analytical chemistry

pubs.acs.org/ac Article

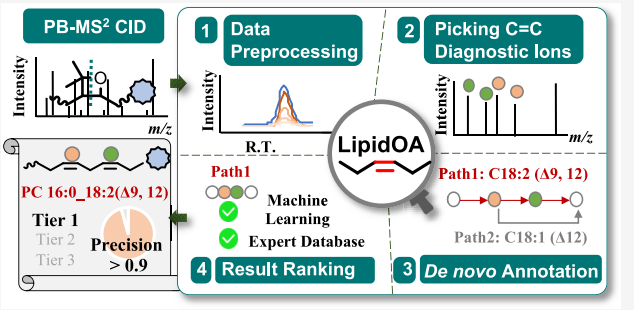
LipidOA: A Machine-Learning and Prior-Knowledge-Based Tool for Structural Annotation of Glycerophospholipids

Donghui Zhang, Qiaohong Lin, Tian Xia, Jing Zhao, Wenpeng Zhang, Zheng Ouyang,* and Yu Xia*

Cite This: *Anal. Chem.* 2022, 94, 16759−16767 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: The Paternò−Büchi (PB) reaction is a carbon−carbon double bond (C=C)-specific derivatization reaction that can be used to pinpoint the location(s) of C=C(s) in unsaturated lipids and quantitate the location of isomers when coupled with tandem mass spectrometry (MS/MS). As the data of PB-MS/MS are increasingly generated, the establishment of a corresponding data analysis tool is highly needed. Herein, LipidOA, a machine-learning and prior-knowledge-based data analysis tool, is developed to analyze PB-MS/MS data generated by liquid chromatography−mass spectrometry workflows. LipidOA consists of four key functional modules to realize an annotation of glycerophospholipid (GPL) structures at the fatty acyl-specific C=C location level. These include (1) data preprocessing, (2) picking C=C diagnostic ions, (3) *de novo* annotation, and (4) result ranking. Importantly, in the result-ranking module, the reliability of structural annotation is sorted via the use of a machine learning classifier and comparison to the total fatty acid database generated from the same sample. LipidOA is trained and validated by four PB-MS/MS data sets acquired using different PB reagents on mass spectrometers of different resolutions and of different biological samples. Overall, LipidOA provides high precision (higher than 0.9) and a wide coverage for structural annotations of GPLs. These results demonstrate that LipidOA can be used as a robust and flexible tool for annotating PB-MS/MS data collected under different experimental conditions using different lipidomic workflows.



Downloaded via PURDUE UNIV on May 31, 2024 at 17:01:13 (UTC).
See https://pubs.acs.org/sharingguidelines for options on how to legitimately share published articles.

INTRODUCTION

The compositions of lipids in biological systems are intricately regulated to maintain proper membrane fluidity[1] and realize diverse functions such as cell signaling[2] and energy storage.[3] The number, geometry, and location of carbon−carbon double bonds (C=Cs) incorporated in the fatty acyls are highly relevant to the physical-chemical property of unsaturated lipids.[1] Increasing evidence reveals that the compositions of lipid C=C location isomers vary in cancer cells or tissues due to dysregulation of key desaturation enzymes in lipid metabolism.[4,5] Lipidome-wide identification and quantitation of lipids at C=C location level thus can facilitate disease phenotyping and lipid marker discovery.[5,6] High-resolution mass spectrometry (HRMS) and tandem mass spectrometry (MS/MS) are routinely used to profile lipids at the sum composition and chain composition levels;[7,8] however, they fail to locate C=Cs in unsaturated lipids via low-energy collision-induced dissociation (CID). In the past decade, new gas-phase dissociation methods have been explored and applied in locating C=Cs in lipids, such as ozone-induced dissociation (OzID),[5,9,10] electron impact excitation of ions from organics (EIEIO),[11,12] ultraviolent photon dissociation (UVPD),[13,14] and so on.[15] Alternatively, C=C specific chemical derivatizations which promote cleavage at the modified site via CID are also found effective to locate C=C without requiring specialized mass spectrometers. These include the Paternò−Büchi (PB) reaction,[16−19] epoxidation,[20,21] singlet oxygen reaction,[22] and so on. Among them, the PB-MS/MS method has been extensively developed for different lipid analysis workflows including shotgun analysis,[16,17] liquid chromatography (LC)-MS,[6,23,24] MS imaging,[25,26] and direct analysis.[27,28]

Automated MS data analysis tools are instrumental to large-scale lipidomic profiling. LIPID MAPS hosts more than 40000 molecular structures of lipids, allowing matching and annotating the lipid at the sum composition level for MS data acquired on HRMS.[29] LipidBlast annotates MS/MS spectra and provides chain information by matching the experimental data to the database composed of both experimental and in silico generated MS/MS spectra (over 210000).[30] MS-DIAL 4 analyzes the LC retention time, collision cross-section, and MS/MS data of lipids; thus, it provides a wide coverage of lipids with a low false discovery

Received: August 11, 2022
Accepted: November 14, 2022
Published: November 22, 2022

ACS Publications © 2022 American Chemical Society

rate.[31] So far, data analysis tools are scarcely developed for annotating lipids at the C=C location level,[12,32] which impedes transfer of the newly developed MS/MS techniques mentioned earlier from expert research laboratories to the larger lipidomics society.

Herein, we aim to develop a data analysis tool for processing PB-MS/MS data acquired by LC-MS/MS workflows. Considering that the synthetic standards are too limited to establish a database for potential C=C location isomers of different classes of lipids, *de novo* identification of C=C is the most preferable approach. This is also because the fragmentation rule for generating the C=C diagnostic ions upon activating the PB products, termed PB-MS/MS, is straightforward and well established.[15,19] As illustrated in Figure 1, two regioisomers of the PB products, P1 and P2 (not considering the stereoisomers) are formed from the [2 + 2] cyclization between the PB reagent and a C=C in lipid. When subjected to MS/MS via CID, the four-membered oxetane ring in the PB products can fragment at the original C=C location and form the diagnostic ions, with one containing an aldehyde group ($^{\Delta}F_A$) and the other containing an olefin group ($^{\Delta}F_O$), in which Δ stands for the location of a C=C counting from ether/ester end of a fatty acyl chain. The *m*/*z* values of the C=C diagnostic fragment ions provide direct evidence for locating C=Cs, while the relative abundances of C=C diagnostic ions from distinct C=C location isomers provide the basis for quantitation. Based on these, Koff et al. expanded Mzmine2 to annotate the C=C diagnostic peaks in PB-MS/MS data acquired for diacylglyceryltrimethylhomoserines (DGTSs).[32] However, the location of C=C in a fatty acyl was not integrated with chain composition analysis and no scoring system was incorporated, leading to a liability of false identification. In this work, we have developed LipidOA for annotating the PB-MS/MS data of glycerophospholipids (GPLs) at the C=C location level, which is also capable of assessing the reliability of structural annotation via a machine-learning and prior-knowledge-based algorithm. The performance of LipidOA was trained and evaluated using data sets of bovine liver and porcine brain polar lipid extracts collected from different MS platforms, PB reagents, and workflows.

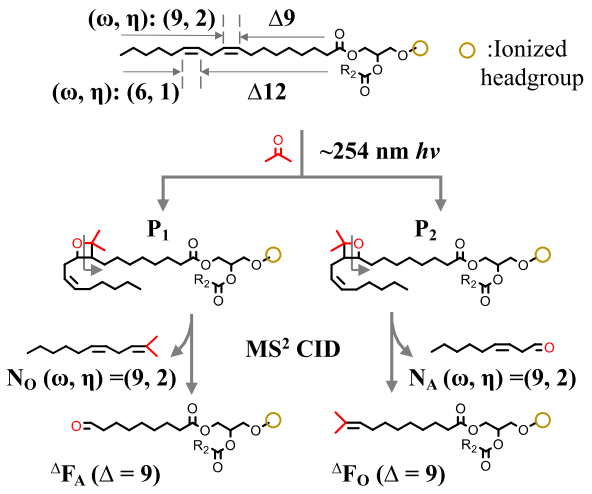


Figure 1. Schematic representation of acetone PB-MS^2 CID of a glycerophospholipid. The C=C diagnostic ions are represented by $^{\Delta}F_A$ and $^{\Delta}F_O$. The complementary neutral losses are represented by N_O and N_A, with "ω" and "η" denoting the location of C=C counting from the methyl end and the degree of unsaturation in the neutral losses, respectively.

