

UC Riverside

UC Riverside Previously Published Works

Title

Evaluation of Data-Dependent MS/MS Acquisition Parameters for Non-Targeted Metabolomics and Molecular Networking of Environmental Samples: Focus on the Q Exactive Platform.

Permalink

<https://escholarship.org/uc/item/4b35w89t>

Journal

Analytical Chemistry, 95(34)

Authors

Stincone, Paolo
Pakkir Shah, Abzer
Schmid, Robin
[et al.](#)

Publication Date

2023-08-29

DOI

10.1021/acs.analchem.3c01202

Peer reviewed

Evaluation of Data-Dependent MS/MS Acquisition Parameters for Non-Targeted Metabolomics and Molecular Networking of Environmental Samples: Focus on the Q Exactive Platform

Paolo Stincone, Abzer K. Pakkir Shah, Robin Schmid, Lana G. Graves, Stilianos P. Lambidis, Ralph R. Torres, Shu-Ning Xia, Vidit Minda, Allegra T. Aron, Mingxun Wang, Chambers C. Hughes, and Daniel Petras*

Cite This: *Anal. Chem.* 2023, 95, 12673–12682

Read Online

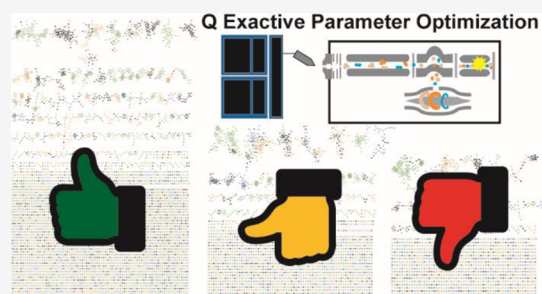
ACCESS |

Metrics & More

Article Recommendations

Supporting Information

ABSTRACT: Non-targeted liquid chromatography–tandem mass spectrometry (LC–MS/MS) is a widely used tool for metabolomics analysis, enabling the detection and annotation of small molecules in complex environmental samples. Data-dependent acquisition (DDA) of product ion spectra is thereby currently one of the most frequently applied data acquisition strategies. The optimization of DDA parameters is central to ensuring high spectral quality, coverage, and number of compound annotations. Here, we evaluated the influence of 10 central DDA settings of the Q Exactive mass spectrometer on natural organic matter samples from ocean, river, and soil environments. After data analysis with classical and feature-based molecular networking using MZmine and GNPS, we compared the total number of network nodes, multivariate clustering, and spectrum quality-related metrics such as annotation and singleton rates, MS/MS placement, and coverage. Our results show that *automatic gain control*, *microscans*, *mass resolving power*, and *dynamic exclusion* are the most critical parameters, whereas *collision energy*, *TopN*, and *isolation width* had moderate and *apex trigger*, *monoisotopic selection*, and *isotopic exclusion* minor effects. The insights into the data acquisition ergonomics of the Q Exactive platform presented here can guide new users and provide them with initial method parameters, some of which may also be transferable to other sample types and MS platforms.



INTRODUCTION

Liquid chromatography–tandem mass spectrometry (LC–MS/MS)-based non-targeted metabolomics is a central tool for the detection, annotation, and relative quantification of small molecules in diverse ecosystems.^{1–4} Electrospray ionization in combination with time-of-flight (ToF) or Orbitrap mass analyzers provides excellent analytical performance with regard to mass resolving power, mass accuracy, sensitivity, and speed.^{5–7} High mass resolving power and mass accuracy are especially needed when analyzing complex samples such as natural organic matter (NOM). Non-targeted LC–MS-based metabolomics can be performed using different acquisition modes such as the data-dependent acquisition (DDA) and data-independent acquisition (DIA) modes, and modified versions of the latter such as the data set-dependent acquisition (DsDA) mode.⁸ Most non-targeted metabolomics experiments are performed in the DDA mode through which both MS1 and MS/MS data are collected, benefiting from a larger variety of tools and workflows for data processing.⁹

Spectral libraries provide the highest confidence level for high-throughput metabolite annotation.^{10–16} Multiple workflows and open-source software solutions exist for feature detection,

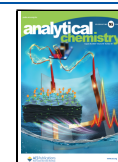
network propagation, spectral matching, and in silico annotation as well as repository scale analysis of MS/MS spectra,^{14,17–35} which have been used in numerous lab-culture, mesocosm, and environmental studies.^{36–43} While non-targeted metabolomics aims to annotate and provide relative quantification of all metabolites present in a sample, most of the MS/MS spectra remain unidentified.⁴⁴

Classical molecular networking (CMN) and feature-based molecular networking (FBMN) emerged as key tools used by the community to propagate annotations through MS/MS spectral similarity networks. While CMN represents clustered, nearly identical MS/MS spectra as nodes, FBMN requires detected LC–MS/MS features. CMN aims to reveal all chemical diversities of a sample set and enables qualitative comparisons between cohorts as well as putative identifications.

Received: March 19, 2023

Accepted: July 19, 2023

Published: August 14, 2023



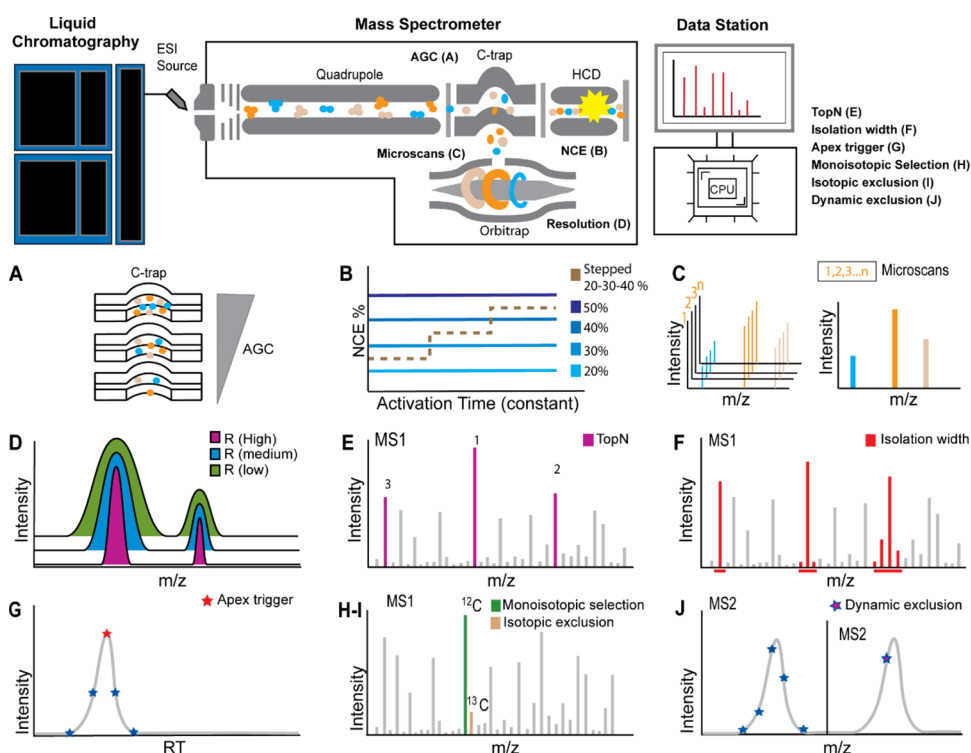


Figure 1. Schematic representation of the LC–MS/MS Q Exactive HF Orbitrap and each of the investigated parameters. (A) Automatic gain control (AGC) represents the number of ions filling the C-trap before being transferred into the Orbitrap. (B) Normalized collision energy (NCE) defines the percentage of energy applied for precursor fragmentation in the HCD cell. (C) Microscans are the number of internal scans that are merged into the written scans. (D) Mass resolving power (R) influences how well m/z signals are resolved. (E) TopN enables the selection of the N most abundant MS1 peaks to be isolated for subsequent MS/MS fragmentation. (F) Isolation width defines the m/z range around the precursor ion for MS/MS fragmentation. (G) Apex trigger allows the mass spectrometer to trigger the MS/MS event at or near the apex of the chromatographic peak rather than at the front. (H) Monoisotopic selection includes only monoisotopic precursors for MS/MS fragmentation and (I) isotopic exclusion allows the exclusion of isotopes for subsequent duty cycles. (J) Dynamic exclusion excludes precursor ion m/z for a specified time frame after the first MS/MS experiment.

