Chemometrics and Intelligent Laboratory Systems 251 (2024) 105169

Contents lists available at ScienceDirect

Chemometrics and Intelligent Laboratory Systems

journal homepage: www.elsevier.com/locate/chemometrics

Lipid Quant 2.1: Open-source software for identification and quantification of lipids measured by lipid class separation QTOF high-resolution mass spectrometry methods

Michaela Chocholoušková [a,b,c,1], Gabriel Vivó-Truyols [d,1], Denise Wolrab [a], Robert Jirásko [a], Michela Antonelli [a], Ondřej Peterka [a], Zuzana Vaňková [a], Michal Holčapek [a,*]

[a] University of Pardubice, Faculty of Chemical Technology, Department of Analytical Chemistry, Studentská 573, 532 10, Pardubice, Czech Republic
[b] Singapore Lipidomics Incubator (SLING), Life Sciences Institute, National University of Singapore, 117456, Singapore
[c] Department of Medicine, Yong Loo Lin School of Medicine, National University of Singapore, 119228, Singapore
[d] Tecnometrix, Finca Mallaui, Ctra. Cala Turqueta s/n, 07760, Ciutadella De Menorca, Spain

ARTICLE INFO

Keywords:
Lipidomics
Lipid class separation
High-resolution mass spectrometry
Quantitation
Lipidomics software
Isotopic correction

ABSTRACT

LipidQuant 2.1 is a software written in Matlab, which is designed for the high-throughput processing of large lipidomic data sets measured by lipid class separation coupled with quadrupole time-of-flight (QTOF) high-resolution mass spectrometry (MS). The software enables the identification of lipid species based on defined mass accuracy. The main focus is on the right lipidomic quantitation using at least one internal standard per lipid class and the implementation of an automated procedure for Type I and Type II isotopic corrections necessary for the determination of accurate molar concentrations, which is not available for the majority of existing software solutions. LipidQuant 2.1 offers three options for peak assignment, visualization of the isotopic pattern, and automated calculation of *m/z* for various adduct ions. The initial lipidomic database covers 31 lipid classes with more than 2900 lipid species that occur primarily in the human lipidome, but users have the full flexibility to modify and extend the database according to their needs. All algorithms and the detailed user manual are provided. The reliability of LipidQuant 2.1 is demonstrated on a set of more than 250 biological samples measured by ultrahigh-performance supercritical liquid chromatography (UHPSFC) coupled with QTOF-MS.

1. Introduction

Cellular and subcellular membranes are formed primarily by lipids. Lipids are highly diverse molecules with hundreds of possible combinations of various backbones, polar head groups, and fatty acyls with different numbers of carbon atoms, double bonds, and bond types [1]. The lipid structure diversity requires various analytical approaches for lipidomic analysis, such as targeted/untargeted methods, lipid class separation, lipid species separation, direct infusion, low- and high-resolution mass spectrometry (MS), and various scan modes (full scan, SIM, MRM, MS/MS, DDA, DIA, *etc.*). These various platforms for lipidomic analyses produce different outputs of lipid information that need to be further processed and quantified. Although lipidomics belongs to the fastest growing omics field, there is still a lack of software for comprehensive lipidomic analysis, including identification, isotopic corrections, and quantification, together with the expectation that the software should include a user-friendly interface, automatization, and relatively short processing time without any particular expertise needed [2,3].

Recently, open-source tools offered a simplification of the processing of lipidomic data, mainly based on targeted and tandem mass spectrometry (MS/MS). Exemplary software, such as the LipidCreator [3], can define fragments, manage stable isotope labeling, optimize collision energies, and generate *in silico* spectral libraries. The LipidCreator is compatible with Skyline [4] for further spectral library matching between experimental and theoretical MS/MS spectra (generated by the LipidCreator), followed by possible quantitation. The LipidMatch tool [5] has been developed for lipid identification in high-resolution tandem MS/MS workflows. The annotation is based on library matching and uses rule-based identification, such as fragments, which must be

* Corresponding author.
E-mail address: Michal.Holcapek@upce.cz (M. Holčapek).
[1] The first two authors should be regarded as **Joint First Authors**.

https://doi.org/10.1016/j.chemolab.2024.105169
Received 22 February 2024; Received in revised form 16 June 2024; Accepted 20 June 2024
Available online 20 June 2024

observed in MS/MS spectra to maximize the correct annotation. LIQUID software [6] works similarly to LipidMatch to identify lipids in LC/MS/MS data. The LipidFinder [7] tool has been developed to search within three independent online databases to obtain alleged lipid identification and lipid class assignment based on the LIPID MAPS system. LipidXplorer [8] interprets lipidomic shotgun data in any mode (MS, PI, NL, data-dependent MS/MS) and low- or high-resolution MS according to resolution, noise threshold, and tolerance settings. ALEX software [9] has been established for the processing, management, and visualization of shotgun data in high-resolution MS. ALEX supports the identification and export of lipid species intensities from mass spectral data files for further use. The Lipid Data Analyzer (LDA) [10] has been created for the unsupervised identification and quantitation of lipids in LC/MS data. The LDA uses a 3D algorithm that confirms the peak boundaries in both *m/z* and retention time orders of measurement and the theoretical isotopic distribution of a molecule as the selection/exclusion criterion. The LDA also supports quantitation, visualization, and data export. The MZmine is mainly focused on LC/MS data, but now the platform supports hybrid datasets, such as gas chromatography-MS data, ion mobility spectrometry (IMS), and MS imagining [11]. This is a brief and unexhausted list of available lipidomic softwares published so far. However, most open source lipidomic softwares are mainly qualitative tools with no quantitative part or missing isotopic correction steps, such as Type II isotopic correction essential for the lipid class separation [12].

Type II isotopic correction is necessary because of the existence of interferences on the *m/z* axis. In lipidomics, molecules with a different number of double bonds may exist and coelute. When this happens, the coeluting isobaric ions cannot be distinguished by the mass spectrometer, resulting in one overlapped signal. The M+2 isotope of a lipid species overlays with the monoisotopic peak of another lipid species with more 2H, as it is the case for two lipid species differing in one double bond. This has consequences at the quantification level, since the measured *m/z* signal contributes to the isotopic patterns of different molecules. The Type I isotopic correction arises from the fact that the relative intensity of the base *m/z* peak decreases with the number of carbon atoms in the molecule (as the abundance of molecules with at least one ^{13}C increases with the number of carbon atoms). This means that the ratio between the concentration of the molecule and the total intensity of *m/z* is not constant, but decreases with the number of carbon atoms. The Type I isotopic correction can be easily addressed by considering the statistical abundance of isotopic peaks in a molecule [12].

