Research Article

Journal of MASS SPECTROMETRY

Received: 21 March 2012 Revised: 14 May 2012 Accepted: 14 May 2012 Published online in Wiley Online Library

(wileyonlinelibrary.com) DOI 10.1002/jms.3033

Computational approach to structural identification of phospholipids using raw mass spectra from nanoflow liquid chromatography–electrospray ionization–tandem mass spectrometry

Sangsoo Lim, Seul Kee Byeon, Ju Yong Lee and Myeong Hee Moon*

A qualitative analysis tool (LiPilot) for identifying phospholipids (PLs), including lysophospholipids (LPLs), from biological mixtures is introduced. The developed algorithm utilizes raw data obtained from nanoflow liquid chromatography–electrospray ionization–tandem mass spectrometry experiments of lipid mixture samples including retention time and *m/z* values of precursor and fragment ions from data-dependent, collision-induced dissociation. Library files based on typical fragmentation patterns of PLs generated with an LTQ-Velos ion trap mass spectrometer are used to identify PL or LPL species by comparing experimental fragment ions with typical fragment ions in the library file. Identification is aided by calculating a confidence score developed in our laboratory to maximize identification efficiency. Analysis includes the influence of total ion intensities of matched and unmatched fragment ions, the difference in *m/z* values between observed and theoretical fragment ions, and a weighting factor used to differentiate regioisomers through data filtration. The present study focused on targeted identification of particular PL classes. The identification software was evaluated using a mixture of 24 PL and LPL standards. The software was further tested with a human urinary PL mixture sample, with 93 PLs and 22 LPLs identified. Copyright © 2012 John Wiley & Sons, Ltd.

Supporting information may be found in the online version of this article.

Keywords: phospholipids; algorithms; CID; nLC–ESI–MS–MS; software tool; LiPilot

INTRODUCTION

Lipidomics is an emerging research area that focuses on characterizing the extensive classes of lipid molecules from cells, tissues, and biological fluids. This field also aims to understand lipid-mediated pathways of metabolism in relation to the discovery of disease biomarkers. Among the lipidomes, phospholipids (PLs) are of considerable interest because they are involved in signal transduction, cell proliferation, and cell death.[1,2] PLs, the major components of cellular membranes, are molecules consisting of a glycerol backbone, a polar head group attached through a phosphate at the sn-3 position and one or two acyl chains attached to the remaining sn-1 and sn-2 positions with different degrees of unsaturation and acyl chain lengths. Due to combinations of different head groups and acyl chains, the composition of PLs in biological samples is very diverse. To understand how PLs vary upon a change in cellular status or disease-related stimuli, comprehensive analysis of PL molecules by both qualitative and quantitative methods is an important and often challenging aspect of lipidomic research.

Through recent advances in electrospray ionization–tandem mass spectrometry (ESI–MS–MS), in which secondary MS analysis provides fragment ion spectra leading to structural identification of lipid molecules, shotgun lipidomic methodology has rapidly grown into a promising tool for lipid analysis.[3–8] Although ESI–MS or ESI–MS–MS methods have been successfully applied to biological tissues, a preliminary separation step is still required to reduce the ionization suppression effect caused by spectral congestion of complicated PL mixtures. High-performance liquid chromatography (HPLC) coupled with ESI–MS[9,10] increases the ability to characterize complicated PL mixtures by separating different polar heads on a normal-phase column while sorting different chain lengths and degrees of acyl chain unsaturation on a reversed-phase column. Capillary LC under microflow or nanoflow regimes is directly coupled with ESI–MS–MS, leading to the simultaneous separation and characterization of PLs from human tissue and urine samples with detection limits as low as 2.2 fmol.[11–14] However, one of the main difficulties in high-throughput analysis of PL molecules arises when vast amounts of MS data are manually processed for structural identification. This analysis bottleneck necessitates easy and accurate tools for identifying PLs.

Automated identification of PL structures from complicated MS and MS–MS spectra has been attempted by developing several programs, including LIMSA,[15] LipidProfiler,[16] Lipid MAPS,[17] LipidQA,[18,19] LipID[20] and LipidXplorer.[21] Although these software programs use different algorithms to identify lipids including PLs, most of them have been customized to work with specific types of MS instruments (some of them require high-

* *Correspondence to: Myeong Hee Moon, Department of Chemistry, Yonsei University, Seoul, 120-749, Korea. E-mail: mhmoon@yonsei.ac.kr*

Department of Chemistry, Yonsei University, Seoul, 120-749, Korea

Journal of MASS SPECTROMETRY

resolution MS). Recently, LipidXplorer has been reported as a useful tool for cross-platform lipidomics.[22] Among the listed programs, LIMSA, LipidMaps and LipID rely on lipid identification by MS scanning alone even with the use of LC. Although LipidProfiler, LipidQA and LipidXplorer operate with a database of reference MS–MS spectra constructed from standard species or rely on a multiple precursor ion scanning method to identify PL species, they are focused to work with data obtained from direct ESI–MS–MS analysis. For an accurate determination of molecular structure from complicated mixtures, separation of complicated PL mixtures (especially of regioisomers) prior to MS analysis is necessary along with data-dependent MS–MS experiment, and it is useful to develop a software to cover the growing needs.

In this study, we report the development of LiPilot, an automated tool for identifying PLs from nanoflow liquid chromatography–electrospray ionization–tandem mass spectrometry (nLC–ESI–MS–MS) experiments. LiPilot can use data inputs from even relatively low-resolution MS instruments. Fractionated PL molecules that are introduced from nLC via ESI are analyzed by MS with data-dependent switching to MS–MS. LiPilot utilizes the entire set of raw data, including precursor ion MS scans and data-dependent MS–MS spectra. The software identifies PLs by comparing observed data with possible fragment ions from a library containing *m/z* values of fragment ions generated in typical fragmentation patterns of PLs experimentally found by ion-trap MS instruments. Candidate PL molecules chosen by the precursor ion search against the database were examined by comparing differences between the theoretical and experimental *m/z* values. A scoring structure developed in this study was used to maximize the identification and differentiation of PL and LPL regioisomers. Although the proposed method provides a targeted identification tool for particular PL classes at the first stage, other classes of lipids such as glycolipids, gangliosides, ceramide and triacylglycerol can be simply integrated in the future. The developed software was evaluated with a mixture of 17 PL and six LPL standards, including all six naturally occurring head groups. LiPilot was also applied to the identification of urinary PLs and LPLs from a prostate cancer patient sample.

EXPERIMENTAL SECTION

Chemicals and PL standards

All synthetic LPLs (16:0-LPC, 18:0-LPE, 18:0-LPS, 18:1-LPI, 18:0-LPG and 18:0-LPA) and PLs (14:0/14:0-PC, 16:0/16:0-PC, 18:0/18:0-PC, 14:0/14:0-PE, 16:0/16:0-PE, 18:0/18:0-PE, 14:0/14:0-PS, 16:0/16:0-PS, 18:0/18:0-PS, 16:0/18:2-PI, 12:0/12:0-PG, 16:0/16:0-PG, 18:0/18:0-PG, 12:0/12:0-PA, 14:0/14:0-PA, 16:0/16:0-PA and 18:0/18:0-PA) were purchased from Avanti Polar Lipids (Alabaster, AL, USA). All solvents used for LC–MS experiments (methanol, acetonitrile, isopropanol and water) were HPLC grade from Avantor Performance Materials (Phillipsburg, NJ, USA). Chloroform, formic acid and ammonium hydroxide were HPLC grade.

