1 Non-targeted **determination of ^{13}C -labeling in the** *Methylobacterium extorquens* AM1
2 **metabolome** using the two-dimensional mass cluster method and principal component analysis

3
4 Brooke C. Reaser,^a Song Yang,^{b,c} Brian D. Fitz,^a Brendon A. Parsons,^a Mary E. Lidstrom,^{d,e} and
5 Robert E. Synovec^{a,*}

6
7 (a) Department of Chemistry, University of Washington, Seattle, Washington 98195-2180,
8 USA

9
10 (b) Shandong Province Key Laboratory of Applied Mycology, School of Life Sciences, Qingdao
11 Agricultural University, Shandong Province 266104, PR China

12
13 (c) Key Laboratory of Systems Bioengineering, Ministry of Education, Tianjin University,
14 Tianjin 300072, PR China

15
16 (d) Department of Chemical Engineering, University of Washington, Seattle, Washington
17 98195-2180, USA

18
19 (e) Department of Microbiology, University of Washington, Seattle, Washington 98195-
20 2180, USA

21
22
23
24 * Corresponding author: Department of Chemistry, Box 351700, University of Washington,
25 Seattle, WA 98195-1700, USA. Tel.: +1 206 685 2328; fax: +1 206 685 8665.

26
27 E-mail address: synovec@chem.washington.edu (R. E. Synovec).

28
29
30 Revised version of JCA-15-1481, prepared for *Journal of Chromatography A*

31
32 December 3, 2015

Abstract

A novel analytical workflow is presented for the analysis of time-dependent ^{13}C -labeling of the metabolites in the methylotrophic bacterium *Methylobacterium extorquens* AM1 using gas chromatography time-of-flight mass spectrometry (GC-TOFMS). Using ^{13}C -methanol as the substrate in a time course experiment, the method provides an accurate determination of the number of carbons converted to the stable isotope. The method also extracts a quantitative isotopic dilution time course profile for ^{13}C uptake of each metabolite labeled that could in principle be used to obtain metabolic flux rates. The analytical challenges encountered require novel analytical platforms and chemometric techniques. GC-TOFMS offers advanced separation of mixtures, identification of individual components, and high data density for the application of advanced chemometrics. This workflow combines both novel and traditional chemometric techniques, including the recently reported two-dimensional mass cluster plot method (2D m/z cluster plot method) as well as principal component analysis (PCA). The 2D m/z cluster plot method effectively indexed all metabolites present in the sample and deconvoluted metabolites at ultra-low chromatographic resolution ($R_s \approx 0.04$). Using the pure mass spectra extracted, two PCA models were created. Firstly, PCA was used on the first and last time points of the time course experiment to determine and quantify the extent of ^{13}C uptake. Secondly, PCA modeled the full time course in order to quantitatively extract the time course profile for each metabolite. The 2D m/z cluster plot method found 152 analytes (metabolites and reagent peaks), with 54 pure analytes, and 98 were convoluted, with 65 of the 98 requiring mathematical deconvolution. Of the 152 analytes surveyed, 83 were metabolites determined by the PCA model to have incorporated ^{13}C while 69 were determined to be either metabolites or reagent peaks that remained unlabeled.

Keywords: GS-MS, metabolomics, time-dependent ^{13}C -labeling, principal component analysis, analyte discovery, chemometrics

1. Introduction

Metabolomics is one of the many "-omics" fields, including proteomics, genomics, and transcriptomics that are of great interest to researchers that employ analytical chemistry methods [1]. The term metabolomics encompasses all studies of metabolites, the small molecules that make up an organism's metabolome and the cellular processes within. The metabolome consists of a wide variety of molecules, organic and inorganic, of varying sizes, polarities and volatilities [2,3]. Metabolomics provides a snapshot of the molecular content of a cell at a given moment. To provide further insight, **time course or flux experiments aim** to explore the dynamic and ever-changing nature of cellular metabolites as they move through the metabolic cycles within the organism. **Time course studies** provide additional insight into metabolic pathways and enzymatic activities. These experiments often utilize stable isotopes for labeling (^{13}C , ^{15}N , ^2H , etc.) to elucidate how concentrations [4], reaction rates [5], or isotopomer distributions [6] change over time [7].

Methyloacterium extorquens AM1 is a methylotrophic bacteria of great interest for its possible uses in the production of biofuels, valuable chemicals and catalysts [3]. Like all methylotrophic bacteria, *M. extorquens* AM1 is capable of using single carbon sources such as methane or methanol as its sole carbon source for all of its carbon and energy needs. Although it is one of the most extensively studied methylotrophic bacteria, there exist metabolic cycles in *M. extorquens* AM1 that are still not completely understood, making it an ideal system for isotopic labeling-based **time course studies**. The pathway-specific rearrangement of molecules in the metabolome is highlighted through the use of an isotopic tracer, specifically ^{13}C in this study, facilitating better understanding of the metabolic processes of pathways or enzymes in a

system [7–10]. For *M. extorquens* AM1, ^{13}C -labeled methanol is incorporated by metabolites in the 0 to 60 minute time frame, constituting the time course of our investigation.

The complex nature of the samples in **time-dependent** studies of the metabolome requires powerful analytical tools that can differentiate between stable isotopes, such as nuclear magnetic resonance (NMR) and mass spectrometry (MS) [9]. MS is often coupled with a separation technique, usually liquid chromatography (LC) [11] or gas chromatography (GC) [12,13]. GC-MS is one of the more popular analytical tools for complex samples such as those seen in metabolomics studies [10,14,15]. GC offers the ability to separate volatile and semi-volatile analytes in complex samples for better structural elucidation via the unique fragmentation pattern of MS with electron impact ionization [16–18]. Chemical derivatization is often required to make non-volatile metabolites ready for GC analysis [19]. Time-of-flight mass spectrometry (TOFMS) is an especially effective detector for GC due to the high acquisition rate (100-500 spectra/s), which allows for better deconvolution and analysis of chromatographically overlapped peaks [20].

Due to the significant data density that results from the use of GC-TOFMS as an instrumental platform, many analytical investigations employ chemometrics, which utilizes mathematics to extract useful information from complex data sets [21]. Commercial software packages are available to implement chemometric techniques for interpretation of complex data sets, including software specifically for the analysis of metabolomics data [12,22,23]. Most of these software packages focus on non-targeted alignment and deconvolution or identification of metabolites in the data set, but many exist as “black boxes” allowing little room for input or interpretation by the researcher. Despite the prevalence of available software, the use of advanced chemometrics is limited in metabolomics. Use of GC-TOFMS for metabolomics

