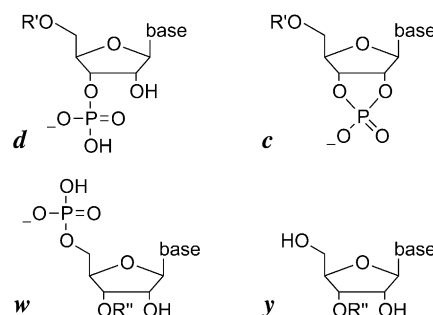**RNA Sequencing**

Characterization of Modified RNA by Top-Down Mass Spectrometry**

Monika Taucher and Kathrin Breuker

Ribonucleic acids (RNA) play a vital role in gene expression and frequently carry post-transcriptional modifications (PTMs) that can be important regulators of physiological processes.[1] For example, a post-transcriptionally modified nucleotide, m^2A (2-methyladenosine), at position 2503 of the 23S ribosomal RNA was recently shown to play a key role in halting translation.[1d] More than 100 different naturally occurring modifications are known to date,[2] but their identification and localization within RNA is generally not possible with high-throughput sequencing methods for the profiling of "transcriptomes", in which PTM information is lost by reverse transcription of RNA into complementary desoxyribonucleic acid (cDNA).[3] Methods for the detection of PTMs that do not involve cDNA usually employ enzymatic or chemical hydrolysis, which annihilates sequence information, and often require laborious labeling reactions.[3b] Herein we present a new, direct approach for the characterization of unknown, modified RNA by top-down mass spectrometry (MS) that immediately reveals the types and sites of all mass-altering modifications without the need for labeling reactions.

Whereas top-down MS of proteins is currently growing into a mature methodology, for which electron capture dissociation (ECD) and collisionally activated dissociation (CAD) of $[M+nH]^{n+}$ ions from electrospray ionization (ESI) provide complementary sequence information,[4] top-down MS of RNA is far from being established. Chait pointed out that if a sufficient number of fragments are observed, top-down MS can provide a complete description of the primary structure and reveal all mass-altering modifications, as well as any correlations that exist between these modifications.[5] However, fragmentation techniques for $[M-nH]^{n-}$ ions of RNA, that is, electron detachment dissociation (EDD)[6] and CAD, are just beginning to be developed.[7] In previous EDD experiments, we obtained 100 and 97% sequence coverage from *d* and *w* fragments (Scheme 1) for 22 and 34 nt sequences, respectively, but only 40% for 61 nt RNA.[7a] Moreover, the yields of *d*, *w* fragment ions were generally

[*] Dipl.-Ing. M. Taucher, Priv.-Doz. Dr. K. Breuker
    Institut für Organische Chemie and Center for Molecular Biosciences Innsbruck (CMBI), Universität Innsbruck
    Innrain 80–82, 6020 Innsbruck (Austria)
    E-mail: kathrin.breuker@uibk.ac.at
    Homepage: http://www.bioms-breuker.at/

Supporting information for this article is available on the WWW under http://dx.doi.org/10.1002/anie.201206232.

Scheme 1. Characteristic fragment-ion types from RNA-backbone cleavage by EDD (*d*, *w*) and CAD (*c*, *y*).

small compared to those of the oxidized molecular ions from electron detachment. In our proposed EDD mechanism for backbone cleavage into *d* and *w* fragments, we attribute this low degree of fragmentation to radical stabilization at the nucleobases.[7a] Furthermore, higher-order RNA gas-phase structure could prevent fragment-ion separation, similar to protein ions withstanding dissociation by ECD.[8] In "activated ion" ECD,[9] this limitation is overcome by the use of energetic collisions or infrared (IR) laser radiation to break noncovalent bonds.[8a]

Herein we first show that vibrational ion activation, either before or after electron detachment, has no appreciable effect on sequence coverage in EDD of 22 nt RNA, whereas increased negative-ion net charge affords 80% sequence coverage for highly modified transfer RNA coding for valine (tRNA^{Val}, 76 nt) in a single spectrum. We then demonstrate 89% sequence coverage in a single CAD spectrum of tRNA^{Val} and show how EDD and CAD data can be combined to provide extensive sequence information for the characterization of modified RNA by top-down MS.

