Analytical and Bioanalytical Chemistry (2019) 411:4349–4357
https://doi.org/10.1007/s00216-019-01709-1

TRENDS

Advancing untargeted metabolomics using data-independent acquisition mass spectrometry technology

Ruohong Wang [1,2] · Yandong Yin [1] · Zheng-Jiang Zhu [1]

Received: 22 November 2018 / Revised: 14 February 2019 / Accepted: 20 February 2019 / Published online: 7 March 2019

Abstract

Metabolomics quantitatively measures metabolites in a given biological system and facilitates the understanding of physiological and pathological activities. With the recent advancement of mass spectrometry (MS) technology, liquid chromatography-mass spectrometry (LC-MS) with data-independent acquisition (DIA) has been emerged as a powerful technology for untargeted metabolomics due to its capability to acquire all MS2 spectra and high quantitative accuracy. In this trend article, we first introduced the basic principles of several common DIA techniques including MS^E, all ion fragmentation (AIF), SWATH, and MSX. Then, we summarized and compared the data analysis strategies to process DIA-based untargeted metabolomics data, including metabolite identification and quantification. We think the advantages of the DIA technique will enable its broad application in untargeted metabolomics.

Keywords Mass spectrometry · Data-independent acquisition · Untargeted metabolomics

Introduction

Metabolome is a collection of all metabolites in the biological system and represents the most downstream molecular entities closest to the phenotype compared to genome and proteome [1, 2]. Metabolomics aims to quantitatively measure metabolites in biological systems and provides the mechanistic insights to understand physiological and pathological activities and events [3]. Metabolites have high diversity of chemical structures and span a broad concentration ranges, therefore, requiring powerful analytical techniques for metabolic profiling [4]. With the recent advancement of mass spectrometry (MS) technology, liquid chromatography-mass spectrometry (LC-MS) has become one of the most commonly used techniques for metabolomics due to its high sensitivity, selectivity, and throughput in data acquisition and is applied for both targeted and untargeted metabolic profiling [1, 4–6]. Targeted metabolomics analyzes a set of pre-selected metabolites related to a specific biological hypothesis using MS techniques such as multiple reaction monitoring (MRM) and parallel reaction monitoring (PRM) [1, 6–8]. Meanwhile, untargeted metabolomics (or called non-targeted metabolomics) focuses on the comprehensive analysis of all the measurable metabolites in biological samples including both known and unknown ones [1, 4].

For LC-MS-based untargeted metabolomics profiling, data-dependent acquisition (DDA) and data-independent acquisition (DIA) MS techniques are two common data acquisition techniques (Fig. 1) [9]. Both DDA and DIA techniques aim to simultaneously acquire MS1 and MS/MS (or MS2) data for all metabolites from the biological samples in one analysis, but using different strategies. In one data acquisition cycle, both DDA and DIA techniques first acquire a full MS1 scan, and followed by one or multiple MS2 acquisition scans. In DDA, all precursor ions acquired in the MS1 scan are ranked by their intensities (and/or charges). Then, top n

Published in the topical collection *Young Investigators in (Bio-)Analytical Chemistry* with guest editors Erin Baker, Kerstin Leopold, Francesco Ricci, and Wei Wang.

Ruohong Wang and Yandong Yin contributed equally to this work.

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s00216-019-01709-1) contains supplementary material, which is available to authorized users.

