J|A|C|S
COMMUNICATIONS

Published on Web 07/26/2006

Scrambling of Sequence Information in Collision-Induced Dissociation of Peptides

Alex G. Harrison,*,† Alex B. Young,† Christian Bleiholder,‡ Sándor Suhai,‡ and Béla Paizs*,‡

Department of Chemistry, University of Toronto, 80 St. George Street, Toronto, Ontario M5S 3H6, Canada, and German Cancer Research Center, Im Neuenheimer Feld 580, D-69120 Heidelberg, Germany

Received April 9, 2006; E-mail: b.paizs@dkfz.de; aharriso@chem.utoronto.ca

Protein identification in proteomics is mostly achieved by tandem mass spectrometry (MS/MS) analysis of peptides produced by enzymatic digestion.[1] In the MS/MS experiments, protonated peptides undergo collision-induced dissociation (CID), and the fragment ion spectrum is used to assign their amino acid sequence. MS/MS-based peptide sequencing strongly relies on various bioinformatics tools, which relate experimental mass spectra to predicted spectra of peptides derived from protein and/or DNA database entries. These bioinformatics tools are based on more or less refined fragmentation models that implement our current understanding of gas-phase peptide chemistry (reviewed recently[2a]). While automated peptide sequencing algorithms have advanced considerably, only a small fraction[1c,d] of acquired MS/MS spectra are assigned unambiguously and further developments are needed to serve the proteomics community with robust protein identification tools.

The peptide MS/MS spectra usually contain a large number of fragment ions from which only a fraction are used for sequencing. If CID is performed in the typical low-energy regime, the latter include *b*, *a*, and *y* fragments[3] and their satellites arising from further backbone fragmentations (internal and immonium ions) or loss of small neutrals (mostly H_2O and NH_3). The structure and mass-to-charge ratio of these ions can directly be derived from the primary structure of the investigated peptides if typical fragmentation is assumed. We will refer to these series as *direct sequence ions*. It is often the case[4] that abundant fragment ions appear in the MS/MS spectra of peptides that do not belong to the direct sequence ion series. These fragments are formed in complex rearrangements and will be termed *nondirect sequence ions* in the following. (Complex rearrangements can also lead to fragments that are formally direct sequence ions.) Current automated peptide sequencing tools consider even nondirect sequence ions as *b*, *y*, or *a* fragments, and this can easily lead to erroneous peptide and protein identification. (For recent editorials on false-positive protein identification, see ref 5.) In the following, we will show that nondirect sequence ions can be formed from linear *b* ions with C-termini oxazolone rings (noted as linear *b* ions) by cyclization and subsequent ring opening and further dissociation to smaller fragments.

We have studied the CID of protonated YAGFL-NH_2 using experimental and modeling techniques. CID of protonated YAGFL-NH_2 (Bachem Biosciences, King of Prussia, PA) was achieved on an electrospray/quadrupole/time-of-flight (QqToF) mass spectrometer (Qstar, MDS Sciex, Concord, Canada). Breakdown graphs of the b_5 ion of YAGFL-NH_2 and cyclo-(YAGFL) (Celtek Peptides, Nashville, TN) were obtained by varying the collision energy in the quadrupole cell with b_5 produced in-source. The potential energy surface of protonated YAGFL-NH_2, including transition structures (TSs) of peptide fragmentation pathways (PFPs), was studied by a conformational search engine developed recently.[2a−c] The final energetics were obtained using total energies at the B3LYP/6-31+G(d,p) level corrected for zero-point energies determined at B3LYP/6-31G(d) computed with the Gaussian program.[6] We have chosen YAGFL-NH_2 because its CID generates abundant *b* and *a* ions and its size makes modeling studies feasible at reliable theoretical levels.

