analytical chemistry

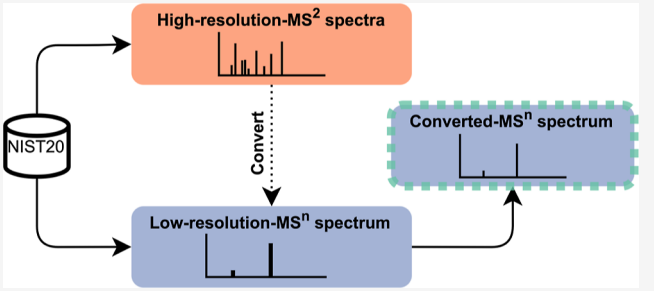
pubs.acs.org/ac
Article

Computational Expansion of High-Resolution-MS^n Spectral Libraries

Brandon Y. Lieng, Andrew T. Quaile, Xavier Domingo-Almenara, Hannes L. Röst, and J. Rafael Montenegro-Burke*

Cite This: *Anal. Chem.* 2023, 95, 17284−17291
Read Online

ABSTRACT: Commonly, in MS-based untargeted metabolomics, some metabolites cannot be confidently identified due to ambiguities in resolving isobars and structurally similar species. To address this, analytical techniques beyond traditional MS^2 analysis, such as MS^n fragmentation, can be applied to probe metabolites for additional structural information. In MS^n fragmentation, recursive cycles of activation are applied to fragment ions originating from the same precursor ion detected on an MS^1 spectrum. This resonant-type collision-activated dissociation (CAD) can yield information that cannot be ascertained from MS^2 spectra alone, which helps improve the performance of metabolite identification workflows. However, most approaches for metabolite identification require mass-to-charge (m/z) values measured with high resolution, as this enables the determination of accurate mass values. Unfortunately, high-resolution-MS^n spectra are relatively rare in spectral libraries. Here, we describe a computational approach to generate a database of high-resolution-MS^n spectra by converting existing low-resolution-MS^n spectra using complementary high-resolution-MS^2 spectra generated by beam-type CAD. Using this method, we have generated a database, derived from the NIST20 MS/MS database, of MS^n spectral trees representing 9637 compounds and 19386 precursor ions where at least 90% of signal intensity was converted from low-to-high resolution.



The metabolome is associated with a multitude of human diseases, such as diabetes mellitus, inflammation, and cancer.[1] Hence, untargeted metabolomics aims to detect and identify as many metabolites comprising a metabolome as possible. To this end, mass spectrometry (MS) is often applied. However, in MS-based untargeted metabolomics experiments, m/z values of ions detected at MS^1 can often correspond to multiple possible metabolites.[2] To resolve such ambiguities, collision-activated dissociation (CAD) is used to fragment ions (herein "precursor ions") and generate MS^2 fragmentation spectra, which represent "fingerprints" of precursor ions.[3−7]

Yet, even with their MS^2 spectra, most metabolites detected in metabolomics experiments still cannot be confidently annotated.[8,9] Ambiguities in identification often arise due to isobaric or structurally similar species that produce similar fragmentation spectra. To address this, other analytical techniques are often employed alongside MS^2 analysis to generate additional structural information. For example, separation techniques such as liquid chromatography and ion mobility spectrometry coupled to MS can determine a metabolite's retention time and collisional cross-section, respectively.[9] Similarly, different fragmentation regimes can be applied to generate complementarily informative fragment ions from the same precursor ion.[10] For example, under a resonant-type CAD regime, only ions of a specific m/z are activated. In contrast, under a beam-type CAD regime, all ions, including fragment ions, are activated. Consequently, different sets of fragment ions can be observed between fragmentation spectra of the same precursor ion generated by either resonant- or beam-type CAD. Despite the availability of these complementary approaches, metabolite identification remains a difficult task.

MS^n fragmentation applies recursive rounds of fragmentation to a precursor ion and its fragment ions. In an MS^n experiment, an MS^n spectrum is generated by selecting ions detected on an MS^{n-1} spectrum for isolation and fragmentation. This produces a "spectral tree" describing the hierarchical relationships between MS^n spectra succeeding a particular MS^1 spectrum.[4,11,12] An example MS^n experiment and its corresponding spectral tree are shown in Figure 1. Since MS^n experiments reveal hierarchical relationships between fragment ions, MS^n spectra encode information about metabolite structures more accurately than MS^2 spectra on their own, suggesting that MS^n fragmentation may help improve the performance of current metabolite identification workflows.[13,14]

Received: July 27, 2023
Revised: October 31, 2023
Accepted: October 31, 2023
Published: November 14, 2023

https://doi.org/10.1021/acs.analchem.3c03343
Anal. Chem. 2023, 95, 17284−17291

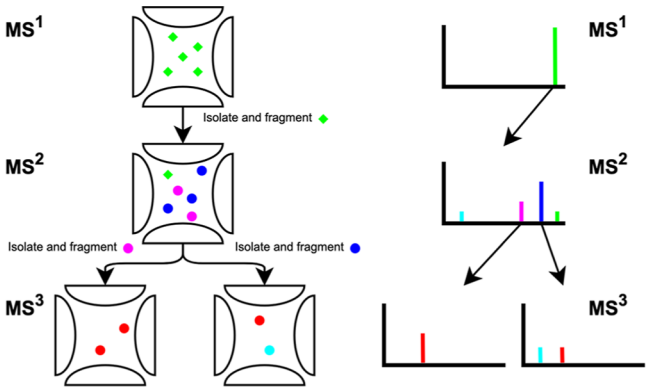


Figure 1. Illustration of an MS^n fragmentation experiment in a linear ion trap (left) and the corresponding MS^n spectral tree (right). Note that the term MS^n typically refers to cases where $n \geq 3$.

Metabolite identification workflows often involve the task of mass decomposition, where molecular formulae are assigned to fragment mass values, requiring mass values measured with high accuracy to reduce ambiguity.[2,15,16] To achieve high mass accuracy, spectra must be collected at high resolution. Unfortunately, MS^n spectra acquired at high resolution are relatively rare in current publicly available spectral databases. The lack of high-resolution-MS^n fragmentation spectra available in public databases can be partly attributed to the fact that MS^n fragmentation requires resonant-type CAD available only via low-resolution and, consequently, low-accuracy (± 0.15 m/z error, equivalent to 375 ppm error for an ion with $m/z = 400$) ion trap instrumentation.[17] On the other hand, high-resolution mass analyzers such as Orbitraps and time-of-flight-type instruments are designed with beam-type CAD capabilities.[18,19] Thus, most MS^n experiments have historically been performed using instruments that are incapable of high-resolution measurements. While Fourier transform-ion cyclotron resonance (FT-ICR) instruments can acquire high-resolution-MS^n data, they are rarely used for untargeted metabolomic profiling due to their high cost and slow acquisition speeds, which are incompatible with ultra-high-performance LC time scales.

