JOURNAL OF AGRICULTURAL AND FOOD CHEMISTRY

Open Access

pubs.acs.org/JAFC

Article

Automated Composition Assessment of Natural Extracts: Untargeted Mass Spectrometry-Based Metabolite Profiling Integrating Semiquantitative Detection

Adriano Rutz* and Jean-Luc Wolfender*

Cite This: J. Agric. Food Chem. 2023, 71, 18010−18023

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Recent developments in mass spectrometry-based metabolite profiling allow unprecedented qualitative coverage of complex biological extract composition. However, the electrospray ionization used in metabolite profiling generates multiple artifactual signals for a single analyte. This leads to thousands of signals per analysis without satisfactory means of filtering those corresponding to abundant constituents. Generic approaches are therefore needed for the qualitative and quantitative annotation of a broad range of relevant constituents. For this, we used an analytical platform combining liquid chromatography−mass spectrometry (LC−MS) with Charged Aerosol Detection (CAD). We established a generic metabolite profiling for the concomitant recording of qualitative MS data and semiquantitative CAD profiles. The MS features (recorded in high-resolution tandem MS) are grouped and annotated using state-of-the-art tools. To efficiently attribute features to their corresponding extracted and integrated CAD peaks, a custom signal pretreatment and peak-shape comparison workflow is built. This strategy allows us to automatically contextualize features at both major and minor metabolome levels, together with a detailed reporting of their annotation including relevant orthogonal information (taxonomy, retention time). Signals not attributed to CAD peaks are considered minor metabolites. Results are illustrated on an ethanolic extract of *Swertia chirayita* (Roxb.) H. Karst., a bitter plant of industrial interest, exhibiting the typical complexity of plant extracts as a proof of concept. This generic qualitative and quantitative approach paves the way to automatically assess the composition of single natural extracts of interest or broader collections, thus facilitating new ingredient registrations or natural-extracts-based drug discovery campaigns.

KEYWORDS: *metabolite profiling, Charged Aerosol Detection, automated composition assessment, liquid chromatography−mass spectrometry, natural extract*

INTRODUCTION

Recent developments in metabolomics and untargeted liquid chromatography−mass spectrometry (LC−MS) analyses allow unprecedented qualitative coverage.[1] Automation and throughput of analyses have also been significantly improved.[2] While these developments are crucial for better qualitative characterization of any sample, quantification remains limited to a few compounds.[3] Accurate quantification is a tedious process that requires high-quality standards, making it unavailable on a large scale of compounds.[4] Furthermore, although highly sensitive, LC−MS is not a universal detection technique.[5] Given the sensitivity achieved by modern MS instruments, only a portion of the acquired signals is relevant for the compositional assessment.[6] Indeed, out of all the signals considered in an LC−MS analysis, the majority are chemical adducts or fragments created during the ionization process and a single "analyte" can lead to dozens of features.[7] This is true, especially for LC−ESI−MS, where some compounds can exhibit very large differences in their ionization behavior.[8] This is a major drawback of current untargeted LC−MS approaches, as a large fraction of the annotated signals cannot be linked to any generic semiquantitative information. This is even worse for very rich matrices such as plant extracts, as a single analysis can suggest hundreds or thousands of features,[9] without satisfactory means of filtering out the analytes that are significant, or that represent a substantial amount of the extracted mass. Despite many advantages in terms of sensitivity and metabolome coverage, untargeted LC−MS analyses are not yet adopted widely for natural extract compositional assessment due to the complexity of the generated data. It is, therefore, crucial to develop methodologies that provide semiquantitative information on a wide range of signals and that estimate annotation confidence.

Confident Annotation of Multiple Features. The introduction of Feature-based Molecular Networking was a major step in deciphering the huge amount of data generated by MS-based metabolomics approaches.[10] While spectral matching of features to spectral libraries had already existed for a long time,[11] calculating the spectral similarity between features to organize and visualize them into spectral families has greatly facilitated the chemical study of complex samples.[12] Further improvements have been made recently by Ion Identity

Received: May 12, 2023
Revised: September 19, 2023
Accepted: September 22, 2023
Published: November 10, 2023

Networking,[13] and by linking positive ionization (PI) and negative ionization (NI) modes.[14] Lately, Sirius[15] has integrated several subtools into its ecosystem. The first, CANOPUS, predicts compound classes from mass spectra.[1] Originally designed using the ClassyFire chemical taxonomy,[16] it now calculates results based on the NPClassifier taxonomy, which is better suited for natural products (NPs).[17] Since it is easier to predict a chemical class than a specific structure, the results of the chemical class annotation are more accurate than the structure annotation. However, until recently, there was no way to estimate the confidence in the annotation. For this, a Support Vector Machine approach was developed to estimate the confidence of metabolites absent from spectral libraries.[18] Performance was then improved using deep neural networks.[19] Another complementary taxonomy-based confidence scoring method was developed in-house.[20] Together, these approaches can facilitate the filtering of highly trusted annotations. Such annotations, even if at a lower confidence level than internal libraries of standards,[21] can be used as additional anchor points before further investigations.

Generic Quantitation of Multiple Components. In MS, simultaneous quantitation on multiple (classes of) chemicals remains challenging. Some approaches have attempted to include optimally selected internal standards from various chemical families to extrapolate the ionization of these compound classes.[22] These approaches remain study-specific and require the use of standards without completely overcoming the intrinsic limitations of compound ionization specificity. Therefore, the use of hyphenation with other types of detectors has emerged as a way to partly overcome these limitations and capture additional complementary information simultaneously. One way to improve the analysis was by combining LC−MS platforms with Photo Diode Array (PDA) detectors.[23] This allowed for the recording of ultraviolet (UV) spectra and UV detectors are now routinely used coupled to LC. More recently, "universal" detectors like the Evaporative Light Scattering Detector (ELSD)[24] and the Charged Aerosol Detector (CAD)[25] have been introduced. These detectors can detect compounds that do not have chromophores and ideally provide a response that is not specific to a particular class of compounds. The CAD, introduced in 2005, has shown multiple advantages in comparison to ELSD.[26] The CAD limit of detection ranges between 0.002[27] and 0.5 μg/mL,[28] depending on the studies, with most values reported around 0.05 μg/mL.[29,30] In most articles, the limits of quantitation were approximately twice the value of the limit of detection. As the aim of this study is to remain as generic as possible, no compound-specific limit of detection will be determined.

Quantitative detectors have been used in gas chromatography (GC) for many years, in GC-Flame Ionization Detector (FID)-MS platforms.[31] However, GC-FID is limited to the analysis of volatile substances. All three detectors (CAD, ELSD, and FID) mentioned above are one-dimensional detectors. They provide intensity measurements over time but no additional spectral dimensions, such as PDA or MS detection. As a result, achieving a high chromatographic resolution is necessary for these detectors, since the data cannot be deconvoluted using the third spectroscopic dimension. Therefore, ideally, all peaks should be separated well from each other. GC inherently offers higher chromatographic resolution than HPLC, making such approximations possible in GC-FID. In LC, achieving very high chromatographic resolution is still a challenge, but current UHPLC platforms have made significant progress in increasing peak capacity.[32]

Combining Untargeted Annotation and Generic Quantitation. Taking advantage of the technological and methodological advancements mentioned above, this work aims to make sense of the abundant qualitative information from current untargeted LC−MS analysis pipelines and link it to universal semiquantitative information, as well as other complementary metadata, such as taxonomy. Such developments should facilitate an automated comprehensive assessment of the natural extract composition. Including the semiquantitative dimension early in the dereplication process can also facilitate the isolation of the compounds for further structural elucidation.[33] The objective is to generate metabolite profiling data from natural extracts, categorizing them based on semiquantitative information that distinguishes between major and minor metabolome. Additionally, by grouping related features, the extensive MS data are condensed into meaningful information for each LC peak. While the hyphenation of both CAD and MS detectors was illustrated by Baker and Regg for botanical constituents determination in 2018, they mostly focused on the analytical aspects.[34]

Our approach aims to simplify the interpretation of the data. Our workflow and reporting mainly prepare the data for final inspection and further targeted confirmation, which is expected to be particularly useful for the registration of botanical ingredients and for the efficient isolation of new NPs. We particularly stress that our tool is only a guide to optimize further experiments required depending on the application (e.g., in applications necessitating accurate quantitation of given compounds to determine the instrumentation-specific limit of detection).

RESULTS AND DISCUSSION

The process leading to the qualitative and semiquantitative compositional assessment is briefly described below. As a minimal input, the workflow requires LC−MS/MS data, supplemented by another detector (here, CAD, as described in Data Acquisition). It then needs to be converted to an open format such as mzML, with all detectors encoded into the same file (Data Conversion). Multiple processing steps then follow and will be detailed in the next sections. First, signals are aligned as described in Signals' Pretreatment. Difficult to integrate properly at first, CAD required additional processing, as described in CAD Processing. After several signal processing steps described below, a qualitative and semiquantitative report is generated, as well as visual means to interrogate the data. In parallel to the above-described semiquantitative processing, MS features are extracted, clustered, and annotated, as described in MS/MS Processing. Afterward, MS features are attributed to CAD peaks as described in the Link Between CAD and MS Data. This allows signals to be categorized as belonging to "major" (linked to CAD) or "minor" metabolites. Finally, to facilitate further investigations, different automated numerical and visual reports are generated as described in Data Visualization. Although the presented method can be adapted to any type of extract, all of the results are illustrated on an ethanolic extract of the aerial parts of *S. chirayita* (Roxb.) H. Karst. (Gentianaceae), presenting a typical compositional complexity.

Data Acquisition. As previously mentioned, one strong requirement for efficient CAD integration is good chromatographic resolution. To achieve the best possible peak capacity

https://doi.org/10.1021/acs.jafc.3c03099
J. Agric. Food Chem. 2023, 71, 18010−18023

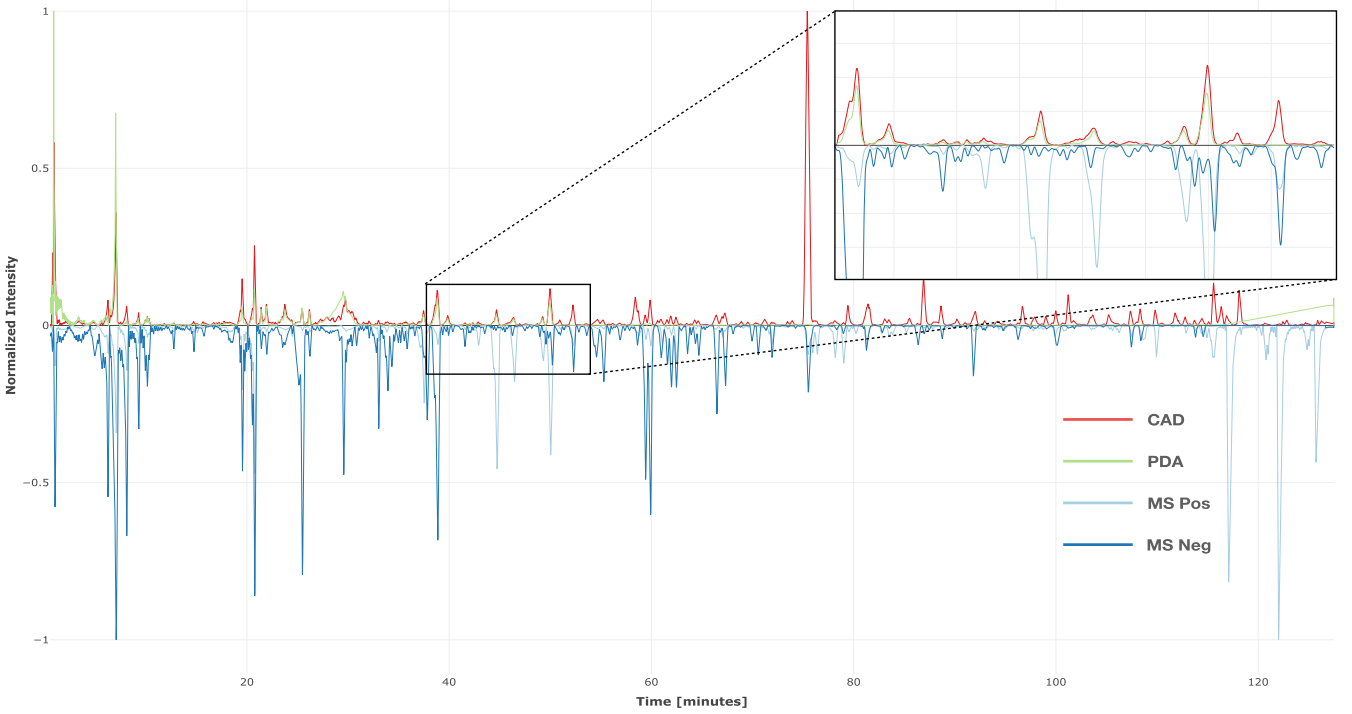

Figure 1. Aligned chromatograms after preprocessing. All signals were preprocessed to improve the downstream steps. CAD and PDA signals are at the top in red and green, respectively. MS signals (positive and negative ionization modes) are at the bottom in light and dark blue, respectively. A zoomed-in area showing the different intensities normalized to the observed signal is also illustrated.

while maintaining generic conditions, multiple trials were performed, with various representative plant extracts, columns, and gradient lengths (data not shown). For the example illustrated below, UHPLC runs of 126 min on a 150 mm column were performed. This corresponded to a peak capacity of 480 at 400 g/mol and 40 °C (for comparison, a run of 7 min on a 50 mm column corresponded to a peak capacity of 199, calculated with HPLC Calculator version 3.1).[35] The details of the data acquisition method are available in Data Acquisition.

