Check for updates

Simulated-to-real benchmarking of acquisition methods in untargeted metabolomics

Joe Wandy[1], Ross McBride[2], Simon Rogers[2], Nikolaos Terzis[3], Stefan Weidt[1], Justin J. J. van der Hooft[4], Kevin Bryson[2], Rónán Daly[1] and Vinny Davies[3]*

[1]Glasgow Polyomics, University of Glasgow, Glasgow, United Kingdom, [2]School of Computing Science, University of Glasgow, Glasgow, United Kingdom, [3]School of Mathematics and Statistics, University of Glasgow, Glasgow, United Kingdom, [4]Bioinformatics Group, Wageningen University and Research, Wageningen, Netherlands

Data-Dependent and Data-Independent Acquisition modes (DDA and DIA, respectively) are both widely used to acquire MS2 spectra in untargeted liquid chromatography tandem mass spectrometry (LC-MS/MS) metabolomics analyses. Despite their wide use, little work has been attempted to systematically compare their MS/MS spectral annotation performance in untargeted settings due to the lack of ground truth and the costs involved in running a large number of acquisitions. Here, we present a systematic *in silico* comparison of these two acquisition methods in untargeted metabolomics by extending our Virtual Metabolomics Mass Spectrometer (ViMMS) framework with a DIA module. Our results show that the performance of these methods varies with the average number of co-eluting ions as the most important factor. At low numbers, DIA outperforms DDA, but at higher numbers, DDA has an advantage as DIA can no longer deal with the large amount of overlapping ion chromatograms. Results from simulation were further validated on an actual mass spectrometer, demonstrating that using ViMMS we can draw conclusions from simulation that translate well into the real world. The versatility of the Virtual Metabolomics Mass Spectrometer (ViMMS) framework in simulating different parameters of both Data-Dependent and Data-Independent Acquisition (DDA and DIA) modes is a key advantage of this work. Researchers can easily explore and compare the performance of different acquisition methods within the ViMMS framework, without the need for expensive and time-consuming experiments with real experimental data. By identifying the strengths and limitations of each acquisition method, researchers can optimize their choice and obtain more accurate and robust results. Furthermore, the ability to simulate and validate results using the ViMMS framework can save significant time and resources, as it eliminates the need for numerous experiments. This work not only provides valuable insights into the performance of DDA and DIA, but it also opens the door for further advancements in LC-MS/MS data acquisition methods.

KEYWORDS

liquid chromatography tandem mass spectrometry, metabolomics, data-dependent acquisition, data independent acquisition, digital twin

1 Introduction

Liquid chromatography tandem mass spectrometry (LC-MS/MS) is one of the dominant analytical platforms for untargeted metabolomics. LC-MS/MS acquisition strategies can be categorised as either Data-Dependent Acquisition (DDA) or Data-Independent Acquisition (DIA). In the former, MS2 scans are scheduled to target particular ions observed in full scan (MS1) survey scans. After each MS1 survey scan, a small number of ions will be prioritised (normally based upon their intensity) and fragmented in a series of MS/MS (MS2) scans. This will be followed by another MS1 survey scan from which the next batch of fragmentation events will be decided. In DIA, fragmentation is not based upon ions observed in survey scans but instead fixed m/z windows are isolated and fragmented, regardless of the ions present. The m/z windows can range from the whole m/z range (All-Ion Fragmentation, or AIF) or can be broken into a series of smaller windows that are iterated over in consecutive scans (e.g., Sequential Window Acquisition of all Theoretical Mass Spectra, or SWATH) (Gillet et al., 2012). DIA therefore removes the need to choose which ions to target at acquisition time at the cost of introducing an additional deconvolution step into the data analysis pipeline. Note that in this paper, we use the term "DIA" to refer to all data-independent acquisition methods, including both SWATH and AIF.

Various different DDA and DIA strategies have been introduced (Kaufmann and Walker, 2016; Guan et al., 2020; Davies et al., 2021; Guo et al., 2021) and although each new method is compared with other approaches, no clear consensus has emerged as to which overall strategy is best in which situation (Fernández-Costa et al., 2020; Guo and Huan, 2020a; Guo and Huan, 2020b). An advantage of DDA is that MS2 spectra are generated nearly ready to use, as each MS2 spectrum targets a particular ion, and we can be reasonably confident that the fragment ions observed do indeed come from the targeted ion. Sometimes multiple ions can end up in the same isolation window (Lawson et al., 2017), but this, in general, is not considered to be a major problem, as most of the times there is one dominant ion species giving rise to the mass fragments. Critics of DDA point to the lack of reproducibility (i.e., due to its stochastic nature, different peaks will be fragmented if the same sample is injected twice) and the low coverage–only a subset of the ions present in the sample are fragmented (Zhang et al., 2020). On the other hand, DIA offers chromatographic traces for all detected fragment ions, which can be useful in the identification of complex samples, particularly those containing isomers, where chromatographic information can help distinguish between them. The increased fragmentation capabilities of DIA can also be beneficial in the case of nearby-eluting isomers, where DDA might trigger only one MS2 spectrum due to dynamic exclusion. As DIA does not prioritise based on the contents of MS1 survey scans, it is more reproducible (we know beforehand exactly the properties of any scan). It also, at least in theory, overcomes the coverage issues as it is able to assign MS2 fragments to any detected MS1 ion.

The analysis of complex samples using DIA presents a unique set of challenges. In DIA, multiple ions from different compounds are fragmented during each MS2 scan, leading to complex spectra that require deconvolution to identify individual components. This can be particularly challenging in untargeted metabolomics where the compounds present and their fragmentation patterns are unknown prior to analysis, making it difficult to set up a pre-determined table of metabolites of interest to compare their fragmentation spectra to. Deconvolution, using software such as MS-DIAL (Tsugawa et al., 2015), is a process that separates the complex spectra generated by the fragmentation of multiple compounds into individual components, enabling the identification of the compounds present in the sample. However, deconvolution is a complex process and it is widely recognised that the high coverage provided by DIA comes at the cost of lower quality spectra (Bern et al., 2010).

Although there have recently been two studies (Guo and Huan, 2020a; Guo and Huan, 2020b) that compare DDA with DIA for its use in untargeted metabolomics, in general comparisons between the two acquisition strategies are still lacking. This is mainly due to the lack of ground truth for real experimental matrices and the cost of running large numbers of injections. While validation on real injections is vital, simulation can also play an important role in answering such questions. For example, the number of ions that can be targeted in a DDA analysis is a complex function of scan times, chromatographic peak widths, and the number of peaks eluting at a particular time. Similarly, the number of spectra that can be accurately deconvoluted in a DIA analysis depends on the number of co-eluting ions in the same isolation window, and how correlated their chromatographic profiles are. In both cases, analysis on real injections is hampered by a lack of knowledge of the true make-up of that sample, or samples being overly simplistic if they consist of just a handful of known standards. Simulation can overcome these limitations by permitting complete control over the ground truth in terms of both the fragment spectra present, the number of chemical ions in the sample, and how and when they elute.

In our previous work we introduced ViMMS (Wandy et al., 2019; Wandy et al., 2022), a virtual metabolomics mass spectrometry simulator framework, and demonstrated how it could be used to develop new, and improve existing DDA strategies. One of the main advantages of using ViMMS to develop new strategies lies in its ability to develop methods without the overhead costs of using a real mass spectrometer (MS). Within the ViMMS framework, it is possible to prototype methods and optimise them *in silico* before transferring the developed method for validation to an actual instrument. New DDA methods, such as WeightedDEW and SmartROI, have been developed on top of ViMMS and shown to outperform Top-N in terms of the number of peaks that were fragmented in both simulated and real experiments (Davies et al., 2021). More recent work has used ViMMS to develop improved methods for multi-sample and -injection DDA strategies (McBride et al., 2023).

In the current work we utilised ViMMS as a method for accurately benchmarking DIA and DDA through simulations. We firstly introduce two DIA controllers (SWATH and AIF) into the ViMMS framework, before conducting extensive simulated experiments to evaluate the comparative performance of DIA and DDA (Top-N) across a range of different simulated conditions, effectively creating a "digital twin" of the real situation. The DIA methods prototyped on the simulator were transferred with ease to run on an actual MS instrument with no code changes in the implementation–evidencing the capability of

ViMMS to develop DIA methods in a simulated-to-real setting. As such, we used ViMMS to validate the *in silico* experiments using complex beer samples in real LC-MS/MS experiments. The real experimental results were first benchmarked using an online reference spectral library (GNPS/NIST14) to assess spectral matches to a database of known molecules. This approach follows that of Guo and Huan (2020a), but importantly compares SWATH as well as AIF. Improving upon Guo and Huan (2020a), we also more systematically compare the results for DDA (Top-N) and DIA (SWATH and AIF) by evaluating our results against a database created on the specific sample using a recently developed, but computationally expensive multiple injection data acquisition method (McBride et al., 2023). This multi-injection dataset was constructed specifically to evaluate the maximum spectral coverage from a realistic experimental setting, complementing the standard approach of evaluating against a database of known molecules.

