/z* and intensities) measured for the fragmentation of a given peptide charge state. Extraction of the relevant data from the database was done in three steps:

(i) Based on the assumed peptide sequence, peaks were assigned to a type of fragment (a, (b-H)$^{\bullet}$, b, (b+H)$^{\bullet}$, (c-H)$^{\bullet}$, c, w, (w+H)$^{\bullet}$, x, (y-H)$^{\bullet}$, y, (y+H)$^{\bullet}$, $z^{\bullet}$, (z+H)) and fragment position following the standard nomenclature using a homemade peak assignment software. Following the original publication [14], the tolerance used on fragment *m/z* values was set to 50 ppm and cysteins were considered in

carbamidomethylcystein form. In case of ambiguity for the assignment, the ambiguity is noted, and the peak is not used for further processing. Differentiation between peaks differing by 1 Da (such as $(c\text{-H})^{\bullet}$/c or $z^{\bullet}$/(z+H)) was done based on intensity comparisons between both peaks: starting from the lowest mass peak that can be assigned to $(c\text{-H})^{\bullet}$ or $z^{\bullet}$, the intensity of the second isotopic peaks is computed based on carbon content of averagine composition [36] for the same mass. If the intensity of the second isotopic peak is larger by over 5% to this calculated peak, this peak pattern is assumed to belong to a doublet, and the difference between the experimental intensity and the averagine calculated intensity is assigned to the higher mass fragment (such as c or (z+H)). With the exception of the preceding calculation, the intensity assigned to a given fragment is that of the monoisotopic peak. This induces a small bias in fragment intensities (as low mass fragments will have higher intensities than high mass fragments) which was left uncorrected, since it did not affect the interpretation of the results.

(ii) With a homemade script, the desired aggregated information (number of fragments per type, intensity weighted average fragment size, amino-acid before/after cleavage, fragment size distributions, …) is computed for each peptide, and output to a csv file, along with peptide size and sequence. When classes of fragments are considered, c and z-types are considered to be comprised of both $(c\text{-H})^{\bullet}$ and c ions (or respectively $z^{\bullet}$ and $(z\text{+H})^{\bullet}$), with intensities added together, and presence meaning at least one of the two sub-type ion present. The text mentions whether complete types or sub-types are to be considered.

(iii) The data is imported in a spreadsheet, and further aggregated by counting the number of peptides which match given properties. The histograms are derived from these results. When intensities are considered, they have been normalized for each peptide by taking the total assigned fragment intensity as 100%.

CID database analysis

For comparison of the ECD data to collision induced fragmentation data, two other databases were used. Doubly charged parent ion data was taken from the CID counterpart of the SwedECD database called SwedCAD [37] (http://www.bmms.uu.se/CAD/), which contains the CID fragments of 15 897

peptides. Since these are also high resolution spectra, the same tolerance and procedure as for the SwedECD analysis was used. Singly charged parent ion data was extracted from the NIST Libraries of Peptide Tandem Mass Spectra, using the *Mus musculus* ion trap data library [38]. This species was chosen, because selection of singly charged parent ions bearing no modifications led to 12 674 peptides, which is close to the number of peptides present in the SwedECD and SwedCAD databases. For the NIST database analysis, the spectra where re-annotated using our in house program, this time with a tolerance of 0.2 Da to take into account the lesser precision of ion trap generated data.

Results and Discussion

The initial question raised in this work stemmed from our observation on the ECD fragmentation of a doubly charged pentapeptide AGWLK [5], leading exclusively to *a*, *b*, *y* and *w* fragments with the exclusion of any *c* or *z* fragment. A first comment can be raised: since this is a quite short peptide, this "atypical" behavior might be simply related to the short size of the peptide. In this previous work [5] we observed that other small peptides (KRQHPG, KRDVY and TPRK) lead to the appearance of *c* fragments as major fragments. Similarly, a recent article has observed the occurrence of a minor fraction of *c* and *z* fragments for a series of doubly charged histidine containing peptides [29]. Even the dipeptide [AK, 2 H]$^{2+}$ was shown to lead to z_1 fragment upon electron transfer from sodium in a beam experiment [39], although loss of hydrogen was the major channel observed. The presence of CID-like fragments (*a*, *b*, *y* fragments) could be explained by a process through which energy is deposited in sufficient amount by the ECD process to open a competitive pathway. Cooper [9] suggested that *b* ions originate from further fragmentation of the reduced species after loss of a hydrogen atom [M, *n-1* H]$^{n-1+}$. Provided that sufficient internal energy is available, the ion would be left in an activated state allowing it to proceed to a standard "internal energy" limited fragmentation. This is in agreement with the observation of a charge reduced / hydrogen loss ion after electron capture. Since internal energy deposition by the electron capture process is, on first order, only related to the recombination energy of the charged site, one can assume that the energy deposited is similar independently of the size of the peptides for a given charge

state. Thus, longer peptides tend to have less internal energy deposited per degree of freedom than shorter ones, which leads to reduce the unimolecular fragmentation kinetics, moving this pathway outside of the observation window for longer peptides. This explanation could be extended to the appearance of *w* ions: provided sufficient energy is available, it could arise from a first N-C_α bond cleavage followed by radical side-chain loss.

Nevertheless, one should note that for AGWLK an intense w_2 ion is observed although z_2 is totally absent. Moreover, if a competitive N-C_α bond cleavage pathway was open, one would expect such fragmentations at other positions. Since a single peptide cannot be considered a general rule, a series of peptide variants, based on the general AGXLK (X=A, D, E, S, W) motif were synthesized, with the objective to see if changing amino-acid composition would open either a competitive *c/z* or a *w* fragmentation at another residue.

Influence of central amino-acid

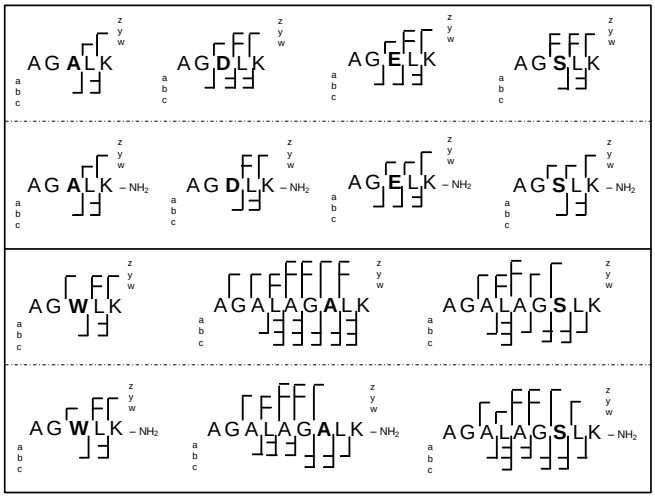


Figure 1. Fragments observed upon ECD dissociation of doubly charged peptides.

The results of the ECD experiments are presented in Figure 1. For all doubly charged pentapeptides, either in the free acid or amidated C-terminal state, the result is similar: only *a*, *b*, *y*, and *w* ions are observed. For the latter, it is worth noticing that the w_2 ion (cleavage in leucine side-chain) is always present, and the w_3 ion is present for X=D, E and S. Moreover, the presence of a *w* fragment for serine is

surprising as this fragmentation is quite energetically unfavorable, since it leads to the loss of a $OH^{\bullet}$ radical. This contrasts with the observation, by Savitski *et al*. [18], that serine rarely leads to *w* fragmentation. Also in contrast with this work, aspartic acid (D) leads to a "standard" *w* ion from a formal loss of $^{\bullet}CHO_2$ and not from a loss of CO_2. This "atypical" fragmentation behavior, leading to *a*, *b*, *y,* and *w* ions with intense *w* ions appears general for the doubly charged AGXLK peptides. It should be noted that although the fragmentation shares some element with fragmentation in hydrogen deficient peptide radicals (such as formation of *a* fragments and specific losses of side-chains) [40], the direct formation of *w* fragments is rarely observed for these radical induced fragmentation.

Influence of peptide size

As discussed above, size in itself does not prevent the observation of *c* or *z* fragments for small peptides, but might alter the fragmentation pathways. This effect was investigated adding four residues to the initial AGXLK motif. It was chosen to add the AGAL motif, since this motif did not contain any basic amino acid, could provide strong *w* ions by the presence of a leucine, did not change the position of the basic amino acid and could be related to the fragmentation of the similar AGALK peptide. As can be seen in Figure 1, the addition of 4 residues leads to the formation of *c*/*z* fragments in addition to *a*, *b*, *y* and *w* fragments which are still present. In terms of intensities, the *c/z* fragments are relatively low intensity fragments compared to *a/b/y* fragments. *w* fragments follow another trend: only the w_6 fragment is observed, and not the smaller w_2 and w_3 fragments observed for the pentapeptides. Furthermore, intensity for the *w* ions is on the same order as the *c/z* fragments, in sharp contrast with the pentapeptides series in which these fragments constituted the major fragmentation channel.

