Published in final edited form as:

Anal Chim Acta. 2017 August 08; 980: 25–32. doi:10.1016/j.aca.2017.05.002.

Analysis of Stable Isotope Assisted Metabolomics Data Acquired by GC-MS

Xiaoli Wei^{1,4,6,7,*}, Biyun Shi^{1,4,6,7}, Imhoi Koo^{1,4,6,7}, Xinmin Yin^{1,4,6,7}, Pawel Lorkiewicz^{2,5}, Hamid Suhail⁸, Ramandeep Rattan⁸, Shailendra Giri⁸, Craig J. McClain^{2,3,6,7,9}, and Xiang Zhang^{1,2,4,6,7}

¹Department of Chemistry, University of Louisville, Louisville, KY 40292. United States

²Department of Pharmacology & Toxicology, University of Louisville, Louisville, KY 40292. United States

³Department of Medicine, University of Louisville, Louisville, KY 40292. United States

⁴Center for Regulatory and Environmental Analytical Metabolomics, University of Louisville, Louisville, KY 40292. United States

⁵Institute of Molecular Cardiology, University of Louisville, Louisville, KY 40292. United States

⁶University of Louisville Alcohol Research Center, University of Louisville, Louisville, KY 40292. United States

⁷University of Louisville Hepatobiology & Toxicology Program, University of Louisville, Louisville, KY 40292. United States

⁸Henry Ford Health System, Detroit, MI 48202. United States

⁹Robley Rex Louisville VAMC, Louisville, KY 40292. United States

Abstract

Stable isotope assisted metabolomics (SIAM) measures the abundance levels of metabolites in a particular pathway using stable isotope tracers (e.g., ¹³C, ¹⁸O and/or ¹⁵N). We report a method termed signature ion approach for analysis of SIAM data acquired on a GC-MS system equipped with an electron ionization (EI) ion source. The signature ion is a fragment ion in EI mass spectrum of a derivatized metabolite that contains all atoms of the underivatized metabolite, except the hydrogen atoms lost during derivatization. In this approach, GC-MS data of metabolite standards were used to recognize the signature ion from the EI mass spectra acquired from stable isotope labeled samples, and a linear regression model was used to deconvolute the intensity of overlapping isotopologues. A mixture score function was also employed for cross-sample chromatographic peak list alignment to recognize the chromatographic peaks generated by the same metabolite in different samples, by simultaneously evaluating the similarity of retention time

CORRESPONDING AUTHOR: Prof. Xiaoli Wei, Department of Chemistry, University of Louisville, 2210 South Brook Street, Louisville, KY 40292, USA. Phone: +01 502 852 8864. Fax: +01 502 852 8149. xiaoli.wei@louisville.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

and EI mass spectrum of two chromatographic peaks. Analysis of a mixture of 16 ^{13}C -labeled and 16 unlabeled amino acids showed that the signature ion approach accurately identified and quantified all isotopologues. Analysis of polar metabolite extracts from cells respectively fed with uniform ^{13}C -glucose and ^{13}C -glutamine further demonstrated that this method can also be used to analyze the complex data acquired from biological samples.

Graphical abstract

Keywords

IsoGC; stable isotope assisted metabolomics; GC-MS; signature ion; isotopologues

INTRODUCTION

The metabolome is a collection of all small metabolites that are associated with each other to form a large network of molecular reactions. The outputs from one enzymatic reaction are inputs to the other. Stable isotope assisted metabolomics (SIAM) is able to measure the levels of metabolites only within a particular pathway using stable isotope tracers (e.g., ^{13}C , ^{18}O and/or ^{15}N). For example, when a primary substrate such as ^{13}C -glucose is added to a biological system, the relative abundances of the formation of particular isotopologues (metabolites that differ only in the isotopic composition of their molecules) provide information on the routes taken by the initial ^{13}C -atoms in ^{13}C -labeled glucose. The unique isotopologue patterns of metabolic products reflect the biosynthetic history of the metabolite under study. Therefore, SIAM follows the fate of the heavy atoms and their incorporation into a multitude of metabolites produced from the labeled primary substrate. For this reason, identification and quantification of heavy atom- containing metabolites lead to exact biochemical pathway assignment.^{1–7}

Currently, each sample in a SIAM project is analyzed by nuclear magnetic resonance (NMR) spectroscopy, liquid chromatography mass spectrometry (LC-MS), or gas chromatography mass spectrometry (GC-MS). A significant feature of SIAM data is that one metabolite in an unlabeled sample is represented by multiple isotopologues in a labeled sample because different numbers of tracer atoms are incorporated into the metabolite through different biosynthesis pathways. This increased number of isotopologues in the SIAM data not only introduces a high degree of isotopic peak overlapping in the mass spectra, but also a high rate of false isotopologue identification owing to the dramatically increased search space of isotopologue candidates. For this reason, unlabeled samples are always prepared in parallel for each labeled sample group under the identical biological experiment conditions, and the metabolites identified from the unlabeled samples are used to

generate all possible theoretical isotopologues for the identification of isotopologues from the labeled samples.⁸

In case of GC-MS based SIAM experiment, the molecular ion of a metabolite may not present in the electron ionization (EI) mass spectrum. Without the molecular ion, the isotopologues of a metabolite must be determined from the fragment ions of an EI mass spectrum acquired from a labeled sample. To quantify these isotopologues, the fragment ion selected from the EI mass spectrum must contain all trace atoms. However, this cannot be directly achieved using either the EI mass spectra of the unlabeled samples or the labeled samples, because the accuracy of current mass spectral matching-based metabolite identification is only about 80% for analysis of GC-MS data.^{9, 10} The poor identification accuracy introduces a high rate of false isotopologue assignment and quantification. For this reason, the GC-MS data of metabolite standards must be used to assist the assignment of isotopologues in the EI mass spectra acquired from labeled samples.

