Accurate Mass Retention Time Locked Metabolomics EI Library and Workflows for Accurate Mass GC/Q-TOF Metabolomics Data Processing

Zijuan Lai[1]; Mine Palazoglu[1]; Sofia Nieto[2]; Nathan Eno[2]; Hong Chen[2]; Vadim Kalmeyer[2]; Aditi Koul[2]; Oliver Fiehn[1]
[1]University of California at Davis, Davis, CA; [2]Agilent Technologies, Inc., Santa Clara, CA
Presenter: Marion Schaeffer Agilent Technologies, Germany

1st GC/Q-TOF Forum
Barcelona. 24 & 25 November, 2016

Agilent Technologies

Introduction

Accurate mass GC/MS methodologies have recently been gaining increased popularity in many applications, including metabolomics, by providing additional evidence of compound identity. While unit mass libraries and relevant screening approaches have been established for many years, accurate mass library creation and screening workflows represent significant challenges, since library curation for accurate mass spectra requires substantial effort. In addition, existing unit mass deconvolution and library search algorithms are not optimized for accurate mass data. To help improve reliability and increase throughput in metabolomics accurate mass applications, we created an accurate mass Retention Time Locked (RTL) EI library of metabolites using accurate mass spectra acquired with high resolution 7200 GC/Q-TOF system. In addition, we also demonstrate metabolomics screening workflow optimized for accurate mass. The screening workflow involves feature detection (FD), accurate mass library search and statistical analysis, where feature detection uses a new profile-based algorithm that is capable to work with extremely complex matrices containing wide range of metabolite concentrations.

Experimental

Sample Preparation

Human breast cancer cell line MCF-7 and MDA-MB-468 cells were grown in DMEM with 10% FBS, 5% PenStrep in 10% CO2, trypsinized, washed in PBS and flash frozen in liquid nitrogen. Prior to extraction, cells were quenched with methanol and dried using speed vacuum concentration system.

Figure 1. 7200 Series GC/Q-TOF System

Metabolites from approximately 1x106 cells were extracted in three replicates with acetonitrile:isopropanol:water (3:3:2). Extracts as well as metabolomics standard mixture solutions were dried under vacuum and derivatized by methoximation using a saturated solution of hydroxylamine HCl in pyridine followed by silylation with N-Methyl-N-(trimethylsilyl) trifluoroacetamide (MSTFA) and 1 % trimethylchlorosilane (TMCS).

Analytical Conditions

EI spectra have been acquired using an Agilent 7890B GC coupled to an Agilent 7200B accurate mass high resolution GC/Q-TOF (Figure 1). Standard metabolomics Fiehn method, retention time locked to d27-myristic acid have been used. GC and MS conditions are described in Table 1.

GC and MS Conditions:	
Column	DB-5 MS UI, 30 meter, 0.25 mm ID, 0.25 µm film
Injection volume	1 µL
Split ratio	10:1
Split/Splitless inlet temperature	250 °C
Oven temperature program	60 °C for 1 min 10 °C/min to 325 °C, 3.5 min hold
Carrier gas	Helium at 1 mL/min constant flow
Transfer line temperature	290 °C
Ionization mode	EI
Source temperature	230°C
Quadrupole temperature	150°C
Mass range	50 to 600 m/z
Spectral acquisition rate	5 Hz

Table 1. GC/Q-TOF conditions.

Data Processing Software Tools

MassHunter Qualitative Analysis (B.07.00 SP1) software tools have been used to automatically convert accurate masses of fragment ions into theoretical masses for all the abundant fragments in the spectrum and to import the spectra into accurate mass metabolomics Personal Compound Database and Library (PCDL). The metabolite screening with the accurate mass metabolomics library was performed in Unknowns Analysis (UA) standalone tool of MassHunter Quantitative Analysis (B.08.00, pre-release) using new profile-based algorithm. Statistical analysis as well as pathway analysis were performed in Mass Profiler Professional (MPP), version 13.1.1.

Results and Discussion

Construction of the Curated Accurate Mass Metabolomics Library

The metabolites for the accurate mass Metabolomics PCDL were selected to make sure that most of the GC/MS-amenable primary metabolites were included. The accurate mass Metabolomics library was created by adding approximately 500 accurate mass EI spectra after correction of each fragment ion m/z into the theoretical m/z. The distribution of the PCDL entries into compound classes is shown on Figure 2.

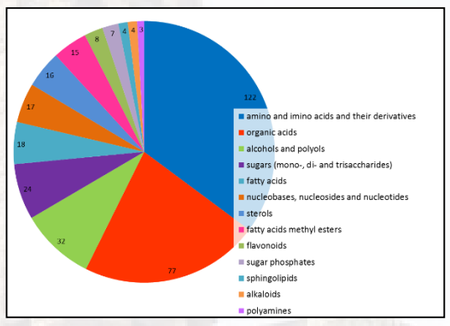

Figure 2. Classification of the compounds added to the accurate mass metabolomics library. Each derivatization state is a separate compound entry. The pie-chart shows only classified compounds.