106 studies, especially the use of stable isotope labeling for time course investigations, should
107 provide several challenges for the analysis of the complex data sets. These challenges include
108 large amounts of data, excessive chromatographic peak overlap, and difficulty in accurate
109 extraction of pure mass spectra for identification of metabolites of interest. In order to
110 successfully extract pure mass spectra, accurate deconvolution of coeluting metabolites is
111 required. Many software packages provide deconvolution steps, but few explain how the
112 deconvolution is mathematically performed or at what chromatographic resolution the algorithm
113 fails. Deconvolution of coeluting/overlapped analytes down to a chromatographic resolution, R_s ,
114 of 0.1 has previously been obtained using advanced chemometrics in [10], however analytes
115 elute at even lower chromatographic resolution with the extreme case of complete overlap. The
116 challenge of coelution as it exists in time-dependent ^{13}C -labeling experiments requires
117 chemometric deconvolution at a lower resolution for a multiplicity of samples (various time
118 points and injection replicates) using less cumbersome but equally robust chemometric
119 techniques, even in especially challenging examples, when more than two metabolites coelute.
120 Successful identification of stable isotope incorporation (or lack thereof) adds an additional
121 challenge. Few mass spectral databases include libraries for isotopically labeled metabolites for
122 mass spectral matching, compounding the already difficult task of studying the hundreds of
123 metabolites that have yet to be identified. Without a library, other methods must be proposed for
124 the accurate identification of metabolites that have incorporated the stable isotope over
125 metabolites that have not. A successful time-dependent stable isotope labeling experiment
126 would ideally address all of these challenges by incorporating the following into the overall
127 method: (1) metabolite retention time indexing, (2) deconvolution of coeluting peaks, (3)
128 extraction and normalization of pure mass spectra, (4) identification of metabolites that fully

incorporated the stable isotope **over the time course**, (5) identification of the number of carbon atoms in these metabolites that became isotopically labeled, and (6) quantitative determination and visualization of the **time course** profile of the ^{13}C -labeled uptake of the metabolite over the time course. Metabolomics investigations use common practices that can become tedious when the researcher is faced with large data sets. Typically, mass spectra from labeled and unlabeled metabolites are plotted head-to-tail so an analyst may visually determine whether the metabolite was labeled or remained unlabeled. The analyst often must visually determine which mass channels (m/z) demonstrate the metabolite's isotopic uptake, and plot the intensity of these m/z individually in order to visualize the kinetic profile of the time course. This process can be arduous and subjective when metabolites are fragmented during MS detection and can show various different isotopomer distributions throughout the mass spectrum. The information from these isotopomer distributions can be readily misinterpreted, leading to an inaccurate determination and visualization of the **time course (or isotopic dilution)** profile and errant conclusions regarding the number of atoms labeled in the metabolite.

Many metabolomics studies, whether or not they involve stable isotope labeling, consider one or more of the analytical challenges outlined above [2,6,10,14,22,24]. Non-targeted identification of isotopically labeled metabolites [6,14] and quantification of isotopic incorporation [14], signal [10], and flux [6] have been performed on various data sets to study complex metabolomes. The current study reported herein aims to address the challenges noted above while integrating and improving upon important aspects of previous studies, including improved deconvolution of coeluting metabolites [10], and the use of principal component analysis PCA to provide key aspects to the analytical method workflow.

Principal component analysis (PCA) is a popular chemometric tool [25] used extensively in many investigations of complex samples, including classification of gasoline [26], chemical forensics of nerve agents [16], species differentiation of tree oils [27], and metabolomics [2]. PCA is a valuable tool for mathematically separating classes of data, visualized in the scores plot, and the loadings plots are often used to provide complementary information to chemically interpret the basis of any observed class separations in the scores plot. PCA is often used in metabolomics, less often for the quantitative information provided by linear algebra, and more often as a visualization tool to separate classes of samples in studies of disease state [28], food decay [28,29], species or strain identification [2,30], and growth condition [2]. However, few studies utilize PCA to elucidate time course information [2,24] when chemical systems in flux would provide a compelling data set for the use of such a chemometric tool. For example, PCA loadings have been used to elucidate changes in metabolite concentrations in cacao beans in order to quantify the time course decay profile over several days [24].

We hypothesize that the quantitative benefits of PCA have yet to be fully exploited in studies using **time-dependent ¹³C-labeling**. The scores plots from PCA can be used to quantitatively indicate whether or not a metabolite has been labeled over time course, eliminating the subjectivity of an analyst deciding by individually plotting each metabolite head-to-tail. The PCA loadings can also elucidate how many carbon atoms per metabolite have become ¹³C-labeled. The PCA scores can also be readily used to quantitatively generate the profile of the time course for ¹³C uptake, a calculation that was previously obtained from manually (i.e., generally subjectively) determining the correct *m/z* to plot the mass spectral information of a metabolite's **time course** profile [7,14,31].

We report an analytical methodology, outlined in Fig. 1, that provides a novel, non-targeted, **time-dependent ^{13}C -labeling** of the methylotrophic bacterium *M. extorquens* AM1, and, in general, provides a high throughput analytical platform that may be more attractive than previously reported methods for **^{13}C -labeling time course investigations**. A complex GC-TOFMS data set, with many metabolites at a low signal-to-noise ratio (S/N) was studied in order to challenge the proposed method for proof-of-principle demonstration of overcoming difficulties in detecting and identifying metabolites. Our method starts with the non-targeted indexing of the retention time of every detectable metabolite in the chromatogram of *M. extorquens* AM1. Overlapped metabolites are deconvoluted, and the pure mass spectra are extracted and normalized for input into PCA. We previously reported the development of a novel data reduction and representation method for GC-TOFMS that significantly facilitates visualization and analyte peak deconvolution, referred to as the two-dimensional mass channel (m/z) cluster plot method (2D m/z cluster plot method) [32]. The 2D m/z cluster plot method provides an accurate and validated determination of the number of analytes (i.e., metabolites) in overlapped peak situations followed by quantitative deconvolution down to a chromatographic resolution of $R_s \sim 0.03$ [32], which is significantly better than typical deconvolution software [10,33]. After mass spectra are extracted via application of the 2D m/z cluster method, PCA is subsequently employed on the first (fully unlabeled) and last (fully ^{13}C -labeled) time points of the time course as a diagnostic tool for the determination of the extent, or lack thereof, of ^{13}C incorporation by each metabolite over the time course. A quantitative metric is implemented from the information gleaned from this primary PCA model to provide statistical insight into stable isotope incorporation. A second PCA model is then performed including all time points in the time

course of the GC-TOFMS data set to quantitatively determine ^{13}C labeling profiles of all metabolites, whether or not they incorporated the ^{13}C label.

2. Experimental

2.1 Standards

A standards mix of metabolites of biochemical interest was created using ^{12}C (naturally abundant) metabolites derivatized for GC-TOFMS analysis, listed in Table 1. Analytical grade metabolite standards for the mix were obtained from Sigma (St. Louis, MO, USA).

2.2 Batch growth conditions of *M. extorquens* AM1

The sample preparation for the time course experiment proceeded as follows (Fig. 1). *M. extorquens* AM1 (rifamycin-resistant strain) was grown in liquid batch cultures in a minimal medium, as previously described [13]. 120 mM of either ^{12}C (natural abundance) methanol or ^{13}C -labeled methanol was used as the sole carbon source, both at 99% purity, purchased from Sigma (St. Louis, MO, USA). In the middle of the exponential phase ($\text{OD}_{600} = 0.60$ to 0.70), 10 ml of the culture was extracted and rapidly passed through a membrane filter (S-PakTM, Millipore, Billerica, MA, USA) using a pipette. The filter was immediately removed and placed on an agar plate of the same medium with ^{12}C -methanol for 20 min, and then transferred to another agar plate with the same concentration of ^{13}C -methanol. At six time points (0, 3, 6, 15, 35, 70 min) for the time course experiment, replicates of the filter were immediately transferred to a petri dish located on the surface of a Cool Beans Chill BucketTM (ISCBioexpress, Kaysville, UT, USA) at 5°C. To collect cells, the following three sequential rinse solutions were applied: (i) 0.5 mL of 25 mM ice cold HEPES buffer (pH 5.2), (ii) 0.5 ml of 20 °C ethanol solution (75/25,v/v, ethanol/aqueous 25 mM HEPES buffer, (pH 5.2), and (iii) 1.5 ml of 20 °C ethanol.