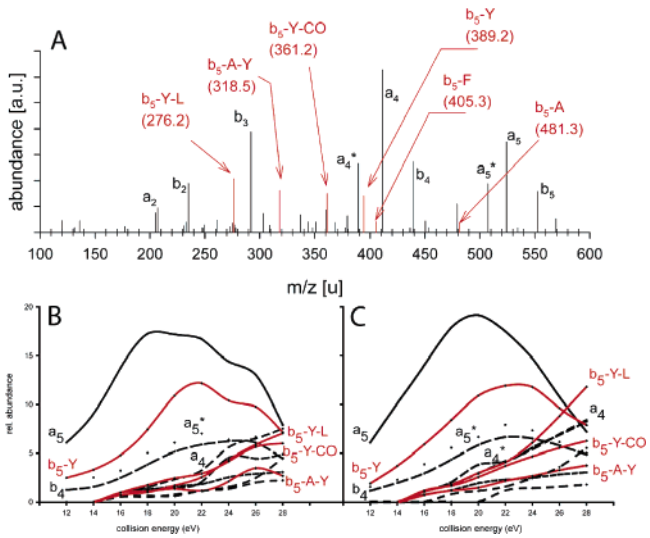


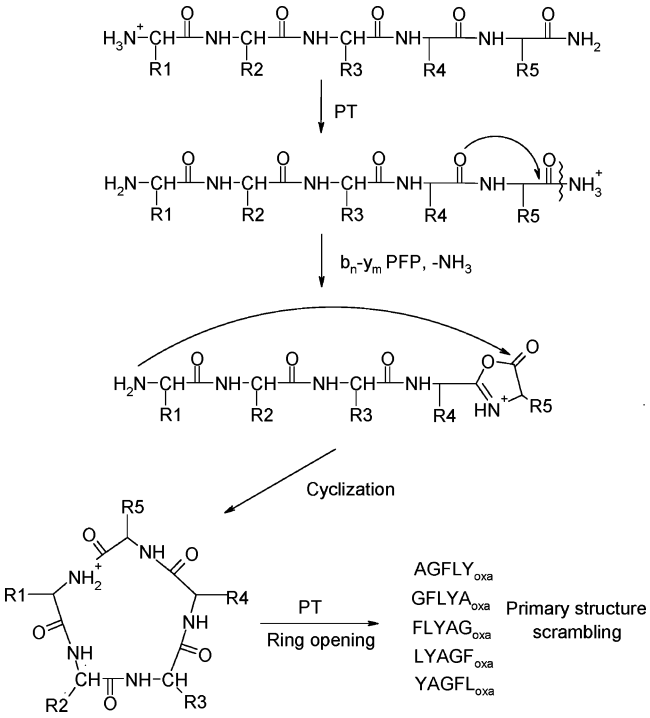
Figure 1. (A) CID mass spectrum (30 eV lab frame collision energy) of protonated YAGFL-NH_2. (B) Breakdown graph of b_5 ion of YAGFL-NH_2. (C) Breakdown graph of cyclo-(YAGFL). Nondirect sequence ion abundances are shown in red.

Protonated YAGFL-NH_2 dissociates (Figure 1A) to produce b_5, b_4, b_3, b_2, a_5, a_5*, a_4, a_4*, and a_2 ions as well as abundant fragments at *m*/*z* 389.2, 361.2, 318.5, and 276.2. The latter can be assigned as b_5-Y, b_5-Y-CO, b_5-A-Y, and b_5-Y-L, respectively. This assignment is supported by the breakdown graph of protonated YAGFL-NH_2 (depicted in Figure S1 in the Supporting Information) that shows that the intact peptide fragments nearly exclusively to b_5 that dissociates further on the a_5 (a_5*) → b_4 → a_4 (a_5*) etc. cascade and to form ions b_5-Y, b_5-Y-CO, b_5-A-Y, and b_5-Y-L. While the latter fragments could be derived from the usual linear b_5 ion ($YAGFL_{oxa}$) as a result of "typical" fragmentation, the corresponding PFPs involve TSs with high barriers[2] (Scheme S1 in the Supporting Information). On the other hand, formation of these ions can easily be explained if one assumes that b_5 is indeed a cyclic peptide or the linear b_5 ion rearranges to a cyclic peptide before further fragmentation. To challenge this "complex rearrangement" hypothesis, the breakdown graphs of b_5 of YAGFL-NH_2 and protonated cyclo-(YAGFL) were obtained. As shown in Figure 1B,C, the two breakdown graphs are rather similar, lending strong support to the previous speculations. Furthermore, the CID spectrum (Figure 1A) of protonated YAGFL-NH_2 displays less abundant ions at *m*/*z* 405.3 and 481.3 assigned to b_5-F and b_5-A, respectively. These ions cannot be formed from $YAGFL_{oxa}$, while they can easily be derived from the cyclic b_5 isomer.