Recent developments in MS instrumentation have yielded hybrid mass spectrometers that combine multiple mass analyzers and fragmentation modules into a single instrument. Hybrid MS instruments enable the acquisition of high-resolution-MS^n spectra, since hybrid instruments containing both a linear ion trap (LIT) and an Orbitrap can perform MS^n fragmentation in the LIT before transferring the resulting fragment ions to the Orbitrap for high-resolution mass analysis. However, generating a spectral library requires the collection of thousands of reference spectra—a laborious and resource-intensive process—which explains the limited availability and accessibility of spectra in current spectra repositories such as GNPS, METLIN, MoNA, and *mzCloud* (www.mzcloud.org).[20−23]

To leverage the additional information MS^n experiments can yield using current metabolite identification tools, a high-resolution-MS^n spectral database must be generated. Here, we describe a computational method to convert low-resolution peaks found throughout existing low-resolution-MS^n spectra in the NIST20 MS/MS database to a higher resolution using complementary high-resolution-MS^2 data generated by beam-type CAD (Figure 2). To achieve this, we first identified precursor ions represented by various types of spectra in NIST20, which we used to demonstrate the validity of two underlying assumptions of the conversion method for a given precursor ion: (1) MS^n fragmentation by resonant-type CAD produced fragment ions that are also observed under MS^2 fragmentation by beam-type CAD and (2) fragmentation spectra acquired using a linear ion trap were comparable to those acquired using an Orbitrap. After verifying these assumptions, we validated our conversion workflow using a subset of precursor ions in NIST20 for which both low- and high-resolution-MS^n spectra were available. Finally, we applied the peak resolution conversion method to all precursor ions in NIST20 represented by both low-resolution-MS^n and high-resolution-MS^2 spectra generated by beam-type CAD, resulting in a ∼9-fold expansion of high-resolution-MS^n spectra.

EXPERIMENTAL SECTION

NIST20 MS/MS Database. Precursor ions represented in NIST20 by high-resolution-MS^2 spectra generated by beam-type CAD (HR-BEAM-MS^2 spectra), high-resolution-MS^n spectra generated by resonant-type CAD (HR-RES-MS^n spectra), and low-resolution-MS^n spectra generated by resonant-type CAD (LR-RES-MS^n spectra) were identified (Figure 2a,b).[12] Here, the term "precursor ion" refers to a unique pairing of a compound and adduct/neutral loss. HR- and LR-RES-MS^n spectra in NIST20 were acquired at a normalized collision energy (NCE) of 35% and at MS levels MS^2 to MS^4. HR-BEAM-MS^2 spectra were acquired at a resolution of $R = 30000$ (full width at half-maximum @ 200 m/z) and across various NCEs.[12] NIST20 contains HR-RES-MS^n, LR-RES-MS^n, and HR-BEAM-MS^2 spectra for 1081, 11797, and 26262 compounds, respectively.

Matching Peaks. Peaks were matched between spectra using a Python script and error tolerances of ± 0.001 m/z when comparing HR-RES-MS^n to HR-BEAM-MS^2 spectra and ± 0.1 m/z when comparing HR-RES-MS^n to LR-RES-MS^n spectra. Peaks on HR-RES-MS^n were rounded down to one decimal place prior to comparisons with LR-RES-MS^n spectra. Percentages of peaks matched within HR- or LR-RES-MS^n spectral trees were calculated by determining the proportion of peaks within each tree that were matched to peaks on corresponding spectra. To calculate percentages of signal intensity matched, matched peaks were weighted by their relative intensities prior to calculating proportions of peaks matched.

Spectral Similarity Scores. Cosine similarity scores were computed using the *matchms* Python package.[24] An error tolerance of ± 0.001 m/z was used when computing scores between HR-RES-MS^n spectra from NIST20 and converted-MS^n spectra, whereas an error tolerance of ± 0.005 m/z was used when computing scores between HR-RES-MS^n spectra experimentally acquired in-house and converted-MS^n spectra.

Peak Resolution Conversion. A computational method was implemented to convert the resolution of peaks found throughout LR-RES-MS^n spectra from low-to-high using the m/z values of matched peaks from corresponding HR-BEAM-MS^2 spectra (Figure 2c). Note that the conversion of a peak from low-to-high resolution also implicitly converts the mass accuracy of the peak from low-to-high. For precursor ions where both LR-RES-MS^n and HR-BEAM-MS^2 spectra were available, LR peaks throughout the LR-RES-MS^n spectral tree were matched to HR peaks on corresponding HR-BEAM-MS^2 spectra. If a matching peak was found, the m/z value of the LR peak was replaced with

