

# FAST MS: Software for the Automated Analysis of Top-Down Mass Spectra of Polymeric Molecules Including RNA, DNA, and Proteins

Published as part of *Journal of the American Society for Mass Spectrometry* special issue "Asilomar: Computational Mass Spectrometry".

Michael Palasser\* and Kathrin Breuker\*



Cite This: *J. Am. Soc. Mass Spectrom.* 2025, 36, 247–257



Read Online

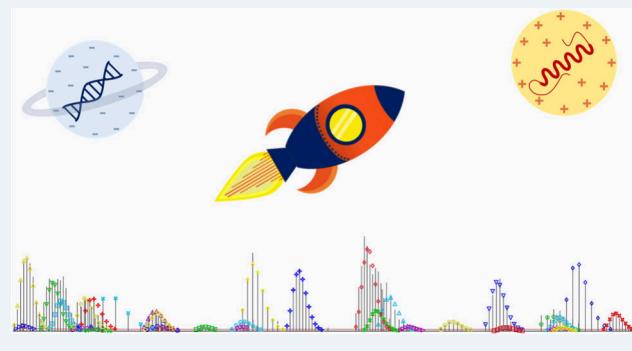
**ACCESS |**

 Metrics & More

 Article Recommendations

 Supporting Information

**ABSTRACT:** Top-down mass spectrometry (MS) enables comprehensive characterization of modified proteins and nucleic acids and, when native electrospray ionization (ESI) is used, binding site mapping of their complexes with native or therapeutic ligands. However, the high complexity of top-down MS spectra poses a serious challenge to both manual and automated data interpretation, even when the protein, RNA, or DNA sequence and the type of modification or the ligand are known. Here, we introduce FAST MS, a user-friendly software that identifies, assigns and relatively quantifies signals of molecular and fragment ions in MS and MS/MS spectra of biopolymers with known sequence and provides a toolbox for statistical analysis. FAST MS searches mass spectra for ion signals by comparing all signals in the spectrum with isotopic profiles calculated from known sequences, resulting in superior sensitivity and an increased number of assigned fragment ions compared to algorithms that rely on artificial monomer units while maintaining the false positive rate on a moderate level (<5%). FAST MS is an open-source, cross-platform software for the accurate identification, localization and relative quantification of modifications, even in complex mixtures of positional isomers of proteins, oligonucleotides, or any other user-defined linear polymer.



## INTRODUCTION

Top-down mass spectrometry (MS) is a powerful approach for the identification, localization, and relative quantification of posttranslational,<sup>1–3</sup> posttranscriptional,<sup>4</sup> and synthetic modifications,<sup>5</sup> and has recently been extended to the analysis of products from reactions for chemical probing of ribonucleic acid (RNA) structure<sup>6</sup> and binding site mapping of RNA-peptide<sup>7,8</sup> and RNA-small molecule<sup>9,10</sup> complexes. By omitting the digestion step characteristic of bottom-up MS and instead dissociating intact molecules in the gas phase, possible correlations between posttranslational or posttranscriptional modifications (PTMs) are preserved in top-down MS, allowing for full characterization of 'proteoforms'<sup>11</sup> or different isomers and forms of RNA.<sup>12</sup> Recent advances in instrumentation, along with the increasing number of FDA-approved protein, peptide and oligonucleotide therapeutics,<sup>13–15</sup> have inspired new applications of top-down MS in academic research and the pharmaceutical industry.<sup>16–20</sup>

While MS/MS spectra can often be acquired within a few minutes, data analysis remains the bottleneck for studies on a larger scale. Depending on the dissociation method and the biomolecule under investigation, up to six (proteins:  $a/x, b/y, c/z$ ) or eight (nucleic acids:  $a/w, b/x, c/y, d/z$ ) different types of

fragments can be formed for each backbone cleavage site, often in multiple charge states and accompanied by neutral losses.<sup>16,21</sup> As a result, top-down mass spectra can be highly complex, especially for larger and heterogeneously modified precursor molecules, making their analysis very challenging. More specifically, a single top-down MS spectrum can contain thousands of isotope peaks representing hundreds of often overlapping isotopic distributions that must be identified, quantified, and assigned to possible fragment species before the data can be used to, for example, localize and quantify modifications or to provide information about ligand binding sites.<sup>7–10,12,22</sup>

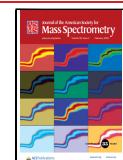
Over the past few decades, considerable effort has been invested in automating the analysis of MS and MS/MS spectra, including the detection and quantification of isotopic distributions and the assignment of fragment ions. Perhaps the most

**Received:** June 2, 2024

**Revised:** December 2, 2024

**Accepted:** December 10, 2024

**Published:** December 23, 2024



**Table 1.** RNA, DNA, and Proteins Studied

name	length	mass/ kDa	sequence
RNA 1	15 nt	4.8	5'-GAAGG UUCGC CUUCG-3'
RNA 2	15 nt	4.8	5'-GAAGG GCAAC CUUCG-3'
RNA 3	17 nt	5.5	5'-GCGAA CCUGhm <sup>5</sup> C GGGUU CG-3' <sup>a</sup>
RNA 4	22 nt	7.0	5'-CGAAG GUUCG CCUUC GCGUC AG-3'
RNA 5	22 nt	7.0	5'-CGUCA GCGAA GGUUC GCCUU CG-3'
NSR RNA	27 nt	8.5	5'-GGCUG CUUGU CCUUU AAUGG UCCAG UC-3'
RRE-IIB-0 RNA	39 nt	12.6	5'-GGUCU GGGCG CAGCG UCAAU GACGC UGACG GUACA GGCC-3'
RRE-I RNA	47 nt	15.2	5'-GGGUU CUUGG GAGCA GCAGG AUUCG UCCUG GCUGU GGAAA GAUAC CC-3'
DNA 1	20 nt	6.0	5'-GCTAC ATTTA TCACG CGCTT-3'
CpG1018 DNA <sup>b</sup>	22 nt	7.1	5'-TGACT GTGAA CGTTC GAGAT GA-3'
ubiquitin	76 aa	8.6	MQIFV KTLTG KTITL EVEPS DTİEN VKAKI QDKEG IPPDQ QRLLF AGKQL EDGRT LSDYN IQKES TLHLV LRLRGG
calmodulin	148 aa	16.7	ADQLT EEQIA EFKEA FSLFD KDGDG TITTK ELGTV MRSLG QNPTE AELQD MINEV DADGN GTIDF PEFLT MMARK MKDTD SEEEL REAFR VFSDKD GNGYI SAAEL RHVMT NLGEK LTDEE VDEMI READI DGDGQ VNYEE FVQMM TAK

<sup>a</sup>hm<sup>5</sup>C = 5-hydroxymethylcytidine. <sup>b</sup>DNA with phosphorothioate backbone.

influential algorithm, THRASH,<sup>23</sup> detects and quantifies protein fragment ions by identifying sets of isotopic peaks in the spectrum and comparing them to isotopic distributions obtained by linear interpolation between two abundance distributions from a look-up table of abundance distributions calculated for polymers of the artificial monomer unit 'averagine' whose elemental composition ( $C_{4.9384}H_{7.7583}N_{1.3577}O_{1.4773}S_{0.0417}$ ) represents an average of the canonical amino acid residues.<sup>24</sup> Algorithms including MS-Decconv,<sup>25</sup> Xtract (Thermo Fisher Scientific), and SNAP (Bruker Daltonics), have adopted this concept for automated data analysis. Additional programs were developed that can analyze protein fragment lists generated by any of these algorithms in terms of sequence coverage and PTMs<sup>26–30</sup> or conduct a database search to identify and characterize proteins and proteoforms.<sup>31,32</sup> Recently, multifunctional software packages were published that incorporate several of these algorithms for applications in the field of proteomics.<sup>33–36</sup> The use of artificial model amino acids is well suited for untargeted approaches but has some avoidable disadvantages when the sequence of the protein under study is already known.<sup>37–39</sup> First, a random search for isotopic distributions in the spectrum increases the number of false positives, and second, fragment ions of low abundance are often missed or assigned to an incorrect mass or charge. Third, the deviation of mass and relative isotope abundance values calculated from artificial model amino acids from actual values generally increases with decreasing fragment mass for statistical reasons and can become intolerable if rare elements such as bromine are present. For these reasons, the analysis of top-down mass spectra often requires some level of time-consuming manual postprocessing.

In the emerging field of top-down MS of RNA with applications in modification analysis of biologically relevant or therapeutic RNA and in structural biology, there is a similar need for automated data analysis. For example, RNA therapeutics are typically highly modified, and the characterization of their sequence and modifications, as well as the assessment of their purity, is critical. Among the noncommercial software developed for MS of RNA and deoxyribonucleic acid (DNA),<sup>40–50</sup> we could not find any that takes into account positional isomers and can relatively quantify site-specific modifications.

The data analysis software FAST MS (Free Analysis Software for Top-Down Mass Spectrometry) was developed to address these challenges. Based on the intact sequence of the

(bio)molecule under study, the program identifies, assigns and relatively quantifies isotopic distributions in peak lists derived from MS or MS<sup>n</sup> spectra and provides several tools for statistical analysis. FAST MS works for RNA, DNA, proteins, and any user-defined linear polymer. Monomer units, modifications, fragment ion types, and even elemental isotope abundances can be defined by the user within the easy-to-use graphical user interface (GUI), providing maximum flexibility for targeted MS applications. FAST MS is open-source and available at GitHub (<https://github.com/michael-palasser/FAST-MS/releases>).