† University of Toronto.
‡ German Cancer Research Center.

10.1021/ja062440h CCC: $33.50

Scheme 1



The b_5 ion of protonated YAGFL-NH_2 can be formed on the cyclic peptide and b_n-y_m PFPs[2] by loss of ammonia from the parent peptide. For the global minimum of protonated YAGFL-NH_2 (Scheme 1, Figure S2 in the Supporting Information), the extra proton resides on the N-terminal amino group. Both ammonia-loss PFPs assume mobilization of the extra proton to form C-terminal amide N-protonated species (relative energy (E_{rel}) of 18.6 kcal/mol). On the cyclic peptide PFP, the N-terminal amino group is proposed to attack the carbon center of the N-protonated amide bond. Contrary to the many attempts made, no low-energy cyclic peptide TS has been found. The b_n-y_m PFP is initiated by nucleophilic attack (Scheme 1) of the O4 amide oxygen on the C-terminal amide carbon and leads to the linear b_5 ion with an oxazolone ring at its terminus ($YAGFL_{oxa}$). (It should be noted here that the b_n-y_m PFP is both energetically and entropically favored against the cyclic peptide one.[2a,c]) We have located a large number of b_n-y_m TSs for which the N-terminal amino group and the leaving NH_3 interact through the $H_2N\cdots H-NH_2^+$ H-bond, bringing the NH_2 group close to the forming C-terminal oxazolone ring. It is worth noting here that the energetically most favored such TS (E_{rel} = 33.7 kcal/mol, Figure S3 in the Supporting Information) is structurally similar to the global minimum, and the former is connected to the latter by only a few proton transfer steps and internal rotations.

After removing ammonia, a linear b_5 isomer with interacting N- and C-termini was located (E_{rel} = 36.5 kcal/mol, Figure S4 in the Supporting Information). Due to its partial positive charge, the carbonyl C of the oxazolone ring is a likely target of nucleophilic attack by the nearby N-terminal amino group. This attack leads to opening of the oxazolone ring, and formation of a cyclic peptide b_5 isomer (the corresponding TS (E_{rel} = 48.5 kcal/mol) is depicted in Figure S5 in the Supporting Information). The barrier to this reaction is only 12 kcal/mol, while the barrier to initiate the $a_5 \rightarrow b_4 \rightarrow a_4$ etc. cascade (CO loss from $YAGFL_{oxa}$ on the $b_5 \rightarrow a_5$ PFP) is approximately 30−35 kcal/mol.[2,7] The latter is an entropically favored "direct bond cleavage"-type reaction, while the former is a rearrangement that can be hindered by entropy effects.

These computational results clearly suggest that, instead of forming cyclic b_5 from the parent peptide directly, a usual linear oxazolone-type isomer ($YAGFL_{oxa}$) is formed. This linear b_5 ion can then further isomerize if appropriate nucleophiles are available to attack on the carbonyl carbon of the charged oxazolone ring. For protonated YAGFL-NH_2, this isomerization is especially favored, leading to nearly exclusive cyclic b_5 formation as suggested by comparison of the breakdown graphs of b_5 of YAGFL-NH_2 and protonated cyclo-(YAGFL) in Figure 1. The cyclic isomer can undergo various proton transfer reactions, and the macro ring can open up at various amide bonds.[2a] Because the cyclization and ring opening reactions occur not necessarily at the same amide bond, the underlying chemistry can lead to linear b_5 ions that have other than the original $YAGFL_{oxa}$ sequence (Scheme 1). For example, ring opening at the A−Y amide bond leads to the $AGFLY_{oxa}$ b_5 isomer that can fragment further on the $b_5 \rightarrow a_5$ PFP and by losing Y to form b_5-Y. On the other hand, ring opening at the A−G amide bond leads to the $GFLYA_{oxa}$ b_5 isomer that can explain formation of the b_5-A-Y fragment in a similar way. A common characteristic of the related PFPs is that they can lead to nondirect sequence ions and scrambling of sequence information in CID of peptides.