EXPERIMENTAL SECTION

Nomenclature. Shorthand notations of lipids follows the guidelines suggested by Liebisch et al.[29] For instance, PC 18:0_18:1(Δ9) denotes a phosphatidylcholine (PC) consisting of a saturated 18-carbon fatty acyl chain and an 18-carbon fatty acyl chain of one degree of unsaturation (Ω = 1), in which Ω represents the degree of unsaturation in a fatty acyl chain. The underscore separator "_" indicates undefined *sn*-positions. The location of a C=C is defined by Δ-nomenclature, counting from the ester/ether end of the fatty acyl chain. The PB products of lipids are signified by "PB" as a superscript, such as in ^{PB}PC 34:1. The C=C diagnostic ions produced in PB-MS/MS are annotated as $^{\Delta}F_A$ and $^{\Delta}F_O$, while the subscripts "A" and "O" represent the cleavage sites that contain an aldehyde group and an olefin group, respectively. Note that, for *de novo* annotation, we also use "ω" to represent the C=C location counting from the methyl end and "η" as the degree of unsaturation in the neutral losses, namely, N_O and N_A, which are complementary to $^{\Delta}F_A$ and $^{\Delta}F_O$ (Figure 1). The chemical formula of these neutral losses can be predicted based on ω, η, and the PB reagent. As an example, when acetone is used as the PB reagent, the chemical formula for the neutral losses complementary to $^{i}F_A$ and $^{i}F_O$ are $C_{\omega}H_{2\omega-2\eta}O$ and $C_{\omega+3}H_{2\omega-2\eta+6}$, respectively.

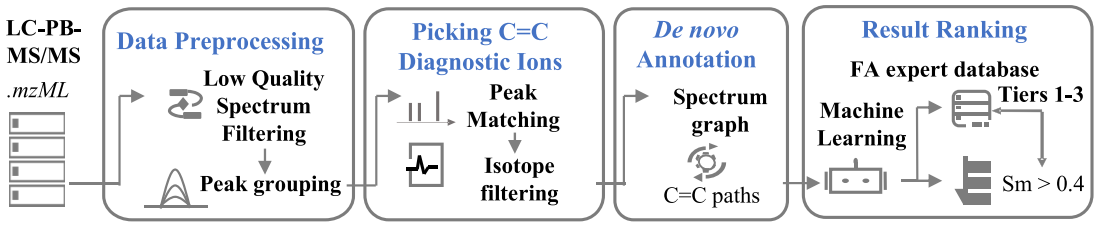
Chemicals. Lipid standards and lipid extracts were purchased from Avanti Polar Lipids (Alabaster, AL, U.S.A.). 2′,4′,6′-Trifluoroacetophenone (triFAP) and acetone were obtained from Aladdin Industrial Corporation (Shanghai, China). HPLC grade acetonitrile (ACN), methanol (MeOH), isopropanol (IPA), acetone, and water were purchased from Fisher Scientific Company (Ottawa, ON, Canada).

LC-MS. A 4500 QTRAP mass spectrometer (Sciex, Toronto, CA) and a trapped ion mobility-quadrupole time-of-flight mass spectrometer (timsTOF, Bruker Daltonics, Bremen, Germany) were used to collect LC-MS and PB-MS/MS data, respectively. Details of MS parameters and LC separation conditions are provided in Supplementary Note 1, SI.

PB-MS/MS Data Sets. Data sets 1 (lipid extracts of bovine liver) and 2 (lipid extracts of porcine brain) were collected in the targeted analysis mode using an online LC-PB-MS/MS system, consisting of hydrophilic-interaction liquid ion chromatography (HILIC) for lipid class separation, acetone as the PB reagent for online reaction, and a QTRAP 4500 mass spectrometer. Data set 1 contains data from four LC runs, while data set 2 contains data from five LC runs. For data sets 3 and 4, triFAP was used as the offline PB reagent for the analysis of lipid extracts from bovine liver. Data set 3 (14 LC runs) was collected in targeted analysis mode on a HILIC-MS system using QTRAP 4500 as the mass analyzer. Data set 4 was collected on a HILIC-MS system using timsTOF as the mass analyzer, in which PI and PG were methylated before the PB derivatization and analyzed in targeted analysis mode (4 LC runs), while PC, PE, and PS were analyzed via a modified data-dependent analysis mode (3 LC runs). Data set 1 was used for training and validating LipidOA, while the other three data sets were used to assess the performance of LipidOA.

https://doi.org/10.1021/acs.analchem.2c03505
Anal. Chem. 2022, 94, 16759−16767

Scheme 1. Major Function Modules in the Analysis Workflow of LipidOA



LipidOA. LipidOA possesses LC-PB-MS/MS data in *.ascii* or *.mzml* format. Data collected on QTRAP 4500 can be converted to *.mzML* format by MSConvert in Proteowizard,[33,34] and data from timsTOF can be exported in *.ascii* format by DataAnalysis from Bruker. LipidOA is written in Python 3.6 and currently available in Windows 8/10, with package pymzml used for *.mzML* data upload,[34] numpy for matrix operation,[35] and scikit-learn for training and validating machine learning models.[36] The graphic user interference is designed using PyQt5, while the line spline and scatter plot are written in pyEcharts. The user guidelines are provided in the Supporting Information. Current version of LipidOA can be found at https://cloud.tsinghua.edu.cn/f/8e52aa89604f4a1fbc67/?dl=1. Raw data of the four data sets have been uploaded to MetaboLights (MTBLS6075).[37]

RESULTS AND DISCUSSION

LipidOA contains four modules (Scheme 1): (1) data preprocessing, (2) picking C=C diagnostic ions, (3) *de novo* C=C annotation, and (4) result ranking. These modules facilitate the identification of C=C diagnostic ions in each PB-MS/MS spectrum, *de novo* structural annotation at the C=C location level with specific chain information, and assessment of the annotation results.

Data Preprocessing. In the data preprocessing module, the raw PB-MS/MS data are first processed by a Low-Quality Spectral Filtering (LQSF) method to remove low quality data. The low quality data are typically associated with lipids of low abundances in complex lipid mixtures, for example, GPLs in the sub-nM range or lower.[15] Taking PC 19:0_18:2 (PC 37:2) from Data set 1 as an example, its relative ion abundance is 3.2% of the most abundant PC in the mixture (PC 36:2), and no C=C diagnostic ions can be confidently distinguished from background noise in its PB-MS/MS spectrum (Figure S1, SI). This type of data will cost LipidOA extra time in analysis and increase the likelihood of a false positive for structural annotation. A spectral quality factor, M_q, which evaluates the dispersion of peak intensity from the average intensity, is calculated for each spectrum. Note that only fragment peaks with $S/N > 3$ are considered. The value of M_q ranges from 0 to 1, with higher value representing a spectrum of higher S/N. The detail for calculating M_q is provided in Supplementary Note 2 and Figure S2, SI.