In order to convert accurate m/z of the fragment ions into the theoretical ones, it is necessary to correctly annotate the mass spectra with the fragment formulas; this relies on the precise measurement of the m/z and the isotope ratios. For all the compound with the significant molecular ion, average observed mass accuracy was below 1 ppm, and average error of isotopic abundances was about 1% (Table 2).

The annotation of the EI fragments was performed using accurate mass information of TOF EI spectra and Molecular Formula Generator (MFG) with Fragment Formula Annotation (FFA) tools of MassHunter Qualitative Analysis software (Figure 3). After the fragment annotation is performed, the theoretical masses of the fragments are automatically calculated and the corrected spectra are directly sent to PCDL (Figure 4).

Name	Formula	m/z	m/z calc	mass error, ppm	MFG Score	M+1 isotope error, %
3-(3-Hydroxyphenyl)propionic acid	C15H26O3Si2	310.1415	310.1415	0.00	97.58	1.3
Creatinine	C13H31N3OSi3	329.1764	329.1769	1.52	94.9	3.2
Phloroglucinol	C15H30O3Si3	342.1495	342.1497	0.58	97.02	0
2'-Deoxyguanosine	C22H45N5O4Si4	254.1151	254.1153	0.79	98.75	0.1
Scopoletin	C13H16O4Si	264.0808	264.0812	1.51	98.6	0.8
2'-Deoxyadenosine	C19H37N5O3Si3	467.2193	467.2199	1.28	95.88	0.8
4-Nitrocatechol	C12H21NO4Si2	299.1005	299.1004	-0.33	98.14	0.1
3-Amino-L-tyrosine	C21H44N2O3Si4	484.2425	484.2424	-0.21	90.91	3.2
Xanthotoxin	C12H8O4	216.0418	216.0417	-0.46	99.59	0.9
trans-Dehydroandrosterone	C23H39NO2Si	389.2747	389.2745	-0.51	99.54	1.4
alpha-Santonin 2	C16H21NO3	275.151	275.1516	2.18	98.08	0.7
alpha-Santonin 1	C15H18O3	246.1249	246.125	0.41	99.51	0.6
Gluconic acid lactone	C18H42O6Si4	466.2049	466.2053	0.86	91.62	0.4
Daidzein	C21H26O4Si2	398.1368	398.1364	-1.00	97.87	1.4
Stearic acid	C21H44O2Si	356.3106	356.3105	-0.28	95.55	0.5
2-Amino-3-methoxybenzoic acid	C14H25NO3Si2	311.1362	311.1367	1.61	98.89	1.5
4-Androsten-3,17-dione	C21H32N2O2	344.2455	344.2458	0.87	99.53	0.1
L-methionine	C11H27NO2SSi2	293.1294	293.1296	0.68	97.87	0.4
Oleic acid	C21H42O2Si	354.295	354.2949	-0.28	95.34	1.1
Fructose	C22H55NO6Si5	569.2864	569.287	1.05	85.96	2.6
Average (abs)				0.82	96.56	1.06

Table 2. Molecular ion mass and relative isotope error for 20 randomly selected metabolites with prominent molecular ion

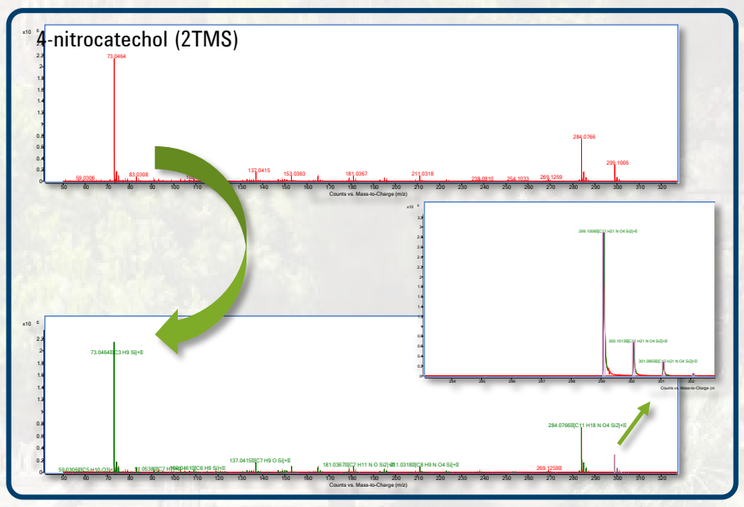

Figure 3. Fragment formula annotation of the spectrum using MassHunter Qualitative Analysis. Fragment formulas are assigned based on the empirical formula of the derivatized metabolite and accurate mass spectral data. Detailed annotation of the molecular ion isotopic cluster with theoretical isotope pattern overlaid is shown on the right.