Rigorous[10] collisional activation of 22 nt RNA before electron detachment did not affect sequence coverage (90% for $[M-8H]^{8-}$ and $[M-9H]^{9-}$, 100% for $[M-10H]^{10-}$ ions, with and without activation), and gave only marginally higher fragment-ion yields (9, 10, and 14% of all EDD products, as compared to 9, 11, and 15% with a laboratory-frame collision energy of 88, 72, and 75 eV for $[M-8H]^{8-}$, $[M-9H]^{9-}$, and $[M-10H]^{10-}$ ions, respectively; Figure 1). The similarity of the spectra with and without collisional activation shows that higher-order gas-phase structure is not a limiting factor in EDD of the 22 nt RNA. The relatively small yield of *d*, *w* fragments at sites other than 15 and 16 for $[M-8H]^{8-}$ and $[M-9H]^{9-}$, and at sites other than 15–19 for $[M-10H]^{10-}$ ions, can instead be attributed to nucleobase ionization energy (IE), as postulated previously[7a] and supported by calculations.[11]
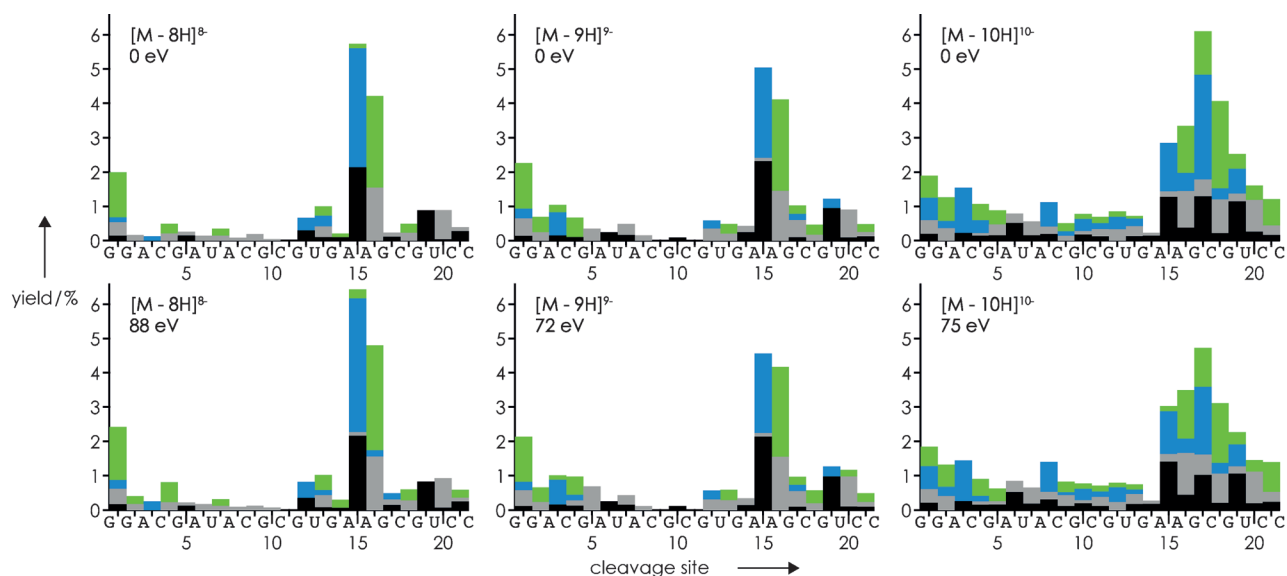
Figure 1. Site-specific yields of *d*, *w* fragments from EDD of $[M-nH]^{n-}$ ions of 22 nt RNA (*d* in black, *w* in gray) and from IR laser activation of $[M-nH]^{(n-1)-\bullet}$ ions formed by electron detachment (*d* in blue, *w* in green), without (top) and with collisional activation prior to EDD (bottom).