Data Processing. Specific parsers were necessary for reading the CAD signal from the RAW file and translating it into the mzML file. The details are available in the section Data Conversion. Then, the signals of the different detectors were aligned. The instrumentation used consisted of a UHPLC connected to a PDA, a CAD, and an orbitrap MS. The nondestructive PDA was placed before an analytical split, and both CAD and MS detectors after. The MS signal was delayed by 0.090 and 0.055 min in comparison to the one of the PDA and the CAD, respectively. These values were used to align the signals as shown in Figure 1.

Figure 1 highlights the complementarity between each signal obtained. While the PDA intensity followed approximately the CAD intensity when a signal was present, some major molecules without a chromophore would have been ignored if PDA was used solely. The same complementarity could be observed between MS ionization modes (see light and dark blue traces). While each CAD major peak had its corresponding peak in MS, the opposite did not hold true, as illustrated by the positive ionization mode at the end of the chromatogram. In order to optimize coverage, it is strongly recommended to use both ionization modes, although the presented approach can also be applied to a single ionization mode.

Charged Aerosol Detector Data. As the CAD acquisition was set to capture small changes in the signal, it required some preprocessing steps before peak picking. These are briefly described hereafter and in more detail in the related Experimental Section. First, a Fourier transformation was performed to reduce noise as in ref 36. Moreover, the chromatographic resolution was improved using the derivative enhancement approach as described in ref 37. This resulted in a sharper signal, where peak detection could be performed with better results than those on the raw signal. To ensure the same processing between the different detectors, we also performed this on the PDA and BPI chromatograms.

As the CAD signal contained only one dimension, peaks were integrated by using simple algorithms. More elaborate techniques exist for multidimensional signals, such as the ones used in MS and PDA peak detection.[13,38−41] A comparison of the peaks detected automatically before and after signal processing is shown in the Supporting Information. This allowed for a good CAD peak-shape comparison with MS features.

High-Resolution Tandem Mass Spectrometry Data. MS features were extracted using MZmine3 v.3.4.0.[42] They were then annotated using the TIMA v.2.8.0,[20,43] encompassing SIRIUS[15,18,44−46] and ISDB annotations.[47] Such an approach allows one to use the strengths of each tool and take advantage of their complementarity. In PI, this led to 3610 features, of which 3254 were annotated, and 1272 confidently. In NI, this led to 2312 features, of which 2003 were annotated, and 637 confidently. All relevant details can be found in the corresponding Experimental Sections.

Link Between CAD and MS Data. CAD and MS peak shapes had to be compared to select the features most likely to correspond to a given semiquantitative signal. As a first mandatory step, all points of CAD and MS peaks were inter- and extrapolated to the same frequency to be defined with approximately the same number of points. This is generally not

https://doi.org/10.1021/acs.jafc.3c03099
J. Agric. Food Chem. 2023, 71, 18010−18023

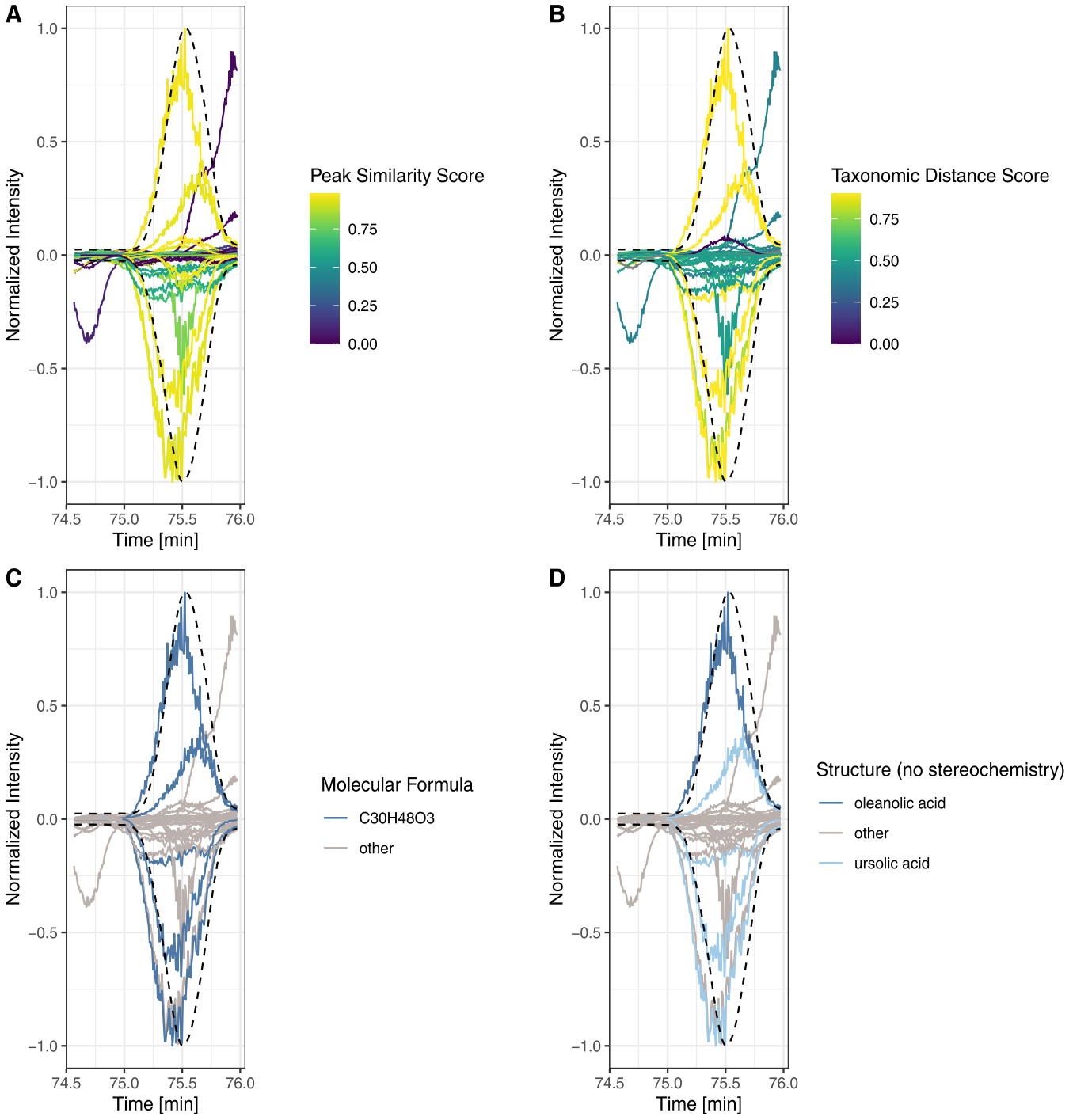

Figure 2. Link between MS features and CAD peak. Each panel is composed of positive and negative ionization modes (negative ionization modes are shown as negative values). The MS features are represented with plain lines and CAD peaks with dashed lines. In panel A, features are colored according to their peak-shape similarity with the CAD peak. In panel B, features are colored according to the `taxonomic distance score` part of the *taxonomically informed scoring*[20] (a value close to one indicates that the structure (without stereochemistry) has already been reported in a similar organism). In panel C, features are colored according to the MF of the best candidate. In panel D, features are colored according to the structure of the best candidate. Overall, this led to features that are more likely to be responsible for the CAD peak observed.

needed, as peak shapes usually compared are coming from the same detector, resulting in a unique frequency of acquisition. Here, resampling was performed since the frequency of the CAD was approximately 30 times higher than the one of MS. Both chromatograms were resampled to a frequency of 2 Hz, which approximated the one of the MS1.

The second step consisted of normalizing the peak's intensity together with a retention time normalization between the two peak's minima (illustrated in green in Figure S1). The aim of this procedure was to improve peak-shape similarity between the compounds whose MS intensity strongly differs from that of the CAD trace. Without this step, the MS features with the highest intensity would by default have a more defined peak shape and, thus, be more likely to be correlated to the CAD signal. Only features with a similarity ≥0.8 were considered, with the possibility to adapt the threshold.

https://doi.org/10.1021/acs.jafc.3c03099
J. Agric. Food Chem. 2023, 71, 18010−18023

If an MS feature had a correlation ≥0.8, its corresponding annotation was considered as a "major" compound, else, it was categorized as a "minor" compound. This allowed for the distinction of major and minor features even if they were found below the same CAD peak. All features outside of a CAD peak were considered "minor" by default. The higher the peak capacity,[48] the better the discrimination between minors and majors.

CAD peaks could envelop multiple or only a few (non)peak shape correlated features. An example of the most intense CAD peak of the *S. chirayita* extract (peak at 75 min in Figure 1) containing multiple MS features in both PI and NI modes is illustrated in Figure 2. Each panel is composed of PI and NI modes (NI mode is shown as negative values). The normalized extracted ion chromatograms of the MS features are represented with plain lines and CAD peaks with dashed lines; the colors of the lines depend on the factor illustrated.

First, in panel A, features were colored according to their peak-shape similarity to the CAD peak. Overall, more than 128 (67 PI and 61 NI) features were found below this CAD peak. Among the features sharing a good peak-shape similarity (≥0.8, in yellow), features with small intensity were found, together with a feature whose apex was slightly shifted in retention time. Features not sharing good peak-shape similarity were clearly distinguished. In panel B, features were colored according to the `taxonomic distance score` part of the *taxonomically informed scoring* (see Experimental Section).[20] For example, a score of 0.9 indicates that the structure has already been reported in the same species as that studied (here, *S. chirayita*). Both chemical structures corresponding to oleanolic acid and ursolic acid without stereochemistry corresponded to structures already reported in *S. chirayita*.[49] In panel C, features were colored according to the molecular formula (MF) of the best candidate. MFs with a good peak-shape similarity and a good taxonomic score are highlighted; other ones were grouped in the "other" category for visibility. This led to a unique MF, $C_{30}H_{48}O_3$ (in dark blue), consistently annotated both in PI and NI, probably corresponding to different isomers, as suggested by the different traces observed. In panel D, features were colored according to the structure of the best candidate. This further refines panel C, showing that some of the features annotated with the same MF, $C_{30}H_{48}O_3$, do not share the same structural annotation, oleanolic acid (dark blue) and ursolic acid (light blue). They were both triterpenoid derivatives, corresponding to the structures of oleanolic acid and ursolic acid, respectively. Together, the different relations between features and their respective CAD peak could help in gaining information while reducing the complexity of the data (from the initial 128 features, two major compounds could be highlighted). Informed filtering of the features considered as major metabolomes is detailed in the next section.

Major Compounds' Filtering. Having all features informed as shown in Figure 2, a guided data reduction could be performed the same way on all peaks. This step was necessary to narrow the amount of information finally reported and analyzed. Basic statistics illustrating the number of features for PI, together with their link to their attributes, are given in Table 1.

As the presented workflow aimed at linking MS features to CAD peaks to categorize them as "major" or "minor" metabolites, the fact that the features are under a CAD envelope was not sufficient. Filtering strategies were therefore developed to narrow down the number of features considered as "major."

