This is a pre-print version associated to paper with doi: 10.1021/jasms.1c00008

LIPIC: an automated workflow to account for isotopologues-related interferences in electrospray ionization high resolution mass spectra of phospholipids

Andrea Castellaneta[1], Ilario Losito[1,2]*, Davide Coniglio[1], Beniamino Leoni[3], Pietro Santamaria[3], Maria A. Di Noia[4], Luigi Palmieri[4], Cosima D. Calvano[2,5], Tommaso R.I. Cataldi[1,2]

[1]Dipartimento di Chimica, [2]Centro Interdipartimentale SMART, [3]Dipartimento di Scienze Agro Ambientali e Territoriali, [4]Dipartimento di Bioscienze, Biotecnologie e Biofarmaceutica, [5]Dipartimento di Farmacia-Scienze del Farmaco,
Università degli Studi di Bari "Aldo Moro", via Orabona 4, 70126 Bari, Italy

ABSTRACT: In the last decade hydrophilic interaction liquid chromatography (HILIC) has emerged as an efficient alternative to reversed-phase chromatography (RPC) for the analysis of phospholipid (PL) mixtures based on mass spectrometric detection. Since the separation of PL by HILIC is chiefly based on their headgroup, the mass spectrum of each class can be obtained by spectral averaging under the corresponding HILIC band. Using experimental *m/z* values resulting from high mass resolution/accuracy instruments the sum compositions of PL in a specific class can be thus inferred but partial overlapping may occur between signals related to the M + 0 isotopologue of one species and the M+2/M+4 isotopologues of species having one/two more C=C bonds in their chemical structures. Here, an automated workflow, named *LIPIC (Lipid Isotopic Pattern Interference Correction)*, is proposed to account for such interferences. Starting from the experimentally verified assumption that peaks in isotope patterns are Gaussian, *LIPIC* predicts, as a function of *m/z* ratio, signal intensities due to M+2 and M+4 isotopologues of species with one or two more C=C bonds than the target one and calculates the corrected intensity for the M+0 isotopologue of the latter. Thanks to an iterative procedure, the suggested algorithm compensates also for slight shifts occurring between experimental and theoretical *m/z* ratios related to isotopologue peaks. Examples of applications to simulated and experimental mass spectra of two PL classes, *i.e.*, phosphatidylcholines (PC) and cardiolipins (CL), emphasize the increased extent of correction at the increase of molecular masses of involved species.

KEYWORDS: phospholipids, isotopologue interference, hydrophilic interaction liquid chromatography, electrospray ionization mass spectrometry

INTRODUCTION

After early studies dating back to the 1990s [1–3], the key role played by mass spectrometry with ElectroSpray Ionization (ESI) in the systematic characterization of lipids in complex biological matrices was recognized at the beginning of the new millennium[4]. Indeed, the application of ESI enabled to overcome the main limitations (*e.g.*, low sensitivity, presence of matrix ions and significant in-source fragmentation) related to previously adopted ionization approaches, like Fast Atom Bombardment (FAB), in the analysis of glycerolipids (GL), glycerophospholipids (GPL) and sphingomyelins (SM)[4,5]. In the subsequent two decades ESI-MS has become a pillar of *lipidomics*[6], contributing to increase its importance among *omics*[7,8]. In fact, lipidomics has proved a powerful approach to face the challenges posed by the inherent structural variability of lipids[7], a key feature that has actually led to a more refined classification of these compounds[9]. More recently, the integration of high-resolution mass spectrometry (HRMS) has allowed a better separation in the *m/z* domain between quasi-isobaric compounds resulting from multiple possible combinations of building blocks (side chains, polar head), typical of some of the aforementioned lipid classes[10].

Since early applications, a rapid characterization of complex lipid mixtures has been obtained by direct-infusion ESI-MS, contributing to the so-called *shotgun lipidomics*[3]. However, the need to reduce the competition for ionization, separate isomeric compounds and get a more reliable identification of lipid species has promoted the integration between ESI-MS and liquid chromatography[11,12]. Specifically, reversed-phase chromatography (RPC) and hydrophilic interaction liquid chromatography (HILIC) have emerged as the most effective separation techniques[13]. Amphiphilic lipid molecules, such as GL, GPL, and SM, are separated as a function of their hydrophobic moieties, namely fatty acyl or alkyl chains linked to the glycerol backbone, when RPC is

adopted. In the case of HILIC, the retention on the polar stationary phase involves the headgroup linked to glycerol, thus resulting in a very efficient class-related separation[10,12,14–16]. In the last years, this feature has been extensively exploited also in our laboratory for the characterization of complex lipid mixtures in clinical and food lipidomics, based on the ESI-mediated coupling between HILIC and low- or high-resolution mass spectrometry[17–26].

As shown in these works, one of the most intriguing aspects of PL analysis by HILIC-ESI-MS is that the ESI–MS spectrum averaged in the usually narrow retention time interval pertaining to a specific lipid class in HILIC can be considered a snapshot, even quantitative of the class. If ionization yields for compounds belonging to that class are proved to be comparable, as we found, for example, for phosphatidylcholines (PC), in one of our previous studies[18], an estimate of the concentration ratio between two lipids of a specific class can be retrieved directly from the ratio of the respective MS responses. Alternatively, calibration of MS responses based on standards can finally lead to concentration estimates. In any case, a reliable evaluation of MS response is required, and it can be initially obtained as a chromatographic peak area calculated from the Extracted Ion Current (EIC) chromatogram referred to each compound (see, for example, Refs. 18, 21, 23, 24). Nonetheless, an equivalent but less time-consuming approach is represented by retrieving the intensity observed for the ion related to each compound in the mass spectrum averaged over the lipid class retention time interval. Specifically, the signal intensity related to the major (M+0) isotopologue of the ion detected for each species can be considered and then a "Type I" correction[27], *i.e.*, an evaluation including also the contribution of the other isotopologues of that ion, can be performed. In fact, the isotopic pattern can slightly change, especially for compounds of the same class with quite different molecular masses, thus the contribution of the major isotopologue to the isotope pattern can be variable[28,29].

However, when lipids like GL, GPL and SM are involved, the evaluation of a correct MS response has also to account for another potential complication. Indeed, signals of ions related to compounds whose molecular masses differ from each other by 2 Da, as a result of the addition/subtraction of a C=C bond along one of their side chains, are often detected in MS spectra. This situation leads to the so-called "double bond overlapping effect"[30] or "Type II" isotopic effect[27], consisting in the overlap between peaks due to the M+0 isotopologue of a species and the M+2 isotopologue of another species bearing one further C=C bond. The interference becomes relevant in the presence of a partial chromatographic co-elution between such couples of species, which is not rare in the case of HILIC separations.

When using mass spectrometers with low resolving power, signals due to the M+0 isotopologue and to interfering isotopologues appear as perfectly overlapped and the intensity of the overall peak can be considered equal to the sum of each contribution. Therefore, the real M+0 intensity can be retrieved by signal subtraction, as described by Han *et al.*[29]. However, when operating under high mass resolving power conditions, like those provided by orbital traps (*i.e.*, about 75000 at *m/z* ratios between 700 and 800, typical of GPL and SM), signals related to the M+0 and to interfering isotopologues are not perfectly overlapped anymore, thus the described correction becomes no longer applicable. On the other hand, peaks are not completely separated too, and their misaligned overlap causes peak broadening and shift, reducing the accuracy needed for a reliable lipid identification through *m/z* database search[30]. The problem becomes even worse in the case of lipids with relatively high molecular weights (higher than 1000), like GPL known as cardiolipins (CL), for which even the mutual interference between the M+0 isotopologue of a specific lipid ion and the M+4 one related to a compound with two less C=C bonds in its structure should be accounted for[31].

To the best of our knowledge, common programs adopted for deisotoping are unable to face this complex problem. A novel algorithm has thus been developed in our laboratory to perform a systematic correction of those signal interferences, including the ones due to M+4 isotopologues, through a "cascade" workflow. Indeed, calculation starts from the signal intensity of the M+0 isotopologue exhibiting the lowest *m/z* ratio (thus not subjected to any interference) in the spectrum averaged for a specific PL class and keeps into account the functional form of signals observed in mass spectra. In the present paper, the main steps of the algorithm and its translation into a new function of the *R* (v. 3.6.3) operating language, named *LIPIC (Lipid Isotopic Pattern Interference Correction)*, including an interactive graphical window based on the *Plotly R Open-Source Graphing Library*[32], will be described. To emphasize the reliability and the extent of correction, applications to simulated and experimental spectra of mixtures of glycerophospholipids from selected classes, namely, phosphatidylcholines and cardiolipins will be reported.

EXPERIMENTAL

Chemicals

LC-MS grade acetonitrile, water and methanol, used for HILIC-ESI-MS separations, HPLC grade methanol and chloroform, used for lipid extraction, and ammonium acetate (reagent grade), used as mobile phase additive, were purchased from Sigma-Aldrich (Milan, Italy). Tetra-myristoyl-(1',3'-bis[1,2-dimyristoyl-sn-glycero-3-phospho]-glycerol) and tetra-oleoyl- (1',3'-bis[1,2-dioleoyl-sn-glycero-3-phospho]-glycerol), used as cardiolipin standards, were purchased from Avanti Polar Lipids (Alabaster, AL, USA).