The resulting solution was transferred to a pre-cooled tube and stored in a -80°C freezer until it was ready for subsequent extractions.

2.3 Preparation of metabolites in *M. extorquens* AM1

Extraction of metabolites from *M. extorquens* AM1 samples was carried out as previously published [11]. Briefly, samples were incubated in a 100 °C water bath for 3 min. The extracted cell suspension was cooled on ice for 5 min, and then cell debris was removed by centrifugation at 5,000 RPM for 5 min. The cell-free metabolite extract was centrifuged at 14,000 RPM for 8 min. The supernatant was dried in a vacuum centrifuge (CentriVap® Concentrator System, Labconco, MO, USA) and stored at -80 °C. For GC-TOFMS analysis, each sample was further derivatized in two steps. First, keto groups were methoximated by adding 50 µl of methoxyamine solution (25 mg/ml methoxyamine hydrochloride in pyridine) and incubated at 60 °C for 30 min. Second, trimethylsilylation was performed by adding 50 µl of a TMS reagent (BSTFA/TMCS, 99:1) and incubated at 30 °C for 90 min.

2.4 GC-TOFMS analysis of metabolites

All GC-TOFMS data were collected using an Agilent 6890N GC (Agilent Technologies, Palo Alto, CA, USA) with a LECO Pegasus III TOFMS as the detector (LECO, St. Joseph, MI, USA). The GC column was a nonpolar BPX-5 (29.9 m x 0.25 mm x 1.0 µm film, SGE Inc., Austin, TX, USA). Ultra high purity helium (Grade 5, 99.999%) was used as the carrier gas (Praxair, Seattle, WA, USA) at a constant flow of 1.5 ml/min, and 1 µl of a given sample was injected via an Agilent 7683 autosampler. All metabolite standards mix samples were injected with a 9:1 split to avoid chromatographic overloading, and all bacteria samples were injected in split-less mode. The GC oven method began at 60 °C with a hold time of 1.25 min after which it was ramped at a rate of 30 °C/min to 280 °C with a hold of 5 min. The ion source was set to 250

°C while the inlet temperature and transfer line were set to 280 °C. After an acquisition delay of 4 min, mass spectra were collected from m/z 50 to 600 at a rate of 100 spectra/s.

2.5 Method of analysis

GC-TOFMS data were imported from the instrument software (ChromaTOF, version 3.32, LECO, St. Joseph, MI, USA) into MATLAB R2012b (Mathworks, Natick, MA, USA) using an in-house software (peg2mat3p8) [34]. Replicates of the standards mix were baseline corrected and smoothed using a Savitzky-Golay filter with a span of 39 spectra (0.39 s) and a third degree polynomial. The mass spectra of the metabolite standards were extracted and averaged. These were imported into MS Search 2.0 (NIST, Gaithersburg, MD, USA) and matched to library entries. A custom library was created from the standards for mass spectral matching to corresponding metabolite peaks detected in the **time course experiment**.

Three replicate chromatograms per time point **of the time course** were aligned using total ion current (TIC) shift function retention time alignment [35]. After alignment, the data were analyzed using the method workflow outlined in Fig 1. The 2D m/z cluster plot method was used to identify and index the retention time of each metabolite by its “mass cluster” [32]. Each mass cluster was identified as either “pure” if a pure mass spectrum was obtained at the cluster center, “re-indexed” if re-indexing was required to obtain a pure mass spectrum, or “deconvoluted” if deconvolution was required to obtain a pure mass spectrum. A signal threshold of 50 was applied at each m/z , corresponding to a S/N of 3. A 70 ms $\times$ 50 ms “cluster box” was used to determine the selective m/z for each analyte present in the cluster plot. A mass spectrum was obtained by averaging 21 spectra across the center of the cluster index for those clusters identified as “pure” and “re-indexed.” Deconvolution was achieved using the 2D m/z cluster method and classical least squares (CLS) to extract the pure mass spectra for the “deconvoluted” clusters. The

application of the method for re-indexing and assisting deconvolution is described in the supplementary content (sections 1 and 2).

Once extracted, the mass spectra (m/z 100-300, excluding m/z 73, and 146-148 so as to exclude peaks coming from the derivatization reagent) for all metabolites were baseline subtracted and normalized, as described in the supplementary content (section 3), prior to analysis using PCA using PLS Toolbox 7.3.1. (Eigenvector Research, Inc., Wenatchee, WA, USA). Two PCA models were constructed, serving as vital steps in the method workflow outlined in Fig. 1, denoted as “PCA on 0 min versus 70 min” and “PCA on Time Course,” respectively. The first PCA step provided an accurate determination of the presence or absence of ^{13}C incorporation over the time course and, in the former, a straightforward method for determining the number of isotopically labeled carbon atoms. The second PCA step provided **time course** profiles of the ^{13}C labeled metabolites. For PCA on 0 min versus 70 min, baseline corrected and normalized mass spectra from the initial (0 min) and final (70 min) time points of the injection replicates of the **time course** were input to PCA for modeling. From the scores of the PCA model, a quantitative metric, degree-of-class separation (DCS) [27,36], was calculated for each metabolite (as indexed by its mass cluster). Based upon the DCS value, each metabolite was then classed as “changing,” indicating ^{13}C uptake by the metabolite, or “not changing,” indicating the lack of detectable ^{13}C uptake. Finally, in the PCA on Time Course model step of the analysis, the mass spectra from the full time course (0, 3, 6, 15, 35, and 70 min) of each metabolite were input into PCA to obtain **time course** profiles for each metabolite.

3. Results and Discussion

3.1 Metabolite standards and M. extorquens AM1 chromatograms

A full chromatogram of the metabolite standards is provided in Fig. 2(A), with each standard peak numbered in the zoom view in Fig. 2(B). The peak number, retention time, and mass spectral match value (NIST library, unless otherwise indicated) for each metabolite standard are summarized in Table 1. Additionally, the analyte number (far right column in Table 1) corresponds to the peak number location of each standard in the target chromatogram of the **time course** on *M. extorquens* AM1 (details of peak number labeling are provided below in the indexing section). The target chromatogram, to which all other chromatograms were aligned, is shown in Fig. 3, which shows the complexity of the bacterial sample, low *S/N* and obvious overlap of many chromatographic peaks. It is this complexity of the bacteria separations that inspired the development and use of the proposed method workflow (provided in Fig. 1) for the non-targeted discovery of metabolites that incorporate ^{13}C over the time course. The overall procedure includes the 2D *m/z* cluster plot method, a novel visualization and deconvolution software developed in-house. The following section describes the application of the 2D *m/z* cluster plot method to accurately extract all of the metabolite mass spectra with sufficient signal from the GC-TOFMS data down to a chromatographic resolution of $R_s \sim 0.04$.

