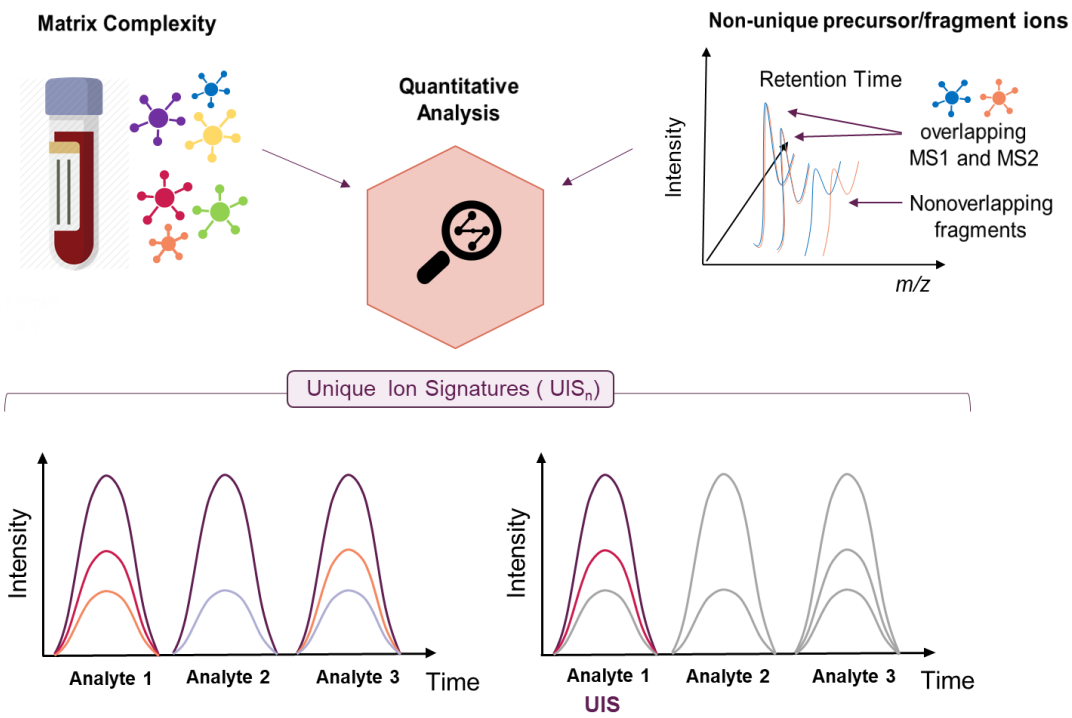
Analyzing Assay Redundancy in Metabolomics using Unique Ion Signatures

Premy Shanthamoorthy[1], Hannes Röst[1,2]

[1]Department of Molecular Genetics, University of Toronto; [2]Department of Computer Science, University of Toronto

Abstract

Targeted, untargeted and data-independent acquisition (DIA) metabolomics workflows are often hampered by ambiguous identification based on either MS1 information alone or relatively few MS2 fragment ions. While DIA methods have enjoyed popularity in proteomics, it is less clear whether they are suitable for metabolomics workflows due to their large precursor isolation windows and complex co-isolation patterns. Here, we quantitatively investigate the conditions necessary for unique metabolite identification in complex backgrounds using precursor and fragment ion mass-to-charge separation, comparing three benchmarked MS methods (MS1, MRM, DIA). We simulated MS1, MRM and DIA using the NIST LC-MS library as a complex background (8274 compounds at collision energy=35) and compared these methods with regards to unambiguous detection using unique ion signatures. Our simulations show that data generated with DIA with 25 Da mass windows outperformed MS1-only and MRM-based methods by a factor of 13.6-fold and 8.7-fold, respectively. Additionally, we use saturation analysis to show that for highly complex samples, a large portion of MS1-only detection (44.9%) is ambiguous while MRM (6.6%) and DIA (0.6%) present lower ambiguity. Our analysis demonstrates the power of using both high resolution precursor and high resolution fragment ion *m/z* for unambiguous compound detection. This work also establishes DIA as an emerging MS acquisition method with high selectivity for metabolomics, outperforming both DDA and MRM with regards to unique compound identification potential.



Introduction

Liquid chromatography coupled to tandem mass spectrometry (LC-MS/MS) allows for the robust analysis of complex samples in metabolomics. LC/MS-based metabolomics allows researchers to explore a large fraction of chemical diversity, uncovering fundamental metabolism, regulation, genetics, and interspecies analyte transfer in complex systems of global importance, such as nutrient cycling, wastewater treatment and the human microbiome.[1-2]

However, the amazing diversity of the metabolome and its lack of a generic polymer template, as in the genome or proteome, greatly complicates the confident identification of unique metabolites required to provide meaningful biological data. Untargeted workflows focused on the precursor ion (MS1-based) are comprehensive, but often unspecific with compound detection hampered by their reliance on a single precursor *m/z* value. An increase in confidence occurs with the addition of fragment ion (MS2) data (an MS2 match in literature).[3] However, MS2 data acquired using data-dependent workflows (DDA) heavily relies on stochastic MS1 data collection, with a high degree of variance based on sample complexity. On the other hand, although targeted methods (multiple/parallel reaction monitoring; MRM/PRM) using MS2 data are highly specific, they are strongly limited in both mass accuracy for traditional MRM and analyte throughput with a focus on few fragments, generally unsuitable for omics type of analysis.

