Automated molecular formula determination by tandem mass spectrometry (MS/MS)
Jarussophon, Suwatchai; Acoca, Stephane; Gao, Jin-Ming; Deprez, Christophe; Kiyota, Taira; Draghici, Cristina; Purisima, Enrico; Konishi, Yasuo
Automated molecular formula determination by tandem mass spectrometry (MS/MS)†

Suwatchai Jarussophon,ab Stephane Acoca,ac Jin-Ming Gao,ad Christophe Deprez,a Taira Kiyota,‡a Cristina Draghici,‡a Enrico Purisima* and Yasuo Konishi*aa

Received 20th October 2008, Accepted 12th January 2009
First published as an Advance Article on the web 5th February 2009
DOI: 10.1039/b818398h

Automated software was developed to analyze the molecular formula of organic molecules and peptides based on high-resolution MS/MS spectroscopic data. The software was validated with 96 compounds including a few small peptides in the mass range of 138–1569 Da containing the elements carbon, hydrogen, nitrogen and oxygen. A Micromass Waters Q-TOF Ultima Global mass spectrometer was used to measure the molecular masses of precursor and fragment ions. Our software assigned correct molecular formulas for 91 compounds, incorrect molecular formulas for 3 compounds, and no molecular formula for 2 compounds. The obtained 95% success rate indicates high reliability of the software. The mass accuracy of the precursor ion and the fragment ions, which is critical for the success of the analysis, was high, i.e. the accuracy and the precision of 850 data were 0.0012 Da and 0.0016 Da, respectively. For the precursor and fragment ions below 500 Da, 60% and 90% of the data showed accuracy within ±0.001 Da and ±0.002 Da, respectively. The precursor and fragment ions above 500 Da showed slightly lower accuracy, i.e. 40% and 70% of them showed accuracy within ±0.001 Da and ±0.002 Da, respectively. The molecular formulas of the precursor and the fragments were further used to analyze possible mass spectrometric fragmentation pathways, which would be a powerful tool in structural analysis and identification of small molecules. The method is valuable in the rapid screening and identification of small molecules such as the dereplication of natural products, characterization of drug metabolites, and identification of small peptide fragments in proteomics. The analysis was also extended to compounds that contain a chlorine or bromine atom.

Introduction

Over the past decades, the major sources of marketed drugs have been natural products, or their semi-synthetic derivatives. Natural products provide larger structural diversity than combinatorial chemistry products and offer significant opportunities for finding novel lead compounds.1–5 Each year, a large number of new natural products are discovered and fully characterized; however, during the courses of isolation, characterization and identification, natural products chemists have faced increasing problems of replication, i.e. re-discovery of known natural products.6,7

†Biotechnology Research Institute, National Research Council Canada, 6100 Royalmount Avenue, Montréal, Québec, Canada H4P 2R2. E-mail: yasuo.konishi@cnrc-nrc.gc.ca; Fax: +1 514 496 5243; Tel: +1 514 496 6339
‡National Nanotechnology Center, National Science and Technology Development Agency, 130 Thailand Science Park, Paholyothin Rd., Klong Luang, Pathumthani 12120, Thailand. E-mail: suwatchai@nanotec.or.th
aDepartment of Biochemistry, McIntyre Medical Sciences Building, McGill University, 3655 Promenade Sir William Osler, Montréal, Québec, Canada H3G 1Y6
bNatural Products Resource Research Centre, College of Science, Northwest A & F University, Yangling, Shaanxi 712100, PR China
cBioactive Natural Product Database, Chapman & Hall’s Dictionary of Natural Products and Natural Products Alert (NAPRALERT). There have been considerable developments of analytical separation techniques such as GC, LC and CE, and of spectroscopic characterization techniques such as PDA, IR, NMR, and MS. Hyphenated techniques couple the separation techniques with online spectroscopic characterization, e.g. LC-MS, GC-MS, CE-MS, LC-UV and LC-NMR. They are expected to resolve the complexity of the natural product extracts. Recent advances of hyphenated techniques and their applications have been reported by several research laboratories.15–18 In addition, multiple combinations of the characterization techniques have been developed, e.g. LC-PDA-MS,19 LC-NMR-MS,20,21 LC-SPE-NMR,22 LC-PDA-IR-NMR-MS,23 and 2D-LC(IEC-RP)-MS.24 These improve the sample separation, structural characterization and elucidation, and also the detection of valuable minor components in natural sources.

Due to high productivity and sensitivity, mass spectrometry has become the most powerful and essential technique providing...
critical information in many phases of drug discovery and development. Examples are structural characterization and identification,25,26 quantitation,27 high-throughput screening,28,29 proteomics,30,31 metabolomics,32–34 metabonomics,35–41 and dereplication of natural products. In the area of dereplication, MS delivers the molecular mass information that can be used as a search query in almost all databases of small molecules.32,41

Accurate mass obtained from high-resolution MS is commonly used to search for candidates in literature databases.44 Unfortunately, in general, several molecular formulas (MFs) are compatible with the observed molecular mass within the experimental accuracy, resulting in a large set of compounds, most of which are false positives. MS/MS measurements often provide more information of some sub-structures, or fingerprints of sub-structures.45,46 There are few databases that provide MS/MS data such as NIST/EPA/NIH Mass Spectral Library which contains 14 802 MS/MS spectra. They may become practical and useful when more MS/MS data are accumulated and systematically analyzed.