3.2 Non-targeted metabolite indexing and spectra extraction for *M. extorquens* AM1

Mass clusters (i.e., analyte peak center locations) in the GC-TOFMS separations were identified using the 2D *m/z* cluster method, in which peak width *W* is plotted versus the retention time, t_R , and further indexed according to the retention time center of the pure *m/z* (see supplementary content for details, sections 1 and 2). They were then classified into one of three groups: pure mass clusters, re-indexed mass clusters, or deconvoluted mass clusters, as shown in Fig. 4. A single rectangle in Fig. 4 represents one pure mass cluster as defined by the cluster box. Due to the dimensions of the axes in Fig. 4, the singular rectangles representing individual

analytes appear as vertical lines along the peak width W dimension. A color-coded version of Fig 4 can be found in the supplementary content (section 2, Figure S3) where each cluster is classified as pure, re-indexed or deconvoluted. A pure mass cluster corresponds to a pure metabolite peak whose spectrum is taken from the center of the cluster. A re-indexed mass cluster corresponds to a metabolite peak that is partially overlapped with a neighboring metabolite or interferent peak and therefore required its spectrum to be taken from a region corresponding to a sufficiently pure region of the chromatographic peak. A convoluted mass cluster refers to a situation in which two or more metabolites (or interferents) are sufficiently overlapped and required deconvolution via Classical Least Squares (CLS) for extraction of the mass spectra.

The 2D m/z cluster plot method provides visualization of the 1D GC-TOFMS data in two dimensions, and ultimately facilitates extraction of the pure mass spectra. In Fig. 4, the separation between mass clusters is not only present in the retention time t_R dimension, but also in the peak width W dimension. An advantage of the method is that it optimizes the number of metabolites found and successfully analyzed, which is a direct consequence of the ability to identify and resolve analytes to an extremely low $R_s \sim 0.04$ (i.e., overlapped peaks that differ in retention time of only 4% of their peak width at base). The mass cluster provides a highly precise and accurate metabolite retention time index, as many metabolites, and therefore many mass clusters, can be clearly identified within the retention time span of one chromatographic peak width. Therefore, the number of mass clusters is a much better estimate of the number of metabolites since the chromatographic resolution in the mass cluster space is ~ 25 times better than in the original chromatogram at unit resolution, $R_s = 1.0$. Additionally, most peak finders fail to resolve peaks at a chromatographic resolution, $R_s < 0.5$. Indeed, while only 101

chromatographic peaks were counted in the TIC chromatogram in Fig. 3, 152 mass clusters were counted in the 2D m/z cluster plot in Fig. 4 (with more details provided in supplementary content, section 4 and Table S1). Several of the metabolites coelute at very low R_s , and appear as a single chromatographic peak in the TIC chromatogram. Differences in peak width due to peak overlap or chromatographic-dependent band broadening differences cause some mass clusters to appear higher in the 2D scatter plot than others due to their larger peak widths, as seen in Fig. 4.

3.3 Principal component analysis to assess ^{13}C incorporation for *M. extorquens* AM1

In order to assess whether or not a metabolite incorporated ^{13}C during the time course, PCA on the 0 min versus 70 min mass spectra was performed. The two time point extremes, 0 min versus 70 min, have been determined to be sufficient to provide the “fully unlabeled” and “fully labeled” states, respectively, for all of the metabolites for *M. extorquens* AM1 [10,14]. The PC2 versus PC1 scores plot of 5-oxoproline, provided in Fig. 5, is a representative metabolite to demonstrate this stage of the method workflow in Fig. 1. Although 5-oxoproline is not expected as a natural metabolite in the methylotrophic bacterium investigated, it can originate as a breakdown product of glutamine and also can be generated enzymatically as a bona fide metabolite [37]. The PCA scores plot for 5-oxoproline demonstrates how the two classes, 0 min and 70 min, separate from each other on PC1, which contains 99.97% of the variance between the data sets. As the only difference between the sample classes is the time exposed to the ^{13}C -methanol, the variance is attributable to changes in the mass spectra that occur as the metabolite incorporates ^{13}C over time. Examination of the loadings plot for PC1 of the PCA model of 5-oxoproline is provided in Fig. 6A. Here, there is an obvious shift in the PC1 loadings intensity from positive to negative as ^{13}C is incorporated at various variables (i.e., the m/z). The TMS-derivatized molecular weight of 5-oxoproline is 273 amu. The intensity of higher

m/z fragments is often too low for analysis due to the use of electron impact ionization (EI) with the TOFMS. Thus, for most metabolites in this study, the focus is on the M-15 fragment, which corresponds to a loss of a methyl group from one of the TMS groups [38]. For TMS derivatized 5-oxoproline, the M-15 peak is $m/z = 258$. A shift in intensity and sign from $m/z = 258$ to $m/z = 263$ in Fig. 6(A) indicates an incorporation of five ^{13}C molecules, and, in fact, 5-oxoproline has that many carbons in its backbone. This loadings plot can be compared to the traditionally applied head-to-tail plot as provided in Fig. 6(B), which shows the 0 min and 70 min mass spectra before PCA. The same mass channel range, $m/z = 258$ to $m/z = 263$, shows a shift in intensity from ^{12}C (red, positive) to ^{13}C (blue, negative), as in the loadings plot. Additionally, the comparison of the loadings and the head-to-tail plots highlights how PCA maximizes differences and minimizes similarities, as mass channels that are unchanged between the two samples are nearly invisible in the loadings plot. Even though 5-oxoproline was a representative metabolite, all metabolites classified as changing were interpolated in the same way, and it was found that the PCA method presented provided a direct means of determining the number ^{13}C atoms incorporated. Other metabolite examples are presented in the supplementary content, section 5. To avoid analyst intervention to identify the number of ^{13}C atoms, a method to survey the metabolites quickly and reduce the number of loadings plots investigated was incorporated into the method, which is presented next.

In order to readily survey the 152 mass clusters (i.e., metabolites and reagent peaks) to assess ^{13}C incorporation, a quantitative metric of the two class separation between the 0 min replicates and 70 min replicates was utilized. Degree-of-class-separation (DCS) is a quantitative metric of the separation between two classes in the PC2 versus PC1 scores plot [27,36],

$$DCS = \frac{D_{0,70}}{\sqrt{s_0^2 + s_{70}^2}} \quad (1)$$

The DCS takes into account both the Euclidean distance between the center of the two classes ($D_{0,70}$) as a measurement of the difference between classes, and the standard deviation of the Euclidean distances from the center of each group to its replicates (s_0 and s_{70}) as the spread within each class. A large DCS value indicates a large difference between the 0 min and 70 min replicates and a strong likelihood of incorporation of ^{13}C . DCS values were calculated for every PCA model of all 152 mass clusters, and those with a DCS value greater than 1.7 were deemed as changing (i.e., incorporating ^{13}C). A DCS value of 1.7 corresponds to a t-test t value of 2.57, with larger DCS values corresponding to larger t-values. With five degrees of freedom, a two sampled t-value less than 2.57 indicates no statistically significant difference between the two sample classes at the 95% confidence level, meaning there is no difference between the 0 min and 70 min mass spectra. For example, 5-oxoproline has a DCS value of 51.24, a value much larger than the DCS value threshold of 1.7, so 5-oxoproline was labeled as changing (see Table 2). DCS values were calculated for all metabolites in the time course based on their PCA models. A histogram of these DCS values is provided in Fig. 7, showing the distribution of DCS values among the metabolites detected. For those metabolites with DCS greater than 1.7, the loadings plots were then investigated for the determination of the number of ^{13}C atoms incorporated. The incorporation of DCS as a quantitative metric helped eliminate analyst intervention for all 152 clusters, emphasizing only those metabolites that were statistically significant in their scores values in the PCA modeling. This, together with the natural way in which PCA loadings plots emphasize changes, greatly reduces analyst intervention and makes unnecessary the close investigation of the entire mass spectrum as is typically performed. Note that it was imperative to use only the mass spectra from the 0 min (first) and 70 min (last) time points for the determination of the number of ^{13}C atoms incorporated, rather than the full time course, as these

time points represent the fully unlabeled and fully labeled metabolite distributions, respectively.

