Published in final edited form as:

Analyst. 2013 September 21; 138(18): 5453–5460. doi:10.1039/c3an00667k.

A Method of Aligning Peak Lists Generated by Gas Chromatography High-resolution Mass Spectrometry

Xiaoli Wei^a, Xue Shi^a, Mark Merrick^b, Peter Willis^b, David Alonso^b, and Xiang Zhang^a

Xiang Zhang: xiang.zhang@louisville.edu

^aDepartment of Chemistry, University of Louisville, KY 40292, USA. Fax: +01 502 852 8149; Tel: +01 502 852 8878

^bLECO Corporation, St. Joseph, MI 49085, USA

Abstract

We report a method for the peak list alignment of gas chromatography high resolution time-of-flight mass spectrometry data. The alignment is performed in a z-score transformed retention time domain to standardize a peak distribution across samples. A mixture score is developed to assess the similarity between two peaks by simultaneously evaluating the mass spectral similarity and the closeness of retention time. An analysis of experimental data acquired under three different flow rates indicates that the proposed method is able to correctly align the heterogeneous data. The effectiveness of method is further validated by analyzing experimental data of multiple mixtures of metabolite extract from mouse liver with 28 spiked-in acids. All of the detected spiked-in acids were correctly aligned. A statistical test correctly detected the concentration differences of the spiked-in compounds between sample groups using the alignment table. The area under curve (AUC) value in the receiver operating characteristic (ROC) curve is larger than 0.85 in all three of the compared sample groups, indicating a high accuracy of peak alignment and supporting the potential application of the proposed method for metabolomics projects such as biomarker discovery.

Introduction

Coupling gas chromatography with mass spectrometry (GC–MS) enables analysis of hundreds to thousands of small molecules, depending on the instrument configuration and the complexity of the samples. GC–MS is well-suited for the analysis of compounds that are naturally volatile or volatile after derivatization. The applications of GC–MS in metabolomics have grown on a global scale in recent years.^{1–3} A large volume of data is usually generated during a metabolomics study and advanced bioinformatics algorithms and software tools are needed to extract the biological information from the experimental data. A number of algorithms and software tools have been developed for processing GC–MS data, including MZmine,⁴ XCMS,⁵ MetAlign,⁶ ADAP,⁷ etc.

With the significant technology advances in mass spectrometry, a powerful analytical platform that couples a gas chromatographic system with a high resolution time-of-flight mass spectrometer (GC–HRTMS) is now available for metabolomic studies.^{8,9} Such a system provides high resolution mass spectra rather than nominal mass resolution mass spectra. It is expected that the high resolution mass spectra can improve the accuracy of deconvoluting overlapping peaks during spectral deconvolution and therefore, aid

compound identification and quantification by removing fragment ions with large m/z deviation from the true value.

A significant application of GC–HRTMS is metabolomics profiling, where tens and even hundreds of biological samples are analyzed in one project. However, while gathering GC–HRTMS data, there is always a shift of retention time. Retention time shifts make it difficult to compare metabolic profiles obtained from multiple samples. In order to correct the retention time shifts, two alignment approaches have been developed: profile alignment and peak list alignment. The profile alignment directly uses the entire chromatographic data, i.e. the raw instrument data, as the input data.^{10,11} In the peak list alignment approach, the raw instrument data are first deconvoluted to peak list, and the peak lists of multiple samples are then employed as the input data to correct retention time shifts.^{5–7} The selection of the two approaches for retention time correction depends on the methods of downstream statistical analysis.

To our knowledge, none of the existing open access or post-acquisition software packages for the analysis of GC–MS data was designed to process data acquired on a GC–HRTMS system. Furthermore, the existing software packages could not process experimental data acquired under different experimental conditions, i.e., heterogeneous data using different temperature ramps or gas flows. In addition, most of the software packages align the experimental data of different samples by retention time alone, even though the fingerprint feature of a compound, i.e., electron ionization (EI) mass spectrum of fragment ions, is readily available in the raw instrument data. Aligning compound peaks solely based on the retention time may introduce a high rate of false alignment because some compounds have similar retention times.^{12–15}

In this study, we developed a method entitled MetHR to align the experimental data generated by a GC–HRTMS. MetHR performs peak list alignment using both retention time and mass spectra and is able to process heterogeneous data acquired under different experimental conditions. The accuracy of MetHR was verified by analyzing two types of samples, a mixture of compound standards and mixtures of metabolite extract from mouse liver with 28 compound standards added at defined concentrations. The mixture of compound standards was analyzed under three different experimental conditions by varying the flow rate.

