
Structural Bioinformatics

LipidQuant 1.0: automated data processing in lipid class separation - mass spectrometry quantitative workflows

Denise Wolrab^{1,#}, Eva Cífková^{1,#,§}, Pavel Čánž², Miroslav Lísa^{1,§}, Ondřej Peterka¹, Michaela Chocholoušková¹, Robert Jirásko¹, Michal Holčapek^{1,*}

¹Department of Analytical Chemistry, University of Pardubice, Faculty of Chemical Technology, Pardubice, Czech Republic, ²DEZA, a.s., Valašské Meziříčí, Czech Republic

#The first two authors should be regarded as Joint First Authors.

§Current address: Department of Chemistry, University of Hradec Králové, Rokitanského 62, 530 03 Hradec Králové, Czech Republic.

*To whom correspondence should be addressed.

Associate Editor: XXXXXXXX

Received on XXXXX; revised on XXXXX; accepted on XXXXX

Abstract

Summary: We present the LipidQuant 1.0 tool for automated data processing workflows in lipidomic quantitation based on lipid class separation coupled with high-resolution mass spectrometry. Lipid class separation workflows, such as hydrophilic interaction liquid chromatography or supercritical fluid chromatography, should be preferred in lipidomic quantitation due to the coionization of lipid class internal standards with analytes from the same class. The individual steps in the LipidQuant workflow are explained, including lipid identification, quantitation, isotopic correction, and reporting results. We show the application of LipidQuant data processing to a small cohort of human serum samples.

Availability and implementation: The LipidQuant 1.0 is freely available at figshare <https://doi.org/10.6084/m9.figshare.14604969.v1> and <https://holcapek.upce.cz/LipidQuant>.

Contact: Michal.Holcapek@upce.cz

Supplementary information: Supplementary data are available at *Bioinformatics* online.

1 Introduction

Lipids are biomolecules present in all cells with a large structural diversity. Lipid species can be classified into 8 main categories and numerous classes and subclasses, as introduced by Lipid MAPS (Fahy *et al.*, 2009), together with a lipid species database containing more than 40,000 entries. Lipidomic analysis aims at the identification and quantitation of all lipids present in biological samples using MS-based methods, either standalone or coupled with chromatographic techniques (Holčapek *et al.*, 2018). Lipid class separation approaches, such as hydrophilic interaction liquid chromatography (HILIC) or ultrahigh-performance supercritical fluid chromatography (UHPSFC), are particularly suited for lipidomic quantitation, because the coelution and coionization of lipid species belonging

to one lipid class together with an exogenous internal standard (IS) of the same lipid class allow the accurate and robust quantitation due to the same matrix effects (Liebisch *et al.*, 2019).

The manual data processing for lipid identification and quantitation for numerous samples and hundreds of lipids is time consuming, hence an automation in data processing is desirable. In recent years, several open source lipidomic software packages were developed, *e.g.*, Lipid Data Analyzer (Hartler *et al.*, 2010), LipidMatch (Koelmel *et al.*, 2019), and Lipid-Creator (Peng *et al.*, 2020), which enable the processing of LC/MS raw data including smoothing of background noise, peak detection, peak identification by comparing to online databases using algorithms, which consider *m/z* values and retention times, and quantitation by relating responses of target analytes to representative lipid standards with known concentra-

tion. Most of them were programmed in C or JAVA and provide visualization of peak shapes, MS or MS/MS spectra, 2D plot of m/z vs. retention time, molecular structure, etc. Lipid Data Analyzer performs a robust standardization algorithm for the correction of suppression effects, while LipidMatch Normalizer assumes the same ionization efficiency for the analyte and IS with the closest retention time. To our best knowledge, none of these software tools has the overall architecture applicable for high-throughput quantitation based on lipid class separation with HILIC-UHPLC/MS or UHPSFC/MS including a type II isotopic correction.

2 Results

LipidQuant 1.0 is a Microsoft Excel based script using visual basic for application (VBA) as a programming language, which is applicable for automated processing of data from all lipid class separation approaches coupled with high-resolution MS. A summary table of all m/z features detected for a lipid class is exported from the vendor software with the corresponding intensities as a txt file, which is then imported into LipidQuant for identification and quantitation of lipid species. Within this study, the MarkerLynx (Waters) was used to generate summary tables for each lipid class. However, other peak picking software can be used as well, as long as the output file contains m/z features in the first column with the heading “ m/z ” followed by individual samples containing the intensities or other quantitative measures for each m/z feature (see example in ReadMe file).