Similar to CMN, FBMN aims to display all the chemical diversity in an experimental sample and, in addition, provides quantitative comparisons between cohorts and resolved isomers, and enables putative identifications.²⁵

The optimization of DDA parameters is crucial for obtaining the most comprehensive molecular insights into a sample.⁴⁵ Optimal DDA settings are particularly important for MS/MS-based analytical approaches such as CMN or the use of spectral counts. However, as DDA settings and the related duty cycle times determine MS acquisition speed and spectral coverage, they are also important parameters that dictate the shape of MS1 chromatographic peaks, which are in turn crucial for effective MS feature detection (often referred to as “feature finding”).^{45–48} In recent studies, both LC and MS parameters were optimized for metabolomic analysis of marine and plant samples using different qTOF platforms and GNPS MN analyses, which showed a differential influence of key acquisition parameters on CMN and FBMN.^{49,50}

The aim of this study was to evaluate the influence of DDA settings for the Q Exactive platform, a widely used MS instrument type for metabolomics analysis. According to dataset statistics in public MS repositories such as MassIVE, 39% of a dataset containing a “GNPS tag” in the title were generated on orbitrap-based instruments (as of June 14, 2023). For the optimization, we focused on complex natural organic matter environmental samples from marine, river, and soil environments, which are considered to be among the most complex samples metabolomics researchers may encounter.⁵¹ During

data acquisition of each sample type, we altered the 10 most common acquisition parameters (explained in detail in the [Experimental Section](#)) which resulted in a total of 35 combinations of settings. The data were then processed by both CMN and FBMN using MZmine 3 and the GNPS platform.^{15,25,27,52} We evaluated the different settings based on the number of library annotations, single nodes, as well as overall feature number and clustering in principal component analysis (PCoA). In addition to the experimental insights gained for the optimization of environmental metabolomics analysis using non-targeted LC–MS/MS, we set out to provide a general overview of key DDA settings of the Q Exactive platform that we consider crucial for effectively optimizing methods and attaining the highest quality results.

EXPERIMENTAL SECTION

A detailed overview of modified data-dependent acquisition parameters highlighted in [Figure 1](#) is provided in the [Supporting Information](#).

Ocean and River Water Sampling. Surface coastal seawater was collected off the Ellen Browning Scripps Memorial Pier (32°52′01.5″N 117°15′26.9″W) on February 26, 2021 between 11:00 and 19:00 PDT. Seawater was stored in acid-washed 20-L HDPE carboys (Nalgene), filtered through an AcroPak 0.8/0.45 μm Supor membrane filter (Pall Corporation), and subsequently acidified to pH \sim 2 (37% HCl, TraceMetal Grade, Fischer Chemical). The seawater was

extracted through PPL cartridges (Bond Elut, Agilent).⁵³ River water samples were collected from the Neckar river in Tübingen, Germany, in the morning of February 16, 2022. The water was collected in glass bottles, previously washed with 100% methanol, via grab sampling. Samples were immediately brought to the laboratory for processing and analysis. Half of the sample volume was acidified to pH \sim 2 (37% HCl, TraceMetal Grade, Fischer Chemical). For solid phase extraction of the non-acidified river water, and acidified river water three different types of resins were used: C18, PPL, and HLB. Extracts were pooled together for the analysis to a final concentration of 10 mg/mL in 50% MeOH.

Solid-Phase Extraction of Water Samples. Solid-phase extraction (SPE) was carried out for river and ocean waters. This extraction methodology allows the capturing of a myriad of natural and anthropogenic organic molecules composing the dissolved organic matter (DOM).^{54,55} The SPE resins were activated with a 3 \times cartridge volume of 100% methanol (LC/MS grade, Fisher Scientific) and conditioned with a 1 \times cartridge volume of the 3% methanol (LC/MS methanol diluted in LC/MS H₂O, Fisher Scientific). The resins did not run dry between the washes. Subsequently, 500 mL of each sample type was loaded into each resin type via vacuum-assisted flow at a rate of 20 mL/min with the help of a vacuum SPE station (Agilent). Following loading, the resins were flushed with 1 \times cartridge volume of 3% methanol and eluted using 2 mL 80% methanol into glass LC vials. The sample extracts were then transferred to preweighed glass vials and dried down using a speedvac at room temperature. Each vial was weighed a second time to quantify the extract and stored at -80 °C until further analysis. The extracts were resuspended to a concentration of \sim 10 mg/mL in 50% methanol/water (LC/MS grade). Samples were analyzed immediately following resuspension.

Sampling and Extraction of Soil Organic Matter. Soil from a depth of 1–50 cm was collected from the Schoenbuch forest in Bebenhausen (Baden-Württemberg, Germany) on December 16, 2021, using a steel shovel, and stored for transportation in a polypropylene bucket. The soil was spread out as a thin layer over aluminum foil and allowed to dry for 48 h at room temperature. The dry material, which was sieved to remove plant debris and gravel, weighed 4 kg. Ethyl acetate (5 L) was added, and the resulting heterogeneous solution was rigorously swirled for 20 min and then filtered through a Büchner funnel (20 cm diameter) fitted with filter paper (qualitative filter paper; Macherey-Nagel 615, 18.5 cm). The clear brown solution was dried to give 3.6 g of extract, from which an aliquot was diluted to 10 mg/mL with 50% methanol in water for LC–MS/MS analysis.

Default Liquid-Chromatography Tandem Mass Spectrometry Method. A Q Exactive HF (Thermo Fisher Scientific) with a heated electrospray ionization (HESI) source, coupled to a Vanquish ultrahigh-performance liquid chromatography (UHPLC) system, was used. Based on a default LC–MS/MS method,⁵³ 35 methods with different DDA settings were created. For each method, 5 μ L of the samples were analyzed in technical triplicates. Gradient elution was performed with H₂O + 0.1% formic acid (FA) as solvent (A) and acetonitrile + 0.1% FA as solvent (B). A C18 porous core-shell column (Phenomenex Kinetex C18, 150 \times 2.1 mm, 1.7 μ m particle size, 100 Å pore size) was used as the stationary phase. The chromatographic conditions included a flow rate of 0.5 mL/min and sample elution with a linear gradient from 0 to 0.5 min, 5% B, 0.5 to 8 min, 5 to 50% B, 8 to 10 min, 50 to 99% B, followed by

isocratic 99% B until 13 min as a “washout phase” and from 13 to 16 min as column re-equilibration phase at 5% B. HESI source parameters were set to 50 AU sheath gas flow rate, 12 AU auxiliary gas flow, 1 L min⁻¹ sweep gas flow, and an auxiliary gas temperature of 400 °C. The spray voltage was set to 3.5 kV and the inlet capillary temperature to 250 °C. Then, a 50 V S-lens RF level was applied. The scan range was set to 150–1500 *m/z* at a default resolution of 120,000, 1 Microscan, and an AGC target of 1E6 with a maximum ion injection time set to 100 ms. The default DDA method included up to 5 MS/MS spectra per MS survey scan (TopN), with a resolution of 15,000, 1 Microscan, and an AGC target of 5E5 with a maximum ion injection time for MS/MS scans set to 50 ms. The other setting parameters used included an isolation width of *m/z* 1, dynamic exclusion set to 5 s, apex trigger between 2 and 15 s, and isotopic exclusion set to ON. The normalized collision energy was stepwise increased from 25 to 35 to 45% with *z* = 1 as the default charge state.

Data Analysis. LC–MS/MS raw data were converted to .mzML file format by selecting the MS/MS peaks using msConvert.⁵⁶ Classical molecular networks (CMNs) were created with GNPS¹⁵ using the following settings: both precursor ion mass tolerance and MS/MS fragment ion tolerance were set to 0.01 Da, edges were filtered for cosine scores >0.7 , and at least 6 matched peaks were required for spectral matching. Edges connect up to 10 most similar other nodes. Spectral matching against the GNPS spectral library was performed with at least 6 matched peaks and a cosine score >0.7 . CMN was investigated separately for each one of the considered settings and the group of samples included in this work (GNPS links are available in the [Supporting Information](#)). PCoA was performed with Qiime2 using the Jaccard distance metric,⁵⁷ which determines the distance between each sample by comparing their shared MS/MS intensities from CMN or MS1 peak area from the feature table from FBMN. The output files “qiime2_emperor.qzv” from GNPS utilized to construct both CMN and FBMN PCoA plots in Qiime2 are available under the job links “Combined CMN and Combined FBMN” ([Table S-3](#)), and the input feature tables are provided in the [Supporting Information](#).

Feature detection was performed with MZmine3 version 3.3.0 and 3.6.0⁵² <https://github.com/mzmine/mzmine3/releases/>. The detailed settings for MZmine3 are provided in the batch file (.xml format) in the MassIVE dataset. In short, MS intensity thresholds were set to 1E5 for MS1 and 1E3 for MS/MS. XICs were built using the ADAP chromatogram builder with a minimum peak intensity of 3E5 for the MS, and a mass window of 5 ppm. XICs were then deconvoluted using “the local minimum resolver” with a chromatographic threshold of 80%, a minimum retention search range of 0.040 min, and a peak minimum absolute height of 3E5. Feature alignment using the join aligner function was carried out after the ¹³C filter (or isotope grouper). For the grouping of isotope peaks and the alignment of the peak features between samples, an *m/z* tolerance of 5 ppm and retention time tolerance of 0.1 min were used. The resulting feature table (.csv) and MS/MS spectra files (.mgf) were exported, uploaded to the MassIVE repository, and used for the FBMN analysis in GNPS. Additionally, the precursor purity, the number of MS/MS fragments per feature, and the proximity to the feature apex in the retention time dimension were exported from MZmine 3. For FBMN, the settings were kept the same as those reported above for the CMN (precursor and fragment ion tolerance: 0.01 Da, cosine scores >0.7 , at least 6 matched peaks, and top 10 edges per

node). Cytoscape 3.7 was used for the visualization of both CMN and FBMN results.⁵⁸ Upset plots were generated with the online tool Intervene.⁵⁹ The Circle packing plots used for inspecting the number of unique library IDs in each one of the sample group settings were obtained from the RAWGraphs online software. The paired t-student test was used to test the significance of the apex proximity values for each setting. CMN and FBMN were used to assess the number of nodes, library IDs, and self-loops and to calculate the annotation and self-loop rates as follows:

$$\text{Annotation rate} = \frac{\text{Number of Library IDs}}{\text{Number of nodes}} \times 100$$

DDA duty cycles were calculated using the XIC of the six internal standards (amitryptiline, sulfamethazine, sulfachloropyridazine, sulfadimethoxine, sulfamethizole, coumarin-314) by dividing the peak width through the number of MS1 data points per peak.