Hydrophilic interaction liquid chromatography (HILIC) [13], normal phase liquid chromatography (NP-LC) [14], and supercritical fluid chromatography (SFC) [15] are well-recognized modes based on the interaction of polar head groups with polar groups in the stationary phase resulting in lipid class separation, the coelution of lipid species with the same polar head group. The advantage of these approaches is the coelution of exogenous internal standards (IS) with endogenous lipids belonging to the same lipid class, thus minimizing matrix effect differences within the lipid class [16]. Consequently, only a molecular level of lipid information, such as the number of carbon atoms and double bonds, can be determined in case of full scan mode of high-resolution MS [17].

LipidQuant 1.0 has been developed using the Excel as a simple tool for the automated processing of lipidomic data acquired by lipid class separation coupled with QTOF high-resolution MS (Full Scan). The software offers identification and quantitation, including Type II isotopic correction, but without Type I correction [18].

LipidQuant 2.1 brings a significantly upgraded version of the previous software. First, we have incorporated the isotopic calculation of Rockwood et al. [19] for accurate isotopic pattern calculation. Second, we have incorporated a new algorithm for Type I and Type II isotopic corrections, which is done more accurately in one step using the regression analysis. Third, the peak assignment has been improved, tackling the problem with the so called linear assignment algorithms. Finally, we also improved data visualization. These features were not present in the previous version.

2. Software BACKGROUND

2.1. Upgrade of LipidQuant 1.0

LipidQuant 1.0 is a simple Excel-based software written in the Visual Basic for Application (VBA) programming language to identify (based on mass accuracy and known lipid class) and quantify (including only Type II isotopic correction) lipidomic data obtained from methods based on the lipid class separation hyphenated with QTOF high-resolution MS, which provided reliable quantitation data in previous large scale studies [20,21]. LipidQuant 2.1 is written in Matlab (version: MATLAB 9.7.0.1471314 (R2019b) Update 7); users do not need any Matlab license) and works on a similar principle but with significant improvements and a compiled friendly graphical user interface. The first upgrade of the LipidQuant software is an isotopic calculation using the Rockwood method (the Mercury algorithm) [19]. This is a significant improvement compared to LipidQuant 1.0, in which external software was needed to accurately calculate the isotopic patterns.

Another upgrade is related to Type I and Type II isotopic corrections. Type I isotopic correction (*i.e.*, the correction applied once the isotopic abundance of each molecule is considered) is easy to correct. It depends only on the accuracy of the theoretical isotopic pattern calculation. However, the Type II isotopic correction can show an important problem with the error propagation, as the correction of the concentration of one molecule M depends on the accuracy of the concentration of the molecule with one additional double bond (M+2). This creates a chain of dependencies, causing the inaccuracy of lighter molecules to significantly affect the accuracy of heavier ones. At the extreme, the accuracy of the concentration of the molecule with a lower *m/z* value of the series is crucial, as the error in such concentration propagates to all others in the series. Traditionally, since the Type II isotopic correction is applied sequentially, the propagation of errors is not well balanced due to the dependencies explained above [12]. Moreover, errors in Type I isotopic correction of lighter molecules can indeed propagate to heavier molecules via Type II error miscorrection. With traditional approaches, Type I and Type II isotopic corrections cannot be generalized, and problems can arise with special cases, in which lipid series show molecules at M+1 and M+2 positions (in case of coelution of PC and SM lipid species). More complicated problems arise with interferences when ethers exist (the substitution of $-CH_2$ by –O and additional double bond results in the appearance of lipids differing by a small increment that only FT-MS instruments can resolve) or more rich isotopic patterns coexist. LipidQuant 2.1 tackles both Type I and Type II isotopic corrections using a more general technique, thus allowing a better balance in error propagation and a model that includes any generalization (*e.g.*, M+1 and M+2) corrections. See Section 2.2.5 for details of the mathematical operation.

The third upgrade of the software is an advanced peak assignment. The task of assigning *m/z* signals to molecules (or isotopes of molecules) is not trivial, even in the case of targeted peak assignment (*i.e.*, the experimental *m/z* values are checked against an existing lipid database). In such cases, the existence of isobaric or nearly isobaric compounds leads to ambiguities that must be resolved in a systematic manner. We tackle the problem of peak assignment as a linear mathematical assignment problem, using the Hungarian algorithm to solve it [22]. To the best of our knowledge, this is the first time that peak assignment has been solved using such a technique, so global (opposed to local) cost functions solve difficult peak assignment situations more efficiently. Such situations would be very difficult (if possible) to solve manually. We refer to Section 2.2.4. for more details.

2.2. Algorithm flow chart

The step-by-step flow of LipidQuant 2.1 software is visualized in Fig. 1 and described step-by-step in a ReadMe file (Supplementary Information) with illustrations. The workflow is divided into two parts. First, the generation of input data in.txt files and second, the data processing using LipidQuant 2.1.

2.2.1. Definition of lipid database

The first step of the software is a definition of the lipid database needed for the calculation of exact masses of lipid species and their adducts in both positive- and negative-ion modes. The lipid database can be prepared from scratch in the software or loaded using the Excel table according to the **ReadMe** file in the Supplementary Information. The example lipid database in Excel format is also provided in Supplementary Information and can be freely used. So far, the Excel lipid database includes 2907 lipid species belonging to 31 lipid classes, namely, phosphatidylcholines (PC), lysophosphatidylcholines (LPC), phosphatidylethanolamines (PE), lysophosphatidylethanolamines (LPE), phosphatidic acid (PA), lysophosphatidic acid (LPA), phosphatidylserines (PS), lysophosphatidylserines (LPS), phosphatidylglycerols (PG), lysophosphatidylglycerols (LPG), phosphatidylinositols (PI), lysophosphatidylinositols (LPI), bis[monoacylglycero]phosphate (BMP), sphingomyelins (SM), ceramides (Cer), hexosylceramides (HexCer), dihexosylceramides (Hex2Cer), trihexosylceramides (Hex3Cer), tetrahexosylceramides (Hex4Cer), sulfohexosylceramides (SHexCer), ceramide phosphoethanolamines (CerPE), monosialodihexosylganglioside (GM3), sphingoid base 1-phosphate (SPBP), sphingoid bases (SPB), acylcarnitines (CAR), cholesteryl esters (CE), sterols (ST), triacylglycerols (TG), diacylglycerols (DG), monoacylglycerols (MG), and fatty acids (FA). However, LipidQuant 2.1 provides the full flexibility to modify the lipid database by add/delete/modify compounds and add/delete lipid classes (see the **ReadMe** file in Supplementary Information).