## ■ EXPERIMENTAL SECTION

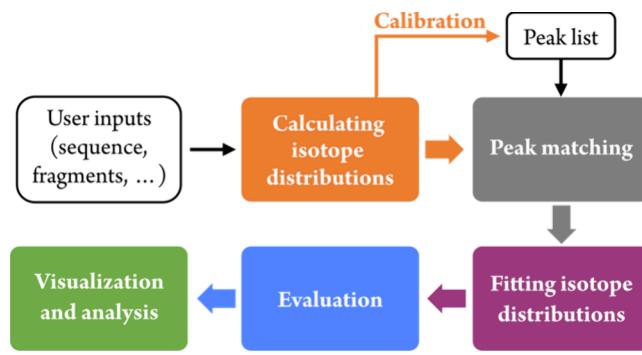
The MS and MS/MS spectra of RNAs, proteins, and DNA 1 (Table 1) were previously recorded on a 7 T Fourier transform ion cyclotron resonance (FT-ICR) instrument (Apex Ultra, Bruker, Austria) equipped with an electrospray ionization (ESI) source, a linear quadrupole for ion isolation, a collision cell for collisionally activated dissociation (CAD) or radical transfer dissociation (RTD), and a hollow dispenser cathode for electron capture dissociation (ECD) or electron detachment dissociation (EDD). ESI spectra were obtained by operation of the linear quadrupole in transmission (radiofrequency-only) mode. For CAD, RTD, ECD, or EDD, ions of interest were isolated in the quadrupole and dissociated in the collision (CAD, RTD) or ICR cell (ECD, EDD); peak picking utilized the FTMS algorithm (Bruker, Austria). The MS/MS spectrum of CpG1018 DNA was recorded on a quadrupole time-of-flight (QTOF) instrument (maXis II, Bruker, Germany) by isolation of (M-9H)<sup>9</sup>- ions in the quadrupole and dissociation in the collision cell; peak picking used the Sum Peak algorithm (Bruker, Germany).

To estimate the number of false positives, the MS/MS spectra of modified and unmodified precursor ions were analyzed using the correct protein, RNA, or DNA sequence and once more with the same settings using an incorrect fake sequence of equal length. The fake sequence for ubiquitin was truncated calmodulin, and the fake sequence for calmodulin was extended ubiquitin. Since there are only four canonical nucleotides, random RNA or DNA sequences would produce many fragments that are isomeric to the correct ones. For example, using a random DNA sequence for analysis of the CAD spectrum of CpG1018 DNA resulted in an unrealistic false positive rate of 33%. Therefore, all-adenosine sequences were used as fake sequences for all RNAs and DNA 1, and a

phosphorothiolated 22 nt fake sequence incorporating 20 adenosines and two terminal uridines was used for CpG1018 DNA. The number of identified *c* and *y* ions (CAD of RNA), *b* and *y* ions (CAD of proteins), *c* and *z*<sup>•</sup> ions (ECD of proteins), *a*, *a*-base, *b*, *c*, *d*, *w*, *x*, *y*, *z* ions (CAD of DNA), and *d* and *w* ions (EDD of DNA) in the analyses with the correct and the fake sequences were used to calculate the false discovery rate.

**Description of the Algorithm.** FAST MS was written in Python 3 using PyCharm 2020 1.4 (JetBrains) and was implemented in a three-tier architecture. It utilizes several open-source Python libraries: Numpy,<sup>51</sup> Numba,<sup>52</sup> Pandas,<sup>53</sup> Scipy,<sup>54</sup> and XlsxWriter (<https://xlsxwriter.readthedocs.io>). For creating the GUI, PyQt5 (<https://pypi.org/project/PyQt5/>), PyQtGraph (<https://www.pyqtgraph.org/>), and Matplotlib (<https://matplotlib.org/>) were employed. Furthermore, PyInstaller (<https://pyinstaller.org/en/stable/>) was used to generate executables that need no Python installation. The main functions and methods were unit-tested using pytest (<https://docs.pytest.org/en/stable/contents.html>) with total line coverage of >60%. The basic workflow is illustrated in Scheme 1.

#### Scheme 1. FAST MS Workflow for Analyzing MS/MS Spectra



FAST MS imports experimental data as a peak list in plain text or CSV file format. This list of *m/z* and associated signal height values can be generated using the MS instrument manufacturer's software, e.g. DataAnalysis (Bruker Daltonics) or Xcalibur (Thermo Scientific). No further data reduction (e.g., deconvolution or charge assignment) is required, as FAST MS does not use deconvoluted data but directly compares calculated isotopic distributions of ions with the peak list of an *m/z* spectrum.

**Calculation of Theoretical Isotope Distributions.** Instead of using artificial monomer units,<sup>23,24,39</sup> FAST MS calculates exact theoretical isotope distributions based on the chemical formula of each species. Isotopic fine structures are computed by applying the polynomial approach.<sup>55</sup> Isotope peak masses and abundances are then calculated as the centroid mass and the sum of the isobaric isotopologue abundances. Because this process can be computationally highly demanding for heavier isotope peaks, FAST MS uses a fast Fourier transform (FFT) based method.<sup>55</sup> The differences between isotope mass values are approximated by

$$\frac{\sum_i \frac{\Delta m_i}{\Delta m_{nom,i}} \cdot P_i \cdot N_{elem}}{\sum_i P_i \cdot N_{elem}} \quad (1)$$

where  $P_i$  is the abundance of the isotope  $i$ ,  $N_{elem}$  is the number of atoms of the corresponding element incorporated in the molecule, and  $\Delta m_i$  and  $\Delta m_{nom,i}$  describe the exact and nominal

mass increases of the isotope relative to its monoisotopic mass; this approach is similar to that proposed by Sadygov.<sup>56</sup> The mass differences between approximated and centroid mass were typically <0.05 ppm for RNA and <0.20 ppm for proteins.

**Ion Detection.** To avoid false positives, FAST MS only searches for the most probable net charges of each fragment, which depend on the precursor ion net charge, the size of the fragment, and any charged modifications. Starting with the most abundant peaks of each theoretical isotope distribution, the program matches the theoretical data with the peak list from the spectrum. Considering that the mass accuracy often decreases with *m/z*, a linear mass error threshold is applied:

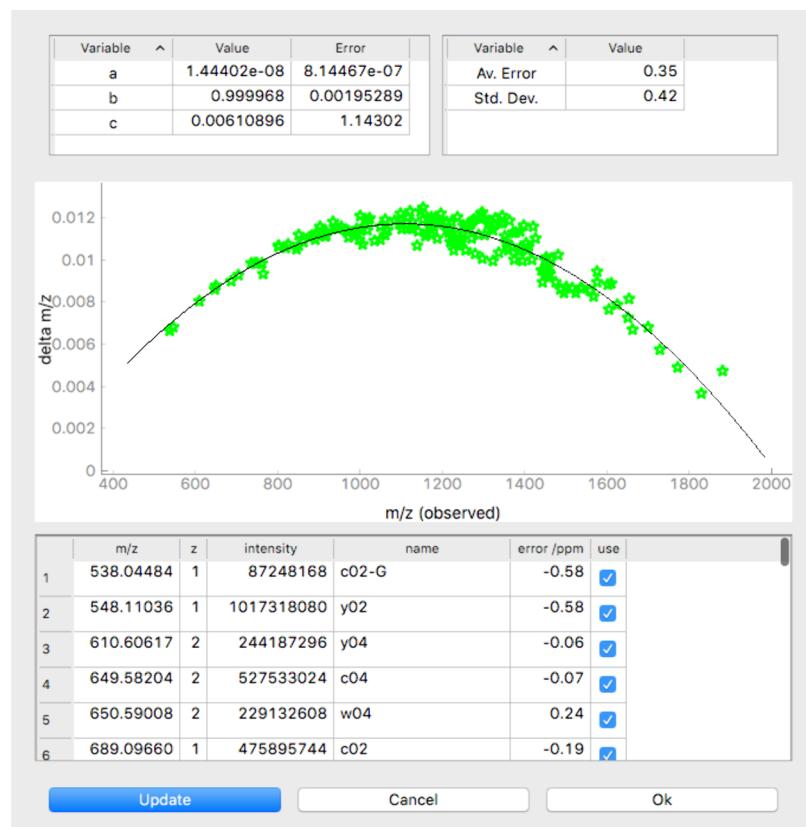
$$\delta_{max} = a \cdot m/z + b \quad (2)$$

where  $a$  and  $b$  are user-definable constants. Because FAST MS processes peak data instead of the profile spectrum, conventional noise detection workflows are not applicable. Therefore, a new algorithm had to be implemented in which the noise in the *m/z* window around an ion signal is determined iteratively by computing the mean intensity within the user-definable window (default width: 3 *m/z*), excluding peaks that are more than 33% higher than the mean value, and calculating the mean again. This procedure is repeated until the mean remains constant. Depending on the remaining peak density  $\rho$  in the window, this mean value is multiplied by the factor

$$\frac{\rho}{2.5 + \rho} \quad (3)$$

(for  $\rho > 5$ ) or by a factor of 0.67 (otherwise) to obtain noise values for each window, which increases the accuracy in regions of the spectrum where the peak density is high enough for this correction. The proposed workflow was developed empirically and worked for different analytes, ion polarities, dissociation techniques, and types of mass spectrometers (Figure S1). Because it relies on the presence of noise peaks in the list of peaks fed into FAST MS, the thresholds used for the peak picking algorithm (e.g., FTMS in DataAnalysis) should be sufficiently low. Otherwise, FAST MS might consider actual peaks as noise, leading to inaccurate S/N values or even missed ion signals.

