

RAMSY: Ratio Analysis of Mass Spectrometry to Improve Compound Identification

Haiwei Gu,[†] G. A. Nagana Gowda,[†] Fausto Carnevale Neto,^{†,‡} Mark R. Opp,[§] and Daniel Raftery^{*,†,||}

[†]Northwest Metabolomics Research Center, Department of Anesthesiology and Pain Medicine, University of Washington, 850 Republican St., Seattle, WA 98109, United States

[‡]Department of Organic Chemistry, Institute of Chemistry, Sao Paulo State University, Rua Francisco Degni 55, Araraquara, Sao Paulo 14800-900, Brazil

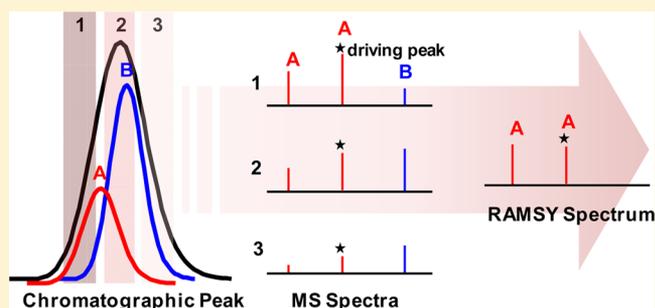
[§]Department of Anesthesiology and Pain Medicine, University of Washington, Harborview Medical Center, 908 Jefferson St., Seattle, WA 98104, United States

^{||}Public Health Sciences Division, Fred Hutchinson Cancer Research Center, 1100 Fairview Ave. N., Seattle, WA 98109, United States

S Supporting Information

ABSTRACT: The complexity of biological samples poses a major challenge for reliable compound identification in mass spectrometry (MS). The presence of interfering compounds that cause additional peaks in the spectrum can make interpretation and assignment difficult. To overcome this issue, new approaches are needed to reduce complexity and simplify spectral interpretation. Recently, focused on unknown metabolite identification, we presented a new approach, RANSY (ratio analysis of nuclear magnetic resonance spectroscopy; *Anal. Chem.* **2011**, *83*, 7616–7623), which extracts the ¹H signals related to the same metabolite based on

peak intensity ratios. On the basis of this concept, we present the ratio analysis of mass spectrometry (RAMSY) method, which facilitates improved compound identification in complex MS spectra. RAMSY works on the principle that, under a given set of experimental conditions, the abundance/intensity ratios between the mass fragments from the same metabolite are relatively constant. Therefore, the quotients of average peak ratios and their standard deviations, generated using a small set of MS spectra from the same ion chromatogram, efficiently allow the statistical recovery of the metabolite peaks and facilitate reliable identification. RAMSY was applied to both gas chromatography/MS and liquid chromatography tandem MS (LC–MS/MS) data to demonstrate its utility. The performance of RAMSY is typically better than the results from correlation methods. RAMSY promises to improve unknown metabolite identification for MS users in metabolomics or other fields.



Mass spectrometry (MS) is an essential analytical tool in complex mixture analysis and enables the detection and quantitation of hundreds of metabolites in biological samples from a single measurement. MS therefore plays a prominent role in the growing metabolomics field.^{1–9} The combination of MS with either gas chromatography (GC) or liquid chromatography (LC) allows for identification of several hundred metabolites through reliable matching of parameters such as *m/z*, retention time, and/or fragmentation patterns from one or more standard MS libraries. In GC/MS, the commonly used electron ionization (EI) provides fairly reproducible and characteristic mass fragments for each metabolite. By matching these data with standard mass spectra in databases, such as the National Institute of Standards and Technology (NIST) database^{10,11} or the Agilent Fiehn GC/MS metabolomics retention time locking (RTL) library,¹² the identity of observed spectra can be established. A match factor (MF), which is a probability measure of accuracy, is often calculated for each identified metabolite.¹³ The use of retention time (RT) and retention index (RI) often provides improved

identification.^{12,14} Meanwhile, LC–MS detected metabolites are typically identified based on exact mass, isotopic ratios, and/or tandem mass spectrometry results.^{15–18} Commonly used exact mass MS or tandem mass spectrometry (MS/MS) databases include the human metabolome database (HMDB)^{19,20} and the METLIN metabolite database.^{21–23}

However, in many practical applications, the identification of compounds using MS is still challenging. For example, in metabolomics^{5,7,8,24–27} the identity of nearly 2/3 of the metabolites detected by current MS methods remains unknown. Moreover, the identity of a large fraction of the metabolites with lower match factors continues to be ambiguous and unreliable. While this situation is not surprising given the high complexity of biological samples, improvements that facilitate reliable identification are highly desirable.

Received: June 27, 2013

Accepted: October 8, 2013

Published: October 8, 2013

A significant challenge arises from the interfering peaks that complicate the MS spectra and make identification difficult. One possible approach is to use correlation methods to help isolate peaks from the same metabolites.²⁸ A number of such methods have been used to analyze nuclear magnetic resonance (NMR) spectroscopy spectra, especially STOCYSY (statistical total correlation spectroscopy).^{29–31} In such methods, peaks from the same metabolites are identified based on correlations between the peaks; however, due to the substantial number of correlations often observed with other metabolites, it can be difficult to distinguish the meaningful correlations among the peaks from a metabolite. More recently, we proposed the method of RANSY (ratio analysis of nuclear magnetic resonance spectroscopy),³² to facilitate the isolation of peaks from the same metabolite and thereby improve compound identification using NMR spectroscopy. RANSY works on the principle that the intensity ratios between the NMR peaks from the same metabolite are fixed. Hence, across a set of samples, the standard deviation of the ratios of peaks from the same metabolite will be small (zero, in principle). On the other hand, ratios of peaks from unrelated metabolites will typically have a large standard deviation, except in those rare cases where the metabolites of interest are highly correlated. We applied RANSY to both one-dimensional (1D) and two-dimensional (2D) NMR data and showed that its performance is generally better than correlation for statistically isolating the peaks associated with a particular metabolite across a set of spectra.

In the current study, we extend the concept of RANSY and present the ratio analysis of mass spectrometry (RAMSY) method, which facilitates improved compound identification using MS. We apply RAMSY to GC/MS and LC tandem MS (LC–MS/MS) data and demonstrate that RAMSY reduces spectral interference and facilitates the identification of individual molecules in overlapped MS spectra without the need for additional experiments.