Although it requires an extra PCA model, it is necessary in order to avoid interferences due to partially labeled isotopomers of the metabolite at earlier time points, which could cause a misidentification of the number of carbons converted to ^{13}C .

3.4 Principal component analysis to determine the time course

To quantitatively visualize and assess the time-dependence of the ^{13}C -labeling, PCA on the entire time course data set was performed. The mass spectra of all injection replicates of the full time course study (0 min, 3 min, 6 min, 15 min, 35 min, and 70 min) were analyzed using PCA, using a method analogous to the previously described step using just the 0 and 70 min samples. The scores on PC1 were plotted versus time as shown in Fig. 8(A), again for the representative metabolite 5-oxoproline. While a time course effect is visually confirmed by this use of PCA, we shall also demonstrate that an accurate quantitative time course is also readily provided by PCA, even elucidating how steep or gradual the change over 0 to 70 minutes for each metabolite as discussed in the supplementary content, section 5. While beyond the scope of the study reported herein, the quantitative information provided by PCA could be used to calculate the metabolic fluxes of each metabolite. Furthermore, when PCA of the full time course was performed on the mass spectra of a metabolite that was classified as not changing based on the DCS value for the 0 min versus 70 min PCA, as in the case of lactate, it was confirmed to have no discernible time course effect when all time points and replicates were analyzed (see supplementary content, sections 5 and 6, Table S2). Using this PCA-based method, the time course effect was quantitatively extracted for all 152 metabolites. Of these, 83 metabolites showed time course effects that indicated incorporation of ^{13}C , and 69 demonstrated no time course effect. Of the 83 changing metabolites, 77 had a DCS of 1.7 or greater as predicted by the

PCA model of 0 min versus 70 min, indicating a time course effect was present prior to construction of the second model. As with any distribution of numbers, as in the case of the DCS histogram provided in Fig. 7, there occur false positives and false negatives, but these account for less than 5% of the metabolites. Seven metabolites displayed time course effects contrary to their designation from the DCS value. Four metabolites were false negatives, where the DCS value appeared below 1.7, but the PCA model of the time course indicated a slight change over time. Three metabolites were false positives having DCS greater than 1.7 but demonstrating no readily observable time course effect. The loadings plot on PC1 of the 0 min versus 70 min model was consulted for the shift in intensity characteristic of a shift from ^{12}C to ^{13}C before being re-classed from changing to not changing or vice versa.

Validation of the workflow method presented for elucidating the time course effect was performed by plotting the commonly applied mass intensity versus time plot, or M+n plot, as provided in Fig. 8(B), where the M-15 peak for 5-oxoproline is plotted as M+0 (no increased abundance of ^{13}C). The intensity of M+0 decreases with time as the other m/z intensity values increase, indicating isotopic dilution of the ^{12}C -labeled molecules by the ^{13}C molecules. Specifically, the M+5 peak increases, showing that 5-oxoproline incorporates five ^{13}C , consistent with what was elucidated in the loadings plot discussed above. This M+n plot, traditionally used to elucidate a time course effect of a metabolite, is both cumbersome and subjective to obtain. It requires the analyst to correctly identify which m/z to plot, a tedious task made difficult for unidentified metabolites, before plotting the intensity of the m/z over time for hopeful elucidation of a time course that can be easily extracted via PCA as demonstrated herein. For metabolites having different time course shapes, the M+n plots demonstrated similar shapes to those seen in the PCA plots, with additional examples provided in the supplementary content, section 5.

Using the information gleaned from both PCA models, the 0 versus 70 min model and the entire time course model, the time course effect was quantitatively reconstructed from the PCA model using Eq. 2,

$$X = 1\bar{x} + TP' \quad (2)$$

In this expression X is the original baseline corrected and normalized mass spectra included in the model, $\bar{x}$ is the mean of the data (since the data is mean centered), T is the scores on PC1 (from Fig 8A) and P' is the loadings on PC1 (from Fig 9). Eq. 2 excludes the error or residuals matrix in order to elucidate how well the PCA model encompasses the variance and models the time course. For this reconstruction, the PCA model on 0 versus 70 min identified the m/z that incorporated ^{13}C , and the PCA model on the full time course provided the values for T and P' for the aforementioned m/z to input into Eq 2. Fig. 10 shows the reconstructed model projected on top of the M+n plots from Fig 8(B) for the M+0 and the M+5 isotopomers. The models for the other isotopomers, (M+1, M+2, M+3, M+4), were calculated but not included for clarity. The coefficient of determination, R^2 , was calculated as a metric for how well the PCA model correlated with the actual data, and the calculation includes all mass isotopomers. As seen in Fig. 10 of 5-oxoproline, $R^2=0.98$, indicating that the reconstruction from PCA is a very good model for the data. This further demonstrates that PCA is as effective as (if not more effective) and less tedious than traditional methods for modeling and quantitatively elucidating time course information in metabolomics. Because the analyst must investigate the loadings plot of the PCA model for 0 versus 70 minutes in order to determine which m/z to include for the determination of labeling, the reconstruction also serves as a validation step. Should the model be reconstructed using m/z that appear to shift, but do not actually contribute to the time course, the reconstruction will not correlate with the time course effect elucidated by the PCA on the time course model,

and will have a lower R^2 value. This concept is shown in Fig. S11(C) and S11(D) in the supplementary content, section 5, where a reconstruction of $m/z = 131$ (M+0) and $m/z = 134$ (M+3) follows the steep trajectory of the time course in Fig. S11(A) and S11(B) with an $R^2 = 0.98$, whereas a reconstruction of $m/z = 243$ (M+0) and $m/z = 248$ (M+5), does not appear to follow the time course effect and has an $R^2 = 0.87$, a much lower value, indicating a poor correlation.

4. Conclusion

We have described a novel analytical method for analysis of a stable isotope time course using GC-TOFMS, based upon the 2D m/z cluster plot method and PCA. The 2D m/z cluster plot method allows for the discovery and deconvolution of coeluting metabolites at low chromatographic resolution. Metabolites were indexed according to their mass cluster locations and were classified as pure or convoluted. Convoluted mass clusters were either re-indexed or deconvoluted according to their proximity to an adjacent mass cluster. Of the 152 mass clusters identified, 54 mass clusters were pure and 98 mass clusters were convoluted. Of these convoluted mass clusters, 33 were re-indexed and 65 were deconvoluted using CLS. Performing PCA on the pure extracted spectra from the mass clusters differentiates those metabolites which are changing from those that are not changing, and the use of DCS as a quantitative metric allows for objective identification of metabolites that incorporated ^{13}C and visualization of the number of carbons converted to ^{13}C with minimal false positives and false negatives. Additionally, PCA performed on the full time course of a changing metabolite elucidates the time course profile for the ^{13}C -uptake of that metabolite. Although it was beyond the scope of this investigation to calculate the metabolic rate of fluxomic uptake, the quantitative information provided by PCA could be used to calculate the rate at which the cellular metabolism

incorporates ^{13}C in a metabolic flux analysis investigation. Of the 152 metabolites surveyed, 83 were identified as changing with time and 69 unchanging with time.