The dereplication process starts by using the information stored in natural product databases. The molecular formula (MF) is one of the most valuable pieces of information as it is available in any natural product database and is independent of the source, of the sample preparation, and of the conditions of the measurements. Conventional elemental composition analysis is typically achieved by high-resolution MS47,48 such as magnetic sector MS, time-of-flight MS (TOF-MS)49 and Fourier transform MS (FTMS).50 These instruments provide a set of MFs that can be further narrowed down in combination with NMR data. However, NMR is much less sensitive than MS, requiring a large quantity of purified sample. Thus, the combination of MS and NMR practically brings little contribution to dereplication. It should be mentioned that the number of possible MFs exponentially increases with the size of the molecules, whereas there has been no drastic progress to improve accurate mass determination.51,52

In our previous paper,53 we analyzed MF and fragmentation pathways of small molecules based on their accurate MS, MS/MS, and MS/MS/MS data. The method provided detailed information on the structure and sub-structures and their fragmentation pathways. The method is useful when a specific molecule is targeted for detailed structural analysis. However, the measurements and the analysis were not automated. Thus, we have now developed simple, automated, and productive software, which determines MF of small molecules and their fragments based on the accurate MS and MS/MS data.

Experimental

Materials

All compounds used in our study were randomly selected from an organo-chemical drug library (Negwer’s database) and those that are commercially available. [Glu]-Fibrinopeptide B was purchased from Sigma (Oakville, ON, Canada) and used as a reference compound for the calibration of the mass spectrometer. All reagents were used without further purification. Water and acetonitrile are of HPLC grade and were purchased from Anachemia (Lachine, QC, Canada) and J. T. Baker (Phillipsburg, NJ), respectively. Formic acid was purchased from Fluka (Oakville, ON, Canada) and was used to aid the positive ion electrospray ionization process. All solvents were degassed for at least 30 minutes before use.

Instrumentation

All MS and MS/MS measurements were performed in positive ion electrospray mode (+ESI) on a Micromass Waters Q-TOF Ultima Global mass spectrometer equipped with a Z-spray ion source and NanoLockSpray (Waters, Mississauga, ON, Canada) source. The m/z was acquired within the mass range of 50–990 m/z for small organic molecules and 100–1990 m/z for organic molecules with >900 Da molecular mass and small peptides. The acquisition time per spectrum was set to 1 s, inter-scan delay was set to 0.1 s, with the lock spray frequency set to 4 s.54–56 The mass spectrometer was set up in V mode with instrument resolution between 9000 and 10 000 based on FWHM. The source and desolvation temperature were set to 80 and 150 °C, respectively. The TOF was operated at an acceleration voltage of 9.1 kV, a cone voltage of 100 V, an RF lens of 45 V, and a capillary voltage of 3.8 kV. Operating parameters of the ESI interface were optimized by infusing standard solutions of [Glu]-fibrinopeptide B, 100 nM in a solution of water–acetonitrile 50 : 50 (v/v) with 0.1% formic acid at a flow rate of 1.0 μL/min. The instrument was carefully calibrated such that the error of the MS/MS fragments of [Glu]-fibrinopeptide B was less than 4 ppm. All measurements were performed at room temperature. The MassLynx 4.0 chromatographic software was used for instrument control and data analysis.

MS/MS experiments

A precursor ion of interest was selected at the quadrupole (Q1). It was fragmented in the hexapole collision cell with argon collision gas. Fragment ions were measured to obtain their m/z and peak area. The collision energy was adjusted for each compound typically from 5 to 40 eV in order to maintain the precursor ion peak in the range of 35–100 counts/scan, while maximizing the peak areas of the product ions. Sometimes, high collision energy was used just to enhance the peak areas of the fragment ions at low m/z. Fragment ions of 20–400 counts/scan were used for most of the analyses. The acquired mass spectra were accumulated for at least 2 min. The mass measurements are most accurate when analyte/lock mass intensity ratio is between 0.5 and 2.0.57,58 In a few cases, fragment ions lower than 20 counts/scan were used for the analysis after accumulating many scans. Typically, analytes were dissolved in water–acetonitrile 50 : 50 (v/v) with 0.1% formic acid and directly infused to the mass spectrometer using a Harvard syringe pump or autosampler direct injection. For the reference channel, freshly prepared [Glu]-fibrinopeptide B (~1 μM) in water–acetonitrile 50 : 50 (v/v) with 0.1% formic acid was continuously infused to maintain the constant concentration of reference solution. Both analyte and reference channels were controlled by NanoLockSpray. The TOF mass correction (accurate mass measurement) parameters were performed with the following parameter sets: no background subtraction;59 Savitsky–Golay’s smooth type; smooth window 3 channels; number of smooths 1;60 minimum peak
width at half-height 4 channels; centroid top 60%; the dead-time correction was turned on. Spectral intensity cut-off threshold setting of 0.1–1.0% was used to simplify and reduce the number of peaks analyzed which have not sufficient intensity to get good accuracy.1,6 The TOF transform was used to exclude all isotope peaks.

The algorithm of molecular formula analysis (MFA)

The forward and reverse MFA algorithms have been described previously.3 Briefly, the MS/MS experiment of a precursor ion A generates several fragment ions, A1, A2, A3, A4, A5, A6, A7, A8, and A9 (Table 1). All possible neutral fragments N1 (i = 1–9) and N1j (i = 1–8; j = 2–9; j > i) are listed in Table 1, i.e. N1 (i = 1–9) are generated from A and N1j (i = 1–8; j = 2–9; j > i) are possibly generated from Ai (i = 1–8). H+ (= 1.0078 Da) is added artificially as the smallest product ion to obtain the corresponding neutral product N, N1H, ..., N9H in Table 1.

