A Data Analysis Algorithm for Automated Relative Quantification of Stable Isotope-Labeled Metabolites

MICHAEL R. SHORTREED; Ju Shin Lee; Brian L. Frey; Lloyd M. Smith
University of Wisconsin, Madison, WI

Introduction

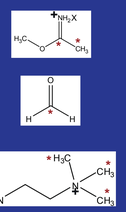
Quantification by liquid chromatography-mass spectrometry (LC-MS) requires special consideration due to the potential for run-to-run variability of retention times and matrix effects such as ionization suppression during the commonly used electrospray ionization. Several different strategies have been employed for quantitative LC-MS of metabolomic samples. Arguably, the most precise method involves the addition of an isotopically-labeled internal standard for every compound of interest (e.g. ^{2}H, ^{13}C, or ^{15}N-labeled). This approach is expensive, but warranted in certain targeted studies. Another strategy for relative quantification, as opposed to absolute quantification, is chemical labeling, which has proven to be useful for quantification in proteomics (e.g., isotope-coded affinity tags). Though relative quantification by labeling has seen limited use for metabolomics due to the lack of a single functional group present in all metabolites, there has been none-the-less a number of recent reports of effective labeling schemes [Guo 2007, Huang 2008, Lamos 2007, Shortreed 2006 and Yang 2007]. Labeled metabolites coelute from the chromatographic separation and appear in the mass spectrum as pairs of peaks with a characteristic mass difference. The peak intensity ratio for each pair yields the relative concentration. Such labeling strategies have a number of advantages including: improved quantitative precision, increased ability for molecular identification, and enhanced detection sensitivity. A major limitation of this strategy is lack of software tools for global identification of labeled compounds and calculation of the peak intensity ratios. Here we report a new software tool for automated identification of isotopically labeled metabolites and quantification of their relative concentration.

Advantages of Isotopic Labeling for Metabolomics

- Non-Targeted approach to metabolomics where the identification of unknown or unpredictable differences between two samples is sought.
- Comprehensive relative quantification can be achieved within a functional group class using judiciously chosen isotopic labels.
- Isotopic labeling facilitates both relative quantification and functional group assignment

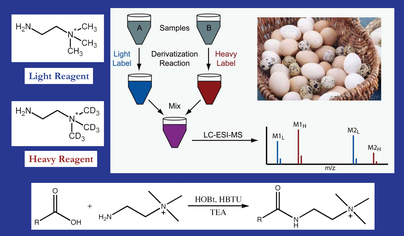
Some examples of labeling reagents.

- **Amines**
  - **Methylacetimidate**
  - **Formaldehyde (Formalin)**
- **Carboxylic acids**
  - **Cholamine**



Relative Quantification of Fatty Acids from Hydrolyzed Egg Lipid

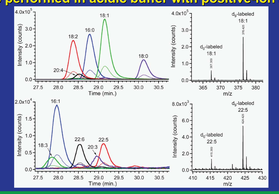
The experiment below is provided as an illustrative example of relative quantification using isotope labels. Cholamine, a carboxylic-acid-reactive quaternary amine, is used to label fatty acids from hydrolyzed egg lipid.



Demonstration of Co-Elution and Characteristic Mass Shift

LC-MS data from a 1:4 mixture of light:heavy cholamine-labeled fatty acids is shown below. Co-elution of both labeled forms of each acid is observed. Two representative mass spectra are shown to illustrate the characteristic mass shift introduced during labeling. These two phenomena (co-elution and characteristic mass-shift) are used, in the newly developed software described below, as key identifiers of labeled compounds.

Analysis performed in acidic buffer with positive-ion mode ESI

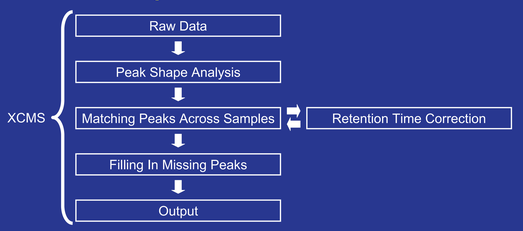


Heavy- and Light-Labeled Products Co-elute (because deuteriums are near the quaternary ammonium group)

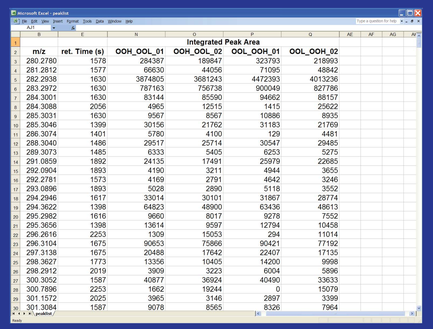
Method

The software described here is compatible with XCMS [Smith 2006], a widely used and freely available software program for processing LC-MS data for metabolite profiling. Both the new software and XCMS are written in the 'R' language and are compatible with standard mzXML data. Data processing begins by applying a number of standard XCMS functions (see flowchart below).