2.2.2. Calculation of isotopic patterns

The software automatically calculates the isotopic patterns for all molecules according to the Rockwood approach [19]. The isotopic patterns for the most common adducts are calculated in both positive and negative modes. The LipidQuant 2.1 software supports the following adducts: $[M+H]^+$, $[M + H–H_2O]^+$, $[M+2H]^{2+}$, $[M+K]^+$, $[M+Na]^+$, $[M+Li]^+$, $[M + NH_4]^+$, $[M - H]^-$, $[M - 2H]^{2-}$, $[M+Cl]^-$, $[M + OAc]^-$, and $[M + HCOO]^-$.

2.2.3. Import of experimental data

The next step is the import of experimental data in.txt format into the software. It should be a summary table with all *m/z* features aligned among measured samples and the corresponding intensities detected in the samples for individual lipid classes exported from the vendor software or another peak picking software. The structure of the input.txt table must be followed as described in **ReadMe** file.

In case of lipid class separation methods coupled with QTOF high-resolution MS, it is important to achieve a satisfactory separation resolution of lipid classes to avoid any mass interference (*e.g.*, PE *vs.* PC) or possible ion-source fragmentation (*e.g.*, TG into DG or DG into MG). Simply speaking, one.txt file corresponds to one lipid class. More than one lipid class can be included in the same.txt file, under the condition that *m/z* values do not overlap. The loading of a collection of.txt files (each containing one or more lipid classes) into the software can be done in one step.

2.2.4. Peak assignment

By peak assignment, we mean the task of assigning experimental *m/z* to the calculated *m/z* values calculated for all lipid species present in the lipid database. This can be mathematically seen as an assignment problem [22]. Assignment problems are a class of combinatorial optimization, in which different "jobs" are assigned to "agents" to minimize the global cost function. For this particular application, "agents" are calculated *m/z* values from the database, and "jobs" correspond to experimental *m/z* values. The assignment consists of assigning unique pairs of "jobs/agents" to minimize the global difference between calculated and experimental *m/z* pairs. An example is shown in Fig. 2A.

Fig. 2A illustrates the necessity to optimize the peak assignment using a global strategy. Using a (more traditional) local strategy, the assignment could start with peak "a", and the optimal assignment is then a/1, b/2, and c/4, having peak "3" unassigned. However, if one starts with peak "b", the assignment is a/2, b/1, c/4, and the peak "3" is not assigned. As can be seen, the solution of the assignment problem depends on the starting peak, which is not acceptable. In other words, the problem should be resolved in a symmetric way: there should be no reason to start with peak "a", "b" or "c", as these starting points should not influence the result. In contrast, a global strategy solves this lack of symmetry. This is tackled by first calculating the cost matrix (M), which contains $\Delta m/z$ values between the calculated and experimental *m/z* values for all possible experimental/theoretical peak pairs. Next, the assignment is made to find the global optimum, i.e., the one that minimizes the global sum of all $\Delta m/z$ assignments. An example of such a cost matrix is shown in Fig. 2B. The green highlighted numbers indicate the selected (optimal) assignment using the Hungarian algorithm. Here, the global optimum is assignment a/1, b/2, and c/4. This is not trivial if one considers the peak "b" alone, for which the theoretical molecule "1"

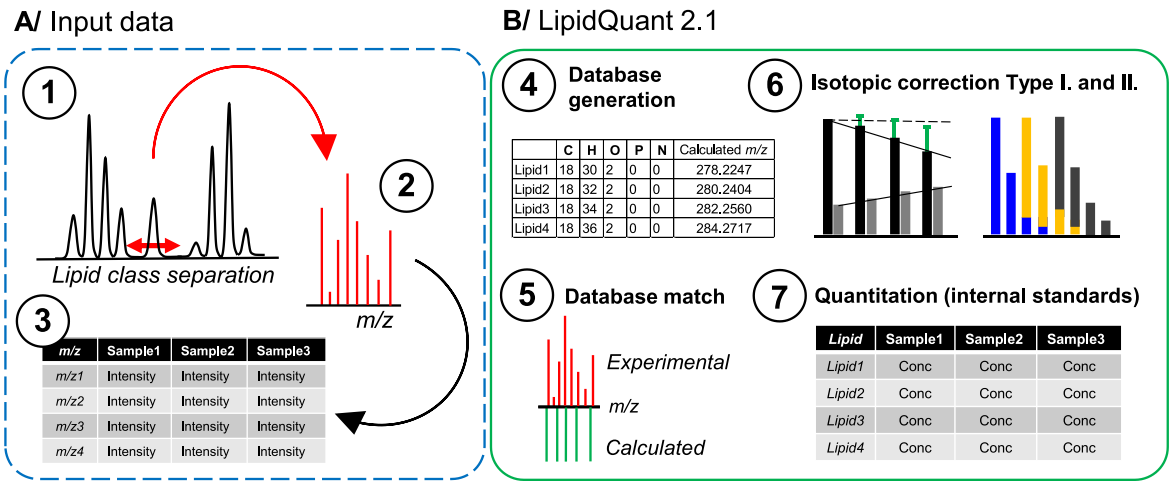
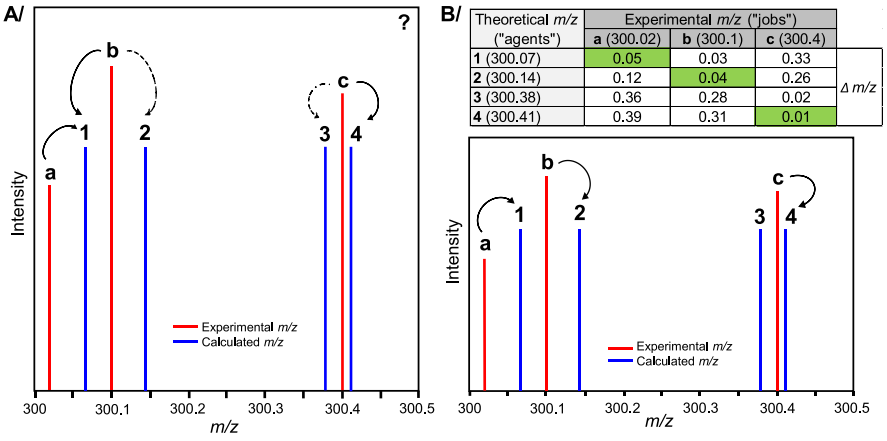


Fig. 1. Brief overview of LipidQuant 2.1 software. IS – internal standards.