RESULTS AND DISCUSSION

Here, we assessed the influence of DDA settings on CMN and FBMN results following non-targeted LC–MS/MS analysis of

Table 1. List of the Modified Parameters^a

settings	parameters (default settings in bold)
mass resolving power	15, 30, 60, 120, 240
TopN	3, 5, 7, 10, 15
normalized collision energy (NCE)	stepped 20-30-40, 20, 30, 40, 50
isolation width (<i>m/z</i>)	1, 2, 3, 4
apex trigger	on, off
dynamic exclusion	on, off
microscans	1, 2, 3, 4
isotopic exclusion	on, off
monoisotopic selection	on, off
AGC target	1E5, 5E5, 1E6, 5E6

^aTen parameters modified for the DDA optimization.

samples from three types of environments: ocean, river, and soil. To optimize the settings, we started from a typical non-targeted DDA MS/MS method with default settings (Table 1, marked as bold) based on our previous studies.^{53–55} We modified 10 key parameters in an iterative fashion, adjusting one parameter at a time, yielding a total of 35 LC–MS/MS methods (Table 1). Some of these settings are commonly found in quadrupole-hybrid- and ion trap-based mass spectrometers with DDA MS/MS capabilities.

Principal Coordinate Analysis of Molecular Network-ing Results. As the first data analysis step after CMN and FBMN processing, we performed principal coordinate analysis (PCoA) using the Jaccard distance metric (Figure 2). In both PCoAs, we observed strong clustering based on sample types (ocean, river, and soil) as expected. Notably, the distribution of each sample group (ocean, river, and soil) in CMN (Figure 2A) was further spread over PCo1 and PCo2, while the different sample types in FBMN (Figure 2B) clustered more closely with higher variance explained through both PCo1 and PCo2. This was anticipated due to the MS1-based chromatographic feature detection used in FBMN. Indeed, MS1-based alignment is less dramatically affected by changed MS/MS settings than MS/MS-level clustering. The highest variance in PCoA space in CMN was observed for settings such as AGC, microscans, mass resolving power, and apex trigger. In CMN, settings with a

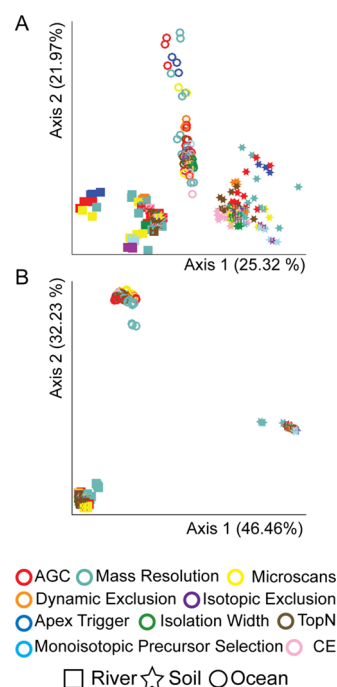


Figure 2. Principal coordinate analysis (PCoA) displays the dissimilarities among sample sources and instrument parameters. Data generated by (A) CMN and (B) FBMN. Squares, stars, and circles are used to distinguish river, soil, and ocean sample sources, respectively. The different colors indicate the 10 settings investigated in this study.

medium level of dissimilarities in PCoA included the precursor filters such as monoisotopic precursor selection and isotopic exclusion and CE. In contrast, the visualization of FBMN data through PCoA showed lower dissimilarity between most of the settings tested. The biggest dissimilarity was observed for mass resolving power (240 K) (see Figure 2B), most likely due to the increase in duty cycle time with the longer transient and therefore lower MS/MS coverage.

Network Statistics of Classical and Feature-Based Molecular Networks. We used the total and unique number of library IDs and network statistics of the CMN results as the first metric to assess the different DDA settings. To determine whether most of the nodes are shared among the settings groups examined and if some of the nodes are uniquely observed under particular settings, we generated Upset plots (Figure 3A) to highlight the intersection between the nodes and the settings groups. In the CMN results, most network nodes (clustered MS/MS spectra) were detected within all settings groups. Nevertheless, a large number of subnetworks appeared to come from specific single-setting groups, such as NCE and isolation width. Inspecting the unique library IDs in CMN per individual settings, the packing graph (Figure 3B) showed that 3 and 4 *m/z* isolation widths, as well as 50 and 20% NCE, were the settings with the highest number of unique IDs. The most straightforward explanation for this is that different collision energies for MS/MS fragmentation might expand the matches of MS/MS spectra in the library to those acquired at a similar collision energy.¹⁰ The 20% NCE method generated higher spectral uniqueness for CMN for all three sample types. For the largest number of the identified compounds, we observed the best results with stepped collision energy at 30% NCE for the three sample types (Table S-1). With increasing precursor isolation

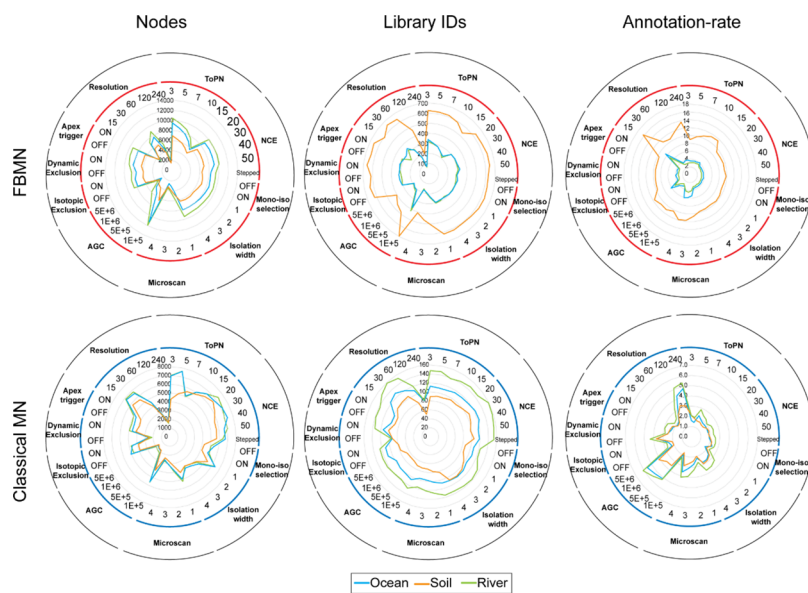
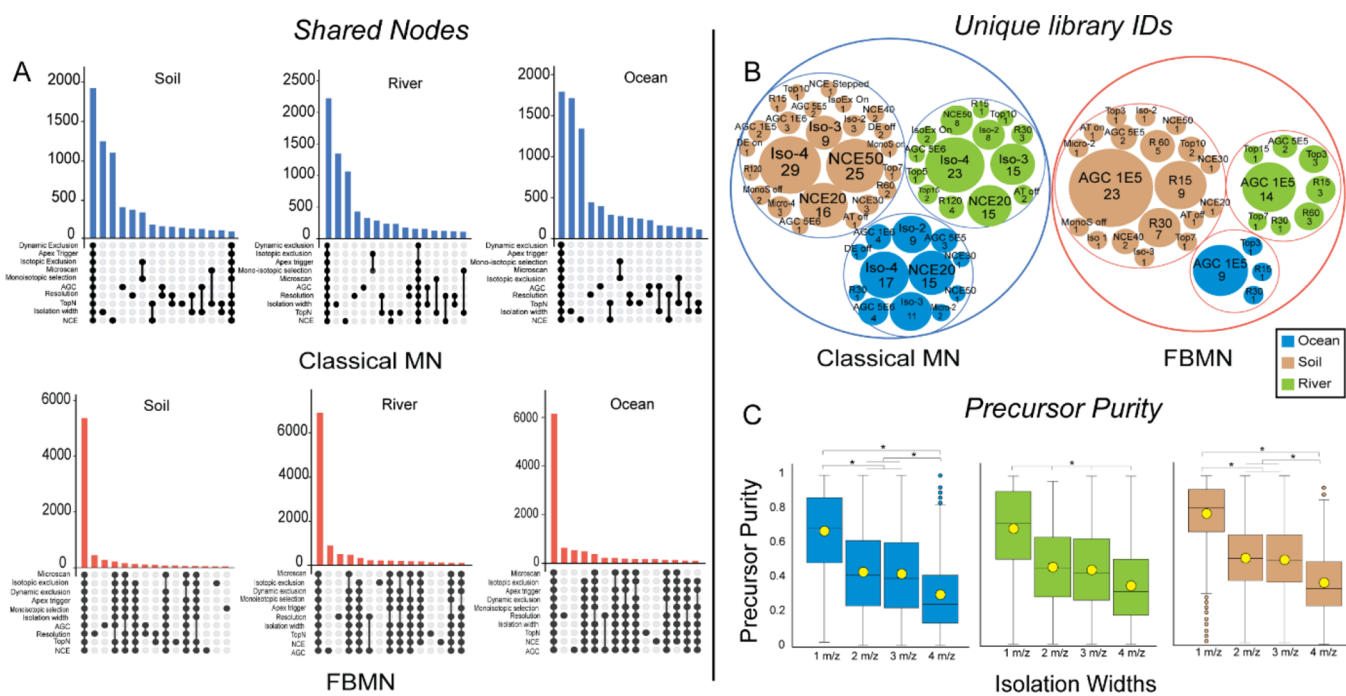