Using a single DCS value as a cutoff for the identification of metabolites that incorporated ^{13}C may result in false positive or false negatives. The use of the 95% confidence interval was used for demonstration of the efficacy of the method, but each investigator should decide what confidence interval is relevant for the data set in question. A range of values near the cutoff value can also be flagged for further investigation, and the researcher can manually decide by examination of the loadings plot whether the metabolite incorporated ^{13}C or not. Further studies into the time-dependent ^{13}C -labeling of mutant and wild type phenotypes as well as applications to comprehensive metabolic flux analysis studies can be investigated using the method described herein.

Acknowledgement. This work was supported by a grant from the DOE (DESC0006871).

Appendix A. Supplementary content associated with this paper has been provided.

References

- [1] J.K. Nicholson, J.C. Lindon, E. Holmes, “Metabonomics”: understanding the metabolic responses of living systems to pathophysiological stimuli via multivariate statistical analysis of biological NMR spectroscopic data, *Xenobiotica* 29 (1999) 1181–1189. doi:10.1080/004982599238047.
- [2] E.M. Humston, K.M. Dombek, J.C. Hoggard, E.T. Young, R.E. Synovec, Time-Dependent Profiling of Metabolites from Snf1 Mutant and Wild Type Yeast Cells, *Anal. Chem.* 80 (2008) 8002–8011. doi:10.1021/ac800998j.
- [3] B.E. Alber, Biotechnological potential of the ethylmalonyl-CoA pathway, *Appl. Microbiol. Biotechnol.* 89 (2011) 17–25. doi:10.1007/s00253-010-2873-z.
- [4] G. Winter, J.O. Krömer, Fluxomics – connecting ‘omics analysis and phenotypes, *Environ. Microbiol.* 15 (2013) 1901–1916. doi:10.1111/1462-2920.12064.
- [5] U. Sauer, Metabolic networks in motion: ^{13}C -based flux analysis, *Mol. Syst. Biol.* 2 (2006) 62. doi:10.1038/msb4100109.
- [6] K. Hiller, C.M. Metallo, J.K. Kelleher, G. Stephanopoulos, Nontargeted Elucidation of Metabolic Pathways Using Stable-Isotope Tracers and Mass Spectrometry, *Anal. Chem.* 82 (2010) 6621–6628. doi:10.1021/ac1011574.
- [7] W. Wiechert, ^{13}C Metabolic Flux Analysis, *Metab. Eng.* 3 (2001) 195–206. doi:10.1006/mben.2001.0187.
- [8] S.B. Crown, M.R. Antoniewicz, Publishing ^{13}C metabolic flux analysis studies: A review and future perspectives, *Metab. Eng.* 20 (2013) 42–48. doi:10.1016/j.ymben.2013.08.005.
- [9] S. Klein, E. Heinzle, Isotope labeling experiments in metabolomics and fluxomics, *WIREs Syst. Biol. Med.* 4 (2012) 261–272.
- [10] S. Yang, J.S. Nadeau, E.M. Humston-Fulmer, J.C. Hoggard, M.E. Lidstrom, R.E. Synovec, Gas chromatography–mass spectrometry with chemometric analysis for determining ^{12}C and ^{13}C labeled contributions in metabolomics and ^{13}C flux analysis, *J. Chromatogr. A* 1240 (2012) 156–164. doi:10.1016/j.chroma.2012.03.072.
- [11] S. Yang, M. Sadilek, M.E. Lidstrom, Streamlined pentafluorophenylpropyl column liquid chromatography–tandem quadrupole mass spectrometry and global ^{13}C -labeled internal standards improve performance for quantitative metabolomics in bacteria, *J. Chromatogr. A* 1217 (2010) 7401–7410. doi:10.1016/j.chroma.2010.09.055.
- [12] R. Goodacre, S. Vaidyanathan, W.B. Dunn, G.G. Harrigan, D.B. Kell, Metabolomics by numbers: acquiring and understanding global metabolite data, *Trends Biotechnol.* 22 (2004)

245–252. doi:10.1016/j.tibtech.2004.03.007.

- [13] S. Yang, M. Sadilek, R.E. Synovec, M.E. Lidstrom, Liquid chromatography–tandem quadrupole mass spectrometry and comprehensive two-dimensional gas chromatography–time-of-flight mass spectrometry measurement of targeted metabolites of *Methylobacterium extorquens* AM1 grown on two different carbon sources, *J. Chromatogr. A* 1216 (2009) 3280–3289. doi:10.1016/j.chroma.2009.02.030.
- [14] S. Yang, J.C. Hoggard, M.E. Lidstrom, R.E. Synovec, Comprehensive discovery of ¹³C labeled metabolites in the bacterium *Methylobacterium extorquens* AM1 using gas chromatography–mass spectrometry, *J. Chromatogr. A* 1317 (2013) 175–185. doi:10.1016/j.chroma.2013.08.059.
- [15] W. Zeng, J. Hazebroek, M. Beatty, K. Hayes, C. Ponte, C. Maxwell, et al., Analytical Method Evaluation and Discovery of Variation within Maize Varieties in the Context of Food Safety: Transcript Profiling and Metabolomics, *J. Agric. Food Chem.* 62 (2014) 2997–3009. doi:10.1021/jf405652j.
- [16] C.G. Fraga, G.A. Pérez Acosta, M.D. Crenshaw, K. Wallace, G.M. Mong, H.A. Colburn, Impurity Profiling to Match a Nerve Agent to Its Precursor Source for Chemical Forensics Applications, *Anal. Chem.* 83 (2011) 9564–9572. doi:10.1021/ac202340u.
- [17] F.J. Santos, M.T. Galceran, Modern developments in gas chromatography–mass spectrometry-based environmental analysis, *J. Chromatogr. A* 1000 (2003) 125–151. doi:10.1016/S0021-9673(03)00305-4.
- [18] Z.S. Khan, R.K. Ghosh, R. Girame, S.C. Utture, M. Gadgil, K. Banerjee, et al., Optimization of a sample preparation method for multiresidue analysis of pesticides in tobacco by single and multi-dimensional gas chromatography-mass spectrometry, *J. Chromatogr. A* 1343 (2014) 200–206. doi:10.1016/j.chroma.2014.03.080.
- [19] O. Fiehn, J. Kopka, R.N. Trethewey, L. Willmitzer, Identification of Uncommon Plant Metabolites Based on Calculation of Elemental Compositions Using Gas Chromatography and Quadrupole Mass Spectrometry, *Anal. Chem.* 72 (2000) 3573–3580. doi:10.1021/ac991142i.
- [20] M.M. van Deursen, J. Beens, H.-G. Janssen, P.A. Leclercq, C.A. Cramers, Evaluation of time-of-flight mass spectrometric detection for fast gas chromatography, *J. Chromatogr. A* 878 (2000) 205–213. doi:10.1016/S0021-9673(00)00300-9.
- [21] K.R. Beebe, B.R. Kowalski, An Introduction to Multivariate Calibration and Analysis, *Anal. Chem.* 59 (1987) 1007A–1017A. doi:10.1021/ac00144a725.
- [22] K. Hiller, C. Metallo, G. Stephanopoulos, Elucidation of Cellular Metabolism Via Metabolomics and Stable-Isotope Assisted Metabolomics, *Curr. Pharm. Biotechnol.* 12

(2011) 1075–1086.