Theoretical *m/z* ("agents")	Experimental *m/z* ("jobs")			
	a (300.02)	**b** (300.1)	**c** (300.4)	
1 (300.07)	0.05	0.03	0.33	Δ *m/z*
2 (300.14)	0.12	0.04	0.26	
3 (300.38)	0.36	0.28	0.02	
4 (300.41)	0.39	0.31	0.01	

Fig. 2. Peak assignment task. See the text for details.

seems optimal. However, this forces the theoretical molecule "2" to be assigned to "a", which is suboptimal.

The LipidQuant 2.1 software offers three possibilities of peak assignment: "Closest-unique", "Closest-sum", and "Highest" (Fig. 3).

The "Closest-unique" assigns the closest experimental *m/z* to the calculated *m/z* value of the lipid species with applied *Δm/z* tolerance; ±ppm (Fig. 3A). This option applies the Hungarian algorithm to the linear assignment problem explained above. The key condition for this option to be equivalent to a linear assignment problem is that only one experimental *m/z* value can be assigned to only one calculated *m/z* value. In other words, in cases where the single experimental *m/z* peak is found close (within *Δm/z* tolerance levels) to more than one theoretical peak, only one of such theoretical candidates can be assigned. This should not be confused with the fact that the linear assignment problems allow for experimental (and/or theoretical) peaks that can be left unassigned (see Fig. 2).

The "Closest-unique" assignment can be too restrictive in certain situations. This is precise because this solution does not allow more than one experimental signal to be the contribution of more than one theoretical *m/z* observation. In other words, the existence of nearly isobaric compounds (indistinguishable at experimental *m/z* tolerance levels) is not contemplated. The "Closest-sum" option solves this problem, as it departs from the assumptions of the linear assignment problem, in which only unique theoretical/experimental peak pairs are allowed. The "Closest-sum" enables the assignment of more experimental *m/z* values to one calculated *m/z*, if the theoretical value is within the *m/z* tolerance interval for two experimental *m/z* (the resolution of mass spectrometry does not allow the separation of these molecules). Finally, the sum concentration (of all possible candidates) is reported (Fig. 3B). However, the existence of low intensity experimental peaks might jeopardize the performance of both the closest-unique and closest-sum algorithms because these algorithms are based only on the difference between calculated and experimental *m/z* values, but not on the signal intensities. It is not rare to observe experimental peaks just above the signal-to-noise ratio (but close to the calculated *m/z* value), probably due to noise artefacts. In contrast, more prominent peaks (probably originating from the targeted compound) are also observed. However, the prominent peaks may appear slightly further away from the calculated *m/z* values than the noise artefacts. If the closest-sum or closest-unique algorithms are applied, the noise artefacts are misassigned to theoretical compounds, whereas the more prominent peaks are not assigned. To avoid such a problem, the user is recommended to use the "Highest-peak" method. It assigns the calculated *m/z* value to the highest experimental value located within *Δm/z* tolerance limits, regardless of whether other (minor) peaks are located closer (Fig. 3C). In all cases, the mass accuracy is calculated using well-known equation: $(ppm) = ((experimental\ m/z - calculated\ m/z)/calculated\ m/z) * 10^6$. The *Δm/z tolerance is always applied as ± ppm (e.g.* ±10 ppm: *from* -10 ppm to 1 0 ppm*).*

2.2.5. *Type I and type II isotopic corrections: nonnegative regression*

Type II isotopic corrections are common in lipidomics against Type I isotopic corrections based on the knowledge of the relative abundance of the base peak to the total abundance. Type II isotopic correction is more

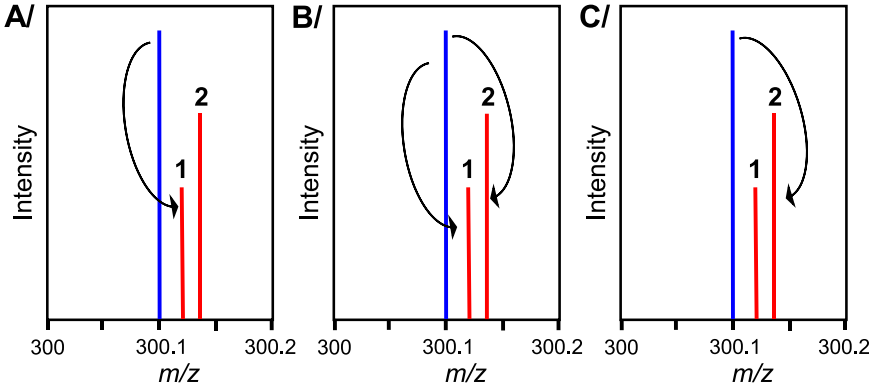


Fig. 3. Peak assignment: Closest unique (A), Closest sum (B), and Highest (C). Blue: calculated *m/z* values, red: experimental *m/z* values. See the text for details. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

challenging to correct due to an error propagation problem. Note that Type II isotopic correction originated from an overlay of M (monoisotopic mass) and M+2 isotope (two ^{13}C) with one more double bond; refer to Fig. 4A: M of TG 54:3 is overlapped with M+2 of TG 54:4 (two ^{13}C in structure), which has one more double bond. Hence, we need to know the concentration of more unsaturated lipid to correct the concentration of more saturated one [12]. However, when long chains are present (with different levels of saturation coeluting), a chain of dependencies appears. This is because the correction for one molecule depends on the concentration of the saturated compound. However, such concentrations should also be corrected as well, as there might be a more saturated molecule that coelutes. Traditionally, solving such interference requires knowing the signal/concentration of the most unsaturated molecule (within the same number of carbons), so the contribution of M+2 can be calculated (and subtracted) from the total signal. This might become more complicated when there are a series of lipids, as the signal at M of the molecule can also suffer interference from other molecules.

The Type II isotopic correction can be influenced by two issues. First, the correction is limited to the interferences due to different levels of saturation, but this is not the general case (*e.g.*, in case of coelution of SM and PC lipid species), and hence, it is difficult to generalize. Second, analysis of the error propagation of such calculations reveals that any error in the first molecule (*i.e.*, molecule with the lowest *m/z* value) will propagate systematically to all adjacent molecules, causing a bias in all of them. In other words, as the solution of the problem is not symmetric, any error in the low *m/z* molecule propagates into the rest.