nLC–ESI–MS–MS

The developed software tool was tested with experimental data obtained from nLC–ESI–MS–MS with PL standards and a urinary PL mixture sample extracted from healthy human urine. A model 1200 series capillary pump system (Agilent Technologies, Palo Alto, CA, USA) equipped with an autosampler and an LTQ Velos ion trap mass spectrometer (Thermo Finnigan, San Jose, CA, USA) equipped with a nano-electrospray ionization source was employed to generate raw mass spectra data. For nLC, silica capillary tubes with 20, 50 and 75 μm inner diameters (i.d.) and a 360 μm outer diameter (Polymicro Technology LLC, Phoenix, AZ, USA) were used. For nLC column preparation, one end of the capillary tube was pulled by flame to create a direct emitter. The pulled-tip capillary column (75 μm i.d. × 5 cm long) was packed with Magic C_{18}, 3 μm–100 Å (Michrom Bioresources Inc., Auburn, CA, USA) in the laboratory. Detailed procedures for column packing and system configuration for gradient elution have been reported previously.[23,24]

Gradient elution was utilized to separate PLs from human urine. The two mobile-phase compositions were as follows: 90/10 (v/v) water/acetonitrile for A and 20/20/60 acetonitrile/methanol/isopropanol for B. Both mobile-phase solutions included 0.1% (v/v) formic acid for the positive ion mode of nLC–ESI–MS–MS to generate $[M+H]^+$ ions without forming $[M+Na]^+$. Both mobile-phase solutions also included 0.05% sodium hydroxide for the negative ion mode to solely generate $[M-H]^-$ ions. The gradient elution protocol increased mobile-phase B from 0% to 40% over 1 min, ramped B to 90% over 44 min and finally reached 100% B for 10 min for both positive and negative ion modes. For ESI, 3 kV was applied. The MS range for the first MS scan was *m/z* 450–950 for positive ion mode and 380–950 for negative ion mode. During collision-induced dissociation (CID) experiments, the normalized collision energy was set at 40% for both positive and negative ion modes. Because regioisomers are in most cases separated by nLC, an exclusion time to skip the MS–MS run of a same precursor ion was set after triplicate measurements. The exclusion time was assigned for ~30 s depending on the peak width of each regioisomer of PL standards in a given LC run condition.

Development of LiPilot software

The software LiPilot was written in Windows C^{++} language to ensure compatibility with any Windows OS-based PC. The program was designed to work with text files that were converted from raw MS data obtained from ion trap mass spectrometers using the XCalibur software from Thermo Scientific. The program is available upon request from the authors free of charge. The program's identification algorithms and features will be explained in detail in the next section.

RESULTS AND DISCUSSION

Generation of PL library

Library files must be generated in .csv (comma separated value) format for each PL head group. Each file contains a list of *m/z* values of theoretical fragment ions for all possible molecules with each head group. Because PL molecules with different head groups exhibit different fragmentation patterns, typical fragment ions for each PL molecule are listed separately in a separate library file. For instance, the PS library contains *m/z* values of observable fragment ion classes (seven different types of fragment ions marked in Table 1a) for all possible PS molecules with the background knowledge that general eukaryotic cells contain PLs with two acyl chains ranging in length from 14 to 22 carbons. Because PS molecules are readily detected in the negative ion mode of MS, the PS library is constructed with negatively charged fragment ions. The same principle is applied to generate libraries for other head groups according to the possible fragment ions

Table 1. List of fragments chosen for the construction of library files and for computational identification of PL and LPL classes (a) in the negative ion mode and (b) in the positive ion mode of ESI–MS–MS from an LTQ Velos mass spectrometer

(a) Negative ion mode	PS	LPS		PG	LPG		PI	LPI		PA	LPA	
		sn-1	sn-2		sn-1	sn-2		sn-1	sn-2		sn-1	sn-2
$[M - H]^-$											○	●
$[M - H\text{–head}]^-$	●	●	●		○	○		○	○			
$[M - H\text{–}R_1COOH]^-$				○			○		○	○		○
$[M - H\text{–}R_1'CH{=}C{=}O]^-$				○			○		○	○		○
$[M - H\text{–}R_2COOH]^-$		○		○			○	●		●	●	
$[M - H\text{–}R_2'CH{=}C{=}O]^-$				○			○	○		○	○	
$[M - H\text{–head–}R_1COOH]^-$	○		○	○		○	○					
$[M - H\text{–head–}R_1'CH{=}C{=}O]^-$	○			○		○	○					
$[M - H\text{–head–}R_2COOH]^-$	●	○		○	●		●					
$[M - H\text{–head–}R_2'CH{=}C{=}O]^-$	○	○		○	○		○					
$[R_1COO]^-$	○			●		●	○		○	○		○
$[R_2COO]^-$	○			○	○		○	○		○	○	
$[M - H\text{–head–}H_2O]^-$								○	●			

(b) Positive ion mode	PC	LPC		PE	LPE	
		sn-1	sn-2		sn-1	sn-2
$[M + H\text{–head}]^+$	○			●	○	○
$[M + H\text{–}R_1COOH]^+$	○			○		○
$[M + H\text{–}R_1'CH{=}C{=}O]^+$	○			○		○
$[M + H\text{–}R_2COOH]^+$	○	○		○	○	
$[M + H\text{–}R_2'CH{=}C{=}O]^+$	●	○	○	●	○	
$[Head + H]^+$		●	○			
$[M + H\text{–}H_2O]^+$		○	●		○	○

○: chosen, ●: chosen as additional I_{factor}. sn-1 and sn-2 indicate locations of the acyl chains in the glycerol backbone of each lysophospholipid.

selected (marked with open and filled circles in Table 1). Fragment ions marked with filled circles are used as indicators of structural differentiation, as will be explained later. The fragment ion libraries for PIs, PGs, PAs and PSs were constructed with fragment ions that can be generated at the negative ion mode of MS, whereas the libraries for PCs and PEs were based on fragment ions at the positive ion mode, as listed in Table 1.

Data pretreatment for programmed analysis

Initial treatment of raw data from nLC–ESI–MS–MS experiments is shown in the schematic flow charts of Fig. 1. The raw MS spectra (.raw file) file contains *m/z* information of each ion with its MS intensity at each retention time slice and *m/z* information of fragment ions along with their corresponding precursor ions. The raw data require conversion from raw format (.raw) into text format (.txt). This conversion is readily achieved with the XCalibur software from Thermo Finnigan directly after data acquisition. This step could take a few minutes depending on the number of files to be converted and the computer's ability to process large amounts of data. Because the converted text file typically occupies at least ten times more space than does the corresponding original raw file, sufficient space for file saving is recommended. The format of .txt file generated from LiPilot is shown in the Supplementary Information. It contains a header containing information of scan number, nLC run time interval and mass range. Each MS data begins with a separate header containing nLC retention time and the type of scan (MS or MS–MS scan) followed by *m/z* value with its intensity.

Programmed analysis with LiPilot is initiated by importing libraries, loading experimental nLC–MS–MS data files in text format and assigning fragment patterns with the calculation parameters shown in the schematic flow chart of Fig. 1. When loading the library, the library files of different PL head groups are loaded together for simultaneous determination of multiple

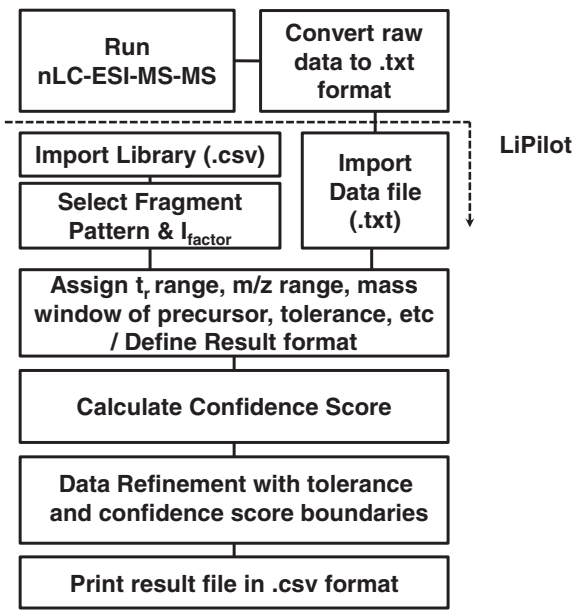


Figure 1. Flow chart of the LiPilot program.