DIA is a next-generation MS method with capabilities of capturing the complete precursor and fragment ion signal (MS1 and MS2) in a single run. It uses a set of pre-programmed isolation windows which span the whole mass range, thus ensuring that any precursor is fragmented. In each cycle, a set of high resolution fragment ion spectra is acquired for every isolation window and the process is repeated once the last isolation window has been acquired, generating a series of fragment ion spectra with the same precursor isolation window. This method allows high analyte throughput with reproducible and consistent quantification to facilitate metabolite discovery.[4-9]

MS1-only, MRM/PRM and DIA methods all use a combination of evidence based on precursor ion signals, fragment ion signals or a combination of the two for compound detection. Here, we will investigate the required conditions to uniquely detect a compound among a given set of expected compounds in a complex sample using precursor and fragment ion signals. While relying on multiple characteristic fragment ions is a generally accepted technique to increase certainty in targeted metabolomics assay generation, it is unclear how many transitions are “enough” and few studies have tried to quantify the effect of the acquisition method on compound detection on a large scale.

Sherman *et al.* first introduced the concept of using information content as criterion to select suitable transitions (the combined representation of a compound with both MS1 and MS2 data) for targeted methods in proteomics.[10-12] This work is based on the concept of unique ion signatures (UIS), referring to the combinations of ions that map uniquely to one analyte (a peptide in their work), for a given analyte background. In both cases, the selection of assays with minimal interference with other analytes averted the high likelihood of ambiguous detection due to multiple analytes sharing a particular combination of transitions. Here, we introduce the UIS concept for metabolomics, and use it to calculate non-redundant theoretical assays for thousands of compounds in a given metabolomic background. Specifically, we quantify the capabilities of metabolite detection using mass accuracy to compare current methods in metabolomics. These analyses assess

the specificity and the power to detect analytes with MS1-only, or with MS2, in addition to assessing the potential of novel combinatorial approaches (DIA) in the field of metabolomics. Methods, such as DDA, where data acquisition is dependent on precursor abundance and sampling biases, are not included in these analyses due to the potential variance in the acquired MS2 data.

Experimental Section

UIS_n

Using the NIST 17 LC/MS library (~12,000 compounds, 500,000 spectra), compounds and spectra were filtered to retain only structurally different compounds measured in positive-ion mode on a high resolution instrument type and a collision energy of 35+/-5 to obtain a "background metabolome" of 8274 compounds.[13] After filtering for common adduct types (H+, Na+), 7669 queries (individual metabolites in the given metabolome) were used for simulations by setting realistic mass tolerances associated with different MS methods. Fragment ion spectra were additionally filtered to only include valid transitions greater than 10% of the maximum relative intensity in the fragment spectrum.

Each query was independently compared against the full 8274 compounds in the background metabolome using the tested MS methods (MS1-only, MRM, DIA) using appropriate values for mass accuracy for both the precursor *m/z* window (Q1) and the fragment *m/z* windows (Q3) based on the resolution of commercially available instrumentation (QQQ or Orbitrap/QTOF). For QQQ instruments we chose an isolation width of 0.7 Da and for high resolution instruments we conservatively assumed a resolution achievable by all high resolution instruments of 40 000 and a corresponding extracted ion chromatogram width of 25 ppm; while high-end Orbitrap instruments can achieve higher resolution than 25 ppm, we show that the difference with respect to differentiating unique compounds using MS1 signal between 25 ppm and 1 ppm is less than 1% (Supplemental Fig. 2). Isolation window sizes were set accordingly, centered at the corresponding precursor and fragment *m/z* values of the query, with the number of transitions defined by a value of *n* (UIS_n). The following MS1/MS2 windows were set (in daltons - Da, or parts per million of 1 Da - ppm): MS1-only: 25 ppm/-; Multiple Reaction Monitoring (MRM): 0.7 Da/0.7 Da; Data-Independent Acquisition (DIA): 25 Da/25 ppm, 25 ppm/25 ppm (Supplementary Table 1). Simulations provided a measure of uniqueness for each metabolite based on the number of compounds found in the background that were not differentiable from the query at the given parameters. The number of hits in the background and the percentage of compounds in the NIST library (queries) with no interference (the background for each query based on the given parameters), was then determined for each method. Analyses and visualization discussed herein were performed using Python 3.7. Visualization of spectral comparisons (Fig. 2 A-B) were performed using R package OrgMassSpecR. All scripts are available through Github at: https://github.com/premyshan/DIAColliderMetabo.

Theoretical Saturation

Matrix Complexity. UIS simulations were conducted to determine the performance of different MS methods (measured by the percentage of unique compounds) in relation to matrix complexity, as indicated by the number of compounds included in the background. For methods using MS2, three transitions were used (UIS_3) to perform simulations using the NIST LC/MS library as background (7669 queries, 8274 compounds in background). The percentage of unique compounds was calculated and saturation effects were then modelled using statsmodels (Python module) with a logarithmic fit.

Number of Transitions. The number of unique ion signatures in the NIST LC/MS library was simulated for different methods that use MS2 information (MRM, DIA). Fragment spectra for each query were filtered to include the top *n* transitions (based on the defined UIS_n) greater than 10% of the maximum relative intensity. Simulations were then conducted using these individual queries and their corresponding background metabolites that interfere within the set MS1/MS2 isolation windows for MRM (0.7 Da/0.7 Da) or DIA (25 Da/25 ppm, 25 ppm/25 ppm). The percentage of unique compounds was calculated for a range of transitions (*n* = 1-8, 100% = 7669 queries).

