Targeting Unique Biological Signals on the Fly to Improve MS/MS Coverage and Identification Efficiency in Metabolomics

Kevin Cho\textsuperscript{a,b}, Michaela Schwaiger-Haber\textsuperscript{a,b}, Fuad J. Naser\textsuperscript{a,b}, Ethan Stancliffe\textsuperscript{a,b}, Miriam Sindelar\textsuperscript{a,b}, Gary J. Patti\textsuperscript{a,b,*}

\textsuperscript{a}Department of Chemistry, Washington University in St. Louis, St. Louis, MO, USA
\textsuperscript{b}Department of Medicine, Washington University in St. Louis, St. Louis, MO, USA

\# These authors contributed equally to this work.

Abstract

When using liquid chromatography/mass spectrometry (LC/MS) to perform untargeted metabolomics, it is common to detect thousands of features from a biological extract. Although it is impractical to collect non-chimeric MS/MS data for each in a single chromatographic run, this is generally unnecessary because most features do not correspond to unique metabolites of biological relevance. Here we show that relatively simple data-processing strategies that can be applied on the fly during acquisition of data with an Orbitrap ID-X, such as blank subtraction and well-established adduct or isotope calculations, decrease the number of features to target for MS/MS analysis by up to an order of magnitude for various types of biological matrices. We demonstrate that annotating these non-biological contaminants and redundancies in real time during data acquisition enables comprehensive MS/MS data to be acquired on each remaining feature at a single collision energy. To ensure that an appropriate collision energy is applied, we introduce a method using a series of hidden ion-trap scans in an Orbitrap ID-X to find an optimal value for each feature that can then be applied in a subsequent high-resolution Orbitrap scan. Data from 100 metabolite standards indicate that this real-time optimization of collision energies leads to more informative MS/MS patterns compared to using a single fixed collision energy alone. As a benchmark to evaluate the overall workflow, we manually annotated unique biological features by independently subjecting \textit{E. coli} samples to a credentialing analysis. While credentialing led to a more rigorous reduction in feature number, on-the-fly annotation with blank subtraction on an Orbitrap ID-X did not inappropriately discard unique biological metabolites. Taken together, our results reveal that optimal fragmentation data can be obtained in a single LC/MS/MS run for >90% of the unique biological metabolites in a sample when features are annotated during acquisition and collision energies are selected by using parallel mass spectrometry detection.

*Corresponding author. gjpattij@wustl.edu.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

G.J.P. is a scientific advisory board member for Cambridge Isotope Laboratories and has a research collaboration agreement with Thermo Fisher Scientific.

Appendix A. Supplementary Material

Supplementary methods, figures, and tables to this article can be found online (PDF).
1. Introduction

Confidently assigning biochemical names and structures to metabolomic features relies on MS/MS fragmentation data. There are two experimental paradigms for acquiring MS/MS data in untargeted metabolomics [1]. MS1 data for global profiling and MS/MS data for structural identification can be collected in separate experiments, or MS1 and MS/MS data can be acquired in a single integrated experiment.

Historically, untargeted metabolomics has most commonly been performed by profiling samples in MS1 mode first. The data are then input into one or more stand-alone software programs for processing and the results are manually inspected [2]. Features of interest are selected, often based on statistical changes between sample groups, and targeted for MS/MS analysis in a subsequent experiment. There are several limitations to this traditional approach of acquiring MS/MS data in untargeted metabolomics. First, processing MS1 data to generate an MS/MS target list is arduous and time consuming. The days, weeks, or even months that can occur between analyses raise issues related to data alignment and sample degradation. Second, when sample material is limited, performing multiple experimental runs to collect fragmentation data separately may not be practical.

An alternative approach that has gained increasing attention in recent years is the collection of MS1 and MS/MS data in a single untargeted metabolomic analysis. These experiments come in different forms: data-dependent acquisition (DDA) and data-independent acquisition (DIA). DDA does not acquire MS/MS data for every feature, but rather only a subset of the total ions as most frequently determined by signal intensity. In contrast to DDA, DIA can obtain fragmentation data for all features detected in a sample. To achieve such broad MS/MS coverage, wide isolation windows are used to fragment multiple precursor ions simultaneously and the resulting chimeric MS/MS spectra must then be
computationally deconvolved [3–8]. The major benefit of DIA is that comprehensive MS1 and fragmentation data are collected in a single experiment, theoretically removing the need for additional analyses.

When considering workflows for acquiring MS1 and non-chimeric MS/MS data simultaneously, a limiting factor is the large number of features typically detected in untargeted metabolomics. Even with the fastest mass spectrometers currently available, it is not possible to target each of the thousands of features individually in separate MS/MS experiments using narrow isolation windows [9]. In practice, however, such comprehensive MS/MS coverage of all features is not necessary to identify all of the metabolites detected in a given experiment because most features do not represent unique metabolites of biological origin. Rather, the majority of features correspond to contaminants introduced through sample handling, artifacts due to informatic errors and instrument noise, and redundant signals of the same metabolite resulting from adducts, oligomers, naturally occurring isotopes, etc. [10, 11]. To date, these classes of features have been annotated after data collection with software that operates independent of acquisition [12–18]. Here we set out to identify non-biological and redundant features in real time, while untargeted metabolomic data are being acquired, so that we can maximize MS/MS coverage and optimize collision energies by focusing only on unique biological signals.