THEORY

The working principle of RAMSY is similar to that for RANSY.³² RAMSY is also designed to work using single data sets that contain multiple MS spectra for the same metabolite. For peaks that originate from the same compound, under the same experimental conditions, their MS peak intensity ratios across the chromatographic peak should be relatively constant. In addition, the standard deviations of those ratios should be small. As shown schematically in Figure 1, during a typical analysis of complex mixtures, the chromatographic elution of a compound of interest (A, red, smaller chromatographic peak) is often interfered by another compound (B, blue, larger chromatographic peak). However, by choosing a driving peak (marked with an *) from compound A, the ratios between all the MS peaks from compound A and the driving peak will show much less variation than ratios between the MS peaks from compound B and the compound A driving peak. Thus, the RAMSY calculation will reduce the interference of compound B's MS peaks from the spectra of compound A.

The procedure for calculating the RAMSY spectrum is largely as described previously.³² Briefly, a driving peak is selected from the mass spectra and then ratios between the driving peak and all the other points (or peaks) in the spectra are calculated, as shown in eq 1.

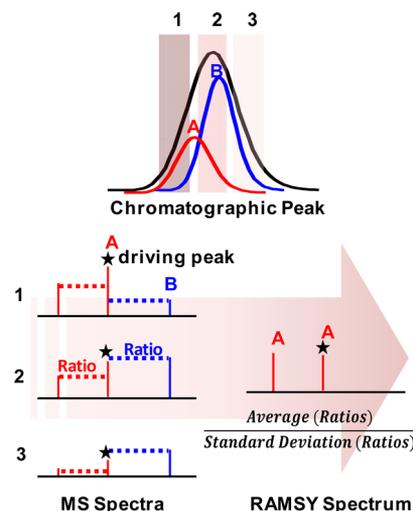


Figure 1. Schematic illustration of RAMSY. In this example, the chromatographic elution of compound A (red, smaller chromatographic peak) is interfered by compound B (blue, larger chromatographic peak). For the MS spectra collected at different retention times (1, 2, and 3), the driving peak (*) and other MS peaks from the same compound (compound A) have ratios that are relatively constant, and the ratios' standard deviations across the spectra are typically small. In contrast, the ratios for MS peaks from compound B and the driving peak (*) from compound A vary and the resulting standard deviations are relatively large. The calculated RAMSY spectrum will de-emphasize peaks from compound B, reducing the interference and aiding compound identification.

$$D_{i,j} = \frac{X_{i,j}}{X_{i,k}} \quad (1)$$

Where, the vector X_i is the i th spectrum of a set of n MS spectra, and the j th data point of m total points in that spectrum is denoted as $X_{i,j}$ ($X_{i,k}$ is the driving peak). D is the ratio matrix of dimension $n \times m$.

The RAMSY values, denoted as an m -element vector R , are the quotients of means and standard deviations across columns of D . The standard deviation is zero for the driving peak itself; therefore, its RAMSY value is predefined (e.g., the value of the highest RAMSY ratio). The other RAMSY values are calculated as elements of the vector R as follows:

$$R_j = \frac{\frac{1}{n} \sum_{i=1}^n D_{i,j}}{\sqrt{\frac{1}{n} \sum_{i=1}^n \left(D_{i,j} - \frac{1}{n} \sum_{i=1}^n D_{i,j} \right)^2}} \quad (2)$$

Since a ratio's standard deviation is used as the denominator, a small standard deviation will produce a large reciprocal value, generating a peak (in principle an MS peak from the same compound as that for the driving peak). In general, the MS peaks from interfering compounds will generate large standard deviations and thus small RAMSY numbers, similar to noise values. Notably, RAMSY values are dimensionless.

EXPERIMENTAL METHODS

Chemicals. Arginine, acetonitrile, methanol, pyridine, and O-methylhydroxylamine hydrochloride were purchased from Sigma-Aldrich (St. Louis, MO). Ammonium acetate was purchased from Fisher Scientific (Hampton, NH). DI water was provided in-house with a Synergy Ultrapure Water System from Millipore (Billerica, MA). FAME (fatty acid methyl ester,

chain lengths from C8–C30) mixture was purchased from Supelco (Bellefonte, PA), and the Fiehn GC/MS Metabolomics Standards Kit was obtained from Agilent Technologies (Santa Clara, CA).

Biological Samples. Sprague–Dawley rat plasma samples for GC/MS were collected in accordance with UW IACUC approved protocols. Human serum samples for LC–MS/MS experiments were obtained from Innovative Research, Inc. (Novi, MI).

GC/MS. We followed the general procedures for the Agilent Fiehn GC/MS Metabolomics RTL library,^{12,33} with minor changes incorporated to improve detection sensitivity. Briefly, proteins in rat plasma samples were first precipitated by the addition of methanol in a ratio of 200 μL of methanol to 100 μL of sample. The mixture was vortexed and then stored at 4 $^{\circ}\text{C}$ for 30 min. While still cold, samples were centrifuged for 10 min at 13000 rpm. The supernatant was then transferred to a clean 1.5 mL Eppendorf tube. Another 200 μL of methanol was added to the protein pellet, mixed well, and centrifuged for 10 min at 13000 rpm. The resulting supernatant was combined with the first, vortexed for 30 s, and then evaporated to dryness using an Eppendorf Vacufuge (Eppendorf, Hauppauge, NY). A solution of myristic acid-d₂₇ (5 μL) from the Fiehn GC/MS Metabolomics Standards Kit was added as an internal standard for retention time locking. For the derivatization process, the samples were first oxidized by adding 10 μL of *O*-methylhydroxylamine hydrochloride solution (in pyridine) at 30 $^{\circ}\text{C}$ for 90 min. The samples were then derivatized using 90 μL *N*-methyl-*N*-trimethylsilyltrifluoroacetamide with 1% chlorotrimethylsilane (MSTFA + 1% TMCS) at 37 $^{\circ}\text{C}$ for 30 min. Subsequently, 2 μL of the FAME (fatty acid methyl ester) mixture was added to each sample; the solution was gently vortexed and transferred to a GC/MS glass vial for analysis.

GC/MS experiments were performed on an Agilent 7890A GC-5975C MSD system (Agilent Technologies, Santa Clara, CA) by injecting 1 μL of the prepared samples with a split ratio of 10:1. Helium was used as the carrier gas with a constant flow rate of 1.2 mL/min. The separation of metabolites was achieved using an Agilent DB5-MS + 10 m Duraguard capillary column (30 m \times 250 μm \times 0.25 μm). The column temperature was maintained at 60 $^{\circ}\text{C}$ for 1.00 min and then increased at a rate of 10 $^{\circ}\text{C}/\text{min}$ to 325 $^{\circ}\text{C}$ and held at this temperature for 10 min. Mass spectral signals were recorded after a 4.90 min solvent delay.