Collision Energy. Compounds from the NIST LC/MS library were filtered by experimental conditions (positive-ion mode and removal of stereoisomers), window size (MS1 = 25 Da), and acquisition instrument (Q-TOF) to calculate the optimal collision energy (CE) for each compound (Supplementary Fig. 4). Individual compounds were compared against each of their interfering compounds using two values to create a similarity matrix - (i) the difference in CE and (ii) their cosine similarity score based on their spectra. A directional traceback procedure in the matrix started at comparisons with a minimum difference in CE and moved towards the minimum similarity score. The mode of these results for each compound against all its interfering compounds (if a singular mode is available) is recorded as an optimal CE (Supplementary Fig. 4).

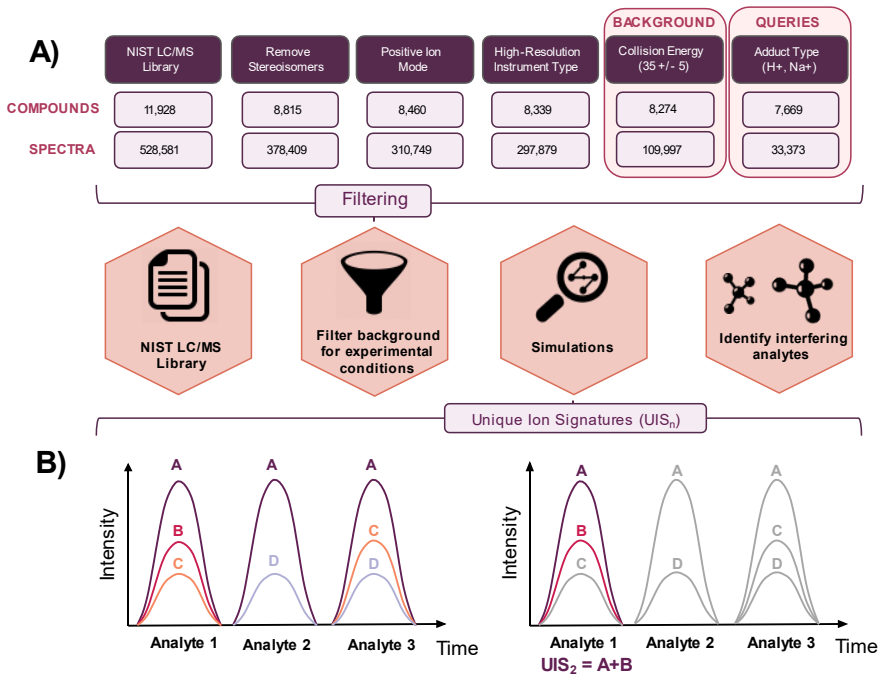


Figure 1. Unique Ion Signatures Analysis Pipeline. (A) Using the NIST LC/MS library (~12,000 compounds, 500,000 spectra), compounds and spectra were filtered for the removal of stereoisomers, and experimental settings of positive-ion mode, high-resolution instrument type and collision energy of 35+/-5 for a background of 8274 compounds. After isolating for common adduct types (H+, Na+), 7669 queries, individual metabolites in the given metabolome, were used for simulations by setting realistic mass tolerances associated with different MS methods. **(B)** The UIS concept was used, explained using three analytes that have some transitions in common. UIS_n is defined as a set of *n* transitions that map exclusively to one analyte in the metabolome to be analyzed. Assuming a metabolome consisting of three analytes that resolve on the chromatography, for analyte 1 there is one UIS_1 (B), one UIS_2 (A and B), and one UIS_3 (A-C). Note that the transition pair A and C is not a unique ion signature because this signal can be explained by either analyte 1 or analyte 3. *Figure adapted from:* Röst, H.; Malmstrom, L.; Aerbersold, R. *Mol. Cell. Proteom.* (2012).

Results and Discussion. To investigate the problems of assay redundancy and specificity in metabolomics, we used computational models to calculate nonredundant theoretical assays using the UIS concept for a given metabolomic background. UIS_n is defined as a set of top *n* analyte transitions (precursor and fragment *m/z*) that map exclusively to one metabolite in the metabolome to be analyzed.[9-10] For this analysis, we simulated different MS methods (MS1-only, MRM, DIA) using the NIST LC-MS library as a background (8274 compounds at collision energy=35), using realistic values for both the precursor *m/z* window (Q1) and the fragment *m/z* window (Q3), measured in daltons (Da), or parts per million of a dalton (ppm) (Fig. 1). In our simulations, we selected a query molecule to be tested against a complex background (e.g. the full NIST library).

We compared the selected precursor and fragment ion coordinates of the query compound against each compound in the background set (excluding stereoisomers which would not be distinguishable by MS). We report any overlap in precursor and fragment ion coordinates at a given *m/z* threshold, which can occur at a level of MS2 only, with separability by high resolution MS1 data (Fig. 2A), or at both levels of MS1 and MS2 (Fig. 2B).