Compared with metabolic profiling data, the increased number of metabolite species in the NMR, LC-MS or GC-MS data greatly increases the difficulty of data analysis. While multiple software packages have been developed for analysis of SIAM data,^{8, 11–18} most of these efforts have focused on analysis of LC-MS data, including X¹³CMS,¹³ geoRge,¹⁴ mzMatch-ISO,¹⁵ and MIRACLE.¹⁶ While Reaser *et al.* developed a workflow to analyze time-dependent ¹³C-labeling of metabolites analyzed by GC-MS,¹⁸ challenges still remain in analyzing SIAM data acquired on GC-MS, including spectrum deconvolution, isotopologue identification, quantification, metabolite association network construction, and pathway assignment and reconstruction.

The objective of this study was to develop a bioinformatics platform for analysis of SIAM data acquired on a GC-MS system that is equipped with an electron ionization (EI) ion source. We developed a signature ion approach for isotopologue assignment and quantification using the fragment ions in the EI mass spectra of the labeled samples, where the GC-MS data of a set of metabolite standards were used as reference. Metabolite retention index and EI mass spectra were employed for metabolite identification and cross-sample chromatographic peak list alignment. The intensities of overlapped isotopologues were deconvoluted using an iterative linear regression model. The developed data analysis system was tested and validated by analyzing a mixture of metabolite standards and a set of SIAM data acquired in a biological project where cells were fed with uniform ¹³C-glucose and ¹³C-glutamine, respectively.

EXPERIMENTAL SECTION

Metabolite standards

165 metabolite standards (amino acids, fatty acids, organic acids, etc.) were purchased from Sigma-Aldrich Corp. (St. Louis, MO, USA). A 1 mM solution was prepared for each metabolite by dissolving the metabolite into water, methanol or ethanol, depending on the solubility of the metabolite. The solution was then diluted 100× using methanol. 10 µL of the solution was then placed in SpeedVac to remove methanol, followed by freeze dry if water was used to dissolve the metabolite.

Mixture of known metabolites

A mixture of known metabolites was created by mixing two commercially-available amino acid mixtures purchased from Cambridge Isotope Laboratories, Inc. (Cambridge, MA, USA): a mixture of unlabeled amino acids (ALGAL amino acid mixture unlabeled) and a mixture of ^{13}C -labeled amino acids (ALGAL amino acid mixture ($\text{U-}^{13}\text{C}$, 97–99%)). Each of these two mixtures contained 16 amino acids, including L-alanine, L-arginine, L-aspartic acid, L-glutamic acid, L-glycine, L-histidine, L-isoleucine, L-leucine, L-lysine, L-methionine, L-phenylalanine, L-proline, L-serine, L-threonine, L-tyrosine and L-valine. After dissolving each amino acid mixture into 0.1 mM HCl, the two amino acid mixtures were combined in a ratio of weight $W_{\text{unlabeled}}:W_{\text{labeled}} = 10:1, 5:1, 2:1, 1:1, 1:2, 1:5$ and $1:10$, respectively. A total of 21 mixtures were prepared, whereby three mixtures were prepared in parallel for each weight ratio.

Biological samples

Mouse mixed glial cell cultures were prepared from the brain cortex of 2–3 days old C57B6 mouse pups as described before.¹⁹ After 15 days of culture, mixed glial cells (3 million cells/plate) were plated in 65 mm plate for 24 h. Cells were washed three times with RPMI without glucose (Glu) and glutamine (Q) (MP Biomedicals, Santa Ana, CA). Cells were cultured in RPMI media supplemented with either ^{13}C -Glu (15 mM; Sigma-Aldrich, St. Louis, MO, USA) plus ^{12}C -Q (2 mM) or ^{12}C -Glu plus ^{13}C -Q (2 mM; Sigma-Aldrich, St. Louis, MO, USA). Following 8 h of incubation, plates were washed with 100 mM sodium acetate (pH 5) and liquid nitrogen was added in plate to stop the reaction. Plates were stored at -80°C until further analysis.

All cells were scrapped off each dish after adding 1.5 mL of 80% methanol. This process was repeated twice and all solutions were collected into a 15 mL Eppendorf tube using a 1 mL F-tip transfer pipette. After 2 min vortex with 3 glass beads, the sample tube was centrifuged at 5,000 rpm, 4°C for 20 min. 1800 μL of the supernatant was transferred into a glass vial and was then placed in SpeedVac to remove methanol, followed by freeze dry to remove water.