✉ Zheng-Jiang Zhu
jiangzhu@sioc.ac.cn

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

[2] University of Chinese Academy of Sciences, Beijing 100049, China

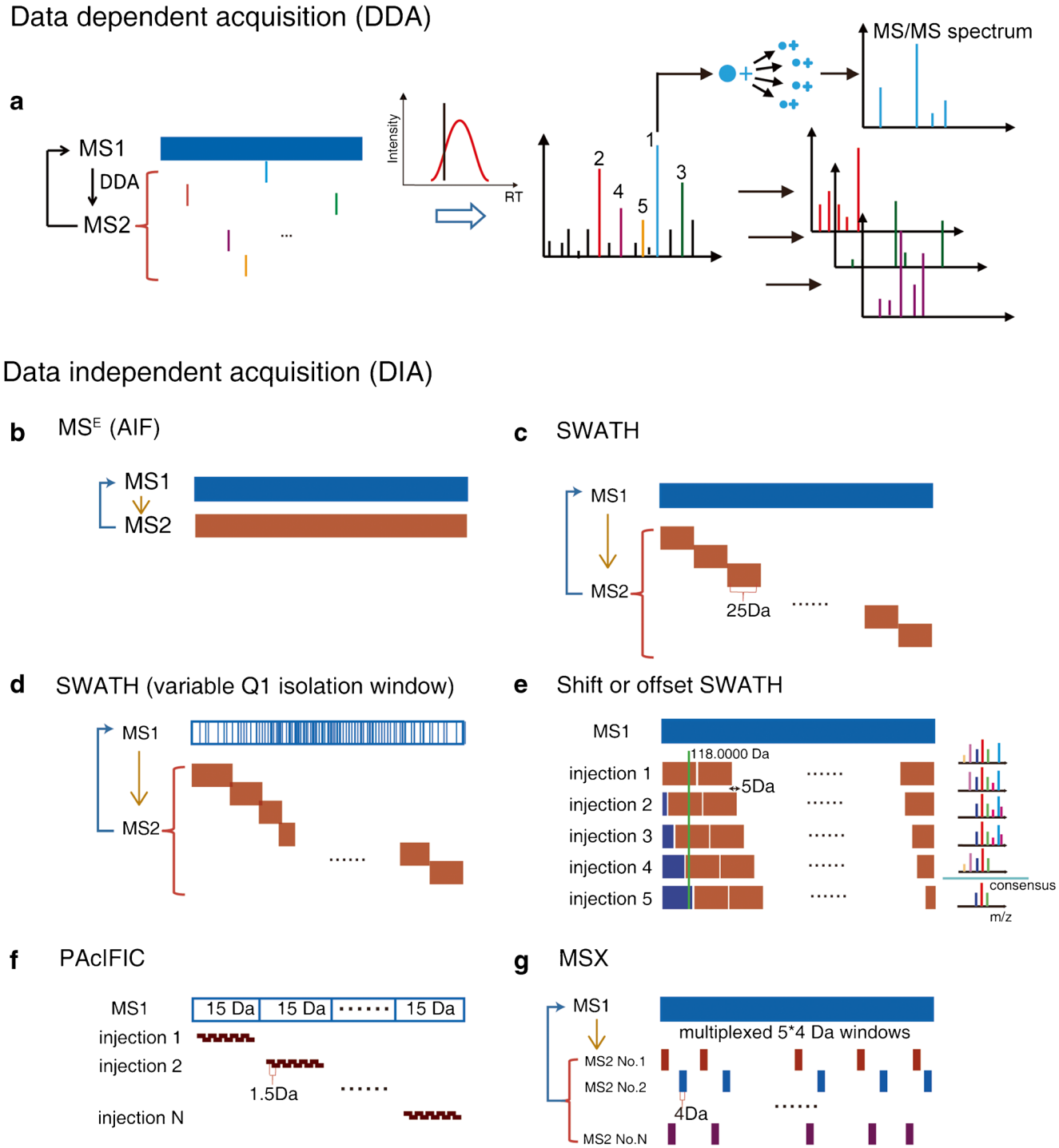

Fig. 1 Mass spectrometry–based data acquisition for untargeted metabolomics: (**a**) data-dependent acquisition (DDA); (**b**–**g**) data-independent acquisition (DIA): (**b**) MS^E and AIF, all co-eluted precursor ions in the whole mass range fragmented to acquire MS2 spectra; (**c**) SWATH, sequential fragmentation of all precursor ions in a serial of quadrupole isolation windows such as 25 Da; (**d**) SWATH with variable Q1 isolation window, assigning each SWATH window with different isolation width based on equalizing the distribution of either the precursor ion population or the total ion current; (**e**) shift or offset SWATH, sequentially shifting the isolation window with a small mass such as 5 Da for five injections; (**f**) PAcIFIC, taking multiple injections for one sample analysis in order to obtain comparable isolation window width as the DDA technique; (**g**) MSX, five separate 4-*m*/*z* isolation windows combined to obtain one MS2 spectrum

ranked precursor ions are sequentially chosen and isolated for fragmentation to acquire their corresponding MS2 spectra one by one (Fig. 1a). The precursor MS1 and its MS2 spectrum are inherently linked, and the metabolite structure can be elucidated by comparing the similarity of *m*/*z* of precursor ion and its MS/MS spectrum to the metabolite standards in the spectral

library [10], such as METLIN [11], MassBank [12], HMDB [5], and MoNA (http://mona.fiehnlab.ucdavis.edu/). However, DDA technique suffered from two major limitations: (1) the low acquisition coverage of MS2 spectra due to its biased selection of high abundant precursor ions for fragmentation and (2) the undefined MS2 spectral quality, which is due to the fact that the MS/MS spectra are not always acquired at the apex of chromatogram peak. In our previous work, we found that < 60% of precursor ions were fragmented for MS2 spectra [13]. The limitations of DDA technique in MS2 spectral coverage and quality restrict its application in untargeted metabolomics. As a comparison, in DIA, all precursor ions in a predefined isolation window (from 5 Da to a full mass range) are isolated to acquire multiplexed MS2 spectra, and this step is repeated until the full mass range is covered. In theory, DIA technique enables to acquire MS2 spectra for all precursor ions. In addition, DIA technique enables to select either MS1 or MS2 ions for metabolite quantification, which potentially increases the quantitative accuracy. However, the direct link between MS1 and MS2 ions in multiplexed MS2 spectra is missing. Shared fragments from the co-isolated precursor ions increased the complexity of multiplex MS2 spectra. We performed a very rough estimation analysis using the human urine SWATH dataset report in our publication [13]. We found that approximately 2838 MS1 ions on average were recorded (considered as co-elution) in one MS1 scan. Therefore, about 118 MS1 ions on average were co-isolated in one 25-Da SWATH window (24 windows in total) (see Electronic Supplementary Material for more details). Both present a great challenge to process and reconstruct MS2 spectrum for metabolite identification in DIA-based untargeted metabolomics.

In this trend article, we first introduced the basic principles of several common DIA techniques and summarized and compared the data analysis strategies to process DIA-based untargeted metabolomics data (mostly for SWATH technique), including both metabolite identification and quantification. Finally, we summarized the future development to overcome the challenges in DIA and advance its application in untargeted metabolomics.