https://doi.org/10.1021/acs.analchem.3c03343
Anal. Chem. 2023, 95, 17284−17291

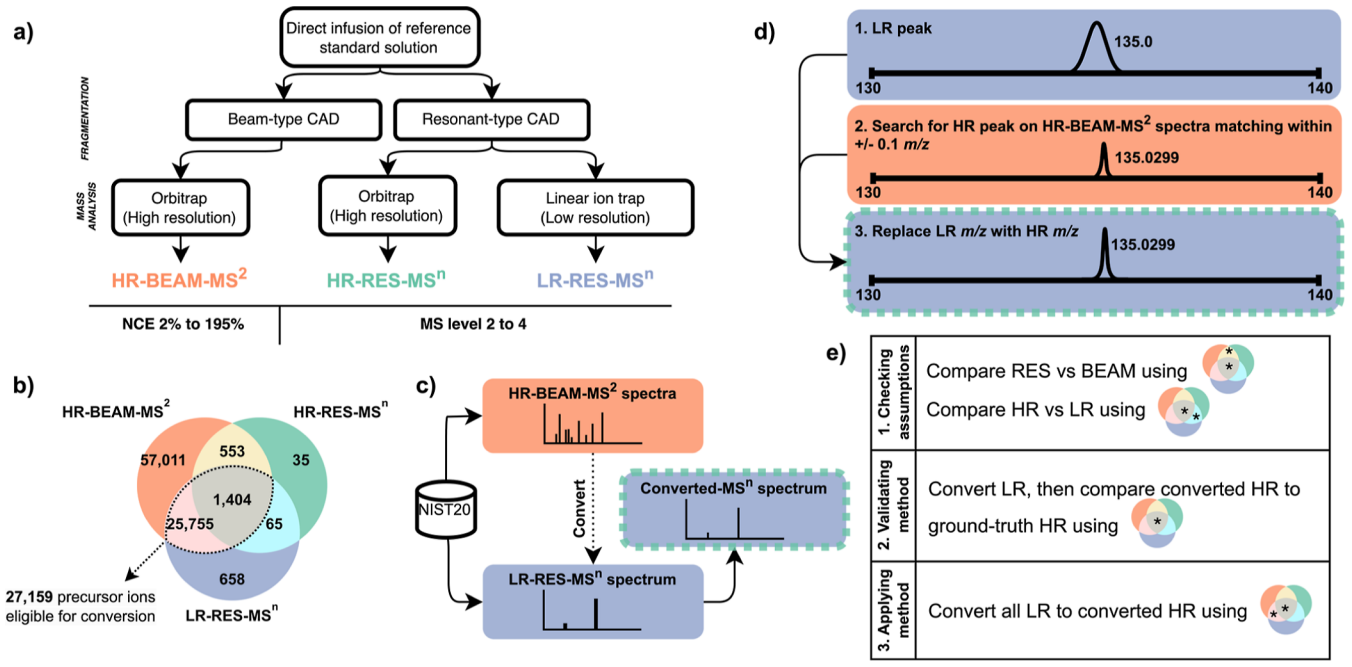


Figure 2. (a) Different combinations of fragmentation methods and mass analyzers were used to acquire three types of spectra in NIST20. (b) Unique precursor ions represented in NIST20 by each of the three types of spectra were identified. 27159 precursor ions were represented by both LR-RES-MS^n and HR-BEAM-MS^2 spectra (dotted outline). (c) To leverage these overlapping data, we proposed a peak resolution conversion method whereby information from an LR-RES-MS^n spectrum is combined with information from HR-BEAM-MS^2 spectra of the same precursor ion to produce a converted-MS^n spectrum. (d) In this method, LR-RES-MS^n peaks are converted to high-resolution peaks using high-resolution m/z values found on HR-BEAM-MS^2 spectra. (e) In this study, we first checked assumptions underlying the method, validated the method using a subset of 1404 precursor ions for which HR-RES-MS^n spectra were already available in NIST20, and then applied the method to the subset of 27159 precursor ions eligible for conversion.

the m/z value of the HR peak, effectively converting the LR peak to a higher resolution (Figure 2d).

When multiple HR peaks were found within the ±0.1 m/z window for an LR peak, the LR m/z value was heuristically converted using the m/z value associated with the most intense HR peak.

HR-RES-MS^n Spectra Generation. HR-RES-MS^n spectra were acquired using standard reference materials. Dry powders of quinine hydrochloride, verapamil hydrochloride, glafenine hydrochloride, ouabain octahydrate, and colchicine (Sigma-Aldrich, St. Louis, MO) were individually solubilized in LC-MS-grade methanol and water (1:1, v/v; Fisher Chemical, Hampton, NH) to a concentration of 1 μg/mL.

Each reference solution was then directly infused at 5 μL/min into an Orbitrap IQ-X Tribrid mass spectrometer (Thermo Scientific, Waltham, MA). HR-RES-MS^n fragmentation was performed in the LIT with parameters matching those used in NIST20 (NCE of 35% and resolution of R = 30000).[12] Spray parameters were set to their default values in Tune based on the 5 μL/min infusion rate. Spectra used in spectral comparisons were each comprised of ten individual scans, averaged using the *SpectraMerger* tool available in OpenMS.[25] After merging, noise peaks (<10% relative intensity) were removed from MS^2 and MS^3 spectra. Similarly, peaks with less than 20% relative intensity were removed from MS^4 spectra.

Generation of Fragmentation Trees. A fragmentation tree for verapamil $[M + H]^+$ was obtained using its LR-RES-MS^n spectral tree from NIST20 and the corresponding converted-MS^n spectral tree. In this fragmentation tree, a node was created for the base peak of each spectrum in the spectral tree. Each spectrum in the spectral tree corresponded to exactly one node in the fragmentation tree; thus, MS^n → MS^{n+1} relationships in the spectral tree were directly translated as edges between nodes. Since all peaks included as nodes in the fragmentation tree were converted from low-to-high resolution using the conversion method, each node was associated with a pair of low-accuracy and converted (high-accuracy) m/z values.

Generation of Molecular Formulas. The ChemCalc Web Service was used to perform all mass decompositions described.[26] Low-accuracy and converted m/z values were decomposed using error tolerances of ±0.1 and ±0.001 m/z units, respectively. To generate molecular formulae for the verapamil $[M + H]^+$ fragmentation tree example, the elements C[0−27], H[0−39], N[0−2], and O[0−4] were allowed based on the molecular formula of the precursor ion. Possible molecular formulae for each node were constrained to subformulae of the respective parent node. To generate molecular formulae for the 1000 MS^3 peaks sampled to assess the effect of conversion on mass decomposition, the elements C[0−105], H[0−180], N[0−20], O[0−75], P[0−5], S[0−5], Br[0−5], F[0−15], Cl[0−10], and B[0−1] were allowed.

RESULTS AND DISCUSSION

NIST20 Spectral Library. First, precursor ions represented by HR-RES-MS^n, LR-RES-MS^n, and HR-BEAM-MS^2 spectra in NIST20 were identified. Notably, some precursor ions were represented by multiple types of spectra and, most interestingly, 27159 precursor ions were represented by both LR-RES-MS^n and HR-BEAM-MS^2 spectra (Figure 2b). These precursor ions represented cases where our peak resolution conversion method could be applied. Instruments used to acquire spectra in these

https://doi.org/10.1021/acs.analchem.3c03343
Anal. Chem. 2023, 95, 17284−17291

subsets of NIST20 are listed in Supporting Information Table S1.

Checking Assumptions of the Conversion Method. Before the peak resolution conversion method was implemented, two assumptions underlying the method had to be tested. First, for a given precursor ion, fragment ion peaks found on RES-MS^n spectra are also found on BEAM-MS^2 spectra. This was important to test as resonant-type and beam-type CAD differ with respect to the frequency of collisions that precursor ions are subjected to under either regime. This could lead to different distributions of internal energy within precursor ions and the formation of different fragment ions.[10] Second, fragment ion peaks found on LR-MS^n spectra are also found on HR-MS^n spectra—it was important to ensure that ions detected in the LIT are also transferred to and detected in the Orbitrap. To test the first assumption, HR-RES-MS^n and HR-BEAM-MS^2 spectra from NIST20 representing the same precursor ion were compared (Figure 3a). Peaks from HR-RES-MS^n spectra were matched to peaks on corresponding HR-BEAM-MS^2 spectra. Following this, the percentage of peaks and intensity matched throughout HR-RES-MS^n spectra was calculated.