Journal of MASS SPECTROMETRY

lipid species. At this stage, text file minimization is critical because the initial text file has a vast amount of redundant data, such as numerous precursor ions without accompanying CID spectra and MS run conditions. These data occupy the majority of the file size. Analysis requires strategic sieves to reduce the initial text file to a temporary file containing only retention time, *m/z* value of each precursor ion, *m/z* values of fragment ions and MS intensity from the corresponding CID spectra.

Setting up calculation parameters

After importing library and data files, calculation parameters are assigned. These parameters are helpful for increasing the probability of PL identification in the search algorithm. Parameter selection is based on experimental observation of fragment ion patterns that are specific to the MS instrument employed. For instance, Fig. 2 shows data-dependent CID spectra of a PI molecule from a human urine sample obtained from the LTQ-Velos ion trap MS during nLC–ESI–MS–MS analysis that was manually identified as 18:0/22:6-PI. The temporary text file compressed from the initial text file contains information on the *m/z* value of 909.6 for the precursor ion ($[M - H]^-$) along with CID spectra representing characteristic fragment ions that indicate the location and length of the two acyl chains and the head group. Fragment ions of *m/z* 643.3 and 599.3 represent the loss of fatty acid in the form of ketenes, $[M - H\text{-}R_1'CH{=}C{=}O]^-$ and $[M - H\text{-}R_2'CH{=}C{=}O]^-$, respectively, from the precursor ion. Similar dissociations of acyl chains in the form of carboxylic acid were observed at *m/z* 625.3 ($[M - H\text{–}R_1COOH]^-$) and 581.2 ($[M - H\text{–}R_2COOH]^-$). Fragment ions that lost the inositol head group, $[M - H\text{–}162]^-$, from the parent ion were not observed with PI. Other PLs such as PC or PE do generate head group fragment ions. However, additional fragment ions dissociated from the parent ion without the inositol head group were observed at *m/z* 481.4 and 437.2, indicative of $[M - H\text{–}162\text{–}R_1'CH{=}C{=}O]^-$ and $[M - H\text{–}162\text{–}R_2'CH{=}C{=}O]^-$, respectively. Likewise, fragment ions at *m/z* 463.2 and 419.2 represent the loss of carboxylic acid from the parent ion without the head group. Moreover, fragment ions observed at *m/z* 283.3 and 327.3 were the free carboxylate anions of the acyl chains, $[R_1COO]^-$ and $[R_2COO]^-$, respectively. In addition, the fragment ion $[M - H\text{–}R_2COOH]^-$ appears in greater abundance than $[M - H\text{-}R_1COOH]^-$, whereas the intensity of the free carboxylate anion $[R_1COO]^-$ was greater than that of $[R_2COO]^-$. These data can be utilized to differentiate the acyl chain position as found in the literature.[25,26]

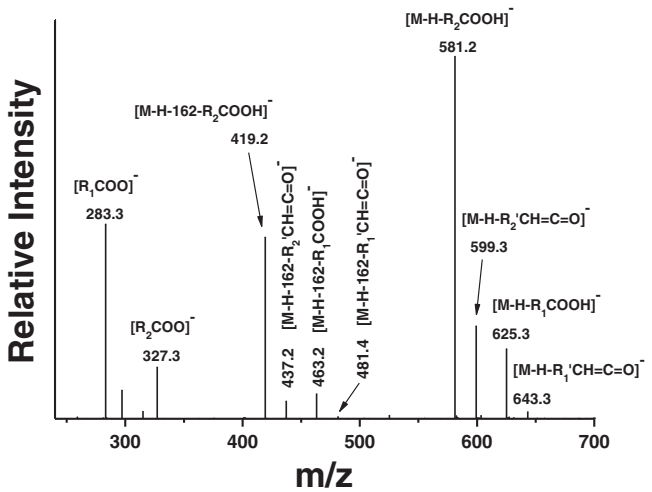


Figure 2. Data-dependent CID spectra of a PI molecule (precursor *m/z*=909.6) obtained by nLC–ESI–MS–MS analysis of a human urinary PL mixture sample. Characteristic fragment ions yield manual identification of 18:0/22:6-PI.

In the first stage of assigning calculation parameters in the LiPilot software, the type of PL head group to be examined should be selected along with the ionization mode (negative or positive) of the MS instrument. Next, the screen shown in Fig. 3a appears with a list of temporary library files to be generated on the right side. For example, PI_LTQ_Velos_N represents the head group (PI), MS instrument utilized (LTQ Velos) and negative ion detection mode (N). Each temporary library file is extracted from the library files loaded for various head groups by including the specific fragment ion types that should be examined by the search algorithm in the Set Peak panel on the left side of Fig. 3a. Typical fragment ions that can be observed by the LTQ Velos ion trap MS and characteristic fragment ion(s) should be selected in the Set Factor panel to distinguish PL regioisomers. For instance, 18:0/22:6-PI and 22:6/18:0-PI are regioisomers with almost identical fragment ions, except for different peak intensity ratios for specific fragment ion pairs such as $[M - H\text{–}R_1COOH]^-/[M - H\text{–}R_2COOH]^-$. Thus, the differentiation factor can be $[M - H\text{–}R_2COOH]^-$ for PI with the current MS instrument, as shown in Fig. 3a, or it can be selected with an additional factor such as $[R_1COO]^-$. The intensity of this factor, I_{factor}, will be incorporated into Eqn 1 (as will be shown later) to calculate the confidence score of each candidate molecule. Regioisomers of PLs (including LPLs) can be distinguished by different retention times in our nLC–ESI–MS–MS analysis in C18 stationary phase.[23,24] However, this was not incorporated in the program. Figure 3b is an example of the PC screen that shows detailed setup for additional parameters used to screen out the unnecessary portion of data. These parameters include selecting the retention time interval to be examined, *m/z* range of precursor ions, dynamic exclusion to avoid generation of the same results for a given period, precursor mass width and fragment ion mass width. Other parameters related to filtering the search results to enhance identification accuracy will be discussed later.

Calculation algorithm

Identification of PL molecules from nLC–ESI–MS–MS data begins with selecting candidate PL molecules from library files for each set of MS–MS data utilizing the LiPilot program with the following procedures. When MS–MS fragment ion spectra from an experiment appear with an experimental *m/z* value of the corresponding precursor ion, LiPilot automatically chooses the first draft results from a group of library files, as shown in Fig. 4a. LiPilot chooses these results with respect to precursor mass and a precursor mass width assigned during the parameter setup in Fig. 3b. The mass width for the precursor ion is a given *m/z* interval around the precursor ion *m/z* value that can be arbitrarily assigned. For example, when an *m/z* width of 0.5 is applied to a precursor ion with *m/z* 788.5, LiPilot extracts all PL molecules that have theoretical precursors with an *m/z* from 788.0 to 789.0 from the libraries of only negative ion mode, as shown in Fig. 4b and 4c. Figure 4c lists the four PS candidates that have the same precursor ($[M - H]^-$) *m/z* value of 788.06 with different acyl chains. Next, LiPilot begins searching for candidate molecules that have an *m/z* value of each fragment ion closest to the theoretical *m/z* of each corresponding fragment ion. For the first row of the extracted list with theoretical precursor mass value 788.06, *m/z* values of $[R_1COO]^-$ and $[R_2COO]^-$

