

Can relative cleavage frequencies in peptides provide additional sequence information?

Bogdan A. Budnik^a, Michael L. Nielsen^{a,1}, Jesper V. Olsen^{a,1}, Kim F. Haselmann^a,
Patric Hörth^b, Wolfgang Haehnel^b, Roman A. Zubarev^{a,*}

^a Department of Chemistry, University of Southern Denmark, Campusvej 55, Odense DK-5230, Denmark

^b Institute of Biology II, University of Freiburg, Freiburg, Germany

Received 26 June 2001; accepted 10 October 2001

Abstract

The ambiguity of attributing N- vs. C-terminal origin to a fragment ion can pose a problem in mass spectrometric de novo peptide sequencing even when all inter-residue bonds are cleaved. This makes additional sequence information highly desirable. The question is investigated whether relative abundances of fragment ions in MS/MS are capable of providing this information. Statistical analysis of data on peptides 10–24 residues long revealed that the frequencies of N–C_α bond cleavage in electron capture dissociation (ECD) are less dependent upon the total number of basic sites, the charge states of the parent ions and the nature of the terminal group than the peptide bond fragmentation in collisionally activated dissociation (CAD). The consensus sequences in peptides of different lengths also showed similar fragmentation patterns, with the typical correlation factor of 0.7. The conclusion is that frequencies of ECD cleavages are promising as a source of additional sequence information provided that strong intra-molecular bonds are absent or broken prior to MS/MS analysis. (Int J Mass Spectrom 219 (2002) 283–294) © 2002 Elsevier Science B.V. All rights reserved.

Keywords: Peptide sequencing; Fragment ions; Tandem mass spectrometry

1. Introduction

Peptide sequencing with tandem mass spectrometry (MS/MS) is based upon obtaining a representative fragment ion population. The mass differences between the fragments with a common terminus reveal the sequence. The challenges are to achieve optimal conditions at which the MS/MS data contain the necessary information and to extract this information.

The ideal situation is when, in every precursor ion, fragmentation of a single inter-residue bond is

achieved, with all such bonds cleaved. If some bonds remain intact, the “gaps” may result in ambiguous assignment: e.g., the absence of the cleavage between two glycine residues can be interpreted as the presence of asparagine (Gly–Gly≡Asn). Even when such sequence gaps, e.g., –AB–, can be “filled” with an unique combination of two residues, A + B, their order remains unclear.

Application of complementary fragmentation techniques, such as collisionally activated dissociation (CAD) and electron capture dissociation (ECD [1]), can help to fill the sequence gaps. For example, CAD preferentially cleaves at the N-side of proline residues, while ECD is inhibited at that site. CAD produces

* Corresponding author. E-mail: rzu@chem.sdu.dk

¹ Present address: MDS Proteomics, Odense, Denmark.

peptide bond (b, y ions) fragmentation, while ECD mostly gives N–C $_{\alpha}$ bond (c, z $^{\bullet}$ ions) cleavages. The mass difference between c and b ions is 17 Da, while that between y and z $^{\bullet}$ ions is 16 Da, which helps to establish the sequence directionality by means of identifying “golden” complementary pairs [2]. However, detection of 1 Da mass difference is not always reliable, since ECD in some cases produces c $^{\bullet}$ and z fragments [3]. Ambiguity can also occur when the masses of the N- and C-terminal fragments coincide or are so close that their isotopic distributions overlap. The difference between the fragments of different directionalities may also accidentally coincide with the mass of an amino acid residue. If this happens, the sequencing algorithm may include the wrong amino acid in the sequence.

Statistical analysis shows that such unlucky coincidences are encountered rather often. Fig. 1 presents the occurrence frequencies of the two undesirable

situations as a function of the peptide length. When the nominal mass of an N-terminal fragment coincides with that of an C-terminal fragment, isotopic distributions overlap. To resolve the overlap, application of very high resolving power and sophisticated software is required [2]. The probability for this type of difficulty to arise reaches an appreciable value for peptides ca. 40 residues long. Much more frequent is the situation when the nominal mass difference between the fragments of different directionalities coincides with the mass of an amino acid residue. This type of problem occurs already in peptides shorter than 10 residues. Although high resolving power and mass accuracy can help to alleviate this problem, additional sequence information is highly desirable.

Here we investigate the question whether the fragmentation pattern (a set of relative abundances of fragment ions) can be a source of such information. If a set of formal rules existed for predicting the