**Ion Quantification and Evaluation.** The basic steps in the quantification process are illustrated in Figure S2. The theoretical abundances of the isotope peaks are iteratively fitted to the observed abundances using a least-squares model. Individual isotope peaks that deviate from a given theoretical isotopic distribution ('outliers') are identified using an empirically adapted version of the Grubbs test.<sup>57</sup> When signals of different ions overlap in *m/z*, a linear combination of the corresponding isotopic distributions is modeled in the next step. To avoid overfitting and false positives, overlapping ions are deleted when their abundance decreases to a user-definable percentage of the previous abundance during an iteration step (default: 80/number of distributions in %, e.g., 40% for two overlapping distributions). After the fit, the program calculates a quality error value for each ion, which is defined by the minimum of the fit's cost function divided by the intensity of the ion or those of overlapping ions. The signal-to-noise ratio (S/N) is calculated by dividing the intensity (sum of the modeled isotope peak signal heights) by the noise next to the ion. The mass and quality errors as well as the S/N values are evaluated, and entries above (errors) or below (S/N) user-defined thresholds are moved to the list of deleted ions.



**Figure 1.** FAST MS dialogue window for internal calibration. FAST MS compares the monoisotopic  $m/z$  values of the signals specified in the list of signals to be used for calibration with the  $m/z$  values of fragment ions calculated from the sequence and calculates fit coefficients, the average mass error, and the standard deviation of the mass errors (both in ppm). The difference between calibrated and uncalibrated  $m/z$  values ( $\Delta m/z$ ) of the signals and the fit function are visualized graphically, and the user can interactively control the quality of the fit by excluding misassigned signals. Once the user is satisfied with the fit, the program applies the calibration to the peak list for further analysis.

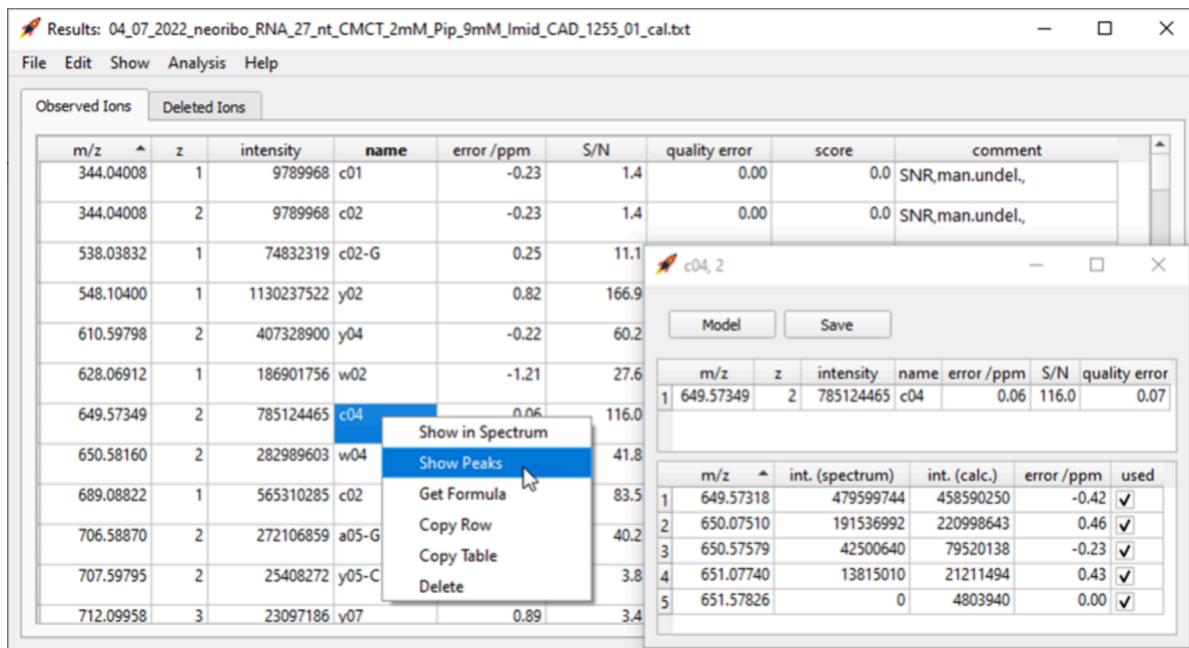
**Statistical Analysis.** Relative ion abundances are calculated by adding the signal heights of all isotopic peaks (reported as “intensity”). For charge-sensitive analyzers (e.g., FT-ICR and Orbitrap), FAST MS includes the option to divide these values by the charge of the ions (reported as “int./ $z$ ”). Since dissociation of a molecular ion typically produces two complementary fragment ions, fragment (but not molecular) ion abundances are halved for calculation of relative fragment yields. Sequence coverage is calculated as the percentage of the number of cleavage sites for which fragments were observed relative to the total number of cleavage sites.

## RESULTS AND DISCUSSION

FAST MS was designed to provide flexibility for analyzing any type of linear polymers and modifications. In addition to amino acids and nucleotides, new monomer units and post-translational, post-transcriptional, or synthetic modifications can be defined by the user. Importantly, FAST MS allows for the analysis of mixtures of positional isomers, including therapeutic peptides, proteins, and RNA with heterogeneous modification profiles (an example is shown in Figure S3). Moreover, isotopically labeled, enriched, or depleted molecules can be analyzed by editing the isotope data. The fragment types that are generated by MS/MS not only depend on the dissociation technique used (e.g., CAD, ECD/ETD, or UVPD)<sup>16</sup> and the type of molecule but also on the applied energy<sup>58–60</sup> and modifications (e.g., 2'-modification of RNA).<sup>60,61</sup> To accommodate all types of MS/MS applications, FAST MS can be

customized according to the dissociation technique used and the compound class. For maximum convenience and runtime stability, sequences, modifications, fragment types, monomer units, and isotope data can be easily accessed and edited within the GUI.

**Analyzing an MS/MS Spectrum.** To demonstrate the features of FAST MS, we used data from chemical probing of the 27 nt NSR RNA (Table 1) with CMCT (*N*-Cyclohexyl-*N'*- $\beta$ -(4-methylmorpholinium)ethylcarbodiimide *p*-toluenesulfonate).<sup>6</sup> Briefly, the CMC<sup>+</sup> cation (*N*-Cyclohexyl-*N'*- $\beta$ -(4-methylmorpholinium)ethylcarbodiimide) reacts with N3 and N1 of exposed uridine and guanosine bases, respectively. To localize the sites and the site-specific extent of CMC<sup>+</sup>-modification, singly CMC<sup>+</sup>-modified NSR ions with a net charge of 7-, (NSR<sup>CMC<sup>+</sup>-8H</sup>)<sup>7-</sup>, were isolated in the mass spectrometer and dissociated by CAD. Analysis of the CAD data using FAST MS then allows conclusions to be drawn about the structure of the NSR RNA in solution.<sup>6</sup> At the beginning of the analysis, the user defines the sequence of the molecule, the precursor net charge, the dissociation technique, the type and number of modifications, and the file path of the peak list. In an optional step, the mass spectrum can be calibrated internally by using a tab-separated list that specifies the measured monoisotopic  $m/z$ ,  $z$ , and intensity values of the signals to be used for calibration. This list can be generated using a data reduction algorithm, such as SNAP, or created manually by selecting monoisotopic signals from the peak list of the spectrum. The algorithm involves several cycles of calibration



**Figure 2.** FAST MS window displaying the list of ions identified in a spectrum. Several options are available by right-clicking on an entry. The window on the right shows values related to the isotope peaks of the  $c_4^{2-}$  ion. After adjusting the measured peak signal heights (“int. (spectrum)”), the isotope peaks can be refitted (button “Model”).

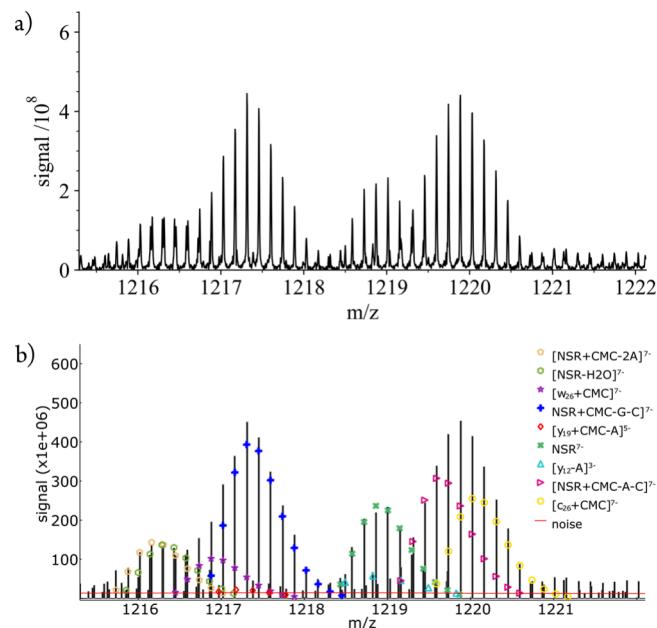
and rejection of values with unrealistic mass errors until the standard deviation between theoretical and calibrated  $m/z$  values falls below a user-defined threshold. Although this process can run entirely automated, it is possible to manually select or remove ions from the list and recalculate the quadratic calibration function (**Figure 1**):

$$m/z_{\text{cal.}} = a \cdot (m/z)_{\text{uncal.}}^2 + b \cdot m/z_{\text{uncal.}} + c \quad (4)$$

where  $a$ ,  $b$ , and  $c$  are the fit coefficients and  $m/z_{\text{cal.}}$  and  $m/z_{\text{uncal.}}$  are the calibrated and uncalibrated  $m/z$  values.