Our results found that over a wide range of experimental conditions, DIA is generally more effective at fragmenting more features, both in simulations and reality. However, DDA outperforms DIA in terms of the number of chemical ions for which high-quality spectra are recovered. Based on simulated and real instrumental results, we were able to provide a clear, actionable guideline on when a particular acquisition method (whether DDA or DIA) should be used.

In summary, the contributions of this paper are as follows.

1) We have introduced two new DIA controllers (SWATH, AIF) into the ViMMS framework.
2) We have conducted extensive simulated experiments to evaluate the performance of DDA (Top-N) vs. DIA with a known *in silico* ground truth.
3) We have validated the simulated results through benchmarking on the actual instrument using two reference datasets: the GNPS/NIST14, and our own Multi-Injection libraries.

2 Materials and methods

2.1 DDA and DIA data acquisition

To validate performance on a real instrument, we performed DDA and DIA acquisition using six beer samples. Each beer sample was acquired once using the Fullscan, Top-N, AIF and SWATH controllers in ViMMS when connected to an actual mass spectrometer (more details in Section 2.4). To create the Multi-Injection reference library (described in Section 2.5), each beer sample was further injected ten times repeatedly for acquisition using the Intensity Non-overlap method in ViMMS (McBride et al., 2023).

For sample extraction, chloroform and methanol were added to beer samples (detailed names in Supplementary Section S1) in a 1:1:3 ratio and mixed with a vortex mixer. The mixture was centrifuged to remove protein and other precipitates, and the supernatant was stored at −80°C. Chromatographic separation was performed with a Thermo Scientific UltiMate 3000 RSLC liquid chromatography system and a SeQuant ZIC-pHILIC column. The gradient elution used 20 mM ammonium carbonate and acetonitrile. 10 μL of each sample was injected with an initial 80% acetonitrile concentration,
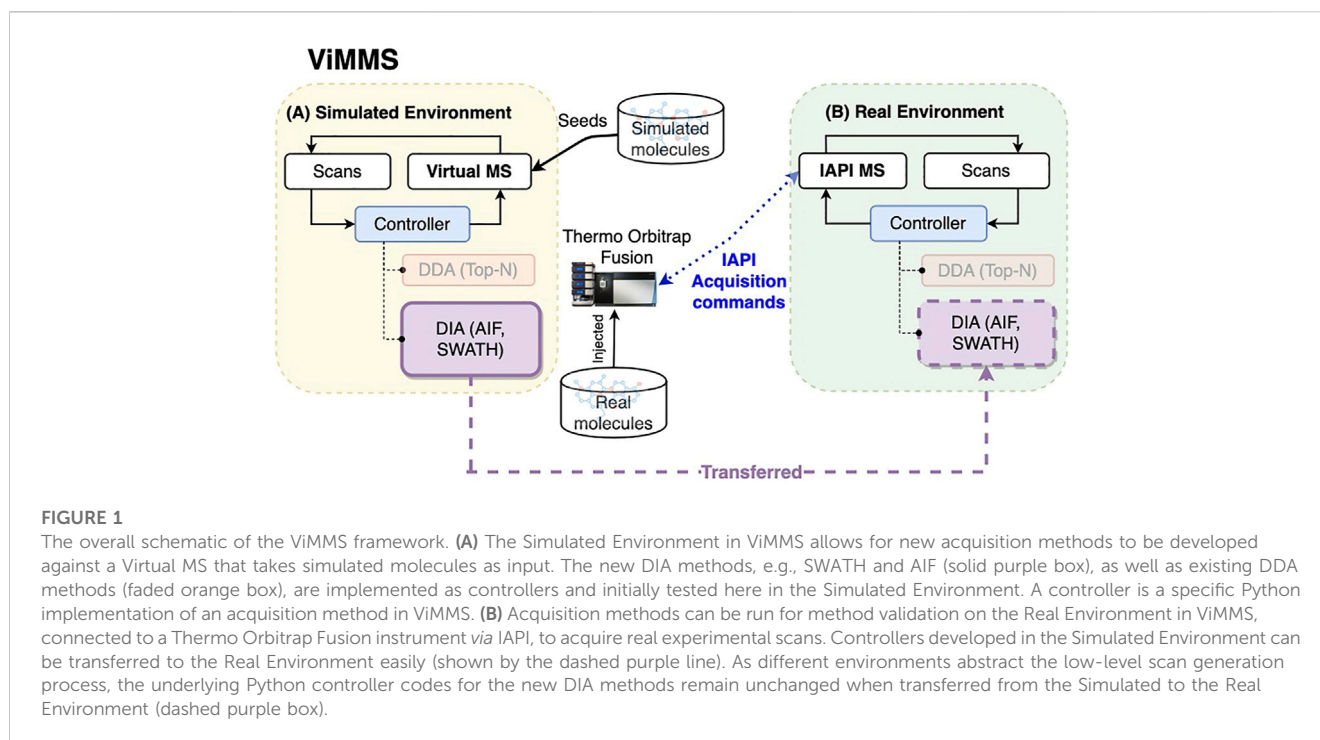
maintaining a linear gradient from 80% to 20% acetonitrile over 15 min, and finally a wash of 5% acetonitrile for 2 min, before re-equilibration at 80% acetonitrile for 9 min. The flow rate was 300 μL/min and the column oven temperature was 40°C.

A Thermo Orbitrap Fusion tribrid-series mass spectrometer was used to generate mass spectra data, controlled through Thermo Instrument Application Programming Interface (IAPI) managed by ViMMS (more details in Section 2.4). Full scan spectra were acquired in positive mode at a resolution of 120,000 and a mass range of 70–1,000 m/z. Fragmentation spectra for both DDA and DIA were acquired using the Orbitrap mass analyser at resolution 7,500. In DDA mode, precursor ions were isolated using 0.7 m/z width and fragmented with fixed HCD collision energy of 25%. The AGC was set as 200,000 for MS1 scans and 30,000 for MS2 scans. N was set to 10 Top-N. The dynamic exclusion window (DEW) was set to 15 s to prevent repeated fragmentation of the same ion. A minimum intensity threshold of 5,000 was also used before a precursor ion can be selected for MS2 fragmentation. For DIA (AIF), an MS1 source CID energy of 25% was used. For DIA (SWATH), a window of 100 m/z was used with no overlap between the windows.

2.2 Developing DIA methods using ViMMS

In previous work (Davies et al., 2021), ViMMS was used to develop DDA methods, but the capability of the framework is not limited to that. Here we introduced two new methods, SWATH and AIF, on top of the framework, demonstrating that DIA methods can also be developed on top of ViMMS. Rather than prioritising ions for fragmentation based on their abundance as is commonly done for DDA, DIA methods operate by fragmenting all precursor ions within a large m/z window. In AIF, all precursors in the entire m/z range are fragmented, whereas in SWATH, a series of smaller and potentially overlapping windows are used to fragment ions in the window. Both AIF and SWATH were implemented as controllers in ViMMS, allowing their performance to be benchmarked in the simulator and validated on an actual instrument easily, as previously done with DDA methods (Davies et al., 2021).

Figure 1 shows a schematic of the DIA method implementations in ViMMS illustrating how the new DIA methods were introduced. The ViMMS framework can be divided into two parts: a "Simulated Environment" where simulated scans are generated by querying synthetic chemicals, and a "Real Environment" where actual scans are generated through measurements using an LC-MS instrument. In the Simulated Environment (Figure 1A), the Virtual MS is seeded with synthetic molecules that are created by either sampling chemical databases or extracted from existing experimental mzML files. Once generated, molecules can be used to produce scans during virtual mass spectrometry. Scans are generated based on which molecules elute at a particular retention time, and generated chemicals are dispatched to the appropriate controllers. In ViMMS, a *controller* class is a specific implementation of an acquisition method in Python that follows a predefined Python interface to receive scans and schedule the next MS1 and MS2 scans. The new DIA methods are implemented as the SWATH and AIF controller classes in ViMMS (solid purple box in Figure 1A),

FIGURE 1
The overall schematic of the ViMMS framework. **(A)** The Simulated Environment in ViMMS allows for new acquisition methods to be developed against a Virtual MS that takes simulated molecules as input. The new DIA methods, e.g., SWATH and AIF (solid purple box), as well as existing DDA methods (faded orange box), are implemented as controllers and initially tested here in the Simulated Environment. A controller is a specific Python implementation of an acquisition method in ViMMS. **(B)** Acquisition methods can be run for method validation on the Real Environment in ViMMS, connected to a Thermo Orbitrap Fusion instrument *via* IAPI, to acquire real experimental scans. Controllers developed in the Simulated Environment can be transferred to the Real Environment easily (shown by the dashed purple line). As different environments abstract the low-level scan generation process, the underlying Python controller codes for the new DIA methods remain unchanged when transferred from the Simulated to the Real Environment (dashed purple box).

extending from the base Controller class (blue box in Figure 1A). Different experiments can be performed in the Simulated Environment, allowing for different data characteristics to be explored.