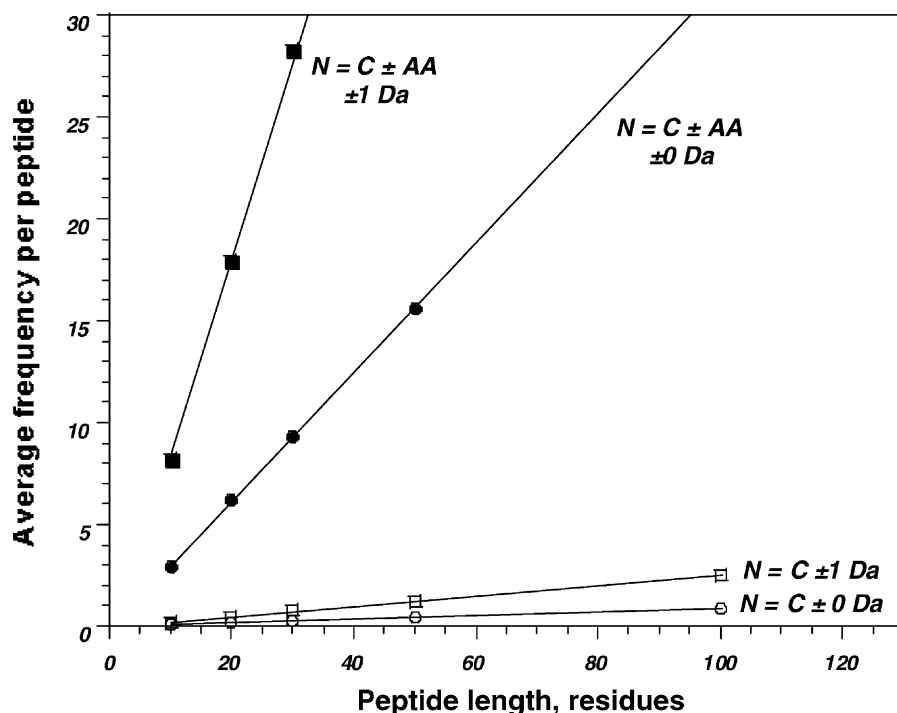


Fig. 1. The occurrence frequencies as functions of the peptide length of the cases when the mass of a N-terminal fragment (N) coincides with that of a C-terminal fragment (C), and when the mass differences between two fragments of different origins coincide with the mass of an amino acid residue (AA). Nominal (integer) masses are used.

relative abundances of fragment ions from the known sequence, several plausible sequence alternatives could be compared using as a criterion how well they account for the observed fragment pattern. For a quick prediction, the set of rules must be simple. Also, the number of operations necessary for prediction must not grow too fast with the size. These requirements are fulfilled for the “reducible” fragmentation process, in which different parts of the molecule fragment independently of each other. Then the cleavage frequency of a certain bond is determined only by groups in the vicinity of that bond. For each combination of these groups (e.g., amino acid residues) one can derive the average cleavage frequency, and then create a library of cleavage frequencies for all such combinations. The expected (predicted) fragmentation pattern would then be a linear combination of such reference frequencies. Thus, the crucial question is whether or not the peptide fragmentation is reducible.

For CAD, the anticipated answer is negative, although some cleavages do appear with higher probabilities than other. Besides the N-side of proline, preferential cleavage occurs at the C-sides of the aspartic acid and glutamic acid residues [4]. But generally speaking, the cleavage pattern is a result of a complex interplay between “global” parameters, such as the availability of the “mobile” protons [5], secondary structure [6] and the presence of certain groups, such as cysteine acid [7]. Instrumental parameters, such as the excitation level, the ionic temperature prior to excitation and the time available for dissociation also influence the outcome.

The mechanism of ECD, unlike that of CAD, is believed to be non-ergodic, i.e., the bond rupture happens prior to energy randomization over the whole molecule [1]. As a result, polypeptide N–C $_{\alpha}$ backbone bonds cleave preferentially even in the presence of labile side-chain groups [8–10]. Moreover, cleavage of a N–C $_{\alpha}$ bond can occur without dissociation of weak inter-molecular bonding [11]. In a non-ergodic process, independent fragmentation behavior of different parts of the polypeptide can be expected, with the cleavage probability for a particular N–C $_{\alpha}$ bond mostly determined by the local sequence and not the

presence or nature of distant groups. Moreover, unlike in CAD where the amount of deposited energy and the rate of collisional and infrared cooling may vary, the kinetic energy of the slow electrons does not play an important role in ECD because of its insignificant value (<0.2 eV) [1,3], and the rate of the collisional and infrared de-excitations is negligible compared to the high cleavage rate ($>10^{12}$ s $^{-1}$ [1]). The recombination energy in ECD has therefore, a fixed value determined by the nature and the charge state of the molecular ion. Even the internal energy of the ions before electron capture is not that important: McLafferty and co-workers have not found significant changes in the ECD cleavage frequencies at the ambient temperature between 20 and 130 °C [12]. Hence ECD a priori has better chances to produce reducible fragmentation than CAD.

The global tendencies of bond cleavage in peptides and proteins have been studied [13]. The ECD cleavage frequencies have been found to be more homogeneous than those in CAD. At the same time, the C-side of tryptophan residue has been found to cleave in ECD preferentially and the N-terminal side of cysteine rarely cleaved [13]. The current study attempts a more detailed approach by investigating series of polypeptides with similar sequences. The parameters known to affect the CAD pattern, such as the charge state and the number of basic groups, are varied; the resulting mass spectra are compared with each other and the corresponding CAD spectra in terms of the relative cleavage frequencies.

2. Experimental

2.1. Mass spectrometry

Experiments were performed on a commercial 4.7 Tesla Ultima Fourier transform (FT) mass spectrometer (IonSpec, Irvine, CA), with a standard ESI hexapole-based interface (Analytica of Branford, MA) additionally equipped with a home-built nano-electrospray ion source. The ions were externally accumulated in the hexapole for 100–500 ms,

and then transported to the cell where they were captured in a gated mode without a cooling gas. A 2-s accumulation resulted in abundant in-hexapole fragmentation [14]; doubly-charged b and y ions were produced this way for a subsequent ECD experiment [11]. The parent ions were selected by on-resonance ejection of undesired ions and irradiated for 5–10 s with subthermal electrons produced by a standard EI filament installed on the backside of the FT cell and biased to +0.85 V. At the end of the irradiation event, 2 ms pulse of a cooling gas (nitrogen or argon) was applied to remove the excess of kinetic and internal energy from the fragment ions. The resulting spectra were processed manually. The relative cleavage frequencies were determined as sums of the abundances of all isotopic peaks of the complementary fragments ($c+z^{\bullet}$ or $b+y$). To measure the degree of spectra similarity quantitatively, we applied the product-moment correlation coefficient R given by [15]:

$$R = \frac{\sum_i \{(x_i - \bar{x})(y_i - \bar{y})\}}{[\sum_i (x_i - \bar{x})^2 \sum_i (y_i - \bar{y})^2]^{1/2}},$$

where x and y are two fragmentation patterns.