Activation by IR laser irradiation (20% laser power, pulse length: 100 ms) dissociated about 50% of the radical $[M-10H]^{9-\bullet}$ ions from electron detachment to virtually reproduce the *d*, *w* fragmentation pattern obtained by EDD of the $[M-10H]^{10-}$ ions (Figure 1). No *c* or *y* ions typically formed by infrared multiphoton dissociation (IRMPD)[12] or CAD[13] of $[M-nH]^{n-}$ ions of RNA were observed under these conditions. In contrast, IRMPD of the even-electron $[M-9H]^{9-}$ ions gave only 30% dissociation at far higher energy (35%, 200 ms), predominantly into *c* and *y* ions (Scheme 1) along with some *a*, *w*[13a] but no *d* ions, and no dissociation at lower energy (25%, 50 ms). Similar results were obtained for $[M-9H]^{9-}$ and $[M-8H]^{8-}$ ions (Figure 1). Rather than opening new dissociation channels, vibrational activation of $[M-nH]^{(n-1)-\bullet}$ ions produced the same *d*, *w* ions as EDD of $[M-nH]^{n-}$ ions, consistent with a high-energy transition state towards *d*, *w* ion formation.[7a]

In our proposed mechanism for RNA-backbone cleavage by EDD,[7a] the $[M-nH]^{(n-1)-\bullet}$ ions from electron detachment undergo significant structural changes initially driven by charge recombination, including a ribose–nucleobase cyclization reaction that transfers the radical to the nucleobase on the 3′ side of the nucleobase from which the electron is initially detached. Backbone cleavage into noncomplementary, even-electron *d* and *w* ions then releases an uncharged, radical nucleoside fragment. Intramolecular radical transfer over longer distances, however, does not appear to take place within the $[M-nH]^{(n-1)-\bullet}$ ions, even after vibrational activation, as the *d*, *w* fragmentation patterns from IR laser irradiation of $[M-nH]^{(n-1)-\bullet}$ and EDD of $[M-nH]^{n-}$ ions are not significantly different. Importantly, the data in Figure 1 show that whereas vibrational activation of 22 nt RNA before or after electron detachment had no appreciable effect, increasing the net charge of the precursor ion significantly increased sequence coverage.

An increase in the net charge of an ion leads to an increase in Coulomb repulsion, which should generally lower the energy required for dissociation. For example, the activation energies for covalent-bond dissociation were 1.3 and 0.8 eV for $[M+H]^+$ and $[M+2H]^{2+}$ ions of bradykinin, respectively,[14] and those for noncovalent-bond dissociation consistently increased from 1.5 to 3.5 eV when the net charge of a protein complex was decreased from 14+ to 11+.[15] Such "Coulomb activation" in EDD of RNA should facilitate fragment-ion formation and separation, and produce more random backbone cleavage by leveling differences between nucleobase IEs.[7a]