XCMS Processing Flowchart



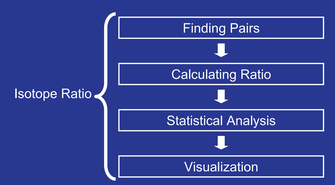
XCMS produces several output files including a grouped and retention time-aligned peak list (shown below in reduced form). This output file is used as input for the new software.



m/z	ret. Time (s)	OOH_OOL_01	OOH_OOL_02	OOL_OOH_01	OOL_OOH_02
		Integrated Peak Area			
280.2780	1578	284387	189847	323793	218993
281.2812	1577	66830	44056	71095	48842
282.2938	1630	3874805	3681243	4472393	4013236
283.2972	1630	787163	756738	900049	827786
284.3001	1630	83144	85590	94662	88157
284.3088	2056	4965	12515	1415	25622
285.3031	1630	9567	8567	10886	8935
285.3046	1399	30156	21762	31183	21769
286.3074	1401	5780	4100	129	4481
288.3040	1486	29617	25714	30547	29485
289.3073	1485	6333	5405	6253	5275
291.0859	1892	24135	17491	25979	22685
292.0904	1893	4190	3211	4944	3655
292.2781	1573	4169	2791	4642	3246
293.0896	1893	5028	2890	5118	3552
294.2946	1617	33014	30101	31867	28774
294.3622	1398	64823	48900	63436	48613
295.2982	1616	9660	8017	9278	7552
295.3656	1398	13614	9597	12794	10458
296.2616	2253	1309	15053	294	11014
296.3104	1675	90653	75866	90421	77192
297.3138	1675	20488	17642	22407	17135
298.3627	1773	13356	10405	14200	9998
298.2912	2019	3909	3223	6004	5896
300.3052	1587	40877	36924	40490	33633
300.7896	2253	1662	19244	0	15079
301.1572	2025	3965	3146	2897	3399
301.3084	1587	9078	8565	8326	7964

The new software scans the output from XCMS for pairs of peaks with nearly identical retention time (co-elution) and selected isotopic mass difference (characteristic mass shift). Data from peak pairs (retention time, observed m/z, neutral unlabeled mass, peak intensity, peak area, peak ratio) is written to a second file. Extracted ion chromatograms and m/z plots are automatically produced for each pair. A flowchart of these operations is provided below.

Isotope Ratio Processing Flowchart

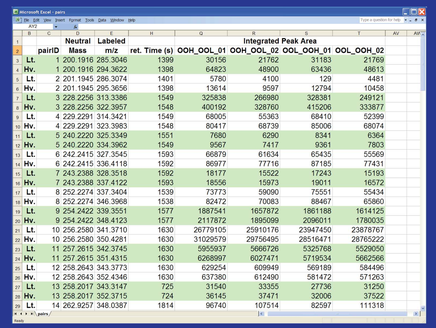


Results

The identification of isotope-labeled metabolites and calculation of relative abundances using the newly developed software is illustrated through analysis of samples of fatty acids that were obtained by saponifying lipids extracted from egg yolks. Fatty acids were reacted with either the heavy(D9)- or the light(D0)-isotopic form of cholamine, which contains a fixed-charged quaternary ammonium group. The heavy and light labeled samples were subsequently mixed in a 1:1 ratio, separated by reverse-phase HPLC and analyzed in positive-ion mode ESI-TOF-MS.