Lipid extraction

High resolution mass spectra of phosphatidylcholine (PC) and cardiolipin (CL) mixtures, exploited to assess the performance of the *LIPIC* program on positive or negative polarity spectra, respectively, were retrieved from HILIC-ESI-MS datasets referred to lipid extracts of a flax microgreen sample, in the first case, and of yeast (*Saccharomyces cerevisiae*) mitochondria, in the second case. Details on the preparation of the two types of samples are provided in Section S1 of the Supporting Information. Lipid extraction was performed in both cases using the Bligh & Dyer protocol[33], appropriately adapted to the two types of samples. Specifically, 0.2 g of lyophilized flax microgreen were re-hydrated

by suspension into 1 mL of LC-MS water and then subjected to manual homogenization into a Potter-Elvehjem homogenizer. The resulting homogenate was suspended into 3 mL of a methanol/chloroform 2:1 (v/v) mixture, keep quiescent for 60 min in the dark and then centrifuged at 4500 g for 10 min in a glass tube, to separate the solid component. Afterwards, a 1 mL volume of the supernatant was withdrawn, and 1.25 mL of water and 1.25 mL of chloroform were added. The resulting mixture was vortexed and then centrifuged at 4500 g for 10 min to facilitate phase separation. The chloroform-rich phase was finally withdrawn from the bottom of the tube and subjected to solvent evaporation under a nitrogen flux; the resulting residue was then re-dissolved into 1 mL of a methanol/chloroform 2:1 (v/v) mixture. The solution was stored in a screw cap glass vial with headspace saturated with nitrogen (to minimize lipid oxidation) and stored at -20°C until the HILIC-ESI-MS analysis was performed. In the case of yeast mitochondria, the organelle pellet was suspended into 1 mL of LC-MS water and then 1 mL of chloroform and 2 mL of methanol were added. After centrifugation at 4500 g for 10 min 1 mL of the supernatant was withdrawn and the lipid extraction proceeded as described before for flax microgreen.

LC-MS instrumentation and operating conditions

An apparatus consisting in an Ultimate 3000 HPLC quaternary chromatograph coupled to a Q-Exactive hybrid quadrupole-Orbitrap mass spectrometer equipped with a HESI source (Thermo Fisher, West Palm Beach, CA, USA) was adopted for LC-MS analyses of lipid extracts. An Ascentis Express column (15 cm length, 2.1 mm internal diameter) packed with core-shell 2.7 μm silica particles (Supelco, Bellefonte, PA, USA) was adopted for HILIC separations, performed at a 300 μL/min flow, with 5 μL sample volumes injected using the autosampler of the Ultimate 3000 HPLC apparatus. The following multi-step binary elution gradient, based on an acetonitrile/water (97:3 v/v) mixture as phase A and a methanol/water (90:10 v/v) mixture as phase B, both containing ammonium acetate 2.5 mM, was finally adopted, after a series of preliminary experiments, for lipid separation: 0-10 min) linear increase of B from 2 to 20%; 10 -15 min) linear increase of B from 20 to 50%; 15 min-20 min) isocratic at 50% B; 20-25 min) return to 2% B; 25-30 min) column reconditioning at 2% B.

The operating HESI and ion optics parameters were set as follow: sheath gas flow rate 35 a.u.; auxiliary gas flow rate 15 a.u.; spray voltage -2.5 kV (negative polarity)/3.5 kV (positive polarity); capillary temperature 320 °C; S-lens RF level 100. The Q-Exactive spectrometer was operated at its maximum resolving power (140000 at *m/z* 200) and HRMS spectra were acquired in a 100-2000 *m/z* interval. The spectrometer was calibrated by HESI-FTMS analysis of calibration solutions provided by the instrument manufacturer for positive or negative polarity acquisitions, infused into the HESI interface at 5 μL/min flow rate. A mass accuracy better than 5 ppm was achieved.

Preliminary processing of lipid MS spectra using the *Alex*^{123} software

During this study, the preliminary processing of HRMS spectra was performed using *Alex*^{123} [34], a freely available software (www.mslipidomics.info) designed for the semi-untargeted analysis of high-resolution shotgun lipidomics data. As recently demonstrated by Ventura *et al.* [35,36], *Alex*^{123} can also handle tab-separated value (TSV) .txt files related to high resolution mass spectra averaged under HILIC chromatographic bands corresponding to specific lipid classes. Specifically, the *Alex*^{123} *extractor* module allows lipid identification (*i.e.*, chemical formula and sum composition determination) through comparison of accurate monoisotopic *m/z* ratios with exact values included in a database. In the present work, the *Alex*^{123} *extractor* output, a new TSV file reporting various information, was exploited to build the input for the *LIPIC* data frame (see next section), *i.e.*, sum composition, chemical formula, *m/z* ratio matching database values and the corresponding intensity. Since the *LIPIC* function requires *m/z* and intensity values to correspond to a peak maximum, and not to a peak shoulder, the *is apex* entry in the *Alex*^{123} *extractor* output TSV file had to be set on "YES". In fact, if this is not done, the user must manually determine the maximum peak coordinates before generating the data frame.

***LIPIC* input**

LIPIC is designed in R workspace and operates on a data frame which can be easily generated by the user as a *Microsoft® Excel®* file with the same structure shown in **Table 1**. In the latter, reporting examples referred to mixtures of CL or PC, the sum compositions and the chemical formulas of ions, in the present work identified through database search using *Alex*^{123} (but they can be obtained using any other approach, including manual annotation) are indicated in the first two columns. The corresponding measured *m/z* ratios and intensity values are reported in the third and in the fifth column, respectively. The fourth column is related to values of the absolute error (*dmz*), *i.e.*, the difference between the measured and the theoretical *m/z* value, initially set to 0 for every element of the fourth column. As discussed in the next section, an estimate of the *dmz* values for each peak is finally provided by *LIPIC*. As shown in **Table 1**, all elements in the data frame must be sorted following the ascending order of the *m/z* values. This is a mandatory

Table 1. Input data frame reporting the *Alex*^{123} *Extractor* output entries used for the *LIPIC* isotope interference correction on simulated spectra obtained for mixtures of cardiolipins (top) and phosphatidylcholines (bottom) after fixing the respective spectral abundances. Here, the "Measured *m/z*" values retrieved by the *Alex*^{123} *Extractor* tool were modified to simulate the contribution of experimental fluctuations in *m/z* ratios (see the Results and discussion section for details).

Sum Composition	Formula (M)	Measured *m/z* ($[M-2H]^{2-}$)	*dmz*	Absolute Intensity
CL 72:6	C81H144O17P2	725.4947	0	185298
CL 72:5	C81H146O17P2	726.5032	0	971557
CL 72:4	C81H148O17P2	727.5090	0	430176
CL 72:3	C81H150O17P2	728.5187	0	940245
CL 72:2	C81H152O17P2	729.5280	0	914815
CL 72:1	C81H154O17P2	730.5324	0	300167
CL 72:0	C81H156O17P2	731.5430	0	26307

Sum Composition	Formula (M)	Measured *m/z* ($[M+H]^+$)	*dmz*	Absolute Intensity
PC 36:6	C44H77NO8P	778.5370	0	283131
PC 36:5	C44H79NO8P	780.5532	0	1420138
PC 36:4	C44H81NO8P	782.5685	0	194221
PC 36:3	C44H83NO8P	784.5837	0	1415877
PC 36:2	C44H85NO8P	786.5990	0	1015926
PC 36:1	C44H87NO8P	788.6167	0	130185
PC 36:0	C44H89NO8P	790.6333	0	3889

Table 2. Output data frames generated by the last *LIPIC_IT* iteration on data described in **Table 1**. Columns labelled as *Type II* and *Type I* report peak intensities referred, respectively, to the M+0 isotopologue of a lipid ion after removing the interference due to the M+2/M+4 isotopologues of lipid ions including 1/2 more C=C bonds, and to the entire isotopic pattern of the ion related to a specific lipid, after correcting for that interference. *Lipid Class Adj-RA/UnAdj-RA* represent percent spectral abundances in the respective class obtained for the considered lipids after performing/not performing Type II correction; *Most Abundant Lipid Adj-RA/UnAdj-RA* represent spectral abundances for lipid ions obtained after performing/not performing Type II correction, expressed as percentages, with the spectral abundance of the prevailing ion corresponding to 100%.

Sum Composition	Formula	Measured m/z	d(m/z)	Measured Intensity	Type II	Type I	Lipid Class Adj-RA (%)	Lipid Class UnAdj-RA (%)	Most Abundant Lipid Adj-RA (%)	Most Abundant Lipid UnAdj-RA (%)
CL 72:6	C81H144O17P2	725.4947	0.0002	185298	185298	469284	6.74	4.91	19.91	19.07
CL 72:5	C81H146O17P2	726.5032	0.0007	971557	930287	2356579	33.86	25.77	99.99	100.00
CL 72:4	C81H148O17P2	727.5090	0.0035	430176	46298	117307	1.69	11.41	4.98	44.29
CL 72:3	C81H150O17P2	728.5187	0.0012	940245	929948	2356800	33.86	24.95	100.00	96.82
CL 72:2	C81H152O17P2	729.5280	0.0035	914815	647451	1641242	23.58	24.28	69.64	94.22
CL 72:1	C81H154O17P2	730.5324	0.0040	300167	1579	4004	0.06	7.97	0.17	30.92
CL 72:0	C81H156O17P2	731.5430	0.0105	26307	5612	14232	0.20	0.70	0.60	2.71
Sum Composition	**Formula**	**Measured m/z**	**d(m/z)**	**Measured Intensity**	**Type II**	**Type I**	**Lipid Class Adj-RA (%)**	**Lipid Class UnAdj-RA (%)**	**Most Abundant Lipid Adj-RA (%)**	**Most Abundant Lipid UnAdj-RA (%)**
PC 36:6	C44H77NO8P	778.5370	-0.0012	283131	283131	469267	6.76	6.34	19.93	19.93
PC 36:5	C44H79NO8P	780.5532	-0.0010	1420138	1420507	2354913	33.93	31.81	100.00	100.00
PC 36:4	C44H81NO8P	782.5685	0.0074	194221	69645	115484	1.66	4.35	4.90	13.68
PC 36:3	C44H83NO8P	784.5837	-0.0015	1415877	1414785	2346504	33.81	31.73	99.64	99.75
PC 36:2	C44H85NO8P	786.5990	-0.0011	1015926	992899	1647159	23.73	22.77	69.95	71.59
PC 36:1	C44H87NO8P	788.6167	0.0096	130185	-2881	0	0.00	2.92	0.00	9.18
PC 36:0	C44H89NO8P	790.6333	0.0012	3889	3938	6536	0.09	0.09	0.28	0.27

requirement for the *LIPIC* approach to isotopic correction (*vide infra*). *LIPIC* requires additional input information, which must be specified using the following syntax in the *R* workspace:

```
Lipic(data,z,Res, t=0.1,span=25,decimals=2,xseq=0.001)
```

