Bioinformatics, 35(4), 2019, 698–700
doi: 10.1093/bioinformatics/bty661
Advance Access Publication Date: 23 July 2018
Applications Note

OXFORD

Systems biology

LipidIMMS Analyzer: integrating multi-dimensional information to support lipid identification in ion mobility—mass spectrometry based lipidomics

Zhiwei Zhou[1,2], Xiaotao Shen[1,2], Xi Chen[1,2], Jia Tu[1,2], Xin Xiong[1] and Zheng-Jiang Zhu[1,*]

[1]Interdisciplinary Research Center on Biology and Chemistry, Shanghai Institute of Organic Chemistry, Chinese Academy of Sciences, Shanghai 200032, China and [2]University of Chinese Academy of Sciences, Beijing 100049, China

*To whom correspondence should be addressed.

Associate Editor: Jonathan Wren

Received on April 23, 2018; revised on July 2, 2018; editorial decision on July 19, 2018; accepted on July 20, 2018

Abstract

Summary: Ion mobility—mass spectrometry (IM-MS) has showed great application potential for lipidomics. However, IM-MS based lipidomics is significantly restricted by the available software for lipid structural identification. Here, we developed a software tool, namely, LipidIMMS Analyzer, to support the accurate identification of lipids in IM-MS. For the first time, the software incorporates a large-scale database covering over 260 000 lipids and four-dimensional structural information for each lipid [i.e. *m/z*, retention time (RT), collision cross-section (CCS) and MS/MS spectra]. Therefore, multi-dimensional information can be readily integrated to support lipid identifications, and significantly improve the coverage and confidence of identification. Currently, the software supports different IM-MS instruments and data acquisition approaches.

Availability and implementation: The software is freely available at: http://imms.zhulab.cn/LipidIMMS/.

Contact: jiangzhu@sioc.ac.cn

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

1 Introduction

Lipids play many vital roles in cell biology such as membrane constitution, energy storage and cell signaling (Han, 2016). The comprehensive study of lipids (i.e. lipidomics) provides mechanistic insights to many important diseases such as diabetes, cancer and neurodegenerative diseases (Han, 2016). The high complexity of lipidome presents a great challenge for lipidomics, which requires powerful analytical techniques to separate and identify lipids. Recently, ion mobility—mass spectrometry (IM-MS) has become a promising technology for lipidomics by providing high separation capacity, high sensitivity and selectivity and capability to distinguish lipid isomers (Hinz *et al.*, 2018; Zheng *et al.*, 2018). Coupling IM-MS with liquid chromatography and data-independent tandem MS techniques (e.g. MS^E and AIF) further enables the comprehensive acquisition of four-dimensional information of lipids in one analysis, including *m/z* of MS1, retention time (RT), collision cross-section (CCS) and MS/MS spectra (Hines *et al.*, 2017; Paglia and Astarita, 2017; Zhou *et al.*, 2018). The integration of multi-dimensional information provides a holistic characterization of lipid structures and supports the large-scale and unambiguous identification of lipids in complex biological samples.

However, IM-MS based lipidomics is significantly restricted by the availability of software tools for data analysis, especially for lipid structural identification. Some tools have been developed to process IM-MS based lipidomics data, such as LC-IMS-MS Feature Finder (Crowell *et al.*, 2013), vender software Mass Profiler (Agilent) and Progenesis QI (Waters). Nevertheless, accurate lipid identification has not been sufficiently achieved since limited

© The Author(s) 2018. Published by Oxford University Press. All rights reserved. For Permissions, please e-mail: journals.permissions@oup.com

Downloaded from https://academic.oup.com/bioinformatics/article/35/4/698/5057157 by Stanford University user on 25 January 2023

information was used for identification (e.g. MS1+CCS and MS1+MS/MS). More importantly, the sizes of embedded libraries in these programs (e.g. CCS library and MS/MS library) are very limited, which further restricts the coverage of lipid identifications in complex biological samples.

Here, we present a freely available webserver, namely, LipidIMMS Analyzer, to integrate four-dimensional information for lipid identification in IM-MS based lipidomics (Fig. 1a). For the first time, the software includes a large-scale database covering over 260 000 lipids and four-dimensional structural information for each lipid (i.e. *m/z*, RT, CCS and MS/MS spectra). Therefore, confidence and coverage of lipid identifications in complex samples is significantly improved. The software supports different data acquisition approaches (e.g. LC-IM-MS and direct infusion-IM-MS) and different IM-MS instruments (e.g. Agilent and Waters).

