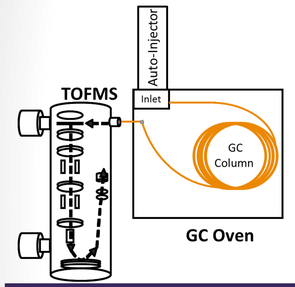
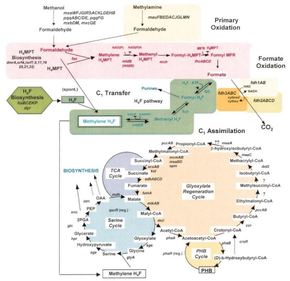
Non-targeted determination of ^{13}C-labeling in *Methylobacterium extorquens* AM1 using principal component analysis

Brooke C. Reaser, Brian D. Fitz, Robert E. Synovec
Department of Chemistry, University of Washington, Seattle WA

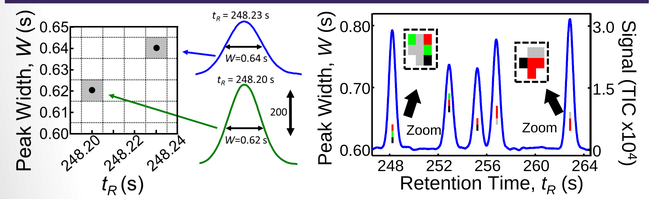
Abstract

A novel analytical workflow for the analysis of *Methylobacterium extorquens* AM1 using ^{13}C-time course analysis and gas chromatography time-of-flight mass spectrometry (GC-TOFMS) is presented. 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 MFA experiment to determine and quantify the extent of ^{13}C uptake. Secondly, PCA modeled the full time course in order to quantitatively extract the fluxomic profile for each metabolite.



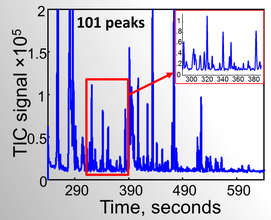


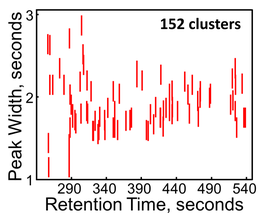
2D *m/z* Cluster Plot Method



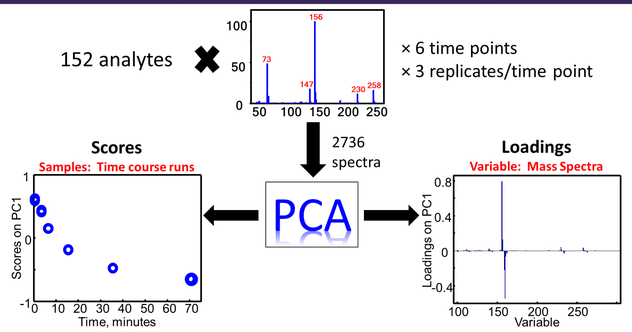
Above left, an illustration of the 2D *m/z* cluster plot method. The algorithm locates peak maxima for all mass channels (*m/z*) and calculates the peak width at half height, *W*, for each peak. Each point is then plotted (*W* vs. t_R) for each peak above a signal threshold. Slight differences in width and retention time arise due to detector noise. Above right, TIC overlaid with the 2D *m/z* cluster plot of the respective region. Selectivity is produced in both width and retention dimensions. Points in the cluster plot with the same t_R and *W* are color coded based upon the frequency of their occurrence: 1 *m/z*: gray; 2 *m/z*: black; 3 *m/z*: green; 4+ *m/z*: red.

Below left, Chromatogram of the TIC of the separation of the metabolites in *M. extorquens* AM1. A traditional peak finding algorithm found 101 peaks in the separation. Below, right the 2D *m/z* cluster plot method for the same separation, with each red box corresponding to an individual analyte. 152 mass clusters were found, a 50% increase in the number of analytes discovered compared to the traditional method.

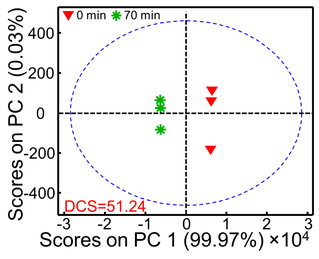




Principal Component Analysis



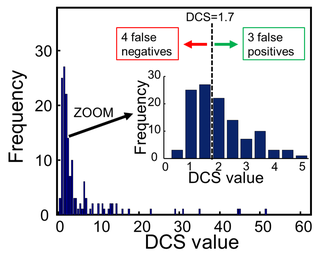
First PCA Model: Determination of ^{13}C Uptake



Left: Scores on PC2 versus scores on PC1 plot of the first PCA model on the mass spectra of 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 of the fluxomic experiment.

