Automatic lipid characterization using SimLipid® from normal phase and reverse phase liquid chromatography - MS, MS/MS data acquired in variable ion modes

Ningombam Sanjib Meitei[1], Arijit Biswas[1], Arun Apte[2], Stephen Madden[3], Mark Sartain[3]
[1]PREMIER Biosoft, Indore, India, [2]PREMIER Biosoft, Palo Alto, U.S.A, [3]Agilent Technologies, Santa Clara, CA, U.S.A.
Corresponding author e-mail: sanjib@premierbiosoft.com

Introduction

LC-MS techniques using electrospray ionization (ESI) and reversed phase chromatography have successfully been used to profile complex lipid samples. Normal phase (NP) LC-MS separates phospholipids into their respective classes based on their respective polar head groups and with little influence from their sn-1 and sn-2 fatty acid substituents. In contrast to NP separations for lipid analysis reverse phase (RP) separations have the signature characteristic of discriminating lipids according to the overall polarity and the fatty acid composition in the sn-1, sn-2, and sn-3 locations. One major challenge faced by researchers in employing such workflows is the lack of software tools to automate the data analysis. We have developed SimLipid to address this challenge.

Methods

Total lipid liver extracts were purchased from Avanti Polar Lipids (Alabaster, AL). Dried lipid extracts were reconstituted in 2:1 chloroform/methanol and further diluted in mobile phase A to a concentration of 200 ng/µL. Injections of 5 µL (1 µg extract) were analyzed by RP or NP phase LC/MS. RP separations were performed on an Agilent Zorbax EclipsePlus RRHD, 2.1 x 100mm, 1.8 µm column with a binary gradient, and NP separations were performed on an Agilent Zorbax Rx-Sil, 2.1 x 100mm, 1.8 µm column with a binary gradient on an Agilent 1290 UHPLC system. The LC eluent flowing at 0.35 ml/min was directly analyzed by mass spectrometry with the following parameters:

6540 Q-TOF with Dual Agilent Jet Stream Source			
Instrument Mode	2 GHz, extended dynamic range, m/z 1700	**Capillary Voltage**	3500 V (+), 3000 (-)
Polarity	positive or negative	**Nozzle Voltage**	0 V
Gas Temperature	300 °C	**Fragmentor**	150 V
Drying Gas (Nitrogen)	11 L/min	**Oct 1 Rf Vpp**	750 V
Nebulizer Gas	35 psi	**Acquisition Speed**	MS-only: 1 spectra/second (MS); Auto MS/MS: 3 spectra/second (MS), 3 spectra/s (MS/MS)
Sheath Gas	300 °C	**Auto MS/MS Parameters**	Isolation Width: Narrow (~1.3 amu) Collision Energy: 20 eV
Sheath Gas Flow	12 L/min	**Reference Correction**	2 points at m/z 121.050873 (+), m/z 922.009798 (+), 2 points at m/z 112.985587 (-), m/z 980.016375 (-)

The first step in data analysis for profiling workflows is to extract molecular features from the results where features are defined by retention time and mass. A feature condenses the abundances from all the specified adducts and isotopes of a compound into a single compound. Molecular features were extracted with MassHunter Qualitative Analysis (Version B.07.00). SimLipid database has been created containing 36,299 lipids and 1,305,386 structure-specific *in-silico* characteristic ions (1-4). Table 1 presents a list of prototype in silico fragments generated by SimLipid for the sodiated lipid species TG(18:1(9Z)/18:1(9Z)/18:1(9Z)) (Figure 1) and corresponding nomenclature used by the program to annotate MS/MS spectra.

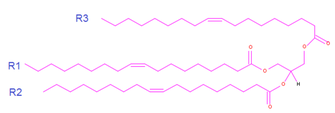


Figure 1: Structure of the TG(18:1(9Z)/18:1(9Z)/18:1(9Z))

SimLipid creates a list of candidate structures for each lipid MS/MS spectrum based on precursor m/z value and other information. For each candidate feature, *in-silico* fragment ions are matched against the experimental MS/MS data. A scoring mechanism was developed in order to differentiate isobaric candidates. Results obtained from two different MS/MS database searches can further be combined together to report those lipids identified in both the runs thereby enabling validation of lipids using diagnostic ions observed in negative ion mode MS/MS spectra as well as those observed only in positive ion mode MS/MS spectra.