2. Methods

Unlabeled and uniformly $^{13}$C-labeled *E. coli* cells were extracted with a 2:2:1 mixture of methanol, acetonitrile, and water. The extract was then vacuum concentrated and resuspended in either a 2:1 mixture of acetonitrile and water or a 2:1:1 mixture of isopropanol, methanol, and water. The unlabeled and labeled extracts were mixed at ratios of 1:1 or 1:2 for credentialing analysis. MCF-7 cells and yeast were extracted with cold methanol, vacuum concentrated, and resuspended in a 2:1 mixture of acetonitrile and water. Zebrafish blood was allowed to clot for 10 min at room temperature and centrifugation was used to obtain serum. Serum was diluted 1:10 with a 2:1 mixture of acetonitrile and water.

Three LC/MS configurations were used in this work: a SeQuant ZIC-pHILIC separation followed by analysis with a Thermo Scientific Orbitrap ID-X, a CORTECS C8 separation followed by analysis with an Agilent 6545 QTOF, and a SeQuant ZIC-pHILIC separation followed by analysis with a SCIEX 5600+ TripleTOF. In all experiments, 2 μL of sample was injected. For iterative DDA, an in-house R script based on MSnbase [19] was used to generate exclusion lists containing $m/z$ values and retention-time windows of ±10 s. To annotate degeneracies on the fly during acquisition of ID-X data, naturally occurring $^{13}$C isotopes and the adducts listed in Supplementary Table 1 were calculated for each $m/z$ value detected to group degenerate features. Analysis of a blank sample in MS1 mode was used to generate an exclusion list. Only features having a natural-abundance $^{13}$C isotope, an intensity threshold greater than 2e4, and a peak width less than 0.5 min were considered as potential candidates for MS/MS analysis. The [M-H]$^-$ ions detected in a full MS1 scan of the research sample were transferred to the inclusion list. Additionally, $m/z$ values from the research sample that were also detected in the blank run were transferred to the exclusion list.
when they were not at least five times more intense than in the blank sample. Both the inclusion and the exclusion lists assumed a mass error \( \leq 5 \) ppm.

Additional method details can be found in the Supplementary Material.

3. Results and discussion

In this work, we aimed to compare the ability of various MS/MS acquisition strategies to provide comprehensive fragmentation coverage of unique metabolites within a complex biological sample. Unless noted, the chromatography was kept constant across all experiments irrespective of the MS/MS acquisition strategy. Given that MS/MS coverage is highly dependent upon metabolite separation, it is important to point out that we applied a well-established chromatographic method in metabolomics using a 25-min gradient and a ZIC-pHILIC column \( \text{[20]} \).

As a benchmark, we first analyzed an unlabeled E. coli sample with a ZIC-pHILIC column and an Orbitrap ID-X mass spectrometer in negative ion mode. In total, 9432 features were detected. To establish which of these features corresponded to unique biological metabolites, we applied the credentialing workflow as described previously \( \text{[21, 22]} \). In brief, E. coli were cultured either in natural-abundance media or uniformly \( ^{13} \)C-labeled media. The extracted samples were then mixed at a 1:1 or a 1:2 ratio and analyzed by LC/MS. All metabolites produced by E. coli cultured under these conditions will occur in their unlabeled and labeled forms, thereby yielding \( ^{12} \)C and \( ^{13} \)C signals that are separated by the number of carbons in the molecule. Additionally, the intensities of the \( ^{12} \)C and \( ^{13} \)C signals approximate the ratio in which the unlabeled and labeled samples were mixed. After processing the data files for peak detection and degeneracy, we systematically filtered features from the unlabeled sample that did not have an appropriate \( ^{13} \)C counterpart. The approach returned 313 credentialed features representing unique metabolites of biological origin (Supplementary Table 2), a percentage consistent with previous reports applying different methods \( \text{[10, 11]} \). From our list of 313 credentialed features, we named 179 unique metabolites on the basis of MS1 and MS/MS data (optimized dot product > 80, Supplementary Figs. 1–27, and Supplementary Table 2). Interestingly, 42 of these metabolites are not currently included in the E. coli Metabolome Database.

We wish to emphasize that credentialing was only used in this work to establish a ground truth of which features in our dataset correspond to unique biological metabolites. Each MS/MS acquisition strategy was compared by analyzing unlabeled samples to determine the number of unique biological metabolites for which MS/MS data were obtained. None of the MS/MS acquisition modes being evaluated relied upon \( ^{13} \)C labeling.