Across all 1957 precursor ions for which both HR-RES-MS^n and HR-BEAM-MS^2 spectra were available (yellow and gray sections in Figure 2b), 77.41% of HR-RES-MS^n peaks (137372 matched/166115 total) were matched to peaks found throughout corresponding HR-BEAM-MS^2 spectra. Although many peaks were not matched from HR-RES-MS^n to HR-BEAM-MS^2 spectra, weighing peaks by their intensity revealed that the majority (94.48%) of total HR-RES-MS^n signal intensity was matched to an HR-BEAM-MS^2 peak. Next, the distribution of the percentages of signal intensity matched within each precursor ion's HR-RES-MS^n spectral tree was visualized (Figure 3b). The median percentage of signal intensity throughout a precursor ion's HR-RES-MS^n spectral tree that was matched to a peak in the same precursor ion's HR-BEAM-MS^2 spectra was 99.33%. Taken together, these data validate the first assumption, showing that HR-RES-MS^n fragments are consistently present on HR-BEAM-MS^2 spectra.

To test the second assumption, LR-RES-MS^n and HR-RES-MS^n spectra from NIST20 representing the same precursor ion and MS^n fragmentation pathway were compared (Figure 3c). Across all 1469 precursor ions for which both LR- and HR-RES-MS^n spectra were available (cyan and gray sections in Figure 2b), 4192 pairings of LA and HA spectra that represented the same MS^n fragmentation pathways were formed. Peaks from LR-RES-MS^n spectra were then matched to peaks on paired HR-RES-MS^n spectra. The percentage of peaks and signal intensity matched throughout LR-RES-MS^n spectra was calculated. Across all comparisons, 54.37% of LR-RES-MS^n peaks (68688 matched peaks/126332 total peaks) were matched to corresponding peaks found throughout paired HR-RES-MS^n spectra. When peaks were weighted by their signal intensity, 91.81% of LR-RES-MS^n signal intensity was matched to a corresponding HR-RES-MS^n peak. In keeping with the previous analysis, the distribution of the percentages of signal intensity matched within each precursor ion's LR-RES-MS^n spectral tree was visualized (Figure 3d). The median percentage of signal intensity throughout a precursor ion's LR-RES-MS^n spectral tree that was matched to a peak found in paired HR-RES-MS^n spectra was 97.70%. These results validate the second assumption, as most of the signal intensity present throughout a precursor ion's LR-RES-MS^n spectra was usually found throughout its HR-RES-MS^n spectra.

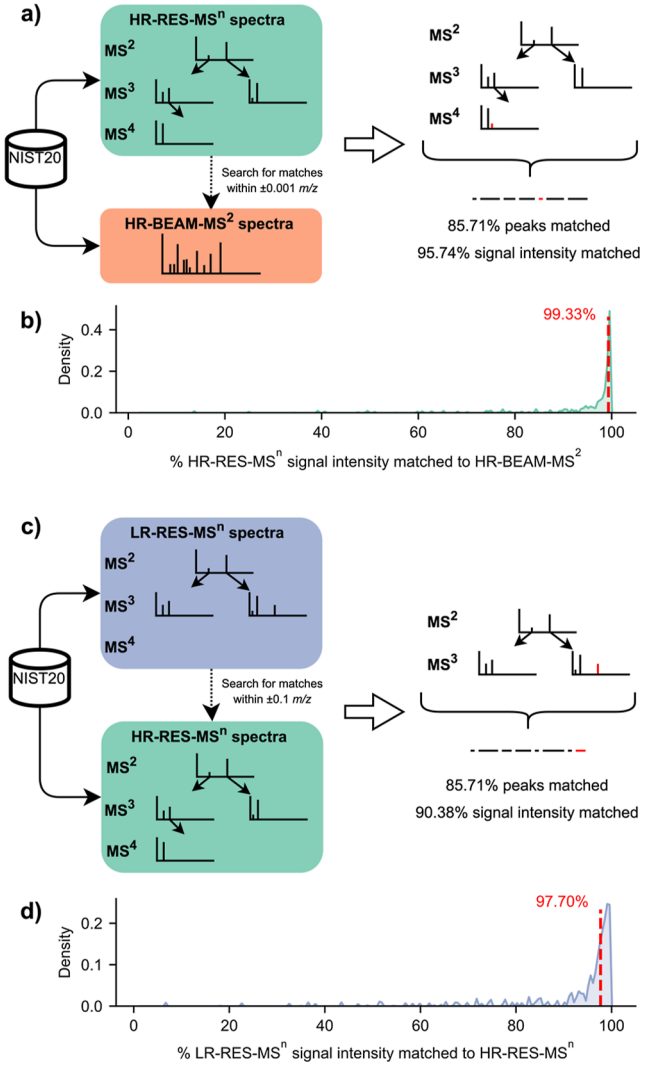


Figure 3. (a) Peaks in HR-RES-MS^2 spectra were matched to corresponding HR-BEAM-MS^n spectra. Peaks shown in black are matched and peaks shown in red are unmatched. (b) The distribution of percentages of signal intensity matched within each precursor ion's HR-RES-MS^n spectral tree to a set of corresponding HR-BEAM-MS^2 spectra. The red dashed line denotes the median value. (c) Peaks in LR-RES-MS^n spectra were matched to HR-RES-MS^n spectra that had an identical MS^n fragmentation pathway. (d) The distribution of percentages of signal intensity matched within each precursor ion's LR-RES-MS^n spectral tree.

