*Bioinformatics*, 31(7), 2015, 1150–1153
doi: 10.1093/bioinformatics/btu796
Advance Access Publication Date: 29 November 2014
Applications Note

OXFORD

Data and text mining

# *Lipid-Pro:* a computational lipid identification solution for untargeted lipidomics on data-independent acquisition tandem mass spectrometry platforms

**Zeeshan Ahmed[1,2,*], Michel Mayr[3], Saman Zeeshan[2], Thomas Dandekar[2], Martin J. Mueller[3] and Agnes Fekete[3]**

[1]Department of Neurobiology and Genetics, Biocenter, [2]Department of Bioinformatics, Biocenter and [3]Department of Pharmaceutical Biology, Biocenter, University of Wuerzburg, Wuerzburg, Germany

*To whom correspondence should be addressed.

Associate Editor: Jonathan Wren

Received on July 15, 2014; revised on November 24, 2014; accepted on November 25, 2014

## Abstract

**Summary:** A major challenge for mass spectrometric-based lipidomics, aiming at describing all lipid species in a biological sample, lies in the computational and bioinformatic processing of the large amount of data that arises after data acquisition. *Lipid-Pro* is a software tool that supports the identification of lipids by interpreting large datasets generated by liquid chromatography—tandem mass spectrometry (LC–MS/MS) using the advanced data-independent acquisition mode $MS^E$. In the $MS^E$ mode, the instrument fragments all molecular ions generated from a sample and records time-resolved molecular ion data as well as fragment ion data for every detectable molecular ion. *Lipid-Pro* matches the retention time-aligned mass-to-charge ratio data of molecular- and fragment ions with a lipid database and generates a report on all identified lipid species. For generation of the lipid database, *Lipid-Pro* provides a module for construction of lipid species and their fragments using a flexible building block approach. Hence, *Lipid-Pro* is an easy to use analysis tool to interpret complex $MS^E$ lipidomics data and also offers a module to generate a user-specific lipid database.

**Availability and implementation:** *Lipid-Pro* is freely available at: http://www.neurogenetics.biozentrum.uni-wuerzburg.de/en/project/services/lipidpro/

**Contact:** zeeshan.ahmed@uni-wuerzburg.de

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

## 1 Introduction

Lipids are compounds of biological origin that are soluble in nonpolar solvents. In biological systems, lipids perform ubiquitous functions such as construction of cell membranes, stabilization of membrane-bound proteins, balancing of energy metabolism or cell signaling (Wymann and Schneiter, 2008). All lipid species in a cell or tissue are referred to as the lipidome. Changes within lipidomes have been closely linked to stress responses or various diseases including diabetes, obesity, heart diseases or neurodegenerative diseases (reviewed in Murphy and Nicolaou, 2013).

Liquid-chromatography coupled to tandem mass spectrometry (LC–MS/MS) is suitable for lipidome analysis since it can cover a broad range of lipids (Harald *et al.*, 2012). Both, data-dependent acquisition (DDA) and data-independent acquisition (DIA) are widely used for LC–MS-based lipidomics. In DDA, the most abundant molecular ions at each retention time (rt) are successively