3.1. DIA coverage of credentialed features

Fragmenting all coeluting features without any \( m/z \) filtering allowed fragmentation data to be collected on all unique biological features in a single ZIC-pHILIC run. This strategy, however, caused the MS/MS data of all features to be chimeric (i.e., contain fragments from more than one precursor in the same MS/MS spectrum). There are several programs available to computationally deconvolve chimeric data, which is typically necessary to
identify metabolites with high confidence from chimeric data [3, 7, 8, 23]. Fragmenting coeluting features in 25 Da bins, as is often done in Sequential Window Acquisition of All Theoretical Fragments (SWATH) experiments, reduced the number of MS/MS scans having fragments from more than one precursor but the percentage of chimeric MS/MS spectra was still 54% (Fig. 1A). Considering only the credentialed features corresponding to unique metabolites, 177 out of 313 still had chimeric MS/MS spectra when filtering with 25 Da bins. In contrast, only 17 out of the 313 credentialed features had chimeric MS/MS spectra when precursors were isolated using narrow 1.5 Da bins. Some representative examples from our E. coli sample showing changes in fragmentation data as a function of the size of the MS/MS isolation window are provided in Fig. 1B and Supplementary Figs. 28–31. A similar trend for signal-to-noise ratios was observed when we collected MS/MS data at increasing scan speeds with a SCIEX 5600+ TripleTOF (Fig. 1C and Supplementary Figs. 32–35), which highlights that scanning narrower MS/MS isolation windows at faster speeds is impractical. These results demonstrate that DIA is highly effective at generating fragmentation spectra for thousands of features in a single chromatographic run (without any requirement for prior analysis of a blank or research sample). The workflow achieves 100% MS/MS coverage in a single run, but at the expense of selectivity. When the experimental objective is to acquire MS/MS data on only those features representing unique biological metabolites, which are a small number of the total ions detected, DIA is relatively inefficient but does reach 100% coverage (Fig. 2A).

3.2. Annotating data during acquisition

Comprehensive annotation of non-biological and redundant features requires advanced computational tools and isotope labeling [11, 12, 24–27]. Many chemical contaminants and degenerate features, however, may be found by using relatively simple data-processing strategies such as blank subtraction and accurate-mass calculations [28–31]. We wished to determine whether these simple data-processing strategies, which can now be applied in acquisition software such as AcquireX, would filter a sufficient number of features to enable comprehensive MS/MS coverage of unique biological metabolites in a single LC/MS/MS run without sacrificing selectivity. Additionally, we wished to determine whether filtering on the basis of blank subtraction would faithfully remove only non-biological features.

We applied a three-step workflow to annotate features corresponding to background and compound degeneracy when acquiring data on the Orbitrap ID-X (Fig. 2B). A blank sample was first analyzed to identify possible contaminant signals. Ideally, the blank sample should go through each step of sample processing (e.g., extraction, concentration, etc.) so that it is exposed to all potential sources of contamination. Second, a research sample was analyzed to identify degenerate features (see Supplementary Table 1), each of which was added to the exclusion list. All features not on the exclusion list or at least 5-fold higher in the sample than in the blank were added to the inclusion list and targeted for MS/MS analysis, representing ions that potentially correspond to unique biological metabolites. In cases where multiple sample groups are being profiled, pooling an aliquot of each sample from the study will increase the likelihood that all potential features are represented for annotation and is likely to prevent having to perform LC/MS/MS analysis on multiple individual samples. Finally, the third step was to analyze the research sample from step 2 in MS/MS
mode with a narrow 1.5 Da isolation window. When all features on the inclusion list could not be targeted for MS/MS analysis in a single chromatographic run by applying user-defined parameters, then we could perform subsequent runs until complete coverage was achieved. Once MS/MS data were acquired for a precursor on the inclusion list, that precursor was moved to the exclusion list for subsequent runs. When extra time was available to perform MS/MS analysis because no ions from the inclusion list were detected, then fragmentation data were acquired in a conventional DDA fashion where the most intense ions not on the exclusion list were targeted.

To evaluate the reliability of this workflow for collecting MS/MS data on unique biological metabolites, we analyzed an unlabeled E. coli sample. The same sample was independently subjected to an off-line credentialing analysis by using stand-alone R scripts to annotate unique biological metabolites manually as described previously [10, 21]. Out of the 9432 total features detected from the unlabeled sample on the ID-X, 9119 were determined to be redundant or non-biological by credentialing (Fig. 3A). Rather than filtering 9119, however, only 8392 were filtered on the fly by using blank subtraction and simple adduct calculations (Fig. 3B). This process left 1040 features for MS/MS analysis (Supplementary Table 3), rather than 313. Importantly, 288 of the 313 features determined to be unique biological metabolites by credentialing were in this list of 1040 (Fig. 3C), yielding a 92% success rate.