Validation of the Conversion Method. Since the above analyses demonstrated the validity of both assumptions underlying the peak resolution conversion method, the method was implemented and validated using a subset of 1404 precursor ions in NIST20 for which LR-RES-MS^n, HR-BEAM-MS^2, and HR-RES-MS^n spectra were available (gray section in Figure 2b). Peaks in LR-RES-MS^n spectra were converted from low-to-high resolution using the implemented method. The percentage of signal intensity converted within each LR-RES-MS^n spectral tree was then calculated. The median percentage of signal intensity within each precursor ion's LR-RES-MS^n spectral tree that was matched to an HR-BEAM-MS^2 peak and converted to high resolution was 97.34% (Figure 4a). Examples of available HR-BEAM-MS^2 spectra and converted-MS^n spectra for the precursor ion avocadene acetate $[M + H-H_2O]^+$ are shown in Figure 4d,e, respectively.

https://doi.org/10.1021/acs.analchem.3c03343
Anal. Chem. 2023, 95, 17284−17291

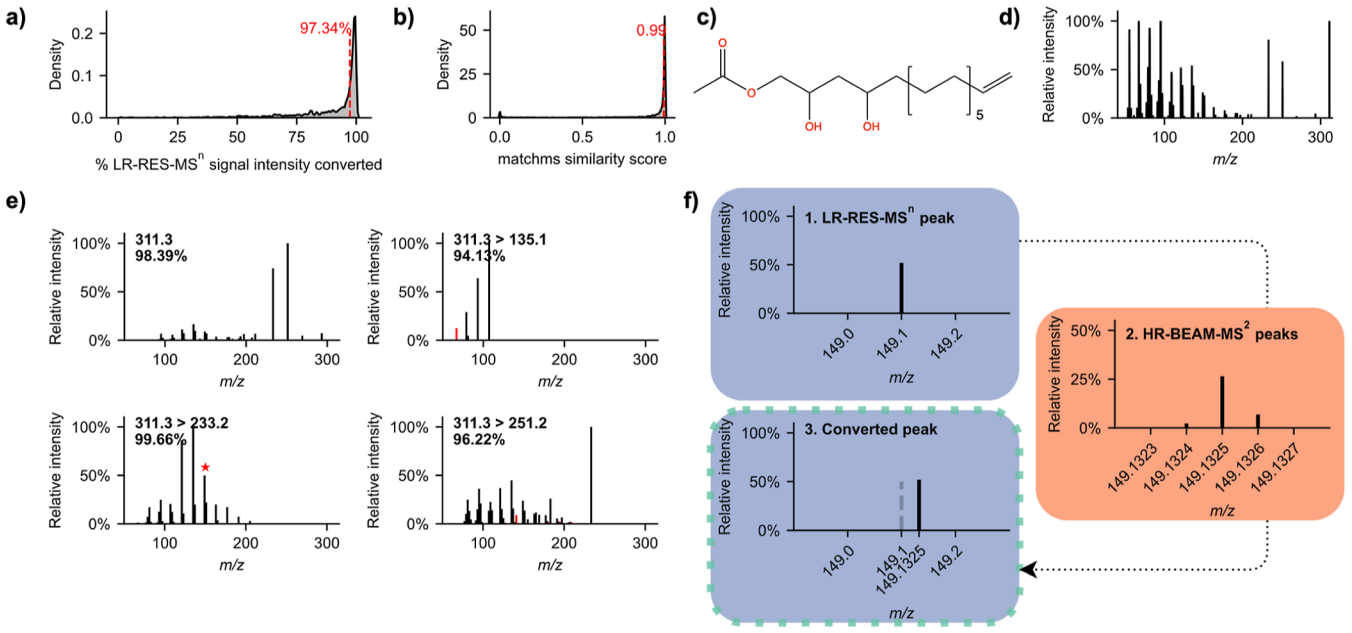


Figure 4. (a) The distribution of percentages of signal intensity converted from LA to HA within each LR-RES-MS^n spectral tree used to evaluate the method. (b) The distribution of cosine similarity scores calculated between pairs of converted-MS^n and HR-RES-MS^n spectra that represent the same MS^n fragmentation pathway. The red dashed line denotes the median value. (c) The structure of avocadene acetate. (d) Composite spectrum of HR-BEAM-MS^2 spectra available for avocadene acetate $[M + H-H_2O]^+$. (e) Converted-MS^n spectra for avocadene acetate $[M + H-H_2O]^+$. Peaks converted from low-to-high resolution are shown in black, whereas unconverted peaks are shown in red. Annotations denote each spectrum's MS^n fragmentation pathway and percentage of signal intensity converted. (f) An example showing the conversion of the LR peak denoted by the red star in (e) to HR. Here, HR-BEAM-MS^2 spectra were searched for a peak with an m/z value in the range (149.0, 149.2). Three candidate m/z values were found, and the one associated with the peak with the highest relative intensity (m/z = 149.1325) was used to convert the LR peak to HR.

Additionally, since HR-RES-MS^n spectra were available in NIST20 for this subset of precursor ions, they were compared against converted-MS^n spectra—between 4012 pairs of converted- and HR-RES-MS^n spectra that represented the same precursor ion and MS^n fragmentation pathway, cosine spectral similarity scores were computed. The median spectral similarity score between converted-MS^n spectra and their corresponding HR-RES-MS^n spectra was 0.99 (Figure 4b). In most cases, multiple HR peaks were found within the ±0.1 m/z window for each converted LR peak and the m/z value of the most intense HR peak was used for conversion (Supporting Information Figures 1 and 2). However, the fact that median spectral similarity scores were high suggested that, in most cases, the most intense peaks were usually converted to the correct HR m/z value or to one within ±0.001 m/z of the correct value (Figure 4b).

These results demonstrate that the peak resolution conversion method performed well, as most of the signal intensity throughout LR-RES-MS^n spectral trees in this subset was converted to HR. Furthermore, these converted-MS^n spectra shared high cosine similarity scores when compared to HR-RES-MS^n spectra using a peak matching tolerance suitable for the comparison of high-resolution spectra.

Generation of a Converted-MS^n Spectral Database. Following validation, the peak resolution conversion method was next applied to all 27203 LR-RES-MS^n spectral trees in NIST20 for which HR-BEAM-MS^2 spectra of the same precursor ions were also available (pink and gray sections in Figure 2b). This generated a database of converted-MS^n spectral trees representing 11562 compounds and 27159 precursor ions (44 precursor ions are represented by two spectral trees each). The proportion of signal intensity converted within each LR-RES-MS^n spectral tree was then calculated. The median percentage of signal intensity within each precursor ion's LR-RES-MS^n spectral tree that was matched to an HR-BEAM-MS^2 peak and converted to high resolution was 97.20% (Figure 5a).

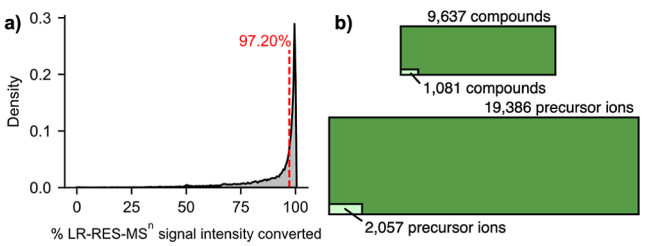


Figure 5. (a) Distribution of percentages of signal intensity converted from LR to HR within each LR-RES-MS^n spectral tree. (b) Counts of compounds and precursor ions represented by high-resolution-MS^n spectra in NIST20 before (light green area) and after (light green area plus dark green area) converting the resolution of LR-RES-MS^n peaks where possible. Areas are proportional to counts.