In particular, the data frame name (*data*), the charge state of ions related to the lipid class (*z*) and the operative resolving power (*Res*) of the mass spectrometer need to be specified. The remaining parameters (*t*, *span*, *decimals*, *xseq*) are set as default values, but, if needed, can be modified. In particular, *t* is a parameter required by the function *Isopattern* included in the *envipat R* package[37]. *Isopattern* calculates the isotope fine structure for a given chemical formula or a set of chemical formulas; here, *t* represents the probability threshold below which isotopic peaks can be omitted. Note that *span*, *decimals* and *xseq* are parameters useful for the calculation of the *dmz* values, as described later in the text.

LIPIC output

After running the *LIPIC* function using the RStudio interface (v. 1.2.5033), two interactive plots and a new data frame, which represents an extension of the input data frame, are obtained. As shown in **Table 2**, the output data frame includes six additional columns, named "Type II", "Type I", "Lipid Class Adj-RA (%)", "Lipid Class UnAdj-RA (%)", "Most Abundant Lipid Adj-RA (%)" and "Most Abundant Lipid Un-Adj-RA (%)".

Specifically, the "Type II" column lists spectral intensity values as obtained after correcting for the eventual interference related to the M+2/M+4 isotopologues of lipid species with one/ two additional C=C bonds ("Type II" isotopic effect). Values in the "Type I" column correspond to intensities related to the entire isotopic patterns of lipid ions. They can thus account for monoisotopic peak intensity reduction due to the increase of higher isotopologues abundance, when, for example, the number of C-atoms in the lipid molecule increases ("Type I" isotopic effect). Columns named "Lipid Class Adj-RA (%)" and "Most Abundant Lipid Adj-RA (%)" refer to different expressions of the intensity of the M+0 peak as obtained after both Type I and Type II corrections. Conversely, columns named "Lipid Class UnAdj-RA (%)" and "Most Abundant Lipid UnAdj-RA (%)" refer to intensity values obtained after only Type I correction (*vide infra*). The values in both "Lipid Class Adj-RA (%)" and "Lipid Class UnAdj-RA (%)" columns are expressed as percentages with respect to the sum of intensity values calculated for all the lipid species in the data frame. Values reported in "Most Abundant Lipid Adj-RA (%)" and "Most Abundant Lipid Un-Adj-RA (%)" columns represent spectral intensities corrected/not corrected for Type II effects and expressed as percentages after assigning 100% to the maximum calculated intensity in the corresponding column.

It is worth noting that *dmz* values estimated at the end of calculations are reported in the fourth column of the output data frame. If differing from the *dmz* values in the input data frame, as it occurs after the first calculation, these can be substituted for the latter, following an iterative process until convergence is reached; this procedure can be automated. To this purpose, a new *R* function named *LIPIC_IT* (*ITerative LIPIC*) was defined. It runs *LIPIC* and

automatically updates the values in the fourth column of the input data frame, until the *dmz* values calculated in a new iteration equal those obtained in the previous one. *LIPIC_IT* requires the same input syntax described before for *LIPIC*.

As previously mentioned, in addition to the output data frame described so far, the *LIPIC* function returns two interactive plots generated through the *plot_ly* function. The first plot is automatically printed into the *RStudio "viewer"* window, and shows, for each lipid sum composition, the values listed in the "Lipid Class Adj-RA (%)"and "Lipid Class UnAdj-RA (%)" columns, using green and blue bars, respectively (*vide infra*). The second plot is generated if prompted by user, who is asked to indicate the .txt file path related to the mass spectrum used for lipid identification through *Alex*[123] software. Thereafter, a plot is obtained containing three overlaid mass spectra: the real mass spectrum and simulated spectra obtained from intensities emerging, respectively, from Type I correction (unadjusted measured intensities) and from both Type II and Type I corrections (adjusted measured intensities). Details and examples of the plots generated after *LIPIC_IT*-based calculations will be provided in the Results and discussion section.

Simulation of the mass spectrum for a lipid mixture

As previously mentioned, the simulation of mass spectra was very important to assess the reliability of the correction performed by *LIPIC* and played an important role in the calculation of the *dmz* values. Simulation assumed peaks in mass spectra to be Gaussian (*vide infra*), thus, each of them was mathematically described by Eq. (1), where σ is the peak standard deviation, x_{max} is the position of the peak maximum on the *m/z* axis (*i.e.*, the theoretical *m/z* ratio for a given isotopologue), I_{max} is the peak maximum intensity, and *I(x)* is the intensity registered at a certain *x* value on the *m/z* axis.

$$I(x) = I_{max} \cdot \exp\left[\frac{(x - x_{max})^2}{2\sigma^2}\right] \qquad (1)$$

The peak standard deviation, σ, can be easily calculated using its mathematical relationship with the peak *Full Width at Half Maximum* (FWHM), which, in turn, is related to the mass spectrometer resolving power, *RP*, since this is meant as the ratio between the *m/z* ratio at the peak maximum and FWHM. Consequently, if the operating resolving power is known, σ can be calculated as shown by Eq. (2):

$$\sigma^2 = \left[\frac{m/z}{RP\sqrt{2\ln 2}}\right]^2 \qquad (2)$$

Considering the L+k isotopologue of the ion of a generic lipid with nominal molecular mass L as the one implying a nominal shift of k Da in that mass, its maximum peak intensity, I_{max}*(L+k)*, can be calculated as the product between the percentual relative abundance of that isotopologue, *%RA(L+k)*, in turn calculated considering as 100% the abundance of the main isotopologue for L (*i.e.*, L+0), and the total intensity distributed over the isotope pattern, $I_{tot}(L)$, as shown in Eq. (3).

$$I_{max}(L+k) = I_{tot}(L) \cdot \frac{\%RA(L+k)}{\sum_{n=0}^{N-1} \%RA(L+n)} \qquad (3)$$

Here, N is the total number of possible isotopologues for L, thus also including the L+0 one, $\sum_{n=0}^{N-1} \%RA(L+n)$ is the sum of the percent relative abundances of all the L isotopologues with respect to the most abundant one, and *%RA(L+k)* represents the *%RA* related to the L+k isotopologue. All the *%RA* values can be easily retrieved using the already cited *isopattern* function, while $I_{tot}(L)$ is calculated by *LIPIC* as described in the first paragraph of the Results and discussion section.

The entire isotope pattern for a lipid with nominal mass L, meant as the function correlating signal intensity I_L to *m/z* ratio (x), can now be mathematically expressed as shown in Eq. (4), *i.e.*, as a sum of gaussian functions related to each isotopologue.

$$I_L(x) = \sum_{i=0}^{N-1} I_{tot}(L) \cdot \frac{\%RA(L+i)}{\sum_{n=0}^{N-1} \%RA(L+n)} \cdot \exp\left\{\frac{[x - x_{max}(L+i)]^2}{2\sigma^2(L+i)}\right\} \qquad (4)$$

Here $x_{max}(L+i)$ and $\sigma^2(L+i)$ represent, respectively, the *m/z* ratio of the maximum and the squared standard deviation of the gaussian peak corresponding to the L+i isotopologue.

Considering a mixture of M lipids with respective nominal molecular masses $L_1, L_2, \ldots, L_j, \ldots, L_M$ and starting from Eq. (4), the corresponding simulated spectrum can be mathematically expressed by $I_{spec}(x)$ in Eq. (5).

$$I_{spec}(x) = \sum_{j=1}^{M} \sum_{i=0}^{N-1} I_{tot}(L_j) \cdot \frac{\%RA(L_j+i)}{\sum_{n=0}^{N-1} \%RA(L_j+\mathrm{n})} \cdot \exp\left\{\frac{\left[x - x_{max}(L_j+i)\right]^2}{2\sigma^2(L_j+i)}\right\} \qquad (5)$$

Eq. (5) is implemented in *LIPIC* for mass spectrum simulation. In particular, if the simulation is meant to account for the isotope correction (which is necessary when the simulated spectrum is used for *dmz* values determination), "adjusted" (*i.e.*, obtained after both Type I and Type II corrections) $I_{tot}(L_j)$ intensity values are used; otherwise, the "unadjusted" (*i.e.*, obtained after only Type I correction) values are implemented. The corrected/uncorrected simulated spectrum is calculated according to the case (*vide infra*). It is worth noting that the simulated spectrum calculation through Eq. (5) is based on equally spaced *x* values along the *m/z* axis, differing each other by the *xseq* value specified in the function input. Specifically, the lower end of the *x* values interval adopted for simulation corresponds to the monoisotopic *m/z* ratio of the lipid having the lowest molecular mass in the mixture reduced by two units, the higher end is equal to the monoisotopic *m/z* ratio of the heaviest isotopologue of the heaviest lipid in the mixture augmented by two units. Once determined, $I_{spec}(x)$ values are normalized and expressed as percentage abundance with respect to the maximum calculated value, like it is usually done by programs performing mass spectra acquisitions. The simulated mass spectrum can be visualized

through the *plot_ly* function, plotting normalized $I_{spec}(x)$ values against *x* values.