LC–MS/MS. Frozen human serum samples were thawed, and proteins were precipitated by mixing 100 μL of serum with 200 μL of cold methanol. The mixture was centrifuged at 13000 rpm for 10 min. The supernatant was transferred to a clean 1.5 mL Eppendorf tube and then dried under vacuum (Eppendorf Vacufuge). The obtained residue was reconstituted in 250 μL DI water prior to LC–MS/MS analysis.

LC–MS/MS experiments were performed using an Agilent 1200 SL LC system coupled online with an Agilent 6520 Q-TOF mass spectrometer (Agilent Technologies, Santa Clara, CA). Each prepared sample (8 μL) was injected onto an Agilent Poroshell 120 EC-C18 column (2.1 \times 50 mm, 2.7 μm), which was heated to 50 $^{\circ}\text{C}$. The flow rate was 0.5 mL/min. Mobile phase A was 5 mM ammonium acetate in water, and mobile phase B was 0.1% water in ACN. The mobile phase composition was kept isocratic at 3% B for 1 min and was increased to 90% B in 4 min; after another 4 min at 90% B, the mobile phase composition was returned to 3% B. Electrospray ionization (ESI) was used in positive mode, and the voltage was

3.5 kV. The collision energy for automatic LC–MS/MS experiments was fixed at 10 V, targeting preselected compounds (such as arginine at m/z 175.1195). The mass accuracy of our LC–MS system is generally less than 5 ppm; the Q-TOF MS spectrometer was calibrated prior to each batch run, and a mass accuracy of less than 1 ppm was often achieved using the standard tuning mixture (G1969–85000, Agilent Technologies, Santa Clara, CA). Throughout the MS measurements, reference masses of m/z 121.0509 and 922.0098 were used to correct any mass errors. The absolute intensity threshold for MS data collection was set to 100, and the relative threshold was 0.001%. The absolute intensity threshold for MS/MS measurements was 5, and the relative threshold was 0.01%. The acquisition rate was 1.5 spectra/s. This Q-TOF system has good resolution for MS measurements; for example, in a typical tuning, the resolution was 4787 and 7315 for the ions at m/z 118.0863 and 322.0481, respectively. After data acquisition, the whole data set obtained for each sample was exported for analysis, without identifying any MS/chromatogram peaks.

Data Analysis. The data were analyzed using Matlab 7.0 (MathWorks, Natick, MA) installed on a personal computer. The data were subjected to analysis using RAMSY as well as correlation algorithms (see Supporting Information; free download available at <http://depts.washington.edu/nwmc/ramsy>). The RAMSY and correlation values were set to zero if all the intensities at a specific m/z were zero across all the spectra selected in the calculation.

In addition, we computed spectral match factors (MFs) using the same algorithms as those reported for the NIST library.¹³ To compute a spectral match factor, we first obtained the “angle” between the two spectra:

$$F_1 = \frac{\sum M(A_S A_U)^{1/2}}{[\sum M A_S \sum M A_U]^{1/2}} \quad (3)$$

M is the m/z value, and A_S and A_U are the base-peak normalized abundances of the peaks in the standard spectrum and unknown spectrum, respectively. Next, F_2 is calculated:

$$F_2 = \left(\frac{1}{N_{U\&S} - 1} \sum_{i=2}^{N_{U\&S}} \left(\frac{A_{S,i}}{A_{S,i-1}} \right)^n \left(\frac{A_{U,i}}{A_{U,i-1}} \right)^{-n} \right) \quad (4)$$

F_2 is based on relative intensities of pairs of adjacent peaks present in both spectra. $N_{U\&S}$ is the number of peaks common to the unknown and standard spectra, and $n = 1$ (–1) if the first abundance ratio is less (larger) than the second. The match factor is then calculated as follows:

$$\text{MF} = \frac{1000}{N_U + N_{U\&S}} (N_U F_1 + N_{U\&S} F_2) \quad (5)$$

where “1000” is the scaling parameter. A perfect match results in an MF value of 1000; spectra with no peaks in common result in a value of 0.

RESULTS AND DISCUSSION

The RAMSY approach was applied to both GC/MS and LC–MS/MS data. To demonstrate the performance of the method, we focused on ion chromatograms that provided overlapping mass spectra due to coeluting metabolites. The GC/MS spectra for methyl caprylate and LC–MS/MS spectra for arginine were thus selected as examples for the analysis. Notably, RAMSY is a