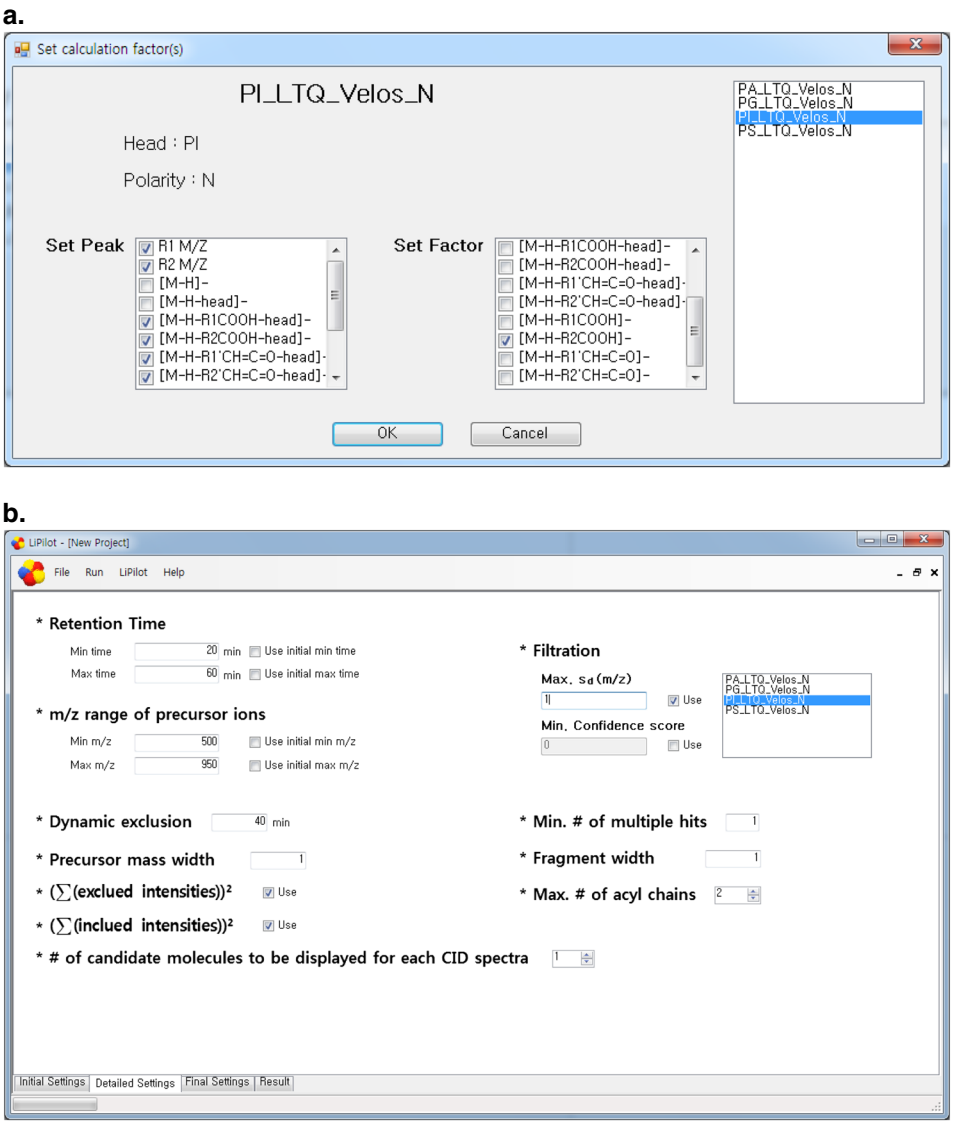


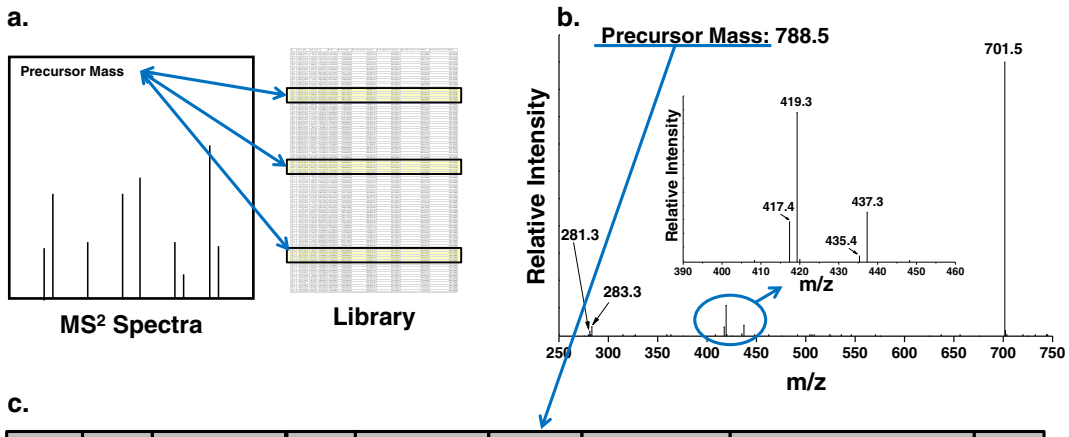
Figure 3. Captured PC screens of the LiPilot program during initial setup of (a) parameters such as type of fragment ions to be examined for each head group and (b) additional parameters for screening purposes, such as *m/z* width for precursor and fragment ions.

are 283.48 and 281.46, which are closest to the experimental *m/z* of 283.3 and 281.3 from the CID spectra (Fig. 4b), respectively. Each candidate molecule is examined to find the closest fit for theoretical *m/z* values of fragment ions with the experimental *m/z* values of observed fragment ions. In the case of the PS molecule listed in the first row of Fig. 4d, the experimental *m/z* value of the fragment ion found closest to the theoretical fragment ion is indicated below the theoretical fragment ion. To select candidate molecules with small differences between experimental and theoretical fragment ion *m/z* values, an elaborate scoring structure, the confidence score, was developed in this study as follows:

$$\text{Confidence score} = \frac{\left(\sum_{k=1}^{i} I_k\right) \times I_{\text{factor}}}{\left(\sum_{l=1}^{N-i} I_l^{\text{E}}\right)^2} \times \frac{1}{\sqrt{\frac{\sum_{k=1}^{i}\left(m_k^e - m_k^l\right)^2}{N-1}}} \quad (1)$$

where i is the number of observed fragment ions chosen in each MS–MS spectrum matching theoretical fragment ions in the library files, N is the number of fragments in each MS–MS spectrum, $\sum_{k=1}^{i} I_k$ is the sum of intensities (I_k) of all experimental fragment ions selected for comparison, I_{factor} is the intensity of the specific fragment factor chosen by the set factor of Fig. 3a to be multiplied for regioisomer differentiation and score optimization, $\sum_{l=1}^{N-i} I_l^{\text{E}}$ is the sum of intensities of all other ions excluding the intensities of peaks (I_k) matching fragment ions in the library, and m_k^e is the *m/z* value of the kth experimental fragment ion that matches an ion with the closest *m/z* value, m_k^l, in the library. $[M - H\text{–head–}R_2COOH]^-$ is experimentally known to be more dominant than $[M - H\text{–head–}R_1COOH]^-$ for PS molecular species. For this reason, the intensity of $[M - H\text{–head–}R_2COOH]^-$ in the case of PS can be incorporated as the I_{factor} into Eqn 1 to differentiate regioisomers of PS molecules that differ in the locations of two different acyl chains at the sn-1 and sn-2 positions. Fragment ions selected for differentiation of regioisomers for each PL group are marked with filled circles in Table 1. For PS, $[M - H\text{–head}]^-$ can be selected as an additional I_{factor} to be multiplied in Eqn 1 because the intensity of the fragment ion

Journal of MASS SPECTROMETRY



c.

