

Automatic *De Novo* Sequencing of peptides by Electron Transfer Dissociation

F. Martin-Maroto, Zhiqi Hao, R.Biringer, J.Vazquez, A. FR Hühmer

Thermo Electron Corporation, San Jose, CA ● Centro de Biología Molecular Severo Ochoa, Madrid, Spain

Overview

New ion activation techniques, such as electron capture dissociation (ECD) and electron transfer dissociation (ETD) have been introduced recently for the analysis of peptides and proteins [1]. These techniques provide a highly efficient method for the sequence analysis of peptides due to formation of homogeneous and complete fragment ion series along the entire peptide backbone. In this report we examine the differences, advantages and disadvantages of ETD fragmentation techniques in comparison with conventional collision-induced dissociation (CID) for *de novo* sequencing applications. We conclude that ETD peptide fragmentation is ideal for *de novo* sequencing of large or modified peptides.

Introduction

ETD fragmentation of peptides is achieved when singly charged fluoranthene anions transfer an electron to multiply protonated peptides. For this to occur, multiple basic residues (K, R, H) should be part of the peptide sequence, which normally requires long peptides or a general basic character of the peptide sequence. The fragmentation occurs following an exothermic reaction depicted in Figure 1 [1]. This pathway is essentially the same than for ECD fragmentation; the difference is only in the origin of the electrons; while in ECD electrons originate from a hot filament and persist in gas phase, in ETD electrons are transferred from anions persisting in gas phase. The advantage is that ETD fragmentation is possible in ion traps while ECD is only possible on FT instruments (hot electrons are not trapped in ion traps). Here we describe the main features of ETD spectra and analyze the advantages and disadvantages of ETD compared with CID for *de novo* sequencing applications. We conclude that the ETD fragmentation technique is ideal for *de novo* sequencing of modified and/or long peptides. It is especially interesting for potential use in studying posttranslational modifications by MS.

Methods

Fragmentation by electron transfer dissociation (ETD) was implemented on a Finnigan™ LTQ™ linear ion trap mass spectrometer prototype (Thermo Electron, San Jose, CA). To compare with CID fragmentation, a classical protein digest was analyzed alternating between CID and ETD, conveniently done in consecutive scans. The best mass spectra (selected manually) where used for comparison. A prototype DeNovoX™ software (Thermo Electron, San Jose, CA) was used to interpret the spectra. The DeNovoX software was enabled to:

- identify complementary *c/z** ion pairs
- deconvolve the complex "isotope envelopes" observed on ETD
- identify secondary *a* and *y* ions, present on ETD spectra

FIGURE 1. Fragmentation pathway of Electron Transfer Dissociation

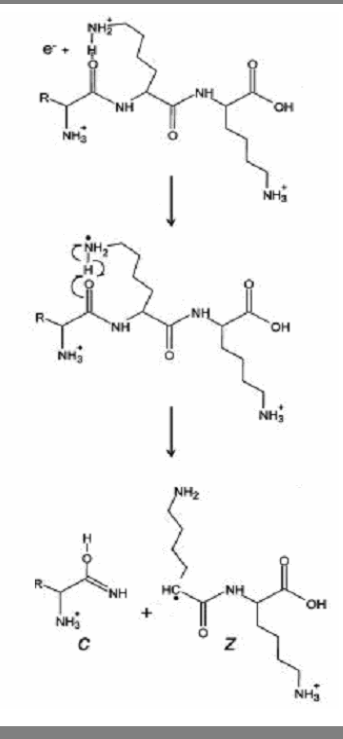
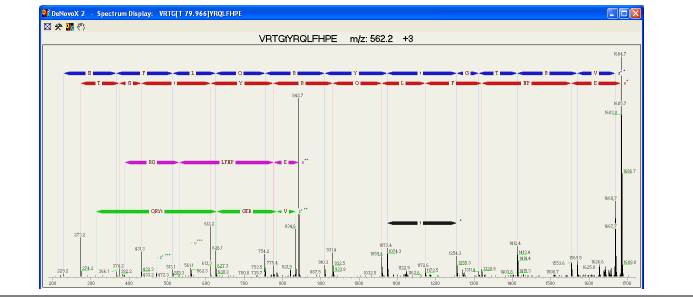


FIGURE 2. Triply charged peptide containing phosphorylated threonine *de novo* sequenced automatically by DeNovoX software. A 10 amino acid tag is unambiguously identified including the phosphorylated residue. No phosphoric acid losses from fragment ions are observable. Highlighted in black on the spectrum is the phosphorylated threonine match "T".