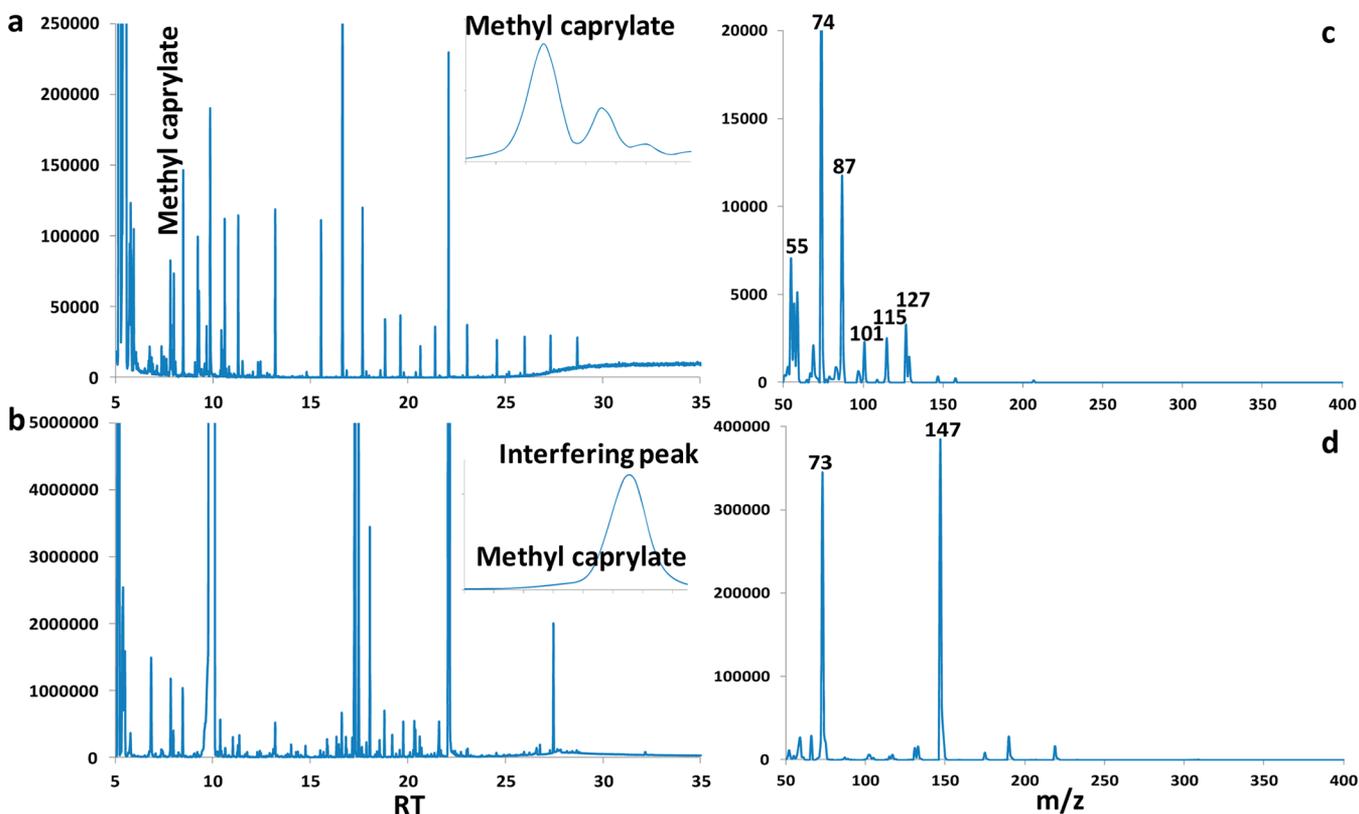


Figure 2. (a) TIC of the FAME mixture; the inset shows a well-resolved peak for methyl caprylate at 7.80 min. (b) TIC of the GC/MS data from a rat plasma sample spiked with the FAME standards; the inset shows the expanded TIC between 7.75 and 7.90 min, and no chromatographic peak can be resolved for methyl caprylate. (c) The EI-MS spectrum at the chromatographic peak maximum (7.80 min, Figure 2a). (d) EI-MS spectrum of the interfering compound(s) at 7.86 min (local maximum, Figure 2b), dominated by the peaks at m/z 73 and 147.

versatile method and can be applied for compound identification using other analytical platforms.

GC/MS. For GC/MS data, we chose a relatively simple example of a compound of interest that is overlapped with other peaks from the biological sample. In this example, the compound is methyl caprylate, a C_8 fatty-acid methyl ester commonly used as one of the 12 retention index (RI) markers, which makes it important to identify correctly.¹² As shown in Figure 2, methyl caprylate appears in the total ion chromatogram (TIC) of the FAME mixture at 7.8 min (see Figure 2a and inset) but is heavily overlapped by interfering compound(s) in the rat plasma sample spiked with the FAME mixture (Figure 2b and inset). The extraction of the mass spectrum of methyl caprylate from the TIC of Figure 2a (the local chromatographic peak maximum at 7.80 min) provides a clean mass spectrum (Figure 2c). Figure 2d shows the EI-MS spectrum of the interfering compound(s) at 7.86 min. In a comparison of Figure 2, panels c and d, it is observed that the peak at m/z 74 (the base peak in Figure 2c) is locally more unique to methyl caprylate (chosen as the driving peak in RAMSY), while the interference is mainly caused by the MS peaks at m/z 73 and 147 (Figure 2d).

We first used the NIST library to provide MFs for the extracted ion chromatograms. For methyl caprylate in the spectrum of the FAME mixture (Figure 2c), the NIST library provided an MF of 904, which is considered an excellent match (this spectrum was selected as the standard spectrum in the following MF calculations). However, in analyzing the rat plasma sample, the best MF for the same compound obtained after scanning all the mass spectra in the TIC peak in the range

of 7.75–7.90 min was 774 (7.81 min; Figure S1 of the Supporting Information); such a value (which lies between 700 and 800) is considered to be a fair match, according to NIST.¹¹ The MF for the average spectrum (Figure S2 of the Supporting Information) calculated from 25 mass spectra over the same time points range was only 195 (vide infra).

To identify methyl caprylate using the RAMSY approach, we again selected 25 mass spectra from the TIC in the range 7.75–7.90 min (Figure 2b). The RAMSY spectrum calculation was performed using the peak at 74 m/z as the driving peak (locally unique to methyl caprylate). Figure 3a shows the RAMSY spectrum in the range of m/z 50–400, and Figure 3b shows the 8 MS peaks with top RAMSY values (including the driving peak; the number of selected peaks is explained below). Figure 3c shows the averaged EI-MS spectrum (Figure S2 of the Supporting Information) after filtering with the RAMSY values (only those MS peaks with top RAMSY values were shown). It can be seen from the comparison of the spectra shown in Figure 2c and Figure 3b that RAMSY correctly identified many of the methyl caprylate peaks, including those at m/z 53 (small peak), 55, 59, 87, 101, and 115. However, RAMSY missed a fragment peak at m/z 127 and a weak molecular ion peak at m/z 158. RAMSY also picked up the wrong peak at m/z 58 (not marked in Figure 3b), since there was no peak observed there in Figure 2c. It may be interesting to note that despite the unit resolution of GC/MS, RAMSY was able to eliminate unrelated peaks quite specifically. The strong overlapping peak at m/z 73 (Figure 2d and Figure S2 of the Supporting Information) did not appear in the RAMSY spectrum (Figure 3b) when using the nearby driving peak at m/z 74, although the MS peak at m/z