Experimental

Mixture of compound standards

A complex mixture of analytes was created by the combination of three commercially-available mixtures. This was achieved using a mixture of 76 compounds each at 1000 $\mu\text{g/mL}$ (Cat. No. 31850, 8270 MegaMix, Restek Corp., Bellefonte, PA), a mixture of $\text{C}_7\text{--C}_{40}$ n -alkanes each at 1000 $\mu\text{g/mL}$ (Cat. No. 49452-U, Sigma-Aldrich Corp., St. Louis, MO) and a mixture of six deuterated semi-volatile internal standards (ISTD) each at 2000 $\mu\text{g/mL}$ (Cat. No. 31206, Restek Corp., Bellefonte, PA). These were combined in a ratio of 2:2:1, respectively, and diluted with dichloromethane to a final concentration of 100 $\mu\text{g/mL}$ for each component.

Spike-in samples

A sample of mouse liver tissue was weighed and homogenized for 2 min after adding water at a concentration of 100 mg liver tissue/mL water. To extract compounds from the mouse liver, 200 μL of homogenized mouse liver sample was mixed with 800 μL of methanol and vortexed for 1 min, followed by centrifugation at 4 °C for 10 min at 15,000 rpm. 0.3 mL of the top solution was aspirated into a plastic tube and dried by N_2 flow. After dissolving the

dried sample with 100 μL of pyridine, 50 μL of N-(tert-butyldimethylsilyl)-N-methyl-trifluoroacetamide (MTBSTFA, Sigma-Aldrich Corp., St. Louis, MO) was added to derivatize the analytes and the sample incubated at 70 $^{\circ}\text{C}$ for 30 min.

A mixture of 28 acid standards was prepared at a concentration of 100 $\mu\text{g}/\text{mL}$ per acid. The acids included glycine, L-alanine, L-proline, L-histidine, L-phenylalanine, L-lysine, L-glutamic acid, L-methionine, L-serine, L-threonine, L-tryptophane, L-valine, L-leucine, adipic acid, butyric acid, fumaric acid, malonic acid, oxalic acid, succinic acid, dodecanoic acid, hexadecanoic acid, heptadecanoic acid, decanoic acid, octadecanoic acid, tetradecanoic acid, tridecanoic acid, nonanoic acid and pentadecanoic acid (Sigma-Aldrich Corp., St. Louis, MO). 100 μL of the acid mixture was added to 50 μL MTBSTFA and held at 70 $^{\circ}\text{C}$ for 30 min for derivatization.

Ten μL of the derivatized acid mixture was added to the first vial, while 20 and 40 μL of the acid mixture were added to the second and the third vials, respectively. After adding 20 μL of derivatized liver extract to each of the three vials, pyridine was added to each of the 3 vials to bring the total volume of each aliquot to 105 μL . This resulted in three samples with spiked-in acid standards. The acid concentration in each of the spiked-in samples is approximately 6.4, 12.8, and 25.6 $\mu\text{g}/\text{mL}$, respectively. The ratio of them is 1:2:4. A blank sample was also prepared in parallel without adding acid standards and liver extract.

GC–HRTMS analysis

All sample analyses were performed on a system consisting of an Agilent 7890A gas chromatograph with an autosampler (Agilent Technologies, Palo Alto, CA) interfaced to a LECO Pegasus® GC–HRT high resolution time-of-flight mass spectrometer with an electron ionization (EI) ion source (LECO, St. Joseph, MI).

For the mixture of compound standards the GC was equipped with a GC $\times$ GC (comprehensive two-dimensional gas chromatography) accessory (LECO, St. Joseph, MI) and a two-column set installed in the modulator and secondary oven. However, the mixture of standards was only run in one-dimensional mode (no modulation). A 10 m $\times$ 0.18 mm i.d. $\times$ 0.2 μm d_f Rtx-5 column (Restek Corp., Bellefonte, PA) connected to a 1 m $\times$ 0.1 mm i.d. $\times$ 0.1 μm d_f Rxi-17 column (Restek Corp., Bellefonte, PA) was used with the following conditions: helium carrier gas at 0.8, 1.0, and 1.2 mL/min; split/splitless inlet at 250 $^{\circ}\text{C}$ and split ratio of 100:1; GC oven temperature program of 45 $^{\circ}\text{C}$ for 2 min and then 18 $^{\circ}\text{C}/\text{min}$ to 315 $^{\circ}\text{C}$ for 5 min; modulator temperature program of +40 $^{\circ}\text{C}$ relative to the GC oven; secondary oven temperature program of +25 $^{\circ}\text{C}$ relative to the GC oven; and a transfer line temperature of 300 $^{\circ}\text{C}$. The MS conditions were as follows: electron energy of 70 eV; ion source temperature of 250 $^{\circ}\text{C}$; acquisition delay of 90 s; acquisition rate of 12 spectra/s; mass range of m/z = 35 – 500; extraction frequency of 2 kHz; and internal mass calibrant of PFTBA (perfluorotributylamine). Five replicate injections of the mixture were performed for each flow condition.