2 Methods and features

2.1 General workflow

LipidIMMS Analyzer webserver provides an interactive workflow for users to perform lipid identification using IM-MS based lipidomics data (Fig. 1b): (i) data import of a MS1 peak table and/or MS/MS data; (ii) database loading; (iii) RT calibration; (iv) *m/z*, RT and CCS match and score; (v) MS/MS spectral match and score; (vi) composite score calculation and (vii) download and browse results of lipid identifications. More details are provided in Supplementary Information.

2.2 Four-dimensional lipid database

In LipidIMMS Analyzer, a large-scale lipid database with four-dimensional information is developed to support lipid identification.

2.2.1 Dimension 1: MS1 library

Lipid structures were created using the template-based combinatorial enumeration (Sud *et al.*, 2012). The template allows the length of acyl chains varying from 2 to 39 and the number of double bonds varying from 0 to 6. A total of 267 716 lipid structures were generated covering four categories (glycerophospholipids, sphingolipids, glycerolipids and fatty acids) and 25 classes (Supplementary Table S1).

2.2.2 Dimension 2: Retention time library

Retension time values for all lipids were predicted using a random forest algorithm. RT on both reverse phase (RP) and hydrophilic liquid chromatography (HILIC) columns were predicted using different training sets (Hines *et al.*, 2017; Zhou *et al.*, 2017) and molecular descriptors (Supplementary Table S2 and Fig. S1). The prediction accuracy was validated using another 78 and 35 RTs with median errors of 7 and 1 s for RP and HILIC columns, respectively (Supplementary Fig. S2). To accommodate different LC conditions, RT values in the library could be re-calculated using the RT calibration method. One example is given to demonstrate the re-calibration error between two different LC systems was about 5 s (median error, Supplementary Tables S3–S5 and Fig. S3).

2.2.3 Dimension 3: Collision cross-section library

The CCS values were predicted using our previous developed software—LipidCCS (Zhou *et al.*, 2017). A total of 375 565 CCS values were predicted for different ion adducts. The coverage of the CCS library is significantly larger than our previous LipidCCS library (63 434 CCS values).

2.2.4 Dimension 4: MS/MS spectrum library

The MS/MS spectra of lipids were predicted using the fragmentation rules (Kind *et al.*, 2013; Tu *et al.*, 2018). These rules were manually summarized according to experimental MS/MS spectra acquired using lipid standards. A total of 375 565 MS/MS spectra were generated covering five common adducts. For each MS/MS spectrum, structural annotations of fragments were also provided.

2.3 Integration of multi-dimensional information

The combination of m/z, RT, CCS and MS/MS spectra for lipid identification can effectively improve the confidence. Frist, library match in each dimension was separately scored. A trapezoidal function was designed to score the RT and CCS matches, and a reverse dot-product function was used to score the MS/MS spectral match. Then different dimensional scores were integrated to calculate the composite score with a linear weighting function. LipidIMMS Analyzer features a high flexibility and versatility. Users could select different combinations to perform lipid identifications according to their requirements (e.g. MS1+MS/MS, MS1+RT+CCS, MS1+RT+CCS+MS/MS and so on).