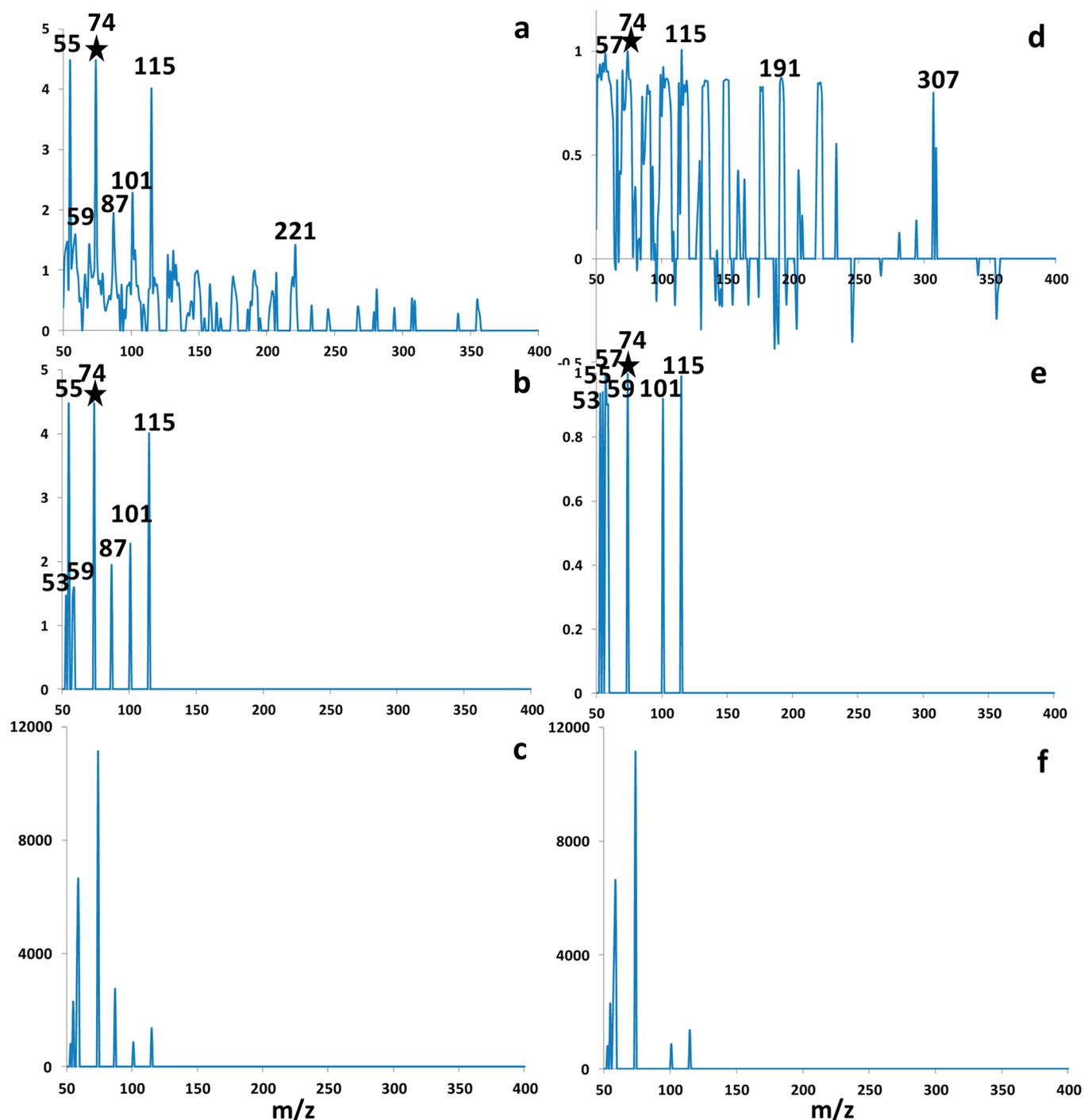


Figure 3. (a) RAMSY spectrum in the range of m/z 50–400, based on the 25 mass spectra from the TIC in the range of 7.75–7.90 min (Figure 2b); the driving peak at m/z 74 for methyl caprylate identification is indicated by the asterisk. (b) Eight peaks with top RAMSY values. (c) Averaged spectrum (Figure S2 of the Supporting Information) after filtering with the selected RAMSY values in Figure 3b. The RAMSY method indicated many correct fragment ions in the averaged EI-MS spectrum. (d) Correlation spectrum in the range of m/z 50–400, based on the 25 mass spectra from the TIC in the range of 7.75–7.90 min (Figure 2b); the driving peak at m/z 74 for methyl caprylate identification is indicated by the asterisk. (e) Eight MS peaks having top correlation values with the peak at m/z 74 (indicated by the asterisk) using the same data as that for Figure 3b. (f) Averaged spectrum (Figure S2 of the Supporting Information) after filtering with the selected correlation values in Figure 3e.

z 73 could generate the isotopic peak at m/z 74. Since RAMSY is based on ratio analysis, it is relatively independent of the peaks' original intensities. As shown in Figure 2d and Figure S2 of the Supporting Information, the peak at m/z 147 is the base peak; however, its RAMSY value was not one of the top eight (Figure 3b).

We then compared the performance of RAMSY with a correlation calculation. Figure 3d shows the correlation spectrum in the range m/z 50–400, based on the same 25 mass spectra from the TIC in the range 7.75–7.90 min (Figure 2b). For a fair comparison, eight peaks were also identified to have high correlation values with the peak at m/z 74 (Figure 3e) which had been used as the driving peak for RAMSY. The

correlation approach correctly identified a number of fragment ions of methyl caprylate such as those at m/z 115, 101, 59, 55, 57, and 53. However, it missed the peak at m/z 87, which is the second highest fragment ion peak for methyl caprylate (Figure 2c). Similar to RAMSY, the correlation approach incorrectly selected the peak at m/z 58.

MFs were also calculated based on the average GC/MS spectra generated using the peaks identified by correlation and peaks identified by RAMSY, separately (Table 1). The MF

Table 1. Match Factor Value Comparisons for Examining the Performance of RAMSY and Correlation for Compound Identification Using the GC/MS and LC–MS/MS Data^a

	GC/MS	LC–MS/MS
molecule	methyl caprylate	arginine
averaged spectrum	195	644
correlation	688	117
RAMSY	752	780

^aDetailed parameters for these calculations are given in the text.

calculated using the averaged MS spectrum (eight highest peaks selected to match the number of peaks used in the calculation) against the standard spectrum of methyl caprylate was 195, which represents a very poor match according to the NIST criteria.¹¹ The MF determined based on eight MS peaks (the driving peak at m/z 74 and seven identified peaks) from the correlation method was 688. The MF calculated using the eight MS peaks (the driving peak at m/z 74 and seven identified peaks) selected by RAMSY was 752, which lies between 700 and 800 and is considered to be a fair match. The MF of RAMSY is comparable to the best MF value (774) that could be obtained in this RT region (7.75–7.90 min) by comparing individual spectra. Both the average spectrum and correlation provided a poorer MF.

Figure 4 shows the MF values of RAMSY and correlation with different numbers of selected peaks in the calculation. The MS peaks with top RAMSY/correlation values were selected from the averaged spectrum (Figure S2 of the Supporting Information) and used in the MF calculation. It is clearly seen

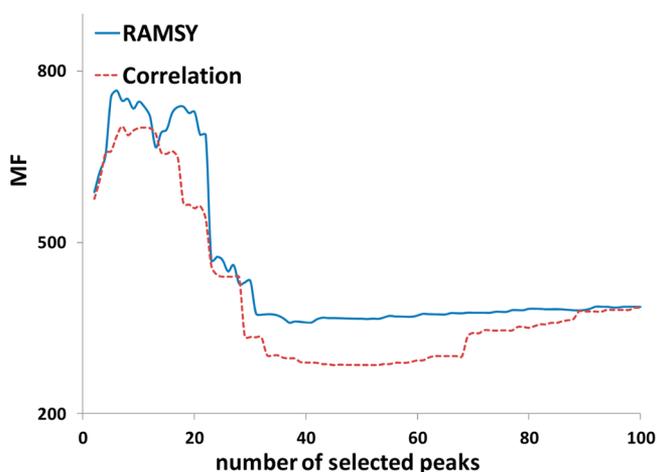


Figure 4. The MF values of RAMSY and correlation with different numbers of selected peaks in the calculation (GC/MS data). The MS peaks with top RAMSY/correlation values were selected from the averaged spectrum (Figure S2 of the Supporting Information) and used in the MF calculation.

that MF values of the RAMSY/correlation are relatively stable when the number of selected MS peaks is between 5 and 10. In general, RAMSY generates a higher MF than correlation; even in a few cases when the correlation generates a higher MF value than RAMSY (e.g., when the number of selected peaks = 13), the MF values are close. Therefore, the selection of 8 peaks in the analysis above represents a typical example.