For the analysis of the spike-in samples, a 30 m $\times$ 0.25 mm i.d. $\times$ 0.25 μm d_f Rxi-5Sil MS column (Restek Corp., Bellefonte, PA) was used with the following GC conditions: helium carrier gas at 1 mL/min; split/splitless inlet at 280 $^{\circ}\text{C}$ and split ratio of 20:1; oven temperature program of 60 $^{\circ}\text{C}$ for 1 min and then 10 $^{\circ}\text{C}/\text{min}$ to 325 $^{\circ}\text{C}$ and hold for 5 min; and a transfer line temperature of 320 $^{\circ}\text{C}$. The MS conditions were as follows: electron energy of 70 eV; ion source temperature of 260 $^{\circ}\text{C}$; acquisition delay of 375 s; acquisition rate of 12 spectra/s; mass range of m/z = 40 – 510; extraction frequency of 2 kHz; and internal mass calibrant of PFTBA (perfluorotributylamine). All samples were injected 5 times for GC–HRTMS analysis.

Instrument data reduction

The raw instrument data were first reduced to a peak list by LECO ChromaTOF-HRT® software (Beta Version 1.59) using the vendor suggested parameters. Internal mass calibration was done with PFTBA using the manufacturer's default calibration matrix applied to each injection. The data processing parameters in ChromaTOF- HRT® for the analysis of the compound mixture were set as: S/N = 50, peak quality = 0.9, and peak confidence = 6.0. The data processing parameters for analysis of the spiked-in data are S/N = 20, peak quality = 0.7, and peak confidence = 6.0.

Theoretical

The peak lists deconvoluted from the instrument data by ChromaTOF-HRT® software are used as the inputs for alignment. The peak features used for alignment are retention time and comprehensive extracted (“Peak True”) mass spectrum of each peak. Fig. 1 depicts the workflow of the MetHR method. MetHR aligns peaks present in different peak lists by simultaneously evaluating the mass spectral similarity and the closeness of retention time using a mixture score. It first transforms the retention time values into z-scores to standardize the distribution of chromatographic peaks in each sample, which enables the alignment of heterogeneous data, i.e., the experimental data acquired under different experimental conditions, such as different flow rates¹³. The retention time value of each peak in a peak list is transformed into a modified z-score as follows:

$$z_{i,j} = \frac{t_{i,j} - \mu}{\sigma} \quad (1)$$

where $t_{i,j}$ denotes the retention time of the j -th peak in the i -th sample, μ and σ are the medians of the means and standard deviations of the retention time values among peaks of all sample set S , respectively.

There are two sequential steps for peak list alignment: full alignment and partial alignment. The full alignment recognizes the landmark peaks that are a set of compound peaks present in every sample, while the partial alignment aligns the peaks in the samples that are not recognized as the landmark peaks.

Full peak list alignment

Let $S = \{S_1, S_2, \dots, S_i, \dots, S_{n+1}\}$ be the sample set and $n+1$ is the total number of samples. Each sample is represented by a peak list, and each peak has two features, retention time and fragment ion mass spectrum (i.e., m/z and intensities). A reference peak list (R_{pl}) is selected from the sample set S by using one-dimensional Kolmogorov-Smirnov (KS) test. The reference sample is a sample that has the maximum similarity in peak distribution with other samples. The sample list can be rewritten as $S = \{R_{pl}, T_1, T_2, \dots, T_i, \dots, T_n\}$, where the samples other than R_{pl} are represented as T_i ($i = 1, 2, \dots, n$) and are named as test samples. Each of the test samples is aligned to R_{pl} , respectively.

For each pair of peak lists $\{R_{pl}, T_i\}$, each peak has a mass spectrum (m/z and intensity pairs) as one of its two features. To calculate the mass spectral similarity between two peaks, the fragment ions between the two mass spectra are first matched using a user defined m/z variation window ($\Delta m/z = 5$ ppm in this study). In case that multiple peaks in one spectrum can be matched to one or multiple fragment ions in the other mass spectrum, discrete convolution is used to find the optimal match.¹⁶ Briefly, given two chromatographic peaks c_1 from R_{pl} and c_2 from T_i , a group of fragment ions $p_1(n_1)$ in the mass spectrum of c_1 can be matched to another group of fragment ions $p_2(n_2)$ in the mass spectrum of c_2 , where n_1

and n_2 are the number of fragment ions in each fragment ion groups respectively. Assuming n_1, n_2 , the convolution at position n is computed as follows:

$$\text{conv}(n) = \sum_{m=1}^n A_1(m) * A_2(n_2 - m) \quad (2)$$

where $n = 1, \dots, n_1 + n_2 - 1$, A_1 and A_2 are the fragment ion intensity in p_1 and p_2 , respectively. By selecting the maximum value among $\text{conv}(n)$, the corresponding peak pair is selected as the optimal match. Then, the mass spectral similarity of all possible peak pairs between R_{pl} and T_i are computed as follows¹⁷