By judging if at least one C=C diagnostic peak in a spectrum has $S/N > 3$ or not, the 156 PB-MS/MS spectra in data set 1 are manually categorized into a high-quality group (120 spectra) and a low-quality group (36 spectra). The M_q values of all above spectra are calculated, and the distribution is shown in Figure 2A. M_q of 13 low-quality spectra and 17 high-quality spectra fall in the range of 0.15 to 0.2. A threshold of M_q at 0.2 or higher would filter out more than 17% of high-quality spectra, thus significantly decreasing true positive in the latter step of *de novo* annotation. On the other hand, a threshold of M_q at 0.15 allows maintaining more than 90% (117/120) of high-quality ones while removing 19% (7 spectra) low-quality spectra, leading to a total of 146 spectra in database 1 after filtering. Therefore, M_q is set at 0.15 to remove low quality data in later studies. As an example, the M_q value of the PB-MS/MS spectrum of PC 19:0_18:2 (Figure S2, SI) is 0.12, lower than the threshold; thus, it is filtered out during this process.

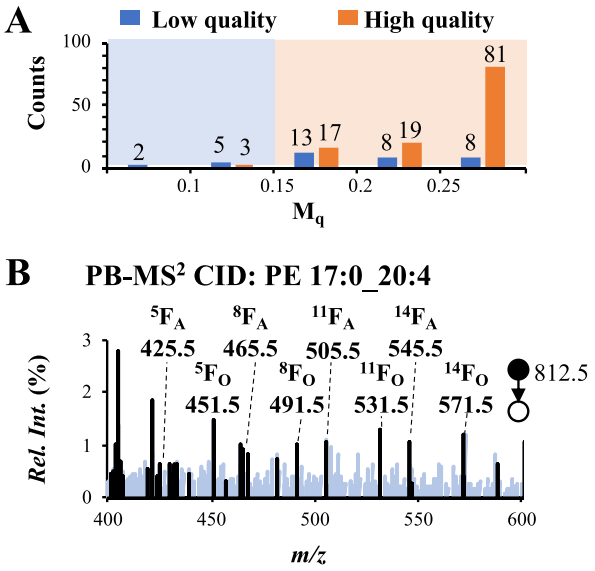


Figure 2. (A) Distribution of M_q values of the 156 PB-MS^2 CID spectra in data set 1. The threshold for removing low-quality spectra is set at $M_q \leq 0.15$. (B) The zoomed-in PB-MS^2 CID spectrum of PE 17:0_20:4 (m/z 400−600). Most of the interference peaks (blue lines) are filtered out after peak grouping. The remaining peaks (black lines) are mainly contributed by the C=C diagnostic ions.

After LQSF, peak grouping (PG) is used to align the precursor ions with the C=C diagnostic ions according to the similarity of their extracted ion chromatograms (EICs) in each PB-MS^2 CID spectrum, while PB-MS^1 data are not analyzed in the current workflow. This process can effectively reduce chemical interference and has been widely used in GC-MS[38] or LC-SWATH[39] data analysis. A non-negative linear regression is used:

$$Y(n) = a + bn + cX(n), \quad n = 1, 2, \ldots, k \tag{1}$$

where k is the number of individual spectra collected for a specific precursor ion in an LC-PB-MS/MS run, with n (an integer number) denoting the nth spectrum; $Y(n)$ represents the peak intensity of a specific fragment ion in the nth spectrum, while the precursor peak intensity in that spectrum is $X(n)$. The coefficients a, b, and c are values to be fitted in

https://doi.org/10.1021/acs.analchem.2c03505
Anal. Chem. 2022, 94, 16759−16767

A Predicated $^{\Delta}F_{A/O}$ (Δ= 7-12) of PC 18:0_18:1

ω	η	Δ	$^{\omega}N_A$ (Da)	$^{\Delta}F_O$ (*m/z*)	$^{\omega}N_O$ (Da)	$^{\Delta}F_A$ (*m/z*)
6	1	12	100.09	**746.57**	126.14	**720.52**
7	1	11	114.11	**732.55**	140.16	**706.50**
8	1	10	128.12	718.54	154.17	692.49
9	1	9	142.14	**704.52**	168.18	**678.48**
10	1	8	156.15	690.51	182.20	664.46
11	1	7	170.17	676.49	196.21	650.45

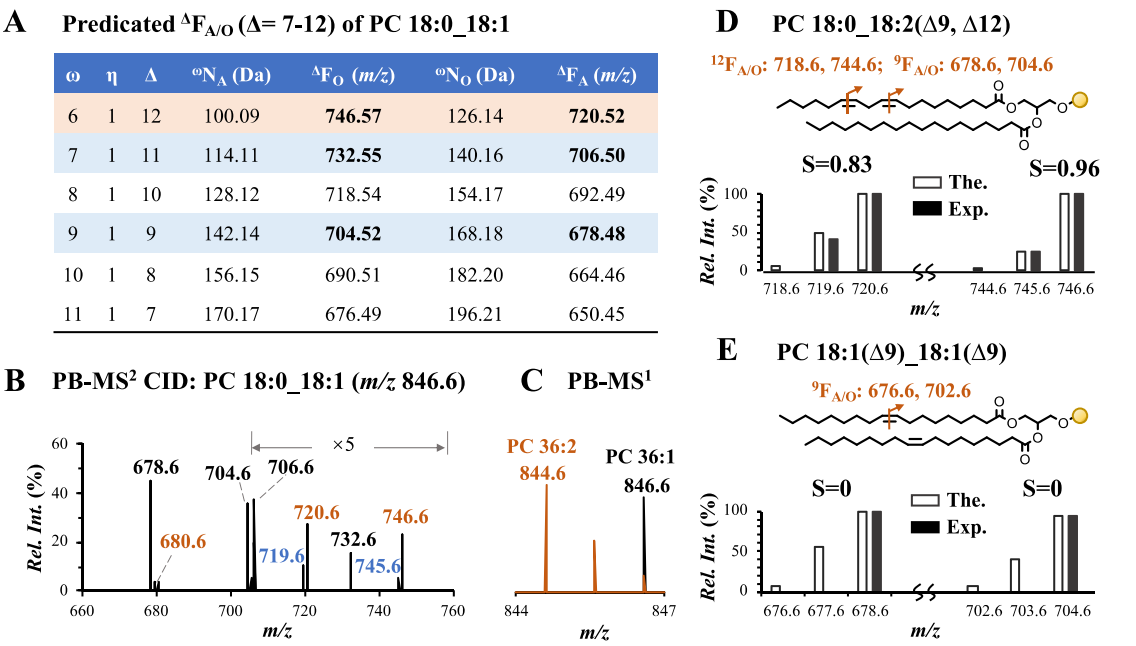


Figure 3. Picking C=C diagnostic ions in PB-MS² CID spectra. (A) Predicted masses of $^{\omega}N_{A/O}$ and *m/z* values of the complementary $^{\Delta}F_{O/A}$ ions in acetone PB-MS² CID of PC 18:0_18:1 with ω set from 6 to 11 and η kept at 1. (B) PB-MS² CID of *m/z* 846.6, which consists of PC 18:0_18:1 and the +2 Da isotope of PC 18:0_18:2 and PC 18:1_18:1 (positive ion mode, data set 1). (C) An overlap of isotopes from the PB products of PC 36:2 (*m/z* 844.6) to that of PC 36:1 (*m/z* 846.6). Comparisons of the theoretical isotopic distributions of (D) $^{12}F_{A/O}$ resulting from the +2 Da isotope of PC 18:0_18:2 to the experimental data and (E) $^{9}F_{A/O}$ resulting from the +2 Da isotope of PC 18:1_18:1 to the experimental data.