Table 1. MS Features-CAD Peaks Statistics (Positive Ionization)

	no filter	peak-shape filter	peak-shape filter and taxon filter
picked features	3161	3161	3161
picked CAD peaks	62	62	62
features linked to CAD peak	1181	392	136
number of features per peak[a]	19.0	7.4	3.0
number of structures per peak[a]	16.6	6.6	2.6
number of MFs per peak[a]	16.6	6.6	2.6
number of chemical classes per peak[a]	13.4	5.9	2.5

[a]On average.

The first step consisted of applying the peak-shape similarity as a filtering criterion as illustrated in Figure 2 (panel A).

As this filtering process was effective but not sufficient (over six structures/formulas per peak were assigned), an additional filter was implemented to further narrow down the number of features considered major. This filter consisted of favoring compounds already known in the genus of the studied extract. If the compound was already isolated from the same biological organism, then the probability of it being a major compound was considered higher. If a compound was found in the studied species, then all other "major" candidate features under the same CAD peak were attributed to the "minor" category. If multiple compounds were found at the species level (taxonomic distance score = 0.9), all of them were kept, as there was no rationale to choose one over another. The same rule was then applied to the genus level (taxonomic distance score of 0.8), giving priority to the compounds found at the species level. This led to fewer features considered "major", with around two features remaining per CAD peak, as described in Table 1.

The corresponding table for NI mode is available in Appendix (Table S1). Roughly, 50 MS features were detected for 1 CAD peak. Overall, more features were detected in PI, compared to NI (3161 vs 2272). Not only the total features were higher in PI, but also the number of features per CAD peak. This is in line with previous conclusions, while PI mode is usually preferred as it is expected to be more generic, the ionization efficiency (ratio to noise) is better in NI mode.[50]

After the complete filtering process, on average, each CAD peak appeared to have an almost unambiguously attributed structure. The number of features and structures that could be attributed to a CAD peak was lowered by almost an order of magnitude (3.0 and 2.6 against 19.0 and 16.6 initially). In fact, even if there are annotated features under a CAD peak envelope before filtering, after filtering, some of the CAD peaks could not be linked any more to any confidently annotated and peak-shape correlated feature. Finally, only CAD peaks with linked features were therefore taken into account in the calculations.

The metrics in Table 1 could also be used to evaluate the CAD peak purity.[51,52] The peak purity could then be used to evaluate the quality of the acquired fragmentation spectra,[53] or even try to deconvolute them.[54] Therefore, to complement the averages shown in Table 1, an alluvial plot summarizing the filtering process applied to the MS features below each CAD peak in the PI data is illustrated in Figure 3.

This stringent filtering allowed one to avoid part of the inconsistent annotations generated by automated annotation

https://doi.org/10.1021/acs.jafc.3c03099
J. Agric. Food Chem. 2023, 71, 18010−18023

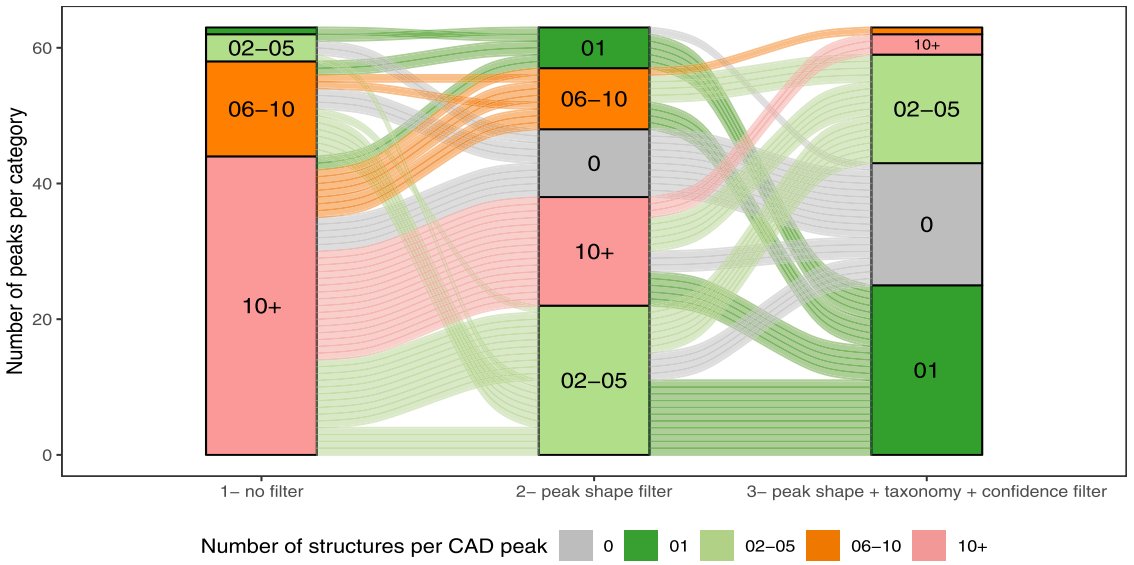

Figure 3. Alluvial plot of the informed data filtering. The underlying data are in the PI mode. The alluvial plot is colored according to five categories. The categories correspond to the number of candidate structures per CAD peak. The first block (from left to right) represents the situation before filtering, the second block after peak-shape filtering, and the last block after peak-shape filtering, taxonomic filter, and annotation confidence filter.

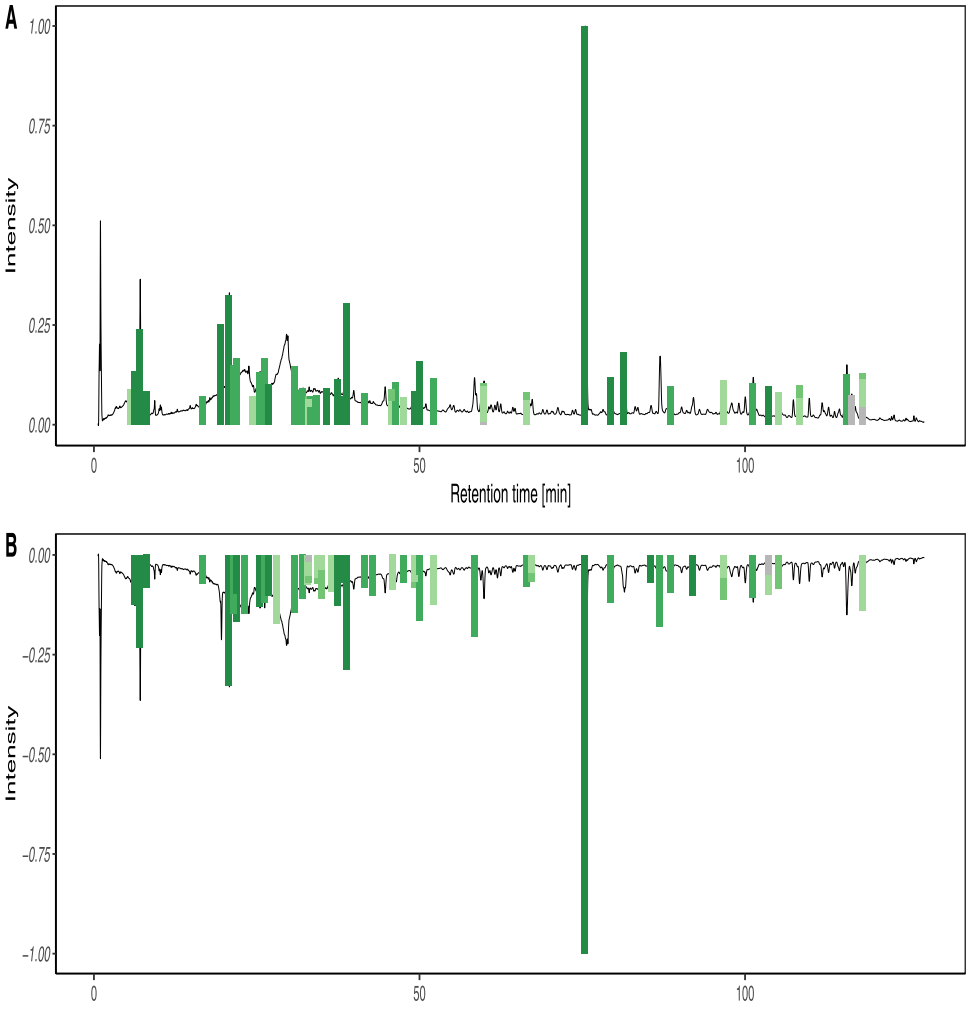

Figure 4. Taxonomically informed pseudochromatograms. The bars are colored according to the taxon level of the closest organism in which the compound was found relative to the biological source of the extract. Dark green bars indicate that the compound was already reported in the species. The remaining categories follow the same logic. Panel A illustrates the annotations in positive ionization mode while panel B illustrates the annotations in negative ionization mode.

https://doi.org/10.1021/acs.jafc.3c03099
J. Agric. Food Chem. 2023, 71, 18010−18023

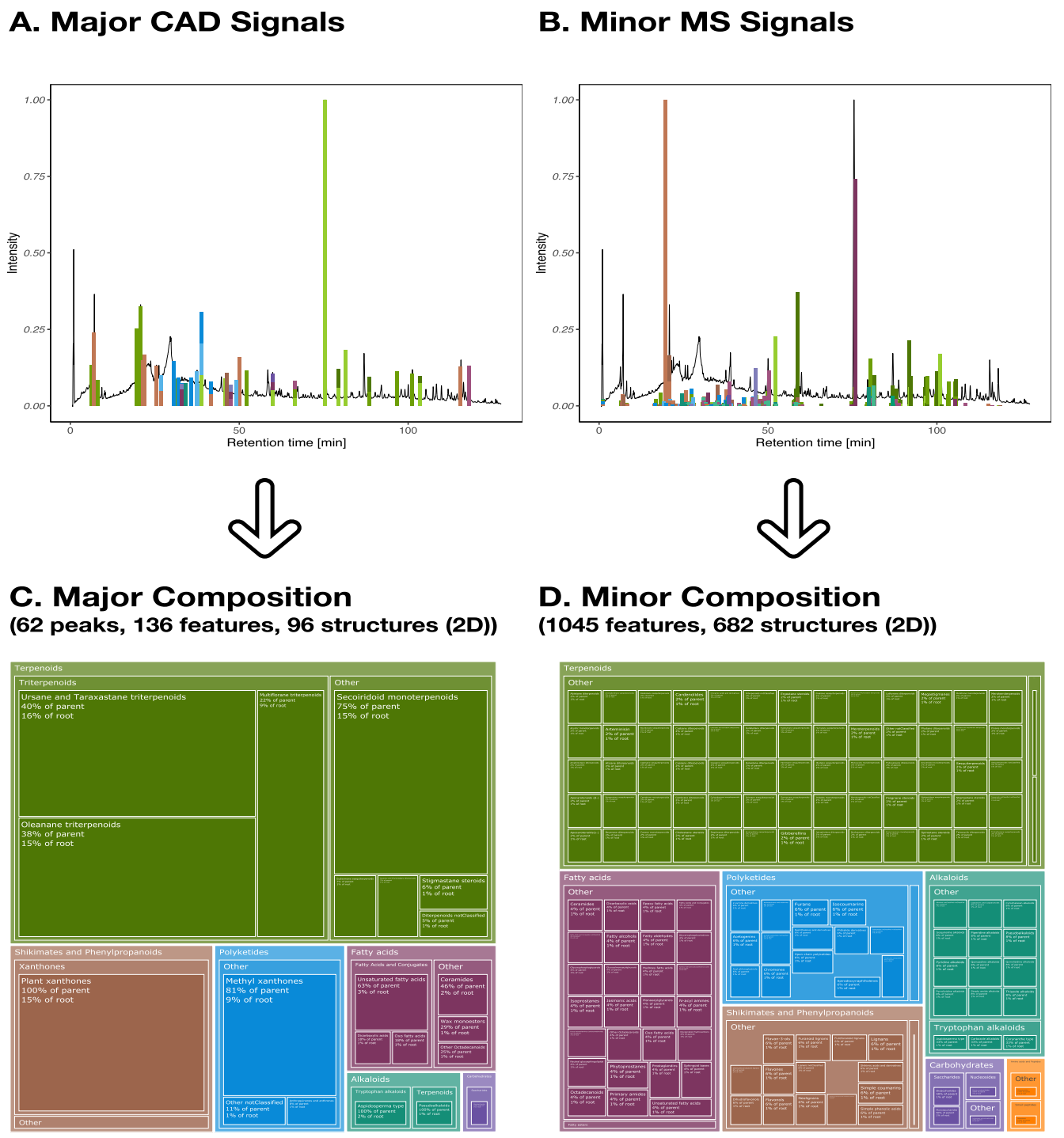

Figure 5. Chemically informed pseudochromatograms and related treemaps. In panels A and B, the bars are colored according to the chemical superclass attributed to the features obtained after filtering. Panel A illustrates the features that were classified as major, while panel B as minor. The data used were the positive ionization mode, and intensities were normalized. In panels C and D, the chemical classes (represented as bars in panels A and B) are regrouped into categories, independently from their position on the chromatogram. For panel C, the area of the rectangle is proportional to the CAD area. For panel D, MS intensity was neglected, and the number of annotations corresponding to a given chemical class is illustrated.

approaches and focus on consistent ones only. The filtering not only reduced the number of candidates linked to a CAD peak but also discarded the ones whose annotation was not confident (see gray block in Figure 3). For notice, features not passing the filtering process are kept as belonging to the "minor" metabolome. Starting from a situation where no CAD peak was unambiguously linked to a single structure, almost half of the CAD peaks were unambiguously linked to a structure at the end. Thus, the unambiguous annotation of the major metabolome could be automated for a portion of the peaks. When ambiguity remained, each plausible candidate was kept while considerably simplifying the initial data. In order to make full use of these filtered metabolite profiling data, different ways of visualizing the major and minor metabolome annotations are suggested below.