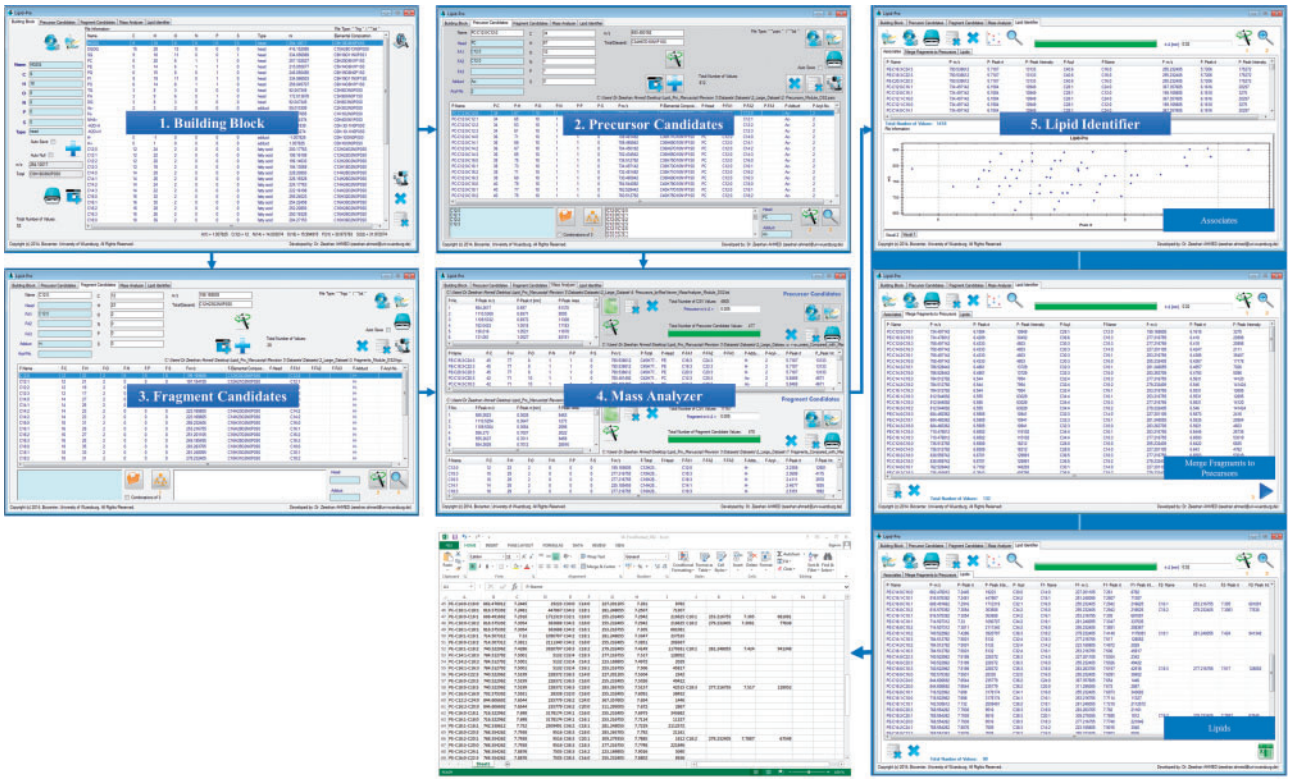


**Fig. 1.** The graphical user interfaces of the five principal modules (*Building Block, Precursor Candidates, Fragment Candidates, Mass Analyzer* and *Lipid Identifier*) of *Lipid-Pro* are shown. Results can be exported in Microsoft Excel format

selected for isolation and fragmentation. Despite the advantages of the DDA approach, there are some inherent limitations such as often irreproducible molecular ion selection, undersampling and long instrument cycle times. DIA overcomes those limitations. $MS^E$, supported by Waters qTOF instrument platforms, acquires intact molecular- and fragment ions in an unbiased manner by continuously switching the energy for collision-induced dissociation between low (MS) and high (MS/MS) energy. DIA analysis requires coherent and intricate data processing since it produces complex data containing information on all molecular and fragment ions at each rt.

Commercial and free downloadable software such as mzMine 2 (Pluskal *et al.*, 2010) or Lipid Data Analyser (Hartler *et al.*, 2011) provide support to align and compare the spectral features of LC–MS data. These applications identify lipid species in LC–MS raw data by matching their monoisotopic masses or elemental compositions with a custom-based or online database. However, these software applications do not consider fragmentation data. LipidBlast (Kind *et al.*, 2013) identifies lipid species by matching the measured MS/MS spectra to an *in silico* generated library after converting raw data into the MGF format. LipidBlast can be used for the analysis of $MS^E$ data, but it is laborious to match hundreds of different lipid species of the database with LC–MS/MS data.

We developed the *Lipid-Pro* software solution for rapid identification of lipids in preprocessed DIA data acquired with LC–MS/MS using rt-aligned monoisotopic masses of molecular ions (MS) and fragment ions (MS/MS). Since *Lipid-Pro* provides a module for construction of lipid species and their characteristic fragments using a flexible building block approach, rare or even non-described lipids beside the common species can be identified in total lipid extracts of biological samples. To meet the technological objectives of this research, we took a step forward in the development of a new user friendly, modular and client-based database management system. The *Lipid-Pro* application can be configured on Microsoft Windows platforms following a simple six-step installation process.