$$C = \frac{\alpha_w \circ \beta_w}{\|\alpha_w\| \cdot \|\beta_w\|} \quad (3)$$

where $w = (x, y)$ is a vector of weight factors of intensity and m/z value, respectively, $\alpha_w = (\alpha_i^w)_{i=1}^I$ and $\beta_w = (\beta_i^w)_{i=1}^I$ are weighted intensities and

$$\alpha_i^w = (\alpha_i)^x \cdot (z_i)^y \text{ and } \beta_i^w = (\beta_i)^x \cdot (z_i)^y \quad (4)$$

where $\alpha = (\alpha_i)_{i=1, \dots, I}$ and $\beta = (\beta_i)_{i=1, \dots, I}$ are the two sequences of intensities of the two matching mass spectra, respectively; I is the total number of intensities; Z_i is the m/z value of the i -th intensity, $i = 1, 2, \dots, I$, and x and y are weight factors. In this study, we used the weight factor $(x, y) = (0.53, 1.3)$.¹⁸

The candidates of the landmark peaks are chosen as follows:

Step 1. $R_{pl} = \{p_{r,1}, p_{r,2}, \dots, p_{r,m}\}$; $p_{r,j} = (s_{r,j}, t_{r,j})$; $p_{r,j}$ is the j -th peak in the reference sample R_{pl} , m is the number of peaks in R_{pl} ; $s_{r,j}$ and $t_{r,j}$ are the fragment ion mass spectrum and z-score transformed retention time of the j -th peak in R_{pl} , respectively; $T_i = \{p_{i,1}, p_{i,2}, \dots, p_{i,q}\}$; $p_{i,j} = (s_{i,j}, t_{i,j})$; $p_{i,j}$ is the j -th peak in the i -th test sample; q is the number of peaks in the i -th test sample; and $s_{i,j}$ and $t_{i,j}$ are the fragment ion mass spectrum and z-score transformed retention time of the j -th peak in the test sample T_i , respectively.

Step 2. For each peak $p_{i,j}$ in the test sample T_i , its spectral similarity with all peaks in R_{pl} are computed using equation (3) and recorded as $C_{i,j} = \{c_{j,1}, c_{j,2}, \dots, c_{j,m}\}$, $c_{j,k}$ denotes the spectral similarity score between the j -th peak in sample T_i and the k -th peak in R_{pl} , where $1 \leq k \leq m$. Peak pairs with spectral similarity scores larger than 0.6 are kept while others are discarded.

Step 3. If a peak $p_{i,j}$ in the test sample T_i has spectral similarity larger than 0.6 with multiple peaks in R_{pl} , the peak pair with the smallest retention time difference is kept while others are discarded.

Step 4. Repeat Steps 2 and 3 for each peak in the test sample T_i to form a peak pair list. It is possible that multiple peaks in the test sample T_i are paired with one peak in R_{pl} . In this case, the peak pair with the smallest retention time difference is kept while others are discarded.

Step 5. Repeat Steps 2-4 for each of the test samples $T_i (i = 1, 2, \dots, n)$. If a peak in R_{pl} is paired to a peak in all of the test samples, this peak is considered as a candidate of landmark peaks. All landmark peak candidates form a $L \times (n+1)$ matrix, where L

denotes the total number of landmark peak candidates, $n+1$ is the total number of samples in sample set S .

A mixture similarity score S_m is used to measure the match quality between two peaks in T_i and R_{pl} as follows:

$$S_m(d_j, s_j|w_i) = w_i * \exp\left(-k * \frac{d_j - d_{\min}}{d_{\text{med}} - d_{\min}}\right) + (1 - w_i) * s_j \quad (5)$$

where d_j is the retention time difference of the z-score transformed retention time between the j -th matched peak pair, $d_{\min}$ and d_{med} are the minimum and median values of the retention time difference among all matched peaks in the two peak lists, respectively, s_j is the spectral similarity between the j -th matched peak pair, k is an empirical value where $k = 0.6$, w_i is a weight factor optimized for the alignment of T_i and R_{pl} ($0 \leq w_i \leq 1$).

All landmark peak candidates recognized between a test peak list T_i and R_{pl} are used to optimize the value of weight factor w_i by maximizing

$$\operatorname{argmax}(\sum_{j=1}^L S_m(d_j, s_j|w_i)) \quad (6)$$

where L is the number of landmark peak candidates and w_i is set as 0.05, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9 and 0.95, respectively.

After optimizing the weight factor w_i , the values of S_m can be calculated for all matched peak pairs between the test peak list T_i and R_{pl} . By iteratively considering the paired sample set $\{R_{pl}, T_i | i = 1, \dots, n\}$, the optimal weight factor set $\{\omega_1, \dots, \omega_n\}$ can be obtained for each test peak list T_i . All landmark peak candidates are removed as outliers if their S_m values are not in the confidence interval of 95%. The potential false-positive landmark peaks are further detected and removed by a rank-order-based filtering method¹³, which assumes that the landmark peaks have the same elution order in GC in the two peak lists T_i and R_{pl} . After outlier removal and rank-order-based filtering, the final list of landmark peaks is determined. The minimum mixture score $S_m^{\min}$ among all the test peak lists is then used as a threshold value for S_m in the partial alignment.