We point out that MS/MS data for each of the 1040 remaining features could be obtained in a single chromatographic run using 1.5 Da isolation windows and 100 ms maximum injection time for each precursor. When no ions from the inclusion list were detected, then fragmentation data were collected on the most intense ion(s) not on the exclusion list as time permitted. This process led to the acquisition of MS/MS data for >95% of all credentialed features corresponding to unique biological metabolites in the first LC/MS/MS run. After four LC/MS/MS runs, fragmentation data for all but 1 of the remaining credentialed features had been acquired. The only credentialed feature not to be triggered for fragmentation after four LC/MS/MS runs had an m/z of 255.2329. We identified this signal to be palmitate, which is known to be a high-concentration contaminant when using plastic consumables [32]. Indeed, the peak in the E. coli sample was not five times higher than the peak in the blank, and m/z 255.2329 was therefore moved to the exclusion list as a non-biological background signal. Notwithstanding, after four LC/MS/MS runs, fragmentation data had been acquired with narrow precursor isolation windows on >99% of the credentialed features representing unique biological metabolites. It is interesting to consider why more of the credentialed features representing unique biological metabolites were not placed on the inclusion list for MS/MS analysis in the first chromatographic run. From manual inspection, we identified two main causes. First, the chromatographic peak widths of these features were greater than 0.5 min, which was the threshold for considering a signal as a candidate for MS/MS analysis. Second, the peak intensities of these features were below 2e4, which we set as the threshold to place ions on the inclusion list. Although we could modify these parameters so that more credentialed features were placed on the inclusion list, this also led to more non-credentialed features on the inclusion list. When the inclusion list gets too large, then all of its ions can no longer be targeted for MS/MS analysis in a single chromatographic run. It is also interesting to consider why non-biological and degenerate features were added to the inclusion list. One major source of non-biological features is artifacts. Unlike
contaminants, these signals do not arise from actual molecules. Instead, artifacts are a result of electronic noise from the mass spectrometer, distorted chromatographic baselines, informatic errors, etc. [33]. Given that artifacts detected in a research sample typically are not also detected in a blank sample, artifacts cannot be simply filtered by blank subtraction. In contrast, most artifacts are successfully filtered by isotope-based credentialing. Uncommon degeneracies led to another source of features that were filtered in our manual processing of the data but that were not filtered on the fly. These included single-analyte multimers, multi-analyte multimers, in-source fragments, etc.

### 3.3. Real-time optimization of fragmentation voltages

One potential challenge of generating MS/MS spectra on the fly during acquisition of an untargeted metabolomic dataset is determining the appropriate fragmentation voltage to use for each precursor. Often in untargeted metabolomics, re-injection of a sample is required to test multiple energy levels for fragmenting each precursor. Achieving comprehensive MS/MS coverage using inappropriate collision energies would not be productive for metabolite identification. To address this issue, we exploited the parallel nature of the Orbitrap ID-X that enables detection of ions not only with the Orbitrap but also with a linear ion-trap. Specifically, we used a series of hidden ion-trap scans to find an optimal collision energy that could then be applied in a subsequent high-resolution Orbitrap scan during acquisition of high-resolution data.

Our approach to finding optimal collision energies was to produce a pseudo-breakdown curve for each precursor by comparing the intensity of the precursor remaining in an ion-trap scan at different collision energies to the precursor intensity in an ion trap scan of 0 collision energy (Fig. 4A). We surmised that a specific percentage of precursor breakdown would lead to optimal fragmentation spectra. To test this prediction, we evaluated a mixture of ~100 metabolite standards with reference spectra in mzCloud using our ZIC-pHILIC method (see Supplementary Information). We used five different collision energies to produce pseudo-breakdown curves for every standard (0, 15, 30, 45, 60% NCE). MS/MS data from each point were then matched against the mzCloud MS/MS database. We considered matches with spectral similarity scores of 80 or higher, as assessed by the normalized dot product, to be high-confidence matches. Matches with spectral similarity scores less than 80 were considered to be low-confidence matches. We concluded that using a collision energy that reduces the precursor intensity to 20% gives the best overall identification confidence and is therefore optimal. These optimized collision energies provided better identification accuracy compared to using a single fixed collision energy (Fig. 4B). As expected, evaluation of MS/MS spectra from 5575 small-molecule standards in the mzCloud database showed that optimal fragmentation voltages are highly compound specific (Fig. 4C). These data underscore the need to optimize collision energies in real time during data collection, rather than simply using a single fixed value.

To increase analysis speed when determining the optimal collision energy, we scanned only a small range around the precursor ion (±2 m/z units) with the ion trap. Additionally, to minimize reductions in duty cycle, the ion-trap scans were performed during the preceding Fourier Transform transient period and used to determine the optimal collision energy for
the next Fourier Transform injection. Overall, the process ensured that fragmentation was performed for all precursors at collision energies that yield suitable MS/MS spectra for library searching. Most importantly, optimization of fragmentation voltages occurred in real time and could be coupled with blank subtraction and adduct annotation so that high-quality MS/MS data were obtained for unique biological metabolites.

When possible, the time gained by targeting only biological features could be used to collect additional collision-induced dissociation (CID) fragmentation spectra. We found that CID can provide complementary information to higher-energy collisional dissociation (HCD) and lead to additional metabolite identifications. Although spectral libraries are currently not well populated with MS/MS spectra from CID fragmentation, this different type of fragmentation can provide useful information. One example is leucine in negative mode. HCD spectra from an ID-X and fragmentation data from a QTOF (see METLIN) contain only the precursor m/z 130.0874. In CID, in contrast, two fragments were observed at m/z 84.0815 and 113.0603. A substructure search in the mzCloud database revealed that these fragments are derived from leucine.

3.4. Extension to other sample types

To demonstrate the applicability of our approach beyond E. coli, we next evaluated additional samples of varying complexity. We first analyzed a 20 μM mixture of 30 chemically diverse standards with the same ZIC-pHILIC method detailed above (Supplementary Table 4). Although we only added 30 standards to the mix, we detected nearly two orders of magnitude more features (2095). Of these features, 2032 were annotated as background or degeneracies during data acquisition. Among the 63 features that remained was the [M-H]− ion of each of the 30 standards (Supplementary Table 5). Thus, consistent with the credentialing analysis of E. coli, our approach resulted in a 10-fold reduction of data during acquisition without losing any metabolite coverage.