Fragmentation channels depend on peptide size

In order to separate the effect of size from other parameters influencing ECD peptide fragmentation, one possibility could have been to synthesize a vast range of peptide sequences and perform ECD experiments on all of these peptides. An alternative option is the use of existing spectra database. Such a

database has been provided by the group of Roman Zubarev and coined the SwedECD database [14, 41]. At this point, one should compare how relevant the use of this database is, compared with the problem at hand. This database is comprised of the ECD spectra of 11 491 doubly charged tryptic peptides. Therefore, all these peptides have the same charge state as our experimental peptides. Furthermore, their tryptic origin ensures that a vast majority (99.1 %) have a basic residue (K or R) at the C-terminal position, which also resembles our experimental peptides.

A first major difference is that peptide sizes in this database range from 6 to 24 residues. A second strong bias in the use of this database is that the peptides present in the database have been identified through the use of protein database identification, and not from a combinatorial peptide library: therefore only peptides that lead to good quality fragmentation spectra, i.e. follow the assumed fragmentation rules, have been retained in the database. The presence of a peptide in the database requires that it leads to at least three *c*, *w* or *z* fragments [14]. Therefore all the peptides following a purely "atypical" pathway, such as the pentapeptides discussed above are in essence not present in the database. On the other hand, fragment intensities have not been used for inclusion or exclusion of peptides from the database, therefore peptides leading to "atypical" fragmentations in parallel with expected fragmentations (such as the nonapeptide series) could be present in the database.

Table 1. Presence of fragment types and average fragment type intensity among the pathways considered for this study.

Fragment type	Complete database		Restricted to peptides composed of			
			7 residues		13 residues	
	Occurrence ratio (%)	Average intensity (%)[a]	Occurrence ratio (%)	Average intensity (%)[a]	Occurrence ratio (%)	Average intensity (%)[a]
a	7.3	1.1	12.3	1.6	4.6	0.6
b-H$^{\bullet}$	9.2	0.7	12.6	0.6	5.4	0.4
b	24.2	2.7	55.3	6.8	9.4	0.8
b+H$^{\bullet}$	11.9	0.5	19.9	0.6	7.1	0.4
c-H$^{\bullet}$	46.0	20.4	13.7	0.7	65.6	7.5
c	82.6	4.6	68.1	21.8	84.6	17.2
w	44.8	8.5	88.5	27.2	12.3	1.4
w+H$^{\bullet}$	27.8	2.4	33.0	2.7	21.9	1.7
x	8.8	0.7	10.8	0.8	7.5	1.3
y-H$^{\bullet}$	17.8	1.3	10.8	0.9	19.0	1.4
y	45.7	3.7	45.0	4.6	48.1	3.7
y+H$^{\bullet}$	17.5	0.6	6.6	0.2	26.4	0.9
z$^{\bullet}$	89.1	30.9	97.4	26.9	97.7	32.5
z+H	87.8	21.7	50.2	4.6	98.3	30.0

[a] Average ratio of the total intensity for a given fragmentation pathway over the total intensity of assigned fragments.

Despite this strong bias, the SwedECD database was examined for the occurrence of various fragment types (and sub-types) as a function of peptide length. First, the occurrence of each fragment type, as well as the average relative intensity of each fragment type in the total fragmentation intensity, was compiled. As can be seen in Table 1, some fragment types (*b*+H$^{\bullet}$, *b*-H$^{\bullet}$, *x*, *y*+H$^{\bullet}$) represent on average low abundance fragmentation pathways (below 1% in average fragmentation intensities). Since they are also

not present in a large fraction of the peptide fragmentation patterns, they will not be examined further. Average fragment intensities change dramatically for varying peptide lengths: for instance, *b* fragments represent on average 6.8 % of total fragment intensity for 7 residue peptides, and are nearly absent (0.8 %) for 13 residue peptides. Occurrence ratios also change greatly between 7 and 13 residues.

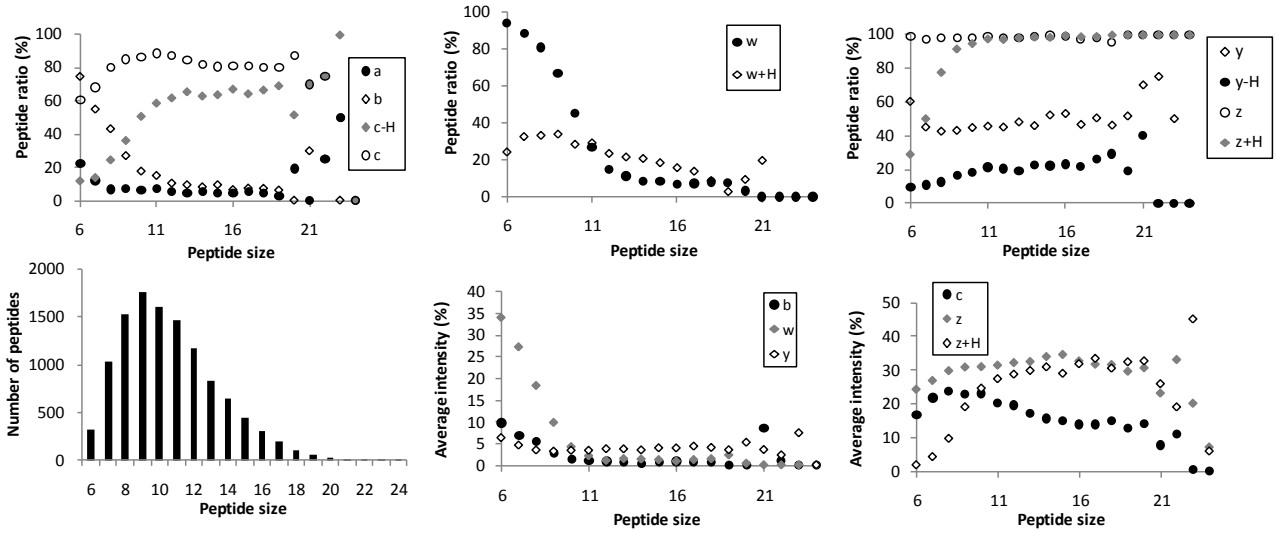


Figure 2. Fragment type occurrences and intensities as a function of peptide size. (top row) Percentage of peptides of a given size leading to at least one fragment of a given type. (bottom row, left panel) Distribution of peptides as a function of size in the database. (bottom row, middle and right panels) Average fragment type intensity (relative to total fragment intensities) as a function of peptide size.

Once plotted on a graph (Figure 2, top row), the trends appear quite clearly: *b* and *w* ions are present mostly for the smallest peptides. The points seem to leave the trend at higher peptide size, but one should note (Figure 2, bottom left) that for these peptide sizes, only few peptides are present in the database, and therefore points for peptide sizes above 20 residues might not be statistically significant. Closer examination of the trend for *b* and *w* ions reveals that they do not seem to follow the same evolution: less than half of the peptides present a *w* fragment for peptides larger than 10 residues, whereas for *b* ions, this level is reached at 8 residues. Such a trend had already been noticed for *b* ions [9] and forms one argument in favor of a mechanism proceeding through a vibrationnally excited MH^+ precursor [9, 10]. Our analysis of the complete database reinforces this argument, as one can clearly

observe that the presence of *b* fragments is strongly related to peptide size for tryptic peptides. The observation of parent peptide size dependence for *w* ions has not been reported elsewhere to our knowledge. If one examines the contribution of the *b* and *w* pathways in the total fragmentation intensities (Figure 2, lower row), one notices that a strong decrease in intensities is also observed. The *b* fragmentation pathway remains minor (< 10% of total intensity) in all instances. On the other hand, the *w* fragmentation pathway represents 34% of fragment intensities for six residue peptides, and is therefore a major contributor to the fragmentation pattern of small peptides.

For the other fragmentation pathways, one can first notice that the w+H$^{\bullet}$ pathway does not follow the same trend as the *w* pathway. As shown by Savitski et al. [18], these w+H$^{\bullet}$ ions mostly originate from CO_2 losses from aspartic acid (D) residues. These ions can only be present if a D residue is present in the peptide sequence. Such a presence is less probable for shorter peptides and increases with peptide size. Such a correction (data not shown) accounts for the initially low proportion of these w+H$^{\bullet}$ fragments. It does not account for the crossing between the w+H$^{\bullet}$ and *w* fragment ratios for peptides larger than 11 residues. This suggests that the pathway leading to w+H$^{\bullet}$ fragments is still active when the pathway leading to *w* fragments has been mostly closed.

It is worth noticing that *y* ions do not follow a trend similar to that of *b* ions, although the general mechanism (either through a vibrationnaly excited intermediate [9, 10] or through a direct $b^{\bullet}/y$ cleavage [12, 42]) accounts for these ions to be produced concomitantly with *b* ions, depending on the entity keeping the charge. There seems to be another process in effect for the fragmentation of larger peptides, as one can see an initial decrease in the percentage of peptides leading to *y* fragments for small peptides as well as an initial decrease in fragment intensity, followed for larger peptides by a stationary level of occurrence and intensities. A possible origin for the *y* ions for large peptides could be the secondary fragmentation of $z^{\bullet}$ ions produced initially: MS^3 experiments on ETD produced $z^{\bullet}$ ions [33, 34] have shown that such a pathway is present for collisionnally excited $z^{\bullet}$ ions. The evolution of the (y-H)$^{\bullet}$ pathway reinforces this hypothesis: at small peptide sizes there are nearly none of these fragments and they increase as peptide size increases. A possible origin for these unusual fragments could be

fragmentation of $z^{\bullet}$ fragments after they have undergone an intramolecular hydrogen atom migration, as was shown by P. B. O'Connor et al [43]. Other evidence in favor of this secondary origin for *y* ions will be presented further.