## 2 *Lipid-Pro*

We developed *Lipid-Pro* as a desktop application (Fig. 1), helpful for the identification of pre-defined lipid species (termed candidates) in large LC–$MS^E$ datasets (or other DIA datasets from other instrument platforms provided in the *.csv data format). For automatic generation of a lipid database, *Lipid-Pro* provides modules for the construction of complex lipid species and their fragments using a building block approach (Yang *et al.*, 2009). By comparison with the self-generated lipid database, pre-defined lipids are identified in the preprocessed datasets by matching the rt-aligned mass-to-charge ratios (*m/z*) of lipid molecular ions and fragment ions in an efficient manner. As output, *Lipid-Pro* delivers information on the identified lipid species including name, rt, *m/z* as well as the peak areas of lipid species and their fragments.

*Lipid-Pro* comprises five modules working in product-line architecture: *Building Block*, *Precursor Candidates*, *Fragment Candidates*, *Mass Analyzer* and *Lipid Identifier* (Fig. 2). Each module performs its task independently and its output is used as an input for the next module until lipid identification is achieved.

### 2.1 Lipid database

The majority of lipid molecular species consist of a linear combination of a small number of chemical entities or building blocks from

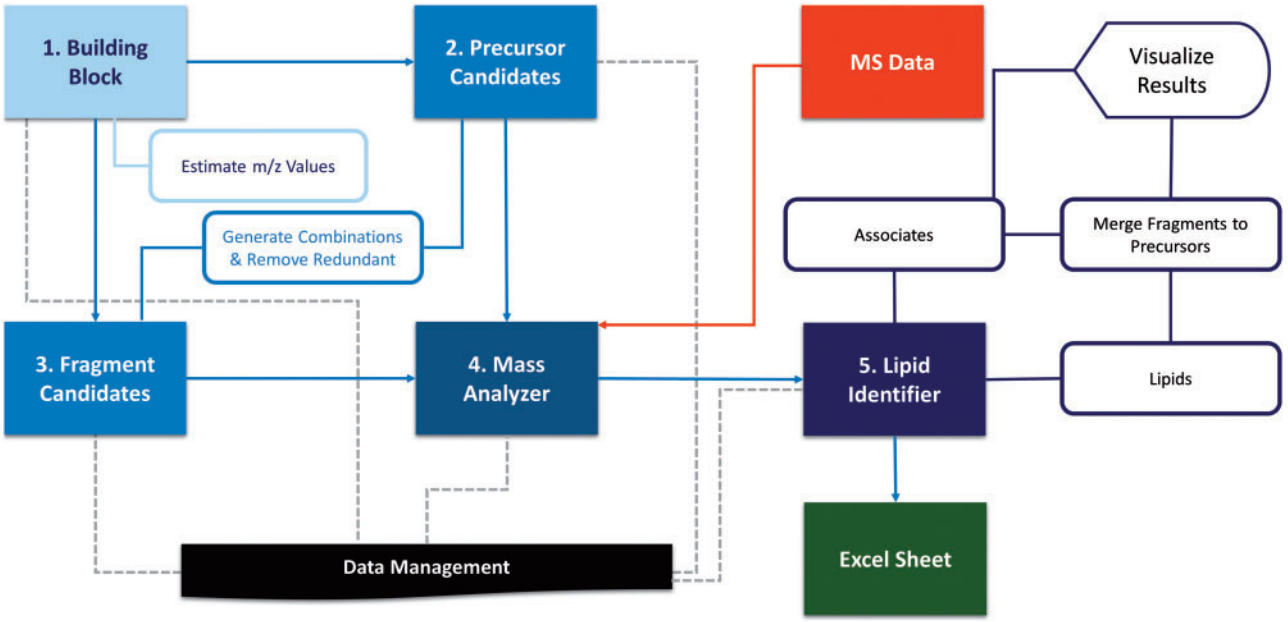


**Fig. 2.** The conceptual architecture and work flow of the five principal modules of *Lipid-Pro* are shown

which naturally occurring lipids can be derived (Foster *et al.*, 2013). Molecular species of all glycerol-based lipids comprise a lipid class specific backbone (here termed as 'head') esterified with one or more fatty acids (termed 'FA1', 'FA2', 'FA3'). For instance, all phosphatidylcholine species consist of glycerophosphocholine (head) and two acyl residues ('FA1' and 'FA2')