We also applied our approach to complex biological samples evaluated with the same analytical method (Fig. 5A). For yeast (Saccharomyces cerevisiae), we detected 10783 total features. Removing background and degeneracies during acquisition reduced the MS/MS list to only 1310 unique signals of biological relevance. We observed a similarly small number of the total features targeted for MS/MS analysis when evaluating human cells in culture (MCF-7) and serum samples from animals (zebrafish). Our results indicate that it is common to detect ten times more features than unique metabolites in our LC/MS profiling experiments, irrespective of sample type and complexity. Determining which features result from background and compound degeneracy during acquisition reduced the MS/MS burden by 90% or more for all samples that we examined.

Lastly, we sought to quantify the benefits of our workflow relative to iterative DDA. In iterative DDA, precursors are fragmented on the basis of their signal intensity as in conventional DDA. Once fragmented, the precursor is then moved to the exclusion list so that it is not targeted again for MS/MS analysis in subsequent chromatographic runs [34, 35]. For comparison, we evaluated the percentage of credentialed compounds in which MS/MS data were collected from our workflow described here, conventional DDA, and iterative DDA (Fig. 5B). When comparing conventional DDA to iterative DDA, we found
that the MS/MS coverage of unique biological compounds increased by ~25% over four LC/MS/MS runs. We note that further increases may be possible after optimization of parameters. In contrast, however, our workflow was able to achieve comprehensive (>95%) MS/MS coverage of unique biological compounds in one LC/MS/MS run following an MS1 analysis of a blank and a research sample. Within four LC/MS/MS runs, the approach obtained MS/MS data for all but one credentialed compound whereas even the more advanced iterative DDA method had only reached ~50% coverage. We also observed that the MS/MS spectra obtained were more often acquired near the apex of analyte elution (Supplementary Fig. 36). Furthermore, given that fewer compounds needed to be fragmented, the time spent on each precursor could be extended to increase the overall quality of fragmentation data for low-abundant features.

3.5. Limitations

A limitation of the workflow described herein is that it is specific to the Orbitrap ID-X. We rely on the Tribrid architecture to optimize MS/MS voltages in a single experiment by applying hidden ion-trap scans. Additionally, the Orbitrap ID-X acquisition software annotates select feature degeneracies during data acquisition and allows a rolling inclusion/exclusion list to be applied across an entire LC/MS worklist. Although it was possible for us to implement an analogous type of workflow on QTOF instruments, some steps (e.g., optimization of MS/MS voltages) had to be broken into multiple experimental runs. Notwithstanding, we validated that we obtained comparable results with a QTOF-based version of the workflow (Supplemental Figs. 37–38).

Another limitation of the approach presented is that it only focuses on unique metabolites originating from the biological sample. Depending on experimental goals, however, other features may also be of interest. Identifying non-biological signals, for example, may reveal unexpected contaminants introduced during sample handling. Further, it has been demonstrated that data from degenerate features can facilitate compound identification in some computational workflows [33].

Finally, to optimize collision energies in a single experiment, we tested five fragmentation voltages with hidden ion-trap scans. These five voltages were selected prior to the analysis and applied to all ions, irrespective of their molecular properties. The goal is to determine when 80% of the precursor has been broken down, but it is conceivable that none of the voltages selected achieve this degree of fragmentation. Although we reached 80% fragmentation for more than 97% of the ~100 polar metabolite standards that we tested, other molecules may not fragment as easily (Fig. 4C). Other work has shown that most metabolites fragment best with NCE values between 10 and 50%, but several compounds require higher collision energies [36]. Thus, we anticipate further improvements to real-time optimization of collision energies in the future where different ranges of voltages can be applied based on experimental context.

4. Conclusions

In summary, we have demonstrated that the thousands of features typically detected by LC/MS in untargeted metabolomics can be reduced by an order of magnitude during data
acquisition by blank subtraction and annotation of common degeneracies (Fig. 3B).

Although this workflow is considerably less rigorous than isotope-based credentialing at annotating unique biological metabolites, it still effectively removes thousands of non-biological and redundant signals on the fly (Fig. 5C). Most importantly, because unique metabolites of biological origin are putatively annotated during acquisition, they can be selectively targeted for MS/MS analysis. The 10-fold reduction in the number of precursors to target for fragmentation enables comprehensive MS/MS coverage to be achieved in a single experiment, while still using narrow isolation windows for improved selectivity. To ensure that the MS/MS data obtained are informative, we established a method to optimize collision energies in real time by using hidden ion-trap scans that maximize the amount of information in fragmentation patterns.