The LipidQuant 2.1 software proposes a different approach to solve Type I and Type II isotopic corrections. Both corrections are tackled simultaneously using regression, applying the constraint of non-negativity and using information from all detected isotopes. This way, a general solution is applied, so there is (i) no need to formulate special cases, and (ii) the problem is solved naturally in a symmetric way, avoiding the problem associated with the error propagation mentioned above.

Fig. 4 illustrates the fitting approach on selected TG $[M + NH_4]^+$. In this figure, the theoretical isotope patterns of TG 54:4 ($C_{57}H_{106}O_6N$), TG 54:3 ($C_{57}H_{108}O_6N$), and TG 54:2 ($C_{57}H_{110}O_6N$) are depicted (Fig. 4A). The bottom plot (Fig. 4B) depicts the experimental signal (blue trace) and the fitted signal (red trace). To find the fitted trace, a regression was performed. We can express this problem mathematically as follows:

$$\mathbf{y} = \mathbf{X\beta} + \boldsymbol{\epsilon} \tag{1}$$

In this equation, **y** is the 9x1 column vector that expresses the experimental traces (each row of this vector corresponds to the *m/z* of 900.80, 901.80 902.81, 903.81, 904.81, 905.82, 906.82, 907.82, 908.82, as seen in Fig. 4). **X** is a 9x3 matrix that collects the theoretical isotopic abundances of TG 54:4, TG 54:3, and TG 54:2. Each column corresponds to a compound, and each row corresponds to the *m/z* enumerated above. The values of X are in this case:

$$\mathbf{X} = \begin{bmatrix} 0.5254 & 0 & 0 \\ 0.3334 & 0 & 0 \\ 0.1105 & 0.5253 & 0 \\ 0.0254 & 0.3335 & 0 \\ 0.0045 & 0.1106 & 0.5253 \\ 0.0007 & 0.0254 & 0.3335 \\ 0.0001 & 0.0045 & 0.1106 \\ 0 & 0.0007 & 0.0245 \\ 0 & 0.0001 & 0.0045 \end{bmatrix}$$

Finally, **β** in Eq. (1) is the 3x1 column vector that expresses the unknown contributions of each of the three molecules to the total signal, and **ε** is the 9x1 matrix that collects the error between experimental observations (**y**, blue trace in Fig. 4B) and the fitted (**Xβ**, red trace in Fig. 4B), theoretical signals. Solving Eq. (1) **β** and considering the least-squares solution becomes:

$$\beta = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \tag{2}$$

The values of **β** using the previous equation are 1.44E8, 1.50E8 and 0.39E8. As can be seen, the fit is quite satisfactory. Note also that Type I and Type II errors are solved here more elegantly. First, the values of the relative abundance of M, M+1, M+2, *etc.*, are considered in the **X** matrix, automatically accounting for Type I error. Moreover, the values of the contribution of each molecule are calculated not only with the value of M (as traditionally done), but also with all isotopes together, making the computation more accurate. For example, the signal at *m/z* = 901.80, which is usually not considered in traditional computations (as it corresponds to the M+1 signal), is considered here to further refine the

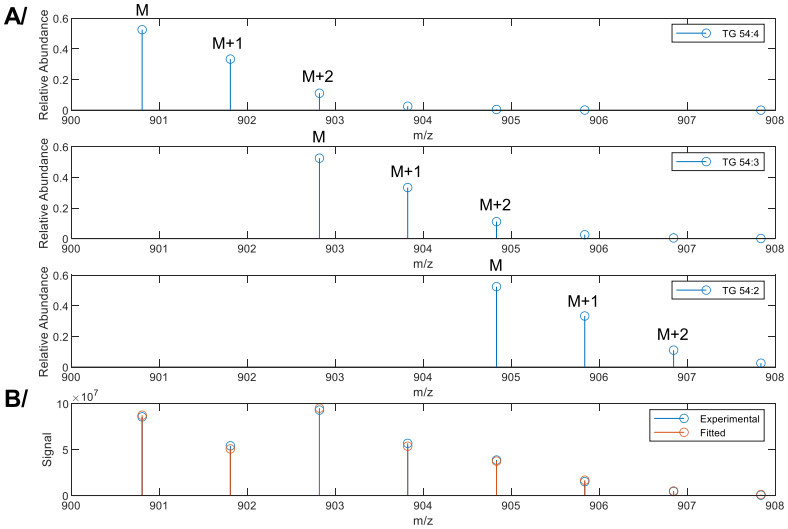


Fig. 4. The fitting approach. A/Relative abundance of isotopic peaks of TG 54:4 ($C_{57}H_{106}O_6N$), TG 54:3 ($C_{57}H_{108}O_6N$), and TG 54:2 ($C_{57}H_{110}O_6N$); adduct $[M + NH4]^+$ (monoisotopic mass, M; molecule includes one ^{13}C, M+1; molecule includes two ^{13}C, M+2), B/Comparison of experimental and fitted isotopic distributions.

calculations. The Type II error is also automatically considered, since different interferents (*i.e.*, columns in **X** matrix) for a given *m/z* value (rows in **X** matrix) are accounted for.

In reality, the number of compounds overlapping usually is larger (*e.g.*, 50 compounds may overlap one after another), and we have restricted our discussion to 3 compounds only for didactical reasons. The least squares solution provided by Eq. (2) is not limited to nonnegative contributions. In other words, negative values of **β** might be found, which does not make sense from a chemical point of view. To avoid this, the algorithm described in Lawson et al. [23] implemented in Matlab™ is finally applied instead of Eq. (2).

2.2.6. Quantification

The **β** values obtained using the procedure described in the previous section represent the best corrected approximation of the signal contributions from each compound in the lipid list. From this, the conversion of the signal contributions into concentrations is performed simply by the internal standards spiked in the sample (one-point calibration). The quantitation results of the determined lipid species can then be inspected graphically or exported to the Excel file for further processing. The exported Excel file includes two sheets containing (i) the estimated signal contributions after isotopic corrections and (ii) the estimated concentrations. A zero value is assigned for those compounds for which no *m/z* peak was found.

3. Results and discussion

LipidQuant 2.1 is demonstrated for the analysis of body fluids (human plasma and serum) by ultrahigh-performance supercritical fluid chromatography (UHPSFC) coupled with quadrupole time-of-flight (QTOF) MS. The details of the method, such as the extraction protocol, internal standards, and instrument parameters, are described in Table S1 in Supplementary Information. The test input.txt files used for software demonstration are also provided.