(1) The first step of the analysis is designated as ‘Forward Analysis’. Briefly, accurate mass measurement uniquely determines the MF of some small fragments. The MFs of these small fragments are then added up sequentially to determine the MF of the precursor ion A. More specifically, the MFA was carried out for the neutral losses N/H (i = 1–9), N, and N9 (i = 1–8). Two restrictions were applied in the search. One is the molecular size of the neutral loss, which is typically limited to 200–400 Da, preferably 200 Da, and the other is the error cut-off, which is 0.002–0.003 Da, preferably 0.002 Da. These restrictions reduce CPU time and enhance the identification of a unique MF, respectively. The neutral molecules that have been uniquely identified are highlighted in bold (N9H, N8H, N7H, ..., etc.). It should be emphasized that all N9 (i = 1–8) are simply listed to fill Table 1 and some of them may not exist mathematically or physically. The MF of fragment ion A9 is assigned from the added MFs (N9H + H+). The observed m/z value of A9 is then replaced with the one calculated from the assigned MF. The molecular masses of N9 (i = 1–8) are also replaced accordingly. Similarly, the MF of fragment ion A8 is assigned from (N9H + H+) and (N99 + A9). The process continues up to the assignment of the MF of A from (N1 + A1), (N2 + A2), (N3 + A3), (N4 + A4), and (N5 + A5). In most of the analysis, the MF of A is uniquely assigned; however, sometimes, two or more MFs are assigned for A. All of the MFs assigned are further examined in the next step, which is designated as ‘Reverse Analysis’, where each of the MFs of the fragment ions and neutral losses assigned in ‘Forward Analysis’ is re-examined.

(2) The second step is designated as ‘Reverse Analysis’. First, the observed m/z value of the precursor ion is replaced with the m/z value calculated from the assigned MF. The MFs of N1 (i = 1–9), in which each element is restricted not to exceed that of A, are analyzed with a cut-off error of 0.002–0.003 Da, but without limitation of the molecular size. The MF of A1 is determined as the difference of those of A and N1. The observed m/z value of A1 is replaced with that of the calculated value. The molecular masses (MM) of N1i (i = 1–9) are also replaced accordingly. Similarly, the MF of the fragment ion A2 is assigned from (A – N2), and (A1 – N12). The process continues until the MF of A9 is assigned from (A – N9), (A1 – N19), (A2 – N29), (A3 – N39), (A4 – A49), (A5 – A59), (A6 – N69), (A7 – N79), and (A8 – N89). If they are not consistent, the MF of Ai is assigned by taking the one most frequently assigned and using it in the following steps. Table 2 shows the outcome of the reverse analysis of brefeldin A.

(3) The third step is designated as ‘Least-squares Index’, where statistical verification is introduced on the outcome of ‘Reverse Analysis’. This step is designed to select a correct MF out of multiple MFs that are occasionally found in ‘Forward Analysis’. It should be recalled that ‘Reverse Analysis’ is performed for all of the MFs derived from ‘Forward Analysis’. The MFA of brefeldin A is an example that resulted in two MFs of C16H30O4 and C17H30N4 in ‘Forward Analysis’. Their ‘Reverse Analysis’ gives quite different MFs of the fragment ions and neutral losses. Tables 2 and 3 show both correct and incorrect MFA in the reverse analysis (the formats of precursor ion, fragment ions and neutral losses are the same as those in Table 1). The two tables have quite different characteristics. The one with the correct MF of the precursor ion produces the MFs of all fragment ions and of most of the potential neutral losses. On the other hand, the one with incorrect MF of the precursor ion fails to assign the MF of four fragment ions and of several neutral losses. In order to quantify the different characteristics of the tables, a least-squares index was introduced in the automated evaluation software to evaluate the difference numerically. The index is based on the observation (illustrated by Table 2) that the number of possible neutral losses, NL_calc, associated with the fragment ions linearly increases as the size of the fragment ions decreases. This gives a range of values for NL_calc. Not all of these neutral losses are observed in the table, i.e. NL_obs is not always equal to NL_calc. Empirically, we observe that the correlation

<table>
<thead>
<tr>
<th>Precursor</th>
<th>A1</th>
<th>A2</th>
<th>A3</th>
<th>A4</th>
<th>A5</th>
<th>A6</th>
<th>A7</th>
<th>A8</th>
<th>A9</th>
<th>H+</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>N1</td>
<td>N2</td>
<td>N3</td>
<td>N4</td>
<td>N5</td>
<td>N6</td>
<td>N7</td>
<td>N8</td>
<td>N9</td>
<td>N</td>
</tr>
<tr>
<td>A1</td>
<td>N12</td>
<td>N13</td>
<td>N14</td>
<td>N15</td>
<td>N16</td>
<td>N17</td>
<td>N18</td>
<td>N19</td>
<td>N1H</td>
<td>N1H</td>
</tr>
<tr>
<td>A2</td>
<td>N23</td>
<td>N24</td>
<td>N34</td>
<td>N35</td>
<td>N36</td>
<td>N37</td>
<td>N38</td>
<td>N39</td>
<td>N3H</td>
<td>N3H</td>
</tr>
<tr>
<td>A3</td>
<td></td>
<td></td>
<td></td>
<td>N45</td>
<td>N46</td>
<td>N47</td>
<td>N48</td>
<td>N49</td>
<td>N4H</td>
<td>N4H</td>
</tr>
<tr>
<td>A4</td>
<td></td>
<td></td>
<td></td>
<td>N56</td>
<td>N57</td>
<td>N58</td>
<td>N59</td>
<td>N5H</td>
<td>N5H</td>
<td>N5H</td>
</tr>
<tr>
<td>A5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>N67</td>
<td>N68</td>
<td>N69</td>
<td>N6H</td>
<td>N6H</td>
<td>N6H</td>
</tr>
<tr>
<td>A6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>N78</td>
<td>N79</td>
<td>N7H</td>
<td>N7H</td>
<td>N7H</td>
</tr>
<tr>
<td>A7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>N89</td>
<td>N8H</td>
<td>N8H</td>
<td>N8H</td>
</tr>
<tr>
<td>A8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>N9H</td>
<td>N9H</td>
<td>N9H</td>
</tr>
<tr>
<td>A9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>N9H</td>
<td>N9H</td>
</tr>
</tbody>
</table>