This is comparable to the median percentage of signal intensity matched and converted within each spectral tree during validation of the method (97.34%; Figure 4a). Spectral trees where more than 90% of LR-RES-MS^n intensity was converted from low-to-high resolution were considered as successfully converted to high resolution. Using this cutoff, 9637 compounds and 19386 precursor ions were represented by converted-MS^n spectral trees in the converted database. This corresponds to an 8.9-fold increase over the 1081 compounds and a 9.4-fold increase over the 2057 precursor ions that were previously represented by HR-RES-MS^n spectral trees in NIST20 (Figure 5b).

https://doi.org/10.1021/acs.analchem.3c03343
Anal. Chem. 2023, 95, 17284−17291

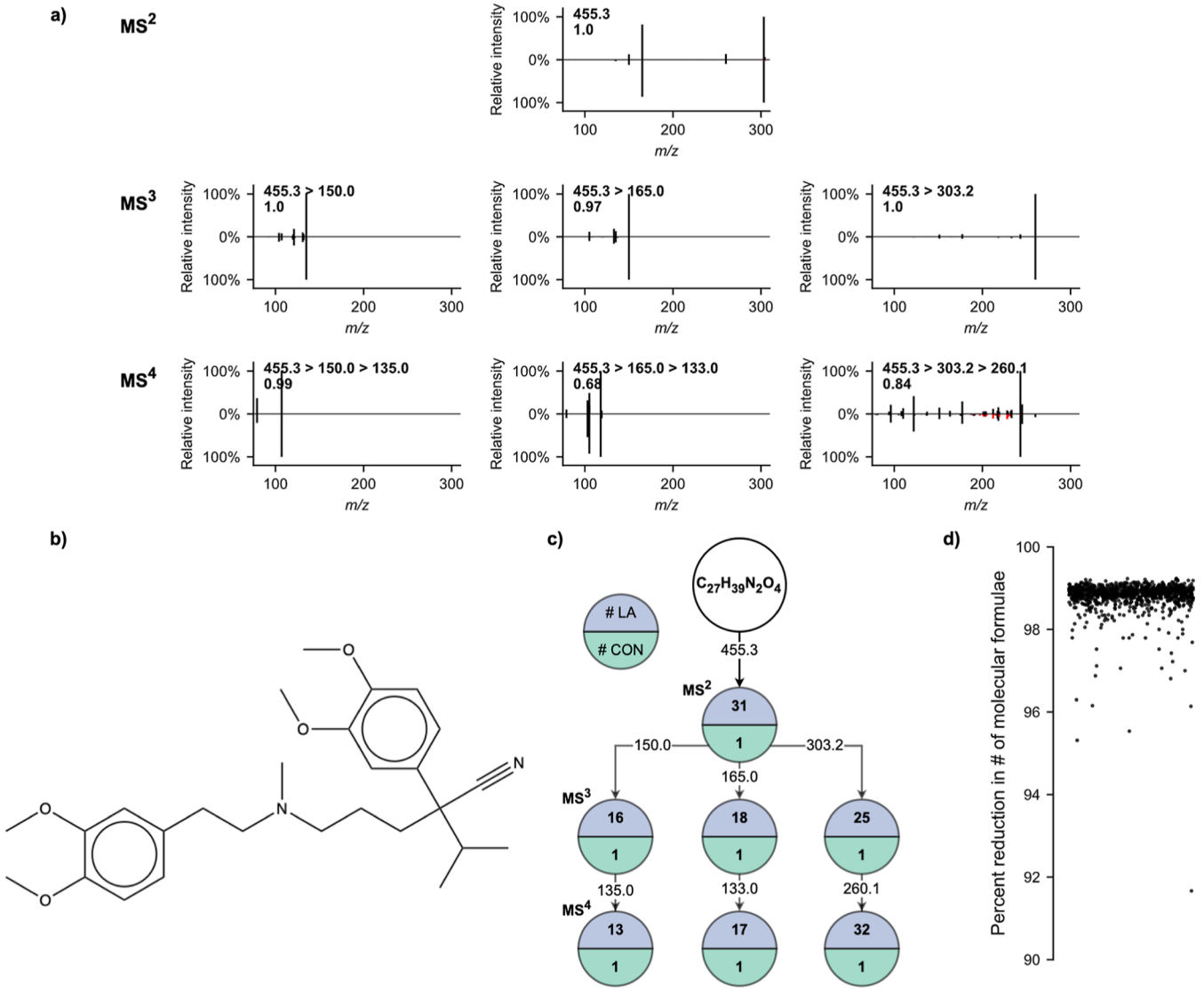


Figure 6. (a) Spectral tree of mirror plots comparing high-resolution-MS^n spectra acquired in-house (above the *x*-axis) to converted-MS^n spectra from NIST20 (below the *x*-axis). Peaks shown in red were not converted and remain as low-resolution peaks. Annotations denote MS^n fragmentation pathways and cosine spectral similarity scores calculated using the *matchms* package and a peak matching tolerance of ±0.005 *m*/*z* units.[24] (b) The precursor ion being fragmented in this example is verapamil $[M + H]^+$. (c) Number of candidate molecular formulae for each node in a fragmentation tree of the precursor ion verapamil $[M + H]^+$. Each node corresponds to a spectrum in the original spectral tree (shown in a) and its base peak. Values along edges represent precursor *m*/*z* values. (d) Percent reductions in the numbers of candidate molecular formulae explaining 1000 sampled MS^3 peaks, depending on whether the LA or converted *m*/*z* value was decomposed. Conversion of a peak from LA to HA reduced the number of candidates by roughly two orders of magnitude in most cases.

Comparison of Experimentally Acquired Spectra to Converted-MS^n Spectra. To demonstrate that experimentally acquired, high-resolution-MS^n spectra would produce high spectral similarity scores to spectra in the converted-MS^n database, spectral similarity was assessed between high-resolution-MS^n spectra acquired in-house and a subset of converted-MS^n spectra. Here, HR-RES-MS^n spectra were collected for five compounds represented by converted-MS^n spectral trees (i.e., >90% intensity converted) and that were not previously represented by HR-RES-MS^n spectra in NIST20 (pink section in Figure 2b).