The assigned and deleted ion signals found by FAST MS are displayed in a list (**Figure 2**). Overlaps with other ions, the reason for the deletion, and editing events by the user are indicated in the comment section for each ion. The program also highlights suspicious values that might suggest an incorrect assignment. To assist the user while inspecting the results, the isotope peaks can be listed and displayed in an interactive line spectrum. **Figure 3** shows that even heavily overlapping isotope clusters are automatically resolved by FAST MS. However, isotope peaks can be manually refitted, deleted, and restored if the user decides that the results need to be corrected. To ensure reproducibility, an audit trail window records the configurations, all inputs and editing events by the user. For later use, the analysis can be saved to a database or exported to Excel.

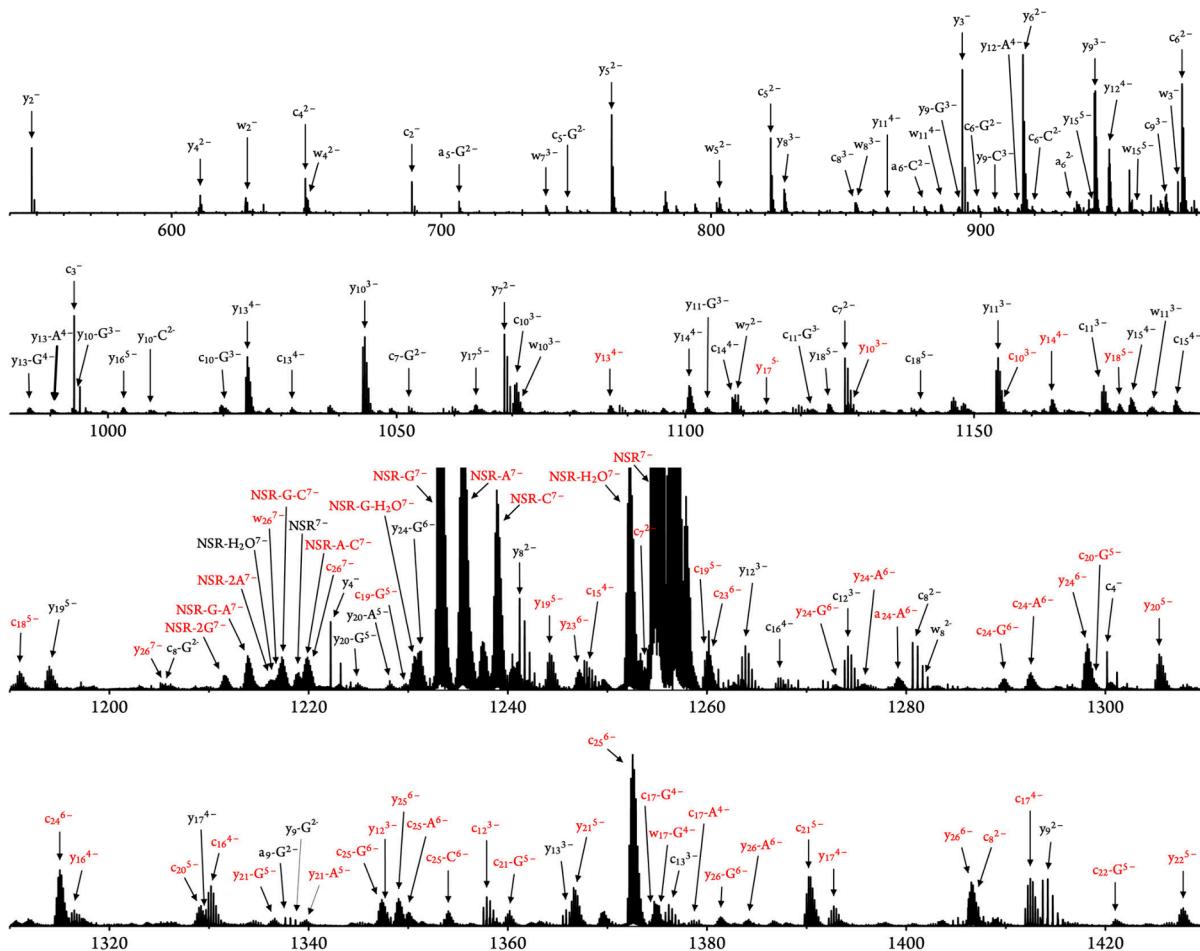
**Figure 4** shows sections of the assigned CAD spectrum of the  $(\text{NSR}^{\text{CMC}+}-8\text{H})^{7-}$  ions. Even though the size of the RNA is only 27 nt, the peak density is high, especially in the  $m/z$  region around the precursor ions. In the  $m/z$  range of 1200–1300 alone, 79 different isotopic distributions were identified. Depending on the secondary structure of the RNA and its sequence, chemical probing typically leads to mixtures of positional isomers from modification at different exposed sites. As a consequence, fragment ions from a given cleavage site can often be observed both with (red) and without (black) CMC<sup>+</sup>-modification (e.g.,  $y_{10}$ ). In addition, many fragments occur in more than one charge state, and some nucleobase loss caused by



**Figure 3.** (a) Section of a profile spectrum from CAD of  $(\text{NSR}^{\text{CMC}+}-8\text{H})^{7-}$  ions and (b) 9 different overlapping isotopic profiles (markers) from fitting the corresponding peak values (vertical black lines) with FAST MS; the calculated noise level is shown as a horizontal red line.

CAD further increases the complexity of the spectrum. It is easy to imagine that manual assignment and quantification would take hours or even days of tedious work. Using FAST MS, however, this work can usually be done within a few minutes.

**Evaluation.** To evaluate the performance of FAST MS, the spectrum shown in **Figure 4** was analyzed by both FAST MS and SNAP. Similar to THRASH, SNAP uses artificial monomer units to detect ion signals in the spectrum. Using a relative intensity threshold of 0.01%, SNAP found 616 ions, of which 261 could be assigned. When the relative intensity threshold was



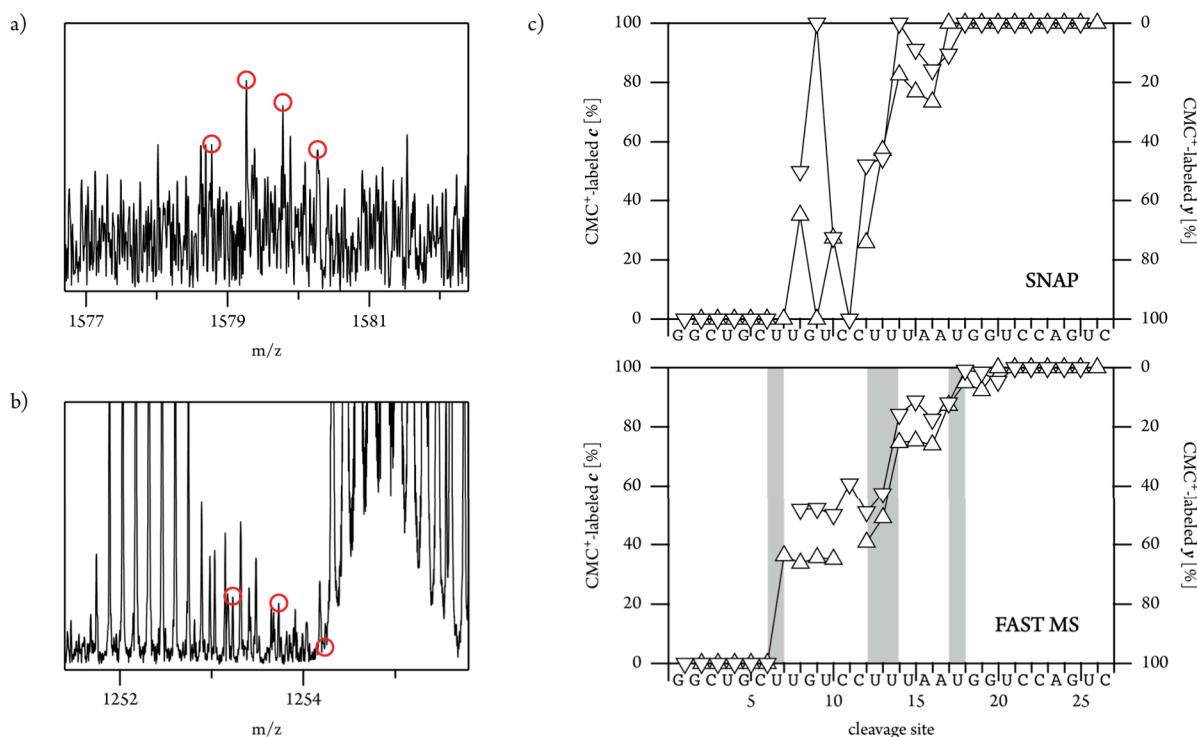
**Figure 4.** Spectrum from CAD of  $(\text{NSR}^{\text{CMC}^+}-8\text{H})^{7-}$  ions with red labels for  $\text{CMC}^+$ -modified fragments and black labels for unmodified fragments; for clarity, only a fraction (~30%) of the ions identified by FAST MS (Table S1) are labeled.

lowered to 0.001%, 1504 ions were detected, but still only 286 could be assigned. Apparently, lowering the threshold leads to an increased number of false positives without significantly increasing the number of assigned ions. Using a comparably low S/N threshold of 2, FAST MS assigned 482 ions (437 correctly from manual inspection; [Table S1](#)) in the same spectrum, which highlights the benefit of searching for isotope distributions calculated from the known sequence.