Data-independent acquisition technology

DIA techniques have been emerged as an alternative approach for untargeted metabolomics owing to its capability to acquire all MS2 spectra and high quantitative accuracy. The concept of data-independent acquisition was first introduced by Venable et al. and applied in proteomics [14]. The method employed the sequential isolation and fragmentation of all precursor ions within a 10-Da window at a time until a desired mass range (e.g., 400–1400 Da) has been covered. The method aimed to acquire MS/MS spectra for multiple ions without the selection of precursor ion and to ensure the full acquisition coverage. The generated dataset had comparable peptide identification coverage and quality as the conventional DDA technique, but higher signal-to-noise ratios and broader dynamic range in quantitative analysis [14]. The method is very time-consuming. Therefore, it has not been applied in untargeted metabolomics. Later, benefited from the recent advancements of mass spectrometers including improved acquisition speed, mass accuracy, and resolution, more research interests have been focused on the development of DIA techniques. Several DIA techniques have emerged and applied in both proteomics and metabolomics. In DIA, all precursor ions in a predefined isolation window (from several Da to a full mass range) are sequentially isolated to acquire multiplexed MS2 spectra and ensure to acquire MS2 spectra for all ions in MS1 scan. According to the width of isolation window, several representative DIA techniques are described, including MS^E [15], all ion fragmentation (AIF) [16], SWATH [17], PAcIFIC [18], and MSX [19].

Both MS^E [15] and AIF [16] techniques transmit all co-eluted precursor ions in the whole mass range for fragmentation and acquiring MS2 spectra, and no precursor isolation is applied (Fig. 1b). The MS^E technique was registered by Waters Corporation. It alternatively acquires the full MS1 scan with low collision energy and MS2 scan from all precursor ions with high collision energy [15]. It enables to obtain complete chromatograms for MS1 and MS2 data due to an efficient duty cycle. The relationship between precursor and product ions can be re-linked by retention times, mass defect, peak shape similarity, or a combination of them. The principle of AIF is similar to MS^E. It was originally developed on Thermo Exactive MS instrument by transmitting all precursor ions into a higher energy collisional dissociation (HCD) cell for fragmentation and acquiring a multiplexed MS2 spectrum after each full MS1 scan [16]. The similar function is also now enabled with different instruments from other vendors, such as Agilent Q-TOF [20].

Since all precursor ions are fragmented all together in MS^E and AIF, the generated multiplexed MS2 spectra are largely complexed. To reduce the complexity, SWATH technique was developed using a narrow isolation window (Fig. 1c). The SWATH technique was first described by Gillet et al. in 2012 and applied in proteomics research [17, 21]. Later, it became popular in the analysis of small molecules [22–26], including metabolomics [27–29]. SWATH-based DIA technique enables the sequential fragmentation of all precursor ions in a serial of quadrupole isolation windows (Q1 window) and records the complete "snapshots" of all metabolite ions (MS1 data) and their product ions (MS2 data) in the entire chromatogram. Depending on the MS1 scan range and the width of isolation window (or SWATH window), multiple multiplexed MS2 spectra were acquired to cover the full mass range in one cycle. The width of isolation window is typically

set as 25 Da, but can be variable from 5 to 100 Da or even larger. Compared with MS^E and AIF, SWATH allows a reduction of simultaneously fragmented precursor ions, therefore, decreasing the complexity of multiplexed MS2 spectra. As a result, the effort to reconstruct the connections between the precursor and product ions is also alleviated. For the quantitative analysis, the reduced complexity of product ions also significantly decreases the interference and improves the quantitative accuracy. Recently, Bonner et al. reviewed the application of SWATH technique in metabolomics [30].

Commonly, SWATH acquisition applies a fixed Q1 isolation window (e.g., 25 Da) to acquire MS2 spectra. A revised version of SWATH technique, called variable Q1 isolation window (Fig. 1d), enables to assign each SWATH window with different isolation width based on equalizing the distribution of either the precursor ion population or the total ion current (TIC) [31]. Therefore, the precursor ion population for fragmentation in one MS2 spectrum is reduced for those from the densely distributed region. This approach further reduces the ion interference from co-elution and co-fragmentation and improves the spectral quality. Alternatively, sequentially shifting the isolation window of SWATH acquisition with a small overlapping mass range (e.g., 5 Da) is another way to reduce the spectral complexity, referred as shift or offset SWATH technique (Fig. 1e) [21]. However, shift SWATH technique typically requires five repetitive injections to cover a 25-Da SWATH window. Specifically, in the first analysis, the SWATH isolation window is set the same as the conventional SWATH method (e.g., 25 Da each; 100–125 Da, 124–150 Da). However, in a second analysis, a given shift (e.g., 5 Da) is set to each isolation window (e.g., 100–105 Da, 104–130 Da, 129–155 Da). Sequentially shifting the isolation window for five analyses provides an effect of spectral deconvolution for acquired multiplexed MS2 spectra. For each feature, multiplexed MS2 spectra at the apex of chromatogram peak in each offset SWATH analysis can be extracted and combined to generate a consensus MS2 spectrum. The consensus spectrum obtained from offset SWATH provided a spectral deconvolution effect similar to the MS2 spectra acquired using a 5-Da isolation window. Therefore, the precursor/product ion relationship can be reconstructed more accurately to improve the accuracy of metabolite identification and quantification.