2.2. Sample preparation

Samples of synthetic peptides were obtained from Sigma or synthesized in-house. Typically, 1–2 μL of 10^{-5} M solution in 49:49:2 (v/v/v) methanol–water–acetic acid mixture was loaded into a nano-electrospray pulled glass capillary (MDS Proteomics, Odense, Denmark) and used as a source of ions lasting for over 60 min.

3. Results and discussion

3.1. Strong intra-molecular hydrogen bonding

For species >8 kDa, McLafferty and co-workers have shown that intra-molecular hydrogen bonding may prevent fragment separation in ECD although the N–C $_{\alpha}$ backbone cleavage seems to occur [3,16]. Moreover, participation of the backbone carbonyl

in hydrogen bonding may prevent hydrogen atom capture and subsequent fragmentation. This of course affects the relative abundances of fragments from the inner regions of the proteins. A similar effect is expected for shorter peptides that possess unusually strong secondary structure, such as, e.g., peptides containing a stretch of 18 alanine residues [17]. The 2+ ions of such a peptide that presumably preserved in the gas phase helical structure gave in ECD varying relative abundances of c ions (Olsen et al., unpublished results). In contrast, CAD of the same species produced much more homogeneous b, y cleavages, due to dissociation of hydrogen bonds prior to backbone fragmentation. To avoid the effect of strong intra-molecular bonding, we limited our study to <3 kDa molecules without repeating amino acid sequences. The results should be applicable to other molecules as well upon fragmentation of the intra-molecular bonding, e.g., by the activated ion technique [16].

3.2. CAD vs. ECD of substance P

In order to study the effect of subtle structure differences on the fragmentation pattern, CAD and ECD spectra of substance P with acidic and amidic C-termini were compared. Fig. 2 shows CAD spectra of MALDI-produced 1+ ions in both forms. While the spectrum of the molecule with the acidic terminus was dominated by the y_9^+ fragment, the other showed an abundant ammonia loss from both the molecular ion and the most abundant fragment, y_{10}^+ . This difference can be explained by the interaction between the basic lysine residue and the acidic C-terminus, and the absence of such interaction with the amidated C-terminus.

As demonstrated by Fig. 3 where CAD spectra of ESI-produced 2+ ions are presented, the cleavage frequencies become more similar at higher charging. Yet the differences in the relative abundances of b_{10}^{2+} , y_9^+ , and especially y_7^+ ions are quite significant. In contrast, the ECD spectra of 2+ ions (Fig. 4) differ mainly in the relative abundances of the C-terminal products, the c_{10}^+ fragment and the reduced