"Coulomb activation" requires highly charged precursor ions. We found in this study that the ESI additives quinuclidine, 7-methyl-1,5,7-triazabicyclo[4.4.0]dec-5-ene, 1,8-diazabicyclo[5.4.0]undec-7-ene, and 1,5-diazabicyclo[4.3.0]non-5-ene, all of which gave highly charged ions of 22 nt RNA,[16] were less effective in producing $[M-nH]^{n-}$ ions of tRNA because of alkali-ion adduction. However, a 1:1 mixture of piperidine, which efficiently suppresses alkali-ion adducts by a mechanism that we have not yet resolved,[16] and quinuclidine, produced abundant $[M-nH]^{n-}$ ions of tRNA^Val with up to 41 net negative charges (0.54 charges/nt; Figure 2a). This new additive mixture also gave ion yields more than an order of magnitude higher than those observed with triethylamine.[7a,16] We isolated tRNA^Val ions with *m/z* values between 700 and 920, corresponding to 27–35 net negative charges (Figure 2b) and subjected them to irradiation with electrons for EDD (Figure 2c), which gave *d* and *w* ions from backbone cleavage at 60 out of 75 possible sites (80% sequence coverage; Figure 3). Collisional activation of the tRNA ions (at laboratory-frame energies of 54–70 eV for ions with 27–35 net charges, respectively) prior to EDD caused about 25% undesired charged-base loss from the $[M-nH]^{n-}$ ions but did not significantly increase sequence coverage (81%; see
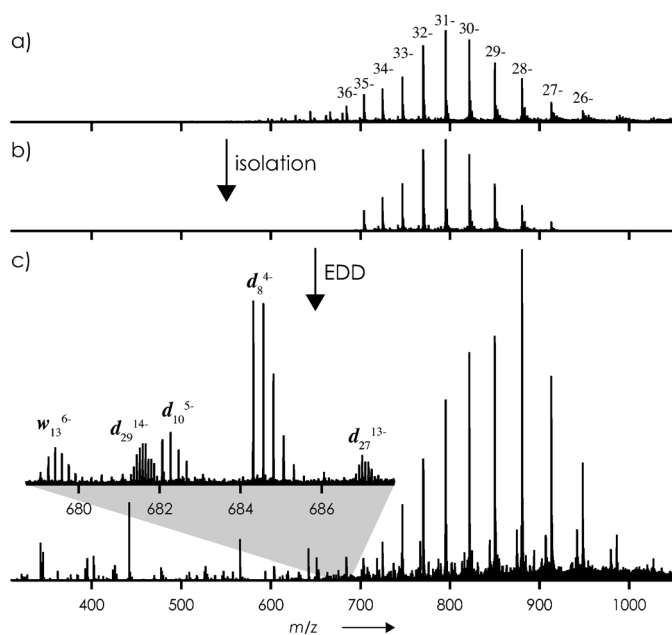
Figure 2. a) ESI mass spectrum of tRNA^{Val} (2 μM) in H_2O/CH_3OH (1:1) with piperidine (10 mM) and quinuclidine (10 mM); b) isolation of ions with m/z values between 700 and 920; c) mass spectrum after exposure of these ions to 28 eV electrons (the inset shows isotopically resolved fragment-ion signals).
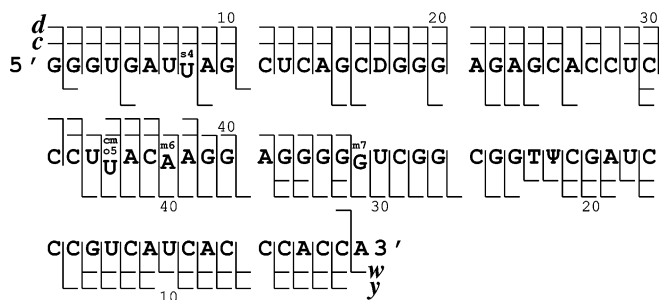

Figure 3. Fragment-ion map illustrating sequence coverage from CAD and EDD of tRNA^{Val}. The numbering of *c*, *d* and *w*, *y* fragments starts from the 5′ and 3′ terminus, respectively.

Figure 4. CAD spectrum of $[M-16H]^{16-}$ ions from ESI of tRNA^{Val} (2 μM) in H_2O/CH_3OH (1:1) with piperidine (100 mM) and imidazole (100 mM).