Partial peak list alignment

After the full alignment, the retention time values of non-landmark peaks in each test peak list are corrected based on the retention time difference of the landmark peaks between a test peak list T_i and R_{pl} . Given two peak lists (R_{pl}, T_i), the chromatographic elution period in the test sample T_i can be segmented into $m+1$ sections by the retention time of landmark peaks, i.e., $[t_{t,0}, t_{t,1}], [t_{t,1}, t_{t,2}], \dots, [t_{t,m}, t_{t,m+1}]$. A local linear fitting method is employed to correct the retention time of peaks in the test sample T_i that are present between two adjacent landmark peaks $t_{t,j}$ and $t_{t,j+1}$ as follows:

$$t'_{t,j} = t_{r,j} \quad (7)$$

$$t'_{t,s} = t_{r,j} + \frac{t_{t,s} - t_{t,j}}{t_{t,j+1} - t_{t,j}} (t_{r,j+1} - t_{r,j}) \quad (8)$$

$$t'_{t,j+1} = t_{r,j+1} \quad (9)$$

where $j = 1, 2, \dots, m-1$, $t_{r,j}$ and $t_{r,j+1}$ are the retention time values of two adjacent landmark peaks, in R_{pl} , respectively, $t'_{t,j}$, $t'_{t,j+1}$ and $t'_{t,s}$ denote corrected retention time values of the two corresponding adjacent landmark peaks and a peak eluted between the two adjacent landmark peaks in the test sample T_i , respectively.

To correct the retention time of chromatographic peaks not eluting between two landmark peaks, i.e., in the range of $[t_{i,0}, t_{i,1}]$ and $[t_{i,m}, t_{i,m+1}]$, an iterative optimization method is applied to each of these two sections of peaks, respectively.¹⁹ In each optimization process, 30% of the landmark peaks are randomly selected from the pool of landmark peaks $\{(t_{r,1}, t_{i,1}), \dots, (t_{r,m}, t_{i,m})\}$ and a polynomial model fitting is used to correct the retention times of peaks in the section of interest. The polynomial fitting error is computed as follows:

$$\varepsilon = \sum_{j=1}^k (t_{t,j}^o - t_{t,j}^f)^2 \quad (10)$$

where $t_{t,j}^o$ is the z-score transformed retention time of the j -th peak in the test sample T_i , $t_{t,j}^f$ is the fitted retention time of the j -th peak, k is the number of peaks in the test peak list at the section of interest. This process is repeated 1000 times and the model having the minimum fitting error is selected and used for retention time correction.

After the retention time correction on the test sample, all the non-landmark peaks present in the test peak list are aligned to the peaks present in R_{pl} . For each peak pair in a test peak list, a mixture score S_m is calculated using equation (5). A peak pair is considered as a match if its mixture score $S_m \geq S_m^{\min}$. If one peak in the test sample is matched to multiple peaks in R_{pl} or *vice versa*, the peak pair with the maximum mixture score is kept while the remaining matches are discarded. If a peak in the test peak list cannot be matched to any peaks in R_{pl} , this peak is considered as a new peak to R_{pl} and is added to R_{pl} . The updated R_{pl} is then used to align the peaks in the next test peak list, and this process is repeated until all test peak lists are aligned.

Evaluation criteria for analysis of spiked-in data

To evaluate the accuracy of both the analytical platform and the data analysis method, the receiver operating characteristic (ROC) curve is applied, which is created by plotting the true positive rate (TPR) vs. the false positive rate (FPR) at various threshold settings. The TPR and the FPR are defined as follows:

$$TPR = \frac{TP}{TP + FN} \quad (11)$$

$$FPR = \frac{FP}{FP + TN} \quad (12)$$

where TP (true-positive) is the number of spiked-in acids that are detected as analytes with significant peak area changes by statistical analysis, FN (false-negative) is the number of spiked-in acids that are not detected as analytes with significant peak area changes, FP (false-positive) is the number of analytes that are not the spiked-in acids but are detected as

having significant peak area changes, and TN (true-negative) is the number of analytes that are not the spiked-in acids and detected without significant peak area changes. The area under curve (AUC) in the ROC curve is equal to the probability of positive outcomes being higher than the negative ones. The higher the AUC score is, the better the observed accuracy of the test for statistical significance.