Once tested and optimised, the developed DIA controllers can be transferred to run on an actual mass spectrometer instrument with no change to their implementations. This is accomplished by swapping the Simulated Environment to a Real Environment (Figure 1B), where an IAPI MS is used in place of the Virtual MS. The IAPI MS class has the same interface as the Virtual MS to ensure code compatibility, however the IAPI MS relies upon the Instrument Application Programming Interface (IAPI) (Thermo Fisher Scientific, 2022) to communicate with an actual Thermo Orbitrap Fusion instrument. All developed controller implementations, whether DDA or DIA that were initially tested against the Simulated Environment in ViMMS, can run without any change in the Real Environment (for the new DIA methods, this is the dashed purple box in Figure 1B). In this manner, real experimental data can be seamlessly acquired using the AIF and SWATH controllers initially developed in the simulator.

2.3 Simulating DDA and DIA methods

2.3.1 Generation of simulated data

Simulations allow us to flexibly define different scenarios to validate hypotheses without costly instrument time. To compare the two types of acquisition methods, we use the Simulated Environment in ViMMS (Figure 1A) to generate simulated data with an increasing number of co-eluting chemicals present to test the limit of deconvolution empirically. While these chemicals are purely synthetic, they allow us to know the ground truth and

evaluate the results of the different methods more accurately without being reliant on inconsistent database matching.

In the simulated study, each chemical is generated by first choosing a formula randomly from the HMDB database, ensuring that its observed monoisotopic mass is between 100–1,000 Da (the mass distribution of e.g., the 5,000 sampled chemicals is shown in Supplementary Section S5). Next the chemical is assigned a uniformly-sampled retention time value between 0–400 s. The choice of uniform distribution here is motivated by the narrow retention time range used. When the number of chemicals are high (e.g., 5,000), this results in dense regions of co-eluting ions throughout the entire simulated injection, which could challenge spectral deconvolution and potentially demonstrate the limit of such approaches.

Finally for each chemical, a chromatogram is generated. The maximum intensity value of the apex of the chromatogram was sampled from a uniform distribution between 1E4–1E7. A Gaussian chromatographic peak shape with a mean centered at the chemical's RT value, and a standard deviation of 5 s is assumed. This chromatographic peak shape assumption matches that of MS-DIAL, thus making the task of peak picking easier. Each chemical's observed MS2 spectra are also generated randomly by sampling fragment peaks' m/z uniformly between 70 m/z to the exact mass of the formula (assuming a positive charge of +1). The intensity of a fragment peak is also set to be between the specified minimum (0.1) and maximum (0.8) proportions of the chemical's apex intensity. The number of fragment peaks in an MS2 scan is generated by sampling from a Poisson distribution with the mean 10. This Poisson mean was chosen to produce sufficient expected numbers of fragment peaks per MS2 scan for spectral matching later.

Simulated samples containing the specified number of chemicals are generated in a case-vs-control setup, where each sample set

consists of 5 case and five control samples, and each sample contains observation from the specified number of chemicals. To reduce the influence of peak picking and alignment during data processing of DIA data, chemicals were generated such that the same chemical has identical m/z and RT values across samples, although their intensity values change across the case and control groups. While simple, this setup represents a realistic case that captures the essence of many real biological mass spectrometry-based experiments. Varying numbers of chemicals are generated during simulation, ranging from 10, 20, 50, 100, 200, 500, 1,000, 2000, and 5,000. The entire experiment is repeated 5 times, resulting in multiple experimental replicates. DDA (Top-N) and DIA (AIF, SWATH) controllers were run for every sample that was simulated, resulting in an mzML file for each acquisition run. Parameters of these controllers were set to be the same as that used for real data acquisition (detailed in Section 2.1).

2.3.2 Processing of simulated data

Once simulated data has been generated in Section 2.3.1, chemicals need to be mapped to their respective fragmentation spectra. For DDA, the association between observed fragmentation scans to chemicals is known unambiguously as their mapping can be read from the simulation state directly. For DIA, spectral deconvolution needs to be performed to assign the deconvoluted fragment peaks to chemicals. We chose MS-DIAL (Tsugawa et al., 2015) for peak picking and spectral deconvolution on the resulting mzML files produced by simulating DIA methods (see Supplementary Section S2 for the MS-DIAL parameter settings). MS-DIAL was chosen as it was widely used by the community for processing DIA data.

From MS-DIAL output, we can extract for each sample set a list of features that were detected and aligned across the DIA mzML files (5 cases and five controls). Each feature is potentially associated with a fragmentation spectra, which has been deconvoluted by MS-DIAL. To assign fragmentation spectra, we matched simulated chemicals to features using the m/z tolerance of 5 ppm and RT tolerance of 10 s, therefore linking chemicals to their deconvoluted spectra. If there are multiple possible candidates during matching, the feature closest in m/z value to the chemical's monoisotopic peak will be chosen. Finally for each simulated chemical, its true fragmentation spectra are known. This can be used to construct a library of true reference spectra for matching. Observed (and deconvoluted, in the case of DIA) spectra are matched to the true spectral library using cosine similarity. For matching, a bin width of 0.05 Da is used and a minimum of at least three matching peaks is required.

2.4 Validation on a real instrument

2.4.1 Generation of real data

The Real Environment in ViMMS was used to validate that the simulated results translate to real experiments (Figure 1B). This environment can be connected to a Thermo Orbitrap Fusion tribrid-series mass spectrometer, allowing us to perform data acquisition on a series of beer samples using the fullscan, Top-N, and the newly introduced SWATH and AIF controllers. The result from data acquisition is a series of mzML files, one for each beer sample

and controller used. For more details on real data acquisition, refer to Section 2.1.

2.4.2 Processing of real data

Peak picking and alignment was performed using MS-DIAL on the real experimentally-derived fullscan mzML files. Unlike in simulations, here the exact chemical composition of the sample is unknown. As such, fullscan features from peak picking on the fullscan data are used as a proxy for chemicals. The fullscan data, which contains the most MS1 information and therefore the best chromatographic peak shapes, was chosen for feature extraction using MS-DIAL. MS-DIAL parameters were chosen by hand to give a reasonable number of peaks comparable to what we have seen from past experiments using this kind of sample on the same instrument (details in Supplementary Section S3). The result from this is a list of fullscan features detected and aligned from the fullscan beer mzML files.

Next DDA fragmentation spectra were to be assigned to fullscan features, with the following procedure employed to bypass the need to do further peak picking on the DDA mzML files (which can be problematic due to the lower number of MS1 scans in fragmentation files). *pymzML* (Bald et al., 2012) was used to load MS2 scans from each DDA mzML file. For each MS2 scan, we used its precursor m/z, isolation window and RT values to assign the scan (and therefore fragmentation spectra) to its corresponding fullscan feature. This assignment was done based on whether the isolation window and RT values of a chemical overlap with the feature's bounding box from peak picking. If a fullscan feature had multiple MS2 scans associated with it, the scan that was fragmented at the highest intensity within the feature's bounding box was chosen.

Finally, just like in simulation, MS-DIAL was used to perform peak picking and spectral deconvolution on the DIA mzML files. Peak picking was performed using the same parameters as the fullscan data, and an additional deconvolution step was done using MS-DIAL (detailed parameters in Supplementary Section S3). From MS-DIAL output we extracted a list of fragmentation features detected in the DIA mzML files and their corresponding MS-DIAL deconvoluted spectra. Fullscan features (without fragmentation information) were matched to the DIA fragmentation features (with deconvolved spectral information) using an m/z tolerance of 5 ppm and RT tolerance of 10 s. If there were several possible candidates during matching, the one closest in m/z value was chosen. The results of this procedure from both the DDA and DIA data is the assignment of a fullscan feature to fragmentation spectra. Such spectra can be used for matching fullscan features to the reference libraries during benchmarking experiments.