Head	R_1	$[R_1COO]^-$	R_2	$[R_2COO]^-$	$[M-H]^-$	[M-H-Head]$^-$	[M-H-Head-R_1COOH]$^-$	…
PS	18:0	283.48	18:1	281.46	788.06	700.98	416.50	…
PS	18:1	281.46	18:0	283.48	788.06	700.98	418.51	…
PS	16:0	255.42	20:1	309.51	788.06	700.98	444.55	…
PS	20:1	331.52	16:0	255.42	788.06	700.98	390.46	…

d.

Head	R_1	$[R_1COO]^-$	R_2	$[R_2COO]^-$	[M-H-Head]$^-$	[M-H-Head-R_1COOH]$^-$	[M-H-Head-R_2COOH]$^-$
PS	18:0	283.48	18:1	281.46	700.98	416.50	418.51
		283.3		281.3	701.5	417.4	419.3
PS	18:1	281.46	18:0	283.48	700.98	418.51	416.50
		281.3		283.3	701.5	419.3	417.4
PS	16:0	255.42	20:1	309.51	700.98	444.55	390.46
		279.8		314.8	701.5	448.2	408.4
PS	20:1	331.52	16:0	255.42	700.98	390.46	444.55
		314.8		279.8	701.5	408.4	448.2

e.

Head	R_1	R_2	$\sum_{k=1}^{i} I_k$	$\sqrt{\frac{\sum_{k=1}^{i}(m_k^e - m_k^t)^2}{N-1}}$	I_{factor}	p-value (t-test)	Confidence Score
PS	18:0	18:1	121.87	0.71	3.00	0.70	131.45
PS	18:1	18:0	121.87	0.71	1.37	0.70	59.89
PS	16:0	20:1	23.78	12.57	0.00	0.79	0.00
PS	20:1	16:0	23.78	12.57	0.03	0.79	0.00

Figure 4. Illustration of the search algorithm in LiPilot with identification of a PS molecule: (a) selection of candidate molecules with *m/z* values that are within a certain interval of the experimental *m/z* value of a precursor ion, (b) CID spectra of 18:0/18:1-PS with the precursor *m/z* value of 788.5, (c) first draft of an imported library with respect to the precursor mass value obtained from experimental CID spectra, (d) lists of *m/z* values of theoretical fragment ions (in the first row of each candidate molecule) matching the *m/z* values of nearby fragment ions from CID spectra in panel b and (e) calculation of confidence scores along with screening parameters such as SDD and I_{factor}, supporting the identification of 18:0/18:1-PS.

$[M - H–head]^-$ is significantly greater than the intensity of other fragment ions. In this case, an additional I_{factor} increases the scores of candidate PS molecules, which also increases the probability of PL identification. The second part of Eqn 1 is an inverse of the standard deviation of differences (SDD) between experimental and theoretical *m/z* values of fragment ions. This value is multiplied to amplify scores depending on how closely the observed *m/z* values match those of the corresponding theoretical fragments. Figure 4e lists the calculated values of parameters in Eqn 1 along with the confidence score of each candidate molecule. For instance, the first two PS molecules show exceedingly high confidence scores compared with the other candidate PS molecules. However, 18:0/18:1-PS has a score of 131.45, which is higher than that of 18:1/18:0-PS due to the introduction of I_{factor}. For the clarification of the selective identification of regioisomers, a pair of regioisomer standards (14:0/16:0-PC and 16:0/14:0-PC) mixed at 1 : 1 ratio was analyzed by nLC–ESI–MS–MS at positive ion mode. Figure 5 shows the BPC of the regioisomers (slight difference in retention time) in the center along with data-dependent CID spectra obtained from the same *m/z*=706.7 at (i) 37.35 min and (ii)

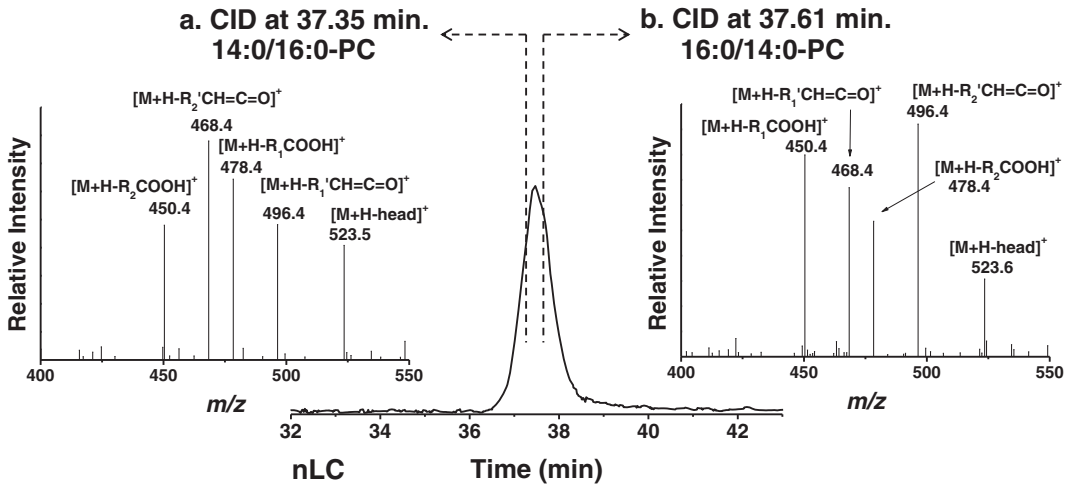


Figure 5. BPC of standard mixture of regioisomers (14:0/16:0-PC and 16:0/14:0-PC: *m/z* 706.7) along with the CID spectra obtained at (a) 37.35 min and (b) 37.61 min during nLC–ESI–MS–MS run.

37.61 min. Although the two species are not completely separated by nLC due to the small difference in geometrical structure, CID patterns of both ions are clearly distinguished by the difference in the relative intensity of $[M+H\text{-}R_2'CH{=}C{=}O]^+$ ion to $[M+H\text{-}R_1'CH{=}C{=}O]^+$, of which formation of the former ion is preferred in positive ion mode as found in the references.[12,27] The search results from LiPilot yield with the unique identification of regioisomers without any candidate molecule suggested (confidence scores were 660.0 for 14:0/16:0-PC and 391.6 for 16:0/14:0-PC). Moreover, the difference in retention times of regioisomers can be utilized as an indicator to distinguish the structural difference in which a PC molecule having a longer chain at the sn-1 position retains longer in RP resin, as can be found from the observation with LPC in literature.[27]

Likewise, all PL classes (PC, PE, PS, PI, PG and PA) and LPL classes (LPC, LPE, LPS, LPI, LPG and LPA) can be successfully identified. It should be noted that the confidence score value may yield significantly different values among lipid classes because the algorithm selects different types of fragment ions for each group of PL or LPL molecules. Therefore, an absolute comparison of calculated scores among the different PL groups is often not meaningful. However, comparing scores of candidate molecules within the same head group is reliable, with a higher score representing higher relevance.