Results

Main ions series: ETD fragmentation leads to complementary *c* (N terminal, with mass $\sum m_R + 18.03$) and *z** ions (C terminal, with mass $\sum m_R + 3$), instead of the *y* and *b* characteristic of CID. The precursor charge frequently gets reduced by electron capturing becoming $[M + nH]^{+(n-1)+}$ or $[M + nH]^{+(n-2)+}$ prior to the dissociation; this charge reduction leads to the superposition of isotope clusters corresponding with ions one or two hydrogen masses heavier than *c/z** ions, producing what looks like an extended "isotope cluster" involving 4 to 6 ions. This phenomenon is more frequently observed for heavier product ions while lighter ones tend to appear isolated without superposition. The less massive ion observed on these "isotope clusters" corresponds normally with the *c/z** ion, although sometimes corresponds with *z* + H/c + H*, or even with *z* - H/c - H*. For *de novo* sequencing this represents a challenge, and is the main disadvantage of ETD compared with CID. Dimers are sometimes present, and can be recognized by a 0.5 da. separation from the singly charged ions of the main clusters.

Good fragmentation of modified peptides: ETD fragmentation is affected very little by chemical modifications such as phosphorylation. In the example of Figure 1, a clean ETD spectrum from a triply charged peptide containing phosphothreonine is successfully sequenced by automatic *de novo* sequencing. No phosphoric acid losses were observed. Peptides containing phosphorylated residues usually produce complicated CID spectra difficult to interpret, with dominant ions corresponding to phosphoric acid losses. This is a clear advantage of the application of ETD to phosphopeptides compared to CID based fragmentation.

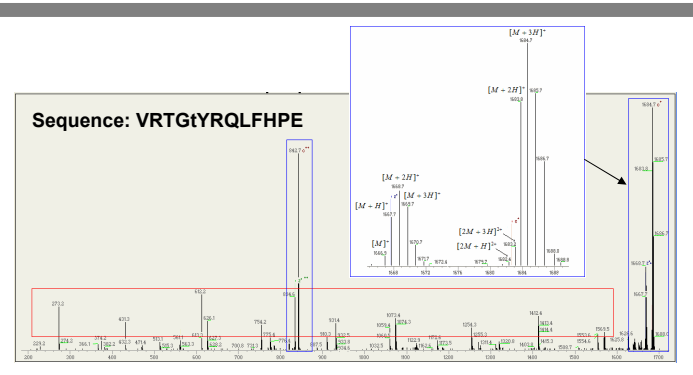


FIGURE 3. Typical ETD spectra; boxed in blue, the *c* and *z** singly and doubly *charge-reduced* precursor ions. In the red box are enclosed the fragment ions, all of which show similar abundances indicative of the non-ergodic nature of the ETD fragmentation process.

Complete and Homogeneous Fragmentation: ETD spectra of peptides display a characteristic sequence-independent fragmentation. Product ions have similar abundances except for those corresponding to different charge states of the intact precursor (Figure 2). Hence, ETD spectra carry little information codified on the ion abundances. In contrast, CID product ions have very different abundances that depend on the neighboring amino acids and other characteristics of the peptide sequence. Nevertheless, abundances of ions of CID spectra are difficult to predict and abundance patterns are less frequently used for *de novo* sequencing of CID spectra [2]. Singly charged *c* and *z** ions can be found from the N to the C terminus all along the peptide backbone. Following singly charged ions, the entire sequence can be read from N to C terminus and from C to N terminus leading to increased confidence for spectrum interpretation. In contrast, interpretation of peptide spectra from CID experiments it is necessary to derive information from *y* and *b* series fragments and from different charge states to derive and read the complete peptide sequence. *De novo* sequencing of the N terminal part of the peptide of a CID spectrum can normally be achieved only from doubly charged *y* ions, while ETD allows N terminal sequencing directly from a complete singly charged *c* fragment ion series.