We conclude that DIA was the only fragmentation approach to achieve complete (i.e., 100%) coverage of all unique biological metabolites in a single chromatographic run, but it came with the complication that nearly 60% of the MS/MS spectra were chimeric and therefore required deconvolution. Conventional DDA provided mostly non-chimeric MS/MS spectra, but achieved less than 40% coverage of unique biological metabolites in a single LC/MS/MS run. By using iterative DDA, coverage could be increased to ~50% but it required three additional LC/MS/MS runs. Integrating on-the-fly filtering and DDA provided a significant improvement in coverage, while still yielding mostly non-chimeric fragmentation data. With the workflow presented, fragmentation data were acquired on 95% of the unique biological metabolites in a single LC/MS/MS run and on >99% of the unique biological metabolites in four LC/MS/MS runs. Notably, however, the presented workflow required running a blank and a pooled research sample in MS1 mode prior to LC/MS/MS analysis whereas DIA, conventional DDA, and iterative DDA did not.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

The authors thank Ioanna Ntai, Amanda Souza, Derek Bailey, John Butler, and Ralf Tautenhahn at Thermo Fisher Scientific for their technical assistance with the Orbitrap ID-X mass spectrometer. This work was supported by funding from the National Institutes of Health grants U01 CA235482 (G.J.P.), R35 ES028365 (G.J.P.), and R24 OD024624 (G.J.P.) as well as financial support from the Edward Mallinckrodt, Jr. Foundation, the Pew Charitable Trusts, and Nestle Purina (G.J.P.).

REFERENCES

[1]. Fenaille F, Barbier Saint-Hilaire P, Rousseau K, Junot C, Data acquisition workflows in liquid chromatography coupled to high resolution mass spectrometry-based metabolomics: Where do we stand?, J Chromatogr A, 1526 (2017) 1–12. [PubMed: 29074071]
[2]. Mahieu NG, Genenbacher JL, Patti GJ, A roadmap for the XCMS family of software solutions in metabolomics, Curr Opin Chem Biol, 30 (2016) 87–93. [PubMed: 26673825]
[3]. Tsugawa H, Cajka T, Kind T, Ma Y, Higgins B, Ikeda K, Kanazawa M, VanderGheynst J, Fiehn O, Arita M, MS-DIAL: data-independent MS/MS deconvolution for comprehensive metabolome analysis, Nat Methods, 12 (2015) 523–526. [PubMed: 25938372]
[4]. Bonner R, Hopfgartner G, SWATH data independent acquisition mass spectrometry for metabolomics, TrAC Trends in Analytical Chemistry, 120 (2019) 115278.
[5]. Simons B, Kauhanen D, Sylvänne T, Tarasov K, Duchoslav E, Ekroos K. Shotgun Lipidomics by Sequential Precursor Ion Fragmentation on a Hybrid Quadrupole Time-of-Flight Mass Spectrometer. Metabolites, 2 (2012) 195–213. [PubMed: 24957374]

[6]. Plumb RS, Johnson KA, Rainville P, Smith BW, Wilson ID, Castro-Perez JM, Nicholson JK. UPLC/MSE: a new approach for generating molecular fragment information for biomarker structure elucidation, Rapid Communications in Mass Spectrometry, 20 (2006) 2234–2234.

[7]. Nikolskiy I, Mahieu NG, Chen YJ, Tautenhahn R, Patti GJ. An untargeted metabolomic workflow to improve structural characterization of metabolites, Anal Chem, 85 (2013) 7713–7719. [PubMed: 23829391]

[8]. Yin Y, Wang R, Cai Y, Wang Z, Zhu ZJ. DecoMetDIA: Deconvolution of Multiplexed MS/MS Spectra for Metabolite Identification in SWATH-MS-Based Untargeted Metabolomics, Anal Chem, 91 (2019) 11897–11904. [PubMed: 31436405]

[9]. Broeckling CD, Hoyes E, Richardson K, Brown JM, Prenni JE. Comprehensive Tandem-Mass-Spectrometry Coverage of Complex Samples Enabled by Data-Set-Dependent Acquisition, Anal Chem, 90 (2018) 8020–8027. [PubMed: 29846054]

[10]. Mahieu NG, Patti GJ. Systems-Level Annotation of a Metabolomics Data Set Reduces 25000 Features to Fewer than 1000 Unique Metabolites, Anal Chem, 89 (2017) 10397–10406. [PubMed: 28914531]

[11]. Wang L, Xing X, Chen L, Yang L, Su X, Rabitz H, Lu W, Rabinowitz JD. Peak Annotation and Verification Engine for Untargeted LC-MS Metabolomics, Anal Chem, 91 (2019) 1838–1846. [PubMed: 30586294]

[12]. Mahieu NG, Spalding JL, Gelman SJ, Patti GJ. Defining and Detecting Complex Peak Relationships in Mass Spectral Data: The Mz.unity Algorithm, Anal Chem, 88 (2016) 9037–9046. [PubMed: 27513885]

[13]. Broeckling CD, Afstar FA, Neumann S, Ben-Hur A, Prenni JE. RAMClust: a novel feature clustering method enables spectral-matching-based annotation for metabolomics data, Anal Chem, 86 (2014) 6812–6817. [PubMed: 24927477]

[14]. Kuhl C, Tautenhahn R, Bottcher C, Larson TR, Neumann S. CAMERA: an integrated strategy for compound spectra extraction and annotation of liquid chromatography/mass spectrometry data sets, Anal Chem, 84 (2012) 283–289. [PubMed: 22111785]