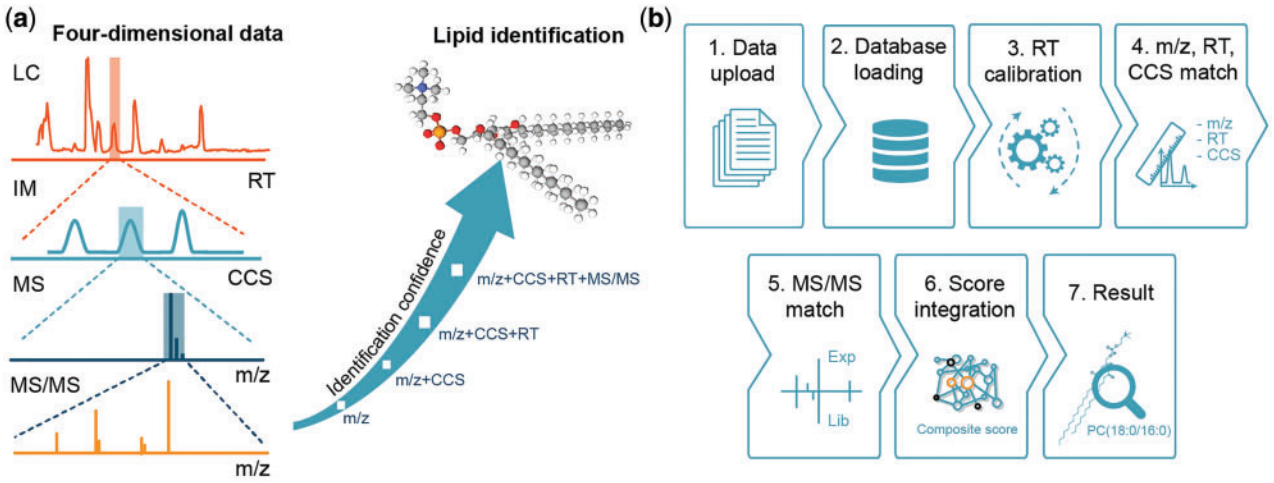

Fig. 1. (**a**) Integration of multi-dimensional information to support lipid identification in IM-MS based lipidomics and (**b**) the data analysis workflow of LipidIMMS Analyzer

Downloaded from https://academic.oup.com/bioinformatics/article/35/4/698/5057157 by Stanford University user on 25 January 2023

3 Results

To demonstrate the performances of LipidIMMS Analyzer, we used a series of lipid standards and multiple biological samples. Results in Supplementary Table S6 demonstrated that the confidence of lipid identifications was effectively improved by matching multi-dimensional information. The software could distinguish different lipid isomers (Supplementary Figs S5–S7) including isomers from different classes, isomers with different acyl chains, and sn1/sn2 positional isomers. Finally, using our software, a total of 500–600 lipids covering four lipid categories could be identified in different biological samples, such as human plasma, mammalian cells, and mouse brain tissue (Supplementary Figs S8–S10).

4 Conclusion

LipidIMMS Analyzer is a freely available webserver to support lipid identification in IM-MS based lipidomics. It incorporated a large-scale lipid database with four-dimensional information to perform accurate lipid identifications with a broad coverage and high confidence.

Acknowledgements

The work is supported by National Natural Science Foundation of China [Grant No. 21575151] and Chinese Academy of Sciences Major Facility-based Open Research Program. Z.-J. Z. is supported by Thousand Youth Talents Program.

Conflict of Interest: none declared.

References

Crowell,K.L. *et al*. (2013) LC-IMS-MS Feature Finder: detecting multidimensional liquid chromatography, ion mobility and mass spectrometry features in complex datasets. *Bioinformatics*, **29**, 2804–2805.

Han,X. (2016) Lipidomics for studying metabolism. *Nat. Rev. Endocrinol.*, **12**, 668–679.

Hines,K.M. *et al*. (2017) Assessment of altered lipid homeostasis by HILIC-ion mobility-mass spectrometry-based lipidomics. *J. Lipid Res.*, **58**, 809–819.

Hinz,C. *et al*. (2018) The potential of Ion Mobility Mass Spectrometry for high-throughput and high-resolution lipidomics. *Curr. Opin. Chem. Biol.*, **42**, 42–50.

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

Paglia,G. and Astarita,G. (2017) Metabolomics and lipidomics using traveling-wave ion mobility mass spectrometry. *Nat. Protoc.*, **12**, 797–813.

Sud,M. *et al*. (2012) Template-based combinatorial enumeration of virtual compound libraries for lipids. *J. Cheminf.*, **4**, 23.

Tu,J. *et al*. (2018) Absolute quantitative lipidomics reveals lipidome-wide alterations in aging brain. *Metabolomics*, **14**, 5.

Zheng,X. *et al*. (2018) Recent advances in lipid separations and structural elucidation using mass spectrometry combined with ion mobility spectrometry, ion-molecule reactions and fragmentation approaches. *Curr. Opin. Chem. Biol.*, **42**, 111–118.

Zhou,Z. *et al*. (2017) LipidCCS: prediction of collision cross-section values for lipids with high precision to support ion mobility-mass spectrometry--based lipidomics. *Anal. Chem.*, **89**, 9559–9566.

Zhou,Z. *et al*. (2018) Advancing the large-scale CCS database for metabolomics and lipidomics at the machine-learning era. *Curr. Opin. Chem. Biol.*, **42**, 34–41.

Downloaded from https://academic.oup.com/bioinformatics/article/35/4/698/5057157 by Stanford University user on 25 January 2023