Ion Type	Fragment Structure Displayed by SimLipid	Description	SimLipid Nomenclature
B - ion		Loss of one neutral fatty acid	M - <C: DB > - H where <C:DB> is the number of carbons and double bonds in the fatty acid E.g., M – 18:1-H
C – - ion		Loss of one sodium carboxylate residue	M - <C:DB > - Na-H E.g., M – 18:1-Na - H
G - ion		Loss of 2 fatty acids; one from position 2 and the other from either position 1 or position 3. It contains all 3 carbons from glycerol backbone. SimLipid generates 3 G fragments by considering loss of any of the fatty acids.	Gk where k represents the fatty acid retained in the fragment structure. E.g., G3
J - ion		Loss of 2 fatty acids on position 1 and 3. It contains 2 carbons out of the 3 in glycerol backbone. In order to generalise the fragmentation pattern, SimLipid generates 3 J fragments by considering loss of any of the 2 fatty acids unconstrained to position 1 and 3.	Jk where k represents the fatty acid retained in the fragment structure. E.g., J3
A - ion		Loss of a part of 1 fatty acid	Fj(Ri); i=1,2 & 3 and j can be any number between 2 to n where n = # carbons in the fatty acid chain – 2 - # double bonds in the fatty chain. E.g., F2(R1)
Acylium ion		Protonated acylium ion (3,4)	RC=O+ Where R represents the alkyl group that is attached to the CO group with a single bond. Simply, it is the number of carbon atoms -1 present in the fatty acid. E.g., 17:1C=O+

Table 1: Prototype in silico fragments generated by SimLipid for the sodiated species of lipid TG(18:1(9Z)/18:1(9Z)/18:1(9Z)).

Results and Discussions

The peaklists containing detected compounds and MS/MS data were generated by qualitative analysis software. Table 2 shows the summary of lipid annotation results of single LC/MS data files from liver lipid extracts by SimLipid software. For each annotated compound, there may be more than one possible matching lipid annotation. SimLipid correctly distinguishes isobaric lipids either based on score or after combining results obtained from varying ion modes.

Data File Type	MFE Results	SimLipid 4.40	SimLipid 4.40 (MS2 data from variable ion modes)
Reverse Phase (+) MS only	4108	1172	
Reverse Phase (-) MS only	1399	430	
Reverse Phase (-) Auto MS/MS	278 (247)	65	35
Reverse Phase (+) Auto MS/MS	972 (612)	226	
Normal Phase (+) MS only	2281	644	
Normal Phase (-) MS only	561	250	
Normal Phase (+) Auto MS/MS	774 (308)	141	25
Normal Phase (-) Auto MS/MS	195 (126)	72	

Table 2. Summary of lipid annotation results comparing single LC/MS data files from liver lipid extracts by SimLipid software. The parentheses denote only compounds with MS/MS scan information.

Overview of the lipids identified by different MS/MS workflows

Table 3 shows the different types of lipids detected by different MS/MS workflows. Some of the identified lipids belonging to bile acids & derivatives and steroids may be false positives since these lipids are identified only in a single type of workflow with low scores and there is no evidence of those lipids in other workflows without any scientific reason. However, there are also cases when lipids belonging to specific classes namely TAG, DAG, PC, SM and PG could not be detected by different workflows. For example, LC-MS/MS experimental runs in the positive ion mode yield highest number of identified PC lipids. However, LC-MS/MS runs in negative ion mode could not detect a single PC lipid. The non-detection of lipids in some types of workflows are investigated and probable reasons along with software remedies are explained in the following sections.