linear regression. The linear regression contains two parts, with "$a + bn$" representing the contribution from a linear baseline in EICs of the fragment ions and $cX(n)$ denoting the contribution of EIC from the PB precursor.[40] Therefore, coefficient c measures the similarity between the fragment peaks and the precursor; the larger the c value suggests a higher similarity between the EICs of the precursor and its fragment ions. Different from Pearson correlation and cosine similarity, peak grouping can analyze the similarity of EIC profiles between the precursor and its fragments while considering the baseline intensity simultaneously.[35] Details of the peak grouping algorithm and threshold selection are provided in Supplementary Note 3 and Figure S3, SI. By keeping fragment ions of the top 10% c values of a spectrum, a majority of chemical interferences can be removed. A relatively low abundance lipid, PE 17:0_20:4 (3.5% of the most abundant PE in data set 1) is used as an example to illustrate the typical performance of PG for the PB-MS² CID data. Figure 2B compares the grouped PB-MS/MS spectrum (black trace) to the PB-MS/MS spectrum without peak grouping (light blue trace) of PE 17:0_20:4. According to the fragmentation rules for PEs in PB-MS/MS, C=C diagnostic ions are expected to appear in *m/z* 400−600;[6] therefore, additional peaks in this range should be contributed by chemical interferences. Peak grouping leads to a dramatic decrease of the count of interference peaks from 192 to 22. The four pairs of C=C diagnostic peaks are all preserved after peak grouping, including $^{14}F_{A/O}$, $^{11}F_{A/O}$, $^{8}F_{A/O}$, and $^{5}F_{A/O}$. Moreover, a low abundance C=C diagnostic peak at *m/z* 465.5 (0.6% relative peak intensity) is preserved after peak grouping, while chemical interferences of higher relative intensities (>0.6%) are successfully removed. The above data demonstrate that peak grouping is useful for keeping PB-MS/MS diagnostic peaks than simply using a threshold according to the relative ion intensity.

Picking C=C Diagnostic Ions. The mass differences between the precursor ions (charged PB products) and the C=C diagnostic ions ($F_{A/O}$) are equal to the masses of the complementary neutral losses ($N_{O/A}$), which can be directly measured in a PB-MS/MS spectrum. The chemical formula and masses of N_O and N_A can be predicted based on the structure of the PB reagent and the paired (ω, η) value of a specific neutral loss. Note that ω stands for the location of C=C counting from the methyl end and η represents the degree of unsaturation in these neutral losses. For example, when acetone is used as the PB reagent, the chemical formula of the neutral losses complementary to $^{\Delta}F_A$ and $^{\Delta}F_O$ are $^{\omega}N_O$ ($C_{\omega}H_{2\omega+2-2\eta}O$) and $^{\omega}N_A$ ($C_{\omega+3}H_{2\omega-2\eta+8}$), respectively (Figure 1). Considering that fatty acyls with a carbon number ≤24 and a degree of unsaturation ≤6 are more commonly found in literature,[29] the values of ω and η are iterated from 1 to 24 and 1 to 6, respectively, to predict all possible pairs of $^{\omega}N_{A/O}$ and calculate the *m/z* values of corresponding $^{\Delta}F_{O/A}$ derived from protonated PB product ions. Figure 3A lists the predicted masses of $^{\omega}N_{A/O}$ and *m/z* values of $^{\Delta}F_{A/O}$ for PC 18:0_18:1 with ω ranging from 6 to 11 and η kept at 1. The mass tolerance for the low-resolution data is set at 0.5 Da. By matching the predicted *m/z* of the diagnostic ions to the PB-MS² CID of PC 18:0_18:1 (*m/z* 846.5) in data set 1, three pairs of C=C diagnostic peaks, $^{9}F_{A/O}$ (*m/z* 678, 704), $^{11}F_{A/O}$ (*m/z* 706, 732), $^{12}F_{A/O}$ (*m/z* 720, 746), are detected in the spectrum (Figure 3B). Consequently, one might draw a hasty conclusion that three C=C location isomers, namely, Δ12, Δ11, and Δ9, coexist for PC 18:0_18:1. However, cautions should be given to data acquired by HILIC-MS or shotgun workflow because the unsaturated lipid of interest often coelutes with its analogue carrying one more degree of unsaturation. The C=C diagnostic ions derived from the +2 Da isotope of latter species cause chemical interference and

https://doi.org/10.1021/acs.analchem.2c03505
Anal. Chem. 2022, 94, 16759−16767

can lead to erroneous assignment for the location of C=C.[41] Given the above considerations, LipidOA incorporates an isotope filtering function. First, the theoretical isotopic distributions of individual C=C diagnostic ions ($^{\Delta}F_{O/A}$) are calculated for the lipid analogue having one more degree of unsaturation. The experimental distribution is compared against the theoretical isotopic distribution (Supplementary Note 4, SI). The contribution by the +2 Da isotope to the experimental data is represented by *S*, the value of which ranges from 0 to 1, with higher value indicating a higher contribution from the +2 isotope.

The PB-MS^2 CID data of PC 18:0_18:1 (data set 1) is an example to illustrate the performance of isotope filtering. Inspection of the MS^1 spectrum of the PB product ions shows that PC 36:1 (*m*/*z* 846.6) contains 18% of the +2 Da isotope of PC 36:2 (Figure 3C, brown trace). MS^2 CID of the underivatized PC 36:2 in negative ion mode suggests it contains PC 18:0_18:2 (major) and PC 18:1_18:1 (minor). The isotopic distributions of all detected C=C diagnostic ions in Figure 3B are matched against that of the theoretical distribution generated from the +2 Da isotopes of PC 36:2 (*m*/*z* 846.4). For the $^{12}F_{A/O}$ ions (*m*/*z* 720, 746), the +2 Da isotope of PC 18:0_18:2 (open column) makes significant contribution to the experimental data (filled columns) (*S* = 0.83 and 0.96) (Figure 3D). These data suggest that $^{12}F_{A/O}$ ions should not be considered as diagnostic ions generated by PC 18:0_18:1. Note that if $^{12}F_{A/O}$ ions are derived from the monoisotope of PC 18:0_18:2 they should be at *m*/*z* 718 and 744 (inset of Figure 3D). On the other hand, the contributions of the +2 Da isotope of PC 18:0_18:2 and PC 18:1_18:1 show negligible contributions to the other two pairs of diagnostic ions. For instance, the *S* values are both 0 by calculating the contributions of +2 Da isotopes of PC 18:1_18:1 to $^{9}F_{A/O}$ (*m*/*z* 704, 678, Figure 3E). Distribution of the *S* values of manually confirmed annotations in data set 1 is provided in Figure S4, SI. More than 97% (214/220) of the *S* values are smaller than 0.2 and the rest falls within 0.2 to 0.4. An inspection of the latter shows that they contain contributions from the lipid of interest and its analog of one more degree of unsaturation. For example, PE 18:0_20:4 (Δ5, 8, 11, 14) coelutes with PE 18:0_20:3 (Δ8, 11, 14). The C=C diagnostic ions derived from the +2 Da isotope of PE 18:0_20:4 are only 0.009 Da smaller than PE 18:0_20:3 (Figure S5, SI). To distinguish these two types of C=C diagnostic ions, it requires mass resolving power above 110,000, which is beyond the performance of MS instruments used in this study. Consequently, the C=C diagnostic ions of PE 18:0_20:3 contain a certain portion of PE 18:0_20:4 which leads to an average *S* value of 0.25 (Figure S5, SI). For this kind of situation, the C=C diagnostic ions should be consider as being derived from PE 18:0_20:3, instead of being rejected. Based on above consideration, the threshold of *S* is set at 0.4 for isotope filtering in later studies.

***De novo* Annotation.** After picking C=C diagnostic ions, the spectrum graph is used as a *de novo* approach to reconstruct each unsaturated chain back to the lipid structure. The spectrum graph is a directed acyclic graph that has been used in *de novo* peptide sequencing.[42,43] In this work it has been revised for lipid analysis. A spectrum graph contains a starting vertex (V_S), an ending vertex (V_E), principle vertices ($V_{(\omega, \eta)}$), and directed edges connecting between neighboring vertices. The (ω, η) of a principle vertex is derived from each identified C=C diagnostic ion pairs (Figure 3A). A path is constructed by linking all edges between V_S and V_E, where the number of principle vertices in each path is equal to the degree of unsaturation in the fatty acyl chain (Ω). In a path the principal vertices are connected from V_S in an increasing order of paired ω and η values, such as from V_S to $V_{(\omega, 1)}$, $V_{(\omega+3, 2)}$, $V_{(\omega+6, 3)}$, ..., V_E. The (ω, η) relationship of neighboring principle vertices is defined as such because C=Cs are typically separated by a methylene group in polyunsaturated fatty acyls. In brief, the spectrum graph just generates all possible unsaturated fatty acyls which would produce the detected C=C diagnostic ions.