- [23] W. Niu, E. Knight, Q. Xia, B.D. McGarvey, Comparative evaluation of eight software programs for alignment of gas chromatography–mass spectrometry chromatograms in metabolomics experiments, *J. Chromatogr. A* 1374 (2014) 199–206. doi:10.1016/j.chroma.2014.11.005.
- [24] E.M. Humston, J.D. Knowles, A. McShea, R.E. Synovec, Quantitative assessment of moisture damage for cacao bean quality using two-dimensional gas chromatography combined with time-of-flight mass spectrometry and chemometrics, *J. Chromatogr. A* 1217 (2010) 1963–1970. doi:10.1016/j.chroma.2010.01.069.
- [25] S. Wold, K. Esbensen, P. Geladi, Principal component analysis, *Chemom. Intell. Lab. Syst.* 2 (1987) 37–52. doi:10.1016/0169-7439(87)80084-9.
- [26] J.S. Nadeau, R.B. Wilson, J.C. Hoggard, B.W. Wright, R.E. Synovec, Study of the interdependency of the data sampling ratio with retention time alignment and principal component analysis for gas chromatography, *J. Chromatogr. A* 1218 (2011) 9091–9101. doi:10.1016/j.chroma.2011.10.031.
- [27] P.J. Dunlop, C.M. Bignell, J.F. Jackson, D.B. Hibbert, Chemometric analysis of gas chromatographic data of oils from Eucalyptus species, *Chemom. Intell. Lab. Syst.* 30 (1995) 59–67. doi:10.1016/0169-7439(95)00036-4.
- [28] H. Gu, Z. Pan, B. Xi, V. Asiago, B. Musselman, D. Raftery, Principal component directed partial least squares analysis for combining nuclear magnetic resonance and mass spectrometry data in metabolomics: Application to the detection of breast cancer, *Anal. Chim. Acta* 686 (2011) 57–63. doi:10.1016/j.aca.2010.11.040.
- [29] V.G. Uarrota, R. Moresco, B. Coelho, E. da C. Nunes, L.A.M. Peruch, E. de O. Neubert, et al., Metabolomics combined with chemometric tools (PCA, HCA, PLS-DA and SVM) for screening cassava (*Manihot esculenta* Crantz) roots during postharvest physiological deterioration, *Food Chem.* 161 (2014) 67–78. doi:10.1016/j.foodchem.2014.03.110.
- [30] I. Olivier, D.T. Loots, A metabolomics approach to characterise and identify various *Mycobacterium* species, *J. Microbiol. Methods* 88 (2012) 419–426. doi:10.1016/j.mimet.2012.01.012.
- [31] K. Nöh, K. Grönke, B. Luo, R. Takors, M. Oldiges, W. Wiechert, Metabolic flux analysis at ultra short time scale: Isotopically non-stationary ¹³C labeling experiments, *J. Biotechnol.* 129 (2007) 249–267. doi:10.1016/j.jbiotec.2006.11.015.
- [32] B.D. Fitz, B.C. Reaser, D.K. Pinkerton, J.C. Hoggard, K.J. Skogerboe, R.E. Synovec, Enhancing Gas Chromatography–Time of Flight Mass Spectrometry Data Analysis Using Two-Dimensional Mass Channel Cluster Plots, *Anal. Chem.* 86 (2014) 3973–3979.

doi:10.1021/ac5004344.

- [33] S.E. Stein, An integrated method for spectrum extraction and compound identification from gas chromatography/mass spectrometry data, *J. Am. Soc. Mass Spectrom.* 10 (1999) 770–781. doi:10.1016/S1044-0305(99)00047-1.
- [34] K.M. Pierce, J.C. Hoggard, Chromatographic data analysis. Part 3.3.4: handling hyphenated data in chromatography, *Anal. Methods* 6 (2014) 645–653. doi:10.1039/C3AY40965A.
- [35] J.S. Nadeau, B.W. Wright, R.E. Synovec, Chemometric analysis of gas chromatography–mass spectrometry data using fast retention time alignment via a total ion current shift function, *Talanta* 81 (2010) 120–128. doi:10.1016/j.talanta.2009.11.046.
- [36] K.M. Pierce, J.L. Hope, K.J. Johnson, B.W. Wright, R.E. Synovec, Classification of gasoline data obtained by gas chromatography using a piecewise alignment algorithm combined with feature selection and principal component analysis, *J. Chromatogr. A* 1096 (2005) 101–110. doi:10.1016/j.chroma.2005.04.078.
- [37] M. Mazelis, H.M. Pratt, In Vivo Conversion of 5-Oxoproline to Glutamate by Higher Plants, *Plant Physiol.* 57 (1976) 85–87. doi:10.1104/pp.57.1.85.
- [38] T. Kind, G. Wohlgemuth, D.Y. Lee, Y. Lu, M. Palazoglu, S. Shahbaz, et al., FiehnLib – mass spectral and retention index libraries for metabolomics based on quadrupole and time-of-flight gas chromatography/mass spectrometry, *Anal. Chem.* 81 (2009) 10038–10048. doi:10.1021/ac9019522.

Figure Captions

Figure 1. A flowchart of the novel method workflow used for the analysis of the time-dependent ^{13}C -labeling of *Methylobacterium extorquens* AM1.

Figure 2. (A) Total ion current (TIC) chromatogram of the standards mix listed in Table 1. The region in the dashed box is shown in Fig. 2(B).

Figure 2. (B) Zoom in of the area in red dashed box in Fig. 2(A), with peak numbers corresponding to those listed in Table 1 to indicate where metabolites in the standards mix elute.

Figure 3. TIC chromatogram of one replicate of the fully unlabeled bacteria *M. extorquens* AM1, which served as the target chromatogram for alignment. The inset shows the complexity of the data, requiring advanced chemometrics for analysis of low intensity and convoluted analyte peaks.

Figure 4. Mass cluster plot of all analytes indexed by the 2D m/z cluster plot method. Each rectangular line represents the mass cluster of one analyte, including multiple selective m/z .

Figure 5. Scores on PC2 versus scores on PC1 plot of the first PCA model on the 0 min vs. 70 min injection replicates for the metabolite 5-oxoproline. The classes (0 min, red triangles; 70 min, green stars) are well separated by the scores on PC1, which encompasses 99.97% of the variance in the data set. The DCS value for 5-oxoproline is 51.24, indicating ^{13}C incorporation over the time course.

Figure 6. (A) Loadings on PC1 for the first PCA model on the 0 min versus 70 min injection replicates of 5-oxoproline. The DCS value based on the scores plot in Fig. 5 indicated isotopic incorporation, prompting analyst intervention to reveal the number of carbons that incorporated. The loadings plot shows clearly a shift in variable (i.e., mass spectral) intensity from m/z 258 to 263, indicating that 5 carbons of the metabolite 5-oxoproline converted from ^{12}C to ^{13}C between 0 and 70 min of the time course. The m/z 258 corresponds to the TMS-derivatized molecular weight of 5-oxoproline, minus a methyl group (M-15) of the TMS group. 5-oxoproline has 5 carbon atoms in its backbone, all of which converted to ^{13}C .

Figure 6. (B) A traditional head-to-tail plot with the average mass spectrum of three injection replicates of the 0 min (red, positive) and 70 min (blue, negative) time points of 5-oxoproline. This traditional view of the data shows the shift in mass spectral intensity from m/z 258 to 263, as in the loadings plot in Fig. 6 (A), validating the use of PCA model loadings plot to discover the extent of isotopic incorporation by a metabolite.

Figure 7. Histogram of DCS values calculated from the scores plots of the first PCA model (0 min versus 70 min) for all 152 clusters indexed as in Fig. 4. DCS value of less than 1.7 indicates no statistical difference between the two sample classes, indicating no ^{13}C incorporation. All analytes were classified based on their DCS value as changing, or not changing as indicated by the black dotted line at $\text{DCS} = 1.7$. Three false positives and four false negatives were discovered by the examination of the second PCA model on the time course.