Figure 4. Radar charts of the main CMN and FBMN results. The total number of nodes, library IDs and the calculated annotation rate for each of the 35 methods investigated. Note the radial scale varies across graphs.

window size, we observed decreasing precursor purity and thus more chimeric MS/MS spectra (Figure 3C). The highest precursor purity we observed was expected with 1 m/z isolation windows. Counterintuitively, the lower precursor purity did not seem to have a negative effect on the total number of library IDs. We assume that the highest abundant compounds in the samples are also the ones that are typically identified, due to higher representation in public spectral libraries as well as higher spectral signal-to-noise ratios, which most likely also applies to

wider isolation windows. On the other hand, widening the isolation window typically increases ion transmission efficiency, which could be the reason why we observed a slightly higher number of unique and total library IDs with wider isolation windows. Although the wider isolation windows slightly increase the number of library annotations, widening the isolation window in complex samples inherently increases the number of chimeric spectra and may result in more false positive and false negative annotations. As a default value, we thus recommend to

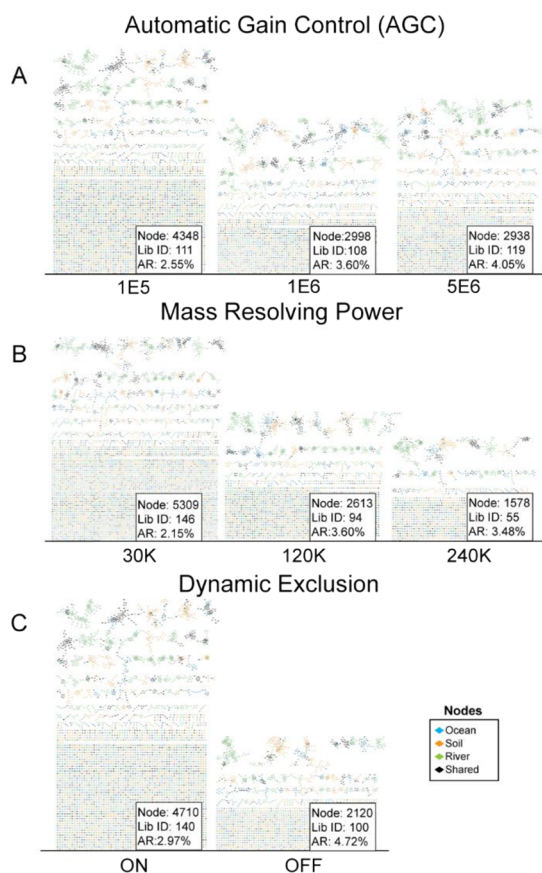


Figure 5. Impact of the three most important settings on classical molecular networks. To differentiate between unique and shared nodes among the sample sources, different colors were used. (A) Effects of high, medium, and low AGC targets; (B) effects of three mass resolving power levels; (C) dynamic exclusion effects when enabled (ON) or disabled (OFF). Network statistics are reported in small boxes at the bottom right of each network, including nodes, library IDs, and annotation rate (AR).

keep the isolation window narrow (e.g., 1 m/z) and only carefully optimize/widen this setting.

Compared with CMN, in FBMN, we observed a strong reduction of individual nodes and subnetworks (Figure 3B). However, there were still some individual nodes among the settings, mostly within the AGC and mass-resolving power settings group. This difference between FBMN and CMN is mainly due to the clustering of consensus generation between MS/MS in FBMN and thus is mainly caused by differences in MS1 chromatographic feature detection. In FBMN, the key settings related to unique library IDs are thus not MS/MS specific settings such as CE or isolation width but rather AGC target value and mass resolving power (Figure 3B). AGC 1E5 as well as mass resolving power of 15 and 30 K showed the highest number of unique library IDs. Generally, the AGC target should be optimized to limit the ions injected into the orbitrap to avoid space-charge effects and to keep the duty cycle time short enough in order to acquire sufficient data points to represent the chromatographic peak shapes. Otherwise, the quality of feature detection in FBMN decreases, picking up more noise or missing relevant features. This effect not only applies to the higher AGC target values but also to other settings that influence scan speed and thus duty cycle times (e.g., mass resolving power, microscans, TopN).

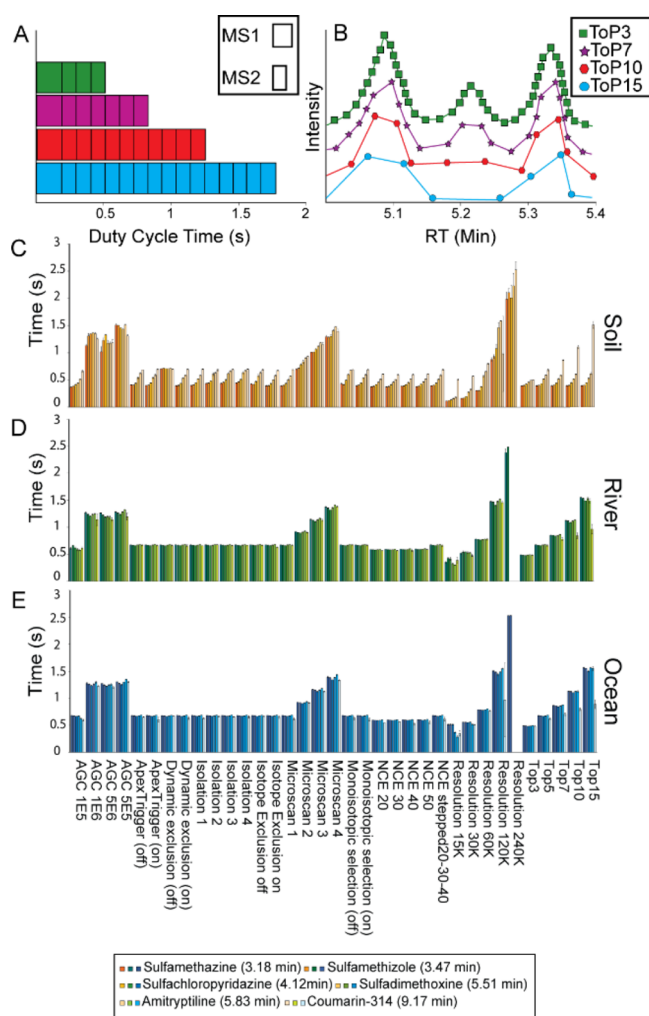


Figure 6. MS/MS duty cycle time. (A) Graphical representation of the time used by the MS scan and the following MS/MS scans needed to conclude a full cycle. (B) TopN time required for the subsequent MS scan and the related MS/MS and the influence on the peak shape. (C–E) XICs of internal standards were used to monitor the duty cycle time at different retention times. The data show the average time that was required to complete each duty cycle on each of the 35 methods.

Next, we investigated the total number of nodes, self-loops, and library IDs obtained by each individual parameter. The network statistics such as connectivity (number of linked nodes) and library annotation rates (number of nodes that matched a library spectrum) are important factors in the assessment of method performances and dataset coverage. We calculated the annotation rate (AR) based on the number of nodes detected with each individual setting (Figure 4 and Tables S-1, S-2). The AR is directly proportional to library IDs and inversely proportional to nodes. According to the number of nodes and the library IDs generated (Figure 4 and Tables S-1, S-2), the following settings (and corresponding values) showed the strongest influence: dynamic exclusion (on), mass resolving power 15 K, AGC (1E5), microscans (1), TopN (5), and isolation width (1 m/z). NCE, monoisotopic selection, isotopic exclusion, and apex trigger, instead, showed little influence. As expected, ocean and river water and soil NOM samples exhibited different molecular network results; however, the trends were similar overall. As shown in the radar charts (Figure 4), FBMN analysis generated slightly more nodes overall than

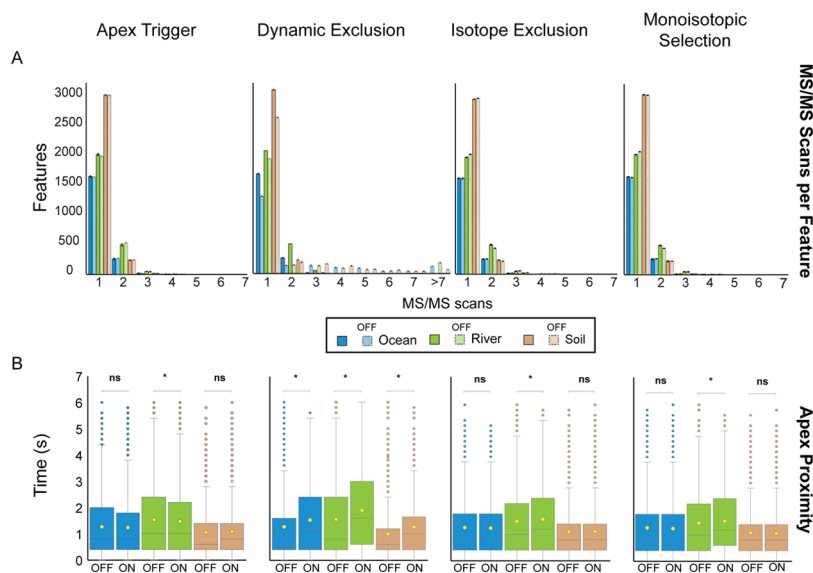


Figure 7. Influence of DDA filters on MS/MS placement. (A) Number of MS/MS scans per feature and their dependence on acquisition parameters (B) Influence of enabling (ON) and disabling (OFF) the indicated settings on the absolute apex proximity, which is expressed in seconds (s). Statistical analysis was performed using the paired t-student test. Asterisks indicate statistical significance (p -value < 0.01); n.s.: no significant difference.