Metabolite derivatization

Each dried sample was dissolved in 30 μL of 20 mg/mL methoxyamine hydrochloride pyridine solution followed by vigorously vortex-mixed for 1 min. Methoxymation was carried out by sonicating for 20 min and incubating at 60°C for 1 h. Derivatization was conducted by adding 20 μL of *N-tert*-Butyldimethylsilyl-*N*-methyltrifluoroacetamide with 1% *tert*-Butyldimethylchlorosilane, *N*-Methyl-*N-tert*-butyldimethylsilyltrifluoroacetamide (MTBSTFA + 1% TBDMSCI) to the glass vial and incubating at 60°C for 1 h. The stock solution was then transferred to a GC vial for analysis. Importantly, the methoxymation and derivatization were carried out just prior to GC-MS analysis.

GC–MS analysis

The derivatized samples were analyzed on a Thermo ITQ 1100 Ion Trap MS equipped with a Trace 1310 GC and an AI 1310 auto sampler. The column was a $60\text{ m} \times 0.25\text{ mm}$ $1\text{dc} \times 0.25\text{ }\mu\text{m}$ 1df , DB-5ms GC capillary column (phenyl arylene polymer virtually equivalent to a

(5%-phenyl)-methylpolysiloxane) purchased from Agilent Technologies (Agilent Technologies J&W, Santa Clara, CA, USA). The helium carrier gas (99.999% purity) flow rate was set to 1.0 mL/min, and the inlet temperature was 280 °C. The column temperature was programmed with an initial temperature of 60 °C for 1.0 min and then ramped at 5 °C/min to 280 °C, and maintained at 280 °C for 15 min. The mass range was set as 29–800 m/z. The ion source chamber was 230 °C with transfer line temperature of 280 °C.

A mixture of C7–C40 *n*-alkanes (Sigma–Aldrich Corp., St. Louis, MO) at a concentration of 2.5 µg/mL was analyzed by GC-MS and the retention times of these *n*-alkanes were used to calculate the retention indices of metabolites detected from the other samples.

THEORETICAL BASIS

Figure 1 depicts the experiment design of a SIAM project and the data analysis workflow of the signature ion approach. Unlabeled samples were prepared in parallel with the labeled samples under identical experiment conditions. The experiment data of metabolite standards and unlabeled samples were used to limit the search space of metabolite candidates for identification of isotopologues, as well as to quantify isotopologues from the labeled samples. To analyze the SIAM data, the experimental data of metabolite standards, unlabeled samples and labeled samples were first deconvoluted into lists of chromatographic peaks, where each peak was represented by its retention time and EI mass spectrum. The lists of chromatographic peaks from the metabolite standards, unlabeled sample and labeled samples were then used as input for further analysis in this study. Details of the data analysis modules are explained in the following sections.

Constructing the database of metabolite standards

After derivatization, each metabolite standard was analyzed by GC-MS under the identical experimental conditions as used for analysis of biological samples. A database of metabolite standards was created and each metabolite was characterized by the name of original metabolite (i.e., metabolite without derivatization), chemical formula of the original metabolite, chemical formula of derivatized metabolite, retention time and EI mass spectrum of the derivatized metabolite. The chemical formula of the original metabolite was used to determine the maximum number of each elements that might be labeled through biosynthesis pathways. The formula of the derivatized metabolite was used to determine the fragment ions in its EI mass spectrum that contain all atoms in the original metabolite (except the hydrogen atoms lost during derivatization), using literature reported empirical fragmentation patterns.²⁰

Aligning the chromatographic peaks of metabolite standards with the chromatographic peaks of unlabeled samples

The purpose of cross-sample chromatographic peak list alignment is to find the chromatographic peaks generated by the same metabolite in different samples. During chromatographic peak list alignment, the chromatographic peak list of metabolite standards was used as a reference peak list. The reference peak list was then aligned with the chromatographic peak lists of the unlabeled samples, one at a time, where a mixture score

S_m was used to simultaneously measure the similarity of the mass spectrum and the retention index between the two chromatographic peaks of interest. S_m is defined as follows,²¹

$$S_m(d_i, S_i|w) = w[1 - \frac{d_i}{\varphi}] + (1-w)S_i \quad (1)$$

where d_i is the Euclidean distance of retention index between the i -th reference peak and the chromatographic peak from the unlabeled sample, φ is the maximum retention index variation of a metabolite between GC-MS analyses, w is a weight factor, S_i is the EI mass spectrum similarity of these two chromatographic peaks calculated using the mixed partial correlation measure developed by Kim et al.¹⁰ The thresholds of S_i , w and φ were respectively set as 0.6, 0.2 and 5 in this study. Any metabolite with $S_i < 0.6$ was not considered as a candidate chromatographic peak for the i -th reference peak. If multiple candidate chromatographic peaks in an unlabeled sample had spectrum similarity scores $S_i > 0.6$ with the i -th reference peak, the candidate chromatographic peak with the largest S_m was aligned with the reference peak.