The data also shows an increase in intensity of hydrogen transfer between *c* and *z* fragments for larger peptides. From the generally accepted model [23, 44], these hydrogen transfers between the initially formed *c* and $z^{\bullet}$ ions occur in an intermediate long-lived hydrogen bound non-covalent complex. This is only another representation, on the same set of data, of results presented and discussed in reference [23] and will not be discussed any further.

Average fragment size depends on peptide size

The data presented up to this point only considered the fragmentation pathways in terms of presence/absence of a given pathway for a peptide, or in terms of branching ratios, studied as a function of peptide size. One result that has been set aside is that a given fragmentation pathway also leads to fragments that have a given size. The difficulty is to make use of this fragment size information on a set of thousands of spectra. A possible measure to keep a global picture of the fragmentation is to consider a given type of fragments as a distribution of fragments over the peptide sequence, and to use the intensity of each fragment as a weight in this distribution. From this hypothesis, two parameters could be extracted for a given peptide and fragment type: the center of the distribution (weighted average of the fragment size with respect to fragment intensity, WA fragment size) and its width. Since the width depends on the type of distribution (random, Gaussian, multimodal), this parameter was not used in this work. Plots of the distribution of WA fragment sizes as a function of peptide size for *c*, $z^{\bullet}$ and *w* fragment sub-types are represented in Figure 3 (see Figure S1 for the complete set of plots for all fragment sub-types). In most instances (*a*, *c*, *w*, *y*, $z^{\bullet}$) the maximum of the distribution follows a roughly linear evolution as a function of peptide size. For the other ion types, the evolution is more complex, with most of the time at least two lines that seem to emerge, parallel in the case of *c*-$H^{\bullet}$ fragments, and

crossing at some point for others. One can extract the apparent slope and intercept for the various fragment type leading to a summary of the WA fragment size dependence to peptide size (Table 2).

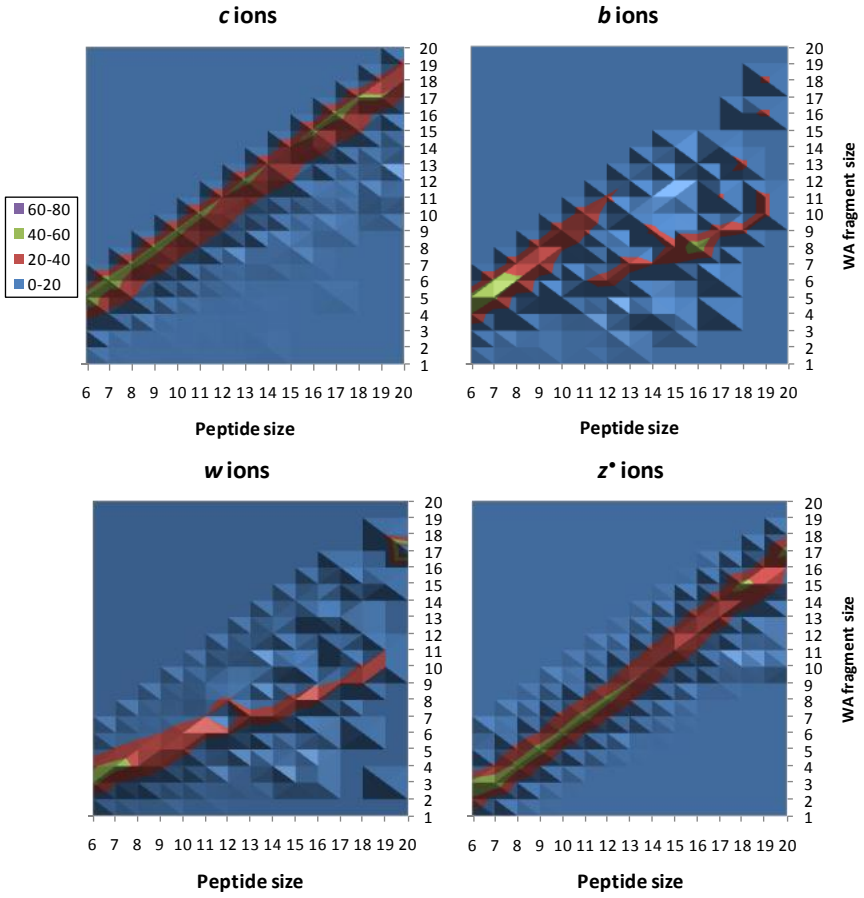


Figure 3. Distribution of the weighted average (WA) fragment size as a function of peptide size for the *c*, *w* and $z^{\bullet}$ fragmentation pathways. The bin size for WA fragment size is set to 1 (see text) and the color code represents the percentage of peptides in a WA fragment size bin within the total number of peptides of the same size leading to at least one fragment of the type considered. (Each column is normalized to a total of 100%.)

Table 2. Linear trends observed for the evolution of weighted average (WA) fragment size as a function of peptide size.

	SwedECD[a]		SwedCAD[b]		NIST Mouse (1+)[c]	
Fragment sub-type	Slope	Intercept	Slope	Intercept	Slope	Intercept
a	0.4	2.1	-[d]	-[d]	0.7	-0.1
b	1.0 0.5	-1.7 -0.2	0.7	-1.3	0.9	-1.1
c-H$^{\bullet}$	1.0 1.0	-1.0 -2.8				
c	1.0	-1.0				
w	0.5	0.7				
w+H$^{\bullet}$	1.1 0.5	-4.4 1.2				
y-H$^{\bullet}$	1.0 0.6	-1.0 -0.9				
y	0.9	-2.4	0.5	1	0.7	-1.1
z$^{\bullet}$	1.0	-4.2				
z+H	1.2[e]	-8.6[e]				

[a]ECD fragmentation data from the SwedECD database. All parent ions are 2+ charged. [b]CID fragmentation data from the SwedCAD database. All parent ions are 2+ charged. [c]CID fragmentation from the NIST Peptide Tandem Mass Spectral Library Information, Mouse library, restricted to 1+ charged, non-modified parent peptides (12 674 peptides). [d]No specific linear trend appears. [e]From the linear part of the graph (peptide sizes ranging from 9 to 20).

The interpretation of this linear behavior dependence is straightforward for three values for the slope. (i) A slope of 0 is as a fragmentation leading to fragments of constant size, independent of the parent peptide size. (ii) A slope of 1 means that the fragment lost has a constant size, or put otherwise, that the fragmentation takes place at a fixed number of residues from the side opposite of the fragment bearing the charge. The intercept is in this case the opposite of the offset from the end. (iii) A slope of ½ could be interpreted either as a purely random fragmentation, or as a fragmentation taking place preferentially at the center of the peptide.

As can be seen in Table 2, the situation is not always as clear cut as the three cases delineated above. But if one looks for instance at the $z^{\bullet}$ ions, an interpretation of the slope of 1 and an intercept of -4.2 is that most of the fragment intensity is located 4.2 residues from the N-terminal end of the parent peptide. This observation is in agreement with previous work on the same database [41], which established that the z_{n-5} fragments (where n is the peptide size) are the most likely to have the maximal intensity of the $z^{\bullet}$ fragment series. In reference [41], this preference in fragmentation position was hypothesized to depend on the parent peptide structure, and mostly on the folding of the N-terminal end of the parent peptide. A similar effect is observed for c ions: although the width of the distribution appears larger than for $z^{\bullet}$ fragments, the pathway towards c ions seems to favor loss of only one residue on the C-terminal end of the peptide. Inspection of the c-$H^{\bullet}$ fragment results (Table 2, Figure 4) led to the identification of C-terminal residue influence on the fragment size. Two parallel lines (with a slope of 1) are observed, with intercepts of respectively -1.0 and -2.8. Since most of the peptides present in the database originate from a tryptic digestion, they mostly bear either an arginine or a lysine residue at their C-terminal position. If one separates the parent peptide population in two sub-population (Figure 4), one with a C-terminal arginine, the other with a C-terminal lysine, the parallel lines for c-$H^{\bullet}$ disappear and form a single line for each type, with an offset of 1 residue for R terminated peptides, and 2.8 residues for K terminated ones. This strong correlation of C-terminal residue with fragment size, which is apparent in the size dependence graph for c-$H^{\bullet}$ fragment, can also be evidenced for c fragments. The same separation on c

ions (Figure 4, bottom) leads to a separation between both subpopulations having a C-terminal lysine or arginine which was not apparent at first since for *c* fragments, the intercepts differ only by one residue. It therefore appears that fragmentation on the C-terminal side is directed and dependent on the nature of the basic residue present. This result is also in good agreement with a structural evidence for the direction of cleavage: the lysine and arginine side-chains vary strongly in length and rigidity. The former is rather elongated and can fold in numerous conformations, whereas the latter is short and the guanidinium group is quite rigid. One could therefore envision that the local folding of the peptidic chain at the C-terminal extremity might be strongly influenced by the proton bearing side chain, and therefore lead to such a clear cut difference in the fragmentation between both types of terminal residues. On the other hand, the chemical properties of lysine and arginine, both as proton bearers (proton affinity), electron acceptor (electron affinity) and hydrogen donors after a putative neutralization vary greatly [30] and such energetic issues could also be partly responsible for the results observed. In addition, the number of hydrogen bonds in which the guanidinium group and the ammonium group could be involved is also different.