Alternative to SWATH technique, other DIA techniques utilizes even smaller isolation windows for the isolation of precursor ions. For example, PAcIFIC, referred as precursor acquisition independent from ion count (Fig. 1f), acquires tandem mass spectra at every *m/z* value on a LTQ-orbitrap instrument [18]. In the first injection, ion trap is used to perform data-independent acquisition at each of ten continuous intervals (each with a 1.5-Da width) across a range of 15 Da using a 2.5 Da isolation width. In the next analysis, the same manner is performed on another 15-Da mass range. The injections are repeated until the whole mass range has been covered. PAcIFIC has comparable isolation window width as the DDA technique and decreases the complexity of acquired MS2 spectra. However, the analysis of PAclFIC takes multiple injections for one sample analysis. Instead, MSX is developed by multiplexing five isolation mass ranges (4 Da each) into one fragmentation (Fig. 1g) [19]. Five isolation mass ranges are randomly chosen from the predefined n possible non-overlapping windows (4 Da each) and combined as one analysis. The random selection is repeated until covering the whole mass range. The generated one multiplexed MS2 spectrum from five isolation mass ranges can be de-multiplexed into five separated MS2 spectra, and each covering a 4-Da window. MSX technique maintains the acquisition efficiency similar to SWATH technique, but higher selectivity similar to PAcIFIC, which is a good combination of data acquisition efficiency and computational deconvolution. Both of PAcIFIC and MSX were only applied in proteomics, but potentially applicable for metabolomics. However, these DIA techniques are not systematically studied for metabolomics, and the availability of data processing tools is another major barrier. Therefore, no untargeted metabolomics studies using two techniques were reported.

Overall, we think DIA technique has the following the advantages to make it suitable for untargeted metabolomics: (1) high transmission efficiency in isolation leading to relative higher intensity and improved MS2 spectra quality; (2) non-restriction of precursor ions leading to the informative and broad coverage of MS/MS spectra; (3) efficient duty cycle ensuring to obtain complete chromatograms for all MS1 and MS2 ions, which increases the quantitative analysis with wider dynamic range, better sensitivity, and higher reproducibility.

Data analysis for DIA data in metabolomics

For untargeted metabolomics, the raw data processing such as peak detection, peak alignment, and grouping is first performed to generate a feature table. Several software tools, such as XCMS [32], MZmine [33], OpenMS [34], and MS-DIAL [27], are the most widely used. All of them are open-source software tools and support users to develop new algorithms and flexible workflows to process both DDA and DIA-based metabolomics data. In addition to generating a feature table, new strategies are required to obtain the corresponding MS2 spectra and perform metabolite identification. However, in DIA-based metabolomics dataset, direct connections between precursor and product ions are missing, which presents a challenging task for data analysis. In DIA, product ions are scanned in each cycle; hence, chromatographic ion profiles are considered the key information to reconstruct the precursor-product relationship. However, the co-elution and co-fragmentation of precursor ions make the eluted

chromatographic ion profiles complicated. In general, two strategies were usually used to process DIA-based metabolomics data: untargeted MS/MS spectral deconvolution (Fig. 2a) and targeted extraction (Fig. 2b) including metabolite-centric and spectrum-centric approaches. MS-DIAL, MetDIA, and MetaboDIA are three representative tools for DIA-based metabolomics.

In 2015, Tsugawa et al. developed the Mass Spectrometry-Data-Independent AnaLysis software, namely, MS-DIAL [27], for untargeted MS/MS spectral deconvolution to support SWATH-based metabolomics. The recent version of MS-DIAL is also applicable for AIF-based metabolomics data. In this software tool, the acquired DIA data are first converted to the Analysis Base File (ABF) data format. Next, the "peak spotting" method is applied to smooth chromatograms of the precursor ions and detect peaks in *m/z* and RT axes. Then, the MS/MS deconvolution algorithm, namely, MS^2Dec, is used to reconstruct the connections between precursor and product ions. In MS^2Dec algorithm, the model peaks of the targeted precursor ion and its shoulder peaks are first extracted from the product ion chromatograms. Then, the least-square optimization method is used for MS2 spectral deconvolution with the model peaks and to reconstruct the pseudo MS2 spectrum. Finally, compound identification is performed using the deconvoluted MS2 spectra, the precursor *m/z*, and RT, through comparing their similarities to the standards in the spectral library. In addition, the peak alignment, filtering, and missing value imputation are also performed to the detected MS1 peaks.

The deconvolution algorithm used in MS-DIAL only considers the model peak of the target precursor ion and its shoulder peaks. This simplified model decreases the computation complexity and has a fast deconvolution speed. However, the co-elution of the co-fragmented product ions from different precursor ions in the same isolation window is far more complexed. Thereby, the determination of component number and the selection of proper model peaks from the elution profiles for the precursor and product ion chromatograms should be carried in a more accurate way. In GC-MS, ADAP-GC software developed by Du Lab provides a better algorithm for the selection of model peak and enables the deconvolution of GC-MS data accurate [35]. In this software, components are automatically determined from the detected peaks using a two-step clustering method, which takes both the apex elution time and shape of EIC peaks into account. Each component represents a potential compound in the analysis and the sharpest peak in the component is selected as model peak. A linear combination of all selected model peaks is applied to each elution profile (i.e., ion chromatogram) for deconvolution. However, this method is only applied for GC-MS data and has not been applied for SWATH-based metabolomics data analysis.