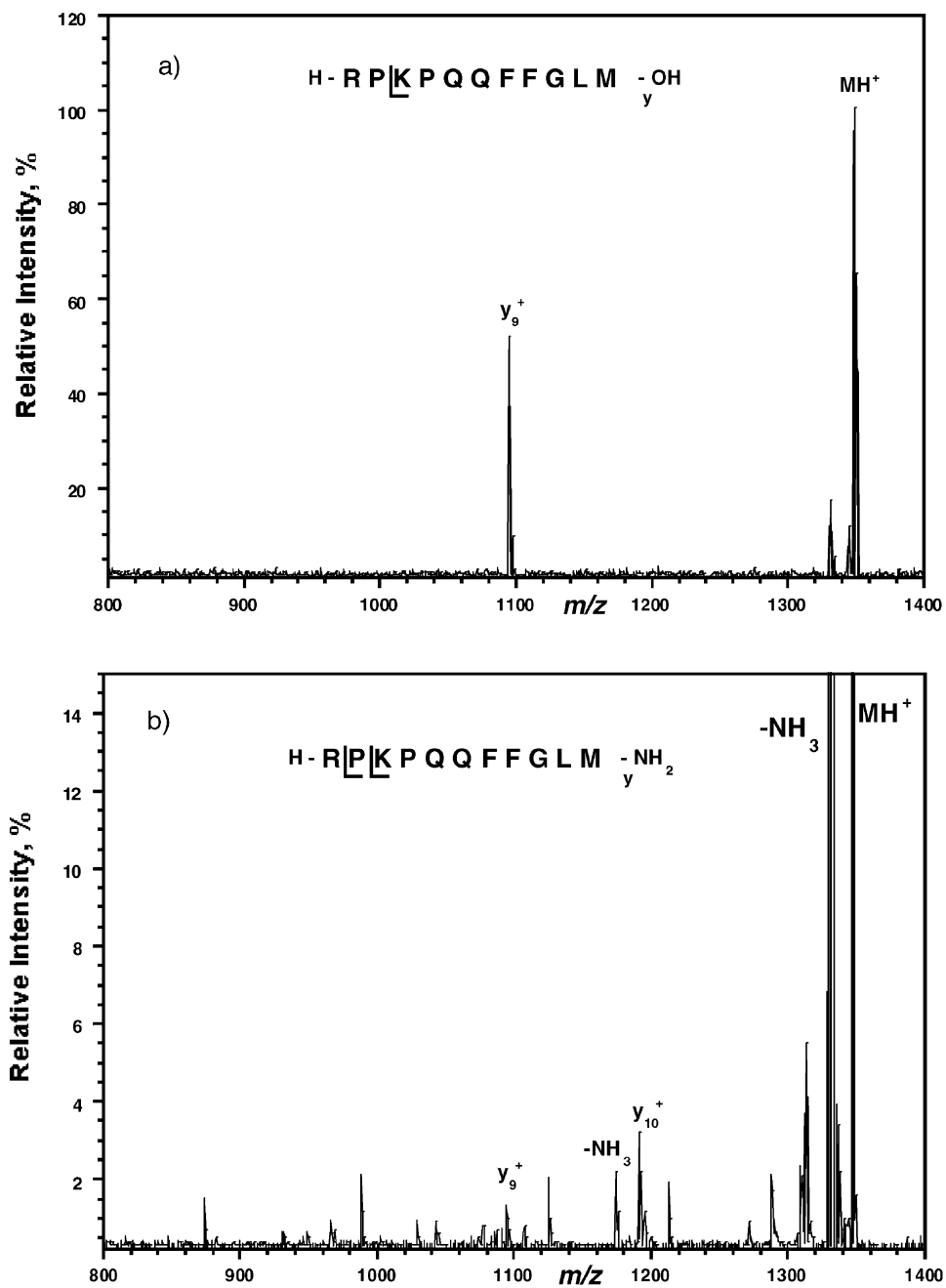


Fig. 2. CAD mass spectra of MALDI-produced 1+ ions of substance P in the acidic (a) and amidated (b) forms.

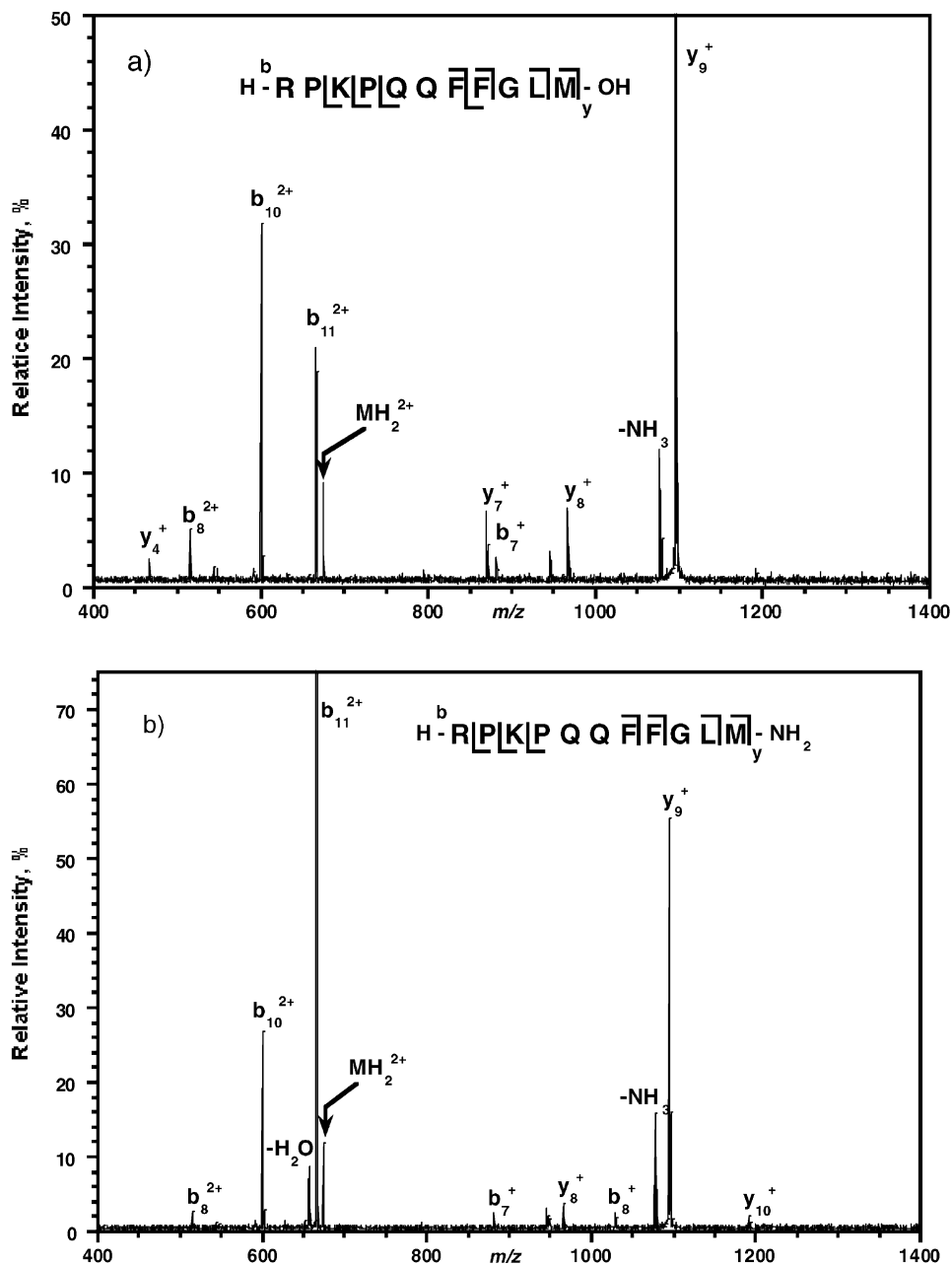


Fig. 3. CAD mass spectra of ESI-produced $2+$ ions of substance P in the acidic (a) and amidated (b) form.