Peptide sequences containing proline residues show resistance to bond cleavage by ETD on the N-terminal side due to the presence of the tertiary nitrogen. All the multiply charged peptides presented in Table 1 could be completely sequenced *de novo* with the exception of the peptide sequence LFTFHADICLTLPDTEK that was incompletely interpreted as LFTFHADIC[210.0]DTEK, based on the incomplete bond cleavage between residue P and L. Absence of a fragment ion may indicate the presence of proline.

Clean Fragmentation: Product ions are clearly separated from internal fragments, side chain losses and chemical noise. In the examples investigated for *de novo* sequencing, multiply charged fragment ions are either not present or have reduced abundances compared with the singly charged ions. All the singly charged product ions of spectra listed in Table 1 are part of the 100 most abundant ions in the spectra, which normally is not the case for all fragment ions in CID spectra. Charge-reduced species derived from the absorption of additional electrons to the precursor ion are usually the most abundant ions of the spectrum. The intact precursor appears in multiple charge states (principally as *c* ion, but present as *z** as well) and it is usually responsible for the most abundant ion of the spectrum. Side chain losses of the singly charged precursor ion are usually present, some of them very intense. These ions can eventually carry valuable sequencing information and they don't obfuscate the spectrum since their masses are not compatible with those of the fragment ions.

Neutral losses and secondary fragmentation pathways: The water, ammonium and carbon monoxide losses that are key for the *de novo* interpretation of CID spectra are not observed in ETD spectra. Nevertheless, ETD delivers equivalent alternative information; besides the fragmentation pathway leading to *c/z** ion series there is a complementary fragmentation pathway leading to a *y/a* ion series. The *y* ion, 16 amu heavier than the *z** ion, is frequently observed. This + 16 amu accompanying peak can serve the same role that the CO neutral loss (b - 28 amu) serves in *de novo* sequencing of CID spectra. The *y* ion is a reliable indicator of *z** ion series in ETD the same as the neutral loss of CO is an indicator of *b* series in CID. Although appearing in lower proportion than *y* ions, some *a* ions accompanying the *c* ions, i.e. *c* - 17, and sometimes even *b* ions, *c* + 11, are observed as well, (3). The neutral loss relationship of the secondary fragmentation pathway was exploited in the detection algorithm of the DeNovoX program and appropriately modified to identify differences of +16, +11 and -17 mass unit. The spectral pre-processing routine identifies the mass differences without knowledge of the actual peptide sequence before performing the *de novo* sequencing. It is normally possible to find many pairs of ions separated by any given mass difference just by chance. Nevertheless, the DeNovoX algorithm deconvolves the isotopic envelopes and correlates the complementary *c* - *z** ion pairs prior to the mass difference analysis achieving a reliable detection. *De novo* sequencing result for some high quality ETD spectra are displayed in Table 1, and show how *y* and *a* ions are detectable in ETD spectra.