Data Visualization. To facilitate further analysis, various visualizations are created using the processed data. One such

https://doi.org/10.1021/acs.jafc.3c03099
J. Agric. Food Chem. 2023, 71, 18010−18023

Table 2. Automated Report of the Major Metabolome of *Swertia chirayita*

Peak ID	Peak RT [min]	Peak Area [%]	Feature ID	Feature RT [min]	Feature *m/z*	Structure	Short InChIKey	Anno-tation Score	Chemical Pathway	Chemical Superclass	Chemical Class	Closest organism*	Ref.
1	5.61	1.02											
2	6.22	1.50	1320	6.27	375.1290		HEYZWPRKKUGDCR	0.74	Terpenoids	Monoterpenoids	Secoiridoid monoter.	S. chirayita	58
3	7.04	2.71	1538	7.09	405.0821		AEDDIBAIWPIIBD	0.65	Shik & PAL	Xanthones	Plant xanthones	S. chirayita	59
4	8.08	0.96	1344	8.13	378.1063		VSJGJMKGNMDJCI	0.45	Terpenoids	Monoterpenoids	Secoiridoid monoter.	S. chirayita	60
5	16.68	0.84	1302	16.7	519.1526		CRZUKDMWTZQWQG	0.48	Terpenoids	Monoterpenoids	Secoiridoid monoter.	Swertia	61
6	18.29	0.94											
7	18.73	1.09											
8	19.51	3.06	3152	19.56	1205.3304		UZYZCCWBBBCDAD	0.45	Terpenoids	Monoterpenoids	Secoiridoid monoter.	S. chirayita	58,60
9	20.73	3.75	2253	20.78	1171.3320		DBOVHQOUSDWAPQ	0.45	Terpenoids	Monoterpenoids	Secoiridoid monoter.	S. chirayita	58,60,62–64
10	21.39	1.72	1139	21.34	495.1524		BSMWQZICWPFTBG	0.42	Terpenoids	Monoterpenoids	Iridoids monoter.	Swertia	65
11	21.92	1.93	11	21.89	201.1136		XRILWTHNIVQPJS	0.55	Terpenoids	Monoterpenoids	Secoiridoid monoter.	S. chirayita	
12	22.82	2.79											
13	23.07	1.31	1905	23.11	683.1276		DVMVOSIIRBPMIY	0.47	Shik & PAL	Xanthones	Plant xanthones	Swertia	
14	23.43	1.01											
15	23.72	2.43											
16	24.39	0.84											
17	25.03	1.36											
18	25.44	1.52	1299	25.51	519.0582		MPXAWSABMVLIBU	0.45	Polyketides	Xanthones	Methyl xanthones	S. chirayita	66
19	26.15	1.64	2270	26.21	1387.2266		HAGKIAPIVRRKSF	0.4	Shik & PAL	Xanthones	Plant xanthones	Swertia	
20	26.77	1.17	361	26.8	275.0553		JDIORNFCMMYMLF	0.64	Polyketides	Xanthones	Methyl xanthones	S. chirayita	66–69
21	27.26	1.27											
22	27.97	1.96											
23	30.74	1.68	1726	30.82	425.1354		KIJTTYRRBNSKMV	0.44	Polyketides	Xanthones	Methyl xanthones	Swertia	
24	31.51	1.00	2821	31.67	707.1969		VPIMEZFXQOIRND	0.4	Terpenoids	Monoterpenoids	Secoiridoid monoter.	Swertia	
25	32.07	1.16	1725	32.13	425.1352		KIJTTYRRBNSKMV	0.44	Polyketides	Xanthones	Methyl xanthones	Swertia	
26	32.74	0.84	486	32.8	367.0832		KNTWJKOMBDNCPG	0.41	Polyketides	Naphthalenes	Spirodioxynaphthalenes	Gentianales	70
27	32.99	0.81	523	33.05	297.2420		VWEWTQPRFPEQEU	0.66	Fatty acids	FA & conj.	Oxo fatty acids	Gentianales	71
28	33.31	0.84	929	33.37	333.0609		IPGXIZABQUBQSI	0.45	Polyketides	Polyc. aro. polyk.	Anthraq. & anthrones	Swertia	72
29	34.26	0.85	334	34.29	329.2343		MDIUMSLCYIJBQC	0.48	Fatty acids	Octadecanoids	Other Octadecanoids	Magnoliops.	73–76
30	34.94	1.25											
31	35.68	1.06	2140	35.69	471.0908		SUZVVVLVMNTJBQ	0.45	Polyketides	notClassified	notClassified	S. chirayita	
32	36.41	1.23											
33	36.99	0.97											
34	37.44	1.38	450	37.48	289.0707		PUECEVJMPDNNHT	0.73	Polyketides	Xanthones	Methyl xanthones	S. chirayita	66
35	37.75	0.82	370	37.84	341.0289		BDBVOZGRVBXANN	0.62	Polyketides	Xanthones	Methyl xanthones	S. chirayita	69
36	38.80	3.42	356	38.86	275.0546		JDIORNFCMMYMLF	0.61	Polyketides	Xanthones	Methyl xanthones	S. chirayita	66–69
37	41.54	0.93	2204	41.67	483.0926		CDRZVFJOMWIKNR	0.42	Shik & PAL	Xanthones	Plant xanthones	Swertia	77
38	42.83	1.09	561	42.84	385.1881		WJMOFJJTTHNUOH	0.4	Terpenoids	Monoterpenoids	Iridoids monoter.	Swertia	78
39	45.79	1.00	2259	45.94	489.3573		JXSVIVRDWWRQRT	0.66	Terpenoids	Triterpenoids	Urs. & Taraxa. triter.	Gentianaceae	79
40	46.33	1.24	429	46.28	283.1093		WRMNZCZEMHIOCP	0.44	Shik & PAL	ϕ eth. (C6-C2)	Phenylethanoids	Swertia	80
41	47.48	0.79	386	47.56	277.2159		PORPGKSIWVNTJC	0.5	Fatty acids	FA & conj.	Unsaturated fatty acids	Gentianales	81
42	49.21	0.97	455	49.30	289.0714		PUECEVJMPDNNHT	0.74	Polyketides	Xanthones	Methyl xanthones	S. chirayita	66
43	49.95	1.87	454	50.05	289.0712		HCOBNXLXVQFNAR	0.63	Shik & PAL	Xanthones	Plant xanthones	S. chirayita	82
44	52.22	1.34	1858	52.38	437.3420		KUFXJZXMWHNCEH	0.53	Terpenoids	Sesquiterpenoids	Eudesmane sesquiter.	Swertia	80
45	58.38	2.41	446	57.82	357.0991		ANFXBCPOMOAXMA	0.4	Polyketides	Xanthones	Methyl xanthones	Swertia	
46	59.84	1.21	408	59.94	279.2329		DTOSIQBPPRVQHS	0.59	Fatty acids	FA & conj.	Unsaturated fatty acids	Gentianales	83
47	66.38	0.93	2160	66.23	473.3634		HFGSQOYIOKBQOW	0.49	Terpenoids	Triterpenoids	Urs. & Taraxa. triter.	Gentianaceae	79
48	67.23	0.79	1017	67.24	473.3644		CJRASULCEXTLGY	0.58	Terpenoids	Triterpenoids	Urs. & Taraxa. triter.	Gentianaceae	
49	75.38	11.52	1884	75.53	439.3602		MIJYXULNPSFWEK	0.77	Terpenoids	Triterpenoids	Oleanane triter.	S. chirayita	49,84
50	79.40	1.38	1929	79.55	443.3899		OMUHPIZFDFBYLO	0.73	Terpenoids	Triterpenoids	Taraxerane triter.	S. chirayita	
51	81.39	2.09	1761	81.14	425.3800		VOFFFRIERAPYGK	0.75	Terpenoids	Triterpenoids	Multiflorane triter.	S. chirayita	
52	85.47	0.80	1181	85.69	501.3606		WCGUUGGRBIKTOS	0.49	Terpenoids	Triterpenoids	Urs. & Taraxa. triter.	S. chirayita	49
53	86.88	2.07	801	86.77	443.3183		UZFLPKAIBPNNCA	0.4	Terpenoids	Apocarotenoids	Apocarotenoids(ε-)	Swertia	80
54	87.89	0.80											
55	88.61	1.09	2030	88.79	455.3535		VORXDCSUUZISOZ	0.69	Terpenoids	Triterpenoids	Hopane & Moret. triter.	Swertia	85
56	91.99	1.17	1177	92.09	501.3605		WCGUUGGRBIKTOS	0.45	Terpenoids	Triterpenoids	Urs. & Taraxa. triter.	S. chirayita	49
57	96.65	1.30	1697	96.77	423.2536		IUVROZVEVDWZSZ	0.42	Terpenoids	Diterpenoids	Kaur. & Phyllocl. diter.	Magnoliops.	86
58	99.94	0.83											
59	101.18	1.22	2260	101.28	1197.8804		NPJICTMALKLTFW	0.4	Terpenoids	Steroids	Stigmastane steroids	Swertia	87
60	103.60	1.13	1750	103.78	425.3787		NINOWLHIYZJIFS	0.74	Terpenoids	Triterpenoids	Multiflorane triter.	S. chirayita	49,88
61	105.18	0.96											
62	108.3	1.10											
63	115.55	1.47	2815	115.45	707.1112		DVMVOSIIRBPMIY	0.44	Shik & PAL	Xanthones	Plant xanthones	Swertia	
64	116.31	0.85											
65	118.09	1.55	2449	117.86	554.5164		DLTXHYMYPKOZOA	>1.00	Fatty acids	Sphingolipids	Ceramides	Eukaryota	

[a]Organism with the shortest taxonomic distance where the structure was reported.

visualization is the pseudochromatogram, which combines chromatograms with histograms representing the integrated peak area. These pseudochromatograms provide insights into the biological origin of the annotated compounds relative to the biological source of the extract, as illustrated in Figure 4. In both panels, the main CAD peaks are related to the green or dark green categories (darker, meaning closely related organisms). This confirms the logic of favoring compounds already isolated from nearby organisms as being attributed to the CAD peaks. Although a majority of CAD peaks are annotated in both modes, an interesting complementarity can be observed. *S. chirayita* is not a well-studied species, nor is the genus *Swertia*, but the family Gentianaceae is. Therefore, its analysis can benefit from previously acquired data on other closely related organisms. In

https://doi.org/10.1021/acs.jafc.3c03099
J. Agric. Food Chem. 2023, 71, 18010−18023