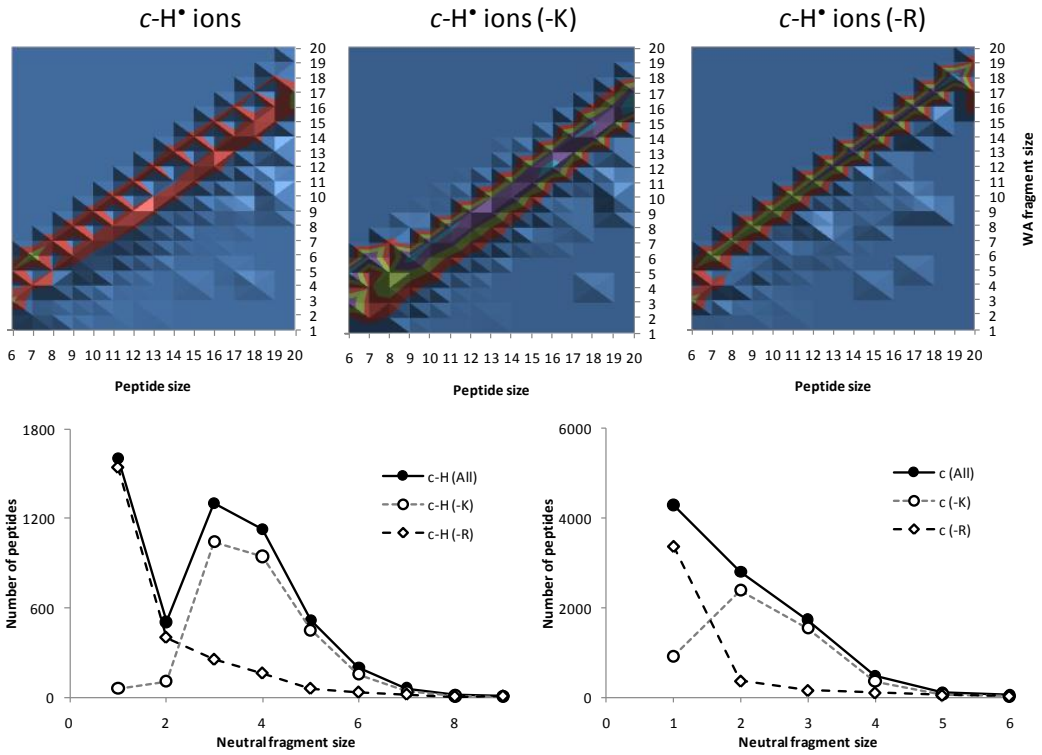


Figure 4. (top) Distribution of the weighted average (WA) fragment size as a function of peptide size for the *c*-H• fragmentation pathway. Middle and right show the distributions restricted to peptides with a C-terminal lysine (middle) and a C-terminal arginine (right). (bottom) Number of peptides as a function of WA neutral fragment sizes (bin size of 1) for the *c*-H• (left) and *c* (right) pathways. Filled circles: all parent peptides. Empty circles and diamonds: parent peptides with a C-terminal lysine and arginine respectively.

w ions originate from an independent pathway

When looking at the w fragmentation pathway, the most striking element in Figure 3 and Table 2, is that the slope of the distribution maximum for w ions is completely different from that of *c* and *z* fragments. It has been generally postulated that w ions are formed by secondary fragmentation of $z^{\bullet}$ fragments [4, 18]. In such a case, on the first order, one would expect that the size distribution of w ions should follow that of *z* fragments, this is obviously not the case. In reference [18], the authors observe

that *w* ions tend to be larger than *z* ions (see Figure 6 of reference [18]) which leads them to postulate that large $z^{\bullet}$ fragments are less likely than smaller ones to interact for long durations with the other moiety, reducing the chance of a hydrogen atom migration which would quench the formation of *w* fragments. This work points out that the size dependence line for *w* fragments (slope 0.5) crosses the *z* fragment one (slope 1) for peptide sizes of about 10 residues. Therefore if one does not separate the peptides according to their sizes and if one considers, as shown above, that *w* fragments are more numerous for small peptides, this also leads to the conclusion that on average *w* fragments are larger than *z* fragments. But this is biased, as our analysis proves that for large peptides *w* fragments are smaller than the *z* fragments, which is not in agreement with the hypothesis proposed by reference [18]. The existence of such a crossing between the size dependence curves for the *w* and *z* fragments also rules out most other systematic biases that could lead to have *w* fragments originate from major *z* fragments: for instance, if only small $z^{\bullet}$ fragments could have sufficient internal energy to lead to *w* fragments, *w* fragments should on average be smaller than the *z* fragments, which is not the case. These elements lead us to postulate that w ions are produced from a distinct process from the one leading to the most abundant *z* ions.

The size dependence of this distinct process is of interest: as it seems to favor the center of peptides and as most of these peptides being tryptic peptides carry a positive charge on each terminal residue, one could think of the alternative "Utah-Washington" ECD or ETD fragmentation process [25, 26, 28]. In this process, the incoming electron is not initially located on the positively charged ammonium or guanidinium groups, but towards carbonyl groups in the intermediate region of the peptide. These mechanisms would favor the locus in the sequence where the negative charge is best stabilized by both positive charges, which is likely to be close to the middle of the peptide. As shown in Figure 5, the distribution of the *w* fragments is rather narrowly centered around the peptide middle, and does not correspond to a purely random distribution. This central distribution appears not to depend on peptide size, and it does not either reflect the statistical distribution of residues for which formation of *w* ions is

favorable (See Figure S3). This element is a good indication that the process at hands for the formation of these ions is truly related to a central position.

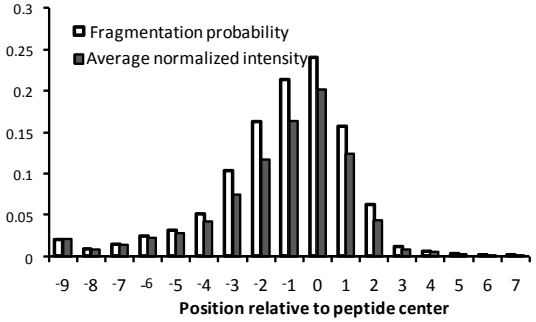


Figure 5. Probability of *w* ion formation and average *w* ion normalized intensity at positions expressed relatively to the peptide center. The positions are ordered N-terminal to C-terminal, taking position 0 as cleavage of the central peptide bond for even sized peptides, and positions 0 and -1 as cleavage of the two peptide bonds surrounding the central residue for odd sized peptides. The probability of *w* ion formation is the ratio of the number of times the given bond was cleaved to the total number of peptides leading to any *w* fragment and having a bond at this position. For calculation of the average intensity at each position, the *w* fragment intensities at each position were normalized to the highest *w* fragment intensity for a given peptide and then averaged for each position.

A distinct process does not mean that these *w* ions are by no means secondary fragments of $z^{\bullet}$ ions: if a minor pathway exists, leading to *z* ions of lesser abundance than those produced by a charge directed mechanism, the size analysis above would have missed these. One should nonetheless point that out of the 9301 *w* fragments, a strong minority (2289, 24.6%) does not present any corresponding *z* fragment in the ECD spectra. Furthermore, a double resonance experiment was conducted on the AGSLK doubly charged peptide: after isolation of the $[MH_2]^{2+}$ parent ion, ECD was performed while at the same time an *rf* pulse at the resonance frequency of a putative parent ion was generated. Varying the excitation power allows to probe varying intermediate lifetimes. Excitation at the highest $V_{p\text{-}p}$ (~ 600 V) for *m/z* 244.18 ($z_2^{\bullet}$) and 373.22 ($z_3^{\bullet}$) did not lead to any significant change in either w_2 or w_3 fragment abundances. This result indicates that either short-lived (< 20 μs) $z^{\bullet}$ ions are precursors for the w

fragments or that a distinct process is involved. In our opinion, this warrants search for an alternative mechanism which would lead to the simultaneous cleavage of two bonds leading to the direct formation of *w* ions from the parent ion. One such proposal could be that electron capture at a carbonyl initiates a side-chain radical loss, followed by N-Cα bond cleavage.