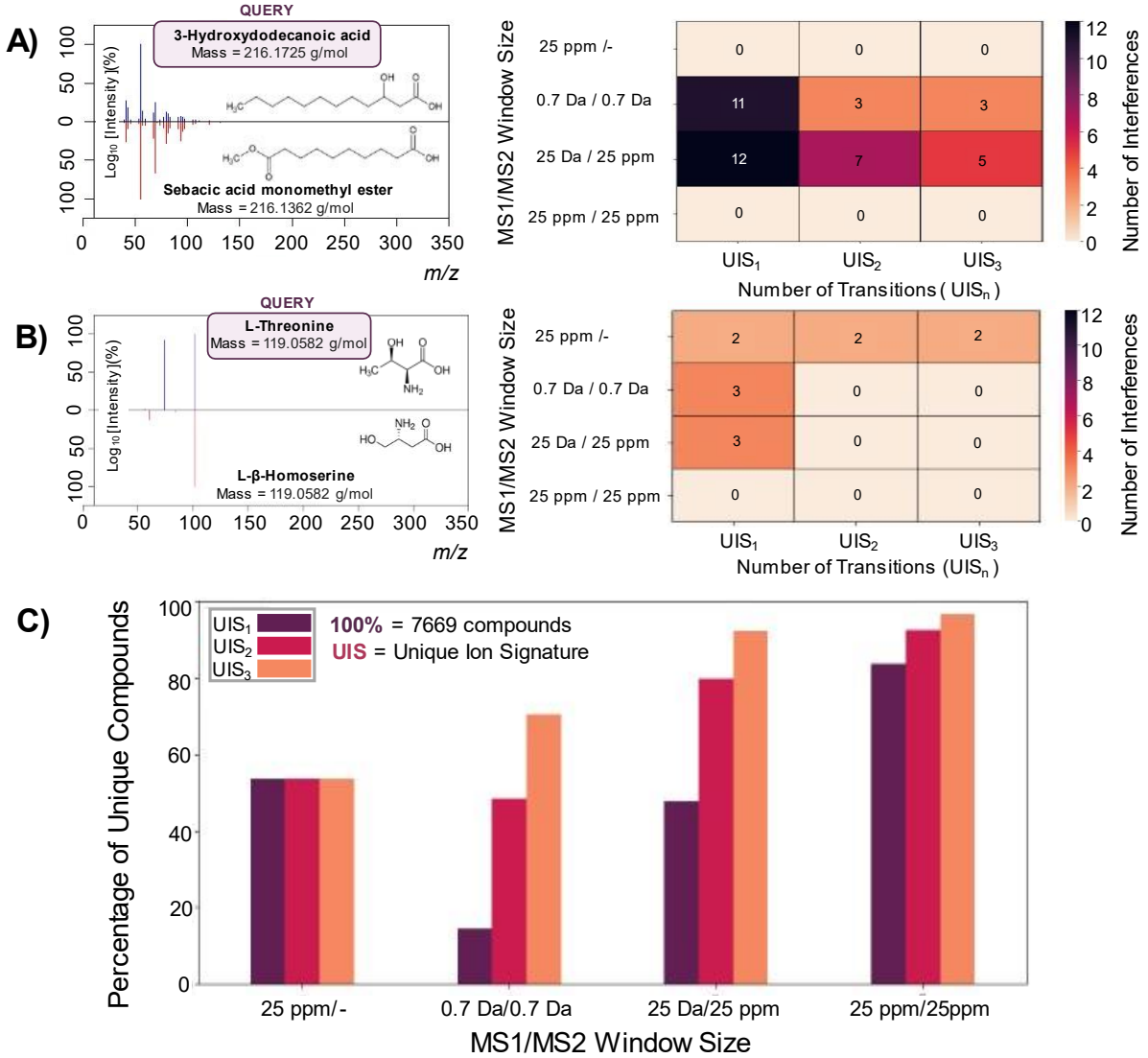


Figure 2. MS1 and MS2 contribute orthogonally to unambiguous compound detection. Pairs of isomers and their corresponding MS2 maps with relative intensity (as a percentage of the most abundant ion) on the y-axis and fragment *m/z* on the x-axis are shown. The number of interferences is represented by the gradient from orange to purple signifying low to high number of interferences for varying MS methods using MS1 and/or MS2 information (MS1-only, MRM, DIA). **(A)** 3-Hydroxydodecanoic acid and Sebacic acid monomethyl ester require high resolution MS1 for its unique identification, demonstrated by no interferences at a MS1 window size of 25 ppm/- and their similar fragmentation patterns. **(B)** L-Threonine and L-β-Homoserine require high resolution MS2, as indicated by no interferences at the presence of all MS2 window sizes at UIS_2 and their different fragmentation patterns. **(C)** Using the NIST LC/MS library (8274 compounds) as background, simulations were conducted for each analyte (7669 queries) to investigate whether an assay can be constructed that uniquely detects the analyte within the background using common acquisition methods. The following MS1/MS2 windows were set (in daltons – Da or parts per million of 1 Da - ppm): MS1 only: 25 ppm/-; Multiple Reaction Monitoring (MRM): 0.7 Da/0.7 Da; Data-Independent Acquisition (DIA): 25 Da/ 25 ppm, 25 ppm/25 ppm. The percentage of compounds in the NIST library with no interference (the background for each query based on the given parameters), known as the percentage of unique compounds (y-axis), is calculated for each method (x-axis).

MS1 and MS2 Contribute Orthogonally to Unambiguous Detection of Complex Metabolites. We simulated four different acquisition modes: MS1, MRM, and DIA with/without a high

resolution MS1 scan. For our MS1 analysis, we assumed that high resolution MS1 scans (at least 40 000 resolution or 25 ppm) would be acquired followed by extracted ion chromatogram (XIC) analyses. For our targeted metabolomics analysis, we assumed a standard QQQ instrument to be used with a 0.7 quadrupole isolation window. For DIA data we first simulated a method using a high resolution MS1 scan, followed by several high resolution fragment ion scans (at least 25 ppm) with a large (25 Da) quadrupole isolation window analyzed by both MS1 and MS2 XIC analysis. A second method was also simulated for DIA, only reliant on the high resolution MS2 scans (no MS1). Our simulations demonstrate that both accurate precursor and fragment ion information contribute orthogonally to unambiguous compound detection in complex samples (Fig. 2). By using high resolution precursor and fragment ion information, we observe improvements in the unique detection of these metabolites, maximizing the overall percentage of unique compounds detected.