LC–MS/MS. RAMSY was also applied to simplify LC–MS/MS spectra for improving compound identification. Using human serum samples, we targeted the TIC region for arginine, a metabolite which had MS/MS spectra containing interfering peaks from other metabolites. A typical MS/MS spectrum at the TIC peak maximum at 0.65 min obtained by targeting arginine at m/z 175.1195 is shown in Figure 5a. On the basis of

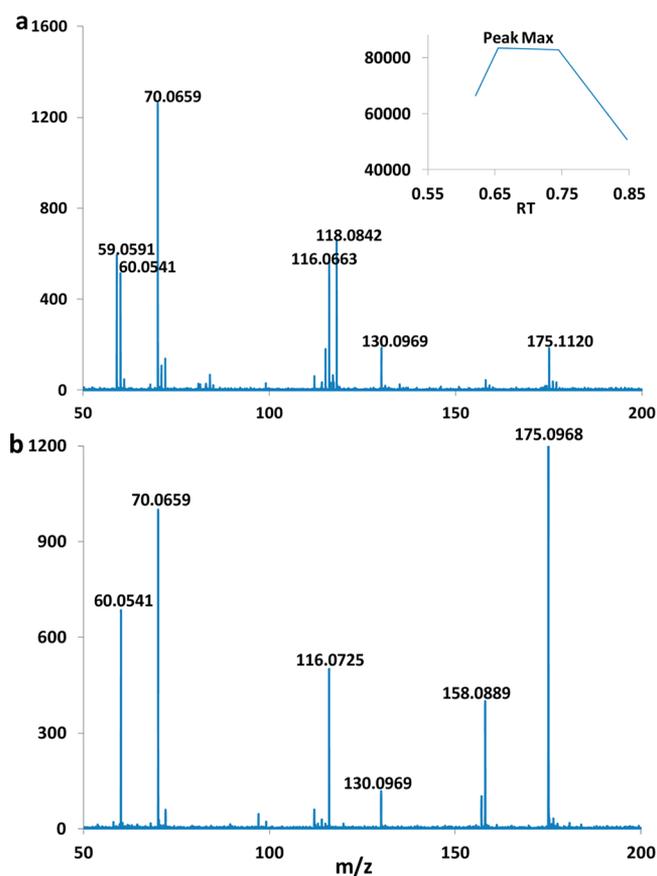


Figure 5. (a) LC–MS/MS spectrum at the peak maximum of the TIC for a human serum sample, targeting arginine at m/z 175.1195. The inset shows the TIC of the LC–MS/MS data. The signals at m/z 59 and 118 are major interfering peaks. (b) Standard LC–MS/MS spectrum of arginine.

the qualitative comparison with the standard spectrum for arginine from the Metlin database,²² a number of interfering peaks, including the strong peaks at m/z 59 and 118, can be seen in the MS/MS spectrum (Figure 5a). Figure 5b shows the MS/MS spectrum for the standard sample of arginine, and it is very similar to the Metlin standard spectrum of arginine and further confirmed the interference peaks in Figure 5a.

For the RAMSY and correlation analysis, eight LC–MS/MS spectra were selected from the TIC region between 0.55 and 0.85 min. Figure 6a shows the average of these spectra. RAMSY and correlation spectra were calculated using the protonated molecular ion peak of arginine at m/z 175.1195 as the driving

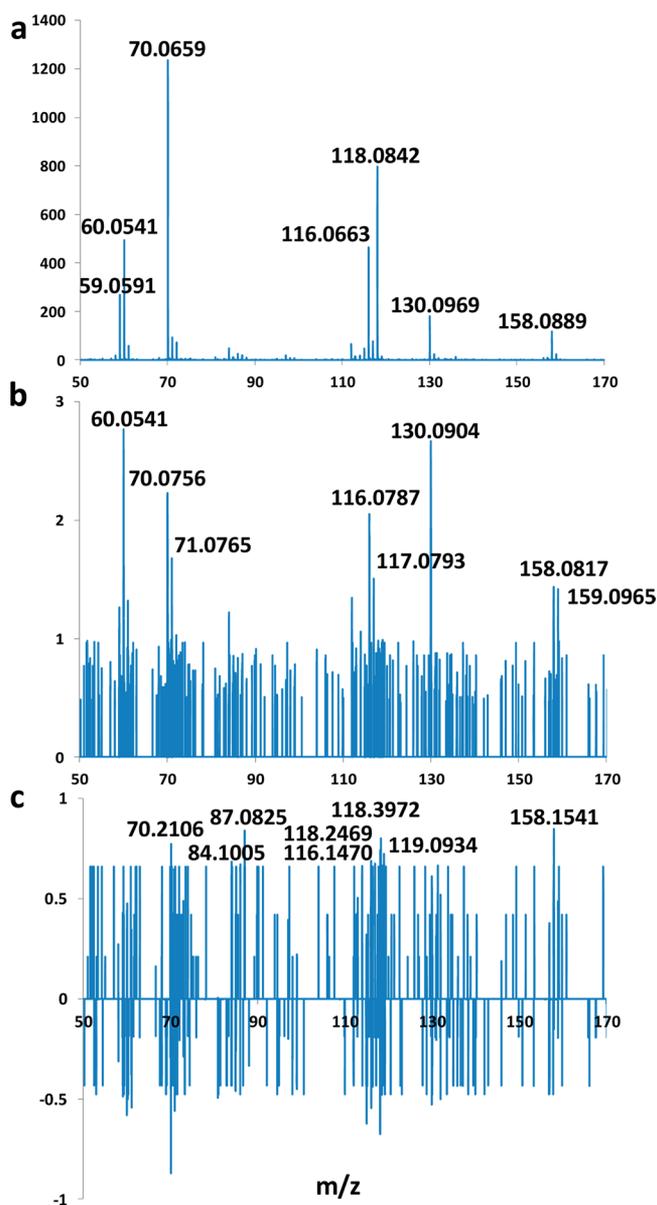


Figure 6. (a) Averaged LC–MS/MS spectrum of the eight spectra from human serum samples selected from the TIC region at 0.55–0.85 min that contains MS/MS spectra for arginine. (b) RAMSY spectrum based on the eight LC–MS/MS spectra of arginine collected from human serum samples. (c) Correlation spectrum based on the eight LC–MS/MS spectra of arginine.