3.1. Determination of lipid species in biological samples

The measured data from 136 human plasma and 133 human serum samples have been used to demonstrate the LipidQuant 2.1 software. In total, 300 lipid species (including 21 internal standards) belonging to 10 lipid classes have been identified and quantified using the present software (Fig. 5A). The identification of lipid species has been based on the retention times of lipid classes (defined in advance by the.txt file generated by peak picking software), mass accuracy (±10 ppm tolerance), and using "closest-unique" as the peak assignment method. The whole list of determined lipid species with experimental *m/z* values, calculated *m/z* values, calculated mass accuracy, adduct ions, and concentrations in all samples is listed in Table S2 in Supplementary information. It is important to remember that Table S2 already includes the lipid species (300 in total) filtered from the original table generated by LipidQuant 2.1 (shown in Table S3). Every user of LipidQuant 2.1 should use their own expertise and experience to subdue their own data QC filtering. In our case, only lipid species detected in more than 75 % of all measured samples were included to demonstrate the software.

Different peak assignment approaches (the "Closest-unique" and the "Closest-sum") have been compared to find any identification/quantitation discrepancies arising by the selected peak assignment method. All identified/quantified lipids (300) were used for correlation plot (Fig. 5B and **C**), and results show the same output. This is a sign of (almost) absence of isobaric peaks.

3.2. Verification of responses

During a run sequence, samples, such as blanks (pure solution, chloroform/methanol), System Standard (SysSTD; neat lipid standards), BQC (Batch Quality Control; pooled sample), TQC (Technical Quality Control; pooled sample), NIST plasma (reference human plasma), and dilution series (diluted TQC sample), were injected regularly to check the quality of measurements (Table S1). Blanks, SysSTD, and dilution series have been used to prove correct and real lipid responses, leading to the verification of trustful results from LipidQuant 2.1. Blank samples contain pure chloroform/methanol solution and should not give us any lipid response compared to biological samples. SysSTD samples contain a mixture of selected lipid species with 18:1 fatty acyl chain(s) at high concentration (Table S1). We should not receive any response from other lipid species. Dilution series demonstrate the behavior of lipids depending on various dilutions. A lower injection volume should be reflected in the lower response of lipids. The LipidQuant 2.1 output provides accurate results reflecting real measurements as we have not found any artificial, suspicious, or changed results in comparison to raw/expected data. As an example, LPC 18:1 D7 (internal standard, $[M+H]^+$, $C_{26}H_{46}O_7NPD_7$) and LPC 18:1 (endogenous lipid, $[M+H]^+$, $C_{26}H_{53}O_7NP$) have been chosen to illustrate the problem. Fig. 6 shows the relationship between the MS response (intensity) and the order of the samples in the run sequence.

Fig. 6A nicely illustrates the behavior of the internal standard LPC 18:1 D7 response in various types of samples, as its intensity in blanks and SysSTD is not detected, reflecting the nature of samples. An experiment using dilution series also brings decisive results, as the diluted samples have been reflected by lower responses. On the other hand, LPC 18:1 gives much higher responses in SysSTD (Fig. 6B), which also corresponds to the nature of the sample. The plots of other internal lipid

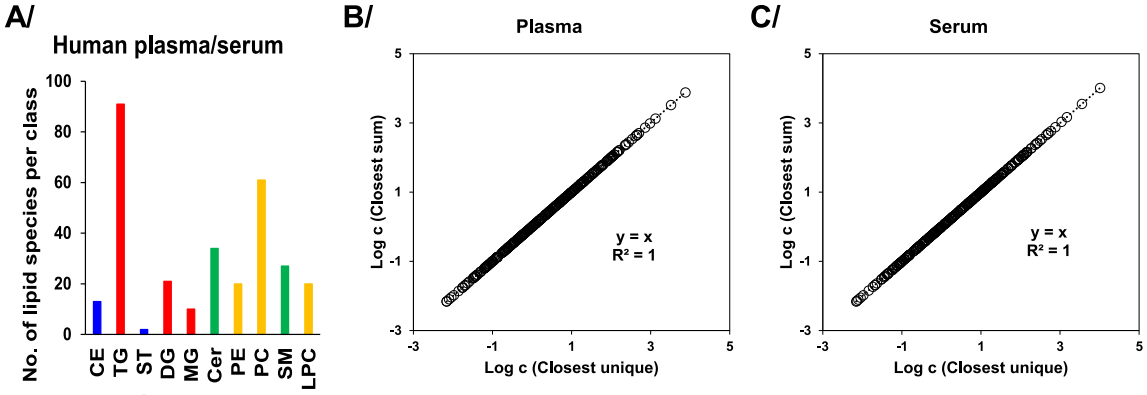


Fig. 5. Lipid distribution identified and quantified using LipidQuant 2.1 software (A). CE – cholesteryl esters, ST – cholesterol, TG – triacylglycerols, DG – diacylglycerols, MG – monoacylglycerols, Cer – ceramides, PE – phosphatidylethanolamines, PC – phosphatidylcholines, SM – sphingomyelins, LPC – lysophosphatidylcholines. Correlation graphs between the Closest-unique and the Closest-sum peak assignment approach for plasma (B) and serum (C) samples.

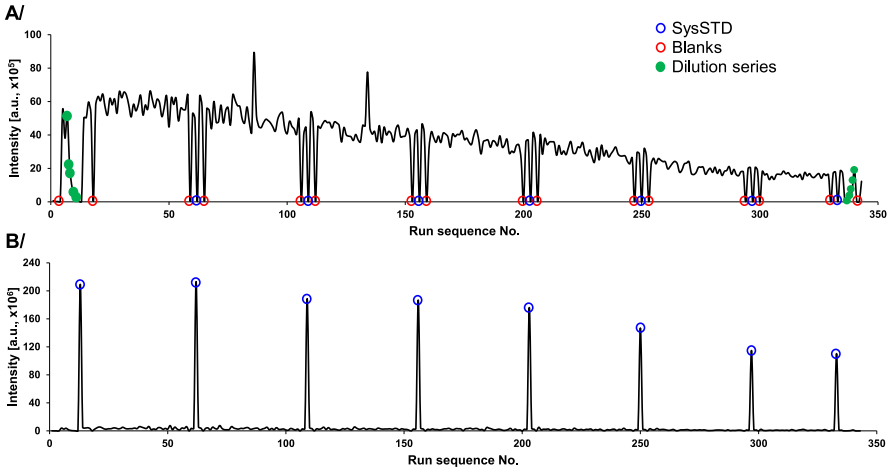


Fig. 6. Relationship between the MS response (mass spectrum peak intensity) and the order of samples in the run sequence for LPC 18:1 D7 (A) and LPC 18:1 (B).

standards and endogenous lipid standards (included in SysSTD samples) can be found in Fig. S1 in Supplementary Information. All these results demonstrate the suitability of the presented software in the matter of lipidomic data analysis using the lipid class separation coupled with QTOF high-resolution MS. The plot of the relation between MS response and the order of samples in the sequence for individual lipid species can be automatically checked in LipidQuant 2.1 software (Peak assignment window or View) for the user's detailed evaluation.