Signature ion recognition

A significant feature of SIAM experimentation is that one metabolite becomes multiple isotopologues, and the intensities of the isotopic peaks of these isotopologues overlap each other. These intensity of overlapping isotopic peaks and their associated m/z values form an isotopologue cluster, i.e., the isotopologue cluster is a collection of isotopic peaks generated from the isotopologues of the same metabolite, $M_{label} = \{M_0, M_1, M_2, \dots, M_k, M_{k+1}, \dots, M_{k+l}\}$, where M_0 is the monoisotopic peak of the unlabeled isotopologue and is the number of trace atoms (e.g., ^{13}C) that were incorporated into the metabolite during biosyntheses. The isotopic peaks of, $M_{k+1}, \dots, M_{k+l}$ are the natural isotopic peaks of the k -th isotopologue whose monoisotopic peak is M_k . For example, assuming a metabolite with 5 carbon atoms having a signature ion $M_0 = 258$, a maximum of 5 carbon atoms might be labeled during biosyntheses. In addition to its monoisotopic peak $M_5 = 263$ ($k = 5, l = 0$), the heaviest isotopologue could also have other naturally occurring isotopic peaks with m/z values of 264 (M_6) and 265 (M_7). The number of natural isotopes l in the EI mass spectrum depends on the elemental composition and the concentration of the derivatized metabolite, as well as the sensitivity of GC-MS instrument.

In order to deconvolute each isotopologue from the intensity overlapping isotopologue cluster, it is critical to correctly recognize the isotopologue cluster in the EI mass spectra of the labeled samples. However, the number of the trace atoms (e.g., ^{13}C) incorporated into a metabolite is unknown in a labeled sample. For this reason, the EI mass spectrum of a metabolite standard must be used as the reference mass spectrum, from which the monoisotopic peak of each isotopologue cluster can be determined based on the chemical formula of the derivatized metabolite standard and the fragmentation pattern of the derivatization reagent.

Different derivatization reagents have different fragmentation patterns.²⁰ For instance, the EI mass spectrum of a MTBSTFA derivatized metabolite may contain fragment ions $[M-57]^+$, $[M-131]^+$, $[M-73]^+$, $[M-15]^+$, $[M-118]^+$, $[M-103]^+$, $[M-133]^+$, $[M-101]^+$, and/or the parent ion $[M]^+$, where M is the m/z value of parent ion of the derivatized metabolite. Each of these ions contains all atoms of the original (underivatized) metabolite except the active hydrogen atoms lost during derivatization. These ions are termed as signature ions in this study.

To use the EI mass spectrum of a derivatized metabolite standard to accurately recognize its signature ions, the m/z values of all possible signature ions were first calculated using the chemical formula of the derivatized metabolite standard and the fragmentation pattern of derivatization reagent. The corresponding signature ions were then recognized from the EI mass spectrum of the derivatized metabolite standard by matching m/z values. If multiple signature ions were found, the one with the largest m/z difference with its nearest fragment ion was selected as the optimal signature ion. Such a selection criterion was used to minimize the chance of intensity overlapping between other fragment ions and the isotopologue cluster in the labeled samples. The threshold of m/z matching was set as $m/z = 0$ in this study because the GC-MS data had nominal mass resolution.

Isotopologue cluster deconvolution

After detecting the optimal signature ion for each metabolite standard that was also detected in the unlabeled samples, the alignment table of the metabolite standards and the unlabeled samples was used to further align the chromatographic peak lists of each labeled sample using equation (1) by setting a minimum mass spectrum similarity to $S_i > 0.4$. A lower threshold of mass spectrum similarity was used here owing to the isotopologues in the labeled sample. If a chromatographic peak in a labeled sample was aligned to a chromatographic peak of a metabolite standard, the metabolite standard and the isotopologues given rise to the chromatographic peak in the labeled sample were the same metabolite.

To quantify the abundance of each isotopologue in the labeled sample, the isotopologue cluster $M_{label} = \{M_0, M_1, M_2, \dots, M_k, M_{k+1}, \dots, M_{k+l}\}$ was first recognized from the EI mass spectrum of a metabolite in the labeled sample, where M_0 is the m/z value of the optimal signature ion that was determined from the EI mass spectrum of the aligned metabolite standard, k is the number of trace atoms in the chemical formula of the original (underivatized) metabolite, l is the number of natural isotopic peaks of the k -th isotopologue. It is possible that some isotopic peaks might have zero intensities. However, each of the natural isotopic peaks $M_{k+1}, \dots, M_{k+l}$ must have non-zero intensity.

It is possible that the isotopic peaks of other fragment ions might overlap with the isotopic peaks in M_{label} . For this reason, the intensities of those fragment ions were deducted from the intensities of the isotopic peaks in M_{label} as follows. All isotopic peaks in the EI mass spectrum of an unlabeled sample with $m/z \in [M_0, M_{k+l}]$ were selected to form an isotopic peak cluster of the unlabeled sample, $M_{unlabel} = \{M_0, M_1, M_2, \dots, M_k, M_{k+1}, \dots, M_{k+l}\}$. The intensity of the monoisotopic peak M_0 in $M_{unlabel}$ was then set to the value of the monoisotopic peak M_0 in M_{label} . The intensities of the remaining isotopic peaks in $M_{unlabel}$

were scaled proportionally. Then, the intensity of each isotopic peak in $M_{unlabel}$ was deducted from the intensity of the corresponding isotopic peak in M_{label} .

