ASMS 2013

ThP-665

An Integrated Workflow for the Identification and Pathway Visualization of Lipids

Stephen Madden, Steven M. Fischer, David Weil, Theodore Sana, Agilent Technologies, Inc., Santa Clara, CA

Introduction

Analysis of lipid samples by LC/MS presents special challenges due to requirements for chromatographic conditions, as well as annotation and identification of compounds found. Untargeted profiling approaches will likely provide a large number of candidate lipids, with scant evidence to confirm the identification or understand the biological context of the findings.

We introduce a workflow for lipid analysis which begins with untargeted profiling of a sample by MS and MS/MS to find as many possible lipid compounds as possible. We use specific lipid identification software to annotate the results, then overlay those compounds on pathways to determine those that might be of interest. Finally, we create a custom database/library from the pathways and use a targeted algorithm to re-interrogate the original data to identify other related compounds in the sample.

Experimental

Sample Preparation

Human serum samples in which the proteins had been removed using an acetonitrile crash were dried down for shipment. Upon arrival at the analysis lab, they were stored in a -80 °C freezer. 400 µL of 50/50/0.1% water/acetonitrile/formic acid was added to the dried down serum resulting in a clear solution. 100 µL of solution was transferred directly to an auto sample vial. Concentrated serum samples were placed in a -20 °C freezer and underwent a maximum of three freeze/thaw cycles.

Instrumentation

Agilent 1260 Infinity Bio-inert quaternary pump, degasser, high performance autosampler with thermostat, and Agilent 6550 iFunnel Q-TOF mass spectrometer.

LC Conditions

Column: Agilent ZORBAX Eclipse Plus C18, 2.1x 150 mm, 1.8 µm

Column temperature: 40 °C

Mobile phase: A = 0.1% formic acid in water, B = 0.1% formic acid in acetonitrile
Flow rate: 0.35 mL/min
Injection volume: 2 µL
Gradient (%B): 0.5%B (0- 5.0 min), 99%B @25 min, 99%B @35 min, 0.5%B @36 min
Stop time: 36 min
Post time: 8 min

Experimental

MS Conditions

Ion source: Dual ESI

Instrument mode: Extended Dynamic Range: Narrow Mass 1700

Scan mode: (-) full MS (2 spectra/s) followed by up to 4 data dependent auto MS/MS (4 spectra/s)

Mass range (MS and MS/MS): *m/z* 50-1000

Drying gas temperature and flow: 200°C @ 16 L/min

Nebulizer pressure: 45 psig

Capillary voltage: 4000 V

Ion Funnel voltages: Optimized for low fragmentation

Collision energy: 10, 20 V

Reference ions: 68.99857 and 966.0007

Data Processing

An acquired data file was loaded into MassHunter Qualitative Analysis rev. B.06.00 and analyzed by an untargeted feature finding algorithm called Molecular Feature Extractor (MFE). MFE finds features by removing ions that are persistent across the retention time, then grouping ions into proposed features using evidence of isotopic clusters, adducts, neutral losses, and multiple charge states.

1035 putative features were found and information about them (including MS and MS/MS spectra) was exported using a Compound Exchange Format (CEF) file. The CEF file was imported into SimLipid 3.50 (PREMIER Biosoft, Palo Alto, California, USA) to run a High Throughput Lipid Search. The HTP Lipid Search annotated 176 compounds using MS1 database search and 11 compounds using MS/MS pattern matching with compound names and LIPID MAPS identifiers. An annotated CEF file was exported by SimLipid and imported into Mass Profiler Professional (MPP) 12.5..

The list of annotated compounds was run against the BioCyc Pathway Database Collection (SRI International, Menlo Park, California, USA) for *Homo sapiens* (version 17) using Pathway Architect 12.5. Compounds were detected in multiple pathways, and the sphingomyelin metabolism pathway was selected for further investigation. Using MassHunter Pathways to PCDL rev. B.05.00, a custom database and library was constructed using compounds from sphingolipid pathways in BioCyc. Targeted mining of the original data file revealed the presence of five previously undetected lipid and related compounds.

Results and Discussion

Untargeted Feature Finding

Figure 1 shows the results of a putative compound found during the feature finding using the MFE algorithm in MassHunter Qualitative Analysis. A chromatogram made from the compound's ions (Extracted Compound Chromatogram or ECC) as well as a spectrum are displayed.

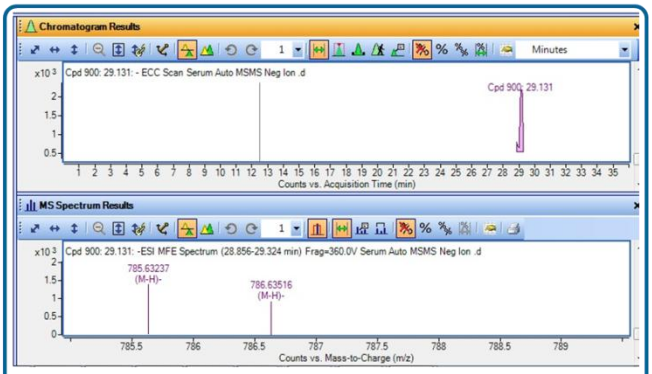


Figure 1. MFE Results for Lipid Compound

Annotation of Lipids

Lipids are annotated with their name, formula, and LIPID MAPS identifier, as well as other available identifiers like KEGG, ChEBI and PubChem. Figure 2 shows results for a lipid in SimLipid 3.50.