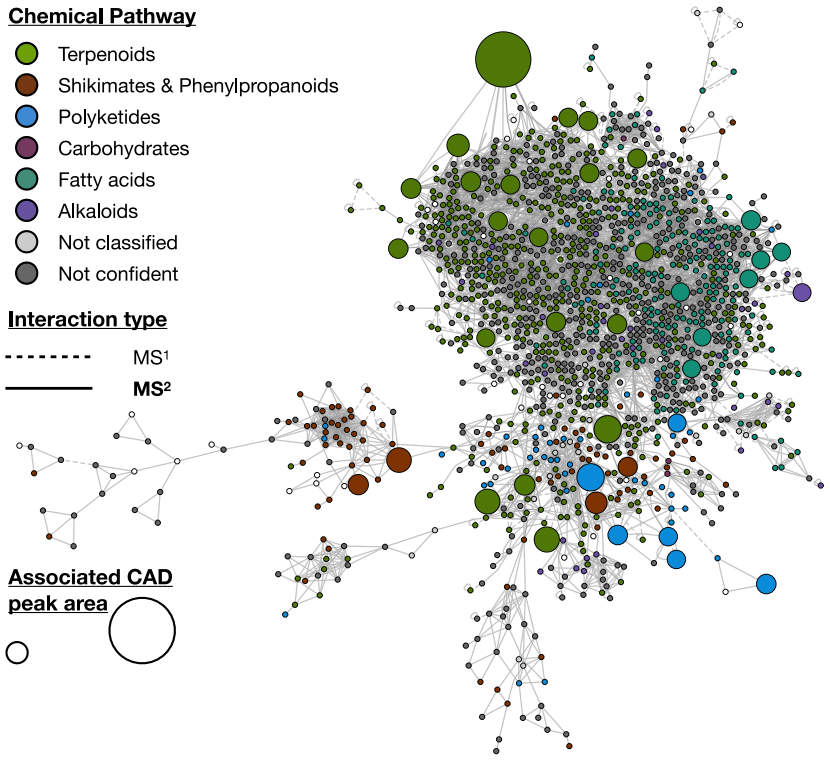

Figure 6. CAD-informed molecular network. Each node represents a feature. Features are colored according to the chemical pathway of their annotation. The size of the nodes corresponds to the peak area of the associated CAD peak. The type of line connecting the nodes corresponds to the interaction used to build the edge. Big- and medium-sized nodes can serve as anchor points for further exploration. Only the largest cluster (in positive ionization mode) is illustrated. The corresponding GNPS job is 36f42cb647c24b0c80e335a44310e27b.

the case of less studied organisms, belonging to a poorly studied branch of the tree of life, it will be necessary to rely more on chemical class attribution and the confidence of the spectral score. The bars can also be adapted to reflect the NPClassifier chemical pathways (and chemical superclasses) present in the extract.[17] Then, these bars can be aggregated together, as shown in Figure 5.

Following the methodology described above, the majority of CAD peaks from the *S. chirayita* extract were annotated with structures belonging to the terpenoid pathway, according to the NPClassifier chemical taxonomy. They accounted for 58% of the total area of the peaks (Figure 5, panel C). Terpenoids also represented most of the "minor" signals. The second most important pathways were the shikimates and phenylpropanoids (16%). A fair amount of peaks were also attributed to the polyketide pathway (15%), because of the ambiguous biosynthetic origin of xanthones. In plants, xanthones are known to be produced by the shikimate pathway,[55] but NPClassifier classified part of them as coming from the polyketide pathway (mainly in Fungi). Fungal endophytes are known to account for many of the NPs found in plants, and also xanthones.[56] As part of a general trend, the proportion of chemical classes in the minor composition has roughly followed that of the major composition, with the exception of fatty acids and alkaloids, which were overrepresented in the minor metabolome. The disproportional amount of alkaloids detected can be attributed to their good ionization in PI. Such compounds are detected at trace levels and do not yield corresponding CAD peaks. The fatty acids form the basis of membranes. Although the extraction process (see Experimental Section) is designed to remove them, it is difficult to deplete them completely. Moreover, they are essential building blocks for more evolved structures, and observing them as a leftover of plants' combinatorial biosynthesis is expected.[57]

Automated Reporting and Literature Matching. In addition to the pseudochromatograms and treemaps presented above, an automated tabular report with the most relevant information per CAD peak can be generated.[58−88] An example of such a table is presented in Table 2. The table shows the 65 integrated CAD peaks (from PI and NI), of which 17 were associated with structures previously reported in *S. chirayita*, and 15 additional in other *Swertia* spp. The majority of the annotated structures were xanthones. Three peaks were associated with a structure found in Gentianaceae, and four were in the Gentianales. Two peaks were already found in Magnoliopsida, but with good spectral matching scores, and were thus kept. Finally, one peak annotated as a ceramide by Sirius was kept even if never found in plants, as its initial score was extremely good. Partial overfitting might explain this score above 1 and should be investigated. Finally, 18 CAD peaks remained without confident annotation, accounting for 22.7% of the overall peak area. These peaks should be prioritized for further investigation by experts with the aim of structural elucidation. An additional column highlighting the closest organism from which the compound was isolated is available. The information is retrieved from Wikidata, which is currently the largest resource for NP occurrences.[89] The studied ethanolic extract of *S. chirayita* underwent a concentration step to remove polar compounds, thus favoring more apolar compounds such as triterpenoids, which represent the most abundant class annotated (see Experimental Section). The annotated triterpenoids correspond to the major red peaks between 75 and 92 min in Figure 1,

https://doi.org/10.1021/acs.jafc.3c03099
J. Agric. Food Chem. 2023, 71, 18010−18023

further reinforcing the reason for using a CAD detector, as the PDA detector would not have detected these compounds.

While our efforts were focused on automation of the confidence in the annotation and peak-shape similarity calculation, some improvements in the quantitation could be made. The error in the quantitation induced by the changing eluent composition could be reduced by implementing an inverse gradient, as in ref 34. A reverse gradient involves adding postcolumn solvent so that the composition of the resulting eluent is constant.[90] Indeed, the proportion of organic solvent has a considerable effect on the CAD response.[91,92] This could also be corrected by using algorithms, such as those suggested by refs 93 and 94. However, the proportion of organic solvent is not the only parameter for improving the precision and accuracy of the quantitation.[95] Even if ELSD and CAD are often considered to respond universally, significant differences have been observed depending on the chemical structures studied.[96] In addition, CAD response is nonlinear and uses a power function that has to be taken into account.[97,98] The chromatographic resolution is another limitation. Indeed, high peak capacities (number of peaks theoretically separated with a resolution of 1 within a chromatographic run)[99] are required to achieve good results. While addressing these limitations could improve the quantitative results, our workflow is already useful for generating comprehensive data on the composition of natural extracts, covering minor and major metabolomes.

CAD-Informed Molecular Network. As the workflow presented enables the annotation of peaks associated with the major and minor metabolome, the molecular networks make it easy to link major NPs and their minor analogues. Figure 6 illustrates the largest cluster of the PI analysis. Features were colored according to the chemical pathway of their annotation. By adding the CAD information as the size of the nodes, the molecular network included an additional semiquantitative dimension on the features obtained after filtering (see Table 1). This dimension could be used to target investigations to nodes correlated to major CAD signals or minor analogues that are linked to them. The semiquantitative dimension brought by the CAD is crucial to guide further efforts, as the amount of data can be misleading toward compounds favored only because of their ionization potential. We expect such visualization to facilitate downstream tasks such as targeted isolation and full structural elucidation of the relevant features. This strategy was already used in a collaborative study to target the isolation of photoactive pigments by taking into account some particular UV signals.[100] While the majority of preparative-scale instruments are hyphenated with UV detectors, the use of more generic detectors such as ELSD or CAD will facilitate the isolation of new NPs without chromophores.

EXPERIMENTAL SECTION

Plant Material. The aerial parts of *S. chirayita* were supplied by Tradall SA (Batch 155174).

Extraction. The plant material was used for other studies; therefore, high amounts of extract were needed.

1.1 kg of plant material was homogenized in a grinder and extracted at room temperature with 5 kg of H_2O and 6 kg of EtOH for 4 days. The extract was then filtered and stored in inox for decantation for 10 days. Finally, the extract was concentrated under reduced pressure, freeze-dried, and stored at −20 °C until further use. From 1.1 kg of plant material, 86.3 g of dried extract was obtained.

Vacuum Liquid Chromatography. Because the extract used for biological experiments was not described here, large amounts were needed. The first fractionation of the extract was undertaken by Vacuum Liquid Chromatography (VLC). 10 g of extract mixed with 20 g of silica was loaded on a chromatographic system made of two layers of silica (50 g and 200 g) separated by sand. This system was first very gently (approximately 2 drops/s) eluted with 3 × 500 mL of 100% H_2O and then 3 × 500 mL of 100% EtOH. The aqueous and ethanolic parts were collected separately, with a "mix" part corresponding to the dead volume of the system between both and the system was washed with DCM. All VLC fractions were then concentrated under reduced pressure, freeze-dried, and stored at −20 °C until further use. This procedure was repeated 3 times. From the initial 30 g of extract, a total of 27.4 g was recovered (91.3%). The mass of VLC_1 (H_2O) was 12.4 g, VLC_2 (mix) was 4.2 g, VLC_3 (EtOH) was 10.4 g. and VLC_4 (wash) was 0.3 g.

A small aliquot of VLC_3 resuspended at 5 mg/mL in EtOH was used for the data acquisition.

Data Acquisition. Chromatographic separation was performed on a Waters Acquity UHPLC system interfaced with a Corona Veo RS Charged Aerosol Detector (CAD) and a Q-Exactive Focus mass spectrometer using a heated electrospray ionization source. Thermo Scientific Xcalibur 3.1 software was used for instrument control. The LC conditions were as follows: column, Waters BEH C_{18} 150 × 2.1 mm, 1.7 μm; mobile phase, (A) water with 0.1% formic acid; (B) acetonitrile with 0.1% formic acid; flow rate, 400 μL·min^{-1}; injection volume, 6 μL; gradient, isocratic at 5% B for 0.5 min linear gradient of 5−100% B over 126.5 min (21.5 min for the short run) and isocratic at 100% B for 12 min (2 min for the short run). The source parameters were as follows: source voltage, 3.5 kV (pos); sheath gas flow rate (N_2), 55 units; auxiliary gas flow rate, 15 units; spare gas flow rate, 3.0; capillary temperature, 350.00 °C; and S-Lens RF Level, 45. The mass analyzer was calibrated using a mixture of caffeine, methionine−arginine−phenylalanine−alanine−acetate, sodium dodecyl sulfate, sodium taurocholate, and Ultramark 1621 in an acetonitrile/methanol/water solution containing 1% formic acid by direct injection. The data-dependent MS/MS events were performed on the three most intense ions detected in full-scan MS (Top3 experiment). The MS/MS isolation window width was 1 Da, and the stepped normalized collision energy (NCE) was set to 15, 30, and 45 units. In data-dependent MS/MS experiments, full scans were acquired at a resolution of 35,000 fwhm (at m/z 200) and MS/MS scans at 17,500 fwhm both with an automatically determined maximum injection time. After being acquired in an MS/MS scan, parent ions were placed in a dynamic exclusion list for 2.0 s. A custom exclusion list was used. An Acquity UHPLC PDA detector was used to acquire UV spectra, which were detected from 200 to 500 nm. An analytical split was used with a split ratio of 9:1 (CAD/MS). CAD parameters were: evaporation temperature at 40 °C, 5 bar N_2, power function 1.

Data Conversion. All raw data files were converted using ThermoRawFileParser v.1.4.0.[101] Currently, only ThermoRawFileParser is available to encode the CAD signal in the mzML file directly from the raw file (with the `--allDetectors` option). This feature was added by the authors of the tool on our request (see https://github.com/compomics/ThermoRawFileParser/pull/87 and https://github.com/compomics/ThermoRawFileParser/issues/60). Once converted, the mzML file contained all requested signals with their related chromatograms and spectra. The chromatograms were encoded as `BasePeak_0` for the BPI chromatogram, `PDA.1_TotalAbsorbance_0` for the PDA chromatogram, and `UV.1_CAD_1_0` for the CAD chromatogram. The generic command used was: `mono ThermoRawFileParser.exe -d $DIRECTORY --allDetectors --format = 2`.

Data Processing. *Signals' Pretreatment.* A manual investigation after the coupling of the different detectors to the LC measured that the MS signal was delayed by 0.090 and 0.055 min in comparison to the one of the PDA and the CAD, respectively. These values were used to align the signals.

To reduce the noise, signals underwent a Fourier transform. To perform the Fourier transform, the `filterFFT` function from the nucleR package[102] was used. Fourier components were set to 0.01. Further signal sharpening was performed as described in ref 37 with the

https://doi.org/10.1021/acs.jafc.3c03099
J. Agric. Food Chem. 2023, 71, 18010−18023

implementation of the function in R (see https://github.com/Adafede/cascade/blob/main/R/signal_sharpening.R).