CMN analysis, especially for ocean and river water, while soil samples generated a reduced number of nodes for both molecular networks analysis. However, soil samples exhibited twice the ARs in comparison to the ocean and river samples, in both CMN and FBMN. The difference in soil, ocean, and river matrices could be attributed to their distinct chemical spaces.

Overall, in CMN, soil samples showed almost 10 times fewer library IDs as compared to FBMN, and consequently, the AR was also reduced. A possible explanation for these variations between the different samples might also be related to the two different extraction methods used (solid phase extraction for ocean and river samples vs liquid–solid extraction for soil). In Figure 5, the global CMNs of individual settings with the highest impact are shown. A drastic overall impact of the effects of AGC, mass resolving power, and dynamic exclusion can be seen between the networks in terms of overall network size (Figure 5A–C).

DDA Duty Cycle Time. The duty cycle defines the time in which the Orbitrap is busy with a single measurement cycle (MS and corresponding DDA MS/MS scans) before it can proceed to the next MS survey scan. In the DDA mode, the cycle time includes the time range needed to acquire the predefined number of MS/MS scans that follow the survey scan MS (Figure 6A). Duty cycle times in all the samples were derived from MS data points from the XICs of the 6 internal standards that were added to all samples. In our DDA optimization experiments, the total number of duty cycles per metabolomics analysis was strongly influenced by multiple parameters. TopN played a key role in defining the number of MS/MS (Figure 6B). Mass resolving power, microscans, and AGC also contributed heavily to the duty cycle time (Figure 6C,D,E). There were similarities among the sample sources, but soil samples showed fewer differences than ocean and river samples. A reason might be related to the lower complexity of soil samples in comparison to ocean and river samples, as described above. It is important to point out that in the ocean and river samples, the internal standards with a long retention time (i.e., sulfadimethoxine, amitryptiline, and coumarin-314) were not detected at MS/MS

level when the mass resolving power was set to 240 K (Figure 6D,E), indicating the importance of appropriate settings and duty cycle time for best chemical coverage.

MS/MS Placement. For the total number and distribution of MS/MS scans along the run, important roles are given to the dynamic exclusion (DE) filter (Figures S-1 and S-2), as well as isotopic exclusion, and apex trigger function. DE highly influences the number of MS/MS scans per feature and their apex proximity as well. Disabling this filter drastically increased the number of MS/MS scans per feature (3 MS/MS scans and more per feature, Figure 7A) for all sample types, and as consequence, also resulted in a closer distance of the top MS/MS scan to the MS XIC apex (Figure 7B). However, when disabled, a considerable reduction of nodes and library IDs are reported for both CMN and FBMN (Tables S-1 and S-2), which indicates that keeping this filter enabled is advisable. The other filters investigated in this work (apex trigger, monoisotopic selection, and isotopic exclusion) did not show strong differences in the total number of MS/MS fragmentation scans per feature between their switch ON and OFF states (Figure 7A). When investigating the MS/MS apex proximity, we observed significant differences across the three sample types, with a slight reduction of the apex proximity when apex trigger was enabled. However, we observed a higher apex proximity with statistical significance only for the river samples (Figure 7B).

In addition, the activation or deactivation of the other two filters (isotopic exclusion and monoisotopic selection) had a significant impact only on the river samples, with higher apex proximity when these filters were disabled. This might be due to the higher feature number of the river samples.

CONCLUSIONS

The goal of this work was to test the influence of different DDA LC–MS/MS settings for non-targeted metabolomics analysis of diverse environmental samples: ocean water, river water, and soil. The degree of optimization was assessed by CMN and FBMN analyses with MZmine3 and GNPS. We defined a group

of different key metrics, which include clustering in PCoA, number of library IDs, nodes, annotation rate, and unique library IDs resulting from the CMN and FBMN analyses. In addition, we took into account the duty cycle time calculated for all the 35 setting methods and the MS/MS placement calculated for the 4 DDA setting filter methods that we examined in this study. According to our results, the most important settings are AGC target, mass resolving power, dynamic exclusion, and microscans. Isolation width, TopN, and collision energy had moderate and apex trigger, monotonic selection, and isotope exclusion minor effects. Settings that directly influenced the duty cycle length (e.g., mass resolving power, microscans, and TopN) had a strong effect on CMN and FBMN and general depth of MS/MS coverage, whereas parameters that only influenced fragmentation and spectrum quality (e.g., isolation width and collision energy) were more critical for CMN and less for FBMN coverage.

As starting settings for the new method development of complex (environmental) samples using C18 UHPLC columns with a gradient time of 10 min, we suggest: dynamic exclusion (ON), AGC (1E5), mass resolving power (30 K), apex trigger (ON), isotope exclusion (ON), isolation width (1 m/z), microscans (1), monoisotopic selection (ON), NCE (20 or stepped NCE), and TopN (5). It is important to point out that some settings, such as NCE, are more important for MS/MS-based clustering in CMN and compound identification through spectral matching, while it is less important for MS1-based feature detection in FBMN.

Finally, our results and guidelines are specific to the three sample matrices and the MS platform tested, and all settings should be adjusted/optimized for different samples and mass spectrometer types. Nevertheless, this work may serve as an initial starting point from which users, especially those new to non-targeted LC-MS/MS, can build upon.

■ ASSOCIATED CONTENT

Data Availability Statement

All raw and processed LC-MS/MS data are available through the MassIVE repository (massive.uscd.edu) under the following identifier: MSV000088937. All molecular networking jobs processed through the GNPS environment are provided in Table S-3 of the Supporting Information. Feature tables from both feature-based and classical molecular networking which were used for downstream statistical analysis are provided as supporting files in .csv file format.

SI Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.analchem.3c01202>.

Detailed description of modified data-dependent acquisition parameters, including automatic gain control, collision energy, microscans, mass resolving power, TopN, isolation width, apex trigger, monoisotopic precursor selection, isotopic exclusion, and dynamic exclusion; MS/MS scatter plots; and a table with all GNPS job links as well as a table with all compounds annotated by the classical molecular network using the GNPS libraries (PDF)

Feature tables from classic and feature-based molecular networks are provided as .csv files with classical molecular network feature intensities (CSV)

Feature tables with feature-based molecular network feature intensities (CSV)

■ AUTHOR INFORMATION

Corresponding Author

Daniel Petras – Cluster of Excellence-Controlling Microbes to Fight Infection, University of Tübingen, Tübingen 72076, Germany; orcid.org/0000-0002-6561-3022;
Email: daniel.petras@uni-tuebingen.de

Authors

Paolo Stincone – Cluster of Excellence-Controlling Microbes to Fight Infection, University of Tübingen, Tübingen 72076, Germany; orcid.org/0000-0002-2214-6655

Abzer K. Pakkir Shah – Cluster of Excellence-Controlling Microbes to Fight Infection, University of Tübingen, Tübingen 72076, Germany

Robin Schmid – Institute of Organic Chemistry and Biochemistry of the Czech Academy of Sciences, 160 00 Praha 6, Czech Republic

Lana G. Graves – Faculty of Mathematics and Natural Sciences, Environmental Systems Analysis, University of Tübingen, Tübingen 72076, Germany; Leibniz Institute of Freshwater Ecology and Inland Fisheries (IGB), Berlin 12587, Germany

Stilianos P. Lambidis – Cluster of Excellence-Controlling Microbes to Fight Infection, University of Tübingen, Tübingen 72076, Germany

Ralph R. Torres – University of California San Diego, Scripps Institution of Oceanography, La Jolla, California 92093, United States

Shu-Ning Xia – Cluster of Excellence-Controlling Microbes to Fight Infection, University of Tübingen, Tübingen 72076, Germany; orcid.org/0000-0001-8317-9488

Vidit Minda – Department of Chemistry and Biochemistry, University of Denver, Denver, Colorado 80210, United States; Department of Pharmacology and Pharmaceutical Sciences, University of Missouri–Kansas City, Kansas City, Missouri 64108, United States