peak. Figure 6 (panels b and c) show the RAMSY and correlation spectra, respectively, based on the eight LC–MS/MS spectra of arginine collected from human serum samples. Although not perfect, it is observed that many peaks in the RAMSY spectrum (Figure 6b) have a better “signal-to-noise ratio,” while the correlation spectrum (Figure 6c) looks more “noisy.” The RAMSY spectrum shows eight peaks above a threshold value of 1.4 (Figure 6b and Figure S3a of the Supporting Information). These eight peaks were selected, since the main purpose of this example is to examine the relationship between arginine and the major peaks in the MS/MS spectra. As seen in Figure 5b and 6b and Figure S3a of the Supporting Information, RAMSY identified almost all the major MS/MS peaks from arginine and greatly de-emphasized the interfering peaks at m/z 59 and 118 (the second highest peak in

Figure 6a). Using the same MS/MS spectra, we made correlation calculations for comparison. As seen in Figure 6c and Figure S3b of the Supporting Information, the correlation analysis identified eight peaks with correlation values above 0.65 (the same number of peaks were selected for a fair comparison between RAMSY and the correlation). Three of the peaks at m/z 70, 116, and 158 were in agreement with the peaks from the standard compound (arginine, Figure 5b) and those identified by RAMSY (Figure 6b and Figure S3a of the Supporting Information). However, a number of peaks, including those at m/z 84, 87, and 118 were not related to the arginine spectrum.

MF calculations were made in a similar manner to those for the GC/MS spectra, and the results are included in Table 1. The MF obtained from the averaged LC–MS/MS spectrum (with a threshold of 90 to match the number of peaks used in the calculation) was 644. The MF calculated for the RAMSY spectrum (with a threshold of 1.4) provided a value of 780, while the MF calculated using the correlation-derived spectrum (with a threshold of 0.65) provided an MF value of 117. In this case, RAMSY also provides a more accurate representation for the identity of the metabolite, arginine. Finally, Figure 7 shows

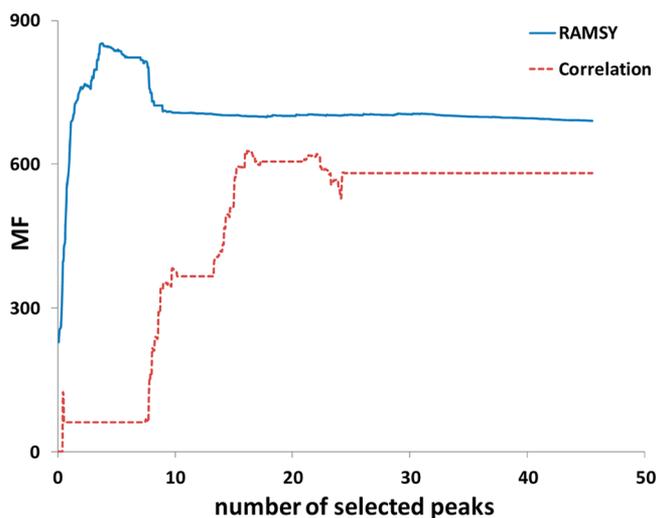


Figure 7. The MF values of RAMSY and correlation with different numbers of selected peaks in the calculation (LC–MS/MS data). The MS peaks with top RAMSY/correlation values were selected from the averaged spectrum (Figure 6a) and used in the MF calculation. On average, each peak has 22 data points in our high-resolution LC–Q-TOF experiments.

the MF values for RAMSY and correlation with different numbers of selected peaks from the averaged LC–MS/MS spectrum of arginine (Figure 6a). It is clearly seen that RAMSY generates higher MFs than correlation, for as many as 45 peaks selected according to the top RAMSY/correlation values.

Given these results, we believe that ratio analysis spectroscopy is a potentially powerful approach to simplify crowded mass spectra for the reliable identification of metabolites within complex mixtures. The RAMSY approach is demonstrated using specific examples for both GC/MS and LC–MS/MS data. RAMSY works when the GC/LC elution of compounds of interest and the interfering compounds are not exactly the same; however, even when the chromatographic peaks are the same, RAMSY is still applicable if the MS intensity ratios of the compounds are changing (e.g., the slopes of the calibration

curves are different). On the basis of our experience, generally more than five spectra and 5–10 MS peaks with top RAMSY values should be utilized. An advantage of using the RAMSY approach is that it is resistant to low peak intensities since the peak identification is based on the ratios. Further, since the RAMSY calculations are primarily based on the ratios of the peak intensities, this approach is anticipated to be less sensitive to altered chromatography conditions. A major requirement for RAMSY is that at least one isolated peak in the mass spectrum is needed to serve as the driving peak, since RAMSY cannot quantitatively differentiate the contribution from coincident interference peaks (i.e., the peak contains contributions from both the compound of interest and the interfering compound). However, it should be possible under most circumstances to find at least one relatively unique peak in the spectra for a given compound of interest. We anticipate that incorporation of the RAMSY approach as a digital filter to metabolite identification algorithms already available will benefit reliable compound identification in the mass spectra of complex biological mixtures.

CONCLUSIONS

Reliable compound identification from the analysis of mass spectra is a considerable challenge. In this work, efforts focused on alleviating this problem showed improved statistical isolation of peaks from the same metabolite in a complex mass spectra using a new approach, RAMSY, which has promise for enhanced metabolite identification. RAMSY identifies mass spectral peaks or fragment ions based on their fixed ratios for a given metabolite under the same experimental conditions. Using specific examples, we have demonstrated RAMSY using both GC/MS and LC–MS/MS data. While RAMSY cannot provide a perfect solution, it fares quite well in terms of peak identification and match factors relative to current correlation or averaging methods. An advantage of RAMSY is that it uses multiple mass spectra from the same chromatograms and does not need additional experiments. Specifically, in combination with advanced methods/software tools, RAMSY shows promise for reliable metabolite identification and reducing the large list of unidentified metabolites in the mass spectra of complex biological samples. We anticipate that the RAMSY algorithm can be incorporated into current analysis software, and that its applications can be made quite broad, including, but not limited to heterospectroscopic applications.

ASSOCIATED CONTENT

Supporting Information

Additional information as noted in text. This material is available free of charge via the Internet at <http://pubs.acs.org>.