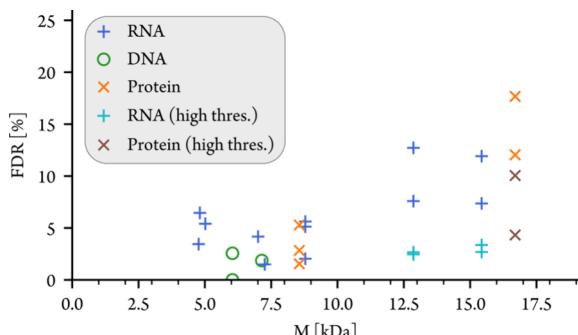
To further test FAST MS, five CAD spectra of modified NSR and RRE-IIB RNA<sup>7</sup> were analyzed by both SNAP and FAST MS (Table S2). On average, FAST MS found more than twice as many *c* and *y* ions as SNAP. We attribute this to the fact that algorithms based on artificial monomer units do not search for specific fragment ions, which makes them more likely to miss ions of lower abundance (Figure 5a), especially in crowded regions of the spectrum (Figure 5b). However, for the analysis of modification profiles or ligand binding, low abundance ions can be very important. More generally, the detection of ions with low S/N is critical when molecules are large,<sup>37</sup> present at low concentrations, and/or when recording time is limited (LC-MS/MS). Figure 5c, in which the site-specific fractions of CMC<sup>+</sup>-modified *c* and *y* fragments are plotted versus cleavage site, illustrates the result of missed isotopic profiles. RNA backbone cleavage produces pairs of complementary fragments, so the fractions of CMC<sup>+</sup>-modified *c* and *y* fragments should add up to 100% for each cleavage site, and they should increase with increasing size of the fragments.<sup>6</sup> Because ions with low S/N

(including CMC<sup>+</sup>-modified  $c_7^{2-}$  and  $c_9^{2-}$ ) were missed by SNAP, the corresponding data show relatively large scatter, making it difficult to accurately assign the modification sites. In contrast, the data obtained by use of FAST MS clearly reveal CMC<sup>+</sup>-modification of U7, U13, U14 and U18.

Manual inspection of the assigned ions in these five CAD spectra (**Table S2**) identified on average 5% as false positives. In the course of a more standardized approach, 20 MS/MS spectra were analyzed in a statistical study similar to a target-decoy approach (**Table S3**). Although FAST MS was not developed for protein identification and database searches, an approach similar to traditional target-decoy studies in bottom-up proteomics<sup>62</sup> can be used to estimate the specificity of the algorithm. For this purpose, each spectrum was analyzed twice by FAST MS, once with the correct sequence that produces both true (TP) and false (FP) positive assignments and once with an incorrect fake sequence that should ideally produce only false positive (FP) assignments. The ratio FP/(TP+FP) then provides a good approximation of the false discovery rate (FDR).<sup>62</sup> The evaluated spectra included 14 CAD spectra of RNA and DNA, an EDD spectrum of DNA, and one CAD spectrum and four ECD spectra of proteins (**Table S3**). **Figure 6** shows that the results of FAST MS are highly accurate for oligonucleotides and proteins with up to ~10 kDa (3.4% FDR on average). The increased FDR for larger molecules is a general phenomenon, not specific to FAST MS, and can be attributed to the increased complexity of their MS/MS spectra as a result of broader isotope



**Figure 5.** Signals of (a) ( $c_9^{CMC+}$ -3H)<sup>2-</sup> and (b) ( $c_7^{CMC+}$ -3H)<sup>2-</sup> in a spectrum from CAD of ( $NSR^{CMC+}$ -8H)<sup>7-</sup> ions (with lower S/N compared to the spectrum in Figure 4 and for which FAST MS identified 173 ions; see Table S2) which were missed by the SNAP algorithm; red circles indicate the isotope peaks calculated by FAST MS. (c) Site-specific fractions of CMC<sup>+</sup>-modified c (upward triangles, left axis) and y (downward triangles, right axis) fragments versus cleavage site from analysis of the same spectrum using SNAP (top) and FAST MS (bottom). Site-specific fractions of CMC<sup>+</sup>-modified c or y fragments were calculated relative to all c or y fragments from a given cleavage site.<sup>6,7,10</sup>

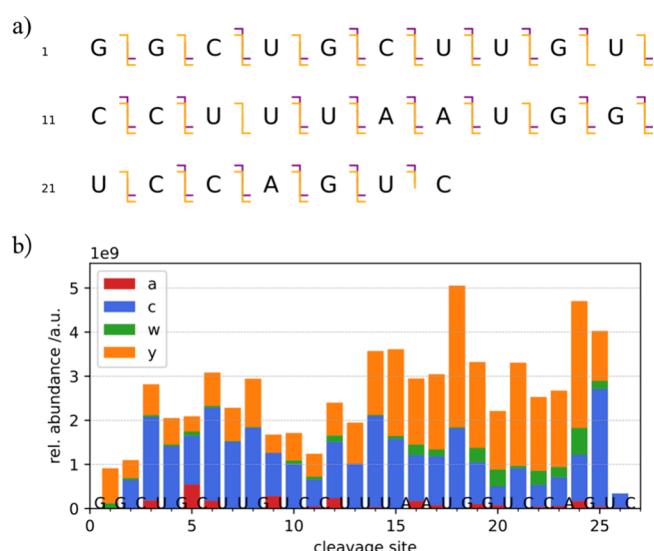


**Figure 6.** False discovery rate (FDR) versus precursor ion mass M from analyses of MS/MS spectra by FAST MS. Thresholds of 0.5 (quality error) and 2 or 3 (S/N) were used for FT-ICR spectra of RNA/DNA or proteins, respectively. Due to the lower resolving power of the QTOF, thresholds of 5 (S/N) and 0.3 (quality error) were used for CpG1018. MS/MS spectra of precursor ion masses above 10 kDa were processed a second time with the higher thresholds ("high thres.") of 5 (S/N) and 0.3 (quality error).

distributions and higher numbers of cleavage sites that lead to more isotope peaks and fragment ions, respectively. Increasing the S/N and quality error thresholds for the larger precursor ions significantly reduced the false discovery rate, while moderately decreasing the number of ions found by ~17%. However, keeping both the FDR low and the number of ions found high for larger molecules would require improved spectral quality and thus spectrometers with higher mass resolving power and sensitivity.

**Analyzing the Ion List.** Processing a spectrum to generate an ion list is typically just the starting point of data evaluation.

For further analysis, FAST MS provides several tools to assist the user in generating meaningful results. For example, the sequence coverage can be determined and graphically displayed in an interactive fragment cleavage map (Figures 7a, S4). Monitoring fragment yields is not only relevant when evaluating instrument or experiment parameters.<sup>59,63</sup> It has also proven critical to



**Figure 7.** FAST MS windows illustrating the statistical analysis of fragment ions from CAD of ( $NSR^{CMC+}$ -8H)<sup>7-</sup> ions: (a) fragment ion map illustrating sequence coverage ( $a$  fragments purple,  $c$  orange,  $w$  purple,  $y$  orange); (b) relative fragment ion abundances versus cleavage site; total fragment yields were 1% ( $a$ ), 10% ( $c$ ), 1% ( $w$ ) and 10% ( $y$ ).

better understand the effect of nucleobases,<sup>60,64</sup> net charge,<sup>60,65</sup> modifications,<sup>12,60,66</sup> and noncovalent interactions<sup>67–69</sup> on RNA or protein dissociation. For these and other purposes, FAST MS can automatically calculate site-specific relative fragment ion abundances (Figure 7b). To investigate the effect of negative or positive charges of modifications, the average charge can also be determined for each fragment.