Y IONS	Sequence	potential pairs (ion X) and (ion X + 16)	confirmed as y = (z* + 16)	detection failure: (c + 16)	detection failure: (noise ion + 16)	z* ions	ions
	EcdCKPLEK	7	6	1	0	10	275
	FESHNFYGATIR	3	2	0	1	12	243
	GLSPRPQTADQAR	8	2	0	0	16	292
	GRHEAEKPLAQSHATK	1	1	0	0	17	252
	KTDGAPFYTDHAK	6	4	1	1	15	307
	KYAAELHLVHWIK	3	2	1	0	14	255
	LFTFHADICLTLPDTEK	7	7	0	0	16	386
	VEADLASHGGEIVR	8	7	0	1	15	265
		42	36	4	2		
A IONS	Sequence	potential pairs (ion X) and (ion X - 17)	confirmed as a = (c - 17)	detection failure: (z* - 17)	detection failure: (noise ion - 17)	c ions	ions
	EcdCKPLEK	1	0	0	1	10	275
	FESHNFYGATIR	2	1	1	0	12	243
	GLSPRPQTADQAR	2	3	1	0	16	292
	GRHEAEKPLAQSHATK	1	1	0	0	17	252
	KTDGAPFYTDHAK	6	4	1	1	15	307
	KYAAELHLVHWIK	2	2	0	0	14	255
	LFTFHADICLTLPDTEK	3	1	1	0	16	386
	VEADLASHGGEIVR	0	0	0	0	15	265
		19	12	4	3		
B IONS	Sequence	potential pairs (ion X) and (ion X + 11)	confirmed as a = (c + 11)	detection failure: (z* + 11)	detection failure: (noise ion + 11)	c ions	ions
	EcdCKPLEK	0	0	0	0	10	275
	FESHNFYGATIR	0	0	0	0	12	243
	GLSPRPQTADQAR	1	1	0	0	16	292
	GRHEAEKPLAQSHATK	1	1	0	0	17	252
	KTDGAPFYTDHAK	1	1	0	0	15	307
	KYAAELHLVHWIK	0	0	0	0	14	255
	LFTFHADICLTLPDTEK	0	0	0	0	16	386
	VEADLASHGGEIVR	0	0	0	0	15	265
		2	2	0	0		

TABLE 1. Automatic identification of *y*, *a* and *b* ions using DeNovoX. The probabilistic algorithm identifies pairs of ions that can correspond with (*c*, *a*) pairs, (*c*, *b*) pairs, and (*z**, *y*) pairs. Results show how the presence of secondary fragment ions of type *y* and *a* allow the program to distinguish *c* from *z** ions before the *de novo* sequencing routine. Breaking the symmetry between *c* and *z** series (otherwise indistinguishable in ETD) is key for the successful interpretation of ETD spectra since it provides the orientation of the sequence tags.

The results of the spectrum pre-processing routine is important since the ion pairs (*c*, *a*) and (*z**, *y*) are separated by a different mass gap which permits distinguishing *c* from *z** ions allowing confident determination of the sequence orientation. In the experiment of Table 1, 36 "y" ions were correctly identified from a total of 42 ion pairs separated by 16 mass units of the (*z**, *y*) pair. An additional 12 *a* ions were correctly identified from a total of 19 pairs, and 2 *b* ions were correctly identified from a total of 2 ion pairs. All the ion pairs were singly charged, even when differences corresponding to multiply charged ions were considered during data processing.

Conclusions

- ETD spectra of peptides display a characteristic sequence-independent fragmentation. Product ions have similar abundances except for those corresponding to different charge states of the intact precursor.
- Singly charged *c* and *z** ions can frequently be found along the peptide backbone. Following singly charged ions, the entire sequence can be read from N to C terminus and from C to N terminus.
- Detecting complementary *c/z** ion pairs together with secondary *y/a* ions allows distinguishing *c* from *z** ions.
- ETD fragmentation is very little affected by chemical modifications of amino acid side chains, such as serine or threonine phosphorylation.
- ETD appears to be the preferred fragmentation method for *de novo* sequencing applications of large peptides or modified peptides containing basic residues.

References

1. Syka JE, Coon JJ, Schroeder MJ, Shabanowitz J and Hunt DF. 2004. Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. *Proc Natl Acad Sci U S A*. 101(26):9528-33.
2. Zhang, Prediction of low energy CID spectra of peptides, 2002 ASMS
3. Helen J. Cooper, Investigation of the presence of *b* ions in ECD mass spectra, JASMS, October 2005
4. Joshua J. Coon *et al*, Anion dependence in the partitioning between proton and electron transfer in ion/ion reactions, IJMS, 2004

Acknowledgements

The authors gratefully acknowledge John, E. P. Syka, for useful comments and advice. All trademarks are the property of Thermo Electron Corporation and its subsidiaries.