In 22 of 31 comparisons, MS^n spectra acquired at high resolution shared high spectral similarity scores (>0.9) with converted-MS^n spectra from NIST20, even when using a stringent error tolerance (±0.005 *m*/*z*) to match peaks. These observations are consistent with spectral similarity scores computed between HR-RES-MS^n already in NIST20 and converted-MS^n spectra (Figure 4b). An example spectral tree of mirror plots and their associated similarity scores is shown in Figure 6a for one of the analyzed precursor ions, verapamil $[M + H]^+$ (Figure 6b). Mirror plots and similarity scores for the other four compounds are shown in Supporting Information Figure 3. This is a notable observation, as spectral similarity scores between reference spectra and experimentally acquired spectra are often considered when annotating detected metabolites with putative identities.

Generation of Molecular Formulas for Low-Accuracy and Converted *m*/*z* Values. To demonstrate that the conversion of peak resolution and mass accuracy facilitates mass decomposition, mass decomposition was assessed using two approaches—one using an example precursor ion from

https://doi.org/10.1021/acs.analchem.3c03343
Anal. Chem. 2023, 95, 17284−17291

NIST20, and the other using a random sample of converted peaks.

In the first approach, a fragmentation tree for the precursor ion verapamil $[M + H]^+$ was manually generated and mass decompositions were performed to annotate each tree node with putative molecular formulae explaining the node's pair of low-accuracy and converted m/z values. Fragmentation trees are useful for metabolite identification, as they model the fragmentation reactions that produced a fragmentation spectrum.[27,28] Using low-accuracy m/z values and an error tolerance of ±0.1 Da, all nodes in the manually generated verapamil $[M + H]^+$ fragmentation tree were decomposed to multiple candidate molecular formulae, including the MS^2 precursor m/z, which decomposed to 31 possible molecular formulae (Figure 6c). However, using converted m/z values and an error tolerance of ±0.001 Da, all nodes were explained by exactly one molecular formula. In this example, conversion of an LR-RES-MS^n spectral tree from NIST20 enabled the unambiguous decomposition of peak m/z values to molecular formulae.

Mass decomposition was also assessed in a more general fashion by sampling 1000 converted MS^3 peaks and computing possible molecular formulae for each peak's low-accuracy and converted m/z values. As expected, the number of candidate decompositions of converted m/z values was always smaller than the number of candidate decompositions of low-accuracy m/z values. On average, conversion of a low-accuracy m/z value allowed for the reduction of the size of the candidate decomposition list by 98.88% or by roughly two orders of magnitude, demonstrating again that converted-MS^n spectra facilitate the task of mass decomposition (Figure 6d).

CONCLUSIONS

Here, we present a computational framework to convert low-resolution peaks found throughout MS^n spectra in NIST20 to a higher resolution using complementary MS^2 spectra generated by beam-type CAD fragmentation. We first demonstrated that fragmentation of precursor ions under a resonant-type CAD regime tends to generate fragment ions that are also observed under a beam-type CAD regime. We also showed that most signals observed throughout low-resolution-MS^n spectra are also found in corresponding high-resolution-MS^n spectra of the same precursor ion and MS^n fragmentation pathway. Spectral similarity scores calculated between MS^n spectra collected with high resolution and corresponding converted-MS^n spectra revealed that, in most cases, such spectra were highly similar. Importantly, our computational methodology led to the expansion of the coverage of compounds and precursor ions represented by high-resolution-MS^n spectra in the NIST20 database by ∼9-fold, using either measure, and did not require the generation of any additional spectral data using reference standard materials, thus saving a considerable amount of material costs and research personnel hours. Our method is available as a Python script, accessible at www.github.com/montenegroburkelab/rescon, allowing users with access to the NIST20 spectral library to generate the converted-MS^n library. The script accepts input fragmentation spectra, subsets for the relevant spectra, converts the spectra, and outputs converted-MS^n spectra in the MSP format. Furthermore, this method is generalizable to other LR-RES-MS^n spectra not in the NIST20 database.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.analchem.3c03343.

> Histograms showing frequencies of instances where multiple candidate HR-BEAM-MS^2 peaks were available during conversion of the ground-truth subset of spectra and all LR-RES-MS^n spectra; spectral trees of mirror plots between HR-RES-MS^n spectra acquired-in-house and converted-MS^n spectra from NIST20; and lists of instruments used to acquire spectra in subsets of NIST20 (PDF)

AUTHOR INFORMATION

Corresponding Author

J. Rafael Montenegro-Burke − *Terrence Donnelly Centre for Cellular and Biomolecular Research, University of Toronto, Toronto, ON M5S 3E1, Canada; Department of Molecular Genetics, University of Toronto, Toronto, ON M5S 1A8, Canada; Institute of Biomedical Engineering, University of Toronto, Toronto, ON M5S 3G9, Canada;* orcid.org/0000-0001-7787-3414; Email: rafa.montenegro@utoronto.ca

Authors

Brandon Y. Lieng − *Terrence Donnelly Centre for Cellular and Biomolecular Research, University of Toronto, Toronto, ON M5S 3E1, Canada; Department of Molecular Genetics, University of Toronto, Toronto, ON M5S 1A8, Canada; Institute of Biomedical Engineering, University of Toronto, Toronto, ON M5S 3G9, Canada;* orcid.org/0000-0002-9269-9247

Andrew T. Quaile − *Terrence Donnelly Centre for Cellular and Biomolecular Research, University of Toronto, Toronto, ON M5S 3E1, Canada; Department of Molecular Genetics, University of Toronto, Toronto, ON M5S 1A8, Canada; Institute of Biomedical Engineering, University of Toronto, Toronto, ON M5S 3G9, Canada;* orcid.org/0000-0002-1194-3712

Xavier Domingo-Almenara − *Centre for Omics Sciences, Eurecat—Technology Centre of Catalonia & Rovira i Virgili University Joint Unit, Reus 43204 Catalonia, Spain; Department of Electrical, Electronic and Control Engineering, Universitat Rovira i Virgili, Tarragona 43007 Catalonia, Spain;* orcid.org/0000-0002-0133-6863

Hannes L. Röst − *Terrence Donnelly Centre for Cellular and Biomolecular Research, University of Toronto, Toronto, ON M5S 3E1, Canada; Department of Molecular Genetics, University of Toronto, Toronto, ON M5S 1A8, Canada;* orcid.org/0000-0003-0990-7488

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.analchem.3c03343

Author Contributions

The manuscript was written through contributions of all the authors. All authors have given approval to the final version of the manuscript.