Table 1 Potential neutral losses in the MS/MS experiment in forward MFA

692 | Analyst, 2009, 134, 690–700

This journal is © The Royal Society of Chemistry 2009
The analysis of glucose neutral loss (C\textsubscript{enriched fragments tends to end up with two MFs. For example, Da) are very close, the MFA of nitrogen-enriched and oxygen-enriched each fragment ion.

Nitrogen-enriched or oxygen-enriched compounds

Bases such as adenine and cytosine are typical nitrogen-enriched groups, and saccharides are typical oxygen-enriched groups. As the molecular masses of CN\textsubscript{4} (68.0122 Da) and H\textsubscript{2}O\textsubscript{4} (68.0109 Da) are very close, the MFA of nitrogen-enriched and oxygen-enriched fragments tends to end up with two MFs. For example, the analysis of glucose neutral loss (C\textsubscript{10}H\textsubscript{18}O\textsubscript{4}, 180.0633 Da) also assigns a nitrogen-enriched neutral loss (C\textsubscript{10}H\textsubscript{14}O\textsubscript{3}, 180.0647 Da) such as p-xanthine, theophylline, and NSC265259 with a molecular mass difference of only 0.0013 Da. In this case, they may be distinguished by another dehydrated neutral loss (C\textsubscript{10}H\textsubscript{12}O\textsubscript{5}, 162.0528 Da), i.e. dehydration should be observed for glucose, but not for the base. In order to minimize the failure of our MFA, a few commonly observed saccharides and bases are pre-assigned in the software and the presence/absence of the dehydrated fragment is manually confirmed later. However, exceptional cases can still occur. For example, the analysis of streptomycin failed as it contains both nitrogen-enriched (C\textsubscript{10}H\textsubscript{12}N\textsubscript{4}O\textsubscript{4}) and oxygen-enriched (C\textsubscript{10}H\textsubscript{12}NO\textsubscript{4} and C\textsubscript{10}H\textsubscript{12}O\textsubscript{4}) sub-structures. Since none of them are commonly present in small molecules, they are not pre-assigned, resulting in no unique MF of the precursor molecule. Nucleoside analogs containing bases and ribose analogs may generally have the same problem.

Results and discussion

Risk of assigning incorrect MF

The first ‘Forward Analysis’ step is typically performed with a molecular mass cut-off of 200 Da and a mass accuracy of 0.002 Da. We have applied ‘Forward Analysis’ to 96 small molecules (138–1569 Da) and observed the following:

(1) ‘Forward Analysis’ of 86 out of 96 compounds (90%) resulted in a unique MF for each precursor molecule. Among them, 83 were assigned the correct MF and 3 were assigned the incorrect MF (97% success rate).
(2) No MF was assigned for two of the compounds (troleandomycin, 813.4511 Da; streptomycin, 581.2657 Da). In the case of troleandomycin, only a few fragment ions were detected. The MS/MS analysis does not work when there are not enough detected peaks of fragment ions. However, for streptomycin, sufficient numbers of fragment peaks were observed and yet the analysis failed. This suggests that our analysis seems to be rather weak in analyzing sugar-containing compounds, requiring further improvement of the method.

(3) Two MFs were assigned for each of the seven compounds. The ‘Least-squares Index’ on the outcome of ‘Reverse Analysis’ assigned the correct MF in all seven cases.

(4) Three MFs were assigned for protoveratrine A (793.4249 Da). The ‘Least-squares Index’ on the outcome of ‘Reverse Analysis’ selected the correct MF.

Thus, the integrated approach reduced the risk of assigning an incorrect MF to 3% (3 out of 96 compounds).

Mass accuracy

The accuracy of the data is crucial for the success of the MFA. The mass spectrometer is calibrated and tuned for peak shape (symmetry and tailing), resolving power (9000–10 000), and ion abundance in the mass ranges of interest. Peaks for which ion abundance was insufficient (<20 counts/scan) or exceeding saturation (400 counts/scan) tended to have large errors, even after accumulation of several scans, and were excluded from the analysis. Alternatively, an isotope peak could be used instead of the over-saturated monoisotope peak. The MFA used a total of 850 peaks of the 96 compounds. After assigning the MF to all of the 850 peaks, the mass accuracy of those peaks was 0.0012 Da with 0.0016 Da precision. Out of them, the mass accuracy of the 734 peaks (86% of 850 peaks) was ±0.002 Da, while 55, 36, 18 and 7 peaks had less accuracy of 0.002–0.003 Da, 0.003–0.005 Da, 0.005–0.010 Da, and 0.010–0.014 Da, respectively. The peaks at high m/z tend to be less accurate. Only 40% and 70% of the peaks above 500 Da achieved the accuracy of ±0.001 Da and ±0.002 Da. In contrast, 60% and 90% of the peaks below 500 Da achieved the accuracy of ±0.001 Da and ±0.002 Da, respectively. There is no need to use many peaks in the analysis. Instead, it is important that the neighboring fragment ions in the MFA are within 200 Da.