Allegra T. Aron – Department of Chemistry and Biochemistry, University of Denver, Denver, Colorado 80210, United States

Mingxun Wang – Department of Computer Science, University of California Riverside, Riverside, California 92507, United States; orcid.org/0000-0001-7647-6097

Chambers C. Hughes – Cluster of Excellence-Controlling Microbes to Fight Infection and Department of Microbial Bioactive Compounds, Interfaculty Institute for Microbiology and Infection Medicine, University of Tübingen, Tübingen 72076, Germany; German Center for Infection Research, Partner Site Tübingen, Tübingen 72076, Germany; orcid.org/0000-0002-7000-6438

Complete contact information is available at:
<https://pubs.acs.org/10.1021/acs.analchem.3c01202>

Author Contributions

P.S. and D.P. conceptualized and designed the study. P.S., A.K.P.S., L.G.G., S.P.L., S.X., C.C.H., R.R.T., and D.P. performed sampling and extractions. P.S. and D.P. performed the MS measurements. R.S. and M.W. contributed to software code. P.S. and A.K.P.S. performed the data analysis. P.S. and D.P. wrote the manuscript. R.S., L.G., R.R.T., V.M., A.T.A., M.W., and C.C.H. critically revised the manuscript. All the authors read, edited, and approved the final manuscript.

Notes

The authors declare the following competing financial interest(s): M.W. is a co-founder of Ometa Labs LLC.

ACKNOWLEDGMENTS

C.C.H. and D.P. were supported by the Deutsche Forschungsgemeinschaft through the CMFI Cluster of Excellence (EXC 2124) and the Collaborative Research Center CellMap (TRR 261). P.S. was supported by the European Union's Horizon 2020 research and innovation programme through a Marie Skłodowska-Curie fellowship n. 101108450-MeStaLeM. S.X. was supported by the Chinese Scholarship Council through the PhD scholarship n. 202008330294. We thank Libera Lo Presti for critical reading of the manuscript.

REFERENCES

- (1) Catalá, T. S.; Shorte, S.; Dittmar, T. *Appl. Microbiol. Biotechnol.* **2021**, *105*, 7225–7239.
- (2) Kido Soule, M. C.; Longnecker, K.; Johnson, W. M.; Kujawinski, E. B. *Mar. Chem.* **2015**, *177*, 374–387.
- (3) Lai, A.; Clark, A. M.; Escher, B. I.; Fernandez, M.; McEwen, L. R.; Tian, Z.; Wang, Z.; Schymanski, E. L. *Environ. Sci. Technol.* **2022**, *56*, 7448–7466.
- (4) Withers, E.; Hill, P. W.; Chadwick, D. R.; Jones, D. L. *Soil Biol. Biochem.* **2020**, *143*, No. 107758.
- (5) Eliuk, S.; Makarov, A. *Annu. Rev. Anal. Chem.* **2015**, *8*, 61–80.
- (6) Longnecker, K.; Kujawinski, E. B. *Org. Geochem.* **2017**, *110*, 92–99.
- (7) Lu, K.; Gardner, W. S.; Liu, Z. *Environ. Sci. Technol.* **2018**, *52*, 7182–7191.
- (8) Broeckling, C. D.; Hoyes, E.; Richardson, K.; Brown, J. M.; Prenni, J. E. *Anal. Chem.* **2018**, *90*, 8020–8027.
- (9) Guo, J.; Huan, T. *Anal. Chem.* **2020**, *92*, 8072–8080.
- (10) Chase, M. W.; National Information Standards Organization (US). *NIST-JANAF Thermochemical Tables*; American Chemical Society: Washington, DC, 1998; vol 9.
- (11) Horai, H.; Arita, M.; Kanaya, S.; Nihei, Y.; Ikeda, T.; Suwa, K.; Ojima, Y.; Tanaka, K.; Tanaka, S.; Aoshima, K.; Oda, Y.; Kakazu, Y.; Kusano, M.; Tohge, T.; Matsuda, F.; Sawada, Y.; Hirai, M. Y.; Nakanishi, H.; Ikeda, K.; Akimoto, N.; Maoka, T.; Takahashi, H.; Ara, T.; Sakurai, N.; Suzuki, H.; Shibata, D.; Neumann, S.; Iida, T.; Tanaka, K.; Funatsu, K.; Matsuura, F.; Soga, T.; Taguchi, R.; Saito, K.; Nishioka, T. *J. Mass Spectrom.* **2010**, *45*, 703–714.
- (12) Montenegro-Burke, J. R.; Guijas, C.; Siuzdak, G. METLIN: A Tandem Mass Spectral Library of Standards. In *Computational Methods and Data Analysis for Metabolomics*; Li, S., Ed.; Methods in Molecular Biology; Springer US: New York, NY, 2020; pp 149–163.
- (13) *mzCloud – Advanced Mass Spectral Database*. <https://www.mzcloud.org/>. (accessed 04 July, 2022).
- (14) Smith, C. A.; Maille, G. O.; Want, E. J.; Qin, C.; Trauger, S. A.; Brandon, T. R.; Custodio, D. E.; Abagyan, R.; Siuzdak, G. *Ther. Drug Monit.* **2005**, *27*, 747–751.
- (15) Wang, M.; Carver, J. J.; Phelan, V. V.; Sanchez, L. M.; Garg, N.; Peng, Y.; Nguyen, D. D.; Watrous, J.; Kaponov, C. A.; Luzzatto-Knaan, T.; Porto, C.; Bouslimani, A.; Melnik, A. V.; Meehan, M. J.; Liu, W.-T.; Crüsemann, M.; Boudreau, P. D.; Esquenazi, E.; Sandoval-Calderón, M.; Kersten, R. D.; Pace, L. A.; Quinn, R. A.; Duncan, K. R.; Hsu, C.-C.; Floros, D. J.; Gavilan, R. G.; Kleigrew, K.; Northen, T.; Dutton, R. J.; Parrot, D.; Carlson, E. E.; Aigle, B.; Michelsen, C. F.; Jelsbak, L.; Sohlenkamp, C.; Pevzner, P.; Edlund, A.; McLean, J.; Piel, J.; Murphy, B. T.; Gerwick, L.; Liaw, C.-C.; Yang, Y.-L.; Humpf, H.