First, we investigated the effects of mass accuracy on the unique detection of compounds on two representative examples: 3-hydroxydodecanoic acid and L-threonine. For the compound 3-hydroxydodecanoic acid (associated with fatty acid metabolic disorders), accuracy at the MS1 level allows unique detection with respect to sebacic acid monomethyl ester, which has a similar fragmentation pattern but differs in precursor *m/z* by over 300 ppm. These compounds are thus resolvable with a high resolution MS1 scan (within 25 ppm) but not separable using a standard quadrupole mass filter (Q1), with masses of 216.1725 Da and 216.1362 Da respectively. The similarity between their MS2 spectra highlights the importance of acquiring high resolution MS spectra, as only a high resolution MS1 precursor scan can distinguish the two analytes, while even utilizing three high resolution MS2 fragments will not uniquely map to a single analyte in the background library (Fig. 2A). In our second example, L-threonine and L-β-homoserine produce highly distinct MS2 spectra that are easily distinguishable using the second most abundant fragment ion even with a low resolution QQQ instrument, but cannot be distinguished using only MS1 information, as these amino acids have the same precursor *m/z* (mass of 119.0582 Da) but differ in their fragmentation patterns. For this pair of metabolites, their unique fragment ion spectra specifically provides optimal discriminating power, as methods using MS2 data show no interferences (Fig. 2B). These examples highlight the combined importance of both high resolution precursor and high resolution fragment ion *m/z* to provide high selectivity for analyte detection.

DIA Outperforms MRM and MS1-only Metabolomics Methods with respect to Unambiguous Detection. Next, we compared our four simulated analytical methods on the full NIST library by using each of the 7669 compounds (filtered for common adducts) as a query against the full set of 8274 compounds in the library. When quantitatively comparing different metabolomics methods, our simulations show that DIA data acquisition followed by the extraction of ion chromatograms with narrow mass tolerances (25 ppm Q1 XIC / 25 ppm Q3 XIC) outperformed both MS1-only and MRM-based methods with respect to unambiguous detection, reducing the number of ambiguous compounds 13.6-fold (46.3% ambiguous compounds for MS1-only, 29.6% for MRM UIS_3, and 3.4% for DIA at UIS_3, Fig. 2C). Interestingly, DIA outperformed both MS1-only and MRM even when not requiring the detection of any MS1 signal and solely relying on narrow mass tolerance fragment ion XICs (25 Da Q1 / 25 ppm Q3 XIC). Therefore, our analysis demonstrates that neither MS1-based extraction at 25 ppm accuracy nor a single transition in MRM (0.7 Da Q1/0.7 Da Q3) is sufficient to uniquely detect most of the compounds in the

NIST library. This demonstrates that neither reliance on few transitions (as in MRM) nor reliance on MS1 signal alone (as done often in untargeted metabolomics) is sufficient for unambiguous compound detection using mass alone, while DIA (both with MS1 and without MS1) outperforms both MS1-only and MRM-based analyses due to its capability of extracting both high resolution precursor and fragment ion traces for any analyte of interest. Interestingly, while extracting a single transition (most abundant fragment ion) at 25 ppm in DIA performed comparably to MS1-based extraction, the selectivity of fragment-ion based analysis can be boosted substantially by simply selecting multiple fragment ions for analysis (UIS_1 to UIS_3). This can also be further enhanced by requiring the presence of a precursor with high mass accuracy (rightmost bars in Fig. 2C). Specifically, the number of non-unique detections using DIA without MS1 decreases by 6.9-fold (from 52.4% to 7.6%) when using the second and third transition (UIS_3). Furthermore, when including the precursor ion trace (UIS_3), the number of ambiguous detections when using DIA decreases by 2.2-fold (from 7.6% to 3.4%). This outperforms both MS1-only and MRM-based methods by 13.6-fold and 8.7-fold, respectively.

Theoretical Saturation of Compound Uniqueness based on Number of Transitions, Matrix Complexity and Collision Energy. In order to estimate the relative contribution of individual factors to ambiguous detection in metabolomics, we performed simulations where we varied the transition number, background complexity and collision energy. Using our UIS framework, we were able to quantify the effects of matrix complexity, the number of transitions and collision energy on unambiguous detection.