RESULTS AND DISCUSSION

The *LIPIC* theoretical approach to the correction of isotopic interference

As anticipated before, the *LIPIC* approach is based on the assumption that mass spectrometric peaks have a Gaussian shape[30]. This assumption was proved in the present work by using a Gaussian curve to fit the peak obtained for the main isotopologue (M+0) of the deprotonated form of cardiolipin annotated as CL 56:0 (1',3'-bis[1,2-dimyristoyl-sn-glycero-3-phospho]-glycerol), *i.e.*, a cardiolipin bearing four identical 14:0 acyl chains, whose exact m/z ratio is 1239.8393 ($[C_{65}H_{125}O_{17}P_2]^-$). The *m/z* and absolute intensity values used to generate the peak were retrieved from a mass spectrum acquired by the Q Exactive mass spectrometer set at its maximum resolving power (*i.e.*, 140000 at *m/z* 200). As shown in **Figure S1** of the Supporting Information, a Gaussian function was suitable to fit experimental data (R^2 = 0.99815) and a similar result (R^2 = 0.99952) was obtained also for the M+1 isotopologue of the same ion (exact *m/z* 1240.8427).

To explain interference effects that can be accounted for by *LIPIC* let us consider L and M ions related to two different compounds belonging to the same lipid class, thus sharing the same ionization mechanism and charge. Suppose that M differs from L in its structure only for the presence on an additional C=C bond. As an example, M and L could refer to doubly charged negative ions of CL 72:2 and CL 72:1, respectively. Let us now consider that the resolving power available at *m/z* ratios corresponding to those ions (*ca.* 730), when the Q-Exactive spectrometer is operated at its maximum resolving power (140000 at *m/z* 200) is *ca.* 75000, due to the inherent decrease with m/z ratio occurring with orbital traps [$RP \propto 1/(m/z)^{1/2}$]. As shown in **Figure 1**, at a resolving power *RP* = 75000 the convolution of peaks due to all isotopologues of the M ion characterized by a +2 nominal shift (designated with M+2), *i.e.*, the red curve in the figure, partially overlaps with the peak related to monoisotopic L (green curve in the figure). The overlap (drawn in blue in the figure) appears thus broader than the latter, with the apex on a *m/z* value (m/z_{meas}) shifted with

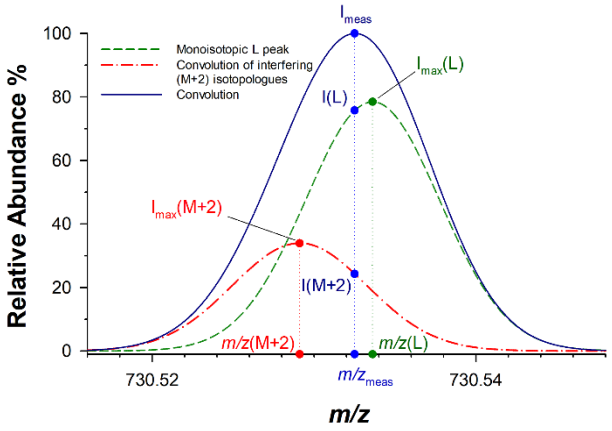

Figure 1. Simulated reconstruction of the peak shift and peak broadening effect occurring for the monoisotopic CL 72:1 $[M-2H]^{2-}$ ion signal (L) caused by its overlap with the contribution of all the M+2 isotopologues of the CL 72:1 $[M-2H]^{2-}$ ion (M). In the simulation, the operating resolving power was set as 75000, which is that available at *m/z* ratios of *ca.* 730 when the Q-Exactive spectrometer is operated at its maximum resolving power (140000 at *m/z* 200).

respect to the theoretical *m/z* ratio of the L isotopologue $m/z\,(L)$. As described by Wang *et al.*[30], this mass shift is expected to increase at the increase of the molar ratio between M and L (and, therefore, between signal intensities due, respectively, to their M+2 and L isotopologues).

Depending on the contribution due to the M+2 peak, a certain degree of asymmetry on the left side is also expected for the resulting peak. The aim of the *LIPIC* algorithm is the calculation of the correct L peak intensity $I_{max}(L)$ starting from the coordinate values (m/z_{meas} and I_{meas}) of the peak obtained experimentally, potentially affected by the presence of the M+2 peak. As shown by Eq. (6), since peaks can be considered Gaussian-shaped, $I_{max}(L)$can be retrieved from the $I(L)$ intensity value assumed for a *m/z* ratio equal to m/z_{meas} using the Gaussian function describing the monoisotopic L peak. Here, σ_L represents the L peak standard deviation, which can be determined using equation (2).

$$I_{max}(L) = I(L) \cdot \exp\left[\frac{\left(m/z_{meas} - m/z\,(L)\right)^2}{2\sigma_L^2}\right] \qquad (6)$$

As shown in **Figure 1**, $I(L)$ can be, in turn, obtained as the difference between I_{meas} and the $I(M+2)$ intensity value of the M+2 Gaussian interfering peak measured at m/z_{meas}. For the sake of simplicity, let us assume that the latter refers to only one of the possible M+2 isotopologues, namely, the most abundant one, containing two ^{13}C atoms. Under this approximation, it is possible to calculate the $I(M+2)$ value as shown by Eq. (7). In this equation σ_{M+2} is the standard deviation of the M+2 peak, $I_{max}(M)$ is the measured intensity of the monoisotopic peak for the M ion, assumed to be free from isotopic interferences due to species with a lower molecular mass, and $\%RA(M+2)$ corresponds to the relative abundance of the M+2 isotopologue, calculated

after setting as 100% the one referred to the most abundant isotopologue in the isotope pattern.

$$I(M+2) = \frac{\%RA(M+2)}{100} \cdot I_{max}(M) \cdot \exp\left[-\frac{\left(m/z_{meas} - m/z\,(M+2)\right)^2}{2\sigma_{M+2}^2}\right] \quad (7)$$

The $\%RA(M+2)$, as well as the relative abundance of every isotopologue in the isotope pattern, can be calculated using the already cited *isopattern* function included in the *envipat R* package[37].

It is worth noting that, for a more accurate calculation, $I(L)$ should be determined considering the interfering contribution of all the possible M+2 isotopologues calculated over the threshold set for the *isopattern* function. In particular, the correction offered by the *LIPIC* algorithm accounts for all the M+2 isotopologues whose m/z ratio differs from that of L for less than $0.4/z$ units, where z is the ion charge value set into the *LIPIC* input. The correction accounts also for the interference eventually introduced by a further lipid ion, M', including two additional C=C bonds in its structure with respect to L. In this case, when operating at a resolving power around 75000, like that pertaining to m/z ratios shown in **Figure 1**, also the signal due to convoluted M'+4 isotopologues would partially overlap with the monoisotopic L+0 peak. Supposing that k is the number of M'+4 isotopologues which differ from the L m/z ratio for 0.4/z units or less, the more general expression of the correction offered by the *LIPIC* function can be summarized in Eq. (8). In this case, each interfering $I(M'+4)_i$ contribution to the I_{meas} value can be calculated adapting Eq. (7), *i.e.*, by substituting the $M+2$ notation with the $M'+4$ one. Here, $\%RA(M'+4)$ would represent the percent relative abundance of a specific M'+4 isotopologue referred to the most abundant M' isotopologue.

$$I_{max}(L) = \left[I_{meas} - \sum_{i=1}^{h} I(M+2)_i - \sum_{i=1}^{k} I(M'+4)_i\right] \cdot \cdot \exp\left[\frac{\left(m/z_{meas} - m/z\,(L)\right)^2}{2\sigma_L^2}\right] \quad (8)$$

In turn, L could interfere, through its L+2 and L+4 isotopologues, with the monoisotopic peaks of ions related to lipids bearing one or two C=C less than L, respectively. In such a case the interfering contributions are calculated by the *LIPIC* function by adapting Eq. (7) as described for $I(M'+4)_i$ and substituting the previously calculated $I_{max}(L)$ value to the $I_{max}(M)$. Thus, the Type II interference correction enabled by *LIPIC* follows a "cascade" workflow, starting from the lipid in the mixture having the lowest molecular mass (which is thus free from isotopic interference) and then applying the correction to lipids with increasingly higher m/z values. Note that all species of a phospholipid class share the same number of heteroatoms (O, N, P, S), since these are mainly present in the polar head, which defines the lipid class itself, and, in the case of O, also on the glycerol backbone, which is obviously common to all species. Consequently, compounds involved in interfering effects would differ just for the number of C and H atoms on their side acyl chains. Because of its low natural abundance (approximately 0.02%) the deuterium influence on the molecule isotopic pattern can be assumed to be negligible with respect to that of the ^{13}C (natural abundance 1.1%). The probability of finding a certain number of ^{13}C atoms increases with the total number of C atoms in the molecule, therefore the monoisotopic peak intensity of a lipid ion decreases when the number of C atoms increases ("Type I" isotopic effect). Accordingly, if an equimolar mixture of two compounds belonging to the same lipid class is ionized in ESI, the ion corresponding to the compound with the lower number of C atoms would show a higher monoisotopic peak intensity. As a consequence, the total intensity $I_{tot}(L)$ distributed over the isotope pattern should be used for quantification purposes, rather than $I_{max}(L)$, in order to account for "Type I" isotopic effect. As shown in Eq. (3'), it is possible to calculate $I_{tot}(L)$ from $I_{max}(L)$ simply by rearranging the terms in equation (3).