-U.; Maansson, M.; Keyzers, R. A.; Sims, A. C.; Johnson, A. R.; Sidebottom, A. M.; Sedio, B. E.; Klitgaard, A.; Larson, C. B.; Boya, P.; Torres-Mendoza, D.; Gonzalez, D. J.; Silva, D. B.; Marques, L. M.; Demarques, D. P.; Pociute, E.; O'Neill, E. C.; Briand, E.; Helfrich, E. J. N.; Granatosky, E. A.; Glukhov, E.; Ryffel, F.; Houson, H.; Mohimani, H.; Kharbush, J. J.; Zeng, Y.; Vorholt, J. A.; Kurita, K. L.; Charusanti, P.; McPhail, K. L.; Nielsen, K. F.; Vuong, L.; Elfeki, M.; Traxler, M. F.; Engene, N.; Koyama, N.; Vining, O. B.; Baric, R.; Silva, R. R.; Mascuch, S. J.; Tomasi, S.; Jenkins, S.; Macherla, V.; Hoffman, T.; Agarwal, V.; Williams, P. G.; Dai, J.; Neupane, R.; Gurr, J.; Rodríguez, A. M. C.; Lamsa, A.; Zhang, C.; Dorrestein, K.; Duggan, B. M.; Almaliti, J.; Allard, P.-M.; Phapale, P.; Nothias, L.-F.; Alexandrov, T.; Litaudon, M.; Wolfender, J.-L.; Kyle, J. E.; Metz, T. O.; Peryea, T.; Nguyen, D.-T.; VanLeer, D.; Shinn, P.; Jadhav, A.; Müller, R.; Waters, K. M.; Shi, W.; Liu, X.; Zhang, L.; Knight, R.; Jensen, P. R.; Palsson, B. Ø.; Pogliano, K.; Linington, R. G.; Gutiérrez, M.; Lopes, N. P.; Gerwick, W. H.; Moore, B. S.; Dorrestein, P. C.; Bandeira, N. *Nat. Biotechnol.* **2016**, *34*, 828–837.
- (16) Sumner, L. W.; Amberg, A.; Barrett, D.; Beale, M. H.; Beger, R.; Daykin, C. A.; Fan, T. W.-M.; 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.; Reilly, M. D.; Thaden, J. J.; Viant, M. R. *Metabolomics* **2007**, *3*, 211–221.
- (17) Bittremieux, W.; Avalon, N. E.; Thomas, S. P.; Kakhkhorov, S. A.; Aksenov, A. A.; Gomes, P. W. P.; Aceves, C. M.; Rodríguez, A. M. C.; Gauglitz, J. M.; Gerwick, W. H.; Jarmusch, A. K.; Kaddurah-Daouk, R. F.; Kang, K. B.; Kim, H. W.; Kondic, T.; Mannochio-Russo, H.; Meehan, M. J.; Melnik, A. V.; Nothias, L.-F.; O'Donovan, C.; Panitchpakdi, M.; Petras, D.; Schmid, R.; Schymanski, E. L.; Van Der Hooft, J. J. J.; Weldon, K. C.; Yang, H.; Zemlin, J.; Wang, M.; Dorrestein, P. C. *bioRxiv* **2022**, No. 2022.05.15.490691.
- (18) Dührkop, K.; Shen, H.; Meusel, M.; Rousu, J.; Böcker, S. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112*, 12580–12585.
- (19) Dührkop, K.; Fleischauer, M.; Ludwig, M.; Aksenov, A. A.; Melnik, A. V.; Meusel, M.; Dorrestein, P. C.; Rousu, J.; Böcker, S. *Nat. Methods* **2019**, *16*, 299–302.
- (20) Dührkop, K.; Nothias, L.-F.; Fleischauer, M.; Reher, R.; Ludwig, M.; Hoffmann, M. A.; Petras, D.; Gerwick, W. H.; Rousu, J.; Dorrestein, P. C.; Böcker, S. *Nat. Biotechnol.* **2021**, *39*, 462–471.
- (21) Giné, R.; Capellades, J.; Badia, J. M.; Vughs, D.; Schwaiger-Haber, M.; Alexandrov, T.; Vinaixa, M.; Brunner, A. M.; Patti, G. J.; Yanes, O. *Nat. Methods* **2021**, *18*, 1370–1376.
- (22) Jarmusch, A. K.; Wang, M.; Aceves, C. M.; Advani, R. S.; Aguirre, S.; Aksenov, A. A.; Aleti, G.; Aron, A. T.; Bauermeister, A.; Bolleddu, S.; Bouslimani, A.; Caraballo Rodriguez, A. M.; Chaar, R.; Coras, R.; Elijah, E. O.; Ernst, M.; Gauglitz, J. M.; Gentry, E. C.; Husband, M.; Jarmusch, S. A.; Jones, K. L.; Kamenik, Z.; Le Gouellec, A.; Lu, A.; McCall, L.-I.; McPhail, K. L.; Meehan, M. J.; Melnik, A. V.; Menezes, R. C.; Montoya Giraldo, Y. A.; Nguyen, N. H.; Nothias, L. F.; Nothias-Esposito, M.; Panitchpakdi, M.; Petras, D.; Quinn, R. A.; Sikora, N.; van der Hooft, J. J. J.; Vargas, F.; Vrbanc, A.; Weldon, K. C.; Knight, R.; Bandeira, N.; Dorrestein, P. C. *Nat. Methods* **2020**, *17*, 901–904.
- (23) Li, Y.; Kind, T.; Folz, J.; Vaniya, A.; Mehta, S. S.; Fiehn, O. *Nat. Methods* **2021**, *18*, 1524–1531.
- (24) Ludwig, M.; Nothias, L.-F.; Dührkop, K.; Koester, I.; Fleischauer, M.; Hoffmann, M. A.; Petras, D.; Vargas, F.; Morsy, M.; Aluwihare, L.; Dorrestein, P. C.; Böcker, S. *Nat. Mach. Intell.* **2020**, *2*, 629–641.
- (25) Nothias, L.-F.; Petras, D.; Schmid, R.; Dührkop, K.; Rainer, J.; Sarvepalli, A.; Protsyuk, I.; Ernst, M.; Tsugawa, H.; Fleischauer, M.; Aicheler, F.; Aksenov, A. A.; Alka, O.; Allard, P.-M.; Barsch, A.; Cachet, X.; Caraballo-Rodríguez, A. M.; Da Silva, R. R.; Dang, T.; Garg, N.; Gauglitz, J. M.; Gurevich, A.; Isaac, G.; Jarmusch, A. K.; Kamenik, Z.; Kang, K. B.; Kessler, N.; Koester, I.; Korf, A.; Le Gouellec, A.; Ludwig, M.; Martin, C. H.; McCall, L.-I.; McSayles, J.; Meyer, S. W.; Mohimani, H.; Morsy, M.; Moyné, O.; Neumann, S.; Neuweger, H.; Nguyen, N. H.; Nothias-Esposito, M.; Paolini, J.; Phelan, V. V.; Pluskal, T.; Quinn, R. A.; Rogers, S.; Shrestha, B.; Tripathi, A.; Van Der Hooft, J. J. J.; Vargas, F.; Weldon, K. C.; Witting, M.; Yang, H.; Zhang, Z.; Zubeil, F.; Kohlbacher, O.; Böcker, S.; Alexandrov, T.; Bandeira, N.; Wang, M.; Dorrestein, P. C. *Nat. Methods* **2020**, *17*, 905–908.
- (26) Petras, D.; Phelan, V. V.; Acharya, D.; Allen, A. E.; Aron, A. T.; Bandeira, N.; Bowen, B. P.; Belle-Oudry, D.; Boecker, S.; Cummings, D. A.; Deutsch, J. M.; Fahy, E.; Garg, N.; Gregor, R.; Handelsman, J.; Navarro-Hoyos, M.; Jarmusch, A. K.; Jarmusch, S. A.; Louie, K.; Maloney, K. N.; Marty, M. T.; Meijler, M. M.; Mizrahi, I.; Neve, R. L.; Northen, T. R.; Molina-Santiago, C.; Panitchpakdi, M.; Pullman, B.; Puri, A. W.; Schmid, R.; Subramaniam, S.; Thukral, M.; Vasquez-Castro, F.; Dorrestein, P. C.; Wang, M. *Nat. Methods* **2022**, *19*, 134–136.