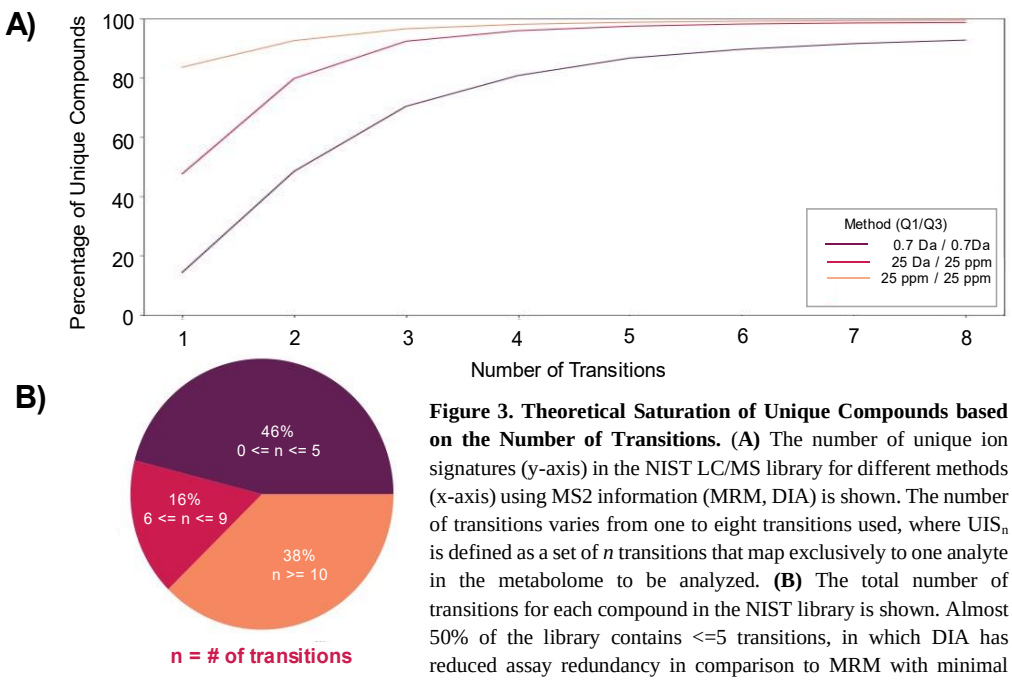


Figure 3. Theoretical Saturation of Unique Compounds based on the Number of Transitions. **(A)** The number of unique ion signatures (y-axis) in the NIST LC/MS library for different methods (x-axis) using MS2 information (MRM, DIA) is shown. The number of transitions varies from one to eight transitions used, where UIS_n is defined as a set of *n* transitions that map exclusively to one analyte in the metabolome to be analyzed. **(B)** The total number of transitions for each compound in the NIST library is shown. Almost 50% of the library contains <=5 transitions, in which DIA has reduced assay redundancy in comparison to MRM with minimal transitions available.

Experimentally, additional selectivity can often be achieved by increasing the number of monitored transitions. However, this may result in a higher limit of detection and lower throughput, thus increasing cost. To evaluate this tradeoff quantitatively, we assessed the theoretical saturation of unique compounds based on the number of transitions utilized for UIS measures (Fig. 3A). While DIA methods are close to saturation already when using the 3 most

abundant transitions and only small gains can be achieved by using more than 3 fragment ion transitions, this was not the case for traditional targeted metabolomics approaches on a QQQ instrument. Even when using 8 transitions, MRM failed to uniquely detect 7.2% of analytes (compared to the 0.5% of ambiguous analytes when using DIA with the precursor ion trace). Since about half of the NIST library compounds have less than 5 high quality transitions, the use of minimal transitions available is important to determine methods that perform best with limited data, a common occurrence in clinical literature. Our findings indicate that specifically for compounds with relatively few fragment ions, DIA strongly outperforms traditional QQQ platforms in terms of assay selectivity, highlighting the potential of this method in untargeted clinical studies (Fig. 3B).

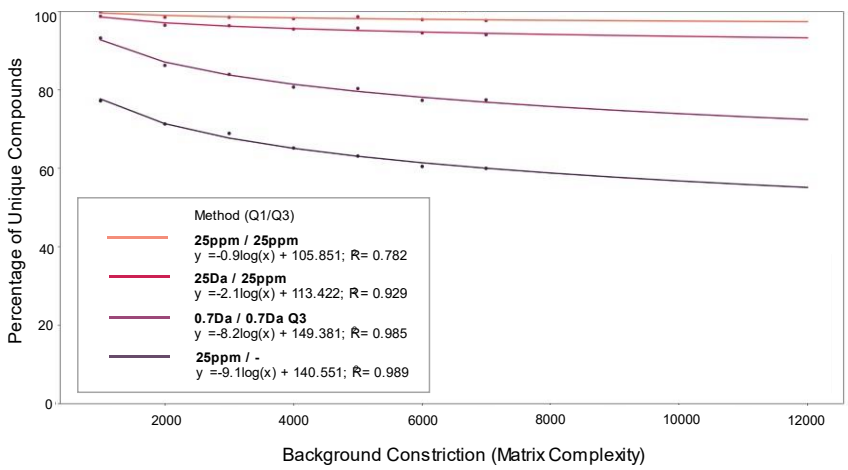


Figure 4. Theoretical Saturation of Unique Compounds based on Sample Matrix Complexity. Simulations were conducted to determine the performance of different MS methods (measured by the percentage of unique compounds, y-axis) in relation to varying library sizes (x-axis). Isolation window sizes for Q1 and Q3 are listed (Q1/Q3) in daltons (Da) or parts per million (ppm) of 1 Da. For methods using MS2, three transitions were used (UIS_3). The simulation was performed (with 7669 queries) using the NIST library as background. A saturation effect is seen at 44.9% non-uniquely detected analytes for MS1-only, 27.6% for MRM, and 2.6% for DIA when including the precursor ion trace at UIS_3.

Further, we investigated whether our findings were dependent on the specific sample matrix (~7000 NIST compounds) chosen and whether our results would change in samples of different background complexity or by restricting the number of considered compounds. Restricting analysis to a particular subset of compounds is equivalent to the practically employed approach of restricting compound detection to a narrow window of chromatographic retention time, thus effectively removing a large amount of potentially interfering background analytes. We simulated this by randomly choosing a subset of the NIST library to produce a sample matrix of lower complexity and used extrapolation to estimate how a more complex sample matrix would behave (Fig. 4). We find that the relative performance of the individual methods (MS1 profiling, MRM using a QQQ and DIA) is independent of the sample complexity, demonstrated by the level of saturation of each method in regards to the percentage of uniquely identified compounds. This demonstrates that our findings are not dependent on the compound library composition or size, but are likely to be generalizable for a wide range of analytical and sample conditions (with corresponding increased sample complexity) and would likely also hold when analysis is restricted to a small region of the chromatography. Analyzing increasing sample complexity, we found saturation beginning at around ~5000 compounds in the background (defined by <2% difference), resulting in 44.9% non-uniquely detected analytes for MS1-only, 27.6% for MRM, and 2.6% for DIA when including the precursor ion trace at UIS_3. These saturation behaviors observed here may indicate that our results would not change even for more complex backgrounds, assuming that the structural composition of more complex samples is comparable to the one studied here.