The *Building Block* module is used to calculate the monoisotopic masses ($m_i$) from the elemental composition of lipid class specific 'heads' and various 'FA's that are the principal building blocks of complex lipids as well as different potential adduct ions (adduct). Next, the *Precursor Candidates* module calculates the monoisotopic masses of hundreds of complex lipid species by combining different 'head' and 'FA' building blocks and also adds potential adduct ions (adduct). In addition, a systematic name is generated that comprises the names of the head and the names of the 'FA's. The *Fragment Candidates* module calculates the $m_i$-values of expected fragments. Since fragmentation predominantly involves cleavage of ester bonds, the characteristic fragments can be derived from the lipid candidates. In addition, a systematic fragment name is generated that comprises the name(s) of the building block(s) of the fragment.

### 2.2 Lipid identification

For lipid identification in LC–MS/MS datasets, *Lipid-Pro* requires two preprocessed files in *.csv format that contain the rt-aligned *m/z*-values of molecular ions (MS data) and rt-aligned fragment ions (MS/MS data). For pre-processing of the raw data, we used MS$^E$ DataViewer$^{TM}$ 1.2 (an external third party application) provided by the manufacturer of the used instrument (UPLC$^{TM}$-Synapt$^{TM}$ G2 HDMS, Waters).

After import of the preprocessed data into the *Mass Analyzer* module, the monoisotopic masses of the putative molecular ions (listed in the *Precursor Candidates* module) and the putative fragments (listed in the *Fragment Candidates* module) are matched to the measured *m/z*-values. The mass tolerance ('*m/z* Δ') is mainly determined by the mass accuracy of the analytical system and can be defined by the user.

Next, in the *Lipid Identifier* module, the time-aligned precursors and fragments that have been found in the database are associated according to their rt-values. The rt-tolerance is dependent on several factors, such as the peak form or the measurement conditions, and can be defined. Finally, all precursor hits are inspected for the presence of all expected fragments. Therefore, the module checks if the fragment's name is a segment of the precursor's name. Lipid candidates missing the expected fragments are removed. The identified lipids can then be visualized by plotting the *m/z*-values against the rt-values. Results can be saved, printed and exported into Excel.

## 3 Conclusion

The most recent available and easy to use version of *Lipid-Pro* has been tested and validated in-house. The advancements in mass spectrometry techniques for lipid analysis combined with this sophisticated computational tool can support various studies such as gene function analysis or biomarker discovery.

## Acknowledgements

*Funding*: The authors thank the Deutsche Forschungsgemeinschaft (DFG), collaborative research center SFB1047 "Insect timing", for funding this research (to ZA, MM, MJM, AF). The authors also thank TR34/Z1 for support (to SZ, TD).

*Conflict of Interest*: none declared

## References

Foster,J.M. *et al.* (2013) LipidHome: a database of theoretical lipids optimized for high throughput mass spectrometry lipidomics. *PLoS One* **8**, e61951.

Harald,C.K., *et al.* (2012) Mass spectrometry based lipidomics: an overview of technological platforms. *Metabolites* **2**, 19–38.

Hartler,J. *et al.* (2011) Lipid data analyser: unattended identification and quantification of lipids in LC-MS data. *Bioinformatics* **27**, 572–577.

Kind,T. *et al.* (2013) LipidBlast in silico tandem mass spectrometry database for lipid identification. *Nat. Methods* **10**, 755–758.

Murphy,S.A. and Nicolaou,A. (2013) Lipidomics applications in health, disease and nutrition research. *Mol. Nutr. Food Res.*, **57**, 1336–1346.

Pluskal,T. *et al.* (2010) MZmine2: modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. *BMC Bioinformatics* **11**, 395–405.

Yang,K. *et al.* (2009) Automated lipid identification and quantification by multidimensional mass spectrometry-based shotgun lipidomics. *Anal. Chem.*, **81**, 4356–4368.

Wymann,M.P. Schneiter,R. (2008) Lipid signalling in disease. *Nat. Rev. Mol. Cell Biol.*, **9**, 162–176.