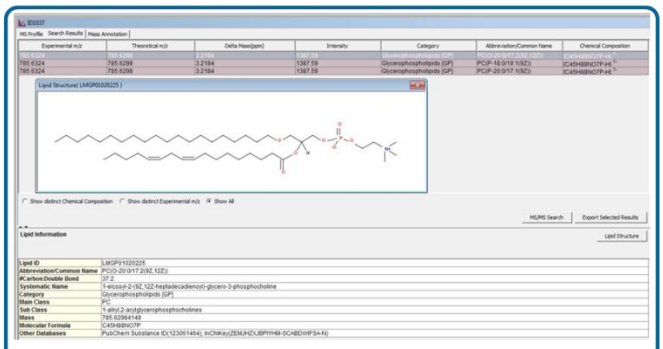


Figure 2. SimLipid Results Including Database Identifiers

Mapping from Compounds to Pathways

Figure 3 shows the mapping of two compounds in the sample to the sphingomyelin metabolism pathway from BioCyc *Homo sapiens* using Pathway Architect. The identification of a phosphatidylcholine is not definitive and the user is able to choose manually among three candidates. This highlights an issue in pathway mapping—many of the pathway sources for lipids use "R groups" to represent multiple compounds of the same type. As a result, it was necessary to manually annotate some of the lipid identifications with KEGG identifiers for these R groups using MPP's ID Browser program.

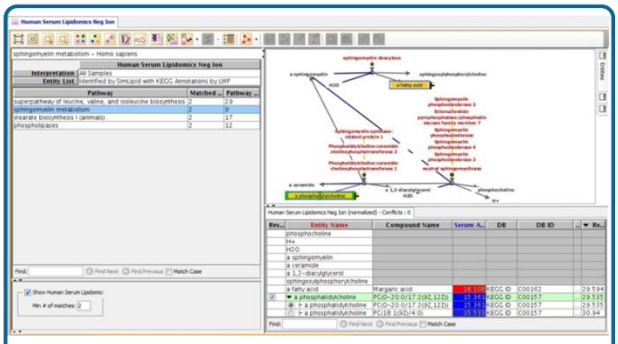


Figure 3. Pathway Analysis Results

Creation of a Custom Database/Library

Figure 4 displays a text search in the MassHunter Pathways to PCDL program against BioCyc *Homo sapiens* for pathways that contain "sphingo." Two pathways appear and are selected to create an accurate mass database/library using content from the METLIN PCDL rev. B.05.00. Figure 5 shows the list of compounds in the database/library in MassHunter PCDL Manager, including that several of them have MS/MS spectra available.

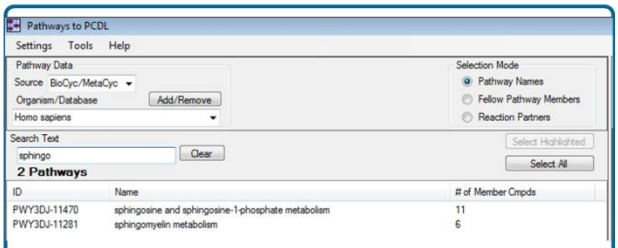
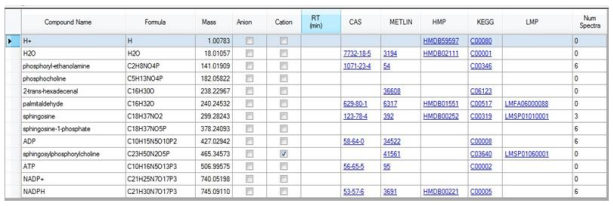