After deducting the overlapping isotopic peaks that did not belong to the isotopologues of interest, each of the remaining isotopic peaks in M_{label} was then tentatively assigned to an isotopologue based on its m/z value. For each assigned isotopologue, the theoretical relative intensities of its natural isotopes were calculated from its chemical formula, by setting the abundance of monoisotopic peak M_0 to be 1.0000.²²

A linear regression model was then used to deconvolute the intensity overlapping isotopic peaks in an isotopologue cluster. Assuming an isotopologue cluster contains n isotopic peaks $Y = \{(m/z_1, y_1), (m/z_2, y_2), \dots, (m/z_n, y_n)\}$ and m isotopologues were assigned to these isotopic peaks, the linear regression model is defined as follows:

$$\begin{bmatrix} x_{11} & \cdots & x_{1m} \\ \vdots & x_{ij} & \vdots \\ x_{n1} & \cdots & x_{nm} \end{bmatrix} \cdot \begin{bmatrix} w_1 \\ \vdots \\ w_m \end{bmatrix} = \begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix} \quad (2)$$

$$\operatorname{argmin}_w \left(\sum_{i=1}^n (y_i - \bar{y}_i)^2 \right) \quad (3)$$

where x_{ij} is the theoretical relative intensity of the j -th isotopologue contributed to the i -th isotopic peak in the isotopologue cluster ($i = 1, \dots, n$ and $j = 1, \dots, m$), w_j is the weight factor of the j -th isotopologue, $\bar{y}_i$ is the fitted intensity of the i -th isotopic peak and y_i is the original intensity of the i -th isotopic peak in the cluster.

To construct the left matrix in equation (2), each x_{ij} was originally assigned with an intensity value of zero. The value of each x_{ij} was then updated after matching the m/z values of the isotopic peaks of the m assigned isotopologues. If the m/z value of an isotopic peak of the j -th isotopologue was the same as the m/z value of the i -th isotopic peak of the isotopologue cluster, the theoretical relative intensity of this isotopic peak of the j -th isotopologue was assigned to x_{ij} . Equation (3) was then applied to find the optimal fitting, and the deconvoluted intensities of isotopologues $\{w_1, \dots, w_m\}$ were calculated by minimizing the fitting error using equation (2). During the linear regression, a tentatively assigned isotopologue was deleted if its weight factor $w_j = 0$. The linear regression was then performed again and the weight factors were updated. This process was repeated until all weight factors were larger than zero.

After the iterative linear regression, the intensity of the j -th isotopologue was represented as $w_j y_i$, where w_j is the optimal weight factor of the j -th isotopologue and y_i is the intensity of the i -th isotopic peak in the isotopologue cluster. The m/z value of the i -th isotopic peak equals to the m/z value of the monoisotopic peak of the j -th isotopologue.

Alignment and normalization of deconvoluted isotopologues

After isotopologue cluster deconvolution, the number of isotopologues in each isotopologue cluster was determined, and an isotopologue list that contained all isotopologues detected from a labeled sample was generated. Each isotopologue was characterized by its retention index, m/z value, intensity, and the number of trace atoms (e.g., ^{13}C), where the number of trace atoms was determined by the m/z difference between the isotopologue and the monoisotopic peak of the unlabeled metabolite.

In order to recognize the abundance alteration of each isotopologue between sample groups, the isotopologue lists of all labeled samples were aligned together by retention index and m/z value matching. The variation windows of retention index and m/z value were set as $\varphi = 5$, $m/z = 0$, respectively.

To quantify the abundance distribution difference of isotopologues between sample groups, an isotope-ratio based normalization was implemented as follows:

$$\widehat{y_{ij}} = \frac{y_{ij}}{\sum_{k=1}^m y_{ik}} \quad (4)$$

where m is the number of isotopologues derived from the i -th metabolite, $\{y_{i1}, y_{i2}, \dots, y_{im}\}$ ($k = 1, 2, \dots, m$) are the intensities of all m isotopologues, y_{ij} is the intensity of the j -th isotopologue generated during deconvolution of isotopologue cluster, and $\widehat{y_{ij}}$ is the normalized intensity of the j -th isotopologue.

After normalization, the conventional statistical significance tests such as two-tailed pairwise t-test with sample permutation or false discovery control (FDR), or partial least square discriminant analysis (PLS-DA), can be used to study the abundance level changes of each isotopologue between sample groups.^{23, 24}

RESULTS AND DISCUSSION

The signature ion method developed in this study was implemented using the MATLAB 2015b software package entitled IsoGC. IsoGC has a modular design, including modules for project information, alignment, normalization and statistical significance tests.