Fragmentation pathways of brefeldin A

The MS/MS spectrum of brefeldin A is shown in Fig. 1. Fig. 2 shows plausible fragmentation pathways analyzed based on the fragment ions used in the MFA (Table 2). Further MS² (n ≥ 3) studies are required to validate them. Also, the fragment ions that are not used in the MFA would provide more detailed fragmentation pathways. Nevertheless, these plausible fragmentation pathways provide useful information on the structure and sub-structure of the target molecules.

McLafferty rearrangement is introduced to explain three consecutive water losses from [C₁₆H₂₅O₃]⁺ (m/z 281) to [C₁₆H₂₃O₂]⁺ (m/z 263), [C₁₆H₂₃O₂]⁻ (m/z 245), and [C₁₆H₂₀O]⁺ (m/z 227) and the loss of CO from [C₁₆H₂₅O₃]⁺ (m/z 245) to [C₁₆H₂₀O]⁺ (m/z 217) and from [C₁₆H₂₃O]⁻ (m/z 227) to [C₁₅H₂₀O]⁻ (m/z 199). Further fragmentations are described in two pathways from [C₁₆H₂₀O]⁻ (m/z 227) and from [C₁₅H₂₃O]⁻ (m/z 217) in Fig. 2. Fang et al. reported other fragmentation pathways of brefeldin A. The MFs of fragment ions at m/z 227, 217, 199, and 185 are different from ours. If we use their MFA, the fragment peaks at 227, 217, 199, and 185 m/z have errors of 0.0005, 0.0370, 0.0368 and 0.0373 Da, respectively, which are highly unlikely based on our mass measurement accuracy.

Molecules with single structural domain

The majority of small molecules consist of a single core structure. An example is prazosin (C₁₉H₂₃N₃O₄, 383.1594 Da) for which the result of the MFA is shown in Table 4. The MF of prazosin is correctly assigned to C₁₉H₂₃N₃O₄. Some of the neutral losses, which are underlined, are easily assigned to the losses of methane (CH₄), water (H₂O), and furan (C₄H₄O), suggesting sub-structures of CH₃– (O or N)– and furan. Since the fragmentation producing free radicals hardly occurs in the ESI Q-TOF instrument, the software does not accept the free radical fragmentations. As a result, MF, of the peak at m/z 232, which was incorrectly assigned to C₁₅H₁₄NO₃⁺, required manual correction to a free radical of C₁₁H₁₁N₃O₂⁺ based on the structure of prazosin (bold in Table 4). Among the fragmentations of the 96 compounds, this was the only case that produced a free radical. It should be emphasized that the software works even if the MFs of a few fragment ions are incorrectly assigned. As the observed m/z values of one or a few peaks may have errors larger than the cutoff error, the software was designed to tolerate such errors.
Table 4  MFA of prazosin

<table>
<thead>
<tr>
<th>MM_logo</th>
<th>383.1581</th>
<th>367.1291</th>
<th>365.1492</th>
<th>349.1183</th>
<th>246.1096</th>
<th>231.0888</th>
<th>230.0805</th>
<th>163.0638</th>
<th>137.0474</th>
</tr>
</thead>
<tbody>
<tr>
<td>C_10H_12N_2O_4</td>
<td>—</td>
<td>CH_4</td>
<td>H_2O</td>
<td>CH_2O</td>
<td>C_9H_11N_2O_4</td>
<td>C_10H_12NO_4</td>
<td>C_9H_11NO_3</td>
<td>C_9H_11NO_3</td>
<td>C_9H_11NO_3</td>
</tr>
<tr>
<td>C_10H_12N_2O_4</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>C_9H_11NO_3</td>
<td>C_10H_12NO_4</td>
<td>C_9H_11NO_3</td>
<td>C_9H_11NO_3</td>
<td>C_9H_11NO_3</td>
</tr>
<tr>
<td>C_9H_11N_2O_4</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>C_9H_11NO_3</td>
<td>C_10H_12NO_4</td>
<td>C_9H_11NO_3</td>
<td>C_9H_11NO_3</td>
<td>C_9H_11NO_3</td>
</tr>
<tr>
<td>C_8H_11NO_2</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>C_9H_11NO_3</td>
<td>C_10H_12NO_4</td>
<td>C_9H_11NO_3</td>
<td>C_9H_11NO_3</td>
<td>C_9H_11NO_3</td>
</tr>
<tr>
<td>C_12H_14N_2O_4</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>C_9H_11NO_3</td>
<td>C_10H_12NO_4</td>
<td>C_9H_11NO_3</td>
<td>C_9H_11NO_3</td>
<td>C_9H_11NO_3</td>
</tr>
<tr>
<td>C_12H_14N_2O_4</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>C_9H_11NO_3</td>
<td>C_10H_12NO_4</td>
<td>C_9H_11NO_3</td>
<td>C_9H_11NO_3</td>
<td>C_9H_11NO_3</td>
</tr>
<tr>
<td>C_12H_14N_2O_4</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>C_9H_11NO_3</td>
<td>C_10H_12NO_4</td>
<td>C_9H_11NO_3</td>
<td>C_9H_11NO_3</td>
<td>C_9H_11NO_3</td>
</tr>
<tr>
<td>C_12H_14N_2O_4</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>C_9H_11NO_3</td>
<td>C_10H_12NO_4</td>
<td>C_9H_11NO_3</td>
<td>C_9H_11NO_3</td>
<td>C_9H_11NO_3</td>
</tr>
</tbody>
</table>

*Initial incorrect assignment to C_13H_13NO_3 required manual correction to C_13H_11N_2O_4.*

Those few incorrectly assigned MFs are easily identified and are corrected manually, once the MFs of other fragment and precursor ions are assigned.