$$I_{tot}(L) = I_{max}(L) \cdot \frac{\sum_{i=0}^{N} \%RA(L+i)}{\%RA(L+0)} \quad (3')$$

Intensities like that expressed by Eq. (3') were involved in calculations leading to values reported in columns named *Lipid Class Adj-RA* and *Most Abundant Lipid Adj-RA* of the output data frame, described in the Experimental section and shown in **Table 2**. Conversely, values reported in columns named *Lipid Class UnAdj-RA* and *Most Abundant Lipid UnAdj-RA* were obtained starting from $I_{tot}(L)$ values in which "Type II" correction was purposely ignored, *i.e.*, values determined as shown in Eq. (9).

$$I_{tot}(L) = I_{meas} \cdot \frac{\sum_{i=0}^{N} \%RA(L+i)}{\%RA(L+0)} \quad (9)$$

It is worth noting that the theoretical approach to isotope interference correction described so far is based on the consideration of "theoretical" mass spectra, *i.e.*, spectra in which exact m/z ratios occur for all involved signals. Indeed, it does not consider the influence of small but not negligible variations occurring on experimental m/z ratios (and, henceforth, in mass spectrometer accuracy) observed for L+0, M+2 and M'+4 isotopologues. In order to adapt the described calculations to the real-life scenario, we made the assumption that all the contributions to a convoluted peak in experimental mass spectra, like those emphasized in **Figure 1**, are shifted of an equal *dmz* amount, which corresponds to the difference between the peak position (m/z_{meas}) in the experimental and in the simulated mass spectrum, *i.e.*, the mass spectrum that would have been acquired if the mass analyzer accuracy had been 0 ppm. The uniformity of the experimental m/z shift is reasonable, since, as evidenced in **Figure 1**, the m/z interval involved is narrower than 0.02 units. Consequently, equations (2), (7) and (8) shown before need to be modified as shown by equations (2'), (7') and (8'). This modification allows the correction operated by *LIPIC* to account for the experimental error in m/z determination.

$$\sigma^2 = \left[\frac{m/z + dmz}{RP\sqrt{2\ln 2}}\right]^2 \qquad (2')$$

$$I(M+2) = \frac{\%RA(M+2)}{100} \cdot I_{max}(M) \cdot \cdot \exp\left[-\frac{\left(m/z_{meas} - m/z\,(M+2) - dmz\right)^2}{2\sigma_{M+2}^2}\right] \qquad (7')$$

$$I_{max}(L) = \left[I_{meas} - \sum_{i=1}^{h} I(M+2)_i - \sum_{i=1}^{k} I(M'+4)_i\right] \cdot \exp\left[\frac{\left(m/z_{meas} - m/z\,(L) - dmz\right)^2}{2\sigma_L^2}\right] \qquad (8')$$

Initially, *dmz* values are obviously unknown, thus, when generating the input *LIPIC* data frame, the user is required to set the corresponding values as equal to 0. When operating on this input data frame, *LIPIC* runs a preliminary "Type II" correction using the experimental values of m/z_{meas} and the theoretical *m/z* values for each lipid ion and for all its possible interfering isotopologues. Referring to **Figure 1**, this corresponds to operating on shifted L and M+2 *m/z* values with respect to those that would stem from the deconvolution of the observed peak. It is like considering the red and green curves in the figure as slightly displaced with respect to the blue one, that is the experimental curve. As a result, this preliminary correction cannot be considered fully reliable. Nevertheless, to obtain a first estimate of *dmz* values, intensity values obtained after "Type II" and "Type I" corrections are used to generate the simulated spectrum. Clearly, the simulated lipid monoisotopic peaks affected by isotopic interference would differ in shape and height from those found in the experimental spectrum. However, the difference between the position of peaks in the real and in the simulated spectrum returns a first estimate of actual *dmz* values. Specifically, the *LIPIC* algorithm performs a peak-picking operation on the simulated mass spectrum, employing the *pick.peaks* function of the *ChemometricsWithR R* package[38]. The *pick.peaks* function divides the simulated spectrum in segments in which local maxima are found. The width of the segments is determined by the *span* parameter value specified in input to *LIPIC*. Once each peak location is determined, the corresponding *m/z* value is compared with the ion *m/z* values reported in the "Measured *m/z*" column in the input data frame. The comparison is made by rounding the *m/z* values to the decimal position specified by the *decimals* parameter value in the *LIPIC* input. A match is found if the rounded *m/z* value for the peak position in the simulated spectrum equals a rounded *m/z* value in the input data frame. Afterwards, the corresponding *dmz* value is calculated by subtracting the peak position *m/z* value in the simulated spectrum to the corresponding measured *m/z* value in the input data frame. The estimated *dmz* values are thus reported into the output data frame; then *LIPIC* performs a second isotope correction implementing these values in equations (2'), (7') and (8'). The user can already assess the reliability of the latter correction by prompting the graphical comparison between the real spectrum and the related simulated one. Nevertheless, the output *dmz* values can be replaced to those in the input data frame, to run a more accurate isotope correction, repeating this process until convergence is reached. As previously mentioned, this process can be automatically being performed using the *LIPIC_IT* function.

As described in the next sections, given a certain profile of compunds in a specific phospholipid class, the *LIPIC* graphical output automatically obtained at the end of the procedure enables a rapid evaluation of the extent of correction.

Assessment of LIPIC performance on simulated mass spectra

To evaluate the incidence of the isotopic interference correction on the reliability of spectral abundances provided by *LIPIC*, appropriately simulated spectra referred to mixtures of 5 CL or 5 PC, in both cases including species differing each other systematically for one C=C bond, were considered. Sum compositions, formulas and spectral relative abundances set for the simulation are reported in **Table 3**. As discussed before, those abundances are expected to be similar to relative concentrations if ESI ionization yields are similar for the compounds involved. Simulated spectra were generated using the Thermo Scientific™ Xcalibur™ (v. 4.1) *Qual Browser* "isotope simulation" tool, referring to the lipid relative spectral abundances indicated in the table and setting a resolving power of 75000. CL and PC were supposed to be ionized as $[M\text{-}2H]^{2-}$ and $[M+H]^{+}$ ions, respectively. The spectrum was exported as a TSV .txt file and subsequently analyzed using the *Alex*^{123} *Extractor* tool for a first detection of phospholipids. As a result, the lists of sum compositions reported in the first columns of **Tables 1** and **2** was inferred, indicating that the *Alex*^{123} *Extractor* software identified, in each case, two additional lipids, CL 72:1/CL 72:0 and PC 36:1/PC 36:0, that should not have been found. Such a fault was triggered by the need of maintaining a relatively large accuracy window in filtering when using the *Alex*^{123} *Extractor*. In fact, if the accuracy window had been reduced, some peaks might have been missed, due to "Type II" isotopic interference, that can lead to a remarkable apparent shift in *m/z* ratios[30]. If, for example, CL 72:4 and PC 36:4 are considered, the abundance of the corresponding ions in the simulated spectrum was significantly lower than the corresponding species having an additional C=C bond (*i.e.*, CL 72:5 and PC 36:5). As a consequence, M+0 peaks for CL 72:4 and PC 36:4 were overwhelmed by peaks due to the M+2 isotopologues of CL 72:5 and PC 36:5, respectively, with which they were convoluted. The *m/z* ratios of the convoluted peaks, 727.5059 and 782.5608, were shifted with respect to the theoretical monoisotopic *m/z* ratios, *i.e.*, 727.5101 and 782.5694, respectively. The apparent mass accuracies in the two cases were thus -5.77 and -10.99 ppm, much higher (in absolute value) than those usually available with the Q-Exactive spectrometer, *i.e.*, better than 2 ppm. As it will be shown later, the erroneous attributions cited before could be eliminated after using *LIPIC* for isotope interferences correction.