The case of b and y ions

The situation for b ions is also of interest. As noticed above, these ions are mostly present for small peptides sizes and both their occurrences and abundances decrease sharply as peptide size increases. The fragment size analysis reveals two distinct regions: fragments of smaller peptides follow a 1.0 slope while those of larger peptides follow a 0.5 slope. The change between both trends occurs for peptide 11-12 residues long, which is also the size at which b fragment occurrences reach a baseline level (Figure 2). One should point out that the number of large peptides leading to b fragments is low enough to consider that the observed 0.5 trend might not be statistically significant. Nevertheless, these results seem to indicate the existence of two separate processes for different peptide lengths. A question arises as to what process could the observed trends be related. Cooper [9] has suggested that b ions originate from fragmentation of reduced species having lost a $H^{\bullet}$ radical (MH^+ ions) but having kept sufficient internal energy to lead to standard "mobile proton" [45, 46, 47] type b/y fragmentation. This interpretation is in agreement with the size-dependence (sharp decrease of the b ions as peptide size increases) for the initial $H^{\bullet}$ loss step: as peptide size increases, improved internal solvation of the charge bearing group could reduce the probability of a $H^{\bullet}$ loss. The same trend is expected if one considers internal energy, the decrease in internal energy per degree of freedom as peptide size increases would lead to a decrease in $H^{\bullet}$ loss. Nevertheless, in such a case, one would expect that the MH^+ ions lead to both b and y fragments, but the size dependent behavior is observed only for the b ion series. In order to clarify this point, a similar fragment size dependence analysis has been carried on other fragment spectra database, for 2+ and 1+ parent ions (Figure S2, Table 2). Neither follows the same trends (slope and intercept) as the b ions produced from ECD activation, but they follow a single straight line. One should

note that the experimental conditions are very different between the sets of data, and that no study on the effect of experimental parameters on the fragment size analysis methodology described here could be done given the limited sets of data available for which precise experimental parameters are known. In these conditions, we can deduce from the data that a single process should induce a single straight line distribution. The presence of two differing lines suggests that two differing processes are in effect. A possible explanation could be that for small peptide size, b ion fragmentation is induced by a purely ECD process, as proposed by Hakansson, Uggerud and Haselmann [10, 12, 42] whereas internal energy induced fragmentation of intermediate MH^+ ions, as suggested by Cooper [9] could be operative for larger peptide sizes, as the slope and intercept observed for large peptide sizes in ECD spectra are closer to those observed for 1+ charged parent ion CID spectra. Based on our results, no definite proof of these hypotheses exists, and one should also add the option of secondary fragmentation of c ions as a possible origin for the onset of b ions. Our data is not in agreement with this possibility for the smaller sized peptides (less than 10 residues long), for which the average size of occurrence of b and c fragments is identical, but cannot be ruled out for larger peptides.

Finally, it was observed that y ions did not follow an efficiency curve decaying as peptide size increased as for b ions (Figure 2) as would be expected for fragments formed by the same process through a common intermediate MH^+ ion. One hypothesis is that these y ions could derive from a secondary fragmentation of $z^\bullet$ ions. It should be noted that in such a case, one should expect the average fragment size to be smaller than that of the parent $z^\bullet$ fragments. Although the slope is slightly reduced, one can consider that even at large peptide size, the average y ions fragment size is of similar size than the $z^\bullet$ fragments, while for small peptides, the y fragments tend to be larger than the $z^\bullet$ fragments. Unfortunately, no easy explanation comes for this trend: as can be observed in Table 2, collision induced fragmentation of neither 1+ nor 2+ charged ions does lead to a fragment size dependence similar to that of those observed after ECD activation.

Relation between peptide size and fragmentation pathways

Extrapolation of the average fragment size to the pentapeptides which were the original object of this study can lead to some interesting observations: the expected average $z^{\bullet}$ fragment size for pentapeptides should be 0.8 residue. This is not possible, as it would mean that some fragments should be of zero-size. (In accordance, the trend for $z^{\bullet}$ fragments slightly tails for small peptides.) But it should be interpreted as an indication that $z^{\bullet}$ fragments require a favorable structure involving 5 N-terminal residues, structure which cannot be achieved by peptides smaller than 5 residues. It is therefore not unexpected that the pentapeptides do not lead to $z^{\bullet}$ fragmentation. These favorable structuring effects for $z^{\bullet}$ ions are also observed, in a lesser extent for c ions. Depending on the nature of the side-chain bearing the C-terminal charge (lysine or arginine), the effect of a local structure of the C-terminal chain, leading to the major c ions, is also observed. These structures, which appear at both the C and N-terminal side of the peptides, are likely to involve solvation of the charges by carbonyl groups, provided either by the amino-acid side-chains or by the peptidic carbonyls. As all these peptides are tryptic and doubly charged, one can postulate that one charge will be located at the C-terminal basic amino-acid, while the other one will be located at the N-terminal amine. The general behavior observed seems to indicate that, on the first order, the peptidic carbonyl functions are mostly involved in this solvation, the amino-acid side-chains being only second order modulators. In these conditions, as peptide size decreases, the two N-terminal and C-terminal structures will come closer and at one point will compete for solvation by available carbonyl groups, in effect disfavoring mechanisms that require such charge solvations.

If other competing mechanisms do not require charge solvation by carbonyl group from the peptide backbone, these will be favored as the other ones are being hindered. Such could be the case of the w fragmentation pathway, for which there does not seem to be a required position of the charge. The w pathway could therefore be in disfavorable competition with c/z fragmentation, and become operative only when the c/z fragmentation pathways become obstructed.

On the other hand, the b fragment size depends, with a slope of 1, on the peptide size, which indicates also a local structure effect. But it could also be that the local structure involved differs from that leading to c and z fragments. As peptide size decreases, and as the two ends come closer, a transition in the local

peptide structure could lead to a new type of folding. Such transitions of peptide structure with peptide size have been observed, such as the evolution from a coiled structure for small Na^+ cationized peptides towards more varied structures as size increases [48], and could be an explanation for the appearance of a pathway leading to b ions.

These elements point towards ECD fragmentation pathways which are strongly correlated with peptide structure, at least in terms of peptide fragment nature and abundances. It does not directly point out towards a given mechanism for the formation of c/z fragments, as each mechanism is dependent on the parent ion structure at some point. Nevertheless, since solvation of the charged site by carbonyl groups seems a key point and that most of the fragment intensity for c and z fragments is located at a position close to the neutralized charge end, it appears in favor of a Cornell-type mechanism, by which electron capture at a charged site leads to the transfer of a hypervalent hydrogen towards a neighboring carbonyl group, initially proton-bound to this charged site. This conclusion is similar to that already reached by R. Zubarev et al. for the z ions series, based on the SwedECD data [41], but we extend this observation to the c ions series.

If one considers the effect of size on the nature of the fragments being formed, there appears to be a strong transition between the b and w pathways and the c and z pathways as peptide size increases. The causes behind this transition could be of two kinds: a change in structure or a variation in internal energy per degree of freedom. The latter might be of importance: if one assumes that the recombination energy (or a constant part of it) is redistributed over the entire peptide before fragmentation (ergodic model), the energy available per degree of freedom decreases as peptide size increases. Therefore, the transition between both fragment series could be simply related to the internal energy available for fragmentation. The data presented here does not allow to draw a definite conclusion on this point, as it is also very possible that the two elements contribute to the final branching ratios between pathways: for instance, if the b fragmentation pathway requires more energy than the c and z pathways, and that a structural change disfavors c and z fragments, the b fragmentation pathway could be favored for small peptides both for structural and energetical reasons.

Conclusions

In this global picture, the "atypical" fragmentation observed for the pentapeptide series resolves to a fragmentation which appears typical for small peptides: extrapolation of the series observed on the SwedECD set of data leads to the conclusion that the pentapeptides studies initially fit quite well within the global picture: high intensities of *w*, *b* and *a* ions are expected for these small peptides. Since the database is biased towards "standard" *c* and *z* fragmentation, the absence of *c* and *z* fragments in some instances for pentapeptide fragmentation does not seem that surprising considering the general trend. In this sense, the pentapeptides used in this study can even be considered as good ground on which to study the electron capture behavior of small peptides. From the conclusions reached in this study, including the extrapolation of the results gained in the SwedECD database analysis, we intend to pinpoint specific changes (in size, composition, structure) that could lead to revert from the "atypical" fragmentation behavior observed for this series towards the formation of *c* and *z* fragments. As shown in Figure 1, change in size is a good start, but it will be of interest to change other parameters, such as the position of the charge bearing group, the nature of the charged group, or construction of sterically constrained structures, preventing or enhancing folding around the charges.

The main result in this work is the demonstration that, far from being a minor, secondary fragmentation pathway, the presence of *w* fragments in ECD spectra originates from a pathway separate from that leading to the major *z* fragments. This is of some importance, as *w* ions constitute the almost exclusive fragmentation by which isolecuine and leucine can be attributed by MS/MS based methods. We observe that *w* ions are often not correlated with *z* fragments: for instance AGALK does not produce any z_3 fragments although a w_3 fragment is observed for the analogous AGSLK peptide, which would indicate that the pathway leading to a *w* fragment is not blocked at the *z* step when formation of *w* fragments is not feasible. Our interpretation is in favor of a pathway leading directly to the *w* fragments, with a direct cleavage of both the N-C_α and C_β-X_γ bonds. Since efficiency seems to depend on the

stability of the radical lost, we would propose that the formation of the side-chain radical is a key step in the *w* fragmentation pathway.