However, proper alignment requires knowledge about whether any given fragment includes either the 3′ or the 5′ terminus; this information is not available from mass values alone. In other words, *c* and *y* fragments in CAD spectra, and *d* and *w* fragments in EDD spectra can generally not be distinguished from each other. In this respect, the different fragmentation pathways in CAD and EDD offer a crucial advantage. Whereas CAD of even-electron RNA ions produces complementary *c* and *y* fragments, the unusual fragmentation of radical RNA ions in EDD gives noncomplementary *d* and *w* fragments and an uncharged, radical nucleoside fragment.[7a] Accordingly, the mass difference between *d* and *c* fragments (which include the original 5′ terminus) is 18.01 Da (H_2O), and that between *w* and *y* fragments (which include the original 3′ terminus) is 79.97 Da (HPO_3). Fragment mass values in the EDD spectrum that are 18.01 and 79.97 Da higher than mass values in the CAD spectrum can thus confidently be assigned as *d* and *w* ions, respectively (see Table S2 in the Supporting Information). Because this comparison at the same time identifies *c* and *y* ions in the CAD spectrum, a total of four RNA-fragment "mass ladders" can be generated (Figure 3). Additional confidence in ion assignments comes from a search in the CAD spectrum for fragments whose mass values add up to that of the intact RNA, identifying complementary *c*, *y* ion pairs.[7b] Likewise, mass values of "quasi-complementary"[7a] *d* and *w* ions in the EDD spectrum add up to the mass of the intact RNA plus 97.98 Da (H_3PO_4). Implementation of the above search criteria in algorithms for de novo sequencing[18] should be straightforward.

The combined CAD and EDD data from the two spectra in Figures 2 and 4 provide virtually complete sequence information (99%) for highly modified tRNA^{Val} (Figure 3). Sample requirements were generally the same as those for top-down MS of proteins using CAD and ECD. Importantly, the data immediately reveal the presence, type, and location of post-transcriptional modifications, although residues of the same mass (e.g., uridine and pseudouridine) cannot be distinguished from each other. However, further RNA characterization can be carried out by using alternative strategies based on mass spectrometry[19] or chemical reac-

Scheme S1 in the Supporting Information). This result is consistent with that observed for 22 nt RNA (Figure 2) and indicates that higher-order structure is not a limiting factor in EDD of such highly charged tRNA ions. Loss of PTMs from *d* or *w* ions was not observed in any of the EDD spectra.

In striking contrast to EDD, CAD of RNA is most informative when the net negative charge of the precursor ions is low.[7a,b,13a] The CAD spectrum of $[M-16H]^{16-}$ ions (0.21 charges/nt) of tRNA^{Val} in Figure 4 shows *c* and *y* ions from backbone cleavage at 67 out of 75 possible sites (89% sequence coverage; Figure 3); no fragments indicate loss of PTMs. By using $[M-24H]^{24-}$ tRNA ions for CAD in a Paul trap instrument, McLuckey and co-workers recently obtained approximately 60% sequence coverage.[7c]

Mass values of fragment ions from dissociation of unknown RNA with unknown modifications can in principle be aligned as "mass ladders"[17] to reveal the exact mass values of all residues and their sequential arrangement (Figure 3).
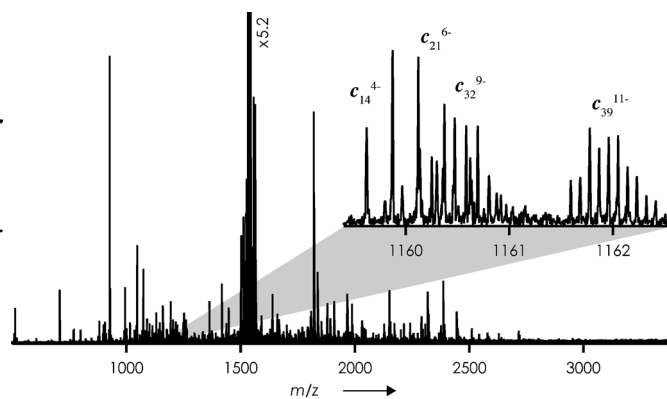
tivity.[3b] Because our MS approach detects changes in residue mass, it can even identify so far unknown types of PTMs.