Filtering results

After the confidence score is calculated for a set of fragment ions within the *m/z* width of fragment ions, redundant results must be filtered out. The precursor mass width value itself is not incorporated into the confidence score calculation. However, the SDD between experimental and theoretical *m/z* values of fragment ions, $\sqrt{\frac{\sum_{k=1}^{i}\left(m_k^e - m_k^l\right)^2}{N-1}}$, defined in Eqn 1, plays a role in finding probable PL molecules. If the SDD values are significantly different for various head groups, the threshold value of SDD can be assigned for each library specifically made for each head group. Table 2 shows an example of search results acquired by LiPilot from a data file obtained with a number of standard PL molecules assigned the *m/z* width of 1.0 for precursor ions and fragment ions. It lists the structure of each PL molecule together with the calculated value of SDD, $\sum_{k=1}^{i} I_k$, and confidence score. When SDD values are examined with a number of standard PL molecules, they are experimentally found to exhibit SDD values less than 1.0. Figure 6 shows plots of the total number of PL molecules reported by the program *versus* different threshold values of SDD. For PS, 32 molecules (white bars) including the three PS standards (marked with black bars) are reported when the SDD value is set to 3. However, when a threshold SDD value of 0.3 is applied, only two of the PS standards are identified. Threshold SDD values for successful identification of standard PL molecules are found to differ according to the type of PL head group, as shown with plots for the other head groups in Fig. 6. The lower left plot of Fig. 6 shows that overall threshold values for the four different head groups should be larger than 1.0 to identify all PL standard molecules. However, this threshold still contains a number of false results that require screening. To improve the identification success ratio, it is helpful to compare the values of $\sum_{l=1}^{N-i} I_l^E$, the sum of intensities of all other ions excluding the intensities of peaks matched with fragment ions specified in each library, which correlates to a great extent with confidence scores. For instance, the confidence score of a PA molecule found to be a false positive (i.e. 20:0/12:0-PA in the fifth row in Table 2) is significantly smaller than the confidence scores of true PA molecules: ~100 *versus* 2–19 $\times$ 10^6. Confidence scores for the other head groups also show differences of more than four orders of magnitude between false-positive and actual PLs. Thus, confidence scores should be well established with sufficient results using standard species or checked manually for speculative species. In the LiPilot program, maximum SDD values and minimum confidence scores can be assigned for each category of a head group in the setup menu, as shown in Fig. 3b.

Application to human urinary PL extract

The software was applied to an authentic sample using human urinary PL extract with the same nLC–ESI–MS–MS run conditions employed for the standards in Fig. 2. From data generated for

Journal of MASS SPECTROMETRY

Table 2. Search results from LiPilot for LC-ESI–MS–MS data obtained with standard phospholipids

		Suggested ID			SDD			
t_r	Observed *m/z*	Head	sn-1	sn-2	$\sqrt{\sum_{k=1}^{i}\left(m_k^e - m_k^l\right)^2 / (N-1)}$	$\sum_{l=1}^{N-i} I_l^E$	Confidence score	Success
18.27	647.4	**PA**	**16:0**	**16:0**	0.71	1.2	1.92E7	o
16.67	591.4	**PA**	**14:0**	**14:0**	0.74	1.6	9.01E6	o
20.53	703.5	**PA**	**18:0**	**18:0**	0.69	2.7	4.27E6	o
15.38	535.4	**PA**	**12:0**	**12:0**	0.9	2.3	2.13E6	o
14.45	647.4	PA	20:0	12:0	0.57	140.1	1.02E2	x
.	.	.	.	.	.	.	.	
25.03	777.5	**PG**	**18:0**	**18:0**	0.68	1.8	4.12E6	o
17.63	609.5	**PG**	**12:0**	**12:0**	0.77	3.1	2.47E6	o
22.13	721.5	**PG**	**16:0**	**16:0**	0.73	2.7	1.98E6	o
18.32	777.5	PG	16:0	20:0	0.75	624.4	0.1	x
.	.	.	.	.	.	.	.	
21.32	833.5	**PI**	**16:0**	**18:2**	0.7	24.9	1.60E4	o
21.32	833.5	PI	18:2	16:0	0.68	403.4	2.2	x
12.77.	777.5.	PI.	16:1.	14:1.	0.98.	336.	0	x
18.07	678.4	**PS**	**14:0**	**14:0**	0.77	0.5	4.93E7	o
23.22	790.5	**PS**	**18:0**	**18:0**	0.71	0.7	4.13E7	o
20.23	734.5	**PS**	**16:0**	**16:0**	0.73	0.8	2.69E7	o
43.31	721.5	PS	16:4	16:3	0.89	340.4	25.8	x
.	.	.	.	.	.	.	.	

PLs marked in bold are standard species employed for validation tests. o: successful match, x: not contained in the standard mixture.

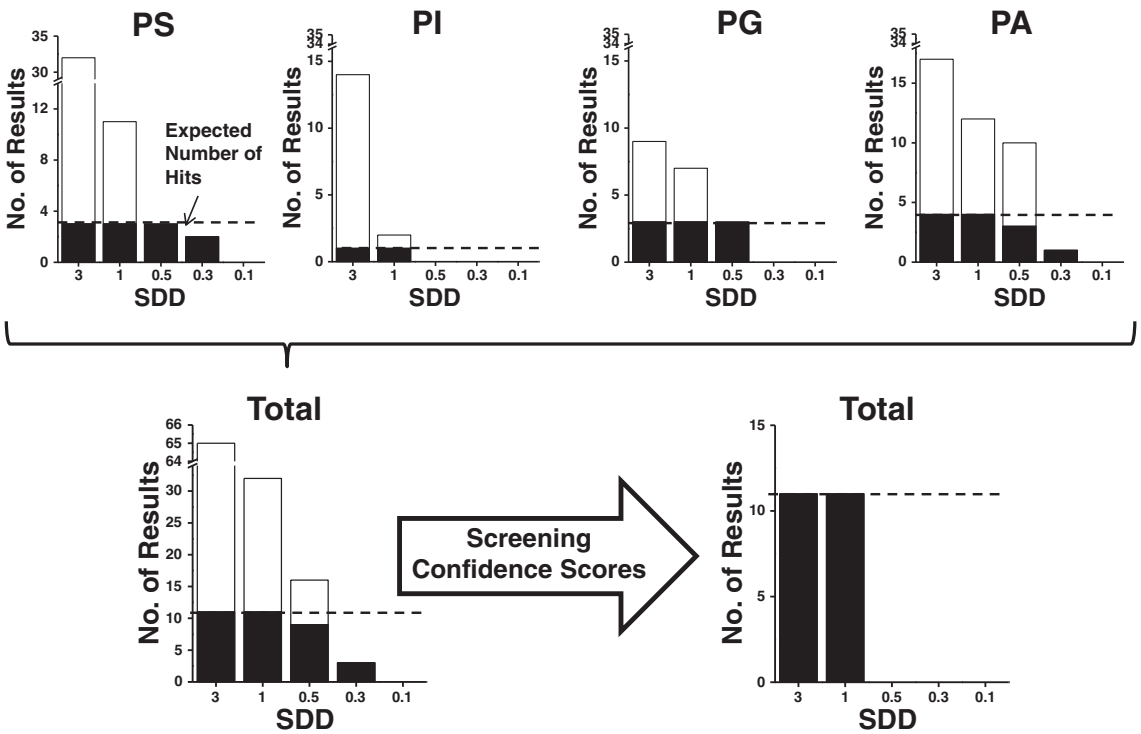


Figure 6. Effects of threshold SDD value and screening confidence scores on the identification of PL standard molecules.

this PL mixture sample, confidence scores varied widely depending on the experimental LC run conditions, which could be used to differentiate isobaric or isomeric species. LiPilot led to the identification of 93 PLs (26 PCs, 15 PEs, 19 PSs, 11 PGs, 20 PI, and 2 PAs) and 22 LPLs (4 LPCs, 7 LPEs, 2 LPSs, 1 LPG, 1 LPI and 7 LPAs), as listed in Table 3. Although most species found by LiPilot had significantly large score values, some exhibited small values (< 100), which were confirmed by manual examination of the CID spectra. Even the isobaric species 18:1/18:1-PS and 18:0/18:2-PS, which have the same *m/z* values, could be identified