Prazosin is predominantly split into two fragment ions with a core of C_12H_13N_2O_2^+ and a side chain of C_7H_8NO_2^+ through the cleavage of two C–N bonds (Fig. 3). The core C_12H_13N_2O_2^+ loses a methyl free radical to form C_11H_11N_2O_2^+. Another type of two C–N bond cleavage of prazosin produces C_10H_11NO_2^+ and C_10H_11NO_2. It is not clear why the charge is localized to C_9H_11NO_3 rather than C_10H_12NO_2 when the charge is easily localized on the similar fragment C_12H_11N_2O_2^+ (m/z = 247.1214). Other plausible fragmentation pathways of the fragment ions used in the MFA are shown in Fig. 3. Erve et al. reported the fragmentation pathway of prazosin that is in agreement with ours. The major structural difference is the fragmentation peak at 231 m/z. Their fragmentation includes a cleavage of a relatively stable C–C bond, whereas our analysis includes the cleavage of the less stable heteroatom C–O bond. It should be mentioned that the fragmentation pathways are analyzed using the fragment ions used in the MFA. Other peaks could be incorporated to get more detailed analysis of the fragmentation pathways; however, they are all speculative and are not worthwhile unless required.

Molecules with multiple core structures

Some molecules consist of two or more core structures. An example is Taxol in which the molecule splits into two core structures in the MS/MS fragmentation as described in our previous paper. Another example is ergot alkaloid dihydroergotamine (C_31H_37N_2O_5, 583.1795 Da) where the result of the MFA of dihydroergotamine is shown in Table 5. The analysis correctly assigns the MF to C_31H_37N_2O_5. However, the MF of a peak at m/z 270 was incorrectly assigned to C_11H_11N_2O_3, as the dehydrated dihydroergotamine split into a fragment ion C_11H_11N_2O_3^+ (m/z = 297) and a neutral loss C_10H_9N_2O (269 Da). The peak at m/z 270 must be the protonated form of C_10H_9N_2O. Thus the MF of the peak at m/z 270 was corrected to C_10H_9N_2O^+ (bold in Table 5). Some small neutral losses underlined in Table 5 are easily assigned to the losses of water (H_2O), carbon monoxide (CO), carbon dioxide (CO_2), and ammonia (NH_3).

The plausible fragmentation pathway of dihydroergotamine is shown in Fig. 4. The precursor molecule is in ketal–keto equilibrium. Both forms undergo a cleavage at an amide bond in the linker, splitting the precursor into an oxonium ion C_12H_12N_2O_2^+ (m/z = 253) and a neutral molecule C_12H_12N_2O_4. Cleavage of the keto form at a C–N bond of the linker splits the precursor ion into a fragment ion C_9H_20N_2O_4^+ (m/z = 270) and a neutral molecule C_9H_19N_2O_4. The ketol form loses formic acid to form C_12H_18N_3O_4^+ (m/z = 538). The ketal form similarly loses water to form C_13H_18N_3O_4^+ (m/z = 566). The linkers of these fragment ions C_13H_16N_3O_4^+ (m/z = 322) and C_16H_15N_3O_4^+ (m/z = 325), and C_16H_15N_3O_4^+ (m/z = 270), respectively. The cleavage of the C–N bond of the dehydrated fragment ion C_13H_16N_3O_4^+ (m/z = 366) also generates a fragment ion C_17H_19N_2O_3^+.

Analysis of structurally-related compounds

An extension of the use of MFA is in analyzing structurally-related compounds. For example, drugs are metabolized in vivo and some metabolites have biological activities and/or toxicity that may cause adverse effects. Since metabolites are structurally related with minor chemical modifications, MFA of metabolites and their fragment ions can provide some structural information to identify/characterize them. The method can also be used to...
follow a series of chemical modifications of natural products in vivo such as in microbial transformation and others. 

As a test of the concept, we first examined dihydroergotamine and dihydroergocristine. These are not metabolically related but they are structurally close homologues. Table 6 lists the precursor and MS/MS fragment ions of dihydroergocristine. The MF of a fragment ion at m/z 594 was not assigned in Reverse Analysis; however, the neutral loss of 18.0096 Da was manually assigned to H₂O (MM₁calc = 18.0106 Da), allowing the assignment of the MF of the fragment ion to C₁₉H₁₉N₂O₄ (bold in Table 6).

With the plausible assumption of comparable fragmentation pathways between these related compounds, we note the existence of two common fragment ions at m/z 270 and 253. We then hypothesize that these correspond to the same molecular entities in both spectra and conclude that these peaks in the dihydroergocristine spectrum correspond to the structures colored red in Fig. 4. The precursor and fragment ions of dihydroergocristine at m/z 612, 594, 350, 325 have an extra C₂H₄ with respect to the precursor and fragment ions of dihydroergotamine at m/z 584, 566, 322, 297, respectively (fragment ion structures shown in blue color in Fig. 4). The structures in Fig. 5 imply that the extra C₂H₄ is attached to the methyl group of dihydroergocristine highlighted in Fig. 5. This suggests that in dihydroergocristine, that methyl group is either a n-propyl group or an isopropyl group. In fact, dihydroergocristine has an isopropyl group.