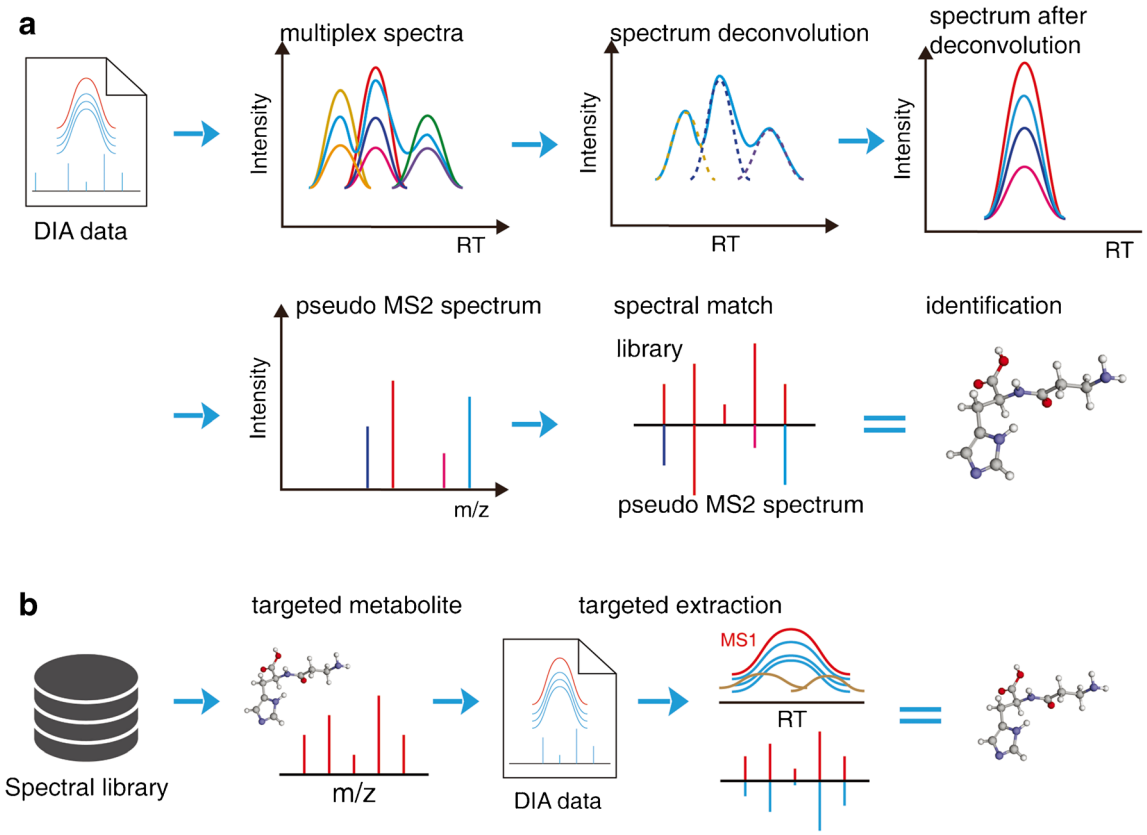

Fig. 2 Illustration of data analysis strategies for DIA-based untargeted metabolomics: (**a**) untargeted MS/MS spectral deconvolution to reconstruct the relationship between the precursor and product ions; (**b**) targeted extraction of metabolites in a given spectral library using a metabolite-centric strategy

Another software tool, namely, MetDIA, is developed by our group in 2016 [28]. MetDIA is an R package that performs targeted extraction of metabolites in a given spectral library from the SWATH data using a metabolite-centric strategy. MetDIA performs the metabolite-centric identifications in the following steps: (1) MS1 peak detection and alignment; (2) targeted MS1 match and extraction of both MS1 and MS2 ion chromatograms using MS2 ions in the spectral library; (3) generation of peak groups with the extracted precursor and product ion chromatograms, and generation of pseudo MS2 spectra; (4) metabolite-centric identification using two orthogonal scores, peak-peak correlation (PPC) score and spectrum-spectrum match (SSM) score; (5) statistical analysis and results output. Specially, for metabolite identification, PPC score is calculated by averaging the highest Pearson correlation coefficient values from half of the product ions. In addition, PPC and SSM scores are averaged to generate the final metabolic-centric identification (MCI) score for targeted metabolite identification. A minimum MCI score was determined as 0.8 for 1% false positive rate (FPR) in metabolite identification. In this work, we constructed a consensus spectral library of 786 metabolites (765 metabolites in positive and 757 metabolites in negative modes). The validation experiments were acquired using a mixture of 30 metabolite standards at two concentrations, and we found that MetDIA correctly identified more metabolites at low concentration with fewer false positive metabolites than the conventional DDA approach. Additionally, MetDIA also provided much more metabolite identifications than DDA in biological samples. Both results proved that the DIA technique is more accurate and sensitive than the DDA technique, especially for low abundant metabolites. Comparing to MS-DIAL, we also found that MetDIA detected 20–70% more metabolites than MS-DIAL. However, the targeted method is restricted by the size and quality of the spectral library. Recently, Bruderer et al. reported a guideline to build high-quality spectral library for SWATH-based metabolomics [36].

MetaboDIA, developed by Chen et al. in 2017 [29], first built a customized consensus spectral library from a DDA-based metabolomics dataset. The generated spectral library consisted of compound identification units (CIUs), each labeled with precursor *m*/*z* value, retention time, and a consensus MS2 spectrum. Then, targeted extraction of MS2 ions was performed in the DIA-based metabolomics dataset for accurate quantification. The targeted extraction strategy is similar to MetDIA [28], but with a large size spectral library. The authors applied MetaboDIA to a clinical serum metabolomics study. In this study, they built a DDA-based spectral library containing MS2 spectra for 1829 metabolites. For metabolite identification, putative molecular formulas were first generated by matching the accurate mass of MS1 peak against a known database. Then, metabolite identification can be further achieved through the MS2 spectral match with an external standard library such as METLIN. MetaboDIA prefers to construct the consensus spectral library from the DDA data because it provides a direct precursor-product ion relationship. Alternatively, MetaboDIA also allows users to build a spectral library directly from the DIA data using the proteomics software DIA-Umpire [37]. However, the ambiguity still existed when mapping the product ions to their precursor ions.