Analysis of the mixtures of metabolite standards

To test and validate the accuracy of IsoGC in deconvoluting overlapping isotopologue peaks, we mixed two commercially-available amino acid mixtures at 7 different weight ratios with three replicates for each weight ratio. One commercial mixture contained 16 unlabeled amino acids, while the other contained the same 16 amino acids, but each of them was uniformly ^{13}C -labeled with a labeling efficiency of 97–99%. Figure 2A shows the EI mass spectrum of L-serine acquired from a sample with mixing weight ratio of $W_{\text{unlabeled}}:W_{\text{labeled}} = 1:5$. IsoGC detected a total of 8 isotopic peaks as the isotopologue peak cluster of L-serine in the EI mass spectrum. After isotopic peak deconvolution, 3 isotopologues were deconvoluted, including L-serine with zero ^{13}C -atom, 2 ^{13}C -atoms and 3 ^{13}C -atoms,

respectively. Among the ^{13}C -labeled isotopologues, $96\% \pm 0.64\%$ of the L-serine was ^{13}C -labeled (Figure 2B). This result agrees with the ^{13}C -labeling efficiency reported by the company.

Table 1 summarizes the analysis results of all 16 amino acids in the mixture. The measured ratio refers to the peak area of an unlabeled amino acid divided by the peak area of ^{13}C -labeled amino acid. The measured ratios of all ^{13}C -labeled amino acids agree with the mixing ratio, with a standard error of mean (SEM) ranging from 0.008 to 0.68. SEM increases as the concentration of labeled amino acids in the mixture decreases. This is reasonable because a low abundance of labeled amino acids introduces high variation in peak detection and isotopic peak deconvolution. Overall, the IsoGC results agree with the expected values, and the small values of SEM demonstrate the high reproducibility of the GC-MS data as well as the accuracy of IsoGC in data analysis.

Analysis of cell culture samples

A total of 11 biological samples were analyzed by GC-MS. These samples formed three sample groups, unlabeled samples ($n = 3$), Glu-group ($n = 4$) and Q-group ($n = 4$). Figure 3A shows the total ion currents (TICs) of three randomly selected samples, one sample per sample group. The high similarity of the TICs demonstrates that ^{13}C -labeling did not introduce notable retention time variation between the isotopologues in the labeled sample and the metabolites in the unlabeled sample. Figure 3B depicts the retention time variation of chromatographic peaks detected in all biological samples. About 95% of peaks have about 0.07% of relative standard deviation (RSD) in retention index. The small RSD of retention index demonstrates that the GC system was very stable during data acquisition.

Figure 4 shows two randomly selected EI mass spectra of pyruvate, in which the upper figure was from an unlabeled sample and the lower one was from a ^{13}C -Glu labeled sample. These two mass spectra have poor spectrum similarity due to ^{13}C -labeling. For this reason, a lower threshold of mass spectrum similarity ($S_i > 0.4$) was used to align the peak lists of the unlabeled samples and the labeled samples. IsoGC correctly aligned the two chromatographic peaks that give rise to these two EI mass spectra using the mixture score function described in equation (1). The signature ion ($m/z = 174$) was also recognized from these two mass spectra. After spectrum deconvolution, 4 isotopologues (0 ^{13}C -atom, 1 ^{13}C -atom, 2 ^{13}C -atoms, and 3 ^{13}C -atoms) were detected for pyruvate in the ^{13}C -Glu sample.

While more than 400 chromatographic peaks were detected in each sample, 66 isotopologues were identified and quantified using the metabolite standards as reference. Using pairwise two-tail t-test with sample permutation, 10 isotopologues were detected with significant changes in their abundance levels between the Glu-group and the Q-group ($p < 0.05$) (data not shown). For instance, Figure 5A shows the relative abundance of each isotopologue of L-alanine in every sample of Glu-group and Q-group, while Figure 5B displays the same data using an error bar plot. The abundance level of isotopologue containing zero ^{13}C -atom was significantly up-regulated in the Q-group with a fold change of 0.76 ($p = 3.5 \times 10^{-3}$). However, the abundance level of the isotopologues with 3 ^{13}C -atoms were dramatically increased in the Glu-group with a fold change of 13.91 ($p = 3.9 \times 10^{-4}$). Isotopologues containing 2 and 3 ^{13}C -atoms have low abundance in both the Glu-group and

the Q-group, and none of them was detected with significant changes in their abundance levels. The biological significance of these discoveries will be reported in a separate paper.

CONCLUSIONS

Stable isotope assisted metabolomics (SIAM) uses stable isotope tracers to support studies of biochemical regulation. To analyze the SIAM data acquired by GC-MS equipped with an electron ionization (EI) ion source, a signature ion approach was developed to quantitatively deconvolute the isotopologue peak clusters from the EI mass spectra acquired from labeled samples. The signature ion contains all atoms present in the metabolite except the hydrogen atoms lost during derivatization. A database of metabolite standards constructed from the GC-MS data of the metabolite standards was used as reference to find the fragmentation pattern of each derivatized metabolite and its signature ion from the labeled samples. A mixture score function was also developed to recognize the chromatographic peaks generated by the same metabolite in different samples. This mixture score function calculates the similarity of two chromatographic peaks by simultaneously evaluating their retention time and EI mass spectra.

Analysis of multiple mixtures containing 16 ^{13}C -labeled amino acids and 16 unlabeled amino acids showed that the developed method can accurately deconvolute the overlapping isotopologue peak cluster. The developed method was also used to analyze polar metabolite extracts of 11 biological samples labeled with ^{13}C -glucose and ^{13}C -glutamine, respectively. A total of 66 isotopologues were detected from the GC-MS data acquired from the labeled samples, of which 10 isotopologues were recognized with significant changes in their abundance levels between the two biological groups with $p < 0.05$.