2.5 Matching experimental spectra to reference libraries

To assess mass spectral quality, we match the fragmentation spectra from DDA and DIA methods to two reference spectra libraries. These reference libraries can be used to match against the observed and deconvoluted spectra from DDA and DIA methods. For matching, a bin width of 0.05 Da is used and a minimum of at least three matching peaks is required.
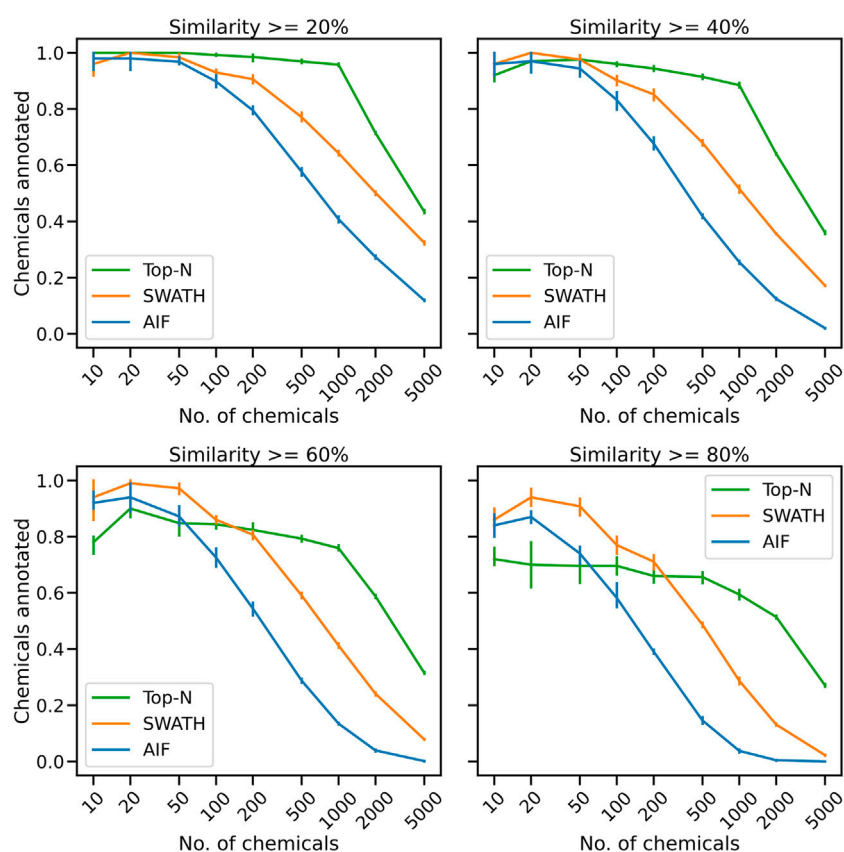
FIGURE 2
The mean proportion of unique chemical annotations at varying numbers of chemicals and similarity thresholds across five replicates. The error bar shows the 95% confidence interval.

The first reference library is the "GNPS Matches to NIST14" dataset obtained from the GNPS library (Wang et al., 2016). This dataset contains 5,763 high confidence matches to NIST14 M/MS library spectra. Filtering by polarity is performed to select spectra in positive mode only that can be used as reference to assess spectral annotation quality. For experiments, we call this the *GNPS/NIST14* library. Its purpose is to assess how many potentially unknown metabolites could be annotated in an untargeted metabolomics experiment using the different acquisition methods.

Additionally, we also introduced our own multiple-injection reference library for spectral matching, which we call the *Multi-Injection* Library. Each beer sample was injected ten times and data acquisition is performed by taking advantage of replicate information using the Intensity Non-overlap method available in ViMMS (detailed in Supplementary Section S4). Intensity Non-overlap is an iterative DDA-based method that detects regions-of-interests (ROIs) in real-time and avoids re-fragmenting the same ROI multiple times across successive injections. An ROI is scored for fragmentation based on its overlapping area weighted by intensity, with higher scoring ROIs selected more often (for more details, refer to Supplementary Section S4 and McBride et al. (2023)). In the presence of multiple injections, Intensity Non-overlap has been shown to outperform Top-N by a large

margin as it is able to target more unique features across injections, while fragmenting each feature closer to its apex.

To convert the acquired Multi-Injection mzML files into spectral library, for each beer sample, we perform peak picking using MS-DIAL on its corresponding fullscan mzML file, generating a list of fullscan feature for that beer sample. Fragmentation spectra, acquired *via* exhaustive fragmentation of each sample using Intensity Non-overlap, were extracted from mzML files and matched to the detected fullscan features following the DDA procedure outlined in Section 2.4.2. The purpose of introducing this Multi-Injection reference library is to assess the coverage of the benchmarked DDA (Top-N) and DIA (AIF, SWATH) methods when only a single replicate is available (a common occurrence) and increasing the number of replicates is not possible due to cost or other constraints.

3 Results

3.1 Simulated results

Acquired fragmentation spectra could help deduce the chemical identities of measured compounds. From the proposed simulated experiment in Section 2.3, for each acquisition method the number
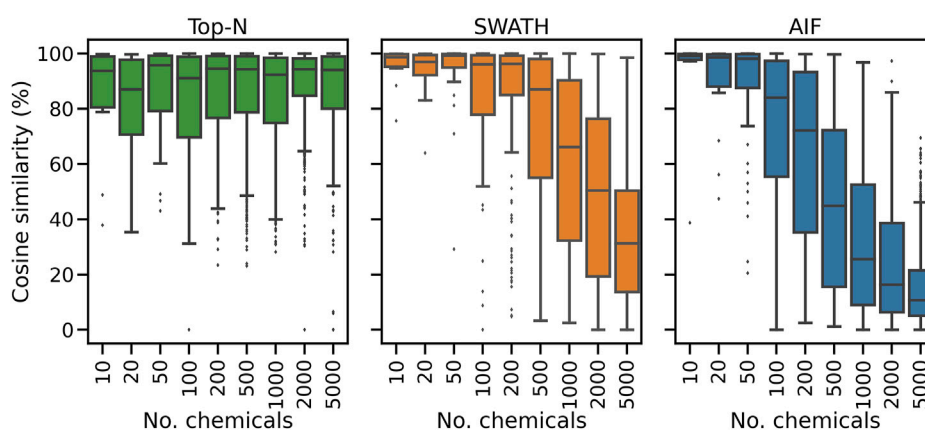
FIGURE 3
The distribution of cosine similarity scores when matching observed spectra to the true reference spectra for varying numbers of chemicals.

of chemicals that could be annotated based on spectral matching is obtained. The number of unique chemical annotations in the dataset is counted at varying thresholds for cosine similarity score of at least 20%, 40%, 60% and 80% (Figure 2). The results across five replicates show that for small numbers of chemicals, all benchmarked methods, whether DDA or DIA, have similar annotation performance. At similarity threshold 60% and with only 100 chemicals, AIF annotated a mean of 72.6% of chemicals, SWATH 86.0% and Top-N 84.4%. As the number of chemicals increases, the gap between the benchmarked methods widen leading to lower annotation rates from DIA. When the number of chemicals increased to 200, at 60% similarity threshold, AIF managed to annotate 54.4% chemicals, SWATH 80.7% and Top-N 82.4%. For 500 chemicals, AIF annotates 28.8% of chemicals, SWATH 59.0% and Top-N 79.3%.

With a greater number of chemicals, Top-N outperforms the two DIA methods across all thresholds. At the highest number of chemicals (5,000), the annotations obtained from SWATH and AIF are nearly 0% while Top-N managed a mean of 31.5% across replicates. It can be observed that increasing the cosine similarity threshold from 20% to 80% lowers the results of all methods, but the overall trend remains. The supplementary section also includes additional results obtained from varying the matching threshold. Supplementary Figure S2 displays the results of matching with a bin width of 0.005 Da, while Supplementary Figure S3 illustrates the results of matching with a bin width of 0.50 Da. The results in Supplementary Figure S2 demonstrate that the tolerance level was too narrow, leading to a decrease in performance for all acquisition methods. Despite this, Top-N still demonstrates superior performance compared to the DIA methods when a high number of chemicals are present. The results in Supplementary Figure S3, obtained using a larger bin width of 0.50 Da, are consistent with those presented in Figure 2 which utilised a bin width of 0.05 Da. This consistency in results highlights that the trend of Top-N outperforming DIA methods holds true for a large number of chemicals (> 200). Supplementary Figure S4 also shows the results from using different window sizes of 100 m/z, 50 m/z and 25 m/z in the SWATH simulation. It can be observed from Supplementary Figure S4 that making the window size smaller produces a slight

annotation improvement in SWATH when the number of chemicals are large (2000 and 5,000). However this improvement is small, and the overall trend of Top-N outperforming both SWATH and AIF in this regime holds.