Other tools have also been emerged for processing DIA-based metabolomics data. For example, OpenSWATH [38], which is a module of OpenMS software tool and designed for proteomics, is capable to perform the DIA analysis with a prebuilt spectral library. We believe OpenSWATH may be used for targeted extraction in DIA-based metabolomics with some proper modifications. Skyline, another popular tool for proteomics [39], can also be utilized to perform targeted extraction using a spectral library and support the analysis of DIA-based metabolomics data [40]. Specific for MS^E-based metabolomics, Prenni lab has reported several publications, including the utilization of the whole metabolomics dataset for reconstructing precursor-product ion relationships [41], and the development of an hierarchical clustering-based approach to group both MS and MS/MS peaks (i.e., RAMClustR) [42]. Other tools provided by instrument vendors such as MasterView and Progenesis QI can also be utilized for processing DIA-based metabolomics data.

Quantification in SWATH-MS data analysis

Accurate metabolite quantification is also important in untargeted metabolomics, especially for the discovery of differential metabolites. In DDA-based metabolomics, MS1 peak areas are commonly used for relative quantification. In DIA-based metabolomics, since both MS1 and MS2 ions are acquired in each sampling point, it enables to select either MS1 or MS2 ions for metabolite quantification. However, metabolite quantification is commonly performed using the MS1 peak area, such as MS-DIAL [27] and MetDIA [28]. Recently, MetaboDIA demonstrated the utility of MS2 ion as a mean for relative quantification [29]. The authors first constructed a DDA library consisting of compound identification units (CIUs). Each CIU was labeled with precursor *m*/*z* value, retention time, and a MS2 spectrum. Then, MetaboDIA re-extracted either MS1 or MS2 ions for each CIU in the library from the DIA dataset for quantification. The authors further demonstrated that the average quantification performances are comparable using MS1 and MS2 ion intensities. However, in certain cases, the use of MS2 ions provided better quantification performances in terms of CV values, and vice versa. In fact, a similar conclusion was also drawn by Li and colleagues using the parallel reaction monitoring (PRM) technique [40]. They claimed that either MS1 or MS2 ions can be used for quantification. For a total of 61 intracellular

metabolites of interests, they evaluated the quantitative reproducibility, sensitivity, and linear response range for each ion and selected the most ideal quantification ion (either MS1 or MS2 ion) for metabolite quantification. However, for a specific metabolite, there is no universal evaluation approach to decide whether MS1 or MS2 ion should be selected as the quantification ion. This presents a challenge to improve the quantification accuracy.

Instead, targeted metabolomics has advantages for quantitation featuring high sensitivity, wide dynamic range, and high reproducibility [4, 7]. But it is limited by the number of metabolites analyzed in one experiment. Recently, we developed a workflow, namely, SWATHtoMRM [13], to combine the broad coverage of SWATH technique for metabolic profiling and accurate quantification of MRM technique. Specifically, we utilized the SWATH-MS technique to profile the pooled biological samples and acquired all MS2 spectra. Then, SWATHtoMRM software was used to generate a large scale MRM transitions from SWATH data. Quantitative analysis was then performed on QqQ with MRM technique. Combining the advantages of SWATH and MRM, SWATHtoMRM has broad metabolites coverage together with high quantification reproducibility, accuracy, sensitivity, and dynamic range. We applied the strategy on the potential metabolite biomarker discovery for colorectal cancer (CRC) diagnosis. In CRC tissue, a total of 1303 metabolites can be measured in one experiment and a total of 20 metabolites were selected as the potential biomarkers for cancer classification. Finally, we validated the performance of these potential biomarkers on human plasma samples. Similarly, Wang et al. developed a "pseudo targeted" method by combining the SWATH and MRM techniques [43]. They applied the method to discover differential serum metabolites related to type 2 diabetes and indicated that mitochondrial dysfunction happened with the patients.

Outlook

Overall, DIA technique has been emerged as a powerful approach for untargeted metabolomics due to its capability to acquire all MS2 spectra and the potential to achieve high quantitative accuracy. However, the generated complex data set for metabolomics heavily relies on sophisticated bioinformatics software for data processing and metabolite identification. Improvement of reconstructed spectral quality is imperative not only for high confident metabolite annotation but also for accurate quantification. From the technology perspective, reducing the data complexity without losing the information is the key to enable the broad application of DIA technique in metabolomics. First, the addition of an additional separation of precursor ions can significantly reduce the spectral complexity. Currently, liquid chromatography-ion mobility-mass spectrometry (LC-IM-MS) coupled with data-independent acquisition method is an emerging technique toward this purpose [44]. Ion mobility technology enables rapid gas-phase separation of metabolite ions through the collisions between ions and buffer gas under an electric field and provides an orthogonal separation to effectively reduce chemical noise. The technique also provides CCS values to aid metabolite identification [45]. Moreover, the product ions eluted at the same drift time are considered to be derived from the same precursor ion. This principle provides an additional constrain to facilitate the spectral devolution in the analysis of DIA-based metabolomics data. Second, other data deconvolution algorithms to construct the precursor and product ion relationship are still needed. Currently, MS2 spectral deconvolution such as MS-DIAL is operated by combining the selected model peaks in a linear way. However, other non-linear deconvolution methods can possibly be employed, such as kernel-based spectral deconvolution. Kernel-based deconvolution is one of the non-parametric deconvolution methods. Recently, Hong et al. reported to employ a machine-learning method for the kernel-based component determination on IMS-MS/MS data [46]. We think the same principle is also applicable for the analysis of DIA-MS data. In DIA data analysis, one can learn the combination of multiple kernels using DIA MS2 extracted ion chromatograms (EICs) and deconvolve the EICs to discover the true spectrum of a certain precursor peak. In summary, DIA technology has demonstrated its advantages and applicability in untargeted metabolomics. With the future advancements in both technology and data processing methods, we believe the advantages with DIA technique will enable its broad application in untargeted metabolomics.