According to the PB-MS^2 CID data of PC 18:0_18:1 (Figure 3B), two principle vertices, $V_{(7, 1)}$ and $V_{(9, 1)}$, can be extracted. Subsequently, a spectral graph is constructed, which contains two paths, V_S-$V_{(7, 1)}$-V_E and V_S-$V_{(9, 1)}$-V_E, suggesting the existence of two C=C location isomers in monounsaturated fatty acyl chain (Ω = 1). By combining the chain composition information, two isomers PC 18:0_18:1(Δ11) and PC 18:0_18:1(Δ9), are annotated (Figure 4A). Figure 4B

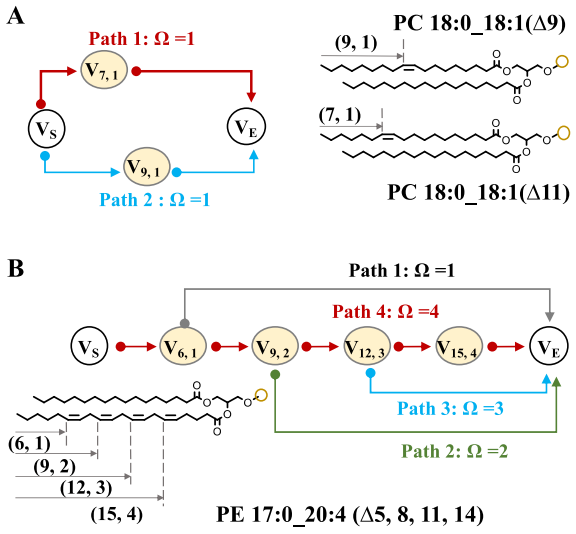


Figure 4. *De novo* annotation of unsaturated GPLs down to C=C location via the spectrum graph method for (A) PC 18:0_18:1 and (B) PE 17:0_20:4.

demonstrates how a spectrum graph is constructed for a polyunsaturated chain. Four principle vertices, $V_{(6, 1)}$, $V_{(9, 2)}$, $V_{(12, 3)}$, $V_{(15, 4)}$, are identified from PB-MS^2 CID data of PE 37:4 (Figure 2B). Four paths are constructed with an increasing degree of unsaturation of a fatty acyl chain (Ω from 1 to 4): Path 1: V_S-$V_{(6, 1)}$-V_E; Path 2: V_S-$V_{(6, 1)}$-$V_{(9, 2)}$-V_E; Path 3: V_S-$V_{(6, 1)}$-$V_{(9, 2)}$-$V_{(12, 3)}$-V_E; Path 4: V_S-$V_{(6, 1)}$-$V_{(9, 2)}$-$V_{(12, 3)}$-$V_{(15, 4)}$-V_E. Chain composition analysis via MS^2 CID in negative ion mode shows that PE 37:4 is primarily contributed by PE 17:0_20:4. Thus, Path 4 is considered as the most likely configuration for C20:4 and the lipid is annotated as PE 17:0_20:4 (Δ5, 8, 11, 14) (Figure 4B).

Result Ranking. Matching experimental MS/MS spectra with those in the database is a widely applied method to score or rank structural annotations. Given that GPL standards having different C=C locations are very limited, it is impossible to generate a PB-MS/MS database of reasonable size. Therefore, the data sets from complex mixtures which multiple experts from our group have carefully annotated are

https://doi.org/10.1021/acs.analchem.2c03505
Anal. Chem. 2022, 94, 16759−16767

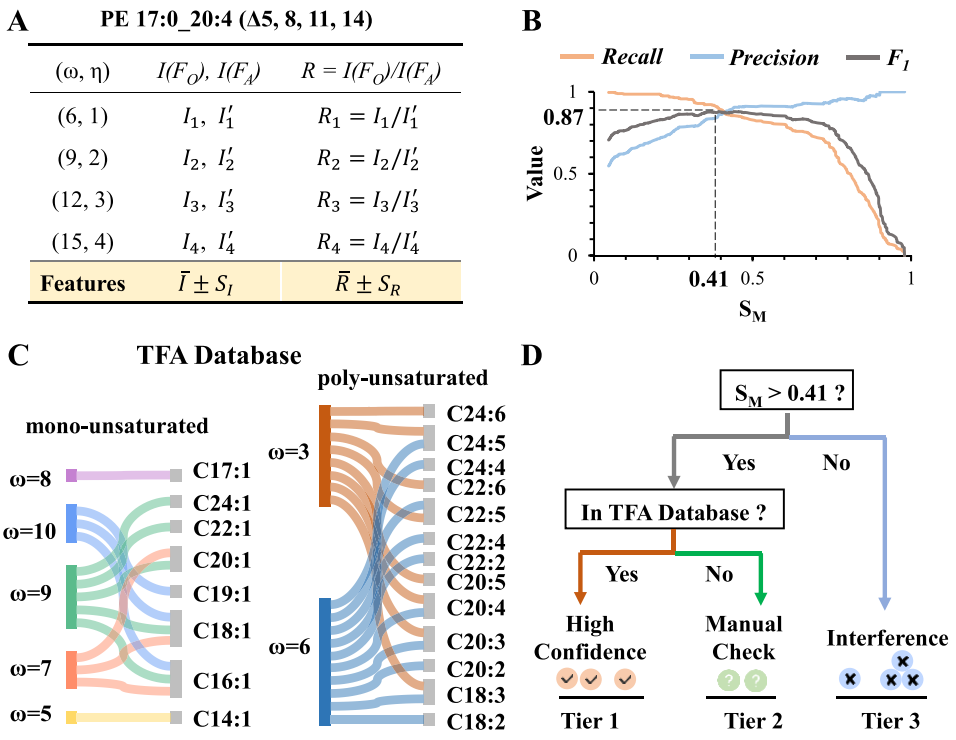


Figure 5. Machine-learning and prior-knowledge based ranking algorithm in LipidOA. (A) Features generated from peak intensities of C=C diagnostic ions, using PE 17:0_20:4 (Δ5, 8, 11, 14) as an example. (B) Plots of precision, recall, and F_1 as a function of S_M in the training data set. (C) A prior knowledge database containing manually identified total fatty acids of bovine liver lipid extracts. (D) A ranking algorithm by combining the use of a Random Forest classifier and a TFA database.

used as data sets. In those data sets, the manual identifications are used as labels to train a machine learning algorithm. The trained machine learning module is then used to score the accuracy of annotated C=C diagnostic peaks in PB-MS/MS. In data set 1, 820 paths are generated from 146 PB-MS/MS spectra resulting from data preprocessing. Among them, 365 and 455 paths are categorized as true positive and negative annotations based on manual identifications, respectively. Combining the chain information obtained from MS^2 CID in negative ion mode, 1068 lipid structures can be annotated while only 220 are true positives. Because data set 1 (as well as other data sets) contains data from multiple LC runs, these identifications correspond to 87 distinct PC and PE molecules. Therefore, it is highly necessary to develop a ranking program to improve annotation accuracy. The Result Ranking module in LipidOA contains a machine learning classifier to score the paths. It also compares the results to an established database of the total fatty acids which are experimentally obtained from the same sample with C=C location manually confirmed.