To further explain the annotation results in Figure 2, here we consider for one replicate the distribution of spectral similarity scores when matching DDA and DIA spectra to the known ground truth of true chemical fragmentation spectra. The results are shown in Figure 3 for the entire range of chemicals tested. From Figure 3, it can be seen that Top-N generally performs best in returning high similarity scores, followed by SWATH then AIF. For Top-N, the cosine similarities of matches generally remain high even with an increasing number of chemicals. For the two DIA methods, their cosine similarities gradually decrease with more co-eluting chemicals. This could be explained by the fact that as the elution profile gets more crowded, more precursor ions are isolated and fragmented in the same window, making deconvolution harder. SWATH performs marginally better than AIF with increasing chemicals. This makes sense as the window used for isolating multiple peaks in SWATH is smaller than AIF where all ions in the entire scan range are used, making the deconvolution problem slightly easier for SWATH.

Inspecting the hardest case of 5,000 chemicals in Figure 4, it can be observed that AIF returns most of its matches at low cosine similarity achieving a median score of 10.7%, Top-N returns most matches at high cosine similarity with a median of 94.1%. SWATH outperforms AIF (median similarity 31.3%) but not as well as Top-N. To explain the decrease in similarity scores, we inspect the pairwise cosine similarity of all observed/deconvolved spectra for each acquisition method (Figure 5). The results show that the pairwise similarity of the ground truth and Top-N spectra is nearly 0 for nearly the entire range of chemicals. However, most likely due to the increased difficulty in deconvolution, pairwise spectral similarities in the SWATH and AIF results are higher with increasing chemicals–resulting in a decrease in identification hits, and fewer matches at high cosine similarity as the experimental mass spectra become more similar to each other and less similar to the true mass spectra of the chemicals.
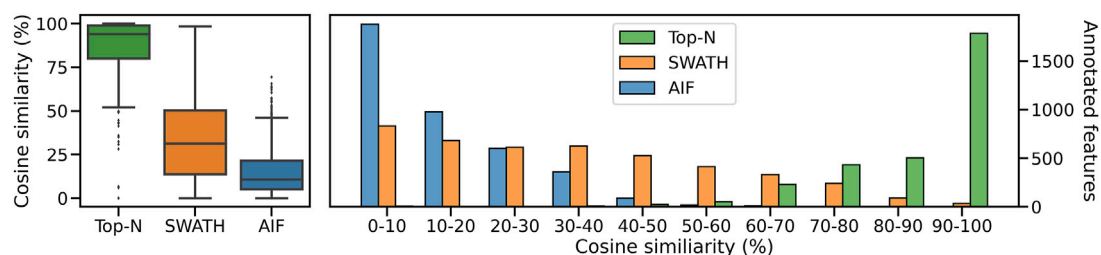
FIGURE 4
The distribution of cosine similarity scores as a boxplot (left) and a histogram (right) when matching observed spectra to the true reference spectra for 5,000 chemicals.
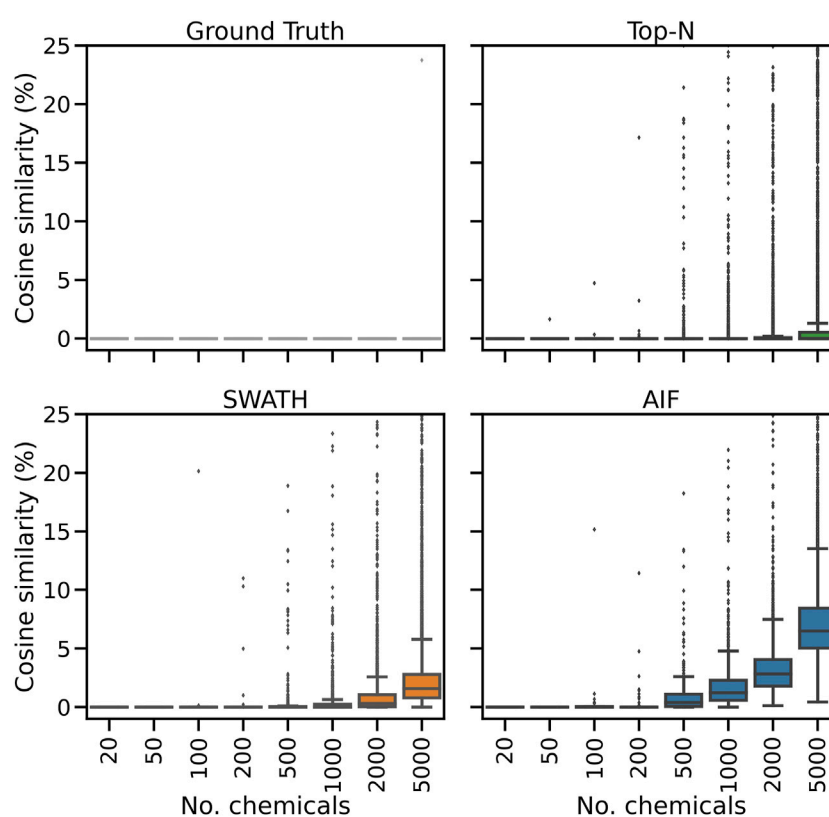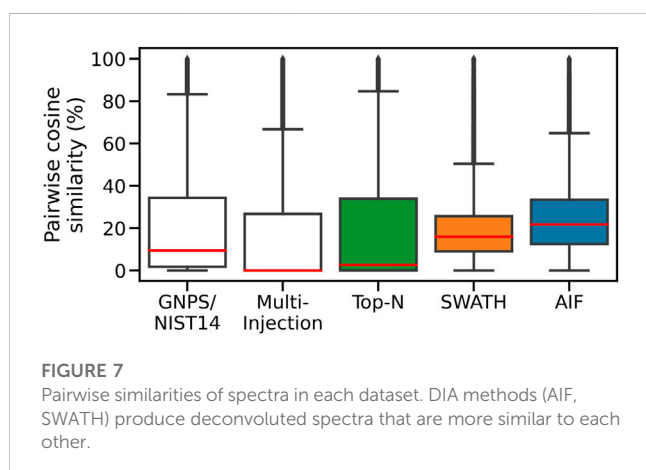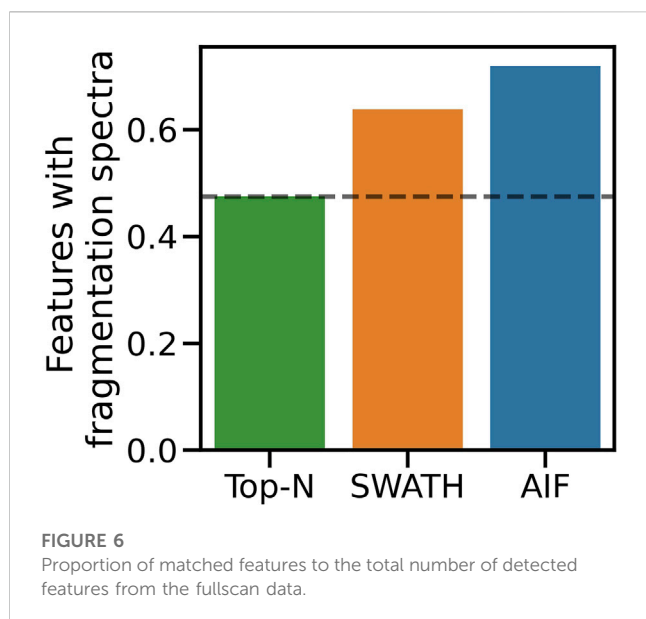

FIGURE 5
The distribution of pairwise cosine similarity scores of the ground truth (true chemical spectra), and fragmentation spectra from Top-N, SWATH and AIF. The plot y-axes are truncated at 25% similarity.

3.2 Real experimental results

3.2.1 Mapping features to fragmentation spectra

Simulated results were further validated on real instruments by running the benchmarked DDA and DIA methods on actual beer samples. Following the procedure in Section 2.4.2 to map fullscan features to fragmentation spectra, 6090 fullscan features initially were detected from the fullscan data after peak picking and alignment. After matching the fullscan features to fragmentation data, DIA methods produce the largest number of matched features, with SWATH at 3889 and AIF at 4381. In contrast, Top-N only has

2,895 matched features. Figure 6 summarises the proportion of fullscan features that can be matched to the fragmentation spectra. Our results here agree with Guo and Huan (2020a) in how DIA (AIF) is able to fragment more features than DDA (Top-N). We note some slight differences in our methodology to (Guo and Huan, 2020a). In this work, we perform peak picking to extract features on the fullscan data, which has more reliable MS1 signals. Fragmentation mzML files are used only to map fragmentation scans to the detected fullscan features. Whereas in (Guo and Huan, 2020a) the peak picking is performed directly on the fragmentation mzML files, potentially leading to poorer extracted chromatographic

FIGURE 6
Proportion of matched features to the total number of detected features from the fullscan data.