Funding information The work has been supported by the National Natural Science Foundation of China (Grants 21575151).

Compliance with ethical standards

The authors declare that they have no conflict of interest.

References

1. Patti GJ, Yanes O, Siuzdak G. Metabolomics: the apogee of the omics trilogy. Nat Rev Mol Cell Biol. 2012;13(4):263–9.
2. Newgard CB. Metabolomics and metabolic diseases: where do we stand? Cell Metab. 2017;25(1):43–56.
3. Johnson CH, Ivanisevic J, Siuzdak G. Metabolomics: beyond biomarkers and towards mechanisms. Nat Rev Mol Cell Biol. 2016;17(7):451–9.
4. Cajka T, Fiehn O. Toward merging untargeted and targeted methods in mass spectrometry-based metabolomics and lipidomics. Anal Chem. 2016;88(1):524–45.

5. Want EJ, Wilson ID, Gika H, Theodoridis G, Plumb RS, Shockcor J, et al. Global metabolic profiling procedures for urine using UPLC-MS. Nat Protoc. 2010;5(6):1005–18.
6. Cai Y, Zhu Z-J. A high-throughput targeted metabolomics workflow for the detection of 200 polar metabolites in central carbon metabolism. Microbial Metabolomics. Berlin: Springer; 2019. p. 263–74.
7. Cai Y, Weng K, Guo Y, Peng J, Zhu Z-J. An integrated targeted metabolomic platform for high-throughput metabolite profiling and automated data processing. Metabolomics. 2015;11(6):1575–86.
8. Zhou J, Liu H, Liu Y, Liu J, Zhao X, Yin Y. Development and evaluation of a parallel reaction monitoring strategy for large-scale targeted metabolomics quantification. Anal Chem. 2016;88(8):4478–86.
9. Schrimpe-Rutledge AC, Codreanu SG, Sherrod SD, McLean JA. Untargeted metabolomics strategies-challenges and emerging directions. J Am Soc Mass Spectrom. 2016;27(12):1897–905.
10. Zhu Z-J, Schultz AW, Wang J, Johnson CH, Yannone SM, Patti GJ, et al. Liquid chromatography quadrupole time-of-flight mass spectrometry characterization of metabolites guided by the METLIN database. Nat Protoc. 2013;8(3):451.
11. Smith CA, O'Maille G, Want EJ, Qin C, Trauger SA, Brandon TR, et al. METLIN: a metabolite mass spectral database. Ther Drug Monit. 2005;27(6):747–51.
12. Horai H, Arita M, Kanaya S, Nihei Y, Ikeda T, Suwa K, et al. MassBank: a public repository for sharing mass spectral data for life sciences. J Mass Spectrom. 2010;45(7):703–14.
13. Zha H, Cai Y, Yin Y, Wang Z, Li K, Zhu ZJ. SWATHtoMRM: development of high-coverage targeted metabolomics method using SWATH technology for biomarker discovery. Anal Chem. 2018;90:4062–70.
14. Venable JD, Dong MQ, Wohlschlegel J, Dillin A, Yates JR. Automated approach for quantitative analysis of complex peptide mixtures from tandem mass spectra. Nat Methods. 2004;1:39–45.
15. Silva JC, Denny R, Dorschel CA, Gorenstein M, Kass IJ, Li GZ, et al. Quantitative proteomic analysis by accurate mass retention time pairs. Anal Chem. 2005;77:2187–200.
16. Geiger T, Cox J, Mann M. Proteomics on an orbitrap benchtop mass spectrometer using all-ion fragmentation. Mol Cell Proteomics. 2010;9:2252–61.
17. Gillet LC, Navarro P, Tate S, Röst H, Selevsek N, Reiter L, et al. Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: a new concept for consistent and accurate proteome analysis. Mol Cell Proteomics. 2012;11:O111.016717.
18. Panchaud A, Scherl A, Shaffer SA, Haller PDV, Kulasekara HD, Miller SI, et al. PAcIFIC: how to dive deeper into the proteomics ocean. Anal Chem. 2011;81:6481–8.
19. Egertson JD, Kuehn A, Merrihew GE, Bateman NW, MacLean BX, Ting YS, et al. Multiplexed MS/MS for improved data-independent acquisition. Nat Methods. 2013;10:744–6.
20. Naz S, Gallart-Ayala H, Reinke SN, Mathon C, Blankley R, Chaleckis R, et al. Development of a liquid chromatography-high resolution mass spectrometry metabolomics method with high specificity for metabolite identification using all ion fragmentation acquisition. Anal Chem. 2017;89(15):7933–42.
21. Ludwig C, Gillet L, Rosenberger G, Amon S, Collins BC, Aebersold R. Data-independent acquisition-based SWATH-MS for quantitative proteomics: a tutorial. Mol Syst Biol. 2018;14(8):e8126.
22. Hopfgartner G, Tonoli D, Varesio E. High-resolution mass spectrometry for integrated qualitative and quantitative analysis of pharmaceuticals in biological matrices. Anal Bioanal Chem. 2012;402(8):2587–96.
23. Siegel D, Meinema AC, Permentier H, Hopfgartner G, Bischoff R. Integrated quantification and identification of aldehydes and ketones in biological samples. Anal Chem. 2014;86(10):5089–100.
24. Roemmelt AT, Steuer AE, Poetzsch M, Kraemer T. Liquid chromatography, in combination with a quadrupole time-of-flight instrument (LC QTOF), with sequential window acquisition of all theoretical fragment-ion spectra (SWATH) acquisition: systematic studies on its use for screenings in clinical and forensic toxicology and comparison with information-dependent acquisition (IDA). Anal Chem. 2014;86(23):11742–9.
25. Bruderer T, Varesio E, Hopfgartner G. The use of LC predicted retention times to extend metabolites identification with SWATH data acquisition. J Chromatogr B. 2017;1071:3–10.
26. Zhu X, Chen Y, Subramanian R. Comparison of information-dependent acquisition, SWATH, and MS(All) techniques in metabolite identification study employing ultrahigh-performance liquid chromatography-quadrupole time-of-flight mass spectrometry. Anal Chem. 2014;86(2):1202–9.
27. Tsugawa H, Cajka T, Kind T, Ma Y, Higgins B, Ikeda K, et al. MS-DIAL: data-independent MS/MS deconvolution for comprehensive metabolome analysis. Nat Methods. 2015;12:523–6.
28. Li H, Cai Y, Guo Y, Chen F, Zhu ZJ. MetDIA: targeted metabolite extraction of multiplexed MS/MS spectra generated by data-independent acquisition. Anal Chem. 2016;88:8757–64.
29. Chen G, Walmsley S, Cheung GCM, Chen L, Cheng CY, Beuerman RW, et al. Customized consensus spectral library building for untargeted quantitative metabolomics analysis with data independent acquisition mass spectrometry and MetaboDIA workflow. Anal Chem. 2017;89:4897–906.
30. Bonner R, Hopfgartner G. SWATH data independent acquisition mass spectrometry for metabolomics. Trends Anal Chem. 2018. https://doi.org/10.1016/j.trac.2018.10.014.
31. Zhang Y, Bilbao A, Bruderer T, Luban J, Strambio-De-Castillia C, Lisacek F, et al. The use of variable Q1 isolation windows improves selectivity in LC-SWATH-MS acquisition. J Proteome Res. 2015;14:4359–71.
32. Smith CA, Want EJ, O'Maille G, Abagyan R, Siuzdak G. XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. Anal Chem. 2006;78(3):779–87.
33. Pluskal T, Castillo S, Villar-Briones A, Orešič M. MZmine 2: modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. BMC Bioinformat. 2010;11(1):395.
34. Rost HL, Sachsenberg T, Aiche S, Bielow C, Weisser H, Aicheler F, et al. OpenMS: a flexible open-source software platform for mass spectrometry data analysis. Nat Methods. 2016;13(9):741–8.
35. Ni Y, Su M, Qiu Y, Jia W, Du X. ADAP-GC 3.0: improved peak detection and deconvolution of co-eluting metabolites from GC/TOF-MS data for metabolomics studies. Anal Chem. 2016;88(17):8802–11.
36. Bruderer T, Varesio E, Hidasi AO, Duchoslav E, Burton L, Bonner R, et al. Metabolomic spectral libraries for data-independent SWATH liquid chromatography mass spectrometry acquisition. Anal Bioanal Chem. 2018;410(7):1873–84.
37. Tsou CC, Avtonomov D, Larsen B, Tucholska M, Choi H, Gingras AC, et al. DIA-Umpire: comprehensive computational framework for data-independent acquisition proteomics. Nat Methods. 2015;12(3):258–64 7 p following 64.
38. Rost HL, Rosenberger G, Navarro P, Gillet L, Miladinovic SM, Schubert OT, et al. OpenSWATH enables automated, targeted analysis of data-independent acquisition MS data. Nat Biotechnol. 2014;32(3):219–23.
39. MacLean B, Tomazela DM, Shulman N, Chambers M, Finney GL, Frewen B, et al. Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. BMC Bioinformat. 2010;26(7):966–8.
40. Li Z, Li Y, Chen W, Cao Q, Guo Y, Wan N, et al. Integrating MS1 and MS2 scans in high-resolution parallel reaction monitoring