Results and discussion

To test the performance of the MetHR method, two sets of data were acquired on a LECO GC–HRTMS instrument. The heterogeneous data were acquired from a mixture of authentic compounds that were analyzed at nominally 0.8, 1.0 and 1.2 mL He/min. The spike-in data were collected from three spike-in samples under the identical instrument conditions. Each sample was analyzed by 5 replicate injections. The instrument data were first reduced into peak lists using ChromaTOF-HRT® software and the peak lists were subjected to alignment and other analysis.

Analysis of heterogeneous data

The mixture of 116 compound standards was analyzed on the GC–HRTMS instrument under three different flow rates. A total of 15 injections were performed with 5 replicate injections $\{S_1^i, S_2^i, S_3^i, S_4^i, S_5^i\}$ for each flow condition, to study the effect of differences in retention times of analytes in the mixture. In order to recognize the same compound from the data acquired under different flow conditions, the retention time values of all peaks detected by ChromaTOF-HRT® software were converted to z-score values so that the z-score transformed retention time values have a normal distribution. Fig. 2A displays the retention time distribution of the compounds acquired under different flow rates. The average retention time difference of each compound between 0.8 and 1.2 mL/min flow rate is 21.5 s. By transforming retention time to z-score values, the retention time distribution difference among the three flow conditions are significantly reduced as depicted in Fig. 2B. The mean z-score difference of the same compound between 0.8 and 1.2 mL/min flow rate is only 0.005.

The peak lists generated by ChromaTOF-HRT® were first manually analyzed to check the detection of the 116 mixed compounds. ChromaTOF-HRT® reported 136–237 chromatographic peaks for each of the 15 samples (injections). The peaks detected in addition to the 116 mixed compounds were primarily low intensity peaks due to either contamination or possible degradation. Of the 116 compounds in the mixture of compound standards, 10 compound standards were not detected in any of the 15 sample injections. Of the remaining 106 compounds, 102 of them were detected in all 15 sample injections, while 2, 1 and 1 specific compound standards were detected in only 12, 11 and 10 sample injections, respectively. The reasons that these compounds were not detected by the ChromaTOF-HRT® software include low concentration, highly overlapping chromatographic peak co-elutions, co-eluting isomers, possible loss to adsorption, peak find issues, and deconvolution issues. For the purposes of the alignment testing, these missing peaks were not investigated further.

Of the 102 compounds detected in all 15 sample injections, 84 compounds were fully aligned in all 15 sample injections. For the other 18 compounds in the set of 102, they were aligned in fewer than all 15 sample injections. The mean value of relative standard deviation of retention time of the fully aligned 84 peaks in the three flow conditions are $1.2 \times 10^{-3} \pm 1.5 \times 10^{-3}$, $1.3 \times 10^{-3} \pm 2.3 \times 10^{-3}$, $6.6 \times 10^{-4} \pm 4.3 \times 10^{-3}$, respectively. A low value of relative standard deviation indicates that the GC–HRTMS system was stable during the experiment and ChromaTOF-HRT® has a high accuracy in determining the

chromatographic peak location. It should be noted, however, that 18 compound standards were not fully aligned in all 15 injections due to the limited spectral similarity between some samples. Manual validation shows that the spectral similarity values of these peaks calculated by equation (3) are sometimes less than 0.75. Fig. 3 depicts two mass spectra of triacontane acquired in this study. The spectral similarity between these two spectra is 0.65. It can be seen that the profiles of the small fragment ions are slightly different between Fig. 3A and Fig. 3B. However, three peaks with large m/z values ($m/z = 401.9734$, 418.9742 and 491.1087) were detected in the spectrum displayed in Fig. 4B while none of these peaks was detected in the spectrum displayed in Fig. 4A. Such a low spectral similarity appears to be due to variation in the spectral deconvolution. By manually lowering the threshold of mass spectral similarity, all these peaks can be fully aligned. However, a lower threshold for mass spectral similarity may introduce a high rate of false positive alignment, especially in analyzing data acquired from complex samples.

Fig. 4 depicts the effectiveness of correcting compound retention time based on the retention time of landmark peaks. Fig. 4A shows the original retention time distribution between samples $S_1^{0.8}$ and $S_3^{0.8}$, while Fig 4B is the distribution of corrected retention time values. Fig. 4C is the original retention time distribution between $S_1^{0.8}$ and $S_1^{1.2}$, two peak lists of the sample analyzed under different flow rates 0.8 mL/min and 1.2 mL/min, respectively. Compared to the peak lists displayed in Fig. 4A, the original retention time values are quite different between $S_1^{0.8}$ and $S_1^{1.2}$ due to the large difference of flow rates. After retention time correction, comparison of the corrected retention time distributions of $S_1^{0.8}$ and $S_1^{1.2}$ is considerably more favorable (Fig. 4D). This demonstrates that retention time correction is effective in aligning the heterogeneous data. Fig. 5 shows the distribution of the relative standard deviation (RSD) of the aligned peaks after the retention time correction for replicates within sets. More than 95% of compounds have an RSD value for retention time across all 15 samples (3 flows and 5 replicates) of less than 0.15%. Heptane has the largest RSD value of 0.69%.