Charged Aerosol Detector Data. Peaks were detected and integrated using the `get_peaks` function from the `chromatographR` package.[103] Default parameters were used.

MS Features' Extraction and Ion Identity Networking. MS features were extracted and informed using MZmine3 v.3.4.0.[104] All parameters were given in the form of an .xml file using the batch mode. The details of the xml file are available in the Appendix. The template xml file was filled with filename and MS mode using a bash command available in the Appendix. Adducts used for the different Ion Identity Networking steps are included in the file.

Feature-Based Molecular Network. In order to keep the retention time, the exact mass information and to allow for the separation of isomers, a feature-based molecular network (https://ccms-ucsd.github.io/GNPSDocumentation/featurebasedmolecularnetworking/) was created using the .mgf file resulting from the MZMine pretreatment step detailed above. When the parameter was not specified, default parameters were used. A network was then created where edges were filtered to have a cosine score above 0.6 and more than six matched peaks. Further edges (MS^2 relationships) between the two nodes were kept in the network if and only if each of the nodes appeared in each other's respective top 20 most similar nodes. The maximum shift between precursors was set to 500 Da. The maximum component size was set to 0. Edges (MS^2 or MS^1 relationships) were added as supplementary pairs. The spectra in the network were then searched against the GNPS' spectral libraries. All matches kept between network and library spectra were required to have a score above 0.7 and at least six matched peaks. The output was visualized using Cytoscape 3.9.1 software.[105] The GNPS job parameters and resulting data are available at the following addresses: 36f42cb647c24b0c80e335a44310e27b for positive and 8c2a1c97698e420d871c70bfdc80940a for negative.

MS Features' Annotation. Sirius. Sirius annotations were performed in batch mode by using Sirius 5.6.3. The details of the commands used for submission are available in the Appendix.

GNPS. GNPS annotations were retrieved from the following jobs:

- PI: 2a49fa45a5a74a8794399619d07359b5,
- NI: ef4b9a47a97c424d8b9411f09018a8e7,
- PI enriched: 2621c3e3af7e48298433fca269d8d1cb,
- NI enriched: 7066d9d60e694957a4798fe3a990a35e,
- PI enriched long: 36f42cb647c24b0c80e335a44310e27b,
- NI enriched long: 8c2a1c97698e420d871c70bfdc80940a

ISDB. The in silico libraries in positive and negative modes were generated as described in ref 47. The SMILES used were from ref 106 CFM 4.0 was used for in silico fragmentation.[107] Parameters are available at https://github.com/mandelbrot-project/spectral_lib_builder. In silico spectral matching was performed using spectral_lib_matcher, based on the matchms library.[108] Parent mass and MS/MS tolerances were 0.01 Da, the similarity method used `ModifedCosine`, and the minimum score and minimum peaks set to 0.

Taxonomically Informed Metabolite Annotation. Taxonomically Informed Metabolite Annotation was performed using all the above-mentioned inputs. It was performed using TIMA v2.8.0.[20,43] All parameters were the default ones, archived at. An up-to-date version of the documentation is available at https://taxonomicallyinformedannotation.github.io/tima-r/.

Annotations were considered confident when the final score after TIMA was above 0.4.

Link Between CAD and MS Data. The extracted MS feature list was reused in R to extract exact MS peak shapes and compare them to the CAD. Both chromatograms were resampled to the same frequency of 2 Hz. Therefore, peak intensity was sampled from the extrapolated traces each 0.5 s. Peak shapes were then compared using the `compareChromatograms` function from the MSnbase package with the `closest` method as an argument.[109,110]

Literature Matching. Compounds were retrieved from Wikidata as described in ref 89. The template of the query used is available at https://github.com/Adafede/wd-np-up-to-date/. Only references published between 1900 and 2023 were kept. Only compounds with an MW between 150 and 1500 Da and an xlogP between −1 and 6 were kept.

■ ASSOCIATED CONTENT

Data Availability Statement

The data are available through the GNPS job IDs (PI: 2a49fa45a5a74a8794399619d07359b5, NI: ef4b9a47a97c424d8b9411f09018a8e7, PI enriched: 2621c3e3af7e48298433fca269d8d1cb, NI enriched: 7066d9d60e694957a4798fe3a990a35e, PI enriched long: 36f42cb647c24b0c80e335a44310e27b, NI enriched long: 8c2a1c97698e420d871c70bfdc80940a). All programs written for this work can be found in the following repository: https://github.com/Adafede/cascade. The main dependencies were described before.

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jafc.3c03099.

Supporting Information 1. Comparison of peak detection on the raw and processed signal; Supporting Information 2. MS features-CAD peaks statistics (Negative Ionization); Supporting Information 3. MZmine commands used: both positive and negative; Supporting Information 4. Sirius commands used: both positive and negative (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

Adriano Rutz − *School of Pharmaceutical Sciences, University of Geneva, 1211 Geneva, Switzerland; Institute of Pharmaceutical Sciences of Western Switzerland, University of Geneva, 1211 Geneva, Switzerland; Institute of Molecular Systems Biology, ETH Zürich, 8093 Zürich, Switzerland;* orcid.org/0000-0003-0443-9902; Email: rutz@imsb.biol.ethz.ch

Jean-Luc Wolfender − *School of Pharmaceutical Sciences, University of Geneva, 1211 Geneva, Switzerland; Institute of Pharmaceutical Sciences of Western Switzerland, University of Geneva, 1211 Geneva, Switzerland;* orcid.org/0000-0002-0125-952X; Email: jean-luc.wolfender@unige.ch

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jafc.3c03099

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors thank Ivano Tonutti from Tradall SA for providing the plant material and insightful feedback. The authors thank the Bright Giant team and the Böcker lab for providing support for custom Sirius use. The authors thank Yann Sagon for his help in submitting the batches to HPC. The authors thank Jonathan Bisson for the molinfo service for structural depiction. The authors thank Sascha Bestmann from Thermo Fisher Scientific for valuable help with the CAD installation. The authors thank caetera for the extraction of the CAD signal directly during mzML conversion. This research was funded in whole or in part by the Swiss National Science Foundation (SNSF) [CRSII5_189921]. For the purpose of Open Access, a CC BY public copyright licence is applied to any Author Accepted Manuscript (AAM) version arising from this submission.

REFERENCES

(1) 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. Systematic Classification of Unknown Metabolites Using High-Resolution Fragmentation Mass Spectra. *Nat. Biotechnol.* **2021**, *39*, 462−471.

(2) Zhang, Y. V.; Rockwood, A. Impact of Automation on Mass Spectrometry. *Clin. Chim. Acta* **2015**, *450*, 298−303.

(3) Liigand, J.; Wang, T.; Kellogg, J.; Smedsgaard, J.; Cech, N.; Kruve, A. Quantification for Non-Targeted LC/MS Screening without Standard Substances. *Sci. Rep.* **2020**, *10*, No. 5808.

(4) Zhou, J.; Yin, Y. Strategies for Large-Scale Targeted Metabolomics Quantification by Liquid Chromatography-Mass Spectrometry. *Analyst* **2016**, *141*, 6362−6373.

(5) Cappiello, A.; Famiglini, G.; Palma, P.; Pierini, E.; Termopoli, V.; Trufelli, H. Direct-EI in LC−MS: Towards a Universal Detector for Small-molecule Applications. *Mass Spectrom. Rev.* **2011**, *30*, 1242−1255.

(6) Mahieu, N. G.; Patti, G. J. Systems-Level Annotation of a Metabolomics Data Set Reduces 25000 Features to Fewer than 1000 Unique Metabolites. *Anal. Chem.* **2017**, *89*, 10397−10406.

(7) Novotny, N. R.; Capley, E. N.; Stenson, A. C. Fact or Artifact: The Representativeness of ESI-MS for Complex Natural Organic Mixtures. *J. Mass Spectrom.* **2014**, *49*, 316−326.

(8) Wolfender, J.-L. HPLC in Natural Product Analysis: The Detection Issue. *Planta Med.* **2009**, *75*, 719−734.

(9) Viant, M. R.; Kurland, I. J.; Jones, M. R.; Dunn, W. B. How Close Are We to Complete Annotation of Metabolomes? *Curr. Opin. Chem. Biol.* **2017**, *36*, 64−69.

(10) Nothias, L.-F.; Petras, D.; Schmid, R.; et al. Feature-Based Molecular Networking in the GNPS Analysis Environment. *Nat. Methods* **2020**, *17*, 905−908.

(11) Stein, S. E. An Integrated Method for Spectrum Extraction and Compound Identification from Gas Chromatography/Mass Spectrometry Data. *J. Am. Soc. Mass Spectrom.* **1999**, *10*, 770−781.

(12) Watrous, J.; Roach, P.; Alexandrov, T.; Heath, B. S.; Yang, J. Y.; Kersten, R. D.; van der Voort, M.; Pogliano, K.; Gross, H.; Raaijmakers, J. M.; Moore, B. S.; Laskin, J.; Bandeira, N.; Dorrestein, P. C. Mass Spectral Molecular Networking of Living Microbial Colonies. *Proc. Natl. Acad. Sci. U.S.A.* **2012**, *109*, E1743−E1752.

(13) Schmid, R.; Petras, D.; Nothias, L. F.; et al. Ion Identity Molecular Networking for Mass Spectrometry-Based Metabolomics in the GNPS Environment. *Nat. Commun.* **2021**, *12*, No. 3832.

(14) Olivier-Jimenez, D.; Bouchouireb, Z.; Ollivier, S.; Mocquard, J.; Allard, P.-M.; Bernadat, G.; Chollet-Krugler, M.; Rondeau, D.; Boustie, J.; van der Hooft, J. J.; Wolfender, J.-L. *bioRxiv*, **2021**.

(15) Dührkop, K.; Fleischauer, M.; Ludwig, M.; Aksenov, A. A.; Melnik, A. V.; Meusel, M.; Dorrestein, P. C.; Rousu, J.; Böcker, S. SIRIUS 4: A Rapid Tool for Turning Tandem Mass Spectra into Metabolite Structure Information. *Nat. Methods* **2019**, *16*, 299−302.

(16) Djoumbou Feunang, Y.; Eisner, R.; Knox, C.; Chepelev, L.; Hastings, J.; Owen, G.; Fahy, E.; Steinbeck, C.; Subramanian, S.; Bolton, E.; Greiner, R.; Wishart, D. S. ClassyFire: Automated Chemical Classification with a Comprehensive, Computable Taxonomy. *J. Cheminf.* **2016**, *8*, 61.

(17) Kim, H. W.; Wang, M.; Leber, C. A.; Nothias, L.-F.; Reher, R.; Kang, K. B.; van der Hooft, J. J. J.; Dorrestein, P. C.; Gerwick, W. H.; Cottrell, G. W. NPClassifier: A Deep Neural Network-Based Structural Classification Tool for Natural Products. *J. Nat. Prod.* **2021**, *84*, 2795−2807.

(18) Hoffmann, M. A.; Nothias, L.-F.; Ludwig, M.; Fleischauer, M.; Gentry, E. C.; Witting, M.; Dorrestein, P. C.; Dührkop, K.; Böcker, S. High-Confidence Structural Annotation of Metabolites Absent from Spectral Libraries. *Nat. Biotechnol.* **2022**, *40*, 411−421.

(19) Dührkop, K. Deep Kernel Learning Improves Molecular Fingerprint Prediction from Tandem Mass Spectra. *Bioinformatics* **2022**, *38*, i342−i349.

(20) Rutz, A.; Dounoue-Kubo, M.; Ollivier, S.; Bisson, J.; Bagheri, M.; Saesong, T.; Ebrahimi, S. N.; Ingkaninan, K.; Wolfender, J.-L.; Allard, P.-M. Taxonomically Informed Scoring Enhances Confidence in Natural Products Annotation. *Front. Plant Sci.* **2019**, *10*, 238−252.

(21) Schymanski, E. L.; Jeon, J.; Gulde, R.; Fenner, K.; Ruff, M.; Singer, H. P.; Hollender, J. Identifying Small Molecules via High Resolution Mass Spectrometry: Communicating Confidence. *Environ. Sci. Technol.* **2014**, *48*, 2097−2098.

(22) Sysi-Aho, M.; Katajamaa, M.; Yetukuri, L.; Orešič, M. Normalization Method for Metabolomics Data Using Optimal Selection of Multiple Internal Standards. *BMC Bioinf.* **2007**, *8*, 93.