Lipid Class	Reverse Phase (-) Auto MS/MS	Reverse Phase (+) Auto MS/MS	Normal Phase (+) Auto MS/MS	Normal Phase (-) Auto MS/MS
Bile acids & derivatives	X	X	Yes	X
Cardiolipin	X	Yes	Yes	X
Ceramides	Yes	Yes	X	X
DAG	X	Yes	X	X
Isoprenoids	X	Yes	Yes	X
N GSP	Yes	Yes	Yes	Yes
OGP	X	Yes	Yes	X
PC	X	Yes	Yes	X
PE	Yes	Yes	Yes	Yes
PG	Yes	X	Yes	Yes
PI	Yes	Yes	Yes	Yes
PS	Yes	Yes	Yes	X
SM	X	Yes	Yes	X
Steroids	X	X	Yes	X
Sterols	X	Yes	X	X
TAG	X	Yes	X	X

Table 3: Lipids detected by different MS/MS workflows. *Abbreviations:N GSP: Neutral glycosphingolipids; OGP: Oxidized glycerophospholipids*

Figure 2 and 3 depicts the number of lipids identified for different classes after subjecting the sample to different workflows.

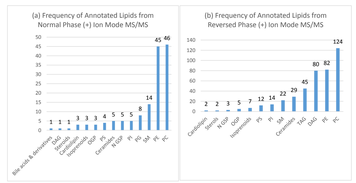


Figure 2: Frequency of lipids categorized on the basis of different classes identified from (a) Normal Phase (+) Ion Mode MS/MS and (b) Reversed Phase (+) Ion Mode MS/MS runs.

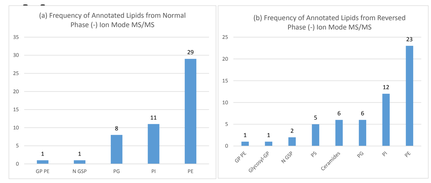


Figure 3: Frequency of lipids categorized on the basis of different classes identified from (a) Normal Phase (-) Ion Mode MS/MS and (b) Reversed Phase (-) Ion Mode MS/MS runs.

Effective workflow for identifying Glycerolipids

From Figure 2 and 3, we have observed that there are 45 TAGs annotated in the RP (+) MS/MS run. However, not a single TAG is detected in the NP (+) MS/MS run. Similarly, only 1 DAG is annotated in NP (+) ion mode MS/MS run as compared to 80 in RP (+) ion mode MS/MS run. This observation can be explained by the lack of retention of neutral lipids such as TAG, DAG, and cholesterol species with the NP silica-based stationary phase in combination with these mobile phases. These unretained neutral lipids would likely be found in the column void fraction (within the first minute of chromatography), and were thus excluded from the retention time window used for feature finding in the extraction algorithm. This observation also highlights one of the disadvantages of NP chromatography versus RP chromatography in that not all lipid classes can be resolved in a single run. i.e. RP separations are more comprehensive when it comes to identifying glycerolipids.

Lipid Identification using product ions from MS/MS spectra

The efficiency of the SimLipid's proprietary scoring algorithm explained in the methods section is depicted using the TAG lipid identified at retention time 34.255 minutes in the reversed phase LC-MS/MS workflow (Figure 4). The TG(16:0/16:0/18:3) is identified as ranked 1 lipid with a score of 0.2453 while isobaric lipid TG(16:1/16:1/16:1) is identified to be the 2nd ranked lipid with a score of 0.2167 out of 1 (figure 5). The MS/MS data does not provide enough information to identify the location of the double bonds in the fatty acyls. However, there is enough data in the spectra to differentiate the fatty acyls with different numbers of carbons and double bonds. Figures 6(a) and 6(b) show the MS/MS spectra featuring 9 major peaks denoted by P1-P9 annotated with fragment ions generated from the first ranked lipid TG(16:0/16:0/18:3) and the second ranked lipid TG(16:1/16:1/16:1). Here, for the lipid TG(16:0/16:0/18:3), P1 corresponding the Acylium ion and P6 corresponding to C – ion explain the presence of the fatty acyl 16:0 and similarly P2 and P5 explain the presence of the chain 18:3. However, for the second ranked lipid TG(16:1/16:1/16:1), the presence of fatty acyl 16:1 is explained by P1 and P7, all the major peaks in the spectrum namely P2, P5 and P6 are left unexplained. SimLipid assigns a penalty for the missed peaks with the magnitude of the penalty decided by the relative intensity of missed out peaks. Hence, SimLipid predicts the lipid TG(16:0/16:0/18:3) as the most probable lipid structure for the MS/MS spectrum at retention time 34.255 minutes.