It is worth noting that simulated mass spectra generated by the *Qual Browser* "isotope simulation" tool are based on theoretical *m/z* values; thus, they cannot consider the contribution of fluctuations occurring on the experimentally measured *m/z* values. We thus decided to also simulate this

contribution, assuming it to be varying between -3 and 3 ppm and modifying the peak *m/z* values returned by the *Alex[123] Extractor* tool accordingly. Briefly, for each lipid class, the *Microsoft™ Excel* "*RANDBETWEEN*" function was used to generate 5 random numbers between -30 and 30. The corresponding values were divided by 10 in order to obtain the accuracy values rounded to the first decimal place; the differences (*dmz*) between the observed and theoretical *m/z* ratios were subsequently calculated and added to the *m/z* values returned by the *Alex[123] Extractor* tool, thus obtaining the "Measured *m/z*" values indicated in the input data frames reported in **Table 2**. The *LIPIC_IT* function was then run using the following code lines for cardiolipins and phosphatidylcholines, respectively:

```
Lipic_it(data,-2,75000,t=0.001,decimals=1)

Lipic_it(data,+1,75000,t=0.001,decimals=1)
```

Here, the parameter "data" symbolically indicates the name given to the input data frame in the *R* workspace. From the output data frames shown in **Table 2** it is apparent that corrected spectral relative abundances, reported in the "Most Abundant Lipid Adj-RA (%)" column, were in excellent accordance with those adopted for the mass spectra simulation (see **Table 3**), thus confirming the correction reliability. As for additional lipids apparently recognized by the *Alex[123] Extractor* tool, *i.e.*, CL 72:1/CL72:0 and PC 36:1/PC 36:0, their corrected relative abundances were equal or close to zero, according to the case. This result clearly confirmed that peaks related to those species were actually referred to M+2 and M+4 isotopologues of abundant species having one or two additional C=C bonds on their structure. The incidence of the isotopic interference correction on calculated relative abundances for species in a PL class can be easily inferred also from the graphical output automatically provided by *LIPIC*, reported in **Figure 2**. The output is based on the coupling, for each of the species in a class, of green and blue bars, corresponding, respectively, to adjusted and unadjusted relative spectral abundances within the class. Consequently, the blue bars are prevailing (or they are even the only visible) in the case of species overestimated in the absence of correction. On the other hand, the green bars are found to prevail for species whose relative abundance is artificially underestimated when the correction is not performed, simply because other species are overestimated, thus making their percentages lower. As apparent from **Figure 2**, and as expected, the isotopic interference correction operated by *LIPIC* is more relevant in the case of cardiolipins, compared to phosphatidylcholines. This effect is clearly due to the stronger interference of M+2 and M+4 isotopologues for cardiolipins, in turn related to the larger number of C atoms included in CL molecular structures.

Assessment of *LIPIC* performance on real mass spectra

Once the reliability of the correction enabled by *LIPIC* was assessed on simulated mass spectra of CL and PC, the program was run on real ESI-FTMS mass spectra, for which no preliminary information on relative abundances was known. In particular, spectra were obtained by spectral averaging under bands detected for CL and PC in HILIC chromatograms referred tolipid extracts of yeast mitochondria and flax microgreens, respectively. As illustrated in **Figure 3**, the focus was put on $[M-2H]^{2-}$ ions in the case of cardiolipins, being their peaks more intense than those of $[M-H]^-$ ones, in accordance with previous studies[39,40].

Table 3. Relative percent spectral abundances employed to simulate mass spectra referring to a mixture of five cardiolipins (A) and five phosphatidylcholines (B). For each lipid species the sum composition and the chemical formula are also reported.

Sum Composition	Formula (M)	Relative Abundance (%)
CL 72:6	$C_{81}H_{144}O_{17}P_2$	20
CL 72:5	$C_{81}H_{146}O_{17}P_2$	100
CL 72:4	$C_{81}H_{148}O_{17}P_2$	5
CL 72:3	$C_{81}H_{150}O_{17}P_2$	100
CL 72:2	$C_{81}H_{152}O_{17}P_2$	70
PC 36:6	$C_{44}H_{77}NO_8P$	20
PC 36:5	$C_{44}H_{79}NO_8P$	100
PC 36:4	$C_{44}H_{81}NO_8P$	5
PC 36:3	$C_{44}H_{83}NO_8P$	100
PC 36:2	$C_{44}H_{85}NO_8P$	70

The occurrence of clusters of peaks usually spaced by 14 *m/z* units, apparent in the spectrum, reflects the presence of CL differing by 28 Da, *i.e.*, two CH_2 moieties, in their nominal molecular masses. The spectral expansion, referred to the base peak of the spectrum, emphasizes the contemporary presence of CL differing by one C=C bond (see peaks at *m/z* 699.480 and 700.484), and the relevant incidence of M+2 and M+4 isotopologues in the case of CL. The systematic occurrence of species differing by one C=C bond is even more evident in the case of phosphatidylcholines, for which $[M+H]^+$ ions, with M representing the zwitterionic form, provided the most intense signals, due to the stable positive charge located on the choline head.

Output data frames obtained after running the *LIPIC* function on both spectra are reported in **Table S1** of the Supporting Information, whereas the corresponding graphical outputs are shown in **Figure 4**. As expected, the extent of "Type II" correction was more relevant in the case of CL, with species 68:3, 70:3 and 72:3 resulting overestimated before the correction (see the prevalence of the blue bars over the green ones), due to the relevant abundance of species 68:4, 70:4 and 72:4, in accordance with the isotopic interference issue explained previously. On the other hand, the correction had generally a negligible effect in the case of PC, with two exceptions, since the correction emphasized that PC 34:1 and PC 36:1 were actually not present in the mixture. Their erroneous identification was due to the relevant incidence of the M+2 isotopologues of PC 34:2 (which is responsible for the most abundant PC ion population in the specific case) and PC 36:2.

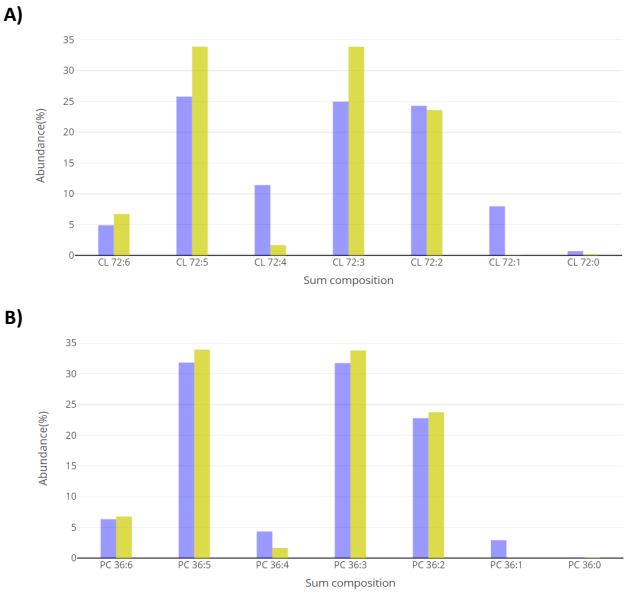

Figure 2. Snapshots of the *LIPIC* bar chart graphical output generated after the last *LIPIC_IT* iterative calculation based on data described in Tables 1 and 3. Charts with coupled bars associated to the class-related relative spectral abundances calculated from the simulated spectra of a mixture of 5 cardiolipins (A) and 5 phosphatidylcholines (B) are shown. The coupled bars refer to values calculated with (green) or without (blue) "Type II" correction. See text for details.

It is worth noting that these results can be inferred also from the *LIPIC* graphical interactive output, *i.e.*, the comparison between the real mass spectra and the simulated ones, with or without isotopic interference correction, reported for CL and PC in **Figures S2** and **S3** of the Supporting Information, respectively. Indeed, the spectrum simulated for CL without considering the correction differs from the real one in many peaks, whereas it can be compared well with it if "Type II" correction is performed. On the other hand, the difference appears much less relevant in the case of PC.

As apparent from the results reported in **Figure 4**, *LIPIC* was able to indicate that some signals assigned preliminarily as further lipids in a class (either automatically, as in our case, or manually) were artificially increased by the contribution of M+2/M+4 isotopologues related to intense signals due to very abundant lipids actually present in the class. In order to verify which extent of interference due to such lipids could be accounted for by the function, a specific, standard addition-like, experiment was devised. Specifically, after HILIC-ESI-FTMS analysis, the lipid extract whose CL and PC spectra are reported in **Figure 3** was divided into two aliquots, that were then spiked with standard CL 72:4 up to an added concentration of 1.8 and 2.6 mg/L, respectively.

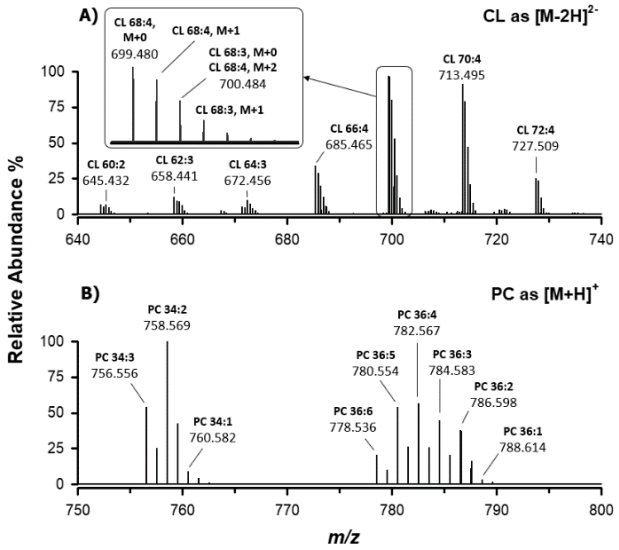

Figure 3. (A) ESI(-)-FTMS mass spectrum averaged under the CL chromatographic band obtained after HILIC separation performed on a yeast mitochondria lipid extract. The *m/z* values refer to the most intense peak for each CL cluster, *i.e.*, for each group of cardiolipins having the same total number of carbon atoms but differing for the number of C=C bonds. An expanded view of the most abundant cluster of CL signals is shown in the inset; (B) ESI(+)-FTMS mass spectrum averaged under the PC chromatographic band obtained after the HILIC separation performed on a flax microgreen lipid extract. The *m/z* values refer to all the identified PC.