Finally, we studied whether collision energy (CE) can be used to optimize the differentiation of structural isomers in fragment spectra. While our previous analyses were restricted to spectra with CE 35, we now performed a pairwise evaluation of spectral similarity across all available collision energies in the analyte library, to answer the question whether specific collision energies are particularly efficient at producing pairwise dissimilar spectra and thus allowing compound differentiation. For this analysis we used a total set of 1088 compounds which had fragment ion spectra acquired with more than 28 collision energy settings, resulting in ~1,600,000 spectral pairs for analysis when using an MS1-only acquisition at 25 Da (Fig. 5). For each pair of interfering compounds, we computed the collision energy at which we achieved maximally different spectra, with ~50% of compounds (with a single optimal CE, 979 total) measured using a Q-TOF instrument having an optimum at CE 20 (Fig. 5B). However, metabolites also varied in the number of pairwise-optimal collision energy (POCE) from 2-13, highlighting the diversity of analyte pairs being compared. We found that the POCE varied substantially between analyte pairs, indicating that different collision energies provide orthogonal information and while one collision energy may be optimal to differentiate a target compound from another compound A, a different collision energy may be needed to differentiate it from a second compound B. The observed optimal discriminating power of collision energy between structural isomers indicates the importance of prior CE optimization in assay design, specific to instrument type and potentially even sample matrix.

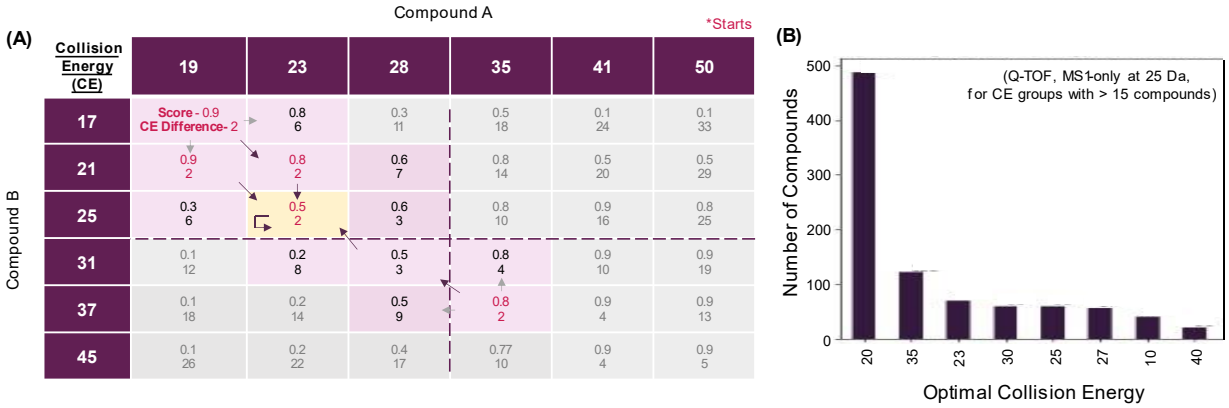


Figure 5. Optimization of Collision Energy between Interferences using Cosine Similarity. Collision energies (CE) were studied within the NIST library through pairwise evaluation of spectral similarity to determine if CE can be used to optimize the differentiation of structural isomers. Compounds from the NIST library measured using a Q-TOF were filtered by condition and window size (MS1-only at 25 Da) to calculate the optimal CE for each compound. Spectra of the remaining 979 compounds and their interferences acquired at 28 collision energies were compared by their difference and their cosine similarity score. **A)** For example, two isomers, Compound A and Compound B were compared and their similarity matrix is shown. Starts are chosen by a minimum difference in CE, as indicated in pink, and a traceback is calculated towards the minimum score. **B)** The mode of this result (if a single mode is available) for each compound in the set of 979 compounds is recorded as a pairwise optimal CE (POCE). POCE groups with more than 15 compounds are demonstrated on the x-axis, with the number of compounds per a specific POCE group measured on the y-axis.

When accounting for all factors of unambiguous detection, the overall theoretical saturation of a compound's uniqueness was observed at 44.9% non-uniquely detected analytes for MS1-only, 6.6% for MRM, and 0.6% for DIA. One important limitation of our study is that we assume that all analytes are separable by retention time (RT) and we do not simulate patterns of co-eluting fragment ions created by more than one analyte. Secondly, additional information is often used for confirmation of compound detection such as accurate retention times (derived from internal or external standards), isotopic intensity patterns or relative fragment ion intensity. However, we argue that using this additional information will likely improve selectivity across all studied

acquisition methods equally and thus not impact our overall conclusions in regards to the relative performance of the individual methods; in fact, our subsampled simulations (Fig. 4) clearly indicate that if the number of eligible interfering analytes are restricted (for example by restricting the analysis to a narrow RT window), our conclusions still hold. Similarly, in practical applications sample matrices will vary and assuming a sample matrix composed of all NIST library compounds may either be too optimistic or too pessimistic for a particular application. However, while the sample matrix will influence the absolute number of unambiguously detectable analytes, it will not impact the relative performance profile of the individual methods, which we have also shown (Fig. 4).