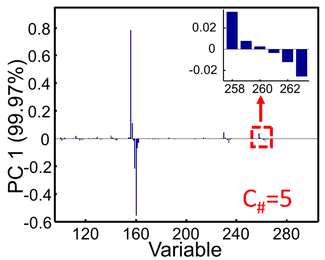
$$DCS = \frac{D_{0,70}}{\sqrt{s_0^2 + s_{70}^2}}$$

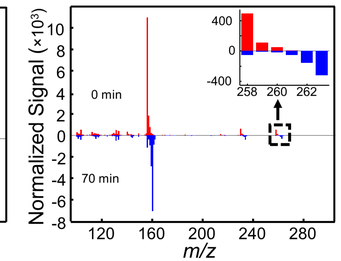
The equation for the Degree-of-Class (DCS) separation, which includes the Euclidean distances between the two classes as well as the standard deviation within the classes. The larger the DCS the more separation between the two classes.



Above: Histogram of DCS values calculated from the scores plots of the first PCA model (0 min versus 70 min) for all 152 clusters indexed. 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 DCS = 1.7. Three false positives and four false negatives were discovered by the examination of the second PCA model on the fluxomic time course.

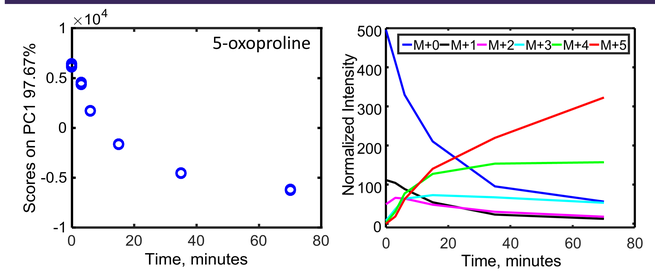
Below: Loadings on PC1 for the first PCA model on the 0 min versus 70 min injection replicates of 5-oxoproline, which clearly shows a shift in variable (i.e., mass spectral) intensity from *m/z* 258 to 263, indicating that 5 carbons of the metabolite 5-oxoproline converted from ^{12}C to ^{13}C between 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.



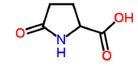


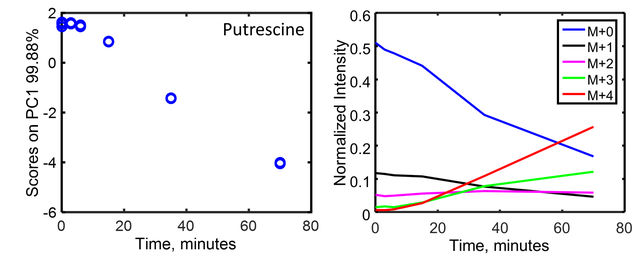
Left: A traditional head-to-tail plot with the average mass spectrum of three injection replicates of the 0 min (red, positive) and 70 min (blue, negative) time points of 5-oxoproline. This traditional view of the data shows the shift in mass spectral intensity from *m/z* 258 to 263, as in the loadings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

Second PCA Model: Time-course Profile

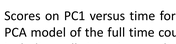


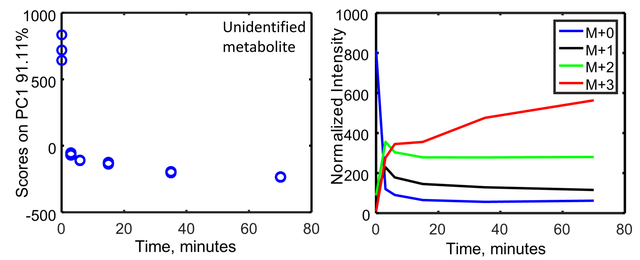
Scores on PC1 versus time for the PCA model of the full time course, for 5-oxoproline (DCS=51.2). The time course is quantitatively elucidated by PCA, indicating ^{13}C incorporation over the time course. This is validated by the traditional M+n plot, where M+0 corresponds to *m/z* 258.





Scores on PC1 versus time for the PCA model of the full time course for putrescine (DCS=44.1). The Scores plot shows a gradual incorporation of ^{13}C, with very little change over the first three time points. This is confirmed by the traditional M+n plot.





Scores on PC1 versus time for the PCA model of the full time course, including all injection replicates for an unidentified metabolite (DCS=8.0). The identity of the metabolite was not required to determine its fluxomic behavior. The Scores plot shows a large change in the mass spectra between the first and second time points, indicating rapid incorporation of ^{13}C by this particular metabolite.

Conclusions and References

Conclusions

We have described a novel analytical method for analysis of stable isotope fluxomics 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 rate at which the cellular metabolism incorporates ^{13}C for that molecule. Of the 152 metabolites surveyed, 83 were identified as changing with time and 69 unchanging with time.

References

B.C. Reaser, S. Yang, B.D. Fitz, B.A. Parsons, M.E. Lidstrom, R.E. Synovec, J. Chromatogr. A 1432 (2016) 111-121

B.D. Fitz, B.C. Reaser, D.K. Pinkerton, J.C. Hoggard, K.J. Skogerboe, R.E. Synovec, Anal. Chem. 86 (2014) 3973.

S. Yang, J.C. Hoggard, M.E. Lidstrom, R.E. Synovec, J. Chromatogr. A 1317 (2013) 175.

S. Yang, J.S. Nadeau, E.M. Humston-Fulmer, J.C. Hoggard, M.E. Lidstrom, R.E. Synovec, J. Chromatogr. A 1240 (2012) 156.