The most important applications of top-down MS involve the localization of modifications of molecules or ligand binding sites in noncovalent complexes. There are already several programs that can localize modifications qualitatively. However, to characterize mixtures of positional isomers,<sup>12</sup> heterogeneous products from chemical probing reactions<sup>6,70–72</sup> or multiple ligand binding sites,<sup>7,9,10</sup> accurate relative quantification of fragments with and without modification or ligand attachment is required. To this end, FAST MS provides a tool that displays the data graphically and can be used to relatively quantify the site-specific extent of modification or ligand attachment. As in Figure 5, the plot in Figure 8a allows for the relative nucleotide reactivities in a chemical probing experiment to be determined. Moreover, the relative abundances of modified and unmodified fragments (Figure 8b) provide a visualized measure of the significance of the data points in Figure 8a. We used this feature

to test the idea that the fixed positive charge of the CMC<sup>+</sup> modification alters the fragmentation behavior of the RNA but comparing the data from CAD of unmodified (NSR-7H)<sup>7-</sup> ions (Figure 8c) with those of the (NSR<sup>CMC<sup>+</sup>-8H)<sup>7-</sup> ions (Figure 8b) showed no significant differences.</sup>

The above examples show how FAST MS can be used for the characterization of products from RNA chemical probing reactions, and the same approach can be used for various other applications including the identification of ligand binding sites, the localization of hydrogen–deuterium exchange sites and PTMs (e.g., protein deamidations or oxidations), as well as the analysis of synthetic peptide and oligonucleotide impurities. The ability to localize and relatively quantify modified sites can eliminate the need for liquid chromatography (LC) separations, which are challenging for positional isomers of oligonucleotides, and thus the use of expensive and environmentally problematic LC eluents such as trifluoroacetic acid or hexafluoroisopropanol.<sup>17</sup>

## CONCLUSIONS

FAST MS is a user-friendly, cross-platform, and open-source software for the analysis of ESI MS and MS/MS spectra. The approach of searching for defined fragment ions from dissociation of ions of known primary structure instead of using artificial monomer units provides superior sensitivity and accurate relative quantification. FAST MS includes several tools for further analysis of ion lists to extract the desired information from top-down MS experiments. Beyond the analysis of MS/MS spectra, FAST MS incorporates several other functionalities. ESI MS spectra can be processed and analyzed in terms of modifications or ligand binding in thermodynamic studies and for monitoring reaction and ligand binding kinetics. In contrast to most other programs, there are no restrictions concerning molecule type, dissociation methods, ion polarity, monomer units, modifications, ligands, adducts, or isotope labels. Moreover, software parameters and thresholds can be edited within the GUI, which provides the user with maximum flexibility and control. We hope that these features will make data analysis easier for MS users and stimulate new applications of top-down MS.

## ASSOCIATED CONTENT

### Supporting Information

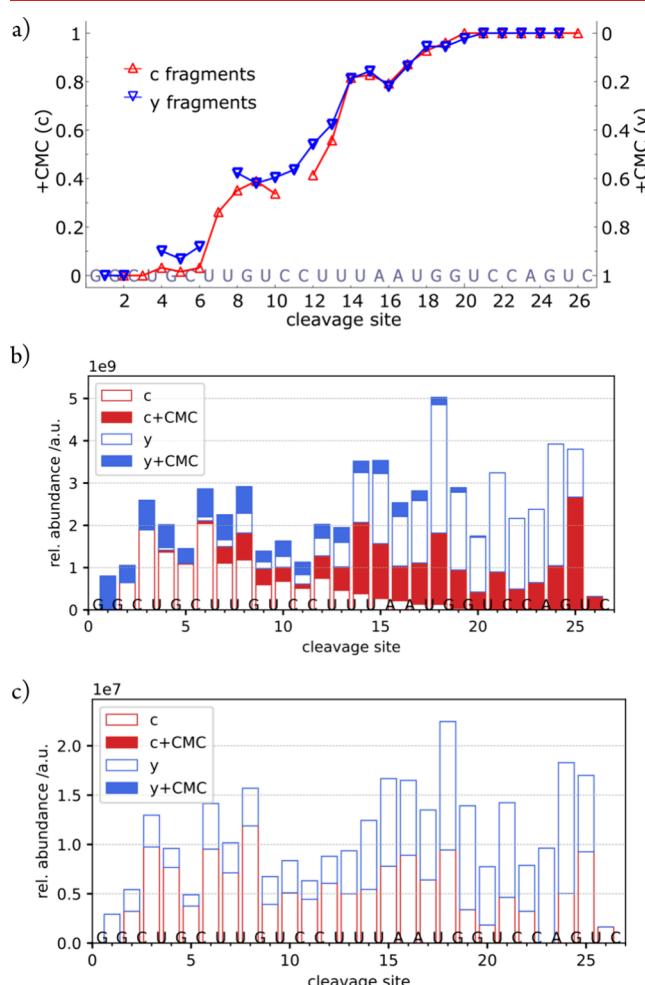
The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/jasms.4c00236>.

Illustration of noise calculation for both FT-ICR and QTOF data, schematic illustration of FAST MS workflow for identifying and quantifying ion signals in a mass spectrum, illustration of the analysis of multiple modifications by FAST MS, fragment ion maps illustrating sequence coverage, list of ions assigned by FAST MS in a spectrum from CAD of (NSR<sup>CMC<sup>+</sup>-8H)<sup>7-</sup> ions, comparison of the performance of FAST MS and SNAP in ion detection, summary of experiments for estimating the FDR of ion assignments by FAST MS (PDF)</sup>

## AUTHOR INFORMATION

### Corresponding Authors

Michael Palasser — Institute of Organic Chemistry and Center for Molecular Biosciences Innsbruck (CMBI), University of Innsbruck, 6020 Innsbruck, Austria; Present



**Figure 8.** FAST MS plots displaying (a) CMC<sup>+</sup>-modified fractions and (b) relative abundances of *c* (red) and *y* (blue) fragments from CAD of singly modified (NSR<sup>CMC<sup>+</sup>-8H)<sup>7-</sup> ions; (c) relative abundances of *c* and *y* fragments of (NSR-7H)<sup>7-</sup> ions versus cleavage site.</sup>

Address: Bachem AG, Hauptstrasse 144, 4416 Bubendorf,  
Email: michael.palasser@bachem.com  
**Kathrin Breuker – Institute of Organic Chemistry and Center  
for Molecular Biosciences Innsbruck (CMBI), University of  
Innsbruck, 6020 Innsbruck, Austria; [ORCID.org/0000-0002-4978-0883](https://orcid.org/0000-0002-4978-0883); Email: kathrin.breuker@uibk.ac.at**

Complete contact information is available at:  
<https://pubs.acs.org/10.1021/jasms.4c00236>

### Author Contributions

The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

### Notes

The authors declare no competing financial interest.

### ACKNOWLEDGMENTS

This research was funded in whole or in part by the Austrian Science Fund (FWF) [grant DOIs 10.55776/P30087 and 10.55776/P36011 to K.B.]. For open access purposes, the author has applied a CC BY public copyright license to any author-accepted manuscript version arising from this submission. M.P. was given time off from his job at Bachem AG to work on the manuscript. We thank Anna Lechner, Giovanni Calderisi, and Patrik Plattner for discussion and critical reading of the manuscript.