This work has also shown a general linear behavior of the weighted average of the fragment size as a function of peptide length. This behavior has been exploited in the present article to gather more information on the fragmentation pathways, to compare these relative to each other. But, if one can show that the parameters of this evolution do not depend on instrumental parameters, these could also be used for fragmentation prediction softwares. For instance, in reference [4] the determination of the backbone cleavage site is based on pseudo-hydrogen affinities for the various sites: the position in the chain is not used as an important parameter, whereas our results show that the dependence of the fragment average size with peptide size could be a basis for prediction of fragmentation spectra. Additional parameters, such as width of the distribution should of course also be taken into account. In this sense, the strong dependence of ion fragment types on peptide size should also be introduced in such models. Furthermore, the methodology presented in this work will now be applied to other fragmentation spectra, and most relevantly to ETD fragmentation: comparison of the ECD and ETD fragmentation behavior both for specific peptides and for entire fragmentation datasets [49] could lead to further knowledge on the similarities and differences between these two kin activation methods.

Finally, in discriminating between the different pathways, our results strongly support a model based on parent ion structure, and more specifically, related to the structure of the peptide and its hydrogen bond network organization around the protonated sites. In this respect, this work has identified various pathways that have to be considered: namely the *c*/*z*, the *b* and the *w* pathways. Further work will pursue finding more experimental details discriminating between pathways. For instance, finding if size also plays a role in terms on distribution of the recombination energy over a larger number of internal degrees of freedom is of interest. The AGXLK pentapeptide system provides an example of a limit case for which no fragment follows one of the pathways (the *c*/*z* pathway): working on variations of this system, further work will focus on finding key elements required to open the *c/z* pathway, and as well to close specifically the *b* or *w* pathways.

Supporting Information Available

Complete size dependence plots for *b*, *c*-H$^{\bullet}$, *c*, *w*, *w*+H$^{\bullet}$, *y*-H$^{\bullet}$, *y*, z$^{\bullet}$, *z*+H pathways from the SwedECD database analysis. Size dependence plots for *b* and *y* fragments from CID databases. *w* fragment probability respective to peptide center for varying sub-population of the SwedECD database.

References

1. Aebersold, R.; Mann, M., Mass spectrometry-based proteomics. *Nature* **2003,** *422*, 198-207.

2. Zhang, Z. Q., Prediction of low-energy collision-induced dissociation spectra of peptides. *Analytical Chemistry* **2004,** *76*, 3908-3922.

3. Zhang, Z. Q., Prediction of low-energy collision-induced dissociation spectra of peptides with three or more charges. *Analytical Chemistry* **2005,** *77*, 6364-6373.

4. Zhang, Z. Q., Prediction of Electron-Transfer/Capture Dissociation Spectra of Peptides. *Analytical Chemistry* **2010,** *82*, 1990-2005.

5. Antoine, R.; Broyer, M.; Chamot-Rooke, J.; Dedonder, C.; Desfrancois, C.; Dugourd, P.; Gregoire, G.; Jouvet, C.; Onidas, D.; Poulain, P.; Tabarin, T.; van der Rest, G., Comparison of the fragmentation pattern induced by collisions, laser excitation and electron capture. Influence of the initial excitation. *Rapid Commun Mass Sp* **2006,** *20*, 1648-1652.

6. Zubarev, R. A.; Kelleher, N. L.; McLafferty, F. W., Electron capture dissociation of multiply charged protein cations. A nonergodic process. *J Am Chem Soc* **1998,** *120*, 3265-3266.

7. Cooper, H. J.; Hakansson, K.; Marshall, A. G., The role of electron capture dissociation in biomolecular analysis. *Mass Spectrom Rev* **2005,** *24*, 201-222.

8. Zubarev, R. A.; Kruger, N. A.; Fridriksson, E. K.; Lewis, M. A.; Horn, D. M.; Carpenter, B. K.; McLafferty, F. W., Electron capture dissociation of gaseous multiply-charged proteins is favored at disulfide bonds and other sites of high hydrogen atom affinity. *J Am Chem Soc* **1999,** *121*, 2857-2862.

9. Cooper, H. J., Investigation of the presence of b ions in electron capture dissociation mass spectra. *Journal of the American Society for Mass Spectrometry* **2005,** *16*, 1932-1940.

10. Haselmann, K. F.; Schmidt, M., Do b-ions occur from vibrational excitation upon H-desorption in electron capture dissociation? *Rapid Commun Mass Sp* **2007,** *21*, 1003-1008.

11. Lee, S.; Chung, G.; Kim, J.; Oh, H. B., Electron capture dissociation mass spectrometry of peptide cations containing a lysine homologue: a mobile proton model for explaining the observation of b-type product ions. *Rapid Commun Mass Sp* **2006,** *20*, 3167-3175.

12. Liu, H. C.; Hakansson, K., Abundant b-type ions produced in electron capture dissociation of peptides without basic amino acid residues. *Journal of the American Society for Mass Spectrometry* **2007,** *18*, 2007-2013.

13. Cooper, H. J.; Hudgins, R. R.; Hakansson, K.; Marshall, A. G., Characterization of amino acid side chain losses in electron capture dissociation. *Journal of the American Society for Mass Spectrometry* **2002,** *13*, 241-249.

14. Falth, M.; Savitski, M. M.; Nielsen, M. L.; Kjeldsen, F.; Andren, P. E.; Zubarev, R. A., Analytical Utility of Small Neutral Losses from Reduced Species in Electron Capture Dissociation Studied Using SwedECD Database. *Analytical Chemistry* **2008,** *80*, 8089-8094.

15. Fung, Y. M. E.; Chan, T. W. D., Experimental and theoretical investigations of the loss of amino acid side chains in electron capture dissociation of model peptides. *Journal of the American Society for Mass Spectrometry* **2005,** *16*, 1523-1535.

16. Kruger, N. A.; Zubarev, R. A.; Horn, D. M.; McLafferty, F. W., Electron capture dissociation of multiply charged peptide cations. *International Journal of Mass Spectrometry* **1999,** *187*, 787-793.

17. Breuker, K.; Oh, H. B.; Cerda, B. A.; Horn, D. M.; McLafferty, F. W., Hydrogen atom loss in electron-capture dissociation: a Fourier transform-ion cyclotron resonance study with single isotopomeric ubiquitin ions. *European Journal of Mass Spectrometry* **2002,** *8*, 177-180.

18. Savitski, M. M.; Nielsen, M. L.; Zubarev, R. A., Side-chain losses in electron capture dissociation to improve peptide identification. *Analytical Chemistry* **2007,** *79*, 2296-2302.

19. Kjeldsen, F.; Haselmann, K. F.; Sorensen, E. S.; Zubarev, R. A., Distinguishing of Ile/Leu amino acid residues in the PP3 protein by (hot) electron capture dissociation in Fourier transform ion cyclotron resonance mass spectrometry. *Analytical Chemistry* **2003,** *75*, 1267-1274.

20. Kjeldsen, F.; Zubarev, R., Secondary losses via gamma-lactam formation in hot electron capture dissociation: A missing link to complete de novo sequencing of proteins? *J Am Chem Soc* **2003,** *125*, 6628-6629.

21. Lin, C.; O'Connor, P. B.; Cournoyer, J. J., Use of a double resonance electron capture dissociation experiment to probe fragment intermediate lifetimes. *Journal of the American Society for Mass Spectrometry* **2006,** *17*, 1605-1615.

22. O'Connor, P. B.; Lin, C.; Cournoyer, J. J.; Pittman, J. L.; Belyayev, M.; Budnik, B. A., Long-lived electron capture dissociation product ions experience radical migration via hydrogen abstraction. *Journal of the American Society for Mass Spectrometry* **2006,** *17*, 576-585.

23. Savitski, M. M.; Kjeldsen, F.; Nielsen, M. L.; Zubarev, R. A., Hydrogen rearrangement to and from radical z fragments in electron capture dissociation of peptides. *Journal of the American Society for Mass Spectrometry* **2007,** *18*, 113-120.

24. Jensen, C. S.; Wyer, J. A.; Nielsen, S. B., Electron capture induced dissociation of dipeptide dications: where does the charge go? *Phys Chem Chem Phys* **2010,** *12*, 12961-12963.

25. Simons, J., Mechanisms for S-S and N-C-alpha bond cleavage in peptide ECD and ETD mass spectrometry. *Chem Phys Lett* **2010,** *484*, 81-95.

26. Syrstad, E. A.; Turecek, F., Toward a general mechanism of electron capture dissociation. *Journal of the American Society for Mass Spectrometry* **2005,** *16*, 208-224.

27. Sawicka, A.; Skurski, P.; Hudgins, R. R.; Simons, J., Model calculations relevant to disulfide bond cleavage via electron capture influenced by positively charged groups. *J Phys Chem B* **2003,** *107*, 13505-13511.

28. Sobczyk, M.; Anusiewicz, W.; Berdys-Kochanska, J.; Sawicka, A.; Skurski, P.; Simons, J., Coulomb-assisted dissociative electron attachment: Application to a model peptide. *Journal of Physical Chemistry A* **2005,** *109*, 250-258.

29. Turecek, F.; Chung, T. W.; Moss, C. L.; Wyer, J. A.; Ehlerding, A.; Holm, A. I. S.; Zettergren, H.; Nielsen, S. B.; Hvelplund, P.; Chamot-Rooke, J.; Bythell, B.; Paizs, B., The Histidine Effect. Electron Transfer and Capture Cause Different Dissociations and Rearrangements of Histidine Peptide Cation-Radicals. *J Am Chem Soc* **2010,** *132*, 10728-10740.

