Peptide Fragmentation

DOI: 10.1002/ange.200601240

Complementary Sequence Preferences of Electron-Capture Dissociation and Vibrational Excitation in Fragmentation of Polypeptide Polycations**

*Mikhail M. Savitski, Frank Kjeldsen, Michael L. Nielsen, and Roman A. Zubarev**

Electron-capture dissociation (ECD)[1] of polypeptide polycations produces preferential cleavage of the backbone N–C_α bond (*c*, *z* fragments), while fragmentation based on vibrational excitation (VE) of the same ions (for example, collision-activated dissociation (CAD), infrared multiphoton dissociation (IRMPD), or surface-induced dissociation (SID)) gives preferential C–N cleavage (*b*, *y* fragment ions). Both types of fragmentation phenomena show preferences for the context of the amino acids. For example, VE has a strong preference for the N-terminal side of proline and the N-terminal side of aspartic acid.[2] More subtle preferences in CAD have been studied recently on a large data set.[3] Cleavage on the N-terminal side of Pro by ECD are absent because of the tertiary amide nitrogen atom.[1] Preferential cleavage on the C-terminal side of tryptophan has been reported,[4] but in general it has been unclear whether amino acid preferences in ECD are similar to those in CAD.[4] This question has both fundamental and application-related aspects. According to the generally accepted theory,[5] cleavage of C–N bonds by VE are induced by a mobile proton solvated by backbone amides. The presence of a proton at the amide nitrogen atom greatly weakens the amide bond,[5] and can also reduce the strength of other nearby bonds.[6] Thus, cleavage of the amide bond by VE is a local phenomenon with respect to the proton location. Both local[1,7] and nonlocal[8–11] cleavage mechanisms have been suggested for ECD. According to the local ECD model, the N–C_α bond next to the position of the neutralized proton is cleaved. Thus, ECD may be expected to have similar preferences as VE for amino acid contexts (except proline). The practical aspect of this issue is the use of CAD and ECD in tandem mass spectrometry, an important tool for proteomics. Application of two fragmentation methods instead of one is only advantageous if the methods are complementary in terms of sequence preferences.

Some CAD/ECD complementarity has been noticed on a limited data set.[4] Here we report the results of an exhaustive comparison performed on a library containing 14967 pairs of ECD and CAD mass spectra of tryptic peptide dications. The high mass accuracy used (± 3 ppm for molecular mass; ± 20 mDa for fragment masses)[12] ensured unique assignment of the fragments in 99+ % of the cases. Most of the identified fragment ions were singly charged C-terminal products: *y* ions in CAD and *z* ions ($z^\bullet$ and z') in ECD. The distributions of the cases when the y_{m-n} (z_{m-n}) fragment (where *m* is the peptide length) had the highest abundance amongst all the same-type fragments are shown in Figure 1. The striking features are the sharp peak at $n=2$ in the CAD histogram and broader peaks at $n\approx 4$ in both the CAD and ECD histograms. The positions of the maxima are independent of the peptide length and the C-terminal residue (Lys or Arg), and they correlate: 47 % of peptides giving the most abundant z_{m-4} ions by ECD produce y_{m-2} ions as the most abundant fragments by CAD (see inset in Figure 2).

Figure 2 presents statistics on the value $\Delta = n_{ECD} - n_{CAD}$ measured for the same peptide sequences. The distribution has a global maximum at $\Delta = +2$ which reflects the correlation between the peaks at $n_{ECD}=4$ and $n_{CAD}=2$ in Figure 1 (the inset in Figure 2 confirms this correlation). The small local maximum at $\Delta = -1$ corresponds to preferential cleavage at $n_{ECD}=4$ and $n_{CAD}=5$. The local minimum at $\Delta = 0$ means that cleavage by ECD at the same position as by CAD is relatively disfavored, contrary to predictions from the local-

[*] M. M. Savitski, Dr. F. Kjeldsen, M. L. Nielsen, Prof. R. A. Zubarev
Laboratory for Biological and Medical Mass Spectrometry
BMC, Uppsala University
Box 583, 75123 Uppsala (Sweden)
Fax: (+46) 18-471-22-44
E-mail: roman.zubarev@bmms.uu.se

[**] This work was supported by the Knut and Alice Wallenberg Foundation and Wallenberg Consortium North (grant WCN2003-UU/SLU-009 to R.Z. and instrumental grant to R.Z. and Carol Nilsson) as well as the Swedish research council (grants 621-2004-4897, 621-2002-5025 and 621-2003-4877 to R.Z.). Thomas Köcher and Christopher Adams are acknowledged for insightful discussions.