- (27) Pluskal, T.; Castillo, S.; Villar-Briones, A.; Oresic, M. *BMC Bioinf.* **2010**, *11*, 395.
- (28) Ruttikies, C.; Schymanski, E. L.; Wolf, S.; Hollender, J.; Neumann, S. *J. Cheminform.* **2016**, *8*, 3.
- (29) Schmid, R.; Petras, D.; Nothias, L.-F.; Wang, M.; Aron, A. T.; Jagels, A.; Tsugawa, H.; Rainer, J.; Garcia-Aloy, M.; Dührkop, K.; Korf, A.; Pluskal, T.; Kamenik, Z.; Jarmusch, A. K.; Caraballo-Rodríguez, A. M.; Weldon, K. C.; Nothias-Esposito, M.; Aksenov, A. A.; Bauermeister, A.; Albarracín Orió, A.; Grundmann, C. O.; Vargas, F.; Koester, I.; Gauglitz, J. M.; Gentry, E. C.; Hövelmann, Y.; Kalinina, S. A.; Pendergraft, M. A.; Panitchpakdi, M.; Tehan, R.; Le Gouellec, A.; Aleti, G.; Mannochio Russo, H.; Arndt, B.; Hübner, F.; Hayen, H.; Zhi, H.; Raffatellu, M.; Prather, K. A.; Aluwihare, L. I.; Böcker, S.; McPhail, K. L.; Humpf, H.-U.; Karst, U.; Dorrestein, P. C. *Nat. Commun.* **2021**, *12*, 3832.
- (30) Smith, C. A.; Want, E. J.; O'Maille, G.; Abagyan, R.; Siuzdak, G. *Anal. Chem.* **2006**, *78*, 779–787.
- (31) Sturm, M.; Bertsch, A.; Gröpl, C.; Hildebrandt, A.; Hussong, R.; Lange, E.; Pfeifer, N.; Schulz-Trieglaff, O.; Zerck, A.; Reinert, K.; Kohlbacher, O. *BMC Bioinf.* **2008**, *9*, 163.
- (32) Tautenhahn, R.; Patti, G. J.; Rinehart, D.; Siuzdak, G. *Anal. Chem.* **2012**, *84*, 5035–5039.
- (33) Tsugawa, H.; Cajka, T.; Kind, T.; Ma, Y.; Higgins, B.; Ikeda, K.; Kanazawa, M.; VanderGheynst, J.; Fiehn, O.; Arita, M. *Nat. Methods* **2015**, *12*, 523–526.
- (34) Wandy, J.; Zhu, Y.; van der Hooft, J. J. J.; Daly, R.; Barrett, M. P.; Rogers, S. *Bioinformatics* **2018**, *34*, 317–318.
- (35) Wang, M.; Jarmusch, A. K.; Vargas, F.; Aksenov, A. A.; Gauglitz, J. M.; Weldon, K.; Petras, D.; da Silva, R.; Quinn, R.; Melnik, A. V.; van der Hooft, J. J. J.; Caraballo-Rodríguez, A. M.; Nothias, L. F.; Aceves, C. M.; Panitchpakdi, M.; Brown, E.; Di Ottavio, F.; Sikora, N.; Elijah, E. O.; Labarta-Bajo, L.; Gentry, E. C.; Shalpour, S.; Kyle, K. E.; Puckett, S. P.; Watrous, J. D.; Carpenter, C. S.; Bouslimani, A.; Ernst, M.; Swafford, A. D.; Zúñiga, E. I.; Balunas, M. J.; Klassen, J. L.; Loomba, R.; Knight, R.; Bandeira, N.; Dorrestein, P. C. *Nat. Biotechnol.* **2020**, *38*, 23–26.
- (36) Crocker, D. R.; Kaluarachchi, C. P.; Cao, R.; Dinasquet, J.; Franklin, E. B.; Morris, C. K.; Amiri, S.; Petras, D.; Nguyen, T.; Torres, R. R.; Martz, T. R.; Malfatti, F.; Goldstein, A. H.; Tivanski, A. V.; Prather, K. A.; Thiemens, M. H. *Environ. Sci. Technol.* **2022**, *56*, 9947.
- (37) Eysseric, E.; Beaudry, F.; Gagnon, C.; Segura, P. A. *Talanta* **2021**, *230*, No. 122293.
- (38) Gamba, A.; Petras, D.; Little, M.; White, B.; Dorrestein, P. C.; Rohwer, F.; Foster, R. A.; Hartmann, A. C. *Front. Mar. Sci.* **2022**, *9*, No. 853554.
- (39) García, S. L.; Nuy, J. K.; Mehrshad, M.; Hampel, J. J.; Sedano-Nuñez, V. T.; Buck, M.; Divne, A.-M.; Lindström, E. S.; Petras, D.; Hawkes, J.; Bertilsson, S. *bioRxiv* **2022**, No. 2022.03.21.485019.
- (40) Molina-Santiago, C.; Vela-Corcía, D.; Petras, D.; Díaz-Martínez, L.; Pérez-Lorente, A. I.; Sopena-Torres, S.; Pearson, J.; Caraballo-Rodríguez, A. M.; Dorrestein, P. C.; de Vicente, A.; Romero, D. *Cell Rep.* **2021**, *36*, No. 109449.
- (41) Oberleitner, D.; Schmid, R.; Schulz, W.; Bergmann, A.; Achten, C. *Anal. Bioanal. Chem.* **2021**, *413*, 5291–5300.
- (42) Tong, Y.; Wang, P.; Sun, J.; Li, X.; Wang, T.; Zhou, Q.; Xie, Z.; Jiang, C.; Wang, J. *Ind. Crops Prod.* **2021**, *174*, No. 114169.
- (43) Wegley Kelly, L.; Nelson, C. E.; Petras, D.; Koester, I.; Quinlan, Z. A.; Arts, M. G. I.; Nothias, L.-F.; Comstock, J.; White, B. M.; Hopmans, E. C.; van Duyl, F. C.; Carlson, C. A.; Aluwihare, L. I.; Dorrestein, P. C.; Haas, A. F. *Proc. Natl. Acad. Sci. U. S. A.* **2022**, *119*, No. e2110283119.
- (44) Chen, L.; Lu, W.; Wang, L.; Xing, X.; Chen, Z.; Teng, X.; Zeng, X.; Muscarella, A. D.; Shen, Y.; Cowan, A.; McReynolds, M. R.; Kennedy, B. J.; Lato, A. M.; Campagna, S. R.; Singh, M.; Rabinowitz, J. D. *Nat. Methods* **2021**, 1377–1385.
- (45) Defossez, E.; Bourquin, J.; von Reuss, S.; Rasmann, S.; Glauser, G. *Mass Spectrom. Rev.* **2023**, *42*, 131–143.
- (46) Davies, V.; Wandy, J.; Weidt, S.; van der Hooft, J. J. J.; Miller, A.; Daly, R.; Rogers, S. *Anal. Chem.* **2021**, *93*, 5676–5683.
- (47) Xu, R.; Lee, J.; Chen, L.; Zhu, J. *Mol. Omics* **2021**, *17*, 665–676.
- (48) Zhang, C.; Liu, M.; Xu, X.; Wu, J.; Li, X.; Wang, H.; Gao, X.; Guo, D.; Tian, X.; Yang, W. *J. Agric. Food Chem.* **2022**, *70*, 5932–5944.
- (49) Afoullouss, S.; Balsam, A.; Allcock, A. L.; Thomas, O. P. *Metabolites* **2022**, *12*, 245.
- (50) Ramabulana, A.-T.; Petras, D.; Madala, N. E.; Tugizimana, F. *Metabolites* **2021**, *11*, 763.
- (51) Bahureksa, W.; Borch, T.; Young, R. B.; Weisbrod, C.; Blakney, G. T.; McKenna, A. M. *Anal. Chem.* **2022**, *94*, 11382–11389.
- (52) Schmid, R.; Heuckeroth, S.; Korf, A.; Smirnov, A.; Myers, O.; Dyrlund, T. S.; Bushuiev, R.; Murray, K. J.; Hoffmann, N.; Lu, M.; Sarvepalli, A.; Zhang, Z.; Fleischauer, M.; Dührkop, K.; Wesner, M.; Hoogstra, S. J.; Rudt, E.; Mokshyna, O.; Brungs, C.; Ponomarov, K.; Mutabdzija, L.; Damiani, T.; Pudney, C. J.; Earll, M.; Helmer, P. O.; Fallon, T. R.; Schulze, T.; Rivas-Ubach, A.; Bilbao, A.; Richter, H.; Nothias, L.-F.; Wang, M.; Orešič, M.; Weng, J.-K.; Böcker, S.; Jeibmann, A.; Hayen, H.; Karst, U.; Dorrestein, P. C.; Petras, D.; Du, X.; Pluskal, T. *Nat. Biotechnol.* **2023**, 447–449.
- (53) Petras, D.; Minich, J. J.; Cancelada, L. B.; Torres, R. R.; Kunselman, E.; Wang, M.; White, M. E.; Allen, E. E.; Prather, K. A.; Aluwihare, L. I.; Dorrestein, P. C. *Chemosphere* **2021**, *271*, No. 129450.
- (54) Cancelada, L.; Torres, R. R.; Garrafa Luna, J.; Dorrestein, P. C.; Aluwihare, L. I.; Prather, K. A.; Petras, D. *Limnol. Oceanogr. Methods* **2022**, *20*, 89–101.
- (55) Petras, D.; Koester, I.; Da Silva, R.; Stephens, B. M.; Haas, A. F.; Nelson, C. E.; Kelly, L. W.; Aluwihare, L. I.; Dorrestein, P. C. *Front. Mar. Sci.* **2017**, *4*, 405.
- (56) Adusumilli, R.; Mallick, P. *Methods Mol. Biol.* **2017**, *1550*, 339–368.
- (57) Bolyen, E.; Rideout, J. R.; Dillon, M. R.; Bokulich, N. A.; Abnet, C. C.; Al-Ghalith, G. A.; Alexander, H.; Alm, E. J.; Arumugam, M.; Asnicar, F.; Bai, Y.; Bisanz, J. E.; Bittinger, K.; Brejnrod, A.; Brislawn, C. J.; Brown, C. T.; Callahan, B. J.; Caraballo-Rodríguez, A. M.; Chase, J.; Cope, E. K.; Da Silva, R.; Diener, C.; Dorrestein, P. C.; Douglas, G. M.; Durall, D. M.; Duvallet, C.; Edwardson, C. F.; Ernst, M.; Estaki, M.; Fouquier, J.; Gauglitz, J. M.; Gibbons, S. M.; Gibson, D. L.; Gonzalez, A.; Gorlick, K.; Guo, J.; Hillmann, B.; Holmes, S.; Holste, H.; Huttenhower, C.; Huttley, G. A.; Janssen, S.; Jarmusch, A. K.; Jiang, L.; Kaehler, B. D.; Kang, K. B.; Keefe, C. R.; Keim, P.; Kelley, S. T.; Knights, D.; Koester, I.; Kosciolk, T.; Kreps, J.; Langille, M. G. I.; Lee, J.; Ley, R.; Liu, Y.-X.; Loftfield, E.; Lozupone, C.; Maher, M.; Marotz, C.; Martin, B. D.; McDonald, D.; McIver, L. J.; Melnik, A. V.; Metcalf, J. L.; Morgan, S. C.; Morton, J. T.; Naimey, A. T.; Navas-Molina, J. A.; Nothias, L. F.; Orchanian, S. B.; Pearson, T.; Peoples, S. L.; Petras, D.; Preuss, M. L.; Pruesse, E.; Rasmussen, L. B.; Rivers, A.; Robeson, M. S.; Rosenthal, P.; Segata, N.; Shaffer, M.; Shiffer, A.; Sinha, R.; Song, S. J.; Spear, J. R.; Swafford, A. D.; Thompson, L. R.; Torres, P. J.; Trinh, P.; Tripathi, A.; Turnbaugh, P. J.; Ul-Hasan, S.; van der Hooft, J. J. J.; Vargas, F.; Vázquez-Baeza, Y.; Vogtmann, E.; von Hippel, M.; Walters, W.; Wan, Y.; Wang, M.; Warren, J.; Weber, K. C.; Williamson, C. H. D.; Willis, A. D.; Xu, Z. Z.; Zaneveld, J. R.; Zhang, Y.; Zhu, Q.; Knight, R.; Caporaso, J. G. *Nat. Biotechnol.* **2019**, *37*, 852–857.
- (58) Shannon, P.; Markiel, A.; Ozier, O.; Baliga, N. S.; Wang, J. T.; Ramage, D.; Amin, N.; Schwikowski, B.; Ideker, T. *Genome Res.* **2003**, *13*, 2498–2504.
- (59) Khan, A.; Mathelier, A. *BMC Bioinf.* **2017**, *18*, 287.