Acknowledgments

The authors thank Mrs. Marion McClain for review of this manuscript. This work was supported by NIH grant nos. 1P20GM113226, 1P50AA024337, 1U01AA021893-01, 1U01AA021901-01, 1U01AA022489-01A1, and 1R01AA023681-01 (CJM) and the Department of Veterans Affairs 1I01BX002996-01A2 (CJM). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. This work was also in part supported by grants R-1508-05912 from the National Multiple Sclerosis Society and internal grant A20020 from the Henry Ford Hospital (SG).

References

1. Creek DJ, Chokkathukalam A, Jankevics A, Burgess KEV, Breitling R, Barrett MP. Stable Isotope-Assisted Metabolomics for Network-Wide Metabolic Pathway Elucidation. *Analytical Chemistry*. 2012; 84:8442–8447. [PubMed: 22946681]
2. Klein S, Heinzle E. Isotope labeling experiments in metabolomics and fluxomics. *Wires Syst Biol Med*. 2012; 4:261–272.
3. Lane AN, Fan TWM, Higashi RM. Stable isotope-assisted metabolomics in cancer research. *Iubmb Life*. 2008; 60:124–129. [PubMed: 18380001]
4. Li F, Pang XY, Krausz KW, Jiang CT, Chen C, Cook JA, Krishna MC, Mitchell JB, Gonzalez FJ, Patterson AD. Stable Isotope- and Mass Spectrometry-based Metabolomics as Tools in Drug Metabolism: A Study Expanding Tempol Pharmacology. *Journal of Proteome Research*. 2013; 12:1369–1376. [PubMed: 23301521]
5. Kyoto Encyclopedia of Genes and Genomes (KEGG).
6. Wei XL, Shi X, Zhong W, Zhao YT, Tang YA, Sun WL, Yin XM, Bogdanov B, Kim S, McClain C, Zhou ZX, Zhang X. Chronic Alcohol Exposure Disturbs Lipid Homeostasis at the Adipose Tissue-

- Liver Axis in Mice: Analysis of Triacylglycerols Using High-Resolution Mass Spectrometry in Combination with In Vivo Metabolite Deuterium Labeling. *Plos One*. 2013; 8
7. Zhong W, Zhao YT, Tang YA, Wei XL, Shi X, Sun WL, Sun XH, Yin XM, Sun XG, Kim S, McClain CJ, Zhang X, Zhou ZX. Chronic Alcohol Exposure Stimulates Adipose Tissue Lipolysis in Mice. *American Journal of Pathology*. 2012; 180:998–1007. [PubMed: 22234172]
8. Wei X, Lorkiewicz P, Salabei JK, Shi B, Hill BG, Kim S, McClain CJ, Zhang X. Analysis of stable isotope assisted metabolomics data acquired by high resolution mass spectrometry. *Anal Methods*. 2017; 9:2275–2283. [PubMed: 28674558]
9. Koo I, Kim S, Zhang X. Comparative analysis of mass spectral matching-based compound identification in gas chromatography mass spectrometry. *J Chromatogr A*. 2013; 1298:132–138. [PubMed: 23726352]
10. Kim S, Koo I, Jeong J, Wu S, Shi X, Zhang X. Compound identification using partial and semipartial correlations for gas chromatography-mass spectrometry data. *Anal Chem*. 2012; 84:6477–6487. [PubMed: 22794294]
11. Yang S, Nadeau JS, Humston-Fulmer EM, Hoggard JC, Lidstrom ME, Synovec RE. Gas chromatography-mass spectrometry with chemometric analysis for determining C-12 and C-13 labeled contributions in metabolomics and C-13 flux analysis. *Journal of Chromatography A*. 2012; 1240:156–164. [PubMed: 22503618]
12. Yang S, Hoggard JC, Lidstrom ME, Synovec RE. Comprehensive discovery of C-13 labeled metabolites in the bacterium *Methylobacterium extorquens* AM1 using gas chromatography-mass spectrometry. *Journal of Chromatography A*. 2013; 1317:175–185. [PubMed: 24007683]
13. Huang XJ, Chen YJ, Cho K, Nikolskiy I, Crawford PA, Patti GJ. (XCMS)-C-13: Global Tracking of Isotopic Labels in Untargeted Metabolomics. *Analytical Chemistry*. 2014; 86:1632–1639. [PubMed: 24397582]
14. Capellades J, Navarro M, Samino S, Garcia-Ramirez M, Hernandez C, Simo R, Vinaixa M, Yanes O. geoRge: A Computational Tool To Detect the Presence of Stable Isotope Labeling in LC/MS-Based Untargeted Metabolomics. *Analytical Chemistry*. 2016; 88:621–628. [PubMed: 26639619]
15. Chokkathukalam A, Jankevics A, Creek DJ, Achcar F, Barrett MP, Breitling R. mzMatch-ISO: an R tool for the annotation and relative quantification of isotope-labelled mass spectrometry data. *Bioinformatics*. 2013; 29:281–283. [PubMed: 23162054]
16. Mashego MR, Wu L, Van Dam JC, Ras C, Vinke JL, Van Winden WA, Van Gulik WM, Heijnen JJ. MIRACLE: mass isotopomer ratio analysis of U-C-13-labeled extracts. A new method for accurate quantification of changes in concentrations of intracellular metabolites. *Biotechnology and Bioengineering*. 2004; 85:620–628. [PubMed: 14966803]
17. Weindl D, Wegner A, Jager C, Hiller K. Isotopologue ratio normalization for non-targeted metabolomics. *Journal of chromatography A*. 2015; 1389:112–119. [PubMed: 25748542]
18. Reaser BC, Yang S, Fitz BD, Parsons BA, Lidstrom ME, Synovec RE. Non-targeted determination of (13)C-labeling in the *Methylobacterium extorquens* AM1 metabolome using the two-dimensional mass cluster method and principal component analysis. *Journal of chromatography A*. 2016; 1432:111–121. [PubMed: 26787164]
19. Singh J, Olle B, Suhail H, Felicella MM, Giri S. Metformin-induced mitochondrial function and ABCD2 up-regulation in X-linked adrenoleukodystrophy involves AMP-activated protein kinase. *Journal of Neurochemistry*. 2016; 138:86–100. [PubMed: 26849413]
20. Schummer C, Delhomme O, Appenzeller BMR, Wennig R, Millet M. Comparison of MTBSTFA and BSTFA in derivatization reactions of polar compounds prior to GC/MS analysis. *Talanta*. 2009; 77:1473–1482. [PubMed: 19084667]
21. Wei XL, Koo I, Kim S, Zhang X. Compound identification in GC-MS by simultaneously evaluating the mass spectrum and retention index. *Analyst*. 2014; 139:2507–2514. [PubMed: 24665464]
22. Wei XL, Sun WL, Shi X, Koo I, Wang B, Zhang J, Yin XM, Tang YN, Bogdanov B, Kim S, Zhou ZX, McClain C, Zhang X. MetSign: A Computational Platform for High-Resolution Mass Spectrometry-Based Metabolomics. *Analytical Chemistry*. 2011; 83:7668–7675. [PubMed: 21932828]