$[M + 2H]^{2+}$ ion. This is despite the fact that these two spectra were obtained by different operators and experimental set-ups, which resulted into very different signal-to-noise ratios. The similar fragmentation

patterns and the presence of dissimilarity near the place of chemical divergence are consistent with the local character of ECD. CAD spectra appear to be reducible to a significantly lesser extent.

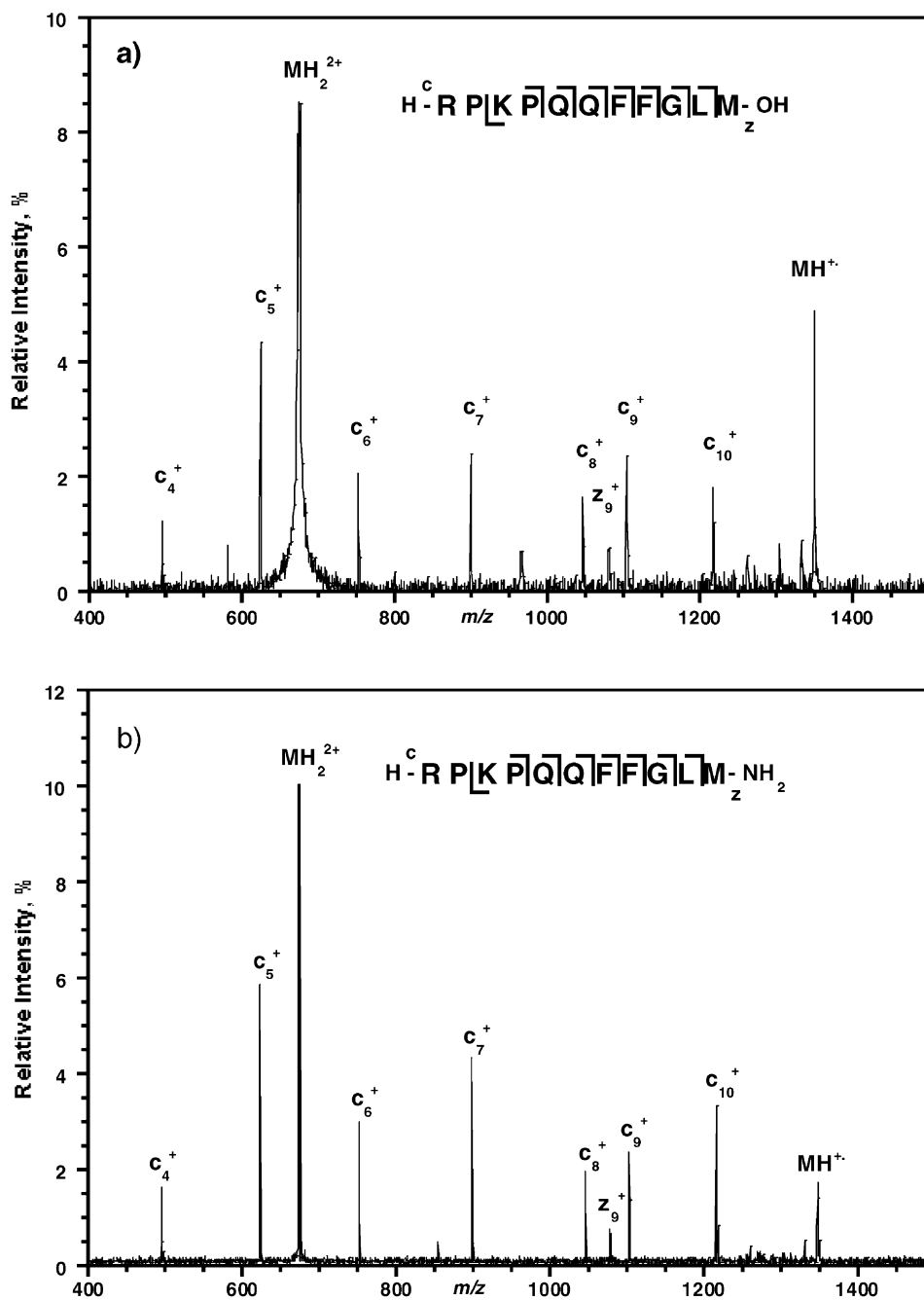


Fig. 4. ECD mass spectra of ESI-produced 2+ ions of substance P in the acidic (a) and amidated (b) form.