FIGURE 7
Pairwise similarities of spectra in each dataset. DIA methods (AIF, SWATH) produce deconvoluted spectra that are more similar to each other.

peak shapes due to sparser MS1 data points in the mzML files. Despite these differences in the experimental set up, the conclusions from both our study and (Guo and Huan, 2020a) agree that AIF outperforms Top-N in the number of fragmented features. Additionally SWATH was not included in (Guo and Huan, 2020a) but it was hypothesised to perform in the middle of Top-N and AIF, as the windowing approach used in SWATH lies between the two extremes of fragmenting everything (AIF) and fragmenting only a few selected precursor ions (Top-N). Our results here confirmed that SWATH achieves a higher coverage than Top-N but lower than AIF.

3.2.2 Pairwise spectral similarity

By following the methodology of Section 2.5, two reference mass spectral libraries were constructed, one based on the GNPS spectra matched to NIST-14 at high reliability (the GNPS/NIST14 library), and another based on an exhaustive multiple-injection approach we constructed ourselves (the Multi-Injection library). The purpose of the GNPS/NIST14 library is to assess how many unknown

molecules can be identified from spectral matching for each acquisition method, whereas the purpose of the Multi-Injection library is to determine the extent of coverage with respect to an exhaustive and expensive multiple-injection method. After filtering by polarity, we obtain 5,274 positive-mode spectra from the GNPS/NIST14 library, whereas for the Multi-Injection library, in total 4987 features were available for matching in this library.
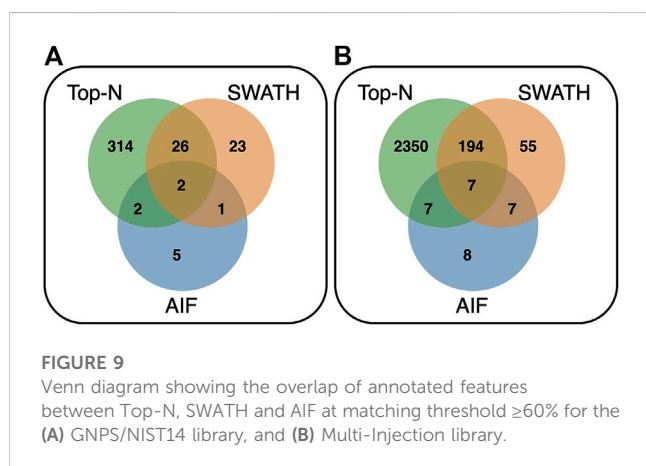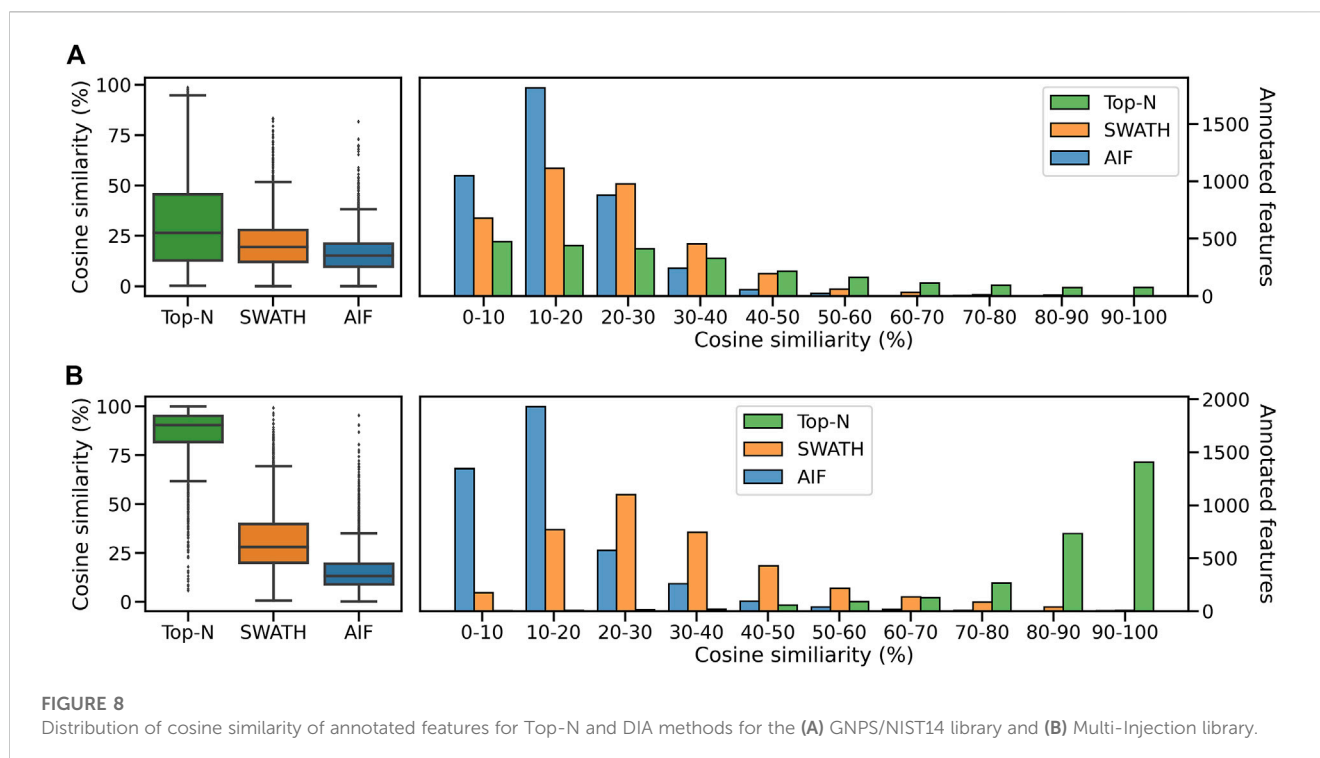
To assess spectral quality, we first compute the pairwise similarity of spectra in the same dataset, and also within each of the two reference libraries. For each feature, we find the matches to fragmentation spectra in the same dataset by computing the cosine similarity (ms2_tol = 0.05 Da, min match peaks = 3), with the results shown in Figure 7. The Multi-Injection reference library has a median of nearly 0% for pairwise similarities, demonstrating that the acquired reference spectra using the Intensity Non-overlap method are sufficiently different from each other. The GNPS reference library has a higher pairwise similarity of 9.5%, suggesting that some of the curated molecules share similar structures and therefore similar mass fragmentation spectra. This is a reasonable assumption as shared structures is the core underlying assumption for models to discover chemical substructures, such as MS2LDA (van Der Hooft et al., 2016) that has been applied to reference libraries like MassBank and GNPS.

The Top-N dataset also has the same low median pairwise similarity (2.7%) in its member spectra. This is expected given the nature of DDA. Assuming that deconvolution works well for DIA data and the resulting spectra are dissimilar to each other in a manner similar to the Top-N results, we expect to observe lower pairwise similarities from the DIA datasets too. However the higher median pairwise similarity scores for SWATH (16%) and AIF (22%) suggest that deconvolved spectra from SWATH and AIF tend to be more similar to each other. This could be due to the difficulty in deconvolving spectra. The results here generally agree with simulated results in Section 3.1, where DIA methods are also shown to exhibit higher pairwise similarity in simulation.

3.2.3 Spectral matching results

We compute the cosine similarity of features to the GNPS/NIST14 reference spectra. Similar to before, features were matched to reference spectra based on their cosine similarity (ms2_tol = 0.05 Da, min match peaks = 3). If there are multiple matches for each feature, the one with the best score (highest) is kept. Figure 8A shows the score distributions for the three methods. From Figure 8A (left panel), it can be observed that Top-N obtains the highest median cosine similarity (26.5%) followed by SWATH (19.5%) and finally AIF (15.2%). Consequently this results in Top-N (DDA) obtaining the most high-scoring matches at ≥ 60% similarity, followed by SWATH and AIF last (Figure 8A). The results here are consistent with the simulated results in Section 3.1 and also with Guo and Huan (2020a) but in that work, only Top-N and AIF were compared. Our results further confirmed the hypothesis in Guo and Huan (2020a) that SWATH should perform in the middle of Top-N and AIF when it comes to spectral quality. This is because deconvolution using SWATH data is easier than AIF due to the smaller window sizes.