The general structure of LipidQuant is based on different table sheets entitled: Start, Database sheets, Support, Results, Average, and Deviation. The first sheet named “Start” is mainly used for data input and processing. The subsequent sheets represent the databases for individual lipid classes containing the exact masses of lipid species, their annotation on the molecular level, the percentage of M+1 and M+2 isotopes for isotopic correction, the tolerance range for lipid identification, and information on the IS. The “Support” sheet summarizes the number of lipid species for individual classes included in LipidQuant 1.0 and allows to define the number of consecutive injections. “Results” sheet summarizes the lipid concentrations for all samples, which can be inserted by pressing the button “Insert data”. For multiple injections, the summary table for the average concentrations together with the standard deviations is generated in “Average” and “Deviation” sheets.

The LipidQuant 1.0 contains the molecular level composition for 23 lipid classes with 1470 lipid species for positive-ion mode, and 24 lipid classes with 1999 lipid species for negative-ion mode, but the user has the full flexibility to modify the list of lipid species or add sheets for an additional lipid classes, as explained in Supplementary information and ReadMe file. The first step is the definition of the mass tolerance window in the individual database sheet according to the instrumental characteristics and measurement conditions. For HILIC and UHPSFC coupled to a QTOF mass analyzer, a tolerance window of ± 0.01 or ± 0.005 m/z is typically applied. The lipid class of interest is selected in the dropdown list, and then the summary table from the peak picking software for the defined lipid class is copied to the cell A1 in the “Start” sheet. The m/z feature filtering is applied when the button “Start” is pressed. The exact masses of lipid species defined in the database are compared with experimentally obtained m/z features and annotated as the corresponding lipid species with a color tag. Lipid species marked as green are within the specified mass tolerance window, and the assignment is unique. Lipid species marked as red are within the mass tolerance range, but more than one m/z feature was found to be within the selected mass tolerance range, so either the mass tolerance window is adjusted, or only one of the species is selected, while the other one is removed by deleting the whole line. The

yellow tag means that the lipid species is within two times the mass tolerance range, therefore the order in the database has to be inserted in the yellow colored column (E), or the feature is removed by deleting the line.

IS has to be defined for the quantitation together with the known concentration in the upper orange colored panel of each lipid class database. Different lipid species can be quantified with different IS, which may be advantageous for lipid species with different response factors, *i.e.*, lipid species with different length of fatty acyl chains or different number of double bonds. This may partially compensate the quantitation errors caused by differences in ionization efficiencies for short vs. very-long or saturated vs. polyunsaturated fatty acyl chains. The automated quantitation is performed by pressing the button “Move” in the “Start” sheet, and then lipid species concentrations are summarized in the lipid class sheet.

The lipid species identification and quantitation have to be performed for each lipid class separately. Therefore, the identified lipid species in the “Start” sheet have to be removed by pressing the button “Clear”, and the next lipid class has to be chosen in the dropdown list and processed in the same way as explained before. If it is intended to remove all concentrations for all lipid classes, “Clear all concentrations” button has to be pressed in the “Start” sheet. If it is aimed to remove only concentrations of one lipid class, “Clear concentrations” button in the corresponding database has to be triggered.

The summary table is generated by triggering the “Insert data” button in the “Results” sheet. Concentrations of all lipid species for all samples will be inserted in the results table. For multiple injections (defined in the “Support” sheet), the average and deviation will be automatically calculated and summarized in the corresponding sheets.

The lipidomic quantitation of human serum samples from healthy volunteers is illustrated with the example of positive-ion UHPSFC/MS data set with the determination of molar concentrations of lipid species from 8 lipid classes (CE, TG, DG, MG, Cer, PC, LPC, and SM).

3 Conclusions

LipidQuant 1.0 represents a simple script freely available for lipidomic identification and quantitation including type II isotopic correction, which can run on every computer with installed Microsoft Excel. All parameters are adjustable, such as the choice of IS, the mass tolerance window, and the extent of embedded lipid database. LipidQuant is vendor-independent, because it works with m/z features and their intensities in a txt table, which may be obtained by any peak picking software.

Funding

This work has been supported by Czech Science Foundation (No. 21-20238S).

Conflict of Interest: none declared.

References

- Fahy, E. et al. (2009) Update of the LIPID MAPS comprehensive classification system for lipids. *J. Lipid Res.*, **50**, S9-S14.
- Hartler, J. et al. (2010) Lipid Data Analyzer: unattended identification and quantitation of lipids in LC-MS data. *Bioinformatics*, **27**, 572-577.
- Holčapek, M. et al. (2018) Lipidomic Analysis. *Anal. Chem.*, **90**, 4249-4257.
- Koelmel, J. P. et al. (2019) Software tool for internal standard based normalization of lipids, and effect of data-processing strategies on resulting values. *Bioinformatics*, **20**, 217.
- Liebisch, G. et al. (2019) Lipidomics needs more standardization. *Nat. Metab.*, **1**, 745-747.
- Peng, B. et al. (2020) LipidCreator workbench to probe the lipidomic landscape. *Nat. Com.*, **11**, 2057.