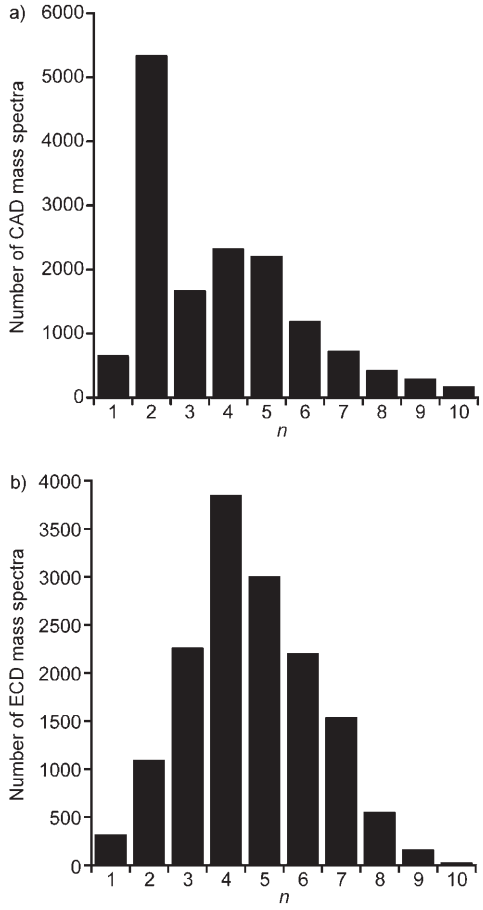

Figure 1. Distributions of tandem mass spectra in which the $(m-n)$th fragment (m is the peptide length) had the highest abundance of all identified C-terminal fragments for tryptic peptide dications fragmented by: a) CAD y'_{m-n} ions; b) ECD z_{m-n} ions (the most abundant of $z^{\cdot}_{m-n}$ and z'_{m-n} species).

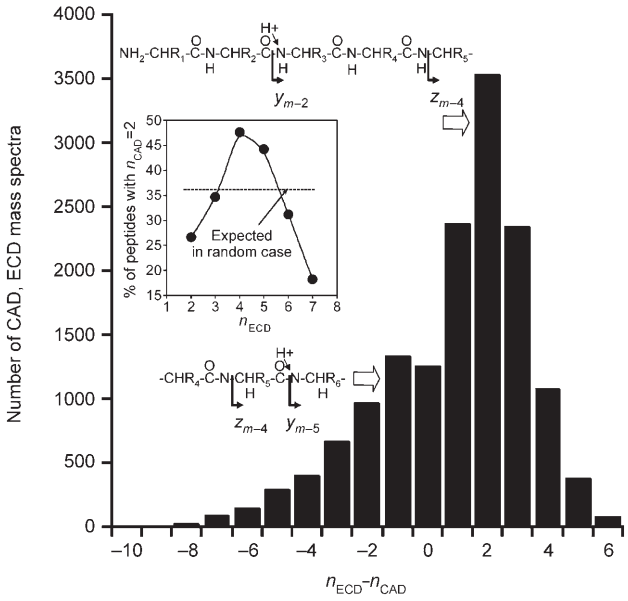

Figure 2. Distribution of the differences $\Delta = n_{ECD} - n_{CAD}$ in the position of the maximum-abundance C-terminal fragment arising from ECD and CAD of the same tryptic peptide dication. Inset: % of peptides with $n_{CAD}=2$ for different n_{ECD}; the value of 36% is expected in a random case. The plot confirms that $n_{ECD}=4$ and $n_{CAD}=2$ are statistically linked.

action ECD mechanism. The preferences for amino acid pairs in ECD and CAD for the same peptide sequences (cleavage positions $n=3$–5) are shown in Figure 3. The abundances of the z_{m-3}, z_{m-4}, and z_{m-5} ion were normalized for each spectrum with respect to the most abundant of these three ions, the same was done for each triplet $(y_{m-3}, y_{m-4}, y_{m-5})$. The CAD and ECD preferences are complementary, with a near-zero correlation of $r=-0.07$. While CAD shows only a small difference between Ile and Leu, cleavages on the N-terminal side of Leu by ECD are much more abundant than for Ile. Furthermore, A–B cleavages in CAD anticorrelate with B–A cleavages, while this feature is much weaker with ECD. Other features of Figure 3 will be discussed in detail in subsequent reports.