[15]. Brown M, Dunn WB, Dobson P, Patel Y, Winder CL, Francis-McIntyre S, Begley P, Carroll K, Broadhurst D, Tseng A, Swainston N, Spasic I, Goodacre R, Kell DB. Mass spectrometry tools and metabolite-specific databases for molecular identification in metabolomics, Analyst, 134 (2009) 1322–1332. [PubMed: 19562197]

[16]. DeFelice BC, Mehta SS, Samra S, Cajka T, Wancewicz B, Fahrmann JF, Fiehn O. Mass Spectral Feature List Optimizer (MS-FLO): A Tool To Minimize False Positive Peak Reports in Untargeted Liquid Chromatography-Mass Spectroscopy (LC-MS) Data Processing, Anal Chem, 89 (2017) 3250–3255. [PubMed: 28225594]

[17]. Xu YF, Lu W, Rabinowitz JD. Avoiding misannotation of in-source fragmentation products as cellular metabolites in liquid chromatography-mass spectrometry-based metabolomics, Anal Chem, 87 (2015) 2273–2281. [PubMed: 25591916]

[18]. Neumann S, Thum A, Böttcher C. Nearline acquisition and processing of liquid chromatography-tandem mass spectrometry data, Metabolomics, 9 (2013) 84–91.

[19]. Gatto L, Lilley KS, MSnbase-an R/Bioconductor package for isobaric tagged mass spectrometry data visualization, processing and quantitation, Bioinformatics, 28 (2011) 288–289. [PubMed: 22113085]

[20]. Spalding JL, Naser FJ, Mahieu NG, Johnson SL, Patti GJ. Trace Phosphate Improves ZIC-phILIC Peak Shape, Sensitivity, and Coverage for Untargeted Metabolomics, J. Proteome Res, 17 (2018) 3537–3546. [PubMed: 30160483]

[21]. Mahieu NG, Huang X, Chen YJ, Patti GJ. Credentialing features: a platform to benchmark and optimize untargeted metabolomic methods, Anal Chem, 86 (2014) 9583–9589. [PubMed: 25160088]

[22]. Wang L, Naser FJ, Spalding JL, Patti GJ. A Protocol to Compare Methods for Untargeted Metabolomics, Methods Mol Biol, 1862 (2019) 1–15. [PubMed: 30315456]
[23]. Du X, Zeisel SH, Spectral deconvolution for gas chromatography mass spectrometry-based metabolomics: current status and future perspectives, Comput Struct Biotechnol J, 4 (2013) e201301013. [PubMed: 24688694]

[24]. Kachman M, Habra H, Duren W, Wigginton J, Sajjakulnukit P, Michailidis G, Burant C, Karnovsky A, Deep annotation of untargeted LC-MS metabolomics data with Binner, Bioinformatics, (2019).

[25]. Stupp GS, Clendinnen CS, Ajredini R, Szewc MA, Garrett T, Menger RF, Yost RA, Beecher C, Edison AS, Isotopic ratio outlier analysis global metabolomics of Caenorhabditis elegans, Anal Chem, 85 (2013) 11858–11865. [PubMed: 24274725]

[26]. Senan O, Aguilar-Mogas A, Navarro M, Capellades J, Noon L, Burks D, Yanes O, Guimerà R, Sales-Pardo M, CliqueMS: a computational tool for annotating in-source metabolite ions from LC-MS untargeted metabolomics data based on a coelution similarity network, Bioinformatics, 35 (2019) 4089–4097. [PubMed: 30903689]

[27]. Domingo-Almenara X, Montenegro-Burke JR, Guijas C, Majumder EL, Benton HP, Siuzdak G, Autonomous METLIN-Guided In-source Fragment Annotation for Untargeted Metabolomics, Anal Chem, 91 (2019) 3246–3253. [PubMed: 30681830]

[28]. Lenfant M, Wdzieczak-Bakala J, Guittet E, Prome JC, Sotty D, Frindel E, Inhibitor of hematopoietic pluripotent stem cell proliferation: purification and determination of its structure, Proc Natl Acad Sci U S A, 86 (1989) 779–782. [PubMed: 29159777]

[29]. Windig W, Phalp JM, Payne AW, A Noise and Background Reduction Method for Component Detection in Liquid Chromatography/Mass Spectrometry, Analytical Chemistry, 68 (1996) 3602–3606.

[30]. Ueno T, Sueyoshi T, Asfiida K, Takegami Y, Computer-Aided Deduction of Mass Spectra Detected on a Photographic Plate.II.Correlation of Intensity, Journal of the Mass Spectrometry Society of Japan, 22 (1974) 103–107.

[31]. Tautenhahn R, Böttcher C, Neumann S, Annotation of LC/ESI-MS Mass Signals, Springer Berlin Heidelberg, Berlin, Heidelberg, 2007, pp. 371–380.

[32]. Yao CH, Liu GY, Yang K, Gross RW, Patti GJ, Inaccurate quantitation of palmitate in metabolomics and isotope tracer studies due to plastics, Metabolomics, 12 (2016).