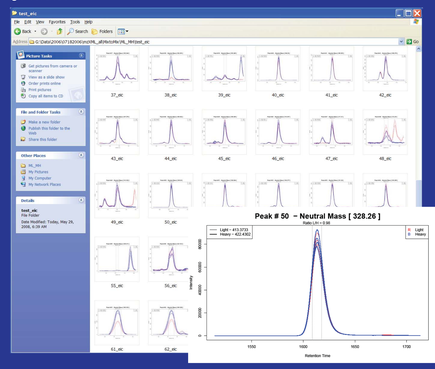
A mixture of all samples is used as the control, which guarantees that all compounds are represented, even if they appear in only one of the samples. Our strategy for identifying labeled metabolites is to analyze a 1:1 mixture of heavy and light labeled control samples. Labeled metabolite pairs coelute, have equal intensity and display a 9Da mass difference. Differences in retention time, intensity or mass shift can be used to eliminate false positives.

Seventy-nine pairs were identified by the new software algorithm for the 1:1 mixture of light- and heavy-labeled control. The first several pairs are displayed in the screen shot below.



pairID	Neutral Mass	Labeled m/z	ret. Time (s)	OOH_OOL_01	OOH_OOL_02	OOL_OOH_01	OOL_OOH_02
				Integrated Peak Area			
Lt.	1	200.1916	285.3046	1399	30156	21762	31183	21769
Hv.	1	200.1916	294.3622	1398	64823	48900	63436	48613
Lt.	2	201.1945	286.3074	1401	5780	4100	129	4481
Hv.	2	201.1945	295.3656	1398	13614	9597	12794	10458
Lt.	3	228.2256	313.3386	1549	325838	266980	328381	249121
Hv.	3	228.2256	322.3957	1548	400192	328760	415206	333877
Lt.	4	229.2291	314.3421	1549	68005	55363	68410	52399
Hv.	4	229.2291	323.3983	1548	80417	68739	85006	68074
Lt.	5	240.2220	325.3349	1551	7680	6290	8341	6364
Hv.	5	240.2220	334.3962	1549	9567	7417	9361	7803
Lt.	6	242.2415	327.3545	1593	66879	61634	65435	55569
Hv.	6	242.2415	336.4118	1592	86977	77716	87185	77431
Lt.	7	243.2388	328.3518	1592	18177	15522	17243	15193
Hv.	7	243.2388	337.4122	1593	18556	15973	19011	16572
Lt.	8	252.2274	337.3404	1539	73773	59090	75551	55434
Hv.	8	252.2274	346.3968	1538	82472	70083	88467	65860
Lt.	9	254.2422	339.3551	1577	1887541	1657872	1861188	1614125
Hv.	9	254.2422	348.4123	1577	2117872	1850099	2096011	1780035
Lt.	10	256.2580	341.3710	1630	26779105	25910176	23947450	23878767
Hv.	10	256.2580	350.4281	1630	31029579	29756495	28516471	28765222
Lt.	11	257.2615	342.3745	1630	5955937	5666726	5325768	5529050
Hv.	11	257.2615	351.4315	1630	6268997	6027471	5719534	5662566
Lt.	12	258.2643	343.3773	1630	629254	609949	569189	584496
Hv.	12	258.2643	352.4346	1630	637380	612490	581472	571263
Lt.	13	258.2017	343.3147	725	31540	33355	27736	31250
Hv.	13	258.2017	352.3715	724	36145	37471	32006	37522
Lt.	14	262.9257	348.0387	1814	96740	107514	82597	111318

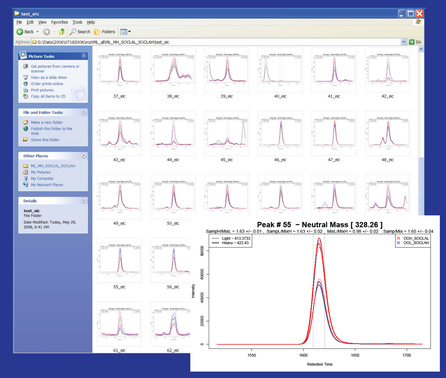
Extracted ion chromatograms for each peak pair are automatically generated.