assays for targeted metabolite quantification and dynamic ^{13}C-labeling metabolism analysis. Anal Chem. 2017;89:877–85.
41. Broeckling CD, Heuberger AL, Prince JA, Ingelsson E, Prenni JE. Assigning precursor–product ion relationships in indiscriminant MS/MS data from non-targeted metabolite profiling studies. Metabolomics. 2013;9(1):33–43.
42. Broeckling CD, Afsar FA, Neumann S, Ben-Hur A, Prenni JE. RAMClust: a novel feature clustering method enables spectral-matching-based annotation for metabolomics data. Anal Chem. 2014;86(14):6812–7.
43. Wang L, Su B, Zeng Z, Li C, Zhao X, Lv W, et al. Ion-pair selection method for pseudotargeted metabolomics based on SWATH MS acquisition and its application in differential metabolite discovery of type 2 diabetes. Anal Chem. 2018:11401–08.
44. Zheng X, Wojcik R, Zhang X, Ibrahim YM, Burnum-Johnson KE, Orton DJ, et al. Coupling front-end separations, ion mobility spectrometry, and mass spectrometry for enhanced multidimensional biological and environmental analyses. Annu Rev Anal Chem. 2017;10(1):71–92.
45. Zhou Z, Tu J, Zhu ZJ. Advancing the large-scale CCS database for metabolomics and lipidomics at the machine-learning era. Curr Opin Chem Biol. 2018;42:34–41.
46. Hong P, Bernstein W, Wei J, Lin C, Kernel-based component decomposition for glycan mixture separation using ion mobility spectrometry-MS/MS, 66th annual ASMS conference on mass spectrometry and allied topics, San Diego, CA, June 3-7; 2018.