Next we perform spectral matching to the Multi-Injection reference library. When matched against Top-N and DIA methods, Top-N performs best (90.5% median) followed by

FIGURE 8
Distribution of cosine similarity of annotated features for Top-N and DIA methods for the **(A)** GNPS/NIST14 library and **(B)** Multi-Injection library.

FIGURE 9
Venn diagram showing the overlap of annotated features between Top-N, SWATH and AIF at matching threshold ≥60% for the **(A)** GNPS/NIST14 library, and **(B)** Multi-Injection library.

SWATH (28.1%) and AIF (13.3%) (Figure 8B). The results here are consistent with the GNPS results above. However we see that the median similarity of Top-N is much higher compared to the two DIA methods here. That is because the reference spectra contains more query spectra since it is based on the same sample, only fragmented exhaustively, and unlike GNPS/NIST14 there are fewer missing matches.

Comparing the overlap of annotated features at matching threshold ≥60% (Figure 9) for both reference libraries, it can be observed that Top-N (DDA) obtains the most hits, being able to annotate the most unique features (295 for GNPS/NIST14, and 2,231 for Multi-Injection). In both cases, there is a large overlap between Top-N and DIA annotations, and most DIA annotations are also recovered by Top-N. Between AIF and SWATH, there is also a lot of overlap with SWATH being able to recover most of the

annotations of AIF. The results here for AIF and Top-N generally agree with Guo and Huan (2020a), with SWATH a new addition in our results that perform in between those two.

4 Discussion and conclusion

In this study, we performed a comprehensive comparison of DDA vs. DIA methods, first in the simulator, followed by validation using real experimental mass spectrometry data. We simulated experiments at various complexity levels to challenge DDA acquision methods and DIA deconvolution methods. Here, we observed that the quality of deconvolution using MS-DIAL is limited by the number of co-eluting chemicals. Identification performance is limited in DIA compared to DDA with a large number (more than 1,000) of simulated chemicals or observed molecular features. We further benchmarked this scenario using a real untargeted metabolomics dataset acquired on the mass spectrometer generated from beer samples. Mass spectral matching of experimental mass fragmentation spectra from this real dataset on two sets of reference mass spectral libraries confirmed our simulated results. This validates our motivation that a simulator framework such as ViMMS can be used to benchmark the performance of both types of methods. Simulation helps to provide an environment that can be used to prototype and validate advanced deconvolution methods–without the need of costly instruments (i.e., a "digital twin"). This encourages the development of better deconvolution methods, since known ground truths were generated *in silico*, thus making benchmarking easier. The ViMMS framework could also be used to simulate various scenarios that we have not covered in this study, for instance simulating different column properties and elution profile of chemicals, and assessing how that affects deconvolution.

In the context of LC-MS/MS analysis, a complex sample would contain a large number of different metabolites, while a simple sample would contain only a few metabolites. Determining the complexity of a sample is important for choosing the appropriate acquisition strategy for LC-MS/MS analysis. ViMMS can generate simulated samples of varying complexity, which can be used to assess whether DDA or DIA methods should be used in specific scenarios. By varying average number of co-eluting ions, we found that DIA being more effective at lower numbers and DDA having an advantage at higher numbers where DIA struggles to handle the large amount of overlapping ion chromatograms. From both simulated and real results, DDA was also found to generally perform better than DIA when it comes to matching unidentified features to spectra in both reference libraries (GNPS and Multi-Injection). DIA fragments more features than DDA but their quality for spectral matching is typically lower. Our results on this are not unique as a similar work in (Guo and Huan, 2020a) confirms our findings. Crucially, our study improves upon that prior work in several key aspects. The first is that SWATH was not included in the comparison of Guo and Huan (2020a), whereas our study did include both AIF and SWATH. Secondly, while spectral matching to a library of known fragmentation spectra (e.g., GNPS, or MassBank) can be done, many compounds present in such databases have no matches thus reducing the identification rate observed from both DDA and DIA methods. Thus we introduce another dataset, constructed using an advanced multiple-injection method, to measure how well DDA and DIA methods perform with respect to exhaustive fragmentation procedures. Exhaustive fragmentation procedures have been shown to perform best with respect to coverage when a large number of replicates are available. However, real experiments are often constrained by cost or time, limiting the number of replicates that could be produced. In this setting, our analysis shows that DDA still performs best compared to DIA in recovering coverage when multiple replicates are not easily available.

Our recommendation on which acquisition method to choose is therefore.

- If the sample complexity is expected to be **low** to **medium**, OR it is preferred to **fragment as many features** as possible, even if they are not all identified, it is recommended to use DIA for data acquisition (keeping in mind the necessary deconvolution step).
- If the sample complexity is expected to be **high**, OR it is preferred to **obtain as many identified features** as possible, it is recommended to use DDA, which targets a specific ion for each MS2 scan thus generating fragmentation spectra that are almost immediately usable for analysis.

It is worth emphasizing that our results do not negate the usefulness of running DIA methods. Using DIA, we obtain the largest number of coverage of features, resulting in the most number of fragmented molecules–many of whom could be used for further investigation in the future. However acquired DIA scans need to be deconvoluted in order to translate the results to actual identification. Improvements in deconvolution methods are therefore needed to fully maximise the usefulness of DIA data. The work introduced here shows that a large margin of improvement is still possible in

spectral deconvolution of DIA results–an avenue for further research that could be explored by the community. The work here also shows that through a simulation framework such as ViMMS, researchers could first test their deconvolution method *in silico*. The portable nature of ViMMS means controllers (fragmentation methods) developed in simulators can be easily ported to run on the real instrument. DIA methods such as AIF and SWATH were implemented as controllers on top of ViMMS, tested in the simulator, and easily deployed to run on the actual instrument. This makes it easy for others to reproduce our results in simulation. It also opens the path for more advanced DIA methods to be developed in the future on top of ViMMS. What makes ViMMS particularly valuable is its ability to simulate a range of different method parameters for both DDA and DIA, allowing researchers to easily evaluate the performance of different LC-MS/MS data acquisition methods. This feature not only provides valuable insights into the strengths and limitations of each acquisition method, but it also opens the door for further advancements in LC-MS/MS data acquisition methods. By using ViMMS, researchers can optimize their choice of acquisition method and obtain more accurate and robust results, while saving significant time and resources as they eliminate the need for numerous experiments. In other fields such as molecular machine learning, having a standardized benchmark dataset (Wu et al., 2018) fosters development and collaborations as now a common reference exists onto which we can compare the performance of different methods. It is our hope that the simulated experiment introduced here could also serve as a standard and reproducible benchmark to which other deconvoluted methods can be compared to.

Our study has several limitations, including the fact that our comparison of DDA and DIA acquisition methods is based solely on the analysis of processed and, in the case of DIA, deconvoluted data. While data processing, particularly the deconvolution step, is a crucial aspect of DIA data analysis, other important factors such as the increased computational demands for deconvolution in DIA are not considered in detail. Additionally, the impact of instrument configuration and sample preparation on the data is not evaluated in our study, which can also influence the performance of both acquisition methods. Further research is necessary to gain a comprehensive understanding of the strengths and limitations of both DDA and DIA in untargeted metabolomics and to determine the optimal approach for various scenarios. Our findings primarily emphasise the MS/MS spectral annotation performance and do not extensively address other factors such as quantitative performance. We only used MS-DIAL for the pre-processing of DIA data. MS-DIAL is de-facto the most popular tool used for spectra deconvolution and analysis of DIA data. However, other deconvolution tools exist, such as those of Yin et al. (2019) and Graca et al. (2022), that we do not include in our comparison. Parameters from MS-DIAL in our study were picked by hand to obtain reasonable results–similar to the limitation in Guo and Huan (2020a). Furthermore, DDA and DIA methods exist that can exploit information across multiple samples (Koelmel et al., 2017; Tada et al., 2020) but these were not included in our evaluation. Using ViMMS, we could generate samples in multiple replicates and assess the performance of DDA and DIA methods when replicates are available. This validation to some extent has been done to compare

standard DDA vs. multiple samples/injections DDA in McBride et al. (2023) but not in the context of comparing to DIA methods.

Another aspect to take into account is the choice of cosine similarity as the metric for spectral matching. While cosine similarity is a commonly used metric in MS-DIAL, there are more advanced techniques available, such as Spec2Vec (Huber et al., 2021) and Tanimoto similarity, which have been shown to better preserve chemical similarity in spectral matching. However, for the purposes of this study, cosine similarity was selected as the metric because it can be applied to both real and synthetic data. The goal of this study was to demonstrate that simulated results can be translated to experimental results, and cosine similarity was chosen as the metric for spectral matching because of its broad applicability to both types of data.