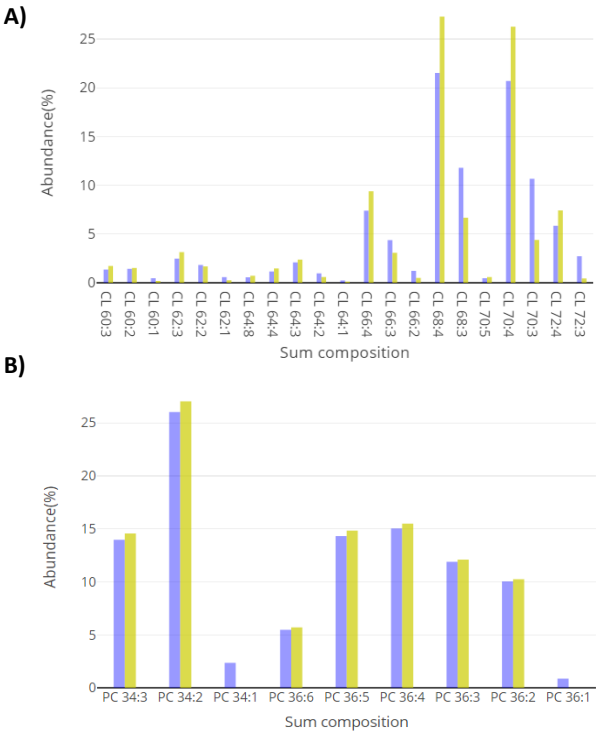

Figure 4. Snapshots of the *LIPIC* bar chart graphical output returned from the last *LIPIC_IT* iterative calculation on spectral data shown in **Figure 3**. Here, class-related lipid spectral

abundances calculated with (green) or without (blue) "Type II" correction refer to mixtures of cardiolipins (A) and phosphatidylcholines (B) in mitochondrial and flax microgreen lipid extracts, respectively. See text for details.

As a result, the spectral relative abundance of the *m/z* 727.509 signal, corresponding to the M+0 isotopologue of CL 72:4 $[M-2H]^{2-}$ ion, was raised from 25% to 50% and then to 60% (see **Figure S4** of the Supporting Information), considering as 100% that of CL 68:4. The effect of CL72:4 addition on the apparent abundance of CL72:3 was striking, since it was raised from 12.6 to 22.7 and then to 28.4%, due to the remarkable contribution of the M+2 isotopologue of CL 72:4. On the other hand, when Type II correction was performed by *LIPIC*, that apparent increase was almost completely accounted for, since the spectral abundance of CL 72:3 changed from the initial 1.7 to 1.8 and finally to 2.0%. Consequently, even in the presence of an isotopically interfering lipid having a spectral abundance more than 30 times higher than the actual abundance of the compound suffering from the interference, calculations made by the function were able to correct the latter efficiently.

It is worth noting that tests described to verify *LIPIC* performance started from input data considering all peaks potentially related to M+0 isotopologues included in clusters due to CL having the same total number of carbon atoms but a different number of C=C bonds. Indeed, if one of those peaks was not assigned preliminarily, *e.g.*, due to the lack of the corresponding compound in lipid databases, and thus not considered in the input, the correction made by *LIPIC* would be inevitably influenced. Nonetheless, a preliminary check of the input assignments should easily avoid this drawback.

A final test of *LIPIC* performance on real samples was made by comparing the outputs obtained when acquiring mass spectra of yeast mitochondria CL at two different resolving powers enabled by the adopted quadrupole-Orbitrap spectrometer, *i.e.*, 140000 and 70000. As evidenced in Table S2 of the Supporting Information, differences observed in corrected spectral abundances were generally negligible, thus confirming that *LIPIC* can be reliably used also when relatively lower resolving powers are adopted.

Conclusions

A function named *LIPIC*, running in *R* environment, was developed to correct "Type I" and "Type II" isotopic effects, the latter being generated by the overlap between peaks referred to the M+0 isotopologue of a specific compound in a phospholipid (PL) class and to the M+2 and M+4 isotopologues of other species, having one or two further C=C bonds in their molecular structure, respectively. The function was applied both to simulated and to real high-resolution/accuracy mass spectra referred to two representative phospholipid classes, obtained upon spectral averaging under the corresponding bands in the HILIC chromatograms of complex lipid extracts. *LIPIC* proved able to minimize interference effects and, as expected, the correction was more relevant in the case of PL with relatively high molecular masses, like cardiolipins, for which the incidence of M+2 and M+4 isotopologues is more remarkable. Even though the extent of interference was less pronounced in the case of PL with lower molecular masses, like phosphatydilcholines, *LIPIC* unveiled mistakes that may occur with the automatic assignment of peaks, when signals due the M+2 isotopologues of abundant compounds in a class are erroneously related to compounds with 2 more H atoms in their molecular structure. Considering that the coupling between HILIC and high resolution/accuracy mass spectrometry is becoming increasingly important in the analysis of complex phospholipid mixtures, *LIPIC* may represent a useful complementary tool when a reliable profile of compounds in a specific PL class is searched for, regardless of their respective abundances. The program will be thus made freely available to researchers interested in applying it to phospholipid classes-related mass spectra.

ASSOCIATED CONTENT

Supporting Information

1) **Section S1**. Description of sample preparation (flax microgreen, yeast mitochondria)
2) **Table S1**. Output data frames obtained by *LIPIC* after iterative runs on real mass spectral data of cardiolipins and phosphatidylcholines.
3) **Table S2**. Output data frames obtained by *LIPIC* after iterative runs on real mass spectral data of cardiolipins acquired after setting resolving powers as 140000 and 70000.
4) **Figure S1**. Gaussian fitting of peaks due to the M+0 and M+1 isotopologues of the singly charged anion of CL 56:0.
5) **Figure S2**. Snapshots of the interactive *LIPIC* graphical output showing corrected/uncorrected simulated spectra along with the real mass spectrum for cardiolipins detected in the lipid extract of yeast mitochondria.
6) **Figure S3**. Snapshots of the interactive *LIPIC* graphical output showing the corrected/uncorrected simulated spectra along with the real mass spectrum for phosphatidylcholines detected in the lipid extract of flax microgreen.
7) **Figure S4**. Comparison between the ESI-(-)-FTMS spectrum shown in Figure 3 for CL in a yeast mitochondria lipid extract and those obtained from the same extract after fortification with 1.8 and 2.6 mg/L CL 72:4, respectively.

File type: PDF

AUTHOR INFORMATION

Corresponding Author

* Ilario Losito, University of Bari "Aldo Moro", Department of Chemistry, e-mail: ilario.losito@uniba.it.

Author Contributions

The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

ACKNOWLEDGMENT

This work was supported by the following grants: PONa3_00395/1 "Bioscienze & Salute (B&H)" and PRIN 20172T2MHH, both funded by the Italian *Ministero per l'Istruzione, l'Università e la Ricerca* (MIUR).