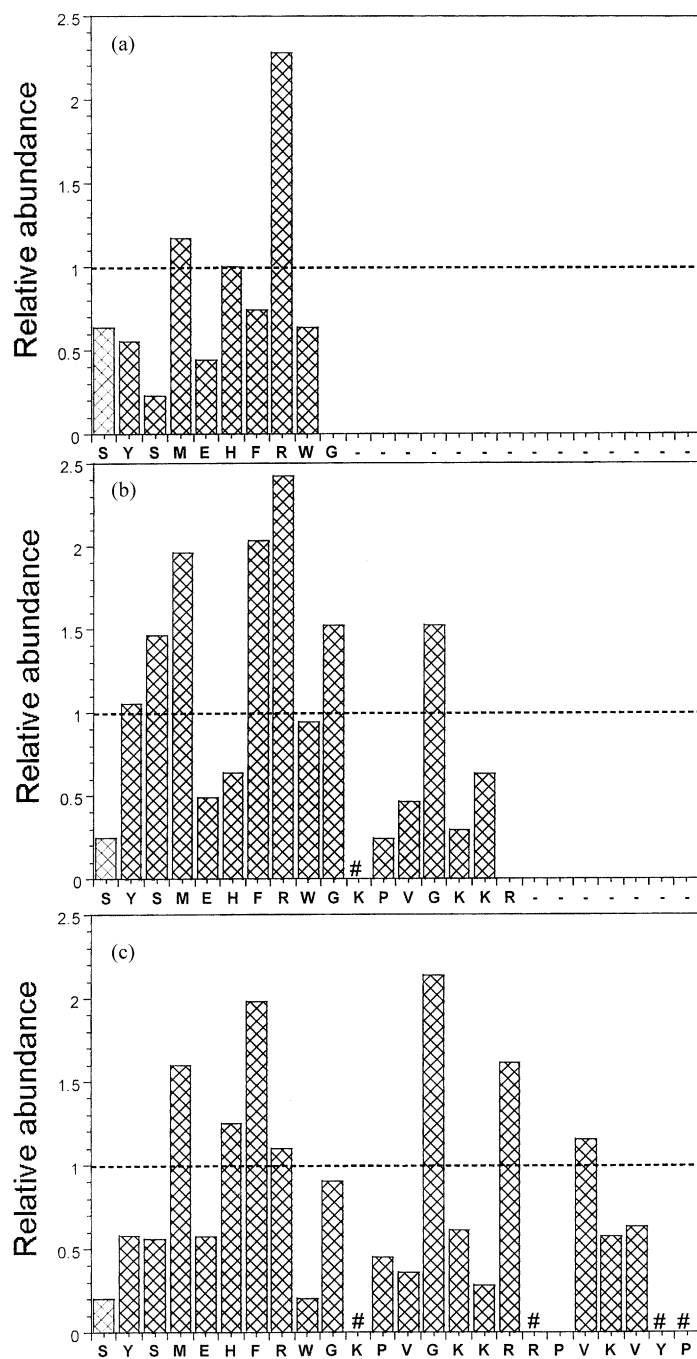


Fig. 5. ECD cleavage frequencies in the molecular ions of ACTH peptides: (a) 2+ of 1–10, (b) 3+ of 1–17 and (c) 5+ of 1–24.

3.3. Repeatability

The repeatability of measuring R was determined from the comparison of two ECD spectra taken with an interval of 3 weeks using the same experimental parameters. The obtained average correlation coefficient for ACTH 1–17, 4+ and ACTH 1–24, 5+ was $R = 0.9$. The value in CAD was similar. Due to the good repeatability, there was no need in distinguishing between the c and y ions and the less stable $z^{\bullet-}$ and b-fragments [1,9].

3.4. Multiple electron capture

The largest product peaks in ECD mass spectra are usually due to the species in the decremented charge state but similar mass as the precursor molecular ions [1]. Subsequent electron capture by these reduced species may produce fragments [3]. To test how this affects the correlation coefficient, we compared conventional ECD spectra with those obtained with sustained on-resonance ejection of the reduced species off the cell, as has been suggested by McLafferty and

co-workers [3]. The obtained correlation coefficients, $R = 0.9$, revealed that multiple electron capture is a minor process, at least in the mass range relevant to the current study. However, to be on the safe side, we used the ejection of the reduced species in the subsequent experiments.

3.5. Charge state effect

The recombination energy RE in ECD is given by [18]:

$$RE = 13.6 \text{ eV} - PA[M_{n-1}] + HA[M_{n-1}],$$

where $PA[M_{n-1}]$ and $HA[M_{n-1}]$ are the proton affinity and the hydrogen atom affinity of the $[M + (n - 1)H]^{(n-1)+}$ ions, respectively. While the latter value is ca. 0.6 eV for H^{\bullet} capture by the carbonyl oxygen (Jensen and Zubarev, unpublished results), the former depends upon the charge state ($n - 1$) and generally decreases by ca. 1 eV per charge for peptide in the 1–2 kDa range [19]. Because of the higher exothermicity of the electron capture, the fragmentation pattern of the higher charge states may

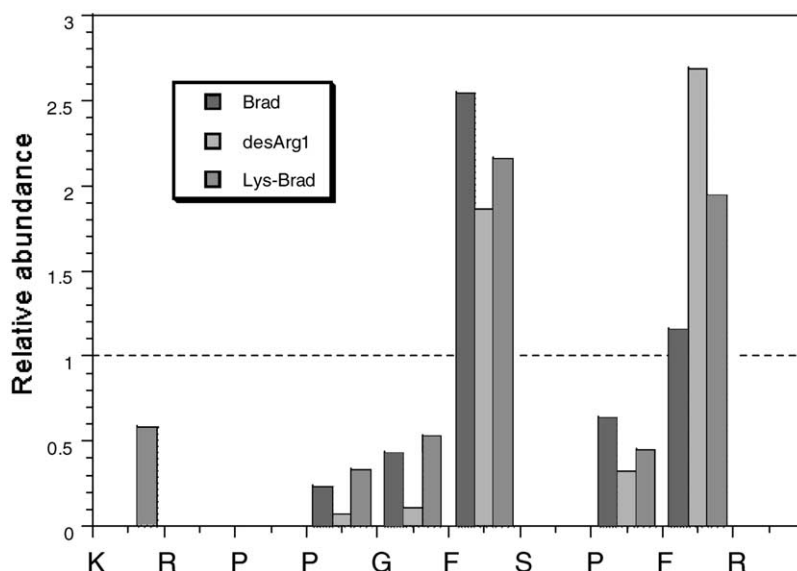


Fig. 6. ECD cleavage frequencies in 2+ ions of bradykinin, des-Arg₉-bradykinin and Lys-bradykinin.