For future work, the developed *in silico* benchmarking pipeline introduced in this work can also serve as the foundation to develop and validate a hybrid method that combines the benefit of both approaches. DDA and DIA methods exist in a spectrum: DDA isolates a certain precursor m/z for fragmentation, and DIA isolates multiple ions in a range of windows. This behaviour is the same across the entire run. It would be interesting to explore the potential of a hybrid method that combines the benefits of both DDA and DIA approaches. Such a method could potentially take advantage of the strengths of both approaches, allowing for more accurate and comprehensive analysis of complex samples. Such work has already been attempted (Guo et al., 2021) but that integration happens in data processing, not during the data acquisition itself. A hybrid acquisition method could use DDA to isolate certain prioritised precursor ions for fragmentation, and use DIA to isolate the remaining ions within a range of windows. This hybrid approach would allow for a more flexible and comprehensive analysis of samples, and could potentially improve the accuracy and reliability of results. This would lead to a more effective use of untargeted metabolomics, as more molecular features will be fragmented with higher-quality mass spectra associated to them. Ultimately, that will improve the biochemical interpretation of metabolomics profiles across the life sciences and other research areas.

Data availability statement

The dataset consisting of the DDA (Top-N, Intensity Non-overlap) and DIA (AIF, SWATH) mzML files generated and analysed for this study can be found under the following DOI: 10.5525/gla.researchdata.1382.

Author contributions

Funding

Conflict of interest

JJJVdH is member of the Scientific Advisory Board of NAICONS Srl. Milano, Italy. All other authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmolb.2023.1130781/full#supplementary-material

References

Bald, T., Barth, J., Niehues, A., Specht, M., Hippler, M., and Fufezan, C. (2012). pymzML—Python module for high-throughput bioinformatics on mass spectrometry data. *Bioinformatics* 28, 1052–1053. doi:10.1093/bioinformatics/bts066

Bern, M., Finney, G., Hoopmann, M. R., Merrihew, G., Toth, M. J., and MacCoss, M. J. (2010). Deconvolution of mixture spectra from ion-trap data-independent-acquisition tandem mass spectrometry. *Anal. Chem.* 82, 833–841. doi:10.1021/ac901801b

Davies, V., Wandy, J., Weidt, S., Van Der Hooft, J. J., Miller, A., Daly, R., et al. (2021). Rapid development of improved data-dependent acquisition strategies. *Anal. Chem.* 93, 5676–5683. doi:10.1021/acs.analchem.0c03895

Fernández-Costa, C., Martínez-Bartolomé, S., McClatchy, D. B., Saviola, A. J., Yu, N.-K., and Yates, J. R., III (2020). Impact of the identification strategy on the reproducibility of the DDA and DIA results. *J. Proteome Res.* 19, 3153–3161. doi:10.1021/acs.jproteome.0c00153

Gillet, L. C., Navarro, P., Tate, S., Röst, H., Selevsek, N., Reiter, L., et al. (2012). 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* 11, O111.016717. doi:10.1074/mcp.O111.016717

Graca, G., Cai, Y., Lau, C.-H. E., Vorkas, P. A., Lewis, M. R., Want, E. J., et al. (2022). Automated annotation of untargeted All-Ion Fragmentation LC–MS metabolomics

data with MetaboAnnotatoR. *Anal. Chem.* 94, 3446–3455. doi:10.1021/acs.analchem.1c03032

Guan, S., Taylor, P. P., Han, Z., Moran, M. F., and Ma, B. (2020). Data dependent-independent acquisition (DDIA) proteomics. *J. Proteome Res.* 19, 3230–3237. doi:10.1021/acs.jproteome.0c00186

Guo, J., and Huan, T. (2020a). Comparison of full-scan, data-dependent, and data-independent acquisition modes in liquid chromatography–mass spectrometry based untargeted metabolomics. *Anal. Chem.* 92, 8072–8080. doi:10.1021/acs.analchem.9b05135

Guo, J., and Huan, T. (2020b). Evaluation of significant features discovered from different data acquisition modes in mass spectrometry-based untargeted metabolomics. *Anal. Chim. Acta* 1137, 37–46. doi:10.1016/j.aca.2020.08.065

Guo, J., Shen, S., Xing, S., and Huan, T. (2021). DaDIA: Hybridizing Data-Dependent and Data-Independent acquisition modes for generating high-quality metabolomic data. *Anal. Chem.* 93, 2669–2677. doi:10.1021/acs.analchem.0c05022

Huber, F., Ridder, L., Verhoeven, S., Spaaks, J. H., Diblen, F., Rogers, S., et al. (2021). Spec2vec: Improved mass spectral similarity scoring through learning of structural relationships. *PLoS Comput. Biol.* 17, e1008724. doi:10.1371/journal.pcbi.1008724

Kaufmann, A., and Walker, S. (2016). Nested data independent MS/MS acquisition. *Anal. Bioanal. Chem.* 408, 5031–5040. doi:10.1007/s00216-016-9607-8

Koelmel, J. P., Kroeger, N. M., Gill, E. L., Ulmer, C. Z., Bowden, J. A., Patterson, R. E., et al. (2017). Expanding lipidome coverage using LC-MS/MS data-dependent acquisition with automated exclusion list generation. *J. Am. Soc. Mass Spectrom.* 28, 908–917. doi:10.1007/s13361-017-1608-0

Lawson, T. N., Weber, R. J., Jones, M. R., Chetwynd, A. J., Rodriguez-Blanco, G., Di Guida, R., et al. (2017). msPurity: automated evaluation of precursor ion purity for mass spectrometry-based fragmentation in metabolomics. *Anal. Chem.* 89, 2432–2439. doi:10.1021/acs.analchem.6b04358

McBride, R., Wandy, J., Weidt, S., Rogers, S., Davies, V., Daly, R., et al. (2023). TopNEXt: Automatic DDA exclusion framework for multi-sample mass spectrometry experiments. *bioRxiv.* doi:10.1101/2023.02.16.527961

Tada, I., Chaleckis, R., Tsugawa, H., Meister, I., Zhang, P., Lazarinis, N., et al. (2020). Correlation-based deconvolution (CorrDec) to generate high-quality MS2 spectra from Data-Independent Acquisition in multisample studies. *Anal. Chem.* 92, 11310–11317. doi:10.1021/acs.analchem.0c01980

Thermo Fisher Scientific (2022). *Thermo Fisher application programming interface.* [Dataset].

Tsugawa, H., Cajka, T., Kind, T., Ma, Y., Higgins, B., Ikeda, K., et al. (2015). MS-DIAL: Data-independent MS/MS deconvolution for comprehensive metabolome analysis. *Nat. methods* 12, 523–526. doi:10.1038/nmeth.3393

van Der Hooft, J. J. J., Wandy, J., Barrett, M. P., Burgess, K. E., and Rogers, S. (2016). Topic modeling for untargeted substructure exploration in metabolomics. *Proc. Natl. Acad. Sci.* 113, 13738–13743. doi:10.1073/pnas.1608041113

Wandy, J., Davies, V., Van Der Hooft, J. J., Weidt, S., Daly, R., and Rogers, S. (2019). *In silico* optimization of mass spectrometry fragmentation strategies in metabolomics. *Metabolites* 9, 219. doi:10.3390/metabo9100219

Wandy, J., Davies, V., McBride, R., Weidt, S., Rogers, S., and Daly, R. (2022). ViMMS 2.0: A framework to develop, test and optimise fragmentation strategies in LC-MS metabolomics. *J. Open Source Softw.* 7, 3990. doi:10.21105/joss.03990

Wang, M., Carver, J. J., Phelan, V. V., Sanchez, L. M., Garg, N., Peng, Y., et al. (2016). Sharing and community curation of mass spectrometry data with global natural products social molecular networking. *Nat. Biotechnol.* 34, 828–837. doi:10.1038/nbt.3597

Wu, Z., Ramsundar, B., Feinberg, E. N., Gomes, J., Geniesse, C., Pappu, A. S., et al. (2018). MoleculeNet: A benchmark for molecular machine learning. *Chem. Sci.* 9, 513–530. doi:10.1039/c7sc02664a

Yin, Y., Wang, R., Cai, Y., Wang, Z., and Zhu, Z.-J. (2019). DecoMetDIA: Deconvolution of multiplexed MS/MS spectra for metabolite identification in SWATH-MS-based untargeted metabolomics. *Anal. Chem.* 91, 11897–11904. doi:10.1021/acs.analchem.9b02655

Zhang, F., Ge, W., Ruan, G., Cai, X., and Guo, T. (2020). Data-independent acquisition mass spectrometry-based proteomics and software tools: A glimpse in 2020. *Proteomics* 20, 1900276. doi:10.1002/pmic.201900276