Table 3. Identified PLs from nLC–ESI–MS–MS analysis of a human urine sample utilizing the developed LiPilot software

Head	Species	*m/z*	t_r	Confidence score	Head	Species	*m/z*	t_r	Confidence score
(a) Negative ion mode									
PS	14:0/14:0	678.6	33.85	7.36E5	**PI**	16:0/16:1	807.6	38.82	5
	16:0/22:6	807.5	38.79	19		16:0/18:2	833.6	39.55	3.24E2
	16:0/18:2	758.6	38.94	1.73E6		18:1/16:1	833.6	39.55	3.24E2
	16:0/20:4	782.5	39.06	1.67E6		16:0/20:4	857.6	39.65	5.58E2
	18:1/22:6	833.6	39.40	48		16:0/22:6	881.6	39.71	4.21E3
	18:0/20:5	808.6	40.46	2.11E6		16:0/22:5	883.6	40.10	12
	16:0/18:1	760.6	40.60	3.63E6		18:1/20:4	883.6	40.10	12
	16:0/22:4	810.6	40.93	5.55E2		16:0/20:3	859.6	41.00	2.01E3
	18:1/18:1	786.6	41.44	9.54E5		16:0/18:1	835.6	41.22	1.01E3
	18:0/18:2	786.6	41.49	2.30E3		18:1/18:1	861.6	41.89	1.77E3
	18:0/20:4	810.6	41.95	1.87E6		18:0/22:6	909.6	42.23	1.60E4
	18:0/22:6	834.6	42.25	1.27E6		18:0/20:4	885.6	42.76	74
	18:0/22:5	836.6	42.62	3.15E6		18:0/22:5	911.6	42.89	3.96E2
	18:0/20:3	812.6	43.04	7.13E6		18:0/18:2	861.6	42.93	1.77E4
	18:1/20:3	810.6	43.63	7.29E2		18:0/20:3	887.6	43.49	3.30E4
	20:1/18:1	814.6	44.03	1.57E6		18:0/16:0	837.6	43.67	3.76E3
	18:0/18:1	788.6	44.27	1.93E7		18:0/18:1	863.6	44.04	1.26E3
	20:0/18:1	816.7	46.55	1.34E5		20:1/16:0	863.6	44.58	6
	18:0/20:1	816.7	46.55	1.34E5		18:0/22:4	913.6	44.70	1
PG	18:2/22:6	817.6	38.02	1.60E3		18:0/20:2	889.6	44.99	1.02E3
	22:6/22:6	865.6	38.58	3.11E3					
	22:5/22:6	867.6	38.90	8.29E2	**PA**	18:0/18:2	699.4	39.08	1.63E3
	18:2/18:2	769.6	39.10	1.11E4		18:0/18:1	701.5	41.12	6.40E2
	18:1/22:6	819.6	39.81	1.78E3					
	18:2/16:0	745.6	40.35	6.91E3	**LPI**	lyso/18:0	599.4	22.02	5.57E2
	18:2/18:1	771.6	40.66	3.87E3					
	18:1/18:2	771.6	40.68	1.16E4	**LPA**	16:1/lyso	407.5	25.71	8.56E2
	18:1/16:0	747.6	42.14	4.04E2		lyso/18:3	431.3	28.64	20
	18:1/18:1	774.6	42.43	0.17		lyso/16:0	409.4	29.28	19
	18:2/18:0	773.6	43.28	24		14:0/lyso	381.6	31.82	2.51E2
LPS	lyso/18:0	524.3	21.58	4.78E6		16:0/lyso	409.6	32.57	2.65E3
	18:0/lyso	524.3	21.79	4.41E5		18:0/lyso	437.5	35.79	2.05E3
LPG	16:0/lyso	483.3	21.31	7.67E2		20:0/lyso	465.6	37.96	3.68E2
(b) Positive ion mode									
PC	14:0/14:0	678.6	35.46	8	**PE**	20:6/18:0	764.7	39.94	8.76E3
	18:2/16:1	756.7	37.80	68		22:6/16:0	764.7	40.81	7.30E3
	20:5/16:0	780.7	38.53	81		16:0/18:2	716.6	41.10	1.07E4
	18:3/16:0	756.7	38.58	1		20:4/16:0	740.7	41.13	4.38E2
	20:6/18:0	806.7	39.08	3		18:1/22:6	790.7	41.18	2.10E4
	18:1/20:5	806.7	39.10	4		22:5/16:0	766.7	41.44	1.07E3
	16:1/16:0	732.7	39.74	5.04E2		18:1/18:2	742.7	41.59	1.43E4
	18:1/16:1	758.7	39.82	13		18:1/16:0	718.7	43.06	1.61E4
	20:4/16:0	782.7	40.20	4		18:1/18:1	744.6	43.46	3.43E3
	22:5/16:0	808.7	40.71	22		22:6/18:0	792.7	43.56	5.02E3
	16:0/18:2	758.7	40.87	51		20:1/16:0	746.7	43.87	2.25E2
	18:2/18:1	784.7	40.97	2.15E3		18:0/20:4	768.7	43.94	1.04E3
	22:5/18:1	834.4	41.32	20		18:0/18:2	744.7	44.03	1.03E4
	20:2/18:2	810.7	41.47	13		20:0/20:4	796.7	44.82	2.77E2
	20:3/16:0	784.7	41.49	2.31E2		18:0/18:1	746.7	45.78	2.08E4
	18:0/20:5	808.7	41.63	17					
	16:0/16:0	734.7	41.66	6.03E2					
	16:0/18:1	760.7	42.57	3.00E2					
	18:1/18:1	786.7	42.60	2.14E5					
	18:0/22:6	834.7	42.95	1.04E2					
	18:0/18:2	786.7	43.14	78					
	18:0/20:4	810.7	43.16	3					

(continues)

Journal of MASS SPECTROMETRY

Table 3. continued

Head	Species	*m/z*	t_r	Confidence score	Head	Species	*m/z*	t_r	Confidence score
	18:0/20:3	812.7	44.09	0.03					
	18:1/18:0	788.7	45.17	8.16E3					
	18:0/20:2	814.7	45.65	3	**LPE**	lyso/14:0	428.5	9.86	2.61E3
	20:0/20:0	846.9	51.49	11		16:0/lyso	454.4	10.24	6.22E3
						18:1/lyso	480.4	10.46	2.77E3
LPC	14:0/lyso	468.3	9.23	53		18:0/lyso	482.4	11.28	5.16E4
	16:0/lyso	496.5	9.95	69		lyso/18:0	482.4	11.29	8.01E4
	lyso/16:0	496.5	9.97	3		lyso/18:1	481.6	14.88	2.82E2
	18:1/lyso	522.3	10.33	1.08E5		lyso/20:1	509.7	17.91	2.13E2

with the developed software, as marked in bold in Table 3a. Another useful feature of the software is the capability of loading several LC–MS–MS data files for simultaneous processing. With this feature, a minimum number of multiple hits for an identified molecule could be assigned such that only molecules found from more than a certain number of different LC–ESI–MS–MS runs would be reported as probable candidates.

CONCLUSION

This report presents a computer-based algorithm that can identify various PLs and LPLs using nLC–ESI–MS–MS analysis. Because manual interpretation of CID spectra for each PL class is extremely laborious and time consuming, it is beneficial to have an automated identification process for lipidomic studies, especially if there are a large number of samples to be analyzed. The developed software is flexible in producing library files for different PL head groups containing customized fragment ion patterns specific to the MS instrument. It provides rapid and accurate identification of different PL classes with a superior capability for differentiating PL and LPL regioisomers. The program also offers a useful means of filtering search results through a user-friendly interface. This program can be further integrated to accommodate MS spectra obtained from various types of mass spectrometers with the proper adaptation of different fragment ion patterns.