Figure 4. Pathways Selected for Custom Database/Library

Results and Discussion

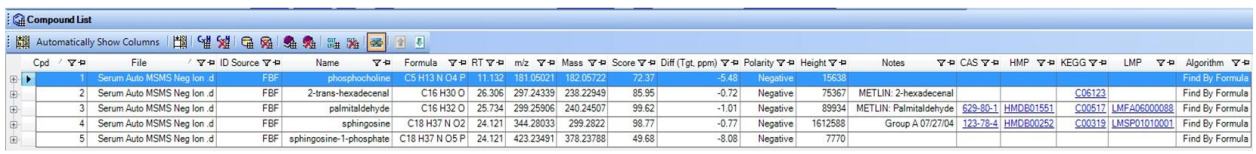


Compound Name	Formula	Mass	Anion	Cation	RT (min)	CAS	METLIN	HMP	KEGG	LMP	Num Spectra
H+	H	1.00783						HMDB59597	C00080		0
H2O	H2O	18.01057				7732-18-5	3194	HMDB02111	C00001		0
phosphoryl-ethanolamine	C2H8NO4P	141.01909				1071-23-4	54		C00346		6
phosphocholine	C5H13NO4P	182.05822									0
2-trans-hexadecenal	C16H30O	238.22967					36608		C06123		0
palmitaldehyde	C16H32O	240.24532				629-80-1	6317	HMDB01551	C00517	LMFA06000088	0
sphingosine	C18H37NO2	299.28243				123-78-4	392	HMDB00252	C00319	LMSP01010001	3
sphingosine-1-phosphate	C18H37NO5P	378.24093									6
ADP	C10H15N5O10P2	427.02942				58-64-0	34522		C00008		6
sphingosylphosphorylcholine	C23H50N2O5P	465.34573		☑			41561		C03640	LMSP01060001	0
ATP	C10H16N5O13P3	506.99575				56-65-5	95		C00002		0
NADP+	C21H25N7O17P3	740.05198									0
NADPH	C21H30N7O17P3	745.09110				53-57-6	3691	HMDB00221	C00005		6

Figure 5. Compounds in the Custom Database/Library

Targeted Feature Finding of the Original Data

Figure 6 is a Compound List from MassHunter Qualitative Analysis after targeted feature finding using the custom database/library. Five new compounds were found by this targeted analysis. For this re-mining of the original data, the Find by Formula (FbF) algorithm was used. Find by Formula starts with the formula of a compound (in this case, from the custom database/library) and determines if it is present in the data file. A probability match score is calculated for each compound based on its mass, isotopic abundance and spacing. If a compound passes the match score filter, it is labeled as identified.



Compound List

Automatically Show Columns

Cpd	File	ID Source	Name	Formula	RT	m/z	Mass	Score	Diff (Tgt, ppm)	Polarity	Height	Notes	CAS	HMP	KEGG	LMP	Algorithm
1	Serum Auto MSMS Neg Ion .d	FBF	phosphocholine	C5 H13 N O4 P	11.132	181.05021	182.05722	72.37	-5.48	Negative	15638						Find By Formula
2	Serum Auto MSMS Neg Ion .d	FBF	2-trans-hexadecenal	C16 H30 O	26.306	297.24339	238.22949	85.95	-0.72	Negative	75367	METLIN: 2-hexadecenal			C06123		Find By Formula
3	Serum Auto MSMS Neg Ion .d	FBF	palmitaldehyde	C16 H32 O	25.734	299.25906	240.24507	99.62	-1.01	Negative	89934	METLIN: Palmitaldehyde	629-80-1	HMDB01551	C00517	LMFA06000088	Find By Formula
4	Serum Auto MSMS Neg Ion .d	FBF	sphingosine	C18 H37 N O2	24.121	344.28033	299.2822	98.77	-0.77	Negative	1612588	Group A 07/27/04	123-78-4	HMDB00252	C00319	LMSP01010001	Find By Formula
5	Serum Auto MSMS Neg Ion .d	FBF	sphingosine-1-phosphate	C18 H37 N O5 P	24.121	423.23491	378.23788	49.68	-8.08	Negative	7770						Find By Formula

Figure 6. Results of Targeted Re-mining of the Original Data

Confirmation Using MS/MS Spectra

Several of the new compounds detected in the original data file were acquired with MS/MS spectra due to the data dependent acquisition. The FbF algorithm can extract the MS/MS spectra for each compound found. Figure 7 displays an averaged product ion spectrum for palmitaldehyde with the precursor *m/z* denoted with a blue diamond and the chemical structure from METLIN. This spectrum provides additional evidence of the identification of the compound.

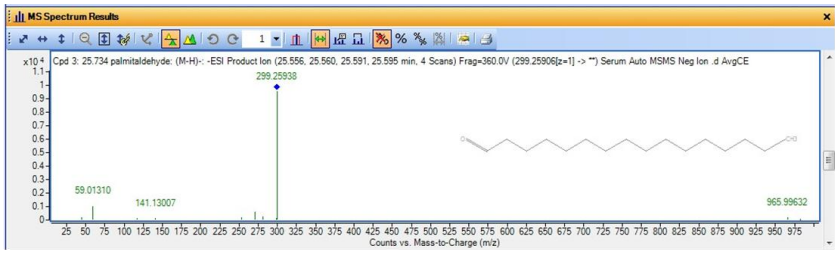


Figure 7. Average Product Ion Spectrum of Palmitaldehyde

Conclusions

We have demonstrated the integrated use of high resolution accurate mass Q-TOF data together with data processing software to identify lipids and understand their biological context.

- MassHunter Qualitative Analysis provides both untargeted (MFE) and targeted (FbF) algorithms for feature finding
- SimLipid performs annotation of compounds found by both MS and MS/MS
- MPP and Pathway Architect perform statistical analysis and mapping of the compounds on BioCyc pathways
- MassHunter Pathway to PCDL creates custom database/libraries for further mining of the original data