3.3. Comparison to the LipidQuant 1.0

Lipid concentrations obtained and calculated using LipidQuant 2.1 were compared with the concentrations provided by published LipidQuant 1.0 [18]. The lipid concentrations calculated using LipidQuant 1.0 are provided in Table S4 in Supplementary Information. The results from both software versions are highly comparable but not the same since different calculation approaches are used in individual softwares (LipidQuant 1.0 does not perform Type I correction). This fact is reflected in the final lipid concentrations. To define the concentration differences, a Residual Standard Deviation (RSD; %) has been calculated (Table S5 in Supplementary). Approximately 82 % of all compared lipid species have stated the RSD <20 %, confirming the high comparability despite different concentration calculations.

TG series $[M + NH_4]^+$ with 54 carbon atoms were chosen to illustrate the different concentration calculations in more detail (Fig. 7). The example is shown for one individual sample (Sample No. 1). Fig. 7A shows the raw experimental data from the MS spectrum. Fig. 7B shows the concentrations obtained by calculation without isotopic correction (green), only Type II isotopic correction (yellow, LipidQuant 1.0), and isotopic correction using regression (blue, LipidQuant 2.1). The importance of Type II isotopic correction is well described for lipid species TG 54:1. We could assume (without isotopic correction) the presence of TG 54:1 in a sample as a signal is observed at $m/z = 906.8490$ which corresponds to TG 54:1. However, it should be considered that this is a contribution of an isotope ion (M+2) of lipid species with one more double bond (TG 54:2; $m/z = 904.8333$). In lipid class separation, Type

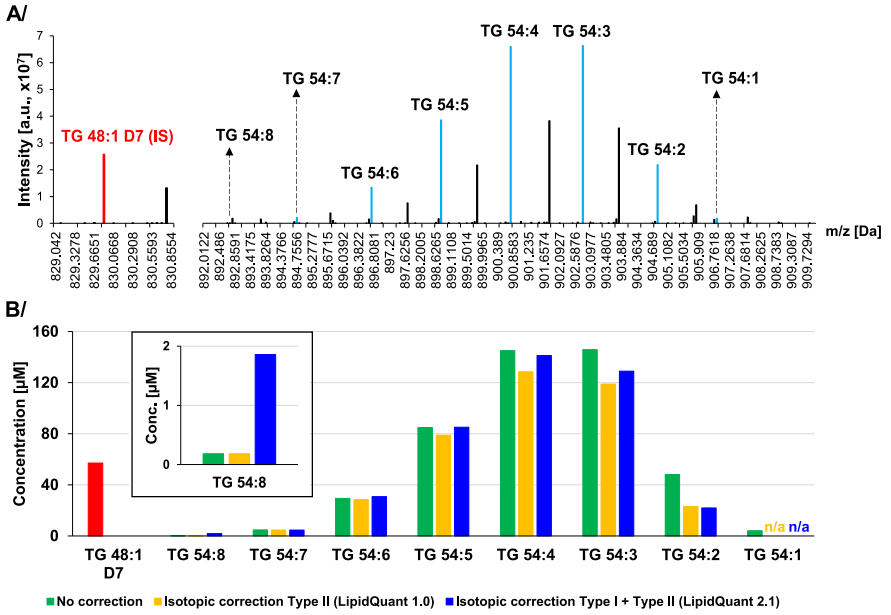


Fig. 7. Raw data obtained from the MS spectrum (A) for sample No.1. Concentration differences obtained by various calculations (B). The internal standard (TG 48:1 D7) is highlighted in red; adduct $[M + NH4]^+$. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

II isotopic correction is essential.

As mentioned, the type II error is compensated in Lipid Quant 2.1 in a different manner, which does not depend only (as it happens in traditional methods) on the concentration of the first molecule not affected by the isotopes of lighter compounds. This is illustrated with TG 54:8, where LipidQuant 2.1 detects a higher contribution of such molecule. In this molecule, the traditional Type II error is not applied, as there is no lighter molecule with which the M+2 isotope overlaps. However, the signal corresponding to M+1 isotope (visible at *m/z* 893.83) reveals that the concentration of such molecule is higher than the one measured using only the M signal. Therefore, the higher results are obtained by LipidQuant 2.1. Moreover, error propagation is more compensated with LipidQuant 2.1, as a miscalculation of such concentration (if any) does not significantly affect the others, unlike the traditional Type II correction used in LipidQuant 1.0.

4. Conclusions

We have introduced the LipidQuant 2.1 software, written in Matlab language, for the identification and quantification of lipid species acquired by lipid class separation hyphenated with QTOF high-resolution MS (full scan). The presented software offers visualization of isotopic patterns, peak assignment, and quantitative data with convenient exporting of final concentrations (and signals) in the Excel file for further user needs. The identification of lipid species is based mass accuracy and the selected peak assignment method (to assign experimental *m/z* features to calculated ones). The LipidQuant 2.1 tool works with input data (.txt file), including *m/z* features followed by intensities detected in all samples. The quantitation part of the presented software also includes both Type I and Type II isotopic corrections, which is not frequently used in other lipidomics softwares available so far. The output data (quantitative data) were compared to the already published first version of LipidQuant 1.0 with no significant differences, but LipidQuant 2.1 offers a more sophisticated approach to calculate isotopic corrections: nonnegative regression. This approach of regression for solving the isotopic correction calculation problems is more stable as there are less problems with the error propagation among lipid species. However, there are still more features we would like to implement in the next generations of LipidQuant, such as the identification of more adducts at the same time, isotopic corrections between different adducts (*e.g.*, overlay of $[M+H]^+$ and $[M+Na]^+$ adducts), statistical tests, uploading of raw data files, more visualizations, and the use of the software even for lipid species separations, to make one software universal for any format of lipidomic data.

5. Independent testing

Prof. Robert Ahrends and Dr. Dominik Kopczynski, University of Vienna, Faculty of Chemistry, Institute of Analytical Chemistry, A-1070 Vienna, Austria.

The authors present LipidQuant software, version 2.1 as an open source software available for lipidomic data processing in lipid class separation workflows. The functionality of the software was tested, and it can be applied for the calculation of molar concentrations of lipid species from various lipid classes based on the known concentrations of exogenous internal standards for these lipid classes. The LipidQuant includes an algorithm for the isotopic correction of both Type I and Type II, which is important, especially for lipid class separation approaches.