Acknowledgements

This study was supported by Grant NRF-2011-0016438 and in part by Grant NRF-2010-0002021 from the National Research Foundation of Korea.

Supporting Information

Supporting information may be found in the online version of this article.

REFERENCES

[1] J. F. H. M. Brouwers, E. A. A. M. Vernooij, A. G. M. Tielens, L. M. G. Van Golde. Rapid separation and identification of phosphatidylethanolamine molecular species. *J. Lipid Res.* **1999**, *40*, 164.
[2] M. M. Wright, A. G. Howe, V. Zaremberg. Cell membranes and apoptosis: role of cardiolipin, phosphatidylcholine, and anticancer lipid analogues. *Biochem. Cell Biol.* **2004**, *82*, 18.
[3] H.-Y. Kim, T.-C. L. Wang, Y.-C. Ma. Liquid chromatography/mass spectrometry of phospholipids using electrospray ionization. *Anal. Chem.* **1994**, *66*, 3977.
[4] M. Koivusalo, P. Haimi, L. Heikinheimo, R. Kostiainen, P. Somerharju. Quantitative determination of phospholipid compositions by ESI-MS: effects of acyl chain length, unsaturation, and lipid concentration on instrument response. *J. Lipid Res.* **2001**, *42*, 663.
[5] X. Han, J. Yang, H. Cheng, H. Ye, R. W. Gross. Toward fingerprinting cellular lipidomes directly from biological samples by two-dimensional electrospray ionization mass spectrometry. *Anal. Biochem.* **2004**, *330*, 317.
[6] S. J. Blanksby, T. W. Mitchell. Advances in mass spectrometry for lipidomics. *Annu. Rev. Anal. Chem.* **2010**, *3*, 433.
[7] X. Han, R. W. Gross. Shotgun lipidomics: electrospray ionization mass spectrometric analysis and quantitation of cellular lipidomes directly from crude extracts of biological samples. *Mass Spectrom. Rev.* **2005**, *24*, 367.
[8] D. Schwudke, J. Oegema, L. Burton, E. Entchev, J. T. Hannich, C. S. Ejsing, T. Kurzchalia, A. Shevchenko. Lipid profiling by multiple precursor and neutral loss scanning driven by the data-dependent acquisition. *Anal. Chem.* **2005**, *78*, 585.
[9] R. Taguchi, J. Hayakawa, Y. Takeuchi, M. Ishida. Two-dimensional analysis of phospholipids by capillary liquid chromatography/electrospray ionization mass spectrometry. *J. Mass Spectrom.* **2000**, *35*, 953.
[10] G. Isaac, D. Bylund, J.-E. Månsson, K. E. Markides, J. Bergquist. Analysis of phosphatidylcholine and sphingomyelin molecular species from brain extracts using capillary liquid chromatography electrospray ionization mass spectrometry. *J. Neurosci. Methods* **2003**, *128*, 111.
[11] R. Taguchi, T. Houjou, H. Nakanishi, T. Yamazaki, M. Ishida, M. Imagawa, T. Shimizu. Focused lipidomics by tandem mass spectrometry. *J. Chromatogr. B* **2005**, *823*, 26.
[12] D. Y. Bang, E. j. Ahn, M. H. Moon. Shotgun analysis of phospholipids from mouse liver and brain by nanoflow liquid chromatography/tandem mass spectrometry. *J. Chromatogr. B* **2007**, *852*, 268.
[13] D. Y. Bang, D. Kang, M. H. Moon. Nanoflow liquid chromatography-tandem mass spectrometry for the characterization of intact phosphatidylcholines from soybean, bovine brain, and liver. *J. Chromatogr. A* **2006**, *1104*, 222.
[14] E. J. Ahn, H. Kim, B. C. Chung, M. H. Moon. Quantitative analysis of phosphatidylcholine in rat liver tissue by nanoflow liquid chromatography/tandem mass spectrometry. *J. Sep. Sci.* **2007**, *30*, 2598.
[15] P. Haimi, A. Uphoff, M. Hermansson, P. Somerharju. Software tools for analysis of mass spectrometric lipidome data. *Anal. Chem.* **2006**, *78*, 8324.
[16] C. S. Ejsing, E. Duchoslav, J. Sampaio, K. Simons, R. Bonner, C. Thiele, K. Ekroos, A. Shevchenko. Automated identification and quantification of glycerophospholipid molecular species by multiple precursor ion scanning. *Anal. Chem.* **2006**, *78*, 6202.
[17] M. Sud, E. Fahy, D. Cotter, A. Brown, E. A. Dennis, C. K. Glass, A. H. Merrill Jr, R. C. Murphy, C. R. H. Raetz, D. W. Russell, S. Subramaniam. LMSD: LIPID MAPS structure database. *Nucleic Acids Res.* **2007**, *35*, D527.
[18] H. Song, F. F. Hsu, J. Ladenson, J. Turk. Algorithm for processing raw mass spectrometric data to identify and quantitate complex lipid molecular species in mixtures by data-dependent scanning and fragment ion database searching. *J. Am. Soc. Mass Spectrom.* **2007**, *18*, 1848.
[19] H. Song, J. Ladenson, J. Turk. Algorithms for automatic processing of data from mass spectrometric analyses of lipids. *J. Chromatogr. B* **2009**, *877*, 2847.

[20] G. Hübner, C. Crone, B. Lindner. LipID – a software tool for automated assignment of lipids inmass spectra. *J. Mass Spectrom.* **2009**, *44*, 1676.
[21] R. Herzog, D. Schwudke, K. Schuhmann, J. L. Sampaio, S. R. Bornstein, M. Schroeder, A. Shevchenko. A novel informatics concept for high-throughput shotgun lipidomics based on the molecular fragmentation query language. *Genome Biol.* **2011**, *12*, R8.
[22] R. Herzog, K. Schuhmann, D. Schwudke, J. L. Sampaio, S. R. Bornstein, M. Schroeder, A. Shevchenko. LipidXplorer: a software for consensual cross-platform lipidomics. *PLoS One* **2012**, *7*, e29851.
[23] H. Kim, H. K. Min, G. Kong, M. H. Moon. Quantitative analysis of phosphatidylcholines and phosphatidylethanolamines in urine of patients with breast cancer by nanoflow liquid chromatography/ tandem mass spectrometry. *Anal. Bioanal. Chem.* **2009**, *393*, 1649.
[24] J. Y. Lee, H. K. Min, D. Choi, M. H. Moon. Profiling of phospholipids in lipoproteins by multiplexed hollow fiber flow field-flow fractionation and nanoflow liquid chromatography-tandem mass spectrometry. *J. Chromatogr. A* **2010**, *1217*, 1660.
[25] F.-F. Hsu, J. Turk. Studies on phosphatidylserine by tandem quadrupole and multiple stage quadrupole ion-trap mass spectrometry with electrospray ionization: structural characterization and the Fragmentation Processes. *J. Am. Soc. Mass Spectrom.* **2005**, *16*, 1510.
[26] H. K. Min, S. Lim, B. C. Chung, M. H. Moon. Shotgun lipidomics for candidate biomarkers of urinary phospholipids in prostate cancer. *Anal. Bioanal. Chem.* **2011**, *399*, 823.
[27] J. Y. Lee, H. K. Min, M. H. Moon. Simultaneous profiling of lysophospholipids and phospholipids from human plasma by nanoflow liquid chromatography-tandem mass spectrometry. *Anal. Bioanal. Chem.* **2011**, *400*, 2953.