PE 17:0_20:4(Δ5, 8, 11, 14) is used as an example to demonstrate how the features are extracted from the C=C diagnostic ions in each PB-MS/MS spectrum for machine learning. First, a list of peak intensities of F_O (I_η) and F_A (I'_η) identified in each path is generated and the ratios ($R_\eta = I_\eta/I'_\eta$) are calculated for each principle vertex (Figure 5A). These values are then transformed into four features: (1) $\bar{I}$, the average of all I_η and I'_η in the path, (2) S_I, the standard derivation of $\bar{I}$, (3) $\bar{R}$, the average of R_η in the path, and (4) S_R, the standard deviation of $\bar{R}$. The values of $\bar{I}$, S_I, $\bar{R}$, and S_R for PE 17:0_20:4 (Δ5, 8, 11, 14) are 0.011, 0.002, 1.4, and 0.5, respectively. Selection of these four features is based on two considerations. First, the higher the peak intensities ($\bar{I}$) of the C=C diagnostic ions, the higher confidence can be achieved for annotation of C=C location. Second, it is empirically observed that R_η of each principle vertex of the same path are close to each other than the wrongly selected diagnostic peaks, leading to a smaller S_R for C=Cs which are in the same fatty acyl chain.

The features from 80% of paths in data set 1 are used as data objects and the manual identifications are used as labels to train three chosen machine learning classifiers, including Naïve Bayes,[44] Support Vector Machines,[45] and Random Forest.[46] True positive (TP), true negative (TN), false positive (FP), and false negative (FN) are calculated under a specific threshold. Precision ($p = \frac{TP}{TP+FP}$), recall ($r = \frac{TP}{TP+FN}$), and F_1 ($F_1 = \frac{2^*p^*r}{p+r}$) are used as performance metrics.[47] Among the three machine learning classifiers, Random Forest outperforms the other two in terms of accuracy (Supplementary Note 5, Figures S6 and S7, SI). A score of Random Forest (S_M) which falls between 0 and 1, is then calculated for each path. S_M is associated with the confidence of annotation, with larger values representing higher possibilities of true positive. A threshold of S_M at 0.41 allows achieving the same values (0.87) of precision and recall in the training data set (Figure 5B). By using this threshold value good performance in precision, recall, and F_1 (all at 0.83) are also obtained in the testing data set. These results demonstrate that by setting S_M at 0.41, Random Forest provides reliable annotation of C=C diagnostic peaks with

https://doi.org/10.1021/acs.analchem.2c03505
Anal. Chem. 2022, 94, 16759−16767

good precision and recall (>0.8), while overfitting of the algorithm is not of a concern.

A prior knowledge database is established by manually annotating the PB-MS/MS data of total fatty acids of the same lipidome.[48] The total fatty acid (TFA) database of bovine liver lipid extracts is shown in Figure 5C, which consists of 31 C=C location isomers ranging from C14:1 to C24:6. During the process of Result Ranking, the annotation results are sorted in three tiers based on Random Forest classification and fatty acyl search in the TFA database (Figure 5D). Paths with $S_M \leq 0.41$ are placed in Tier 3, most likely resulting from chemical interferences, and will not be further considered. Paths with $S_M > 0.41$ are further matched against the TFA database. The ones being found in the TFA database are placed in Tier 1 and considered as high confidence annotations. For those are not found in the TFA database, they are placed in Tier 2, which requires manual interpretation to check whether they are false positives.

By applying such a ranking procedure to data set 1, LipidOA successfully removes 96% false positive annotations (from a total of 848); thus, it places 189 true positive and 7 false positive annotations in Tier 1, resulting in a precision of 0.96 and recall of 0.86. Tier 2 contains 26 annotations, which all proved false positive results. Tier 3 contains mostly false positive results (815) and 31 true positive results (out 220). The numbers of (1) spectra being processed, (2) paths (Vs to Ve) being generated in spectrum graphs, and (3) lipid identifications before and after each step of LipidOA processing are summarized in Table S1, SI. It is important to emphasize that LipidOA fills the gap from annotating C=C diagnostic ions to annotating C=C location in a specific acyl chain with high accuracy. This function is not achieved by Mzmine 2 which also aims to process the PB-MS/MS data (Supplementary Note 6 and Figure S8, SI).[32]

Application of LipidOA to Different PB-MS/MS Data Sets. The performance of LipidOA was further tested with three other data sets, DS2-4 (Figure 6A), in which the PB-MS/MS data were acquired from different biological samples (bovine liver vs porcine brain), using different PB reagents (acetone vs triFAP), and different mass spectrometers (low resolution vs high resolution). All PB-MS/MS data were collected via targeted analysis except for PC, PE, and PS in DS4-2, which were collected via a modified data-dependent acquisition (DDA) method. In this method, the limit on the number of times that a precursor can be fragmented was removed; the PB products can thus be fragmented in continuous cycles and the EICs can be obtained for the PB-MS/MS data. All data sets have been manually interpreted for C=C location and these results are considered true positives.

Good precisions (>0.9) and reasonable recall values (0.7−0.95) are obtained for annotations in Tier-1 of DS2, DS3, and DS4-1, similar to that achieved for DS1 (Figure 6B). The lower recall value in DS2 (0.72) is caused by a severe overlap of HILIC retention time between PE and PS in porcine brain lipid extracts, which decreases the effectiveness of Peak Grouping. However, low abundance isomers of PC 34:1 and PS 36:1, such as the n-10 C=C location isomers of C18:1 are identified with high confidence in addition to the more frequently reported n-7 and n-9 C=C location isomers. On the other hand, PS presents in very low abundance in bovine liver extracts, thus having a negligible impact on the annotation of PE in the other data sets. DS4-2 comprises PB-MS/MS data of PC, PE, and PS, collected via the modified DDA. High

A

Datasets	Applied	Sample	MS	PB reagent
DS1	TR, VA	B (PC, PE)	LR	acetone
DS2	TE	P (PC, PE, PS)	LR	acetone
DS3	TE	B (PC, PE)	LR	triFAP
DS4-1	TE	B (PG, PI)	HR	triFAP
DS4-2	TE	B (PC, PE, PS)	HR	triFAP

TR: training; VA: validating; TE: testing; B: bovine liver; P: porcine brain; LR: low resolution; HR: high resolution

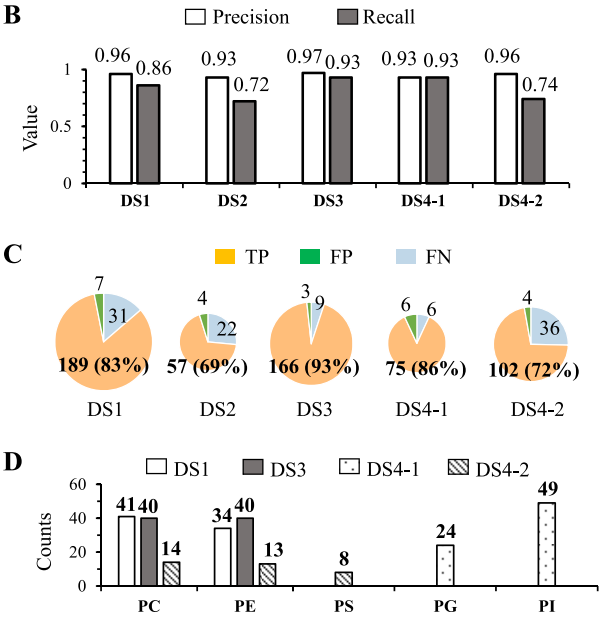


Figure 6. (A) Data sets used for training, validating, and testing LipidOA. (B) Precision and recall for Tier 1 identification of each data set by LipidOA. (C) Pie chart of the number of structural annotations categorized in true positive (TP), false positive (FP), and false negative (FN) of each data set made by LipidOA. (D) The number of distinct unsaturated lipid structures annotated by LipidOA for DS1, DS3, DS4-1, and DS4-2.

precision (0.96) is achieved; however, the recall value is lower (0.74). This lower recall value results from the fact that lipids of low ion abundance are acquired with fewer data points via the modified DDA. This downgrades the performance of Peak Grouping which requires at least 9 MS/MS data points to construct EICs. Overall, LipidOA provides 613 structural annotations (589 true positives and 24 false positives) across five classes of GPLs from the four data sets, with a precision higher than 0.9 achieved. The itemized precision and recall values for each class of GPLs are provided in Figure S9, SI.