30. Chen, X. H.; Turecek, F., The arginine anomaly: Arginine radicals are poor hydrogen atom donors in electron transfer induced dissociations. *J Am Chem Soc* **2006,** *128*, 12520-12530.

31. Chamot-Rooke, J.; Malosse, C.; Frison, G.; Turecek, F., Electron capture in charge-tagged peptides. Evidence for the role of excited electronic states. *Journal of the American Society for Mass Spectrometry* **2007,** *18*, 2146-2161.

32. Chamot-Rooke, J.; van der Rest, G.; Dalleu, A.; Bay, S.; Lemoine, J., The combination of electron capture dissociation and fixed charge derivatization increases sequence coverage for O-glycosylated and O-phosphorylated peptides. *Journal of the American Society for Mass Spectrometry* **2007,** *18*, 1405-1413.

33. Han, H. L.; Xia, Y.; McLuckey, S. A., Ion trap collisional activation of c and z(center dot) ions formed via gas-phase ion/ion electron-transfer dissociation. *Journal of Proteome Research* **2007,** *6*, 3062-3069.

34. Chung, T. W.; Turecek, F., Backbone and Side-Chain Specific Dissociations of z Ions from Non-Tryptic Peptides. *Journal of the American Society for Mass Spectrometry* **2010,** *21*, 1279-1295.

35. Tsybin, Y. O.; Witt, M.; Baykut, G.; Kjeldsen, F.; Hakansson, P., Combined infrared multiphoton dissociation and electron capture dissociation with a hollow electron beam in Fourier transform ion cyclotron resonance mass spectrometry. *Rapid Commun Mass Sp* **2003,** *17*, 1759-1768.

36. Senko, M. W.; Beu, S. C.; Mclafferty, F. W., Determination of Monoisotopic Masses and Ion Populations for Large Biomolecules from Resolved Isotopic Distributions. *Journal of the American Society for Mass Spectrometry* **1995,** *6*, 229-233.

37. Falth, M.; Savitski, M. M.; Nielsen, M. L.; Kjeldsen, F.; Andren, P. E.; Zubarev, R. A., SwedCAD, a database of annotated high-mass accuracy MS/MS spectra of tryptic peptides. *Journal of Proteome Research* **2007,** *6*, 4063-4067.

38. Stein, S. E.; Rudnick, P. A., NIST Peptide Tandem Mass Spectral Libraries. Human Peptide Mass Spectral Reference Data, *M. musculus*, ion trap. In Official Build Date October 12, 2009 ed.; National Institute of Standards and Technology, Gaithersburg, MD, 20899: 2009.

39. Bernigaud, V.; Cederquist, H.; Haag, N.; Holm, A. I. S.; Huber, B. A.; Hvelplund, P.; Kadhane, U.; Larsen, M. K.; Manil, B.; Nielsen, S. B.; Panja, S.; Ptasinska, S.; Rangama, J.; Reinhed, P.; Schmidt, H. T.; Streletskii, A. V.; Stochkel, K.; Worm, E. S.; Zettergren, H., Electron capture-induced dissociation of AK dipeptide dications: Influence of ion velocity, crown-ether complexation and collision gas. *International Journal of Mass Spectrometry* **2008,** *276*, 77-81.

40. Sun, Q. Y.; Nelson, H.; Ly, T.; Stoltz, B. M.; Julian, R. R., Side Chain Chemistry Mediates Backbone Fragmentation in Hydrogen Deficient Peptide Radicals. *Journal of Proteome Research* **2009,** *8*, 958-966.

41. Savitski, M. M.; Kjeldsen, F.; Nielsen, M. L.; Zubarev, R. A., Complementary sequence preferences of electron-capture dissociation and vibrational excitation in fragmentation of polypeptide polycations. *Angew Chem Int Edit* **2006,** *45*, 5301-5303.

42. Bakken, V.; Helgaker, T.; Uggerud, E., Models of fragmentations induced by electron attachment to protonated peptides. *European Journal of Mass Spectrometry* **2004,** *10*, 625-638.

43. Leymarie, N.; Costello, C. E.; O'Connor, P. B., Electron capture dissociation initiates a free radical reaction cascade. *J Am Chem Soc* **2003,** *125*, 8949-8958.

44. Lin, C.; Cournoyer, J. J.; O'Connor, P. B., Use of a double resonance electron capture dissociation experiment to probe fragment intermediate lifetimes (vol 17, pg 1605, 2006). *Journal of the American Society for Mass Spectrometry* **2007,** *18*, 165-165.

45. Wysocki, V. H.; Tsaprailis, G.; Smith, L. L.; Breci, L. A., Special feature: Commentary - Mobile and localized protons: a framework for understanding peptide dissociation. *J Mass Spectrom* **2000,** *35*, 1399-1406.

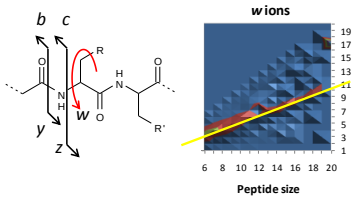
46. Summerfield, S. G.; Gaskell, S. J., Fragmentation efficiencies of peptide ions following low energy collisional activation. *International Journal of Mass Spectrometry* **1997,** *165*, 509-521.

47. Paizs, B.; Suhai, S., Fragmentation pathways of protonated peptides. *Mass Spectrom Rev* **2005,** *24*, 508-548.

48. Semrouni, D.; Ohanessian, G.; Clavaguera, C., Structural, energetic and dynamical properties of sodiated oligoglycines: relevance of a polarizable force field. *Phys Chem Chem Phys* **2010,** *12*, 3450-3462.

49. Sun, R. X.; Dong, M. Q.; Song, C. Q.; Chi, H.; Yang, B.; Xiu, L. Y.; Tao, L.; Jing, Z. Y.; Liu, C.; Wang, L. H.; Fu, Y.; He, S. M., Improved Peptide Identification for Proteomic Analysis Based on Comprehensive Characterization of Electron Transfer Dissociation Spectra. *Journal of Proteome Research* **2010,** *9*, 6354-6367.

Graphical Abstract



Peptide fragmentation database analysis reveals that peptide size is a key factor in the selection of the *b* and *w* pathways in electron capture dissociation.

Supporting information: Dissociation channel dependence on peptide size observed in electron capture dissociation of tryptic peptides, by G. van der Rest, R. Hui, G. Frison and J. Chamot-Rooke

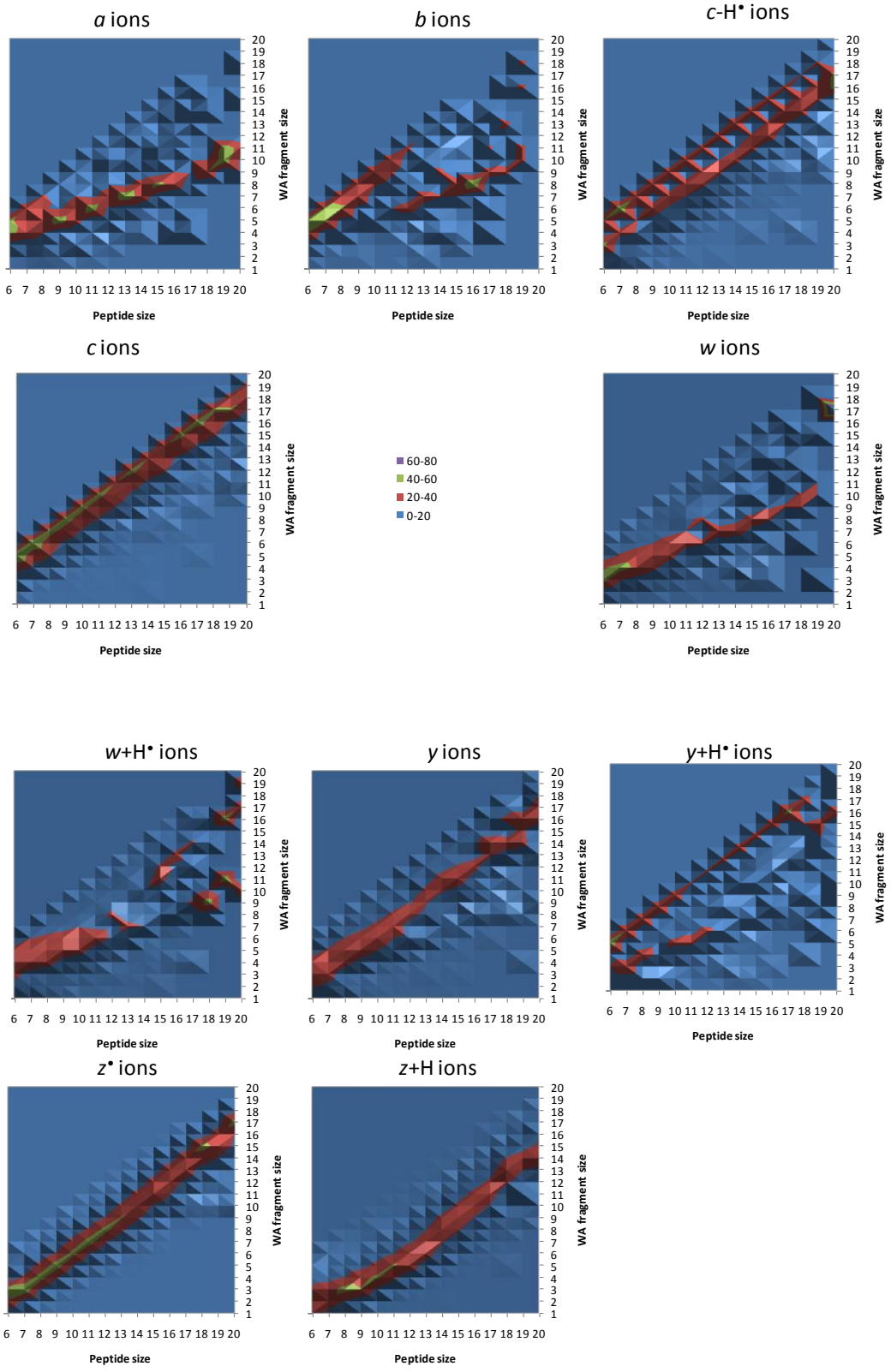


Figure S1. Distribution of the weighted average (WA) fragment size as a function of peptide size for the various fragmentation pathways. The bin size for WA fragment size is set to 1 (see text) and the color code represents the percentage of peptides in a WA fragment size bin within the total number of peptides of the same size leading to at least one fragment of the type considered. (Each column is normalized to a total of 100%.)