In conclusion, we have described herein a straightforward approach for the detailed characterization of highly modified RNA by top-down MS that can readily be implemented in de novo sequencing strategies. In a first step, the exact (to within 1 ppm) mass of an unknown RNA is determined by ESI MS with internal calibration. Next, two separate spectra from RNA dissociation by CAD (of RNA with a low net charge, ca. 0.2 charges/nt) and EDD (of highly charged RNA, ca. 0.5 charges/nt) are recorded and calibrated (see the Supporting Information). Analysis of the two spectra gives two separate lists with monoisotopic fragment mass values. Any values in the EDD mass list that are 18.01 and 79.97 Da higher than values in the CAD mass list are assigned to d and w ions, respectively. The corresponding values in the CAD mass list, that is, those that are 18.01 and 79.97 Da lower than values in the EDD mass list, are assigned to c and y ions, respectively. Next, values for each ion type (c, d, w, y) are separately sorted by mass. This process results in four "fragment mass ladders" that reveal both the exact mass values of RNA residues and their alignment within the sequence. Our new approach can be used to characterize modified as well as unmodified RNA, but should be especially useful for synthetic or post-transcriptionally modified RNA whose analysis by chemical or biochemical methods is rather limited.[3b]

Experimental Section

Experiments were performed on a 7 T Fourier transform ion cyclotron resonance mass spectrometer (Bruker, Austria) equipped with an ESI source, a collision cell for CAD, a hollow dispenser cathode for EDD (electron energy: 28 eV), and a CO_2 laser (10.6 μm, 35 W) for IR activation. Chemicals and tRNA^{Val} (from *Escherichia coli*, GGGUG AUs⁴UAG CUCAG CDGGG AGAGC ACCUC CCUc-mo⁵UA Cm⁶AAGG AGGGG m⁷GUCGG CGGTΨ CGAUC CCGUC AUCAC CCACC A, modified residues in bold; s⁴U: 4-thiouridine, D: dihydrouridine, cmo⁵U: uridine-5-oxyacetic acid, m⁶A: N^6-methyladenosine, m⁷G: 7-methylguanosine, T: thymidine, Ψ: pseudouridine) were purchased from Sigma–Aldrich (Austria). The 22 nt RNA (GGACG AUACG CGUGA AGCGU CC) was prepared by solid-phase synthesis.[20] For increased statistics, 500 scans were added for each spectrum. For experimental details, see the Supporting Information. SNAP2 (Bruker) and mMass[21] software were used for data reduction. For Figure 1, the EDD data were normalized to the sum of all EDD products (100% = [oxidized molecular ions] + ([d ions] + [w ions])/2), and EDD/IR data were normalized to the abundance of the $[M-\mathrm{n\,H}]^{(n-1)-\bullet}$ ion in the corresponding EDD spectrum.

[1] a) A. Hüttenhofer, P. Schattner, N. Polacek, *Trends Genet.* **2005**, *21*, 289–297; b) S. R. Eddy, *Nat. Rev. Genet.* **2001**, *2*, 919–929; c) C. S. Chow, T. N. Larnichhane, S. K. Mahto, *ACS Chem. Biol.* **2007**, *2*, 610–619; d) N. Vázquez-Laslop, H. Ramu, D. Klepacki, K. Kannan, A. S. Mankin, *Embo J.* **2010**, *29*, 3108–3117; e) K. D. Meyer, Y. Saletore, P. Zumbo, O. Elemento, C. E. Mason, S. R. Jaffrey, *Cell* **2012**, *149*, 1635–1646.

[2] W. A. Cantara, P. F. Crain, J. Rozenski, J. A. McCloskey, K. A. Harris, X. N. Zhang, F. A. P. Vendeix, D. Fabris, P. F. Agris, *Nucleic Acids Res.* **2011**, *39*, D195–D201.

[3] a) Z. Wang, M. Gerstein, M. Snyder, *Nat. Rev. Genet.* **2009**, *10*, 57–63; b) S. Kellner, J. Burhenne, M. Helm, *RNA Biol.* **2010**, *7*, 237–247.

[4] W. D. Cui, H. W. Rohrs, M. L. Gross, *Analyst* **2011**, *136*, 3854–3864.

[5] B. T. Chait, *Science* **2006**, *314*, 65–66.