Analysis of spike-in data

A total of three spiked-in samples were prepared and each sample was analyzed on the instrument via five replicate injections. Each group of the five replicate injections identifies a sample group, and a total of three sample groups are formed, i.e., G10, G20 and G40. The concentration of the spiked-in compounds is identical between the replicate injections in a sample group but different between sample groups. By experimental design, the concentration ratio of the spiked-in compounds between the three sample groups should be $G10:G20:G40 = 1:2:4$ if the spiked-in compounds are not present in the liver extract. A total of 15 peak lists were generated by ChromaTOF-HRT® from the spike-in experiments. The number of chromatographic peaks detected from each of these samples ranges from 1000 to 1200.

Manual verification of the peak lists generated by ChromaTOF-HRT® shows that 23 of the 28 spiked-in compounds were automatically detected. Among the 23 compounds, 17 compounds were detected and fully aligned in all 15 samples, while the remaining six compounds (L-valine, glycine, L-methionine, L-serine, L-glutamic acid and L-threonine) were not detected in all the 15 injections. Possible causes of the variation in detecting these six compounds include instrument variation, inaccuracy of spectrum deconvolution, and sample stability. Compound L-valine was detected in 14 samples and was correctly aligned in these 14 samples. Compounds glycine, L-methionine, L-serine, L-glutamic acid and L-threonine were detected in 13, 13, 11, 10 and 4 samples, respectively. All these five compounds were also correctly aligned from the corresponding samples.

By design, the concentrations of the spiked-in compounds in the three sample groups are different from each other. To test whether the concentration differences of the spiked-in compounds can be recognized from the alignment table, a two-tailed *t*-test was used to check the mean difference of the peak area of each compound between sample groups, by setting different *p*-values. Fig. 6 depicts the ROC curve of recognizing the concentration difference of the spiked-in compounds between sample groups using the alignment results. As expected, the FPR increases with the increase of TPR. The TPR levels off at 1.0 when the FPR reaches 0.6 between all the comparing sample groups. The AUC of the ROC curve of G10 vs. G20 is 0.86, while the AUC of the ROC curve of G20 vs. G40 and G10 vs. G40 is 0.92 and 0.88, respectively. A high value of AUC indicates that a high accuracy of recognizing the concentration difference of the spiked-in compounds between sample groups, which is achieved on the basis of correct alignment of the spiked-in compounds.

The results of aligning the peak lists from the spike-in experiments demonstrate that MetHR approach is able to handle the complex data for untargeted analysis. The statistical significance test demonstrates that the aligned results can be used to detect differences in compound concentration between sample groups, which has broad applications in differential metabolomic analyses including disease biomarker discovery. It, however, should be noted that multiple data analysis steps are involved in metabolomics analysis. The variation introduced in any of these analysis steps can affect the overall accuracy of the study. In this study, not all of the spiked-in compounds were detected by ChromaTOF-HRT® even though all detected spiked-in compounds were correctly aligned. Therefore, improving the quality of the entire spectrum of data analysis process are still critical research topics for the application of GC-HRTMS in metabolomics.

The weight factor (0.53, 1.3) used in equation (3) was derived from NIST MS library, which contains mass spectra with unit mass resolution. By analyzing the spike-in samples with different literature reported mass weight factors including (0.53, 1.3), (0.5, 3) and (0.5, 2), a total of 86, 32 and 71 compounds were fully aligned, indicating that the weight factor (0.53, 1.3) has the best performance in alignment. However, it is possible that this weight factor may still not be the optimal one to the high resolution mass spectra. Study in high resolution mass spectral matching could further improve the alignment accuracy.

Conclusions

Significant advances in time-of-flight mass spectrometry have produced a GC-MS system – the GC-HRTMS – that is capable of high resolution mass spectrometric analysis of metabolomic samples. The complexity of the data requires comparable advances in data analysis to allow for the effective utilization of the information provided. A method entitled MetHR has been developed to align the peak lists generated from GC-HRTMS data based on the retention time and fragment ion mass spectrum of each peak. MetHR performs peak list alignment in a *z*-score transformed retention time domain to ensure a normal peak distribution across the samples. It further employs a mixture score to assess the similarity between two peaks by simultaneously evaluating their mass spectral similarity and the closeness of retention time.

The capabilities of MetHR have been tested in two controlled experiments. Analysis of experimental data acquired under three different flow rates indicates that MetHR is able to correctly align the heterogeneous data. The effectiveness of MetHR is further validated by the analysis of the experimental data of multiple mixtures of metabolite extract from mouse liver with 28 spiked-in acids. All of the spiked-in acids detected by ChromaTOF-HRT® software were correctly aligned by MetHR. The statistical significance test was able to correctly recognize the concentration differences of the spiked-in compounds between

sample groups from the alignment table, with the area under curve (AUC) value larger than 0.85, indicating the potential application of MetHR for metabolomics projects such as biomarker discovery.