Figure 6C shows the pie-chart presentation of Tier-1 structural annotations in each data set, categorized into true positive, false positive, and false negative annotations, respectively. The true positive annotations are close to or higher than 70%, suggesting that LipidOA can provide accurate and good coverage of identification under various experimental conditions.

We also summarize the number of distinct structure identifications in bovine liver lipid extracts from true positive results in Tier 1 in DS1, DS3, DS4-1, and DS4-2. As shown in Figure 6D, the number of identified PC and PE are close between DS1 and DS3 (41 PCs, 34 PEs in DS1 and 40PCs, 40PEs in DS3) via the use of acetone and triFAP as the PB

https://doi.org/10.1021/acs.analchem.2c03505
Anal. Chem. 2022, 94, 16759−16767

reagents, respectively. The identity of the lipids from both data sets shows more than 80% overlap. A couple of more polyunsaturated PCs and PEs are identified from DS3.

The number of identifications is lower in DS4-2 because of DDA cannot be performed on the low abundance lipids. DS4-1 hosts targeted analysis data specialized for PG and PI; therefore, the number of identified structures provided by LipidOA is identical to those made by manual inspection. All above structural annotations as well as associated MS data are curated into a database, which is open to public search and comparison (Figure S10 and in "An Introduction to LipidOA", SI).

CONCLUSIONS

In this work, LipidOA, a machine-learning and prior-knowledge-based tool, has been developed to analyze PB-MS/MS data and annotate structures of GPLs down to the C=C location level. LipidOA carries two distinct traits. First, it reconstructs single or multiple C=Cs back to a specific fatty acyl chain via a *de novo* approach, thus allowing the detection of novel lipid species which may not have been reported before. Second, it contains a ranking system that sorts many annotations in three tiers and effectively places high precision (>0.9) annotations in Tier 1. The latter function is made possible by combining a machine learning classifier and a total fatty acid database of the same lipidome. It is also robust and flexible, which can be applied to analyze PB-MS/MS data collected on different MS instruments and using different workflows. As shown in this study, the performance of LipidOA is comparable to the annotations obtained by manual interpretation from experts. We expect that LipidOA will aid in the identification of low abundance lipid C=C location isomers which have been less frequently reported before. LipidOA is just a first step toward establishing a data analysis tool compatible with detailed structural analysis of lipids. In future work, the function of LipidOA will be expanded to accommodate user-defined PB reagents, a wider variety of lipid classes, and data collected from HRMS. More importantly, a function for relative quantitation of lipid C=C location isomers will be incorporated to facilitate lipidomic phenotyping and applications in biomedical research.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.analchem.2c03505.

Details of coupling the PB reaction with various LC-MS/MS workflows; low quality spectra filtering; peak grouping; isotope filtering; machine learning; an example of structural annotation by Mzmine 2; performance of LipidOA; an introduction to LipidOA (PDF)

AUTHOR INFORMATION

Corresponding Authors

Yu Xia − *MOE Key Laboratory of Bioorganic Phosphorus Chemistry and Chemical Biology, Department of Chemistry, Tsinghua University, Beijing 10084, China;* orcid.org/0000-0001-8694-9900; Email: xiayu@mail.tsinghua.edu.cn

Zheng Ouyang − *State Key Laboratory of Precision Measurement Technology and Instruments, Department of Precision Instrument, Tsinghua University, Beijing 100084, China;* orcid.org/0000-0002-7932-0933; Email: ouyang@mail.tsinghua.edu.cn

Authors

Donghui Zhang − *State Key Laboratory of Precision Measurement Technology and Instruments, Department of Precision Instrument, Tsinghua University, Beijing 100084, China; MOE Key Laboratory of Bioorganic Phosphorus Chemistry and Chemical Biology, Department of Chemistry, Tsinghua University, Beijing 10084, China;* orcid.org/0000-0002-0358-3040

Qiaohong Lin − *MOE Key Laboratory of Bioorganic Phosphorus Chemistry and Chemical Biology, Department of Chemistry, Tsinghua University, Beijing 10084, China*

Tian Xia − *MOE Key Laboratory of Bioorganic Phosphorus Chemistry and Chemical Biology, Department of Chemistry, Tsinghua University, Beijing 10084, China*

Jing Zhao − *MOE Key Laboratory of Bioorganic Phosphorus Chemistry and Chemical Biology, Department of Chemistry, Tsinghua University, Beijing 10084, China*

Wenpeng Zhang − *State Key Laboratory of Precision Measurement Technology and Instruments, Department of Precision Instrument, Tsinghua University, Beijing 100084, China;* orcid.org/0000-0002-7018-930X

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.analchem.2c03505

Author Contributions

All authors have given approval to the final version of the manuscript.

Notes

The authors declare the following competing financial interest(s): Z.O. is the founder of PURSPEC Technologies, Inc., which is developing the miniature mass spectrometry system and an Omega PB reactor.

ACKNOWLEDGMENTS

Financial support from National Key R&D Program of China (2018YFA0800903) and the National Natural Science Foundation of China (Nos. 21874081, 21934003) are greatly appreciated.

REFERENCES

(1) Harayama, T.; Riezman, H. *Nat. Rev. Mol. Cell Biol.* **2018**, *19*, 281−296.
(2) Shevchenko, A.; Simons, K. *Nat. Rev. Mol. Cell Biol.* **2010**, *11*, 593−598.
(3) Olzmann, J. A.; Carvalho, P. *Nat. Rev. Mol. Cell Biol.* **2019**, *20*, 137−155.
(4) Vriens, K.; et al. *Nature* **2019**, *566*, 403−406.
(5) Young, R. S. E.; Bowman, A. P.; Williams, E. D.; Tousignant, K. D.; Bidgood, C. L.; Narreddula, V. R.; Gupta, R.; Marshall, D. L.; Poad, B. L. J.; Nelson, C. C.; Ellis, S. R.; Heeren, R. M. A.; Sadowski, M. C.; Blanksby, S. J. *Cell Rep.* **2021**, *34*, 108738.
(6) Zhang, W.; Zhang, D.; Chen, Q.; Wu, J.; Ouyang, Z.; Xia, Y. *Nat. Commun.* **2019**, *10*, 79.
(7) Cajka, T.; Fiehn, O. *TrAC, Trends Anal. Chem.* **2014**, *61*, 192−206.
(8) Yang, K.; Han, X. *Trends Biochem. Sci.* **2016**, *41*, 954−969.
(9) Thomas, M. C.; Mitchell, T. W.; Harman, D. G.; Deeley, J. M.; Nealon, J. R.; Blanksby, S. J. *Anal. Chem.* **2008**, *80*, 303−311.
(10) Claes, B. S.; Bowman, A. P.; Poad, B. L.; Young, R. S.; Heeren, R. M.; Blanksby, S. J.; Ellis, S. R. *Anal. Chem.* **2021**, *93*, 9826−9834.