[6] a) B. A. Budnik, K. F. Haselmann, R. A. Zubarev, *Chem. Phys. Lett.* **2001**, *342*, 299–302; b) B. Ganisl, T. Valovka, M. Hartl, M. Taucher, K. Bister, K. Breuker, *Chem. Eur. J.* **2011**, *17*, 4460–4469; c) F. E. Leach, J. J. Wolff, T. N. Laremore, R. J. Linhardt, I. J. Amster, *Int. J. Mass Spectrom.* **2008**, *276*, 110–115.

[7] a) M. Taucher, K. Breuker, *J. Am. Soc. Mass Spectrom.* **2010**, *21*, 918–929; b) M. Taucher, U. Rieder, K. Breuker, *J. Am. Soc. Mass Spectrom.* **2010**, *21*, 278–285; c) T. Y. Huang, J. A. Liu, S. A. McLuckey, *J. Am. Soc. Mass Spectrom.* **2010**, *21*, 890–898.

[8] a) K. Breuker, M. Jin, X. M. Han, H. H. Jiang, F. W. McLafferty, *J. Am. Soc. Mass Spectrom.* **2008**, *19*, 1045–1053; b) K. Breuker, S. Brüschweiler, M. Tollinger, *Angew. Chem.* **2011**, *123*, 903–907; *Angew. Chem. Int. Ed.* **2011**, *50*, 873–877.

[9] D. M. Horn, Y. Ge, F. W. McLafferty, *Anal. Chem.* **2000**, *72*, 4778–4784.

[10] The collision energies used for 22 nt RNA caused fragmentation of approximately 50% of the ion population; only the remaining undissociated but highly activated $[M-n\mathrm{H}]^{n-}$ ions were subjected to EDD. Because the deeper potential well for collisional activation trapped about twice as many ions, precursor-ion abundances for EDD with and without collisional activation were similar, and the spectra comparable in terms of their signal-to-noise ratio.

[11] V. V. Zakjevskii, S. J. King, O. Dolgounitcheva, V. G. Zakrzewski, J. V. Ortiz, *J. Am. Chem. Soc.* **2006**, *128*, 13350–13351.

[12] M. W. Gardner, N. Li, A. D. Ellington, J. S. Brodbelt, *J. Am. Soc. Mass Spectrom.* **2010**, *21*, 580–591.

[13] a) T. Y. Huang, A. Kharlamova, J. Liu, S. A. McLuckey, *J. Am. Soc. Mass Spectrom.* **2008**, *19*, 1832–1840; b) J. M. Tromp, S. Schürch, *J. Am. Soc. Mass Spectrom.* **2005**, *16*, 1262–1268.

[14] W. D. Price, P. D. Schnier, E. R. Williams, *Anal. Chem.* **1996**, *68*, 859–866.

[15] N. Felitsyn, E. N. Kitova, J. S. Klassen, *Anal. Chem.* **2001**, *73*, 4647–4661.

[16] B. Ganisl, M. Taucher, C. Riml, K. Breuker, *Eur. J. Mass Spectrom.* **2011**, *17*, 333–343.

[17] B. T. Chait, R. Wang, R. C. Beavis, S. B. H. Kent, *Science* **1993**, *262*, 89–92.

[18] D. M. Horn, R. A. Zubarev, F. W. McLafferty, *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 10313–10317.

[19] a) T. Brückl, D. Globisch, M. Wagner, M. Müller, T. Carell, *Angew. Chem.* **2009**, *121*, 8074–8077; *Angew. Chem. Int. Ed.* **2009**, *48*, 7932–7934; b) M. Taucher, B. Ganisl, K. Breuker, *Int. J. Mass Spectrom.* **2011**, *304*, 91–97.

[20] R. Micura, *Angew. Chem.* **2002**, *114*, 2369–2373; *Angew. Chem. Int. Ed.* **2002**, *41*, 2265–2269.

[21] M. Strohalm, D. Kavan, P. Novak, M. Volny, V. Havlicek, *Anal. Chem.* **2010**, *82*, 4648–4651.