(23) El-Elimat, T.; Figueroa, M.; Ehrmann, B. M.; Cech, N. B.; Pearce, C. J.; Oberlies, N. H. High-Resolution MS, MS/MS, and UV Database of Fungal Secondary Metabolites as a Dereplication Protocol for Bioactive Natural Products. *J. Nat. Prod.* **2013**, *76*, 1709−1716.

(24) Yang, J.; Liang, Q.; Wang, M.; et al. UPLC-MS-ELSD-PDA as a Powerful Dereplication Tool to Facilitate Compound Identification from Small-Molecule Natural Product Libraries. *J. Nat. Prod.* **2014**, *77*, 902−909.

(25) Eom, H. Y.; Park, S.-Y.; Kim, M. K.; Suh, J. H.; Yeom, H.; Min, J. W.; Kim, U.; Lee, J.; Youm, J.-R.; Han, S. B. Comparison between Evaporative Light Scattering Detection and Charged Aerosol Detection for the Analysis of Saikosaponins. *J. Chromatogr. A* **2010**, *1217*, 4347−4354.

(26) Schilling, K.; Holzgrabe, U. Recent applications of the Charged Aerosol Detector for liquid chromatography in drug quality control. *J. Chromatogr. A* **2020**, *1619*, No. 460911.

(27) Russell, J. J.; Heaton, J. C.; Underwood, T.; Boughtflower, R.; McCalley, D. V. Performance of charged aerosol detection with hydrophilic interaction chromatography. *J. Chromatogr. A* **2015**, *1405*, 72−84.

(28) Li, J.; Hu, D.; Zong, W.; Lv, G.; Zhao, J.; Li, S. Determination of Inulin-type Fructooligosaccharides in Edible Plants by High-Performance Liquid Chromatography with Charged Aerosol Detector. *J. Agric. Food Chem.* **2014**, *62*, 7707−7713.

(29) Long, Z.; Guo, Z.; Acworth, I. N.; Liu, X.; Jin, Y.; Liu, X.; Liu, L.; Liang, L. A non-derivative method for the quantitative analysis of isosteroidal alkaloids from Fritillaria by high performance liquid chromatography combined with charged aerosol detection. *Talanta* **2016**, *151*, 239−244.

(30) Magnusson, L.-E.; Risley, D. S.; Koropchak, J. A. Aerosol-based detectors for liquid chromatography. *J. Chromatogr. A* **2015**, *1421*, 68−81.

(31) Tranchida, P. Q.; Salivo, S.; Franchina, F. A.; Bonaccorsi, I.; Dugo, P.; Mondello, L. Qualitative and Quantitative Analysis of the Unsaponifiable Fraction of Vegetable Oils by Using Comprehensive 2D GC with Dual MS/FID Detection. *Anal. Bioanal. Chem.* **2013**, *405*, 4655−4663.

(32) Grata, E.; Guillarme, D.; Glauser, G.; Boccard, J.; Carrupt, P.-A.; Veuthey, J.-L.; Rudaz, S.; Wolfender, J.-L. Metabolite Profiling of Plant Extracts by Ultra-High-Pressure Liquid Chromatography at Elevated Temperature Coupled to Time-of-Flight Mass Spectrometry. *J. Chromatogr. A* **2009**, *1216*, 5660−5668.

(33) Hettiarachchi, K.; Streckfuss, E.; Sanzone, J. R.; Wang, J.; Hayes, M.; Kong, M.; Greshock, T. J. Microscale Purification with Direct Charged Aerosol Detector Quantitation Using Selective Online One- or Two-Dimensional Liquid Chromatography. *Anal. Chem.* **2022**, *94*, 8309−8316.

(34) Baker, T. R.; Regg, B. T. A multi-detector chromatographic approach for characterization and quantitation of botanical constituents to enable in silico safety assessments. *Anal. Bioanal. Chem.* **2018**, *410*, 5143−5154.

(35) Guillarme, D.; Nguyen, D. T.; Rudaz, S.; Veuthey, J.-L. Method Transfer for Fast Liquid Chromatography in Pharmaceutical Analysis: Application to Short Columns Packed with Small Particle. Part II: Gradient Experiments. *Eur. J. Pharm. Biopharm.* **2008**, *68*, 430−440.

(36) Wahab, M. F.; Gritti, F.; O'Haver, T. C. Discrete Fourier Transform Techniques for Noise Reduction and Digital Enhancement of Analytical Signals. *TrAC, Trends Anal. Chem.* **2021**, *143*, 116354.

(37) Wahab, M. F.; O'Haver, T. C.; Gritti, F.; Hellinghausen, G.; Armstrong, D. W. Increasing Chromatographic Resolution of

https://doi.org/10.1021/acs.jafc.3c03099
J. Agric. Food Chem. 2023, 71, 18010−18023

Analytical Signals Using Derivative Enhancement Approach. *Talanta* **2019**, *192*, 492−499.
(38) Smith, C. A.; Want, E. J.; O'Maille, G.; Abagyan, R.; Siuzdak, G. XCMS: Processing Mass Spectrometry Data for Metabolite Profiling Using Nonlinear Peak Alignment, Matching, and Identification. *Anal. Chem.* **2006**, *78*, 779−787.
(39) Ni, Y.; Su, M.; Qiu, Y.; Jia, W.; Du, X. ADAP-GC 3.0: Improved Peak Detection and Deconvolution of Co-eluting Metabolites from GC/TOF-MS Data for Metabolomics Studies. *Anal. Chem.* **2016**, *88*, 8802−8811.
(40) Porter, S. E. G.; Stoll, D. R.; Rutan, S. C.; Carr, P. W.; Cohen, J. D. Analysis of Four-Way Two-Dimensional Liquid Chromatography-Diode Array Data: Application to Metabolomics. *Anal. Chem.* **2006**, *78*, 5559−5569.
(41) Wilde, M. J.; Zhao, B.; Cordell, R. L.; Ibrahim, W.; Singapuri, A.; Greening, N. J.; Brightling, C. E.; Siddiqui, S.; Monks, P. S.; Free, R. C. Automating and Extending Comprehensive Two-Dimensional Gas Chromatography Data Processing by Interfacing Open-Source and Commercial Software. *Anal. Chem.* **2020**, *92*, 13953−13960.
(42) Schmid, R.; Heuckeroth, S.; Korf, A.; et al. Integrative analysis of multimodal mass spectrometry data in MZmine 3. *Nat. Biotechnol.* **2023**, *41*, 447−449.
(43) Rutz, A.; Bisson, J.; Allard, P.-M. Taxonomically Informed Scoring Enhances Confidence in Natural Products Annotation. (accessed April, 17, 2023).
(44) Böcker, S.; Letzel, M. C.; Lipták, Z.; Pervukhin, A. SIRIUS: Decomposing Isotope Patterns for Metabolite Identification†. *Bioinformatics* **2009**, *25*, 218−224.
(45) 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. Publisher Correction: Database-independent Molecular Formula Annotation Using Gibbs Sampling through ZODIAC. *Nat. Mach. Intell.* **2020**, *2*, 727.
(46) Dührkop, K.; Shen, H.; Meusel, M.; Rousu, J.; Böcker, S. Searching Molecular Structure Databases with Tandem Mass Spectra Using CSI:FingerID. *Proc. Natl. Acad. Sci. U.S.A.* **2015**, *112*, 12580−12585.
(47) Allard, P.-M.; Bisson, J.; Rutz, A. ISDB: In Silico Spectral Databases of Natural Products. (accessed July 29, 2022).
(48) Horvath, C. G.; Lipsky, S. R. Peak Capacity in Chromatography. *Anal. Chem.* **1967**, *39*, 1893.
(49) Chakravarty, A. K.; Mukhopadhyay, S.; Das, B. Swertane Triterpenoids from Swertia Chirata. *Phytochemistry* **1991**, *30*, 4087−4092.
(50) Liigand, P.; Kaupmees, K.; Haav, K.; Liigand, J.; Leito, I.; Girod, M.; Antoine, R.; Kruve, A. Think Negative: Finding the Best Electrospray Ionization/MS Mode for Your Analyte. *Anal. Chem.* **2017**, *89*, 5665−5668.
(51) Bryant, D. K.; Kingswood, M. D.; Belenguer, A. Determination of Liquid Chromatographic Peak Purity by Electrospray Ionization Mass Spectrometry. *J. Chromatogr. A* **1996**, *721*, 41−51.
(52) Bylund, D.; Danielsson, R.; Markides, K. E. Peak Purity Assessment in Liquid Chromatography− Mass Spectrometry. *J. Chromatogr. A* **2001**, *915*, 43−52.
(53) Lawson, T. N.; Weber, R. J. M.; Jones, M. R.; Chetwynd, A. J.; Rodríguez-Blanco, G.; Di Guida, R.; Viant, M. R.; Dunn, W. B. msPurity: Automated Evaluation of Precursor Ion Purity for Mass Spectrometry-Based Fragmentation in Metabolomics. *Anal. Chem.* **2017**, *89*, 2432−2439.
(54) Nikolskiy, I.; Mahieu, N. G.; Chen, Y.-J.; Tautenhahn, R.; Patti, G. J. An Untargeted Metabolomic Workflow to Improve Structural Characterization of Metabolites. *Anal. Chem.* **2013**, *85*, 7713−7719.
(55) Remali, J.; Sahidin, I.; Aizat, W. M. Xanthone Biosynthetic Pathway in Plants: A Review. *Fronti. Plant Sci.* **2022**, *13*, 98−112.
(56) Khattab, A. R.; Farag, M. A. Marine and Terrestrial Endophytic Fungi: A Mine of Bioactive Xanthone Compounds, Recent Progress, Limitations, and Novel Applications. *Crit. Rev. Biotechnol.* **2022**, *42*, 403−430.
(57) Pollier, J.; Moses, T.; Goossens, A. Combinatorial Biosynthesis in Plants: A (p)Review on Its Potential and Future Exploitation. *Nat. Prod. Rep.* **2011**, *28*, 1897.
(58) Suryawanshi, S.; Asthana, R. K.; Gupta, R. C. Assessment of Systemic Interaction between Swertia Chirata Extract and Its Bioactive Constituents in Rabbits. *Phytother. Res.* **2009**, *23*, 1036−1038.
(59) Mandal née Sarkar, S.; Chatterjee, A. Structure of chiratanin, a novel dimeric xanthone. *Tetrahedron Lett.* **1987**, *28*, 1309−1310.
(60) Ray, S.; Majumder, H. K.; Chakravarty, A. K.; Mukhopadhyay, S.; Gil, R. R.; Cordell, G. A. Amarogentin, a Naturally Occurring Secoiridoid Glycoside and a Newly Recognized Inhibitor of Topoisomerase I from Leishmania Donovani. *J. Nat. Prod.* **1996**, *59*, 27−29.
(61) Zhou, Y.; Di, Y.-T.; Gesang, S.; Peng, S.-L.; Ding, L.-S. Secoiridoid Glycosides fromSwertia mileensis. *Helv. Chim. Acta* **2006**, *89*, 94−102.
(62) Wang, C.-Z.; Maier, U. H.; Eisenreich, W.; Adam, P.; Obersteiner, I.; Keil, M.; Bacher, A.; Zenk, M. H. Unexpected Biosynthetic Precursors of Amarogentin - A Retrobiosynthetic13C NMR Study. *Eur. J. Org. Chem.* **2001**, *2001*, 1459−1465.
(63) Inouye, H.; Nakamura, Y. Zwei stark bittere glucoside aus swertia japonica makino: Amarogentin und amaroswerin. *Tetrahedron Lett.* **1968**, *9*, 4919−4924.
(64) Saha, P.; Das, S. Regulation of hazardous exposure by protective exposure: Modulation of phase II detoxification and lipid peroxidation byCamellia sinensis andSwertia chirata. *Teratogenesis, Carcinogenesis, and Mutagenesis* **2003**, *23*, 313−322.
(65) Ikeshiro, Y.; Tomita, Y. A New Iridoid Glucoside of *Swertia japonica*. *Planta Med.* **1984**, *50*, 485−488.
(66) Ghosal, S.; Sharma, P.; Chaudhuri, R.; Bhattacharya, S. Chemical Constituents of the Gentianaceae V: Tetraoxygenated Xanthones of Swertia Chirata Buch.-Ham. *J. Pharm. Sci.* **1973**, *62*, 926−930.
(67) Tian, L.-Y.; Bai, X.; Chen, X.-H.; Fang, J.-B.; Liu, S.-H.; Chen, J.-C. Anti-diabetic effect of methylswertianin and bellidifolin from Swertia punicea Hemsl. and its potential mechanism. *Phytomedicine* **2010**, *17*, 533−539.
(68) Shi, G.; Lu, R.; Yang, Y.; Li, C.; Yang, A.; Cai, L. Crystal Structure of 1,5,8-Trihydroxyl-3-Methoxy Xanthone from *Swertia Chirayita*. *J. Chem. Crystallogr.* **2005**, *35*, 135−139.
(69) Singh, P. P.; Ambika; Chauhan, S. M. S. Activity-Guided Isolation of Antioxidant Xanthones from *Swertia Chirayita* (Roxb.) H. Karsten (Gentianaceae). *Nat. Prod. Res.* **2012**, *26*, 1682−1686.
(70) Jiang, H.; Ma, S.-G.; Li, Y.; Liu, Y.-B.; Li, L.; Qu, J.; Yu, S.-S. Spirobisnaphthalenes and lactones from the seeds of *Strychnos angustiflora* with potential anti-inflammatory activity. *Bioorg. Med. Chem. Lett.* **2016**, *26*, 4832−4836.
(71) Mahmood, C.; Daulatabad, J. D.; Mulla, G. M. M.; Mirajkar, A. M.; Hosamani, K. M. 7-keto-octadec-cis-11-enoic acid from Gardenia lucida seed oil. *Phytochemistry* **1991**, *30*, 2399−2400.
(72) Sakamoto, I.; Tanaka, T.; Tanaka, O.; Tomimori, T. Xanthone glucosides of *Swertia japonica* Makino and a related plant: Structure of a new glucoside, isoswertianolin and structure revision of swertianolin and norswertianolin. *Chem. Pharm. Bull.* **1982**, *30*, 4088−4091.
(73) Abbet, C.; Slacanin, I.; Hamburger, M.; Potterat, O. Comprehensive analysis of Phyteuma orbiculare L., a wild Alpine food plant. *Food Chem.* **2013**, *136*, 595−603.
(74) Farag, M. A.; El-Ahmady, S. H.; Elian, F. S.; Wessjohann, L. A. Metabolomics driven analysis of artichoke leaf and its commercial products via UHPLC−q-TOF-MS and chemometrics. *Phytochemistry* **2013**, *95*, 177−187.
(75) He, Z.-D.; Ma, C.-Y.; Tan, G. T.; Sydara, K.; Tamez, P.; Southavong, B.; Bouamanivong, S.; Soejarto, D. D.; Pezzuto, J. M.; Fong, H. H.; Zhang, H.-J. Rourinoside and rouremin, antimalarial constituents from Rourea minor. *Phytochemistry* **2006**, *67*, 1378−1384.
(76) Oueslati, M. H.; Jannet, H. B.; Mighri, Z.; Chriaa, J.; Abreu, P. M. Phytochemical Constituents from *Salsola tetrandra*. *J. Nat. Prod.* **2006**, *69*, 1366−1369.