The correlated maxima at $n_{ECD}=4$ and $n_{CAD}=2$ in Figure 1 require explanation. The most likely reason for the correlation is the presence of an N-terminal structure that solvates the charge, as in the simulations of Ala_nH^+ ions by Samuelson and Martyna.[13] Such a structure can be further stabilized by neutral hydrogen-bonding interactions. According to the recently suggested nonlocal mechanism,[11] electron capture on the backbone nitrogen atom involved in neutral hydrogen-bonding interactions with the carbonyl group of the fourth residue will induce H· transfer to that carbonyl group. The thus created aminoketyl radical will rapidly relax through cleavage of the N–C_α bond next to the carbonyl group[10,11] to

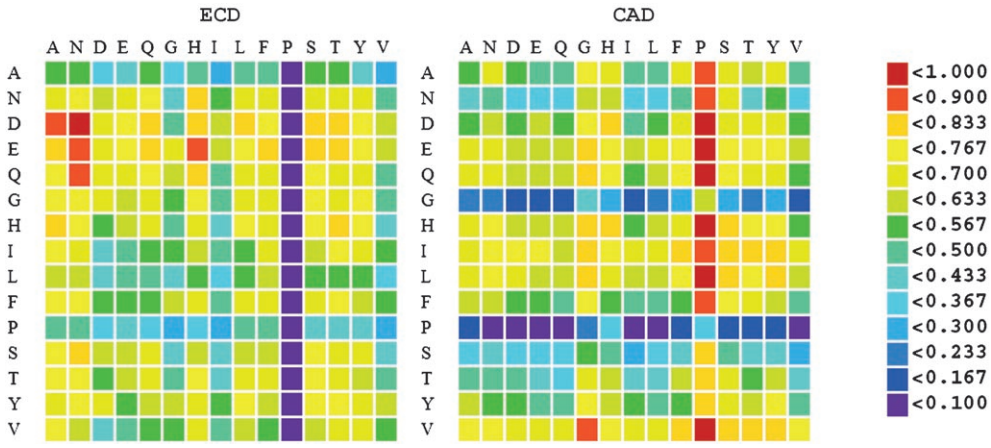

Figure 3. Amino acid pair preferences in ECD (left panel) and CAD (right panel) for C-terminal $(m-n)$th fragments, $n=3.5$. Columns correspond to residue B in an amino acid pair A–B.

give neutral N-terminal c_4 fragments and singly charged z_{m-4} ions, thus accounting for the maximum at $n \approx 4$ in Figure 1 b. Vibrational activation of the same structure will result in the solvated proton preferentially attacking the adjacent backbone amide, thus effecting C–N bond cleavage[14] to form y_{m-n} and b_n ions. The maximum peak at $n_{CAD} = 2$ in Figure 1 a hints at frequent charge solvation on the second carbonyl group. The enhanced stability of the b_2 ion[14] may additionally increase the cleavage rate for $n_{CAD} = 2$. Vibrational excitation can also unfold the structure and induce proton migration along the backbone.[15] The migrating proton will lead to heterogeneous population of unfolded structures, as suggested by Gaskell and co-workers,[15] which will result in a broad distribution of the C–N cleavage sites (broad maximum in Figure 1 a). On the basis of the fraction of peptide sequences that produced y_{m-2} ions by CAD and z_{m-4} by ECD (inset in Figure 2), we estimate that 30 to 40 % of all doubly charged tryptic peptides form the N-terminal charge-solvating structure.

Tables similar to those in Figure 3, but created separately for $n = 3$, 4, and 5 (with the intensities still normalized to the most abundant ion in the triplet), vary with n much less for CAD than for ECD. For example, correlation between the CAD tables for $n = 3$ and $n = 4$ is $r_{3,4} = 0.88$, for $n = 4$ and $n = 5$ it is $r_{4,5} = 0.85$, and for $n = 3$ and $n = 5$ it is $r_{3,5} = 0.75$. The corresponding figures for ECD are $r_{3,4} = 0.75$, $r_{4,5} = 0.62$, and $r_{3,5} = 0.47$, that is, sequence preferences for forming z_3 differ markedly from those for z_5. This result is consistent with the presence of a nonrandom secondary N-terminal structure, which is known to affect rates of cleavage by ECD.[11] Since this structure is charge-stabilized, the presence of a strongly basic residue at the N terminus should destabilize it, as the proton will be sequestered by the basic side chain. Consistent with that prediction, peptides with an N-terminal arginine (residue formed in trypsinolysis as a result of a missed cleavage) gave a maximum peak at $n = 1$ instead of $n \approx 4$ in the ECD spectra, and the sharp maximum peak at $n = 2$ in the CAD spectra disappeared.