Notes

The authors declare no competing financial interest.

https://doi.org/10.1021/acs.analchem.3c03343
Anal. Chem. 2023, 95, 17284−17291

■ ACKNOWLEDGMENTS

J.R.M.-B. and H.L.R. hold Canada Research Chair awards. This work was funded by the Canadian Institutes of Health Research (487000, 487740, and 419634), the Natural Sciences and Engineering Research Council of Canada (452229), the Canadian Cancer Society Research Institute (707420), and the Canada Foundation for Innovation (417332). X.D.-A is supported by the Spanish State Research Agency (AEI/10.13039/501100011033) grant PID2019-106277RR-I00; the "la Caixa" Foundation (ID 100010434) via the Junior Leader Fellowship (LCF/BQ/PR21/11840001); and the European Commission's Horizon 2020 Research and Innovation Program via the GLOMICAVE project (no. 952908).

■ REFERENCES

(1) Guijas, C.; Montenegro-Burke, J. R.; Warth, B.; Spilker, M. E.; Siuzdak, G. *Nat. Biotechnol.* **2018**, *36* (4), 316−320.
(2) Kind, T.; Fiehn, O. *BMC Bioinf.* **2006**, *7* (1), 234.
(3) Bristow, A. W. T.; Webb, K. S.; Lubben, A. T.; Halket, J. *Rapid Commun. Mass Spectrom.* **2004**, *18* (13), 1447−1454.
(4) Dunn, W. B.; Erban, A.; Weber, R. J. M.; Creek, D. J.; Brown, M.; Breitling, R.; Hankemeier, T.; Goodacre, R.; Neumann, S.; Kopka, J.; Viant, M. R. *Metabolomics* **2013**, *9* (S1), 44−66.
(5) McLuckey, S. A. *J. Am. Soc. Mass Spectrom.* **1992**, *3* (6), 599−614.
(6) Blaženović, I.; Kind, T.; Ji, J.; Fiehn, O. *Metabolites* **2018**, *8* (2), 31.
(7) 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.; Reily, M. D.; Thaden, J. J.; Viant, M. R. *Metabolomics* **2007**, *3* (3), 211−221.
(8) da Silva, R. R.; Dorrestein, P. C.; Quinn, R. A. *Proc. Natl. Acad. Sci. U.S.A.* **2015**, *112* (41), 12549−12550.
(9) Monge, M. E.; Dodds, J. N.; Baker, E. S.; Edison, A. S.; Fernández, F. M. *Annu. Rev. Anal. Chem.* **2019**, *12* (1), 177−199.
(10) Ichou, F.; Schwarzenberg, A.; Lesage, D.; Alves, S.; Junot, C.; Machuron-Mandard, X.; Tabet, J.-C. *J. Mass Spectrom.* **2014**, *49* (6), 498−508.
(11) Sheldon, M. T.; Mistrik, R.; Croley, T. R. *J. Am. Soc. Mass Spectrom.* **2009**, *20* (3), 370−376.
(12) Yang, X.; Neta, P.; Stein, S. E. *J. Am. Soc. Mass Spectrom.* **2017**, *28* (11), 2280−2287.
(13) Scheubert, K.; Hufsky, F.; Böcker, S. Multiple Mass Spectrometry Fragmentation Trees Revisited: Boosting Performance and Quality Brown, D., Morgenstern, B., Eds.; Lecture Notes in Computer Science: Berlin, Heidelberg, 2014; Vol. *8701*, pp 217−231.*Algorithms in Bioinformatics*
(14) Scheubert, K.; Hufsky, F.; Rasche, F.; Böcker, S. *J. Comput. Biol.* **2011**, *18*, 1383−1397.
(15) Kind, T.; Fiehn, O. *BMC Bioinf.* **2007**, *8* (1), 105.
(16) Xing, S.; Shen, S.; Xu, B.; Li, X.; Huan, T. *Nat. Methods* **2023**, *20*, 881−890.
(17) Makarov, A.; Denisov, E.; Kholomeev, A.; Balschun, W.; Lange, O.; Strupat, K.; Horning, S. *Anal. Chem.* **2006**, *78* (7), 2113−2120.
(18) Hu, Q.; Noll, R. J.; Li, H.; Makarov, A.; Hardman, M.; Graham Cooks, R. *J. Mass Spectrom.* **2005**, *40* (4), 430−443.
(19) Perry, R. H.; Cooks, R. G.; Noll, R. J. *Mass Spectrom. Rev.* **2008**, *27* (6), 661−699.
(20) Wang, M.; Carver, J. J.; Phelan, V. V.; Sanchez, L. M.; Garg, N.; Peng, Y.; Nguyen, D. D.; Watrous, J.; Kapono, 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.; Kleigrewe, 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, C. A.; Torres-Mendoza, D.; Gonzalez, D. J.; Silva, D. B.; Marques, L. M.; Demarque, 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* (8), 828−837.
(21) Guijas, C.; Montenegro-Burke, J. R.; Domingo-Almenara, X.; Palermo, A.; Warth, B.; Hermann, G.; Koellensperger, G.; Huan, T.; Uritboonthai, W.; Aisporna, A. E.; Wolan, D. W.; Spilker, M. E.; Benton, H. P.; Siuzdak, G. *Anal. Chem.* **2018**, *90* (5), 3156−3164.
(22) MassBank of North America. https://mona.fiehnlab.ucdavis.edu/(accessed Apr 30, 2022).
(23) 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* (7), 703−714.
(24) Huber, F.; Verhoeven, S.; Meijer, C.; Spreeuw, H.; Castilla, E.; Geng, C.; van der Hooft, J.; Rogers, S.; Belloum, A.; Diblen, F.; Spaaks, J. *JOSS* **2020**, *5* (52), 2411.
(25) Röst, H. L.; Sachsenberg, T.; Aiche, S.; Bielow, C.; Weisser, H.; Aicheler, F.; Andreotti, S.; Ehrlich, H.-C.; Gutenbrunner, P.; Kenar, E.; Liang, X.; Nahnsen, S.; Nilse, L.; Pfeuffer, J.; Rosenberger, G.; Rurik, M.; Schmitt, U.; Veit, J.; Walzer, M.; Wojnar, D.; Wolski, W. E.; Schilling, O.; Choudhary, J. S.; Malmström, L.; Aebersold, R.; Reinert, K.; Kohlbacher, O. *Nat. Methods* **2016**, *13* (9), 741−748.
(26) Patiny, L.; Borel, A. *J. Chem. Inf. Model.* **2013**, *53* (5), 1223−1228.
(27) 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* (4), 299−302.
(28) Böcker, S.; Dührkop, K. *J. Cheminf.* **2016**, *8* (1), 5.

https://doi.org/10.1021/acs.analchem.3c03343
Anal. Chem. 2023, 95, 17284−17291