Figure 8. (A) Scores on PC1 versus time for the PCA model of the full time course, including all injection replicates at all time points for 5-oxoproline. A clear time course effect is quantitatively elucidated through PCA, indicating steady ^{13}C incorporation over the time course.

Figure 8. (B) A traditional M+n line plot of the average mass spectral intensity of the injection replicates of all time points of 5-oxoproline. This shows the same time course effect as the scores on PC1 vs time elucidated by the PCA model in (A), validating PCA as a quantitative method for time-dependent ^{13}C -labeling elucidation.

Figure 9. The loadings on PC1 of the second PCA model of the full time course of 5-oxoproline. A clear shift in variable (i.e., mass spectral) intensity is seen at m/z 258-263 as in Fig. 6(A).

Figure 10. The reconstruction of the PCA model data of 5-oxoproline overlaid with the original mass spectral intensity. The M+0, represented by circles, corresponds to m/z 258 and no ^{13}C incorporation. The blue solid circles and solid line correspond to the M+0 normalized mass spectral intensity versus time as in Fig. 9(B), while the red rings correspond to the M+0 intensity versus time as calculated based on the PCA model results. The same is done for M+5, represented by triangles, which corresponds to m/z 263 and all 5 carbons converted to ^{13}C . The blue solid triangles with the dotted line is the original data, the red hollow triangles are reconstructed from the PCA model. The correlation coefficient was calculated, and an $R^2 = 0.98$ indicates the PCA model is a good representation of the original data, quantitatively extracts the time course profile without need for analyst intervention, and validates the choice of m/z 258 as the base peak for the carbon count.

Figure 1

Figure 2A Black and White

Figure 2A Color

Figure 2B Black and White

Figure 2B Color

Figure 3 Black and White

Figure 3 Color

Figure 4

Figure 5 Black and White

Figure 5 Color

Figure 6A Black and White

Figure 6A Color

Figure 6B Black and White

Figure 6B Color

Figure 7 Black and White

Figure 7 Color

Figure 8A Black and White

Figure 8A Color

Figure 8B Black and White

Figure 8B Color

Figure 9 Black and White

Figure 9 Color

Figure 10 Black and White

Figure 10 Color

Tables: Non-targeted determination of ^{13}C -labeling in *Methylobacterium extorquens* AM1 metabolome using the two-dimensional mass cluster method and principal component analysis

Table 1. List of 25 metabolites in the standards mixture listed in order of elution as indicated by the peak number, which corresponds to the peak indicated in Fig. 2(B) at the retention time t_R (s). Match Value (MV) to the NIST library, unless otherwise indicated, for each compound are listed. The analyte number # corresponds to the elution location of that metabolite in the bacterial fluxomic chromatograms. * = match to Kegg library, + = match to Feihn library.

Peak #	Metabolite Standard	t_R (s)	MV	Analyte #
1	Pyruvate	320.0	952*	43
2	Lactate	321.0	661 ⁺	44
3	Alanine	338.0	885	52
4	Oxalate	347.0	918	59
5	3-Hydroxybutyrate	357.0	903	63
6	Methylmalonate	376.0	892	
7	Valine	377.0	869	70
8	Leucine	393.0	827	65
9	Ethylmalonate	396.5	884	
10	Isoleucine	401.0	844	
11	Threonine	403.5	874	80
12	Proline	408.0	722	
13	Glycine	408.0	746	82
14	Succinate	408.0	538*	83
15	Glycerate	408.0		
16	Methylsuccinate	411.0	898	85
17	Serine	417.0	844	
18	Fumarate	418.0	861	88
19	Methylmaleate	432.0	908	
20	Malate	452.5	861	102
21	Methionine	470.0	848	
22	5-Oxoproline	474.0	846	112
23	Phenylalanine	500.5	852	
24	Isocitrate	532.5	777	
25	Citrate	534.0	802	135

Table 2. List of metabolite successfully identified by match value, sorted by their rank based on DCS value, with the percent variance encompassed by both the first and second PCA models, the analyte number indicating elution location in the chromatogram, identification and match value to the NIST, flux of Changing (C) or Not Changing (NC), and carbon count (C#).

Rank	DCS	PC1 %	Analyte #	Name	MV	Flux	C#	Fluxomic PC1 %
1	51.2	99.97	112	5-oxoproline	914	C	5	97.67
2	44.5	99.90	52	Alanine	848	C	3	95.88
3	44.1	99.96	130	Putrescine	871	C	4	99.88
4	34.6	99.91	100	l-aspartic acid	807	C	4	95.94
6	28.3	99.63	139	c00029d: UDP glucose fragment shared with G1P	909	C	4	92.86
7	22.5	99.78	129	D-glucose 1-phosphate main TMS derivative	896	C	5	85.61
8	17.6	99.57	80	L-threonine	760	C	4	91.81
9	15.3	99.68	70	Valine	860	C	5	94.11
12	13.2	99.41	85	Methylsuccinate	771	C	5	89.18
24	7.3	98.04	43	Pyruvate	851	C	3	92.43
27	6.5	98.45	76	acetyl-CoA TMS phosphoric acid fragment	937	C	4	90.16
29	5.9	96.34	82	Glycine	852	C	2	62.87
31	5.9	96.44	135	Citrate	849	C	6	81.44
37	4.4	93.41	102	Malate	928	C	4	86.91
41	3.6	86.70	65	Leucine	743	C	6	71.40
44	3.2	86.70	63	3-hydroxybutyrate	806	C	4	44.73
45	3.2	85.90	88	fumarate	588	C	4	60.95
49	2.9	96.65	42	benzene, 1,2,3-trimethyl-	899	C	2	97.15
53	2.7	95.13	83	succinate	814	C	3	84.52
59	2.3	93.53	40	benzene, 1-ethyl-2-methyl-	878	C	3	91.95
66	1.9	69.79	141	galactose	856	C	3	88.63
68	1.9	88.05	59	oxalate	710	C	2	83.63
69	1.9	88.92	13	trifluoromethyl-bis-(trimethylsilyl)methyl ketone	896	C	4	81.48
77	1.7	82.62	26	disilathiane, hexamethyl-	813	C	2	80.95
79	1.6	74.56	45	pyridine, 3-trimethylsiloxy-	767	NC		74.79
83	1.6	96.89	98	trimethylsilyl 4-methylbenzoate	905	NC		95.84
96	1.2	61.70	57	pentasiloxane, dodecamethyl-	822	NC		93.29

98	1.2	53.26	107	3,6,9,12-tetraoxa-2,13-disilatetradecane, 2,2,13,13-tetramethyl-	874	NC		56.71
101	1.2	41.62	145	D-sorbitol 6TMS	898	NC		57.63
105	1.1	56.33	27	1,2-bis(trimethylsiloxy)ethane	830	NC		48.85
106	1.1	92.82	69	butanoic acid, 3-methyl-3-[(trimethylsilyl)oxy]-, trimethylsilyl ester	873	C	4	77.32
107	1.1	62.04	41	silane, (cyclohexyloxy)trimethyl-	869	NC		61.69
116	0.9	76.40	47	glycolic Acid	917	NC		56.06
124	0.7	90.06	44	lactate	932	NC		79.46
132	0.6	67.77	68	valeric	661	NC		60.05