It was also studied how the proton affinity PA_k of the kth amino acid from the N terminus affected the position n of the most abundant ECD fragment z_{m-n}. A clear anticorrelation was found between PA_k and n, which means that the more basic N terminus sequestered the proton to a higher degree and led to a smaller size of the N-terminal compact structure. The correlation between PA_k and n was equally strong for all cases with $k < 5$, but dropped twofold at $k = 5$ and disappeared at $k > 5$. It was concluded that the N-terminal structure involves four or five residues, in agreement with the size of the complementary part to the most abundant ECD fragment (Figure 1 b). The effect of the peptide size on n was only noticeable for $m = 7$, thus indicating that for larger peptides the N-terminal structure is independent of the C-terminal charge.

In summary, this study proves for the first time that preferences for the cleavage site and amino acid context in ECD and CAD are truly complementary. This finding supports nonlocal ECD mechanisms that decouple the proton location and the position of the N–C_α bond that is cleaved. Thus, broader use of ECD in proteomics as a complement to CAD is warranted. The strong indication that an N-terminal charge-solvating structure is a common feature in tryptic peptide dications encourages further investigation, for example, using molecular dynamics simulations.

Experimental Section

Data for Figures 1–3 were collected from the proteomics analysis of lysates of human cell lines and *E. coli* (all samples from Sigma) performed on an LTQ FT mass spectrometer (Thermo) using consecutive ECD and CAD fragmentation of peptides eluting from the analytical column of a nano-LC system (Agilent 1100). The peptides were identified by the Mascot search engine (Matrix Sciences). The details of the experimental procedure and data processing are described in Ref. 12.

Received: March 30, 2006
Revised: June 5, 2006
Published online: July 17, 2006

Keywords: collision-activated dissociation · electron-capture dissociation · ion–electron reactions · mass spectrometry · peptide fragmentation

[1] R. A. Zubarev, N. L. Kelleher, F. W. McLafferty, *J. Am. Chem. Soc.* **1998**, *120*, 3265.
[2] W. Yu, J. E. Vath, M. C. Huberty, S. A. Martin, *Anal. Chem.* **1993**, *65*, 3015.
[3] Y. Y. Huang, J. M. Triscari, G. C. Tseng, L. Pasa-Tolic, M. S. Lipton, R. D. Smith, V. H. Wysocki, *Anal. Chem.* **2005**, *77*, 5800.
[4] N. A. Kruger, R. A. Zubarev, B. K. Carpenter, N. L. Kelleher, D. M. Horn, F. W. McLafferty, *Int. J. Mass Spectrom.* **1999**, *183*, 1.
[5] A. R. Dongre, A. Somogyi, V. H. Wysocki, *J. Mass Spectrom.* **1996**, *31*, 339.
[6] A. Somogyi, V. H. Wysocki, I. Mayer, *J. Am. Soc. Mass Spectrom.* **1994**, *5*, 704.
[7] R. A. Zubarev, K. F. Haselmann, B. Budnik, F. Kjeldsen, F. Jensen, *Eur. J. Mass Spectrom.* **2002**, *8*, 337.
[8] K. Breuker, H. B. Oh, C. Lin, B. K. Carpenter, F. W. McLafferty, *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 14011.
[9] M. Sobczyk, W. Anusiewicz, J. Berdys-Kochanska, A. Sawicka, P. Skurski, J. Simons, *J. Phys. Chem. A* **2005**, *109*, 250.
[10] E. A. Syrstad, F. Turecek, *J. Am. Soc. Mass Spectrom.* **2005**, *16*, 208.
[11] A. Patriksson, C. Adams, F. Kjeldsen, J. Raber, D. van der Spoel, R. A. Zubarev, *Int. J. Mass Spectrom.* **2005**, *248*, 124–135.
[12] M. L. Nielsen, M. M. Savitski, R. A. Zubarev, *Mol. Cell. Proteomics* **2005**, *4*, 835.
[13] S. Samuelson, G. J. Martyna, *J. Phys. Chem. B* **1999**, *103*, 1752.
[14] T. Yalcin, C. Khouw, I. G. Csizmadia, M. R. Peterson, A. G. Harrison, *J. Am. Soc. Mass Spectrom.* **1995**, *6*, 1165.
[15] O. Burlet, C. Y. Yang, S. J. Gaskell, *J. Am. Soc. Mass Spectrom.* **1992**, *3*, 337.