be expected to differ from that of the lower charge states. Besides, the secondary structure of ions is also charge-dependent. While the first proton added to a linear molecule can stabilize its conformation via

intra-molecular proton bonding, additional protons eventually destabilize it due to the growing Coulombic repulsion. However, the average R value for different charge states of the same molecules (ACTH 1–10,

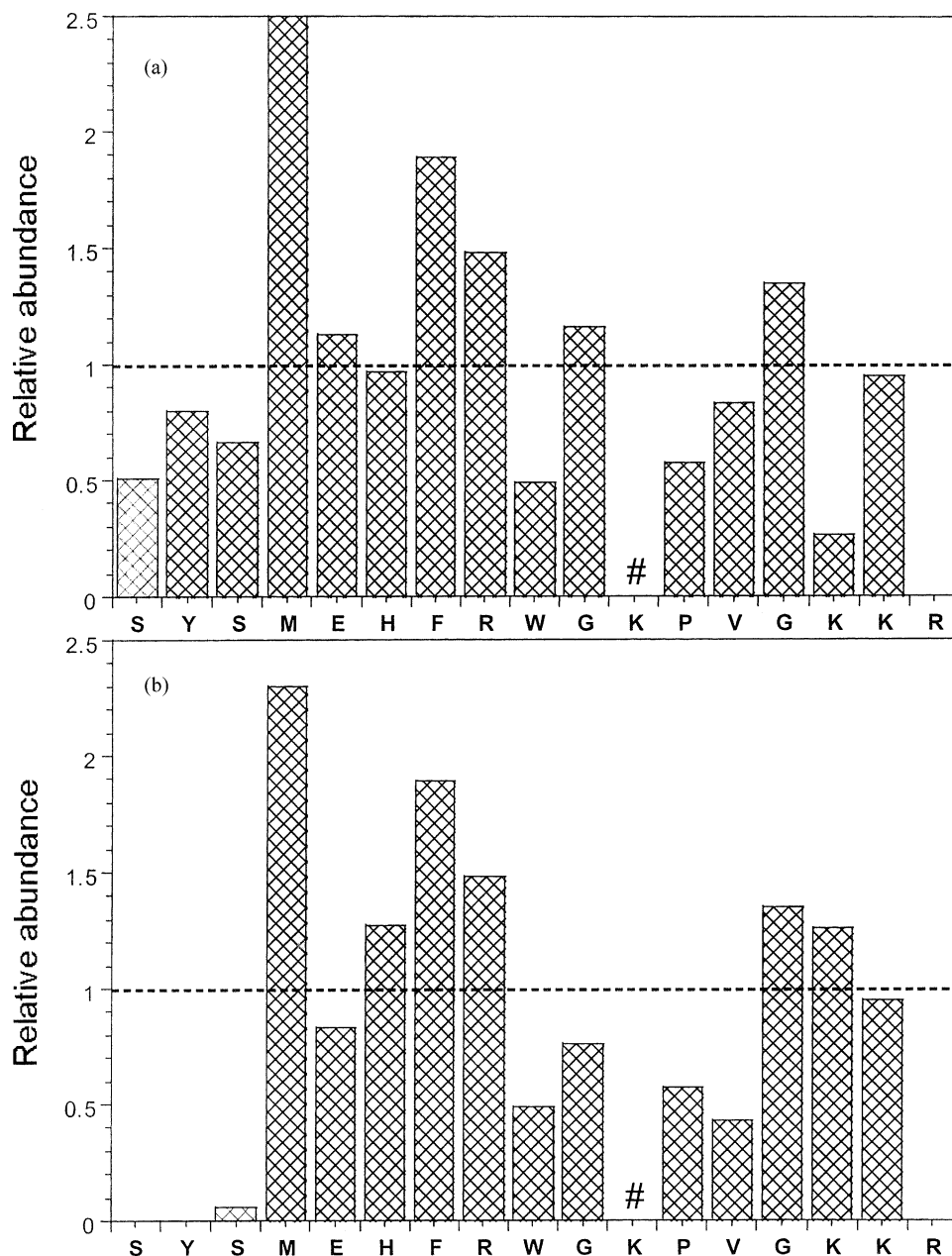


Fig. 7. The ECD cleavage frequencies in (a) 4+ molecular ions of ACTH 1–17 and (b) y_{15}^{4+} ions of the same molecule.

2+ and 3+, ACTH 1–17, 3+ through 5+ and ACTH 1–24, 4+ through 6+) reached a respectable value of 0.6 compared to $R = 0.2$ for CAD of the same species.

3.6. Size effect

The cleavage frequencies were compared in the consensus sequence regions of the ACTH 1–10, 2+, 1–17, 3+ and 1–24, 5+ ions. When the charge states were chosen such that the m/z values were similar, the average correlation coefficient was $R = 0.6$ for ECD (Fig. 5) and $R < 0.2$ for CAD.