Acknowledgments

The authors thank Drs. Wei Chen, Jihong Wang, Joe Binkley and Jeffrey S. Patrick at LECO for their discussion and support on this project. This work was supported by LECO Corporation, St. Joseph, MI and National Institute of Health (NIH) grant 1RO1GM087735 through the National Institute of General Medical Sciences (NIGMS).

Notes and references

1. Bando K, Kunimatsu T, Sakai J, Kimura J, Funabashi H, Seki T, Bamba T, Fukusaki E. *J Appl Toxicol*. 2010
2. Kim HJ, Kim JH, Noh S, Hur HJ, Sung MJ, Hwang JT, Park JH, Yang HJ, Kim MS, Kwon DY, Yoon SH. *Journal of proteome research*. 2011; 10:722. [PubMed: 21047143]
3. Constantinou C, Chrysanthopoulos PK, Margarity M, Klapa MI. *Journal of Proteome Research*. 2011; 10:869. [PubMed: 21028881]
4. Pluskal T, Castillo S, Villar-Briones A, Oresic M. *BMC Bioinformatics*. 2010; 11:395. [PubMed: 20650010]
5. Tautenhahn R, Patti GJ, Rinehart D, Siuzdak G. *Analytical chemistry*. 2012
6. Lommen A, Kools HJ. *Metabolomics: Official journal of the Metabolomic Society*. 2012; 8:719. [PubMed: 22833710]
7. Jiang W, Qiu Y, Ni Y, Su M, Jia W, Du X. *Journal of proteome research*. 2010; 9:5974. [PubMed: 20825247]
8. Patrick JS. *LC/GC*. 2012
9. Alonso AE, Binkley J, Siek K. *Current Trends in Mass Spectrometry*. 2011
10. Nicole F, Guitton Y, Courtois EA, Moja S, Legendre L, Hossaert-McKey M. *Bioinformatics*. 2012; 28:2278. [PubMed: 22782550]
11. Eilers PHC. *Analytical Chemistry*. 2004; 76:404. [PubMed: 14719890]
12. Oh C, Huang X, Regnier FE, Buck C, Zhang X. *Journal of chromatography A*. 2008; 1179:205. [PubMed: 18093607]
13. Wang B, Fang A, Heim J, Bogdanov B, Pugh S, Libardoni M, Zhang X. *Anal Chem*. 2010; 82:5069. [PubMed: 20476746]
14. Kim S, Fang AQ, Wang B, Jeong J, Zhang X. *Bioinformatics*. 2011; 27:1660. [PubMed: 21493650]
15. Kim S, Koo I, Fang AQ, Zhang X. *BMC Bioinformatics*. 2011; 12
16. Zhang X, Asara JM, Adamec J, Ouzzani M, Elmagarmid AK. *Bioinformatics*. 2005; 21:4054. [PubMed: 16150809]
17. Stein SE, Scott DR. *J Am Soc Mass Spectr*. 1994; 5:859.
18. Kim S, Koo I, Wei XL, Zhang X. *Bioinformatics*. 2012; 28:1158. [PubMed: 22333245]
19. Wei X, Shi X, Kim S, Zhang L, Patrick JS, Binkley J, McClain C, Zhang X. *Anal Chem*. 2012; 84:7963. [PubMed: 22931487]

Fig. 1.
Workflow of MethHR software.

Fig. 2.

The retention time distributions of the mixture of authentic compounds acquired under different flow rates. (A) is the original retention time distribution. (B) is the distribution of z-score transformed retention time.

Fig. 3.
Two mass spectra of compound triacontane acquired from the mixture of compound standards.

Fig. 4.

The effectiveness of correcting compound retention time values based on the retention time values of landmark peaks. (A) is the original retention time distribution between $S_1^{0.8}$ and $S_3^{0.8}$, two peak lists of two replicate injections under flow rate of 0.8 mL/min. (B) is the distribution of corrected retention time values of compounds between $S_1^{0.8}$ and $S_3^{0.8}$. (C) is the original retention time distribution between $S_1^{0.8}$ and $S_1^{1.2}$, two peak lists of the samples analyzed under different flow rates 0.8 mL/min and 1.2 mL/min, respectively. (D) is the retention time distribution between $S_1^{0.8}$ and $S_1^{1.2}$ after retention time correction.

Fig. 5. The distribution of relative standard deviation (RSD) of the aligned peaks after the retention time correction. Y-axis is accumulative probability. That is a value on y-axis representing the fraction of compounds that have relative standard deviation less than a corresponding value on x-axis.

Fig. 6. The ROC curve of the alignment results. Blue curve denotes the ROC curve of G10 vs. G20, the red one denotes the ROC curve of G20 vs. G40, and the green one is the ROC curve of G10 vs. G40.