[33]. Sindelar M, Patti GJ, Chemical Discovery in the Era of Metabolomics, J Am Chem Soc, 142 (2020) 9097–9105. [PubMed: 32275430]

[34]. Koelmel JP, Kroeger NM, Gill EL, Ulmer CZ, Bowden JA, Patterson RE, Yost RA, Garrett TJ, Expanding Lipidome Coverage Using LC-MS/MS Data-Dependent Acquisition with Automated Exclusion List Generation, J Am Soc Mass Spectrom, 28 (2017) 908–917. [PubMed: 28265968]

[35]. Ivanisevic J, Want EJ, From Samples to Insights into Metabolism: Uncovering Biologically Relevant Information in LC-HRMS Metabolomics Data, Metabolites, 9 (2019).

[36]. Barbier Saint Hilaire P, Rousseau K, Seyer A, Dechaumet S, Damont A, Junot C, Fenaille F, Comparative Evaluation of Data Dependent and Data Independent Acquisition Workflows Implemented on an Orbitrap Fusion for Untargeted Metabolomics, Metabolites, 10 (2020) 158.
Fig. 1.
Tradeoff between selectivity, sensitivity, and coverage when collecting metabolomic MS/MS data. (A) Untargeted metabolomics was performed by using a 25-min ZIC-pHILIC separation and an Orbitrap ID-X mass spectrometer to analyze an *E. coli* extract. When 1.5 m/z isolation windows were used to collect MS/MS data, only 15% of the MS/MS scans were convolved with fragments from coeluting precursors. In contrast, when a 25 m/z MS/MS isolation window was used, 54% of the MS/MS scans were convolved with fragments from coeluting precursors. (B) Fragmentation of glucose on an Orbitrap ID-X as a representative example of MS/MS spectra when wider isolation windows are used. Given that glucose coelutes with many other features in the applied method, larger MS/MS isolation windows result in chimeric spectra that must be computationally deconvolved before they can potentially be used to support metabolite identification. With a 1.5 m/z isolation window, a 99.4% match was returned by mzCloud. Match scores dropped below 20% when the m/z isolation window was 25 m/z or larger and no computational deconvolution was applied. (C) Fragmentation data of serine on a SCIEX 5600+ TripleTOF as a representative example of the decrease in signal to noise that occurs when MS/MS acquisition time was decreased.
Fig. 2.
Workflow to improve the efficiency of acquiring non-chimeric MS/MS data on unique biological metabolites in untargeted metabolomics. (A) Contrasting the efficiency of DIA, conventional DDA, and the three-step workflow described here. The intersection of dotted lines represents an experiment where non-chimeric MS/MS data are acquired only for unique metabolites originating from the biological sample. Conventional DDA refers to a workflow where MS/MS fragmentation is triggered on ions with the highest intensity. (B) Schematic of the three-step workflow. The workflow reduces the MS/MS burden during acquisition of an untargeted metabolomic dataset, thereby providing more time to acquire non-chimeric MS/MS data for unique biological metabolites.
Fig. 3.
Integrating on-the-fly filtering and DDA effectively reduces the MS/MS burden by 90% during acquisition of an untargeted metabolomic dataset without sacrificing selectivity or coverage of unique biological metabolites. (A) Number of features detected as a function of retention time for *E. coli* analyzed with a ZIC-pHILIC method. The total number of features at any given time is substantially higher than the number of features that pass filtering or credentialing analysis. (B) The three-step workflow described selects a small fraction (~10%) of the total number of features detected in the *E. coli* experiment. (C) Ninety-two percent of the biological metabolites determined to be unique by isotope-based credentialing were included within the subset selected during data acquisition. All but one were fragmented after four analytical runs.
Fig. 4. Real-time optimization of fragmentation voltages in untargeted metabolomics. (A) Hidden ion-trap scans are used to analyze the precursor intensity after fragmentation with different collision energies. The first to yield an 80% reduction in precursor intensity is chosen for the Orbitrap scan. (B) Using an assisted collision energy target of 20% remaining precursor yields optimal fragmentation for identification. See supplementary methods for the formula used to calculate identification accuracy. (C) Analysis of 5575 small-molecule standards from the mzCloud database shows that optimal fragmentation voltages are highly compound specific.
Fig. 5.
Comparing MS/MS acquisition methods. (A) The workflow presented here led to similar reductions in the MS/MS burden for MCF-7 human breast cancer cells in culture, S. cerevisiae, and zebrafish serum. Data are presented as mean ± standard deviation (n = 3 replicates per group). (B) Percentage of credentialed compounds in which MS/MS data are obtained when fragmentation is performed by using three different methods. (C) Contrasting the “traditional” workflow for performing untargeted metabolomics to DIA and the on-the-fly filtering workflow. Traditionally, MS1 data are collected first and processed manually to select a small number of features for MS/MS analysis in a subsequent experiment. In DIA, MS1 and MS/MS are collected in the same experiment. MS/MS coverage is 100%, but data processing is complicated by chimeric MS/MS data that require deconvolution. When on-the-fly filtering is coupled with DDA, MS1 and MS/MS data are also acquired in the same experiment, but MS/MS data are mainly collected for unique metabolites of biological origin. The approach relies on putative annotation of background and degenerate features during execution of the LC/MS worklist. The workflow achieves >95% MS/MS coverage without sacrificing selectivity or sensitivity, which decreases post-acquisition analysis time.