### REFERENCES

- (1) Zabrouskov, V.; Ge, Y.; Schwartz, J.; Walker, J. W. Unraveling Molecular Complexity of Phosphorylated Human Cardiac Troponin I by Top Down Electron Capture Dissociation/Electron Transfer Dissociation Mass Spectrometry. *Mol. Cell. Proteomics* **2008**, *7*, 1838–1849.
- (2) Wu, S.; Brown, J. N.; Tolic, N.; Meng, D.; Liu, X.; Zhang, H.; Zhao, R.; Moore, R. J.; Pevzner, P.; Smith, R. D.; Pasa-Tolic, L. Quantitative analysis of human salivary gland-derived intact proteome using top-down mass spectrometry. *Proteomics* **2014**, *14*, 1211–1222.
- (3) Melani, R. D.; Gerbasi, V. R.; Anderson, L. C.; Sikora, J. W.; Toby, T. K.; Hutton, J. E.; Butcher, D. S.; Negrao, F.; Seckler, H. S.; Srzentic, K.; et al. The Blood Proteoform Atlas: A reference map of proteoforms in human hematopoietic cells. *Science* **2022**, *375*, 411–418.
- (4) Taucher, M.; Breuker, K. Characterization of Modified RNA by Top-Down Mass Spectrometry. *Angew. Chem., Int. Ed.* **2012**, *51*, 11289–11292.
- (5) Bereiter, R.; Renard, E.; Breuker, K.; Kreutz, C.; Ennifar, E.; Micura, R. 1-Deazaguanosine-Modified RNA: The Missing Piece for Functional RNA Atomic Mutagenesis. *J. Am. Chem. Soc.* **2022**, *144*, 10344–10352.
- (6) Palasser, M.; Breuker, K. RNA Chemical Labeling with Site-Specific, Relative Quantification by Mass Spectrometry for the Structural Study of a Neomycin-Sensing Riboswitch Aptamer Domain. *ChemPlusChem* **2022**, *87*, No. e202200256.
- (7) Schneeberger, E. M.; Halper, M.; Palasser, M.; Heel, S. V.; Vusurovic, J.; Plangger, R.; Juen, M.; Kreutz, C.; Breuker, K. Native mass spectrometry reveals the initial binding events of HIV-1 rev to RRE stem II RNA. *Nat. Commun.* **2020**, *11*, 5750.
- (8) Schneeberger, E. M.; Breuker, K. Native Top-Down Mass Spectrometry of TAR RNA in Complexes with a Wild-Type tat Peptide for Binding Site Mapping. *Angew. Chem., Int. Ed.* **2017**, *56*, 1254–1258.
- (9) Heel, S. V.; Bartosik, K.; Juen, F.; Kreutz, C.; Micura, R.; Breuker, K. Native Top-Down Mass Spectrometry Uncovers Two Distinct Binding Motifs of a Functional Neomycin-Sensing Riboswitch Aptamer. *J. Am. Chem. Soc.* **2023**, *145*, 15284–15294.
- (10) Heel, S. V.; Juen, F.; Bartosik, K.; Micura, R.; Kreutz, C.; Breuker, K. Resolving the intricate binding of neomycin B to multiple binding motifs of a neomycin-sensing riboswitch aptamer by native top-down mass spectrometry and NMR spectroscopy. *Nucleic Acids Res.* **2024**, *52*, 4691–4701.
- (11) Catherman, A. D.; Skinner, O. S.; Kelleher, N. L. Top Down proteomics: Facts and perspectives. *Biochem. Biophys. Res. Commun.* **2014**, *445*, 683–693.
- (12) Glasner, H.; Riml, C.; Micura, R.; Breuker, K. Label-free, direct localization and relative quantitation of the RNA nucleobase methylations m<sup>6</sup>A, m<sup>5</sup>C, m<sup>3</sup>U, and m<sup>5</sup>U by top-down mass spectrometry. *Nucleic Acids Res.* **2017**, *45*, 8014–8025.
- (13) Leader, B.; Baca, Q. J.; Golan, D. E. Protein therapeutics: a summary and pharmacological classification. *Nat. Rev. Drug Discovery* **2008**, *7*, 21–39.
- (14) Wang, L.; Wang, N.; Zhang, W.; Cheng, X.; Yan, Z.; Shao, G.; Wang, X.; Wang, R.; Fu, C. Therapeutic peptides: current applications and future directions. *Signal Transduct. Target Ther.* **2022**, *7*, 48.
- (15) Roberts, T. C.; Langer, R.; Wood, M. J. A. Advances in oligonucleotide drug delivery. *Nat. Rev. Drug Discovery* **2020**, *19*, 673–694.
- (16) Chen, B.; Brown, K. A.; Lin, Z.; Ge, Y. Top-Down Proteomics: Ready for Prime Time? *Anal. Chem.* **2018**, *90*, 110–127.
- (17) Pourshahian, S. Therapeutic Oligonucleotides, Impurities, Degradants, and Their Characterization by Mass Spectrometry. *Mass Spectrom. Rev.* **2021**, *40*, 75–109.
- (18) Chen, G.; Warrack, B. M.; Goodenough, A. K.; Wei, H.; Wang-Iverson, D. B.; Tymak, A. A. Characterization of protein therapeutics by mass spectrometry: recent developments and future directions. *Drug Discovery Today* **2011**, *16*, 58–64.
- (19) Lian, Z.; Wang, N.; Tian, Y.; Huang, L. Characterization of Synthetic Peptide Therapeutics Using Liquid Chromatography-Mass Spectrometry: Challenges, Solutions, Pitfalls, and Future Perspectives. *J. Am. Soc. Mass Spectrom.* **2021**, *32*, 1852–1860.
- (20) Macias, L. A.; Garcia, S. P.; Back, K. M.; Wu, Y.; Johnson, G. H.; Kathiresan, S.; Bellinger, A. M.; Rohde, E.; Freitas, M. A.; Madsen, J. A. Spacer Fidelity Assessments of Guide RNA by Top-Down Mass Spectrometry. *ACS Cent. Sci.* **2023**, *9*, 1437–1452.
- (21) Santos, I. C.; Brodbelt, J. S. Recent developments in the characterization of nucleic acids by liquid chromatography, capillary electrophoresis, ion mobility, and mass spectrometry (2010–2020). *J. Sep. Sci.* **2021**, *44*, 340–372.
- (22) Heel, S. V.; Breuker, K. Investigating the Intramolecular Competition of Different RNA Binding Motifs for Neomycin B by Native Top-Down Mass Spectrometry. *ChemPlusChem* **2024**, *89*, No. e202400178.
- (23) Horn, D. M.; Zubarev, R. A.; McLafferty, F. W. Automated reduction and interpretation of high resolution electrospray mass spectra of large molecules. *J. Am. Soc. Mass Spectrom.* **2000**, *11*, 320–332.
- (24) Senko, M. W.; Beu, S. C.; McLafferty, F. W. Determination of Monoisotopic Masses and Ion Populations for Large Biomolecules from Resolved Isotopic Distributions. *J. Am. Soc. Mass Spectrom.* **1995**, *6*, 229–233.
- (25) Liu, X.; Inbar, Y.; Dorrestein, P. C.; Wynne, C.; Edwards, N.; Souda, P.; Whitelegge, J. P.; Bafna, V.; Pevzner, P. A. Deconvolution and database search of complex tandem mass spectra of intact proteins: a combinatorial approach. *Mol. Cell. Proteomics* **2010**, *9*, 2772–2782.
- (26) Frank, A. M.; Pesavento, J. J.; Mizzen, C. A.; Kelleher, N. L.; Pevzner, P. A. Interpreting top-down mass spectra using spectral alignment. *Anal. Chem.* **2008**, *80*, 2499–2505.
- (27) Liu, X.; Sirotnik, Y.; Shen, Y.; Anderson, G.; Tsai, Y. S.; Ting, Y. S.; Goodlett, D. R.; Smith, R. D.; Bafna, V.; Pevzner, P. A. Protein identification using top-down. *Mol. Cell. Proteomics* **2012**, *11*, No. M111.008524.
- (28) Fellers, R. T.; Greer, J. B.; Early, B. P.; Yu, X.; LeDuc, R. D.; Kelleher, N. L.; Thomas, P. M. ProSight Lite: graphical software to analyze top-down mass spectrometry data. *Proteomics* **2015**, *15*, 1235–1238.