CRediT authorship contribution statement

Michaela Chocholoušková: Writing – original draft, Investigation, Conceptualization. **Gabriel Vivó-Truyols:** Writing – original draft, Software. **Denise Wolrab:** Writing – review & editing, Validation. **Robert Jirásko:** Writing – review & editing, Investigation. **Michela Antonelli:** Validation, Investigation. **Ondřej Peterka:** Writing – review & editing, Investigation. **Zuzana Vaňková:** Investigation. **Michal Holčapek:** Writing – review & editing, Resources, Project administration, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

Acknowledgement

This work was supported by the project No. 21-20238S sponsored by the Czech Science Foundation and ERC Adv grant No. 101095860 sponsored by the European Research Council. Authors would like to acknowledge the careful reading of the manuscript and program testing by other group members and cooperators (Karol Parchem, Mária Kanásová, Zuzana Minarechová, and Malena Manzi).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.chemolab.2024.105169.

References

[1] M.R. Wenk, Lipidomics: new tools and applications, Cell 143 (2010) 888–895, https://doi.org/10.1016/j.cell.2010.11.033.
[2] G. Liebisch, et al., Lipidomics need more standardization, Nat. Metab. 1 (2019) 745–747, https://doi.org/10.1038/s42255-019-0094-z.
[3] B. Peng, et al., LipidCreator workbench to probe the lipidomic landscape, Nat. Com. 11 (2020) 2057, https://doi.org/10.1038/s41467-020-15960-z.
[4] B. Peng, R. Ahrends, Adaptation of Skyline for targeted lipidomics, J. Proteome Res. 15 (2016) 291–301, https://doi.org/10.1021/acs.jproteome.5b00841.
[5] J.P. Koelmel, et al., LipidMatch: an automated workflow for rule-based lipid identification using untargeted high-resolution ttandem mass spectrometry data, BMC Bioinf. 18 (2017) 331, https://doi.org/10.1186/s12859-017-1744-3.
[6] J.E. Kyle, et al., LIQUID: an-open source software for identifying lipids in LC-MS/MS-based lipidomics data, Bioinformatics 33 (2017) 1744–1746, https://doi.org/10.1093/bioinformatics/btx046.
[7] A. O'Connor, et al., LipidFinder: a computational workflow for discovery of lipids identifies eicosanoids-phosphoinositides in platelets, JCI insight 2 (2017) e91634, https://doi.org/10.1172/jci.insight.91634.
[8] R. Herzog, et al., LipidXplorer: a software for Consensual cross-platform lipidomics, PLoS One 7 (2012) e29851, https://doi.org/10.1371/journal.pone.0029851.
[9] P. Husen, et al., Analysis of lipid experiments (ALEX): a software Framework for analysis of high-resolution shotgun lipidomics data, PLoS One 8 (2013) e79736, https://doi.org/10.1371/journal.pone.0079736.
[10] J. Hartler, et al., Lipid Data Analyzer: unattended identification and quantitation of lipids in LC-MS data, Bioinformatics 27 (2011) 572–577, https://doi.org/10.1093/bioinformatics/btq699.
[11] R. Schmid, et al., Integrative analysis of multimodal mass spectrometry data in MZmine 3, Nat. Biotechnol. 41 (2023) 447–449, https://doi.org/10.1038/s41587-023-01690-2.
[12] M. Wang, et al., Selection of internal standards for accurate quantitation of complex lipid species in biological extracts by electrospray ionization mass spectrometry – what, how and why? Mass Spectrom. Rev. 36 (2017) 693–714, https://doi.org/10.1002/mas.21492.
[13] E. Cífková, et al., Lipidomic differentiation of human kidney tumor and surrounding normal tissues using HILIC-HPLC/ESI-MS and multivariate data analysis, J. Chromatogr. B 1000 (2015) 14–21, https://doi.org/10.1016/j.jchromb.2015.07.011.
[14] M. Holčapek, et al., Determination of nonpolar and polar lipid classes in human plasma, erythrocytes and plasma lipoprotein fractions using ultrahigh-performance liquid chromatography - mass spectrometry, J. Chromatogr A 1377 (2015) 85–91, https://doi.org/10.1016/j.chroma.2014.12.023.
[15] M. Lísa, M. Holčapek, New High-throughput and comprehensive approach for lipidomic analysis using ultrahigh-performance supercritical fluid chromatography/electrospray ionization – mass spectrometry, Anal. Chem 87 (2015) 7187–7195, https://doi.org/10.1021/acs.analchem.5b01054.
[16] M. Chocholoušková, et al., Intra-laboratory cross-comparison of four lipidomic quantitation platforms using hydrophilic interaction liquid chromatography vs. supercritical fluid chromatography coupled to two quadrupole - time-of-flight mass

spectrometers, Talanta 231 (2021) 122367, https://doi.org/10.1016/j.talanta.2021.122367.
[17] D. Wolrab, et al., Validation of lipidomic quantitative methods based on lipid class separation - mass spectrometry: comparison of supercritical fluid chromatography and hydrophilic interaction liquid chromatography, Anal. Bioanal. Chem. 412 (2020) 2375–2388, https://doi.org/10.1007/s00216-020-02473-3.
[18] D. Wolrab, et al., LipidQuant tool for automated data processing in lipid class separation-mass spectrometry workflows, Bioinformatics 37 (2021) 4591–4592, https://doi.org/10.1093/bioinformatics/btab644.
[19] A.L. Rockwood, S.L. Van Orden, R.D. Smith, Rapid calculation of isotope distributions, Anal. Chem. 67 (15) (1995) 2699–2704, https://doi.org/10.1021/ac00111a031.
[20] D. Wolrab, et al., Lipidomic profiling of human serum enables detection of pancreatic cancer, Nat. Commun. 13 (2022) 124, https://doi.org/10.1038/s41467-021-27765-9.
[21] D. Wolrab, et al., Plasma lipidome profiles correlate with kidney, breast, and prostate cancer, Sci. Rep. 20322 (2021), https://doi.org/10.1038/s41598-021-99586-1.
[22] H.W. Kuhn, The Hungarian Method for the assignment problem, Nav. Res. Logist. Q. 2 (1955) 83–97, https://doi.org/10.1002/nav.3800020109.
[23] C.L. Lawson, R.J. Hanson, Solving Least-Squares Problems, Prentice Hall, Upper Saddle River, NJ, 1974, p. 161 (Chapter 23).