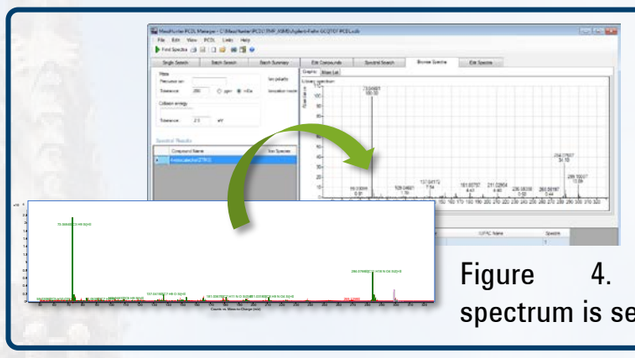

Figure 4. Annotated spectrum is sent into PCDL

To perform metabolomics screening we used the novel profile-based algorithm for feature detection (FD) optimized for accurate mass. First, to test the metabolomics PCDL we run the FD against the metabolite standards data and used the newly created metabolomics PCDL for the library search. For 56 randomly selected metabolites the library match score ranged from 89.9 to 99.1 with an average of 96.2.

Metabolomics screening workflow incorporated in Unknowns Analysis tool has been further tested using metabolite extracts from two different cell lines: MCF-7 and MDA-MB-468. The results are shown on Figure 5.

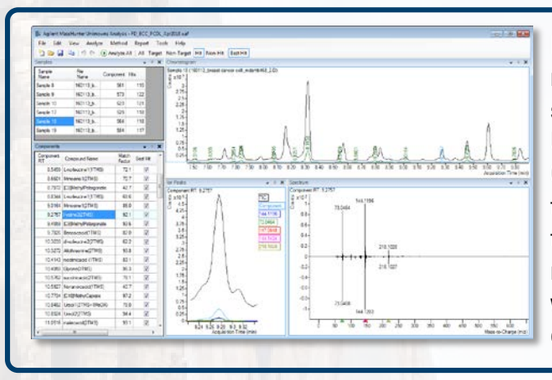

Figure 5. Accurate mass metabolomics screening in Unknowns Analysis. On average, more than 100 hits from the accurate mass metabolomics PCDL were detected in each sample.

The novel profile-based algorithm is also capable of extracting features with high quality spectra and accurate mass from highly saturated regions of the chromatogram, where conventional deconvolution algorithms typically fail (figure 6).

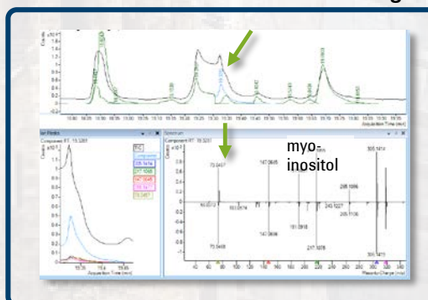

Figure 6. Shown is highly saturated region of the chromatogram where new FD algorithm is capable of extracting high quality spectrum.

To evaluate the differences between MCF-7 and MDA-MB-468, the data processed in UA were imported as .CEF files into MPP for statistical as well as pathway analyses (Figures 7 a and b). Pathway analysis revealed differences in a number of biochemical pathways for MCF-7 and MDA-MB-468. An example for one of the pathways (methionine salvage cycle) is shown on Figure 7b.

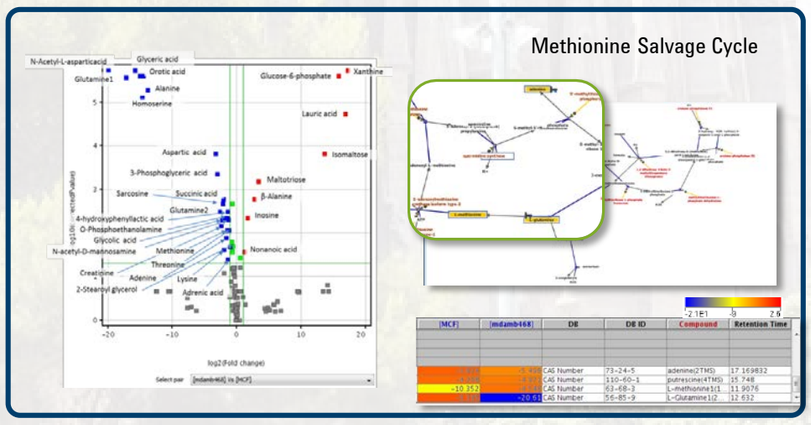

Figure 7. a) Volcano plot of log of fold change *vs* log of p-Value for MDA-MB-468 vs. MCF-7 cells demonstrating the major differences in metabolite concentrations between the two cell lines. b) Pathway analysis example showing differences in methionine salvage cycle between MCF-7 and MDA-MB-468.

Conclusions

- A curated accurate mass metabolomics PCDL containing ~500 compound entries has been created using a GC/Q-TOF and MassHunter software tools.
- The accurate mass metabolomics PCDL was further tested using the MCF-7 and MDA-MB-468 cell extracts and metabolomics screening workflow using a novel profile based algorithm, followed by statistical analysis in MPP.