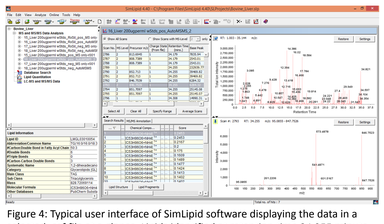


Figure 4: Typical user interface of SimLipid software displaying the data in a selected .cef file at a glance. Lipids identified at retention time 34.255 minute in the reversed phase LC-MS/MS workflow is displayed.

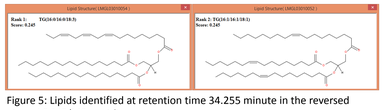


Figure 5: Lipids identified at retention time 34.255 minute in the reversed phase LC-MS/MS workflow.

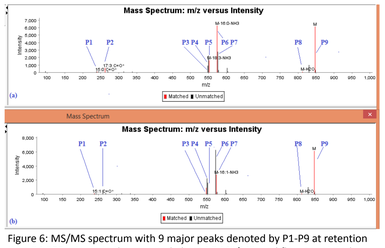


Figure 6: MS/MS spectrum with 9 major peaks denoted by P1-P9 at retention time 34.255 minute in the reversed phase LC-MS/MS workflow annotated with (a) TG(16:0/16:0/18:3) which is ranked 1 lipid and (b) TG(16:1/16:1/16:1) which is ranked 2.

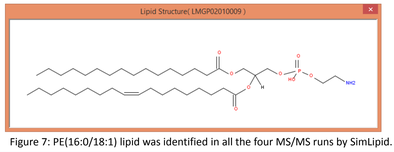


Figure 7: PE(16:0/18:1) lipid was identified in all the four MS/MS runs by SimLipid.

Hence, in order to achieve structural information from tandem mass spectrometry data, both positive and negative ionization modes need to be utilized. The PE(16:0/18:1) lipid was identified in all the four experimental MS/MS runs (Figure 7). Table 4 shows the MS/MS spectra from different experimental workflows wherein peaks are annotated with corresponding fragment ions by SimLipid software. The MS/MS spectra in negative ion mode feature peaks corresponding to fatty acids while those in positive ion mode feature head groups.

Utilizing MS/MS data from variable ion modes for identification of Glycerophospholipids and Sphingolipids

Structural characterization of lipids from GP and SP categories using tandem mass spectrometry data requires identification of head group, sn-1 chain and sn-2 chain. Traditionally, the degree of structural information obtained as a result of this analysis varies by the type of instrumentation used. In negative ionization mode tandem mass spectra tend to yield sn-1 and sn-2 fatty acid residue fragments, which reveal the fatty acid composition of the lipid, whereas in positive ion mode MS/MS spectra are more likely to feature fragment ions corresponding to only head groups, thereby only head group information was routinely obtainable from positive mode fragmentation.

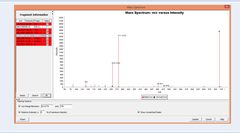
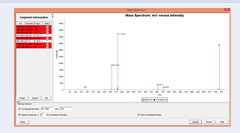
Experiment	Precursor m/z	Adduct	MS/MS spectra annotated by SimLipid
Normal Phase (+) MS/MS	718.5378	[M+H]+	
Normal Phase (-) MS/MS	716.5243	[M-H]-	
Reverse Phase (+) MS/MS	718.5378	[M+H]+	
Reverse Phase (-) MS/MS	716.5243	[M-H]-	

Table 4: MS/MS spectra from different experimental workflows wherein peaks are annotated with corresponding fragment ions from PE (16:0/18:1) lipid by SimLipid software. M-NL+H is the fragment ion generated due to the loss of head group i.e., loss of phosphoethanolamine from the lipid.

SimLipid software streamlines the process of combining lipids identified by different experimental workflows and exports the fragment ions that are observed in different spectra thereby facilitating easy review of the identified lipids. Figure 8 shows the prototype results of lipids identified from normal phase LC-MS/MS data of both the ion modes exported into MS excel file.