AUTHOR INFORMATION

Corresponding Author

*E-mail: draftery@uw.edu. Tel: 206-543-9709. Fax: 206-616-4819.

Notes

The authors declare the following competing financial interest: Daniel Raftery reports holding equity and an executive position at Matrix Bio, Inc.

ACKNOWLEDGMENTS

This work was supported by the NIH/NIGMS (Grant 2R01 GM085291).

REFERENCES

- (1) Dunn, W. B.; Ellis, D. I. *TrAC, Trends Anal. Chem.* **2005**, *24*, 285–294.
- (2) Zhang, A. H.; Sun, H.; Wang, P.; Han, Y.; Wang, X. J. *Analyst* **2012**, *137*, 293–300.
- (3) Scalbert, A.; Brennan, L.; Fiehn, O.; Hankemeier, T.; Kristal, B. S.; van Ommen, B.; Pujos-Guillot, E.; Verheij, E.; Wishart, D.; Wopereis, S. *Metabolomics* **2009**, *5*, 435–458.
- (4) Pan, Z. Z.; Raftery, D. *Anal. Bioanal. Chem.* **2007**, *387*, 525–527.
- (5) Fiehn, O. *Plant Mol. Biol.* **2002**, *48*, 155–171.
- (6) Dettmer, K.; Aronov, P. A.; Hammock, B. D. *Mass Spectrom. Rev.* **2007**, *26*, 51–78.
- (7) Gowda, G. A. N.; Zhang, S. C.; Gu, H. W.; Asiago, V.; Shanaiah, N.; Raftery, D. *Expert Rev. Mol. Diagn.* **2008**, *8*, 617–633.
- (8) Reaves, M. L.; Rabinowitz, J. D. *Curr. Opin. Biotechnol.* **2011**, *22*, 17–25.
- (9) Lindon, J. C.; Holmes, E.; Nicholson, J. K. *Pharm. Res.* **2006**, *23*, 1075–1088.
- (10) Babushok, V. I.; Linstrom, P. J.; Reed, J. J.; Zenkevich, I. G.; Brown, R. L.; Mallard, W. G.; Stein, S. E. *J. Chromatogr., A* **2007**, *1157*, 414–421.
- (11) Stein, S. E. In *NIST/EPA/NIH Mass Spectral Database (NIST 11) and NIST Mass Spectral Search Program*, version 2.0g; National Institute of Standards and Technology: Gaithersburg, MD, 2011.
- (12) Kind, T.; Wohlgemuth, G.; Lee, D. Y.; Lu, Y.; Palazoglu, M.; Shahbaz, S.; Fiehn, O. *Anal. Chem.* **2009**, *81*, 10038–10048.
- (13) Stein, S. E. *J. Am. Soc. Mass Spectrom.* **1994**, *5*, 316–323.
- (14) Mallard, W. G.; Reed, J. In *AMDIS-User Guide*; National Institute of Standards and Technology: Gaithersburg, MD, 1997.
- (15) Liang, Y.; Wang, G. J.; Xie, L.; Sheng, L. S. *Curr. Drug Metab.* **2011**, *12*, 329–344.
- (16) Rivier, L. *Anal. Chim. Acta* **2003**, *492*, 69–82.
- (17) Prasad, B.; Garg, A.; Takwani, H.; Singh, S. *TrAC, Trends Anal. Chem.* **2011**, *30*, 360–387.
- (18) Rojas-Cherto, M.; Peironcelly, J. E.; Kasper, P. T.; van der Hooft, J. J. J.; de Vos, R. C. H.; Vreeken, R.; Hankemeier, T.; Reijmers, T. *Anal. Chem.* **2012**, *84*, 5524–5534.
- (19) Wishart, D. S.; Jewison, T.; Guo, A. C.; Wilson, M.; Knox, C.; Liu, Y. F.; Djoumbou, Y.; Mandal, R.; Aziat, F.; Dong, E.; Bouatra, S.; Sinelnikov, I.; Arndt, D.; Xia, J. G.; Liu, P.; Yallou, F.; Bjorn Dahl, T.; Perez-Pineiro, R.; Eisner, R.; Allen, F.; Neveu, V.; Greiner, R.; Scalbert, A. *Nucleic Acids Res.* **2013**, *41*, D801–D807.
- (20) Human Metabolome Database. <http://www.hmdb.ca/> (accessed Oct 23, 2013).
- (21) Tautenhahn, R.; Cho, K.; Uritboonthai, W.; Zhu, Z. J.; Patti, G. J.; Siuzdak, G. *Nat. Biotechnol.* **2012**, *30*, 826–828.
- (22) Scripps Center for Metabolomics. <http://metlin.scripps.edu/> (accessed Oct 23, 2013).
- (23) Smith, C. A.; Want, E. J.; O'Maille, G.; Abagyan, R.; Siuzdak, G. *Anal. Chem.* **2006**, *78*, 779–787.
- (24) Gu, H.; Gowda, G. A. N.; Raftery, D. *Future Oncol.* **2012**, *8*, 1207–1210.
- (25) Nicholson, J. K.; Holmes, E.; Kinross, J. M.; Darzi, A. W.; Takats, Z.; Lindon, J. C. *Nature* **2012**, *491*, 384–392.
- (26) Bain, J. R.; Stevens, R. D.; Wenner, B. R.; Ilkayeva, O.; Muoio, D. M.; Newgard, C. B. *Diabetes* **2009**, *58*, 2429–2443.
- (27) Fan, T. W. M.; Lane, A. N. *J. Biomol. NMR* **2011**, *49*, 267–280.
- (28) Borrás, S.; Kaufmann, A.; Companyo, R. *Anal. Chim. Acta* **2013**, *772*, 47–58.
- (29) Cloarec, O.; Dumas, M. E.; Craig, A.; Barton, R. H.; Trygg, J.; Hudson, J.; Blancher, C.; Gauguier, D.; Lindon, J. C.; Holmes, E.; Nicholson, J. *Anal. Chem.* **2005**, *77*, 1282–1289.
- (30) Blaise, B. J.; Navratil, V.; Emsley, L.; Toulhoat, P. *J. Proteome Res.* **2011**, *10*, 4342–4348.
- (31) Holmes, E.; Cloarec, O.; Nicholson, J. K. *J. Proteome Res.* **2006**, *5*, 1313–1320.
- (32) Wei, S. W.; Zhang, J.; Liu, L. Y.; Ye, T.; Gowda, G. A. N.; Tayyari, F.; Raftery, D. *Anal. Chem.* **2011**, *83*, 7616–7623.

(33) In *Agilent Fiehn GC/MS Metabolomics RTL User Guide*; Agilent Technologies: Wilmington, DE, 2008.