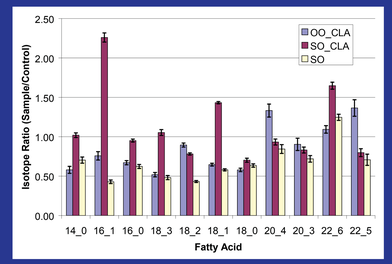
There were three samples in the analysis (SO, SO_CLA and OO_CLA) in addition to the control sample (OO). Each sample:control mixture was analyzed by LC-MS in duplicate and the 1:1 control mixture was analyzed in quadruplicate. In one example run of the complete algorithm, LC-MS data from $SO_L{-}OO_H$, $SO_H{-}OO_L$ and $OO_L{-}OO_H$ are processed as a group. This yielded 95 identified peak pairs. This was repeated for the SO_CLA and OO_CLA samples yielding 101 and 98 peak pairs, respectively. The total processing time for each group was ~175s (111s from XCMS and 64s from the new algorithm). The desktop machine used for data processing was an Intel Pentium 4 CPU (3.20GHz) with Hyper-Threading Technology. The total RAM memory was 1.49GB DDR.

All data from the three runs are manually curated in and Excel spreadsheet. Each peak pair produces features in the mass spectrum corresponding to the light- and heavy-labeled derivatives of the completely ^{12}C form of the acid as well as smaller peaks 1 and 2 Da higher from ^{13}C and $2x^{13}C$ according to the normal isotopic abundances. These peaks are manually eliminated from the data. Further elimination of false positives is achieved by careful analysis of all of the extracted ion chromatograms. This step is quite rapid because all extracted ion chromatograms for each peak pair is automatically generated by the software algorithm.

A screen-shot of some of the extracted ion chromatograms generated for the SO_CLA sample is displayed below with one example enlarged better show the detail of the product.



The data from eleven commonly observed fatty acids from hydrolyzed egg lipid is displayed in the histogram below. The values represented in the graph are the isotope ratios for each sample relative to the OO sample. The measurement precision ranged from 1-10% with a median value of 4%, which highlights the effectiveness of the isotopic labeling method for relative quantification.



Conclusion

The new software algorithm automatically identifies pairs of isotopically labeled metabolites, calculates the intensity ratio and generates extracted-ion chromatograms for each pair. The total processing time for all three sample groups (24 total files, 12 Mbyte each) analyzed for this work was under 10 minutes. In the future, we plan to enhance the functionality of this program by enabling it to collate data from all sample groups into a single spreadsheet and remove satellite peak pairs that emanate from natural isotopic abundances of ^{13}C. Furthermore, we plan to evaluate the performance of the software algorithm using data from other instrument systems (e.g. GC-MS). The algorithm was designed with compatibility to other instruments in mind. However, this functionality remains to be evaluated.

Stable Isotope Labeling

Guo, K., Ji, C. and Li, L. (2007) Stable-isotope dimethylation labeling combined with LC-ESI MS for quantification of amine-containing metabolites in biological samples. *Anal. Chem.*, 79, 8631-8638.

Huang, X. and Regnier, F.E. (2008) Differential metabolomics using stable isotope labeling and two-dimensional gas chromatography with time-of-flight mass spectrometry. *Anal. Chem.*, 80, 107-114.

Lamos, S.M., Shortreed, M.R., Frey, B.L., Belshaw, P.J. and Smith, L.M. (2007) Relative quantification of carboxylic acid metabolites by liquid chromatography-mass spectrometry using isotopic variants of cholamine. *Anal. Chem.*, 79, 5143-5149.

Shortreed, M.R., Lamos, S.M., Frey, B.L., Phillips, M.F., Patel, M., Belshaw, P.J. and Smith, L.M. (2006) Ionizable isotopic labeling reagent for relative quantification of amine metabolites by mass spectrometry. *Anal. Chem.*, 78, 6398-6403.

Yang, W.C., Adamec, J. and Regnier, F.E. (2007) Enhancement of the LC/MS analysis of fatty acids through derivatization and stable isotope coding. *Anal. Chem.*, 79, 5150-5157.

XCMS

Smith, C.A., Want, E.J., O'Maille, G., Abagyan, R. and Siuzdak, G. (2006) XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. *Anal. Chem.*, 78, 779-787.

Acknowledgements

Financial support for this work was provided by NIH/NHGRI 2 R01 HG002298, Wisconsin Institutes for Discovery, and the UW-Madison School of Medicine and Public Health Medical Education Research Committee (MERC).

We are grateful to Paul Benton, Colin Smith, Ralf Tautenhahn and Dong-Gee Hong for there assistance in this project.

Egg lipid samples were a generous gift from Mark Cook, Shane Huebner and Vanessa Leone.