Identifying Glycerophosphocholine lipids

Despite having PC lipids identified with the most frequency in positive ion mode for both NP and RP workflows (Figure 2), no PC lipid could be detected in negative ion mode. On further investigation of the data, we observed that the problem originates from the parameters we chose in the feature finding algorithm. Due to the modifier ammonium acetate used in the LC mobile phases, PCs in negative ion mode form mainly acetate adducts [M+OAc]-, and there are no [M-H]- ions or other adducts (ex. [M+Cl]-) typically observed. Therefore, the feature finding incorrectly assumed that the [M+OAc]- ion is an [M-H]- ion and assigns the wrong neutral mass to the feature, causing database searches to fail for these ions. This problem is addressed by SimLipid by allowing the user to select the desired ion species to override the default ion species provided by MFE (Figure 9). On performing the search using this software feature, 22 PC lipids each from the NP – MS/MS run and RP – MS/MS run are identified (generated report in Figure 10). 14 PC lipids could be verified by combining MS/MS data from both the ion modes for NP-LC MS/MS workflow while 55 PC lipids could be verified by combining MS/MS data from both the ion modes for RP-LC MS/MS workflow. Similar workflows can be employed to identify SM and PG lipids.

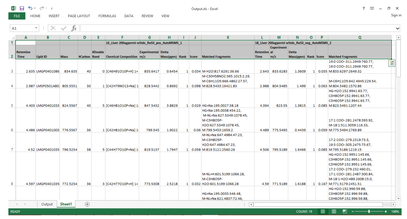


Figure 8: Lipid identified using MS/MS data from both the ion modes exported into MS excel file.

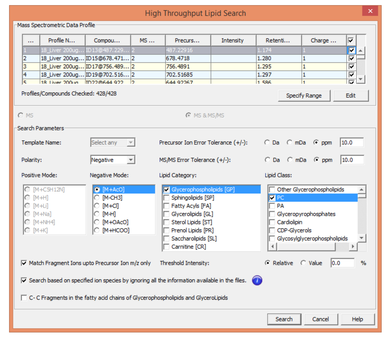


Figure 9: Typical SimLipid software interface to select ion species that we want, to override the ion species provided by molecular feature finding algorithm.

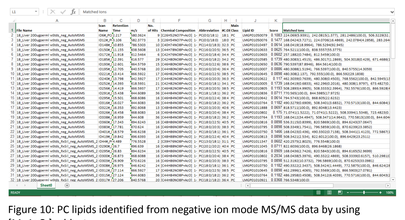


Figure 10: PC lipids identified from negative ion mode MS/MS data by using [M+AcO-] adduct.

Conclusion

Mass spectrometry based experimental workflows, despite being one of the most popular platforms for qualitative and quantitative lipid data analysis, are challenging because different experimental workflows are suitable for different classes of lipids. At times, data from multiple experimental workflows need to be mined to characterize lipids definitively. This necessitates dynamic software tools that not only support different types of data analysis protocols based on experimental workflows but also has the capability to simultaneously process data from multiple workflows. SimLipid software has been redesigned to address these challenges and extensive data analysis was carried out using its automated features. The ability to identify lipids structures accurately using data from variable ion modes and sophisticated experimental workflows provided by Agilent's Q-TOF instrument series enables significant improvement in the quality of lipidome data analysis. The entire process is expedited including both the experimental and bioinformatics analysis. This increases the accuracy and confidence in the identification of lipids.

Reference

[1]. Murphy, Robert C. *Mass spectrometry of phospholipids: tables of molecular and product ions*. Illuminati Press, 2002.
[2] Cheng, C., Gross, M. L., Pittenauer, E. Complete structural elucidation of triacylglycerols by tandem sector mass spectrometry. Anal. Chem. 1998, 70, 4417-4426.
[3] McAnoy A.M., Wu C.C., and Murphy R.C. Direct Qualitative Analysis of Triacylglycerols by Electrospray Mass Spectrometry Using a Linear Ion Trap. J Am Soc Mass Spectrom 2005, 16, 1498–1509.
[4] Murphy R.C., James P.F., McAnoy A.M., Krank J., Duchoslav E., Barkley R.M. Detection of the abundance of diacylglycerol and triacylglycerol molecular species in cells using neutral loss mass spectrometry. Analytical Biochemistry 2007, 366, 59–70.