REFERENCES

1. Han, X., Gross, R.W.: Electrospray ionization mass spectroscopic analysis of human erythrocyte plasma membrane phospholipids. *Proc. Natl. Acad. Sci.* **91**, 10635–10639 (1994). https://doi.org/10.1073/pnas.91.22.10635
2. Kerwin, J.L., Tuininga, A.R., Ericsson, L.H.: Identification of molecular species of glycerophospholipids and sphingomyelin using electrospray mass spectrometry. *J. Lipid Res.* **35**, 1102–14 (1994)
3. Brugger, B., Erben, G., Sandhoff, R., Wieland, F.T., Lehmann, W.D.: Quantitative analysis of biological membrane lipids at the low picomole level by nano-electrospray ionization tandem mass spectrometry. *Proc. Natl. Acad. Sci.* **94**, 2339–2344 (1997). https://doi.org/10.1073/pnas.94.6.2339
4. Han, X., Gross, R.W.: Global analyses of cellular lipidomes directly from crude extracts of biological samples by ESI mass spectrometry : a bridge to lipidomics. *J. Lipid Res.* **44**, 1071–1079 (2003). https://doi.org/10.1194/jlr.R300004-JLR200
5. Blanksby, S.J., Mitchell, T.W.: Advances in Mass Spectrometry for Lipidomics. *Ann. Rev. Anal. Chem.* **3**, 433–465 (2010). https://doi.org/10.1146/annurev.anchem.111808.073705
6. Han, X.: Lipidomics for studying metabolism. *Nat. Rev. Endocrinol.* **12**, 668–679 (2016). https://doi.org/10.1038/nrendo.2016.98
7. Wang, J., Wang, C., Han, X.: Tutorial on lipidomics. *Anal. Chim. Acta.* **1061**, 28–41 (2019). https://doi.org/10.1016/j.aca.2019.01.043
8. Rustam, Y.H., Reid, G.E.: Analytical Challenges and Recent Advances in Mass Spectrometry Based Lipidomics, *Anal. Chem.* **90**, 374-397 (2018). http://pubs.acs.org/doi/10.1021/acs.analchem. 7b04836
9. Fahy, E., Subramaniam, S., Brown, H.A., Glass, C.K., Merrill, A.H., Murphy, R.C., Raetz, C.R.H., Russell, D.W., Seyama, Y., Shaw, W., Shimizu, T., Spener, F., Van Meer, G., VanNieuwenhze, M.S., White, S.H., Witztum, J.L., Dennis, E.A.: A comprehensive classification system for lipids. *J. Lipid Res.* **46**, 839–861 (2005). https://doi.org/10.1194/jlr.E400004-JLR200
10. Züllig, T., Köfeler, H.C.: High Resolution Mass Spectrometry in Lipidomics. *Mass Spectrom. Rev.* **21627** (2020). https://doi.org/10.1002/mas.21627
11. Cajka, T., Fiehn, O.: Comprehensive analysis of lipids in biological systems by liquid chromatography-mass spectrometry. *TrAC - Trends Anal. Chem.* **61**, 192–206 (2014). https://doi.org/10.1016/j.trac.2014.04.017
12. Zheng, L., T'Kind, R., Decuypere, S., von Freyend, S.J., Coombs, G.H., Watson, D.G.: Profiling of lipids in Leishmania donovani using hydrophilic interaction chromatography in combination with Fourier transform mass spectrometry. *Rapid Commun. Mass Spectrom.* **24**, 2074–2082 (2010). https://doi.org/10.1002/ rcm.4618
13. Holčapek, M., Liebisch, G., Ekroos, K.: Lipidomic Analysis. *Anal. Chem.* **90**, 4249–4257 (2018). https://doi.org/10.1021/acs.analchem. 7b05395
14. Zhu, C., Dane, A., Spijksma, G., Wang, M., van der Greef, J., Luo, G., Hankemeier, T., Vreeken, R.J.: An efficient hydrophilic interaction liquid chromatography separation of 7 phospholipid classes based on a diol column. *J. Chromatogr. A*. **1220**, 26–34 (2012). https://doi.org/10.1016/j.chroma.2011.11.034
15. Schwalbe-Herrmann, M., Willmann, J., Leibfritz, D.: Separation of phospholipid classes by hydrophilic interaction chromatography detected by electrospray ionization mass spectrometry. *J. Chromatogr. A.* **1217**, 5179–5183 (2010). https://doi.org/10.1016/ j.chroma.2010.05.014
16. Okazaki, Y., Kamide, Y., Hirai, M.Y., Saito, K.: Plant lipidomics based on hydrophilic interaction chromatography coupled to ion trap time-of-flight mass spectrometry. *Metabolomics*. **9**, 121–131 (2013). https://doi.org/10.1007/s11306-011-0318-z
17. Losito, I., Patruno, R., Conte, E., Cataldi, T.R.I., Megli, F.M., Palmisano, F.: Phospholipidomics of human blood microparticles. *Anal. Chem.* **85**, 6405–6413 (2013). https://doi.org/10.1021/ ac400829r
18. Losito, I., Conte, E., Cataldi, T.R.I., Cioffi, N., Megli, F.M., Palmisano, F.: The phospholipidomic signatures of human blood microparticles, platelets and platelet-derived microparticles: A comparative HILIC-ESI-MS investigation. *Lipids*. **50**, 71–84 (2015). https://doi.org/10.1007/s11745-014-3975-7
19. Losito, I., Facchini, L., Diomede, S., Conte, E., Megli, F.M., Cataldi, T.R.I., Palmisano, F.: Hydrophilic interaction liquid chromatography–electrospray ionization-tandem mass spectrometry of a complex mixture of native and oxidized phospholipids. *J. Chromatogr. A.* **1422**, 194–205 (2015). https://doi.org/10.1016/j.chroma.2015.10.023
20. Granafei, S., Losito, I., Palmisano, F., Cataldi, T.R.I.: Identification of isobaric lyso-phosphatidylcholines in lipid extracts of gilthead sea bream (Sparus aurata) fillets by hydrophilic interaction liquid chromatography coupled to high-resolution Fourier-transform mass spectrometry. *Anal. Bioanal. Chem.* **407**, 6391–6404 (2015). https://doi.org/10.1007/s00216-015-8671-9
21. Facchini, L., Losito, I., Cianci, C., Cataldi, T.R.I., Palmisano, F.: Structural characterization and profiling of lyso-phospholipids in fresh and in thermally stressed mussels by hydrophilic interaction liquid chromatography-electrospray ionization-Fourier transform mass spectrometry. *Electrophoresis*. **37**, 1823–1838 (2016). https://doi.org/10.1002/elps.201500514
22. Facchini, L., Losito, I., Cataldi, T.R.I., Palmisano, F.: Ceramide lipids in alive and thermally stressed mussels: an investigation by hydrophilic interaction liquid chromatography-electrospray ionization Fourier-transform mass spectrometry. *J. Mass Spectrom.* **51**, 768–781 (2016). https://doi.org/10.1002/jms.3832
23. Losito, I., Facchini, L., Catucci, R., Calvano, C., Cataldi, T., Palmisano, F.: Tracing the Thermal History of Seafood Products through Lysophospholipid Analysis by Hydrophilic Interaction Liquid Chromatography–Electrospray Ionization Fourier Transform Mass Spectrometry. *Molecules*. **23**, 2212 (2018). https://doi.org/10.3390/molecules23092212
24. Facchini, L., Losito, I., Cataldi, T.R.I., Palmisano, F.: Seasonal variations in the profile of main phospholipids in Mytilus galloprovincialis mussels: A study by hydrophilic interaction liquid chromatography–electrospray ionization Fourier transform mass spectrometry. *J. Mass Spectrom.* **53**, 1–20 (2018). https://doi.org/10.1002/jms.4029
25. Bianco, M., Calvano, C.D., Huseynli, L., Ventura, G., Losito, I., Cataldi, T.R.I.: Identification and quantification of phospholipids in strawberry seeds and pulp (Fragaria × ananassa cv San Andreas) by liquid chromatography with electrospray ionization and tandem mass spectrometry. *J. Mass Spectrom.* **55**, e4523 (2020). https://doi.org/10.1002/jms.4523
26. Calvano, C.D., Bianco, M., Ventura, G., Losito, I., Palmisano, F., Cataldi, T.R.I.: Analysis of Phospholipids, Lysophospholipids, and Their Linked Fatty Acyl Chains in Yellow Lupin Seeds (Lupinus luteus L.) by Liquid Chromatography and Tandem Mass

Spectrometry. *Molecules*. **25**, 805 (2020). https://doi.org/10.3390/ molecules25040805
27. Han, X., Gross, R.W.: Quantitative Analysis and Molecular Species Fingerprinting of Triacylglyceride Molecular Species Directly from Lipid Extracts of Biological Samples by Electrospray Ionization Tandem Mass Spectrometry. *Anal. Biochem.* **295**, 88–100 (2001). https://doi.org/10.1006/abio.2001.517
28. Günther, E., Katussevani, B., Koal, T., Ramsay, S.L., Weinberger, K.M., Graber, A.: Isotope correction of mass spectrometry profiles. *Rapid Commun. Mass Spectrom.* **22**, 2248–2252 (2008). https://doi.org/10.1002/rcm.3591
29. Han, X., Yang, J., Yang, K., Zhongdan, Z., Abendschein, D.R., Gross, R.W.: Alterations in Myocardial Cardiolipin Content and Composition Occur at the Very Earliest Stages of Diabetes : A Shotgun Lipidomics Study. *Biochemistry*. **46**, 6417–6428 (2007). https://doi.org/10.1021/bi7004015
30. Wang, M., Huang, Y., Han, X.: Accurate mass searching of individual lipid species candidates from high-resolution mass spectra for shotgun lipidomics. *Rapid Commun. Mass Spectrom.* **28**, 2201–2210 (2014). https://doi.org/10.1002/rcm.7015
31. Scherer, M., Schmitz, G., Liebisch, G.: Simultaneous Quantification of Cardiolipin, Bis(monoacylglycero)phosphate and their Precursors by Hydrophilic Interaction LC–MS/MS Including Correction of Isotopic Overlap. *Anal. Chem.* **82**, 8794–8799 (2010). https://doi.org/10.1021/ac1021826
32. Plotly Technologies Inc.: *Collaborative data science*, https://plotly.com/ (2015)
33. Bligh, E.G., Dyer, W.J.: A rapid method of total lipid extraction and purification. *Can. J. Biochem. Physiol.* **37**, 911-917 (1959) https://doi.org/10.1139/o59-099
34. Husen, P., Tarasov, K., Katafiasz, M., Sokol, E., Vogt, J., Baumgart, J., Nitsch, R., Ekroos, K., Ejsing, C.S.: Analysis of Lipid Experiments (ALEX): A Software Framework for Analysis of High-Resolution Shotgun Lipidomics Data. *PLoS One*. 8, e79736 (2013). https://doi.org/10.1371/journal.pone.0079736
35. Ventura, G., Calvano, C.D., Porcelli, V., Palmieri, L., De Giacomo, A., Xu, Y., Goodacre, R., Palmisano, F., Cataldi, T.R.I.: Phospholipidomics of peripheral blood mononuclear cells (PBMCs): the tricky case of children with autism spectrum disorder (ASD) and their healthy siblings. *Anal. Bioanal. Chem.* (2020). https://doi.org/10.1007/s00216-020-02817-z
36. Ventura, G., Bianco, M., Calvano, C.D., Losito, I., Cataldi, T.R.I.: HILIC-ESI-FTMS with All Ion Fragmentation (AIF) Scans as a Tool for Fast Lipidome Investigations. *Molecules*. **25**, 1–16 (2020). https://doi.org/10.3390/molecules25102310
37. Loos, M., Gerber, C., Corona, F., Hollender, J., Singer, H.: Accelerated Isotope Fine Structure Calculation Using Pruned Transition Trees. *Anal. Chem.* **87**, 5738–5744 (2015). https://doi.org/10.1021/acs.analchem.5b00941
38. Wehrens, R.: *Chemometrics with R*. Springer Verlag, Berlin Heidelberg (2020)
39. Murphy, R.C.: *Tandem Mass Spectrometry of Lipids*. Royal Society of Chemistry, Cambridge (2014)
40. Hsu, F.F., Turk, J., Rhoades, E.R., Russell, D.G., Shi, Y., Groisman, E.A.: Structural characterization of cardiolipin by tandem quadrupole and multiple-stage quadrupole ion-trap mass spectrometry with electrospray ionization. *J. Am. Soc. Mass Spectrom.* **16**, 491–504 (2005). https://doi.org/10.1016/ j.jasms.2004. 12.015