![Diagram](image)

**Fig. 4** Proposed fragmentation pathways of dihydroergotamine. The fragment ions that are also observed in the MS/MS spectrum of dihydroergocristine are shown in red. The precursor ion and the fragment ions that are different from those in the MS/MS spectrum of dihydroergocristine are shown in blue. The fragment ion corresponding to the one in black was not observed in the MS/MS spectrum of dihydroergocristine.

![Diagram](image)

**Fig. 5** Structures of dihydroergotamine and dihydroergocristine. The structural difference is encircled in red.

<table>
<thead>
<tr>
<th>Table 6 MFA of dihydroergocristine</th>
</tr>
</thead>
<tbody>
<tr>
<td>MMobs</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>C₃H₄N₂O₂</td>
</tr>
<tr>
<td>C₅H₆N₂O₂</td>
</tr>
<tr>
<td>C₅H₆N₂O₂</td>
</tr>
<tr>
<td>C₁₉H₁₉N₂O₄</td>
</tr>
<tr>
<td>C₁₉H₁₉N₂O₄</td>
</tr>
<tr>
<td>C₁₉H₁₉N₂O₄</td>
</tr>
<tr>
<td>C₁₉H₁₉N₂O₄</td>
</tr>
<tr>
<td>C₁₉H₁₉N₂O₄</td>
</tr>
<tr>
<td>C₁₉H₁₉N₂O₄</td>
</tr>
<tr>
<td>C₁₉H₁₉N₂O₄</td>
</tr>
<tr>
<td>C₁₉H₁₉N₂O₄</td>
</tr>
</tbody>
</table>

Note: Initial incorrect assignment to C₁₉H₁₉N₂O₃ required manual correction to C₁₉H₁₉N₂O₄.

This journal is © The Royal Society of Chemistry 2009
Cyclazocine and N-allylnormetazocine

Another example of metabolites is cyclazocine and N-allylnormetazocine, where we take cyclazocine as a known compound and N-allylnormetazocine as its analog. Tables 7 and 8 list the precursor and fragment ions of benzomorphan opioids cyclazocine and N-allylnormetazocine, respectively.

The plausible fragmentation pathway of cyclazocine is shown in Fig. 6. The fragment ions of cyclazocine at m/z 175, 173 and 159 are the same as those of N-allylnormetazocine. Another fragment ion at m/z 216, which was not used in the MFA, is also common and is added in the fragmentation pathway. The MF difference between cyclazocine and N-allylnormetazocine is CH₂.

Since the peak at m/z 216 is the largest fragment ion of N-allylnormetazocine, it is very likely that the nitrogen is alkylated in a similar way as the cyclopropylmethyl group of cyclazocine with a cyclopropyl or allyl group. Indeed, N-allylnormetazocine has an N-allyl group.

Peptides

The MFA of peptides is useful for proteomics and identification of bioactive peptide metabolites. 5-Leucine enkephalin was analyzed as an example (ESI† Table S1). Fourteen fragment ions were used for the analysis, more than used for small organic molecules as the analysis was extended from MFA to peptide sequencing. The MFs of the fragment ions at m/z 538 and 510 were assigned manually (shown in bold) because the observed m/z values have errors larger than the allowed cut-off error of 0.002 Da. The MFs of the neutral losses that are assigned manually are also shown in bold.

The use of MFA for peptides was not the final goal. Proteomics requires further sequence analysis, which was carried out with the following stepwise analysis of the MS/MS data. The first step is the analysis of the amino acid composition of the peptide.
resulting in Gly, Leu, Phe and Tyr. Leu could be Ile as they are not distinguished by our MS/MS fragmentation energy. Other amino acids are not found in the neutral losses. Adding the molecular masses of the four amino acid residues and H$_2$O for the C-terminal residue resulted in the molecular mass of (2Gly + Leu + Phe + Tyr + H$_2$O) matching the precursor molecule. Thus, the amino acid composition of the peptide is 2Gly, Leu (or Ile), Phe, and Tyr with a C-terminal free carboxyl group. The MFs of the fragment ions are assigned to the amino acid residues, H$_2$O, CO, NH$_3$, and HCO$_2$H reflecting different amide bond cleavage sites and N- or C-terminal residues. The MF of the fragment C$_{11}$H$_{12}$N$_2$O$_2$ (204 Da) is, for example, assigned to (Gly + Phe) and other assignments are listed in the second row from the bottom (see ESI† Table S1). The bottom row of Table S1 lists the amino acid residues, H$_2$O, CO, NH$_3$ and HCO$_2$H lost from the precursor molecule to form the corresponding fragment ions.

Alternatively, the amino acid composition can be obtained by tracing fragmentation pathways, which release amino acid residues (or di-peptides). In the case of 5-leucine enkephalin, one such fragmentation pathway is shown in Fig. 7A, giving a composition of 2Gly, Leu, Phe, and Tyr. The second step is the identification of the N- or C-terminal residues (Fig. 7B). Since Leu and Tyr are the first amino acid residues that can be fragmented off from the precursor ion, they are the N- or C-terminal residues. The third step is the identification of the C-terminal residue. As peptides often have a free carboxyl group or are blocked with an amide, fragment ions that contain Leu-OH, Leu-NH$_2$, Tyr-OH, or Tyr-NH$_2$ were searched, resulting in a fragment ion of (Phe, Leu)-OH. Therefore, Leu was assigned to the C-terminal residue, and, automatically, Tyr was assigned to the N-terminal residue. The fourth step is the sequencing of internal amino acid residues. Starting from the C-terminal Phe-Leu-OH fragment ion, in the fragmentation pathway amino acid residues are sequentially added (or di-peptides if necessary) as of Fig. 7C. Similarly, starting from the N-terminal Tyr-Gly fragment ion, amino acid residues (or di-peptides if necessary) are sequentially added as shown in Fig. 7D.