- (29) Kou, Q.; Xun, L.; Liu, X. TopPIC: a software tool for top-down mass spectrometry-based proteoform identification and characterization. *Bioinformatics* **2016**, *32*, 3495–3497.
- (30) Lantz, C.; Zenaidee, M. A.; Wei, B.; Hemminger, Z.; Ogorzalek Loo, R. R.; Loo, J. A. ClipsMS: An Algorithm for Analyzing Internal Fragments Resulting from Top-Down Mass Spectrometry. *J. Proteome Res.* **2021**, *20*, 1928–1935.
- (31) LeDuc, R. D.; Taylor, G. K.; Kim, Y. B.; Januszyk, T. E.; Bynum, L. H.; Sola, J. V.; Garavelli, J. S.; Kelleher, N. L. ProSight PTM: an integrated environment for protein identification and characterization by top-down mass spectrometry. *Nucleic Acids Res.* **2004**, *32*, W340–W345.
- (32) Zamdborg, L.; LeDuc, R. D.; Glowacz, K. J.; Kim, Y. B.; Viswanathan, V.; Spaulding, I. T.; Early, B. P.; Bluhm, E. J.; Babai, S.; Kelleher, N. L. ProSight PTM 2.0: improved protein identification and characterization for top down mass spectrometry. *Nucleic Acids Res.* **2007**, *35*, W701–W706.
- (33) Guner, H.; Close, P. L.; Cai, W.; Zhang, H.; Peng, Y.; Gregorich, Z. R.; Ge, Y. MASH Suite: a user-friendly and versatile software interface for high-resolution mass spectrometry data interpretation and visualization. *J. Am. Soc. Mass Spectrom.* **2014**, *25*, 464–470.
- (34) Cai, W.; Guner, H.; Gregorich, Z. R.; Chen, A. J.; Ayaz-Guner, S.; Peng, Y.; Valeja, S. G.; Liu, X.; Ge, Y. MASH Suite Pro: A Comprehensive Software Tool for Top-Down Proteomics. *Mol. Cell. Proteomics* **2016**, *15*, 703–714.
- (35) Wu, Z.; Roberts, D. S.; Melby, J. A.; Wenger, K.; Wetzel, M.; Gu, Y.; Ramanathan, S. G.; Bayne, E. F.; Liu, X.; Sun, R.; et al. MASH Explorer: A Universal Software Environment for Top-Down Proteomics. *J. Proteome Res.* **2020**, *19*, 3867–3876.
- (36) Park, J.; Piehowski, P. D.; Wilkins, C.; Zhou, M.; Mendoza, J.; Fujimoto, G. M.; Gibbons, B. C.; Shaw, J. B.; Shen, Y.; Shukla, A. K.; et al. Informed-Proteomics: open-source software package for top-down proteomics. *Nat. Methods* **2017**, *14*, 909–914.
- (37) Compton, P. D.; Zamdborg, L.; Thomas, P. M.; Kelleher, N. L. On the scalability and requirements of whole protein mass spectrometry. *Anal. Chem.* **2011**, *83*, 6868–6874.
- (38) Hogan, J. D.; Klein, J. A.; Wu, J. D.; Chopra, P.; Boons, G. J.; Carvalho, L.; Lin, C.; Zaia, J. Software for Peak Finding and Elemental Composition Assignment for Glycosaminoglycan Tandem Mass Spectra. *Mol. Cell. Proteomics* **2018**, *17*, 1448–1456.
- (39) Kaur, P.; O'Connor, P. B. Algorithms for automatic interpretation of high resolution mass spectra. *J. Am. Soc. Mass Spectrom.* **2006**, *17*, 459–468.
- (40) Ni, J. S.; Pomerantz, S. C.; Rozenski, J.; Zhang, Y. H.; McCloskey, J. A. Interpretation of oligonucleotide mass spectra for determination of sequence using electrospray ionization and tandem mass spectrometry. *Anal. Chem.* **1996**, *68*, 1989–1999.
- (41) Oberacher, H.; Wellenzohn, B.; Huber, C. G. Comparative sequencing of nucleic acids by liquid chromatography-tandem mass spectrometry. *Anal. Chem.* **2002**, *74*, 211–218.
- (42) Rozenski, J.; McCloskey, J. A. SOS: a simple interactive program for ab initio oligonucleotide sequencing by mass spectrometry. *J. Am. Soc. Mass Spectrom.* **2002**, *13*, 200–203.
- (43) Kellersberger, K. A.; Yu, E.; Kruppa, G. H.; Young, M. M.; Fabris, D. Top-down characterization of nucleic acids modified by structural probes using high-resolution tandem mass spectrometry and automated data interpretation. *Anal. Chem.* **2004**, *76*, 2438–2445.
- (44) Yu, E. T.; Hawkins, A.; Kuntz, I. D.; Rahn, L. A.; Rothfuss, A.; Sale, K.; Young, M. M.; Yang, C. L.; Pancerella, C. M.; Fabris, D. The Collaboratory for MS3D: A New Cyberinfrastructure for the Structural Elucidation of Biological Macromolecules and Their Assemblies Using Mass Spectrometry-Based Approaches. *J. Proteome Res.* **2008**, *7*, 4848–4857.
- (45) Kretschmer, M.; Lavine, G.; McArdle, J.; Kuchimanchi, S.; Murugaiah, V.; Manoharan, M. An automated algorithm for sequence confirmation of chemically modified oligonucleotides by tandem mass spectrometry. *Anal. Biochem.* **2010**, *405*, 213–223.
- (46) Hari, Y.; Stucki, S. R.; Nyakas, A.; Blum, L.; Reymond, J. L.; Schürch, S. OMA & OPA - A Software Tool for Mass Spectrometric Sequencing of Nucleic Acids. *Chimia* **2014**, *68*, 86.
- (47) Nyakas, A.; Blum, L. C.; Stucki, S. R.; Reymond, J. L.; Schürch, S. OMA and OPA—software-supported mass spectra analysis of native and modified nucleic acids. *J. Am. Soc. Mass Spectrom.* **2013**, *24*, 249–256.
- (48) Nakayama, H.; Akiyama, M.; Taoka, M.; Yamauchi, Y.; Nobe, Y.; Ishikawa, H.; Takahashi, N.; Isobe, T. Ariadne: a database search engine for identification and chemical analysis of RNA using tandem mass spectrometry data. *Nucleic Acids Res.* **2009**, *37*, No. e47.
- (49) Yu, N.; Lobue, P. A.; Cao, X.; Limbach, P. A. RNAModMapper: RNA Modification Mapping Software for Analysis of Liquid Chromatography Tandem Mass Spectrometry Data. *Anal. Chem.* **2017**, *89*, 10744–10752.
- (50) Wein, S.; Andrews, B.; Sachsenberg, T.; Santos-Rosa, H.; Kohlbacher, O.; Kouzarides, T.; Garcia, B. A.; Weisser, H. A computational platform for high-throughput analysis of RNA sequences and modifications by mass spectrometry. *Nat. Commun.* **2020**, *11*, 926.
- (51) Harris, C. R.; Millman, K. J.; van der Walt, S. J.; Gommers, R.; Virtanen, P.; Cournapeau, D.; Wieser, E.; Taylor, J.; Berg, S.; Smith, N. J.; et al. Array programming with NumPy. *Nature* **2020**, *585*, 357–362.
- (52) Lam, S. K.; Pitrou, A.; Seibert, S. Numba: a LLVM-based Python JIT compiler. In *Proceedings of the Second Workshop on the LLVM Compiler Infrastructure in HPC*; Association for Computing Machinery, 2015. DOI: [10.1145/2833157.2833162](https://doi.org/10.1145/2833157.2833162).
- (53) McKinney, W. *Data Structures for Statistical Computing in Python*; 2010. DOI: [10.25080/Majora-92bf1922-00a](https://doi.org/10.25080/Majora-92bf1922-00a).
- (54) Virtanen, P.; Gommers, R.; Oliphant, T. E.; Haberland, M.; Reddy, T.; Cournapeau, D.; Burovski, E.; Peterson, P.; Weckesser, W.; Bright, J.; et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat. Methods* **2020**, *17*, 261–272.
- (55) Rockwood, A. L.; Palmblad, M. Isotopic distributions. *Methods Mol. Biol.* **2013**, *1007*, 65–99.
- (56) Sadygov, R. G. Poisson Model To Generate Isotope Distribution for Biomolecules. *J. Proteome Res.* **2018**, *17*, 751–758.
- (57) NIST/SEMATech e-Handbook of Statistical Methods. <https://www.itl.nist.gov/div898/handbook/eda/section3/eda35h1.htm> (accessed 15 Jun 2022).
- (58) Tromp, J. M.; Schürch, S. Gas-phase dissociation of oligoribonucleotides and their analogs studied by electrospray ionization tandem mass spectrometry. *J. Am. Soc. Mass Spectrom.* **2005**, *16*, 1262–1268.
- (59) Taucher, M.; Rieder, U.; Breuker, K. Minimizing Base Loss and Internal Fragmentation in Collisionally Activated Dissociation of Multiply Deprotonated RNA. *J. Am. Soc. Mass Spectrom.* **2010**, *21*, 278–285.
- (60) Riml, C.; Glasner, H.; Rodgers, M. T.; Micura, R.; Breuker, K. On the mechanism of RNA phosphodiester backbone cleavage in the absence of solvent. *Nucleic Acids Res.* **2015**, *43*, 5171–5181.
- (61) Gao, Y.; McLuckey, S. A. Collision-induced dissociation of oligonucleotide anions fully modified at the 2'-position of the ribose: 2'-F-H and 2'-F-/H-/OMe mix-mers. *J. Mass Spectrom.* **2012**, *47*, 364–369.
- (62) Elias, J. E.; Gygi, S. P. Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. *Nat. Methods* **2007**, *4*, 207–214.
- (63) Calderisi, G.; Glasner, H.; Breuker, K. Radical Transfer Dissociation for De Novo Characterization of Modified Ribonucleic Acids by Mass Spectrometry. *Angew. Chem., Int. Ed.* **2020**, *59*, 4309–4313.
- (64) Fuchs, E.; Falschlunger, C.; Micura, R.; Breuker, K. The effect of adenine protonation on RNA phosphodiester backbone bond cleavage elucidated by deaza-nucleobase modifications and mass spectrometry. *Nucleic Acids Res.* **2019**, *47*, 7223–7234.
- (65) McLuckey, S. A.; Vaidyanathan, G.; Habibigoudarzi, S. Charged vs Neutral Nucleobase Loss from Multiply-Charged Oligonucleotide Anions. *J. Mass Spectrom.* **1995**, *30*, 1222–1229.

- (66) Dang, X.; Singh, A.; Spetman, B. D.; Nolan, K. D.; Isaacs, J. S.; Dennis, J. H.; Dalton, S.; Marshall, A. G.; Young, N. L. Label-Free Relative Quantitation of Isobaric and Isomeric Human Histone H2A and H2B Variants by Fourier Transform Ion Cyclotron Resonance Top-Down MS/MS. *J. Proteome Res.* **2016**, *15*, 3196–3203.
- (67) Schennach, M.; Schneeberger, E. M.; Breuker, K. Unfolding and Folding of the Three-Helix Bundle Protein KIX in the Absence of Solvent. *J. Am. Soc. Mass Spectrom.* **2016**, *27*, 1079–1088.
- (68) Schennach, M.; Breuker, K. Proteins with Highly Similar Native Folds Can Show Vastly Dissimilar Folding Behavior When Desolvated. *Angew. Chem., Int. Ed.* **2014**, *53*, 164–168.
- (69) Breuker, K.; Brüschweiler, S.; Tollinger, M. Electrostatic stabilization of a native protein structure in the gas phase. *Angew. Chem., Int. Ed.* **2011**, *50*, 873–877.
- (70) Chen, J.; Cui, W.; Giblin, D.; Gross, M. L. New protein footprinting: fast photochemical iodination combined with top-down and bottom-up mass spectrometry. *J. Am. Soc. Mass Spectrom.* **2012**, *23*, 1306–1318.
- (71) Polak, M.; Yassaghi, G.; Kavan, D.; Filandr, F.; Fiala, J.; Kukacka, Z.; Halada, P.; Loginov, D. S.; Novak, P. Utilization of Fast Photochemical Oxidation of Proteins and Both Bottom-up and Top-down Mass Spectrometry for Structural Characterization of a Transcription Factor-dsDNA Complex. *Anal. Chem.* **2022**, *94*, 3203–3210.
- (72) Polak, M.; Palasser, M.; Kadek, A.; Kavan, D.; Wootton, C. A.; Delsuc, M. A.; Breuker, K.; Novak, P.; van Agthoven, M. A. Top-Down Proteoform Analysis by 2D MS with Quadrupolar Detection. *Anal. Chem.* **2023**, *95*, 16123–16130.