It is worth noting here that the ion population at *m*/*z* 552.3 noted as b_5 so far can contain six different isomers (five linear and one cyclic). Each linear isomer can fragment further by losing CO, and therefore, the a_5 ion population can contain again various isomers. Such processes can result in a very large number of fragment ions with mostly small abundance. The fact that the MS/MS spectrum of protonated YAGFL-NH_2 does not show such behavior suggests that not all cyclization−reopening PFPs are active. The energetic and kinetic details of *b* ion cyclization−reopening and other PFPs leading to nondirect sequence ions will be presented in a forthcoming paper.

Acknowledgment. This work was financially supported by the NSERC (Canada, A.G.H.) and the DFG (S.S., SU 244/3-1). C.B. was supported by a fellowship of the DKFZ International PhD-Program.

Supporting Information Available: Structures and total energies of the species presented in the text, breakdown graph of protonated YAGFL-NH_2, and complete ref 6. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) (a) Hunt, D. F.; Yates, J. R., III; Shabanowitz, J.; Winston, S.; Hauer, C. R. *Proc. Natl. Acad. Sci. U.S.A.* **1986**, *83*, 6233. (b) Biemann, K.; Scoble, H. A. *Science* **1987**, *237*, 992. (c) Aebersold, R.; Goodlett, D. R. *Chem. Rev.* **2001**, *101*, 269. (d) Steen, H.; Mann, M. *Nat. Rev. Mol. Cell Biol.* **2004**, *5*, 699.

(2) (a) Paizs, B.; Suhai, S. *Mass Spectrom. Rev.* **2005**, *24*, 508. (b) Paizs, B.; Lendvay, G.; Vékey, K.; Suhai, S. *Rapid Commun. Mass Spectrom.* **1999**, *13*, 525. (c) Paizs, B.; Suhai, S. *J. Am. Soc. Mass Spectrom.* **2004**, *15*, 103.

(3) For peptide fragment ion nomenclature, see: (a) Roepstorff, P.; Fohlmann, J. *J. Biomed. Mass Spectrom.* **1984**, *11*, 601. (b) Biemann, K *Biomed. Environ. Mass Spectrom.* **1988**, *16*, 99.

(4) (a) Tang, X.-J.; Thibault, P.; Boyd, R. K. *Anal. Chem.* **1993**, *65*, 2824. (b) Tang, X.-J.; Boyd, R. K. *Rapid Commun. Mass Spectrom.* **1994**, *8*, 678. (c) Vachet, R. W.; Bishop, B. M.; Erickson, B. W.; Glish, G. L. *J. Am. Chem. Soc.* **1997**, *119*, 5481. (d) Yagüe, J.; Paradela, A.; Ramos, M.; Ogueta, S.; Marina, A.; Barahona, F.; López de Castro, J. A.; Vázquez, J. *Anal. Chem.* **2003**, *75*, 1524. (e) Olsen, J. V.; Mann, M. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 13417.

(5) (a) Taylor, G. K.; Goodlett, D. R. *Rapid Commun. Mass Spectrom.* **2005**, *19*, 3420. (b) http://www.mcponline.org.

(6) Frisch, M. J.; et al. *Gaussian 03*; Gaussian Inc.: Pittsburgh, PA, 2003.

(7) (a) Yalcin, T.; Khouw, C.; Csizmadia, I. G.; Peterson, M. R.; Harrison, A. G. *J. Am. Soc. Mass Spectrom.* **1995**, *6*, 1165. (b) Yalcin, T.; Csizmadia, I. G.; Peterson, M. B.; Harrison, A. G. *J. Am. Soc. Mass Spectrom.* **1996**, *7*, 233. (c) Paizs, B.; Szlávik, Z.; Lendvay, G.; Vékey, K.; Suhai, S. *Rapid Commun. Mass Spectrom.* **2000**, *14*, 746.

JA062440H