The final step applies the sequence Tyr-Gly-Gly-Phe-Leu-OH to all fragment ions in order to confirm the sequence. Each line shown in Fig. 8 represents the sequence region of the fragment of which its MF is shown in the left column. As all of the fragments used in the MFA fit to the sequence, the Tyr-Gly-Gly-Phe-Leu-OH sequence is confirmed.

**Chloro- or bromo-containing compounds**

The MFA was extended to organic compounds that contain chlorine and/or bromine atoms in addition to C, H, N and O. The inclusion of chlorine and bromine atoms in the automated MFA is not possible with the precision of 0.002 Da. However, as chlorine and bromine have characteristic isotopes of $^{37}$Cl (24.47% natural abundance) and $^{81}$Br (49.48% natural abundance), chlorine and/or bromine atoms in precursor and fragment ions estimated based on their isotope peaks are replaced with hydrogen atom(s). The molecular masses of these ions are modified accordingly and are submitted to the MFA. Chlorine

![Fig. 8 Overall detail of analysis of 5-leucine enkephalin.](image)

![Fig. 9 Plausible fragmentation pathways of quinacrine, where the cleavages occurred at C–N bonds.](image)

<table>
<thead>
<tr>
<th>MM$_{obs}$</th>
<th>399.2080</th>
<th>326.1167</th>
<th>258.0573</th>
<th>243.0446</th>
<th>141.1510</th>
</tr>
</thead>
<tbody>
<tr>
<td>C$<em>{23}$H$</em>{36}$ClN$_2$O</td>
<td>C$<em>{19}$H$</em>{19}$ClN$_2$O</td>
<td>C$<em>{11}$H$</em>{11}$ClN$_2$O</td>
<td>C$<em>{9}$H$</em>{10}$N</td>
<td>C$<em>{7}$H$</em>{6}$N</td>
<td>C$<em>{5}$H$</em>{4}$N</td>
</tr>
<tr>
<td>C$<em>{19}$H$</em>{19}$ClN$_2$O</td>
<td>C$<em>{15}$H$</em>{15}$ClN$_2$O</td>
<td>C$<em>{11}$H$</em>{11}$ClN$_2$O</td>
<td>C$<em>{9}$H$</em>{10}$N</td>
<td>C$<em>{7}$H$</em>{6}$N</td>
<td>C$<em>{5}$H$</em>{4}$N</td>
</tr>
<tr>
<td>C$<em>{11}$H$</em>{11}$ClN$_2$O</td>
<td>C$<em>{7}$H$</em>{7}$ClN</td>
<td>C$<em>{5}$H$</em>{5}$N</td>
<td>C$<em>{3}$H$</em>{3}$N</td>
<td>C$<em>{1}$H$</em>{1}$N</td>
<td>C$<em>{1}$H$</em>{1}$ClN$_2$O</td>
</tr>
<tr>
<td>C$<em>{9}$H$</em>{10}$N</td>
<td>C$<em>{7}$H$</em>{7}$ClN</td>
<td>C$<em>{5}$H$</em>{5}$N</td>
<td>C$<em>{3}$H$</em>{3}$N</td>
<td>C$<em>{1}$H$</em>{1}$N</td>
<td>C$<em>{1}$H$</em>{1}$ClN$_2$O</td>
</tr>
<tr>
<td>C$<em>{7}$H$</em>{7}$ClN</td>
<td>C$<em>{5}$H$</em>{5}$N</td>
<td>C$<em>{3}$H$</em>{3}$N</td>
<td>C$<em>{1}$H$</em>{1}$N</td>
<td>C$<em>{1}$H$</em>{1}$ClN$_2$O</td>
<td>C$<em>{1}$H$</em>{1}$ClN$_2$O</td>
</tr>
</tbody>
</table>

Table 9 MFA of quinacrine
and/or bromine atoms are then restored to the MFs manually. The method was applied to five compounds beclometasone dipropionate (C26H32Cl2O3, 520.2228 Da), benzamil (C13H24ClN3O, 319.0948 Da), brimonidine (C11H10BrN5, 291.0120 Da), phenoxybenzamine (C13H23NO, 303.1390 Da), and quinacrine (C23H18ClN2O, 399.2077 Da), resulting in correct MF assignments for all of them.

Table 9 shows the MFA after restoring a chlorine atom and Fig. 9 shows the chemical fragmentation pathway of quinacrine based on the analysis in Table 9. Since 5 compounds containing Cl or Br may not be enough to generalize the analysis, these compounds are not included in the 96 compounds analyzed in this article.

Conclusions

The automated software developed to determine the MFs (C, H, N, and O) of precursor molecules demonstrated a high success rate of 95%. The software also determined the MFs of fragment ions and neutral losses. Although a few MFs of the fragment ions were incorrectly assigned due to the large errors of the observed mi/z values, the software managed to get the correct MF of the precursor ion. The incorrectly assigned MFs of fragment ions and neutral losses were then easily corrected manually. Using the MFs of the precursor ions, fragment ions and neutral losses, plausible fragmentation pathways were estimated. These were used for the structural characterization of homologous compounds with possible applications to dereplication of natural products and identification of drug metabolites. The analysis was further successfully extended to compounds containing Cl and Br.

Acknowledgements

We acknowledge Beata Usakiewicz for her technical assistance, and Misato Konishi for her useful suggestions. S. A. was supported by a McGill Chemical Biology Scholarship. NRCC Publication 50655.

References