3.7. Basicity effect

Fig. 6 shows the relative abundances of ECD fragments of 2+ ions of bradykinin and its variants. The average correlation coefficient was $R = 0.8$, despite the fact that one variant was without the C-terminal arginine and another one was with additional lysine at the N-terminus, which changed the number of strongly basic sites from two to one and three, respectively. Note that the most abundant cleavages occurred in all cases after phenylalanine residues, perhaps due to the high propensity of its side chain towards hydrogen atom capture [20]. This propensity must be affected by the following residue, because the global analysis in [13] have not found higher frequency of the ECD cleavage after phenylalanine residues.

3.8. Fragment ions vs. molecular ions

In order to obtain full cleavage, MS/MS data must sometimes be supplemented by ECD spectra of abundant multiply-charged fragments [11]. Therefore, it is important to determine whether the correlation is preserved in MS^3 . To this end, ECD fragmentation patterns of y_{15}^{4+} ions of ACTH 1–17 (Fig. 7) and b_{11}^{2+} ions of bombesin [11] were compared with fragmentation patterns of the corresponding molecular ions. The resulting average correlation factor was $R = 0.6$, somewhat lower than in MS/MS data but still usable.

4. Conclusions

The first, preliminary step towards utilizing fragment ion abundances for deriving sequencing information on peptides has been made. The results of the present statistical analysis of ECD cleavage frequencies, despite its limited scope, are encouraging. It appears that these frequencies are significantly less affected by the molecular size, basicity and the charge state than CAD data, and determined primarily by the local sequence. Although the degree to which neighboring chemical groups do affect the cleavage frequencies have not been quantitatively determined, the importance of the current findings is in the indication that ECD spectra may be reducible. This, in turn, encourages creation of a database of relative ECD cleavage frequencies. In our lab, we plan to build such a library and to test with its help the utility of the library-based sequencing approach.

Acknowledgements

P. Roepstorff and T.J.D. Jørgensen are acknowledged for useful discussions. This work has been supported by the Danish National Research Council (Grants 51-00-0238, 51-00-0358).

References

- [1] R.A. Zubarev, N.L. Kelleher, F.W. McLafferty, J. Am. Chem. Soc. 120 (1998) 3265.
- [2] D.M. Horn, R.A. Zubarev, F.W. McLafferty, Proc. Natl. Acad. Sci. U.S.A. 97 (2000) 10313.
- [3] R.A. Zubarev, E.K. Fridriksson, D.M. Horn, N.L. Kelleher, N.A. Kruger, B.K. Carpenter, F.W. McLafferty, Anal. Chem. 72 (2000) 563.
- [4] W. Yu, J.E. Vath, M.C. Huberty, S.A. Martin, Anal. Chem. 65 (1993) 3015.
- [5] A.R. Dongre, A. Somogyi, V.H. Wysocki, J. Mass Spectrom. 31 (1996) 339–350.
- [6] G. Tsapralis, H. Nair, A. Somogyi, V. Wisocki, W. Zhong, J.H. Futrell, S.G. Summerfield, S.J. Gaskell, J. Am. Chem. Soc. 121 (1999) 5142.
- [7] O. Burlet, C.Y. Yang, S.J. Gaskell, J. Am. Soc. Mass Spectrom. 3 (1992) 337.

- [8] N.L. Kelleher, R.A. Zubarev, K. Bush, B. Furie, B.C. Furie, F.W. McLafferty, C.T. Walsh, *Anal. Chem.* 71 (1999) 4250.
- [9] E. Mirgorodskaya, P. Roepstorff, R.A. Zubarev, *Anal. Chem.* 71 (1999) 4431.
- [10] A. Stensballe, O.N. Jensen, J.V. Olsen, K.F. Haselmann, R.A. Zubarev, *Rapid Commun. Mass Spectrom.* 14 (2000) 1793.
- [11] K.F. Haselmann, B.A. Budnik, J.V. Olsen, M.L. Nielsen, C.A. Reis, H. Clausen, A.H. Johnson, R.A. Zubarev, *Anal. Chem.* 73 (2001) 2998.
- [12] K. Breuker, D.M. Horn, B. Cerda, F.W. McLafferty, in: *Proceedings of the 48th ASMS Conference on Mass Spectrometry and Allied Topics*, Long Beach, CA, 2000.
- [13] N.A. Kruger, R.A. Zubarev, B.K. Carpenter, N.L. Kelleher, D.M. Horn, F.W. McLafferty, *Int. J. Mass Spectrom. Ion Processes* 182/183 (1999) 1.
- [14] K. Sannes-Lowery, R.H. Griffey, G.H. Kruppa, J.P. Speir, S.A. Hofstadler, *Rapid Commun. Mass Spectrom.* 12 (1998) 1957.
- [15] J.C. Miller, J.N. Miller, *Statistics for Analytical Chemistry*, 3rd Edition, Ellis Horwood, New York, 1993.
- [16] D.M. Horn, Y. Ge, F.W. McLafferty, *Anal. Chem.* 72 (2000) 4778.
- [17] M.F. Jarrold, *Ann. Rev. Phys. Chem.* 51 (2000) 179.
- [18] R.A. Zubarev, in: M. Larsson, J.B.A. Mitchell, I.F. Schneider (Eds.), *Dissociative Recombination, Theory, Experiments and Applications IV*, World Scientific, Singapore, 2000, pp. 214–217.
- [19] I.A. Kaltashov, D. Fabris, C.C. Fenselau, *J. Phys. Chem.* 99 (1995) 10046.
- [20] R.A. Zubarev, N.A. Kruger, E.K. Fridriksson, M.A. Lewis, D.M. Horn, B.K. Carpenter, F.W. McLafferty, *J. Am. Chem. Soc.* 121 (1999) 2857.