We demonstrate that an acquisition method using both precursor and fragment ion XIC at 25 ppm accuracy is sufficient to unambiguously detect a large number of structurally heterogeneous analytes in a complex sample matrix of over 8000 compounds. Here we show quantitatively the impact that orthogonal information contributed from both MS1 and MS2 levels have on compound detection and differentiation from background compounds using *m/z* separation. Since our method solely relies on the accuracy of *m/z* values as physical constants, our conclusions are independent of lab-specific factors and chromatographic setup. We further demonstrate that the relative performance of the studied acquisition methods is consistent over a wide number of parameters, such as sample complexity and chromatographic constriction of the search space. In addition to the criterion of *m/z*, we expect additional orthogonal sources of information to maximize unique detection in the metabolome. This is supported by previous studies demonstrating the benefits of utilizing additional information for detection, such as retention time, collisional cross section and isotopic abundance patterns (Supplementary Fig. 2).[14]

Conclusions. We use unique ion signature analyses to study the performance characteristics of several widely used acquisition methods in metabolomics and demonstrate the benefit of using both high resolution precursor and high resolution fragment ion *m/z* for unambiguous compound detection on a set of over 8000 compounds. Our study highlights the potential of DIA for unambiguous compound detection (and quantification) in complex samples. Overall, we provide a global perspective on unambiguous compound detection and present a robust framework to study this phenomenon quantitatively.

Acknowledgements. This project was funded by the Government of Canada through CIHR (FRN 166094) as well as NSERC (CREATE grant #528163-2019). P.S and H.R. conceived the project. All authors supplied ideas to the experiment design. P.S. performed the data analysis and prepared the publication. H.R. supervised the project. All authors discussed the results and edited the paper, with additional thanks to Olga Zaslaver and Oliver Alka for further discussion and edits. H.R. is supported by the Canadian Foundation for Innovation and is the Canada Research Chair in Mass Spectrometry-based Personalized Medicine.

References

1. Zhou, B.; Xiao, J. F.; Tuli, L.; Ressom, H. W. LC-MS-based metabolomics. *Mol. BioSyst.* **2012**, 8, 470-481.
2. da Silva, R. R.; Dorrestein, P. C.; Quinn, R. A. Illuminating the dark matter in metabolomics. *Proc. Natl. Acad. Sci.* **2015**, 112, 12549–12550.
3. Sumner, L. W.; Amberg, A.; Barrett, D.; Beale, M. H.; Beger, R.; Daykin, C. A.; Fan, T. W.; Fiehn, O.; Goodacre, R.; Griffin, J. L.; Hankemeier, T.; Hardy, N.; Harnly, J.; Higashi, R.; Kopka, J.; Lane,

A. N.; Lindon, J. C.; Marriott, P.; Nicholls, A. W.; Reily, M. D.; Thaden, J. J.; Viant, M. R. Proposed minimum reporting standards for chemical analysis. *Metabolomics*. **2007**, 3, 211-221.

4. Gillet, L. C.; Navarro, P.; Tate, S.; Röst, H.; Selevsek, N.; Reiter, L.; Bonner, R.; Aebersold, R. Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: a new concept for consistent and accurate proteome analysis. *Mol. Cell Prot*. **2012**, 11(6).
5. Zhu, X.; Chen, Y.; Subramanian, R. Comparison of Information-Dependent Acquisition, SWATH, and MSAll Techniques in Metabolite Identification Study Employing Ultrahigh-Performance Liquid Chromatography–Quadrupole Time-of-Flight Mass Spectrometry. *Anal. Chem.* **2014**, 86, 1202–1209.
6. Wang, R.; Yin, Y.; Zhu, Z. J. Advancing untargeted metabolomics using data-independent acquisition mass spectrometry technology. *Anal Bioanal Chem*. **2019**, 411(19), 4349-4357.
7. Bonner, R.; Hopfgartner, G. SWATH data independent acquisition mass spectrometry for metabolomics. *Trends Anal. Chem.* **2019,** 120.
8. Fenaille, P.; Barbier Saint-Hilaire, K.; Rousseau, C. J. Data acquisition workflows in liquid chromatography coupled to high resolution mass spectrometry-based metabolomics: where do we stand? *J. Chromatogr. A*. **2017**, 1526, 1-12.
9. Zhou, J.; Chen, X.; Yin, Y. Development of data-independent acquisition workflows for metabolomic analysis on a quadrupole-orbitrap platform. *Talanta*. **2017**, 164, 128-136.
10. Sherman, J.; McKay, M. J.; Ashman, K.; Molloy, M. P. How specific is my SRM?: The issue of precursor and production redundancy. *Proteomics*. **2009**, 9, 1120-1123.
11. Rost, H.; Malmstrom, L.; Aerbersold, R. A computational tool to detect and avoid redundancy in selected reaction monitoring. *Mol. Cell. Proteom.* **2012**, 11, 540-549.
12. Helsens, K.; Mueller, M.; Hulstaert, N.; Martens L. Sigpep: calculating unique peptide signature transition sets in a complete proteome background. *Proteomics*. **2012**, 12, 1142-1146.
13. Yang, X.; Neta, P.; Stein, S. Extending a Tandem Mass Spectral Library to Include MS2 Spectra of Fragment Ions Produced In-Source and MSn Spectra. *J. Am. Soc. Mass. Spectrom.* **2017**, 28, 2280-2287.
14. Kind, T.; Fiehn, O. Metabolomic database annotations via query of elemental compositions: mass accuracy is insufficient even at less than 1ppm. *BMC. Bioinform.* **2006**, 7.