23. Wei X, Shi X, Koo I, Kim S, Schmidt RH, Arteel GE, Watson WH, McClain C, Zhang X. MetPP: a computational platform for comprehensive two-dimensional gas chromatography time-of-flight mass spectrometry-based metabolomics. *Bioinformatics*. 2013; 29:1786–1792. [PubMed: 23665844]
24. Wei X, Shi X, Kim S, Zhang L, Patrick JS, Binkley J, McClain C, Zhang X. Data preprocessing method for liquid chromatography-mass spectrometry based metabolomics. *Anal Chem*. 2012; 84:7963–7971. [PubMed: 22931487]

1. A signature ion approach is developed for analysis of stable isotope GC-MS data.
2. GC-MS data of compound standards are used for selection of the signature ion.
3. Linear regression model is used to deconvolute the overlapping isotopologue peaks.
4. The developed method was tested by known compounds and biological samples.

Figure 1.

The schema of a SIAM project. The upper section is the project design of a SIAM project and the lower section is the workflow of IsoGC software.

Figure 2. Sample results of deconvoluting metabolite in the mixture of known metabolites. (A) is the EI mass spectrum of L-serine acquired from a sample with weight ratio of $W_{unlabeled}:W_{labeled} = 1:5$. (B) is the relative abundance distribution of three isotopologues of L-serine after isotopic peak deconvolution. Each color coded bar represents the relative abundance of each isotopologue and the error bar refers to the variation of the relative abundance among the three replicates.

Figure 3. Retention time variation among the GC-MS data acquired from the 11 biological samples. (A) Three TICs of three samples randomly selected from the unlabeled samples (top), Glu-group (middle) and Q-group (bottom). (B) The RSD distribution of the retention index of metabolites identified in all 11 biological samples.

Figure 4. Sample EI mass spectra of unlabeled and corresponding labeled metabolite. The upper figure is an EI mass spectrum of pyruvate acquired from an unlabeled sample and the lower EI mass spectrum shows ^{13}C -labeled pyruvate acquired from a labeled sample in Glu-group.

Figure 5.
The abundance distribution of L-alanine isotopologues automatically generated by IsoGC.
(A) 3D plot of the abundance of each isotopologue detected in every sample. (B) Histogram of each isotopologue grouped by sample groups.

Table 1

The analysis results of 16 amino acids mixed with known mixing ratios.

Mixed ratio (unlabeled/labeled)	Measured ratio (unlabeled/labeled)	
	All ^{13}C -labeled isotopologues	Fully ^{13}C -labeled isotopologues
10:1	8.98 ± 0.68	11.00 ± 0.59
5:1	4.22 ± 0.32	5.16 ± 0.26
2:1	1.86 ± 0.15	2.34 ± 0.16
1:1	1.05 ± 0.08	1.17 ± 0.07
1:2	0.49 ± 0.03	0.58 ± 0.03
1:5	0.21 ± 0.01	0.24 ± 0.01
1:10	0.11 ± 0.008	0.12 ± 0.009