(11) Baba, T.; Campbell, J. L.; Le Blanc, J. C. Y.; Baker, P. R. S. *J. Lipid Res.* **2016**, *57*, 2015−2027.
(12) Baba, T.; Campbell, J. L.; Le Blanc, J. C. Y.; Baker, P. S.; Ikeda, K. *J. Lipid Res.* **2018**, *59*, 910−919.
(13) Williams, P. E.; Klein, D. R.; Greer, S. M.; Brodbelt, J. S. *J. Am. Chem. Soc.* **2017**, *139*, 15681−15690.
(14) Macias, L. A.; Garza, K. Y.; Feider, C. L.; Eberlin, L. S.; Brodbelt, J. S. *J. Am. Chem. Soc.* **2021**, *143*, 14622−14634.
(15) Zhang, W.; Jian, R.; Zhao, J.; Liu, Y.; Xia, Y. *J. Lipid Res.* **2022**, *63*, 100219.
(16) Ma, X.; Chong, L.; Tian, R.; Shi, R.; Hu, T. Y.; Ouyang, Z.; Xia, Y. *Proc. Natl. Acad. Sci. U. S. A.* **2016**, *113*, 2573−2578.
(17) Cao, W.; Cheng, S.; Yang, J.; Feng, J.; Zhang, W.; Li, Z.; Chen, Q.; Xia, Y.; Ouyang, Z.; Ma, X. *Nat. Commun.* **2020**, *11*, 375.
(18) Ma, X.; Xia, Y. *Angew. Chem.* **2014**, *53*, 2592−2596.
(19) Ma, X.; Zhang, W.; Li, Z.; Xia, Y.; Ouyang, Z. *Acc. Chem. Res.* **2021**, *54*, 3873−3882.
(20) Feng, Y.; Chen, B.; Yu, Q.; Li, L. *Anal. Chem.* **2019**, *91*, 1791−1795.
(21) Kuo, T. H.; Chung, H. H.; Chang, H. Y.; Lin, C. W.; Wang, M. Y.; Shen, T. L.; Hsu, C. C. *Anal. Chem.* **2019**, *91*, 11905−11915.
(22) Zhou, Y.; Park, H.; Kim, P.; Jiang, Y.; Costello, C. E. *Anal. Chem.* **2014**, *86*, 5697−5705.
(23) Lin, Q.; Li, P.; Fang, M.; Zhang, D.; Xia, Y. *Anal. Chem.* **2022**, *94*, 820−828.
(24) Ren, H.; Triebl, A.; Muralidharan, S.; Wenk, M. R.; Xia, Y.; Torta, F. *Analyst* **2021**, *146*, 3899−3907.
(25) Bednařík, A.; Bölsker, S.; Soltwisch, J.; Dreisewerd, K. *Angew. Chem.* **2018**, *57*, 12092−12096.
(26) Wäldchen, F.; Spengler, B.; Heiles, S. *J. Am. Chem. Soc.* **2019**, *141*, 11816−11820.
(27) Deng, J.; Yang, Y.; Liu, Y.; Fang, L.; Lin, L.; Luan, T. *Anal. Chem.* **2019**, *91*, 4592−4599.
(28) Zhang, W.; Chiang, S.; Li, Z.; Chen, Q.; Xia, Y.; Ouyang, Z. *Angew. Chem.* **2019**, *131*, 6125−6130.
(29) Liebisch, G.; Fahy, E.; Aoki, J.; Dennis, E. A.; Durand, T.; Ejsing, C. S.; Fedorova, M.; Feussner, I.; Griffiths, W. J.; Köfeler, H.; Merrill, A. H., Jr.; Murphy, R. C.; O'Donnell, V. B.; Oskolkova, O.; Subramaniam, S.; Wakelam, M. J. O.; Spener, F. *J. Lipid Res.* **2020**, *61*, 1539−1555.
(30) Kind, T.; Liu, K. H.; Lee, D. Y.; Defelice, B.; Meissen, J. K.; Fiehn, O. *Nat. Methods* **2013**, *10*, 755−758.
(31) Tsugawa, H.; Ikeda, K.; Takahashi, M.; Satoh, A.; Mori, Y.; Uchino, H.; Okahashi, N.; Yamada, Y.; Tada, I.; Bonini, P.; Higashi, Y.; Okazaki, Y.; Zhou, Z.; Zhu, Z. J.; Koelmel, J.; Cajka, T.; Fiehn, O.; Saito, K.; Arita, M.; Arita, M. *Nat. Biotechnol.* **2020**, *38*, 1159−1163.
(32) Korf, A.; Jeck, V.; Schmid, R.; Helmer, P. O.; Hayen, H. *Anal. Chem.* **2019**, *91*, 5098−5105.
(33) Kessner, D.; Chambers, M.; Burke, R.; Agus, D.; Mallick, P. *Bioinformatics* **2008**, *24*, 2534−2536.
(34) Martens, L.; Chambers, M.; Sturm, M.; Kessner, D.; Levander, F.; Shofstahl, J.; Tang, W. H.; Römpp, A.; Neumann, S.; Pizarro, A. D.; Montecchi-Palazzi, L.; Tasman, N.; Coleman, M.; Reisinger, F.; Souda, P.; Hermjakob, H.; Binz, P. A.; Deutsch, E. W. *Molecular and cellular proteomics: MCP* **2011**, *10*, R110.000133.
(35) Van Der Walt, S.; Colbert, S. C.; Varoquaux, G. *Comput. Sci. Eng.* **2011**, *13*, 22−30.
(36) Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; Vanderplas, J.; Passos, A.; Cournapeau, D.; Brucher, M.; Perrot, M.; Duchesnay, E. *J. Machine Learning Research* **2011**, *12*, 2825−2830.
(37) Haug, K.; Cochrane, K.; Nainala, V. C.; Williams, M.; Chang, J.; Jayaseelan, K. V.; O'Donovan, C. *Nucleic acids research* **2019**, *48*, D440−D444.
(38) Smirnov, A.; Qiu, Y.; Jia, W.; Walker, D. I.; Jones, D. P.; Du, X. *Anal. Chem.* **2019**, *91*, 9069−9077.
(39) Bonner, R.; Hopfgartner, G. *TrAC, Trends Anal. Chem.* **2019**, *120*, 115278.
(40) Hiller, K.; Hangebrauk, J.; Jäger, C.; Spura, J.; Schreiber, K.; Schomburg, D. *Anal. Chem.* **2009**, *81*, 3429−3439.
(41) Gao, L.; Ji, S.; Burla, B.; Wenk, M. R.; Torta, F.; Cazenave-Gassiot, A. *Anal. Chem.* **2021**, *93*, 3163−3171.
(42) Yang, H.; Chi, H.; Zeng, W. F.; Zhou, W. J.; He, S. M. *Bioinformatics* **2019**, *35*, i183−i190.
(43) Ma, B.; Zhang, K.; Hendrie, C.; Liang, C.; Li, M.; Doherty-Kirby, A.; Lajoie, G. *Rapid Commun. Mass Spectrom.* **2003**, *17*, 2337−2342.
(44) Taheri, S.; Mammadov, M. *Int. J. Appl. Math. Comput. Sci.* **2013**, *23*, 787−795.
(45) Noble, W. S. *Nat. Biotechnol.* **2006**, *24*, 1565−1567.
(46) Probst, P.; Wright, M. N.; Boulesteix, A. L. *Wiley Interdisciplinary Reviews: data mining and knowledge discovery* **2019**, *9*, e1301.
(47) Powers, D. M. *International Journal of Machine Learning Technology* **2021**, *2*, 37−63.
(48) Zhao, J.; Xie, X.; Lin, Q.; Ma, X.; Su, P.; Xia, Y. *Anal. Chem.* **2020**, *92*, 13470−13477.

Recommended by ACS

Visible-Light Paternò–Büchi Reaction for Lipidomic Profiling at Detailed Structure Levels
Hengxue Shi, Yu Xia, *et al.*
MARCH 10, 2023
ANALYTICAL CHEMISTRY
READ

Omic-Scale High-Throughput Quantitative LC–MS/MS Approach for Circulatory Lipid Phenotyping in Clinical Research
Jessica Medina, Julijana Ivanisevic, *et al.*
JANUARY 30, 2023
ANALYTICAL CHEMISTRY
READ

Segmented Flow Strategies for Integrating Liquid Chromatography–Mass Spectrometry with Nuclear Magnetic Resonance for Lipidomics
Jiajun Lei, Richard A. Yost, *et al.*
JANUARY 11, 2023
ANALYTICAL CHEMISTRY
READ

Rapid Multi-Omics Sample Preparation for Mass Spectrometry
Laura K. Muehlbauer, Joshua J. Coon, *et al.*
JANUARY 02, 2023
ANALYTICAL CHEMISTRY
READ

Get More Suggestions >

https://doi.org/10.1021/acs.analchem.2c03505
Anal. Chem. 2022, 94, 16759−16767