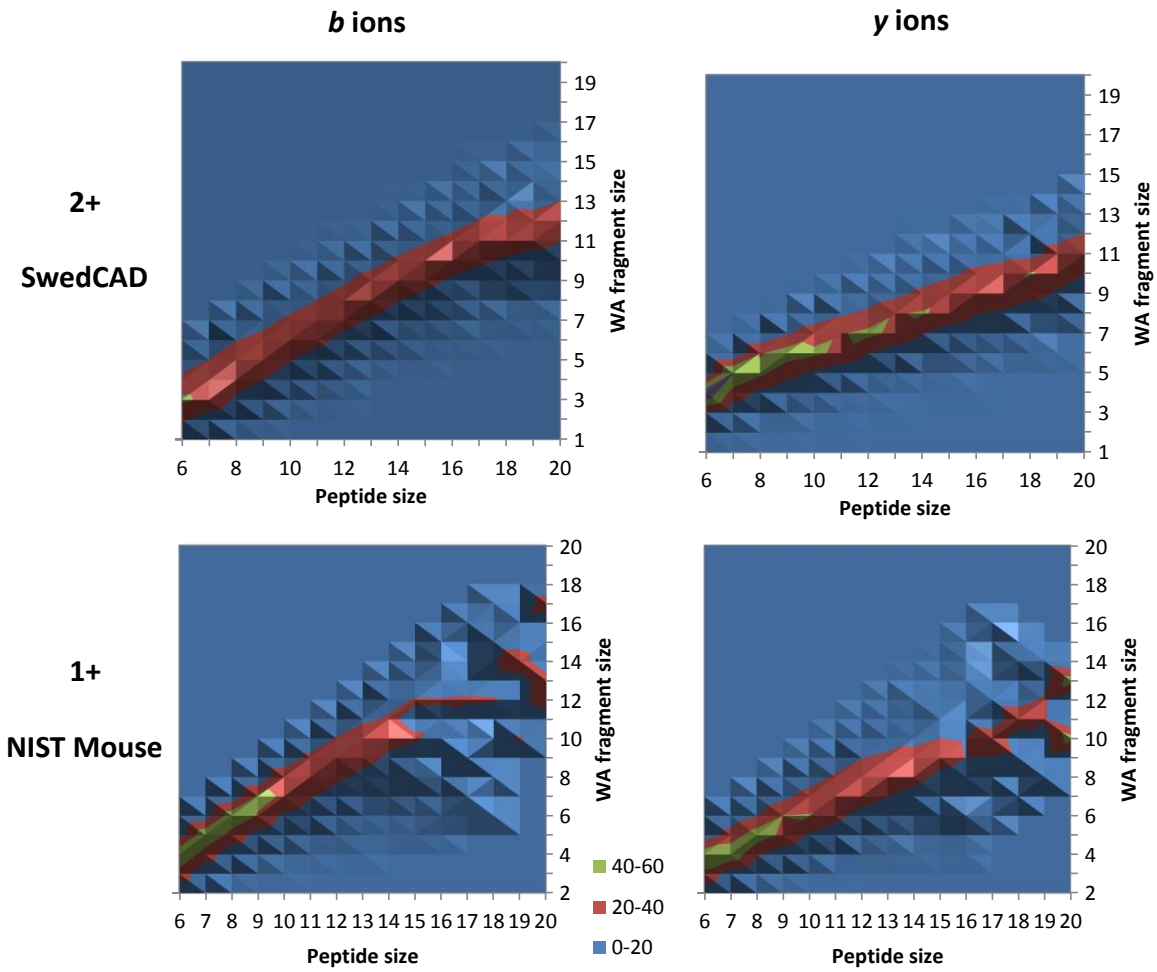


Figure S2. Distribution of the weighted average (WA) fragment size as a function of peptide size for the *b* (left) and *y* (right) fragmentation pathways. The database are the SwedCAD database for 2+ charged parent ions (top row) and the NIST *Mus musculus* database restricted to 1+ charged parent ions (bottom row). The bin size for WA fragment size is set to 1 (see text) and the color code represents the percentage of peptides in a WA fragment size bin within the total number of peptides of the same size leading to at least one fragment of the type considered. (Each column is normalized to a total of 100%.)

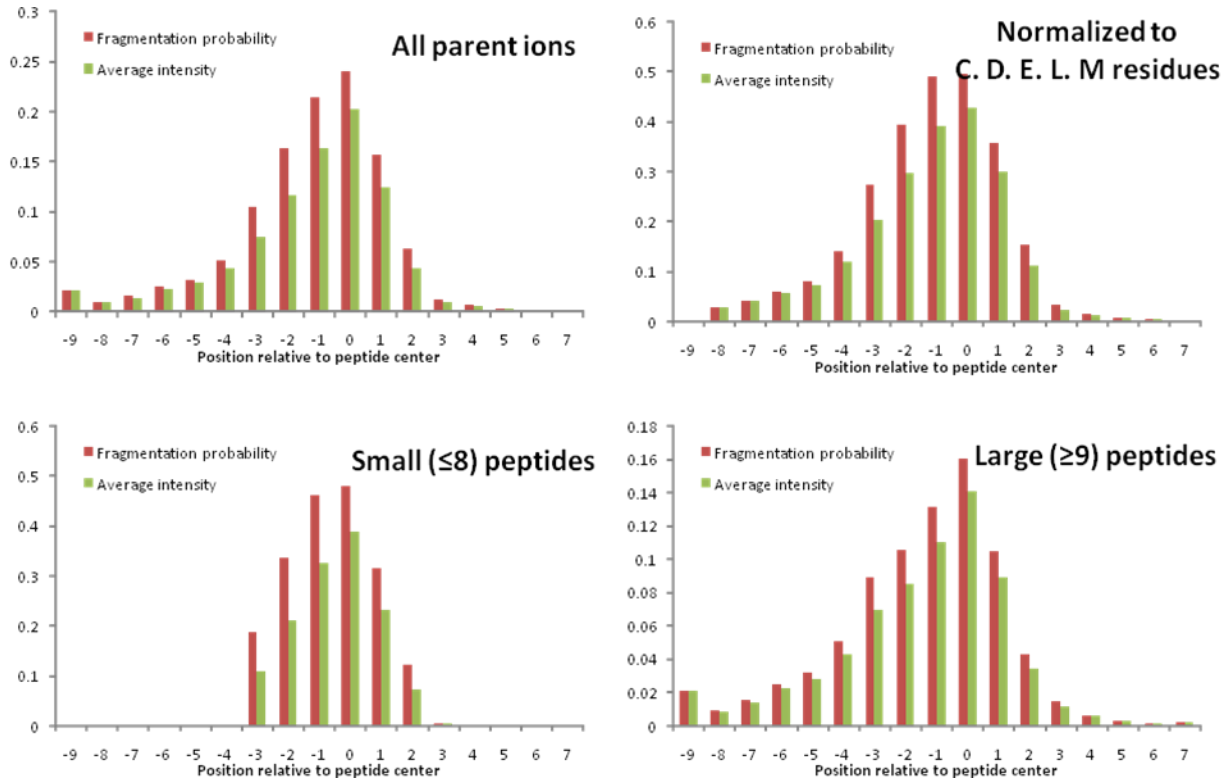


Figure S3. Probability of *w* ion formation and average *w* ion normalized intensity at positions expressed relatively to the peptide center (see Figure 5). Top left: same as figure 5, for comparison. Top right: fragmentation probabilities and intensities are expressed relative to the total number of amino-acid presenting a C, D, E, L or M residue at the given position. These residues were chosen as those that lead to frequent (> 20%) *w* ion formation. This allows to see that the observed distribution is not related to a bias towards favorable residues located at peptide center. Bottom row: the entire peptide population has been split in two sets: one for small (≤8 residues; left) peptides and large (≥9 residue; right) peptides. In both sets the *w* ions appear to favor central residues, although one can observe that the probability to lead to *w* ions is much reduced for larger peptides.