https://doi.org/10.1021/acs.jafc.3c03099
J. Agric. Food Chem. 2023, 71, 18010−18023

(77) Hostettmann, K.; Jacot-Guillarmod, A. Identification de xanthones et de nouveaux arabinosides de C-glucosides flavoniques dans Swertia perennis L. *Helv. Chim. Acta* **1976**, *59*, 1584−1591.
(78) Kikuchi, M.; Kikuchi, M. Studies on the Constituents of *Swertia Japonica* MAKINO II. On the Structures of New Glycosides. *Chem. Pharm. Bull.* **2005**, *53*, 48−51.
(79) Fan, H.; Zang, Y.; Zhang, Y.; Zhang, H.-F.; Zhao, Z.; Hu, J.-F. Triterpenoids and Iridoid Glycosides from Gentiana dahurica. *Helv. Chim. Acta* **2010**, *93*, 2439−2447.
(80) Sakai, T.; Nakagawa, Y.; Iwashita, T.; Naoki, H.; Sakan, T. Semburins and Swertiols, Novel 2,8-Dioxabicyclo[3.3.1]Nonanes and Their Precursory Alcohols, from *Swertia Japonica* Makino. *Bull. Chem. Soc. Jpn.* **1983**, *56*, 3477−3482.
(81) Minquan, H. AC18 conjugated tetraenoic acid from Ixora chinensis seed oil. *Phytochemistry* **1990**, *29*, 1317−1319.
(82) Asthana, R. K.; Sharma, N. K.; Kulshreshtha, D. K.; Chatterjee, S. K. A xanthone from *Swertia chirayita*. *Phytochemistry* **1991**, *30*, 1037−1039.
(83) Elkahoui, S.; Smaoui, A.; Zarrouk, M.; Ghrir, R.; Limam, F. Salt-induced lipid changes in Catharanthus roseus cultured cell suspensions. *Phytochemistry* **2004**, *65*, 1911−1917.
(84) Bhan, S.; Kumar, R.; Kalla, A. K.; Dhar, K. Triterpenoids from Swertia petiolata. *Phytochemistry* **1988**, *27*, 539−542.
(85) Bhan, S.; Kumar, R.; Kalla, A. K.; Dhar, K. Isomeric 2, 3-seco-hopene lactones from Swertia petiolata. *Phytochemistry* **1987**, *26*, 3363−3364.
(86) Ohkoshi, E.; Makino, M.; Fujimoto, Y. Studies on the Constituents of Mikania hirsutissima (Compositae). *Chem. Pharm. Bull.* **1999**, *47*, 1436−1438.
(87) Tian, L. Y.; Chen, J. C.; Fang, J. B.; Zhou, Q.; Bai, X.; Zhou, J. Q.; Chen, X. H. A New Monosaccharide from Swertia Punicea Hemsl. *Chin. Chem. Lett.* **2009**, *20*, 684−686.
(88) Chakravarty, A. K.; Das, B.; Pakrashi, S. C.; McPhail, D. R.; McPhail, A. T. X-Ray Crystal Structure of Swertanone, a Triterpene of New Skeletal Type from Swertia Chirata Buch.-Ham. *J. Chem. Soc., Chem. Commun.* **1989**, 438−440.
(89) Rutz, A.; Sorokina, M.; Galgonek, J.; Mietchen, D.; Willighagen, E.; Gaudry, A.; Graham, J. G.; Stephan, R.; Page, R.; Vondrášek, J.; Steinbeck, C.; Pauli, G. F.; Wolfender, J.-L.; Bisson, J.; Allard, P.-M. The LOTUS Initiative for Open Knowledge Management in Natural Products Research. *eLife* **2022**, *11*, No. e70780.
(90) Vehovec, T.; Obreza, A. Review of operating principle and applications of the charged aerosol detector. *J. Chromatogr. A* **2010**, *1217*, 1549−1556.
(91) Khandagale, M. M.; Hutchinson, J. P.; Dicinoski, G. W.; Haddad, P. R. Effects of Eluent Temperature and Elution Bandwidth on Detection Response for Aerosol-Based Detectors. *J. Chromatogr. A* **2013**, *1308*, 96−103.
(92) Hutchinson, J. P.; Dicinoski, G. W.; Haddad, P. R. *Charged Aerosol Detection for Liquid Chromatography and Related Separation Techniques*; John Wiley & Sons, Inc., 2017; pp 191−219.
(93) Hutchinson, J. P.; Li, J.; Farrell, W.; Groeber, E.; Szucs, R.; Dicinoski, G.; Haddad, P. R. Universal Response Model for a Corona Charged Aerosol Detector. *J. Chromatogr. A* **2010**, *1217*, 7418−7427.
(94) Górecki, T.; Lynen, F.; Szucs, R.; Sandra, P. Universal Response in Liquid Chromatography Using Charged Aerosol Detection. *Anal. Chem.* **2006**, *78*, 3186−3192.
(95) Hutchinson, J. P.; Li, J.; Farrell, W.; Groeber, E.; Szucs, R.; Dicinoski, G.; Haddad, P. R. Comparison of the Response of Four Aerosol Detectors Used with Ultra High Pressure Liquid Chromatography. *J. Chromatogr. A* **2011**, *1218*, 1646−1655.
(96) Pawellek, R.; Krmar, J.; Leistner, A.; Djajić, N.; Otašević, B.; Protić, A.; Holzgrabe, U. Charged Aerosol Detector Response Modeling for Fatty Acids Based on Experimental Settings and Molecular Features: A Machine Learning Approach. *J. Cheminf.* **2021**, *13*, No. 53.
(97) Pawellek, R.; Muellner, T.; Gamache, P.; Holzgrabe, U. Power Function Setting in Charged Aerosol Detection for the Linearization of Detector Response − Optimization Strategies and Their Application. *J. Chromatogr. A* **2021**, *1637*, No. 461844.
(98) Haidar Ahmad, I. A.; Blasko, A.; Tam, J.; Variankaval, N.; Halsey, H. M.; Hartman, R.; Regalado, E. L. Revealing the Inner Workings of the Power Function Algorithm in Charged Aerosol Detection: A Simple and Effective Approach to Optimizing Power Function Value for Quantitative Analysis. *J. Chromatogr. A* **2019**, *1603*, 1−7.
(99) Fekete, S.; Schappler, J.; Veuthey, J.-L.; Guillarme, D. Current and future trends in UHPLC. *TrAC, Trends Anal. Chem.* **2014**, *63*, 2−13.
(100) Hammerle, F.; Quirós-Guerrero, L.; Rutz, A.; Wolfender, J.-L.; Schöbel, H.; Peintner, U.; Siewert, B. Feature-Based Molecular Networking−An Exciting Tool to Spot Species of the Genus Cortinarius with Hidden Photosensitizers. *Metabolites* **2021**, *11*, 791.
(101) Hulstaert, N.; Shofstahl, J.; Sachsenberg, T.; Walzer, M.; Barsnes, H.; Martens, L.; Perez-Riverol, Y. ThermoRawFileParser: Modular, Scalable, and Cross-Platform RAW File Conversion. *J. Proteome Res.* **2020**, *19*, 537−542.
(102) Flores, O.; Orozco, M. nucleR: A Package for Non-Parametric Nucleosome Positioning. *Bioinformatics* **2011**, *27*, 2149−2150.
(103) Bass, E. ethanbass/chromatographR: v0.4.7. 2023. (accessed April 13, 2023).
(104) Pluskal, T.; Castillo, S.; Villar-Briones, A.; Orešič, M. MZmine 2: Modular Framework for Processing, Visualizing, and Analyzing Mass Spectrometry-Based Molecular Profile Data. *BMC Bioinf.* **2010**, *11*, 395.
(105) Shannon, P.; Markiel, A.; Ozier, O.; Baliga, N. S.; Wang, J. T.; Ramage, D.; Amin, N.; Schwikowski, B.; Ideker, T. Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Res.* **2003**, *13*, 2498−2504.
(106) Rutz, A.; Bisson, J.; Allard, P.-M. The LOTUS Initiative for Open Natural Products Research: frozen dataset union wikidata (with metadata). (accessed July 29, 2022).
(107) Wang, F.; Liigand, J.; Tian, S.; Arndt, D.; Greiner, R.; Wishart, D. S. CFM-ID 4.0: More Accurate ESI-MS/MS Spectral Prediction and Compound Identification. *Anal. Chem.* **2021**, *93*, 11692−11700.
(108) Huber, F.; Verhoeven, S.; Meijer, C.; Spreeuw, H.; Castilla, E.; Geng, C.; van der Hooft, J.; Rogers, S.; Belloum, A.; Diblen, F.; Spaaks, J. Matchms - Processing and Similarity Evaluation of Mass Spectrometry Data. *J. Open Source Software* **2020**, *5*, 2411.
(109) Gatto, L.; Lilley, K. S. MSnbase-an R/Bioconductor Package for Isobaric Tagged Mass Spectrometry Data Visualization, Processing and Quantitation. *Bioinformatics* **2012**, *28*, 288−289.
(110) Gatto, L.; Gibb, S.; Rainer, J. MSnbase, Efficient and Elegant R-Based Processing and Visualization of Raw Mass Spectrometry Data. *J. Proteome Res.* **2021**, *20*, 1063−1069.

https://doi.org/10.1021/acs.jafc.3c03099
J. Agric. Food Chem. 2023, 71, 18010−18023