analytical chemistry

Open Access
This article is licensed under CC-BY-NC-ND 4.0

pubs.acs.org/ac
Article

NP^3 MS Workflow: An Open-Source Software System to Empower Natural Product-Based Drug Discovery Using Untargeted Metabolomics

Cristina F. Bazzano, Rafael de Felicio, Luiz Fernando Giolo Alves, Jonas Henrique Costa, Raquel Ortega, Bruna Domingues Vieira, Raquel Peres Morais-Urano, Luciana Costa Furtado, Everton L. F. Ferreira, Juliana R. Gubiani, Roberto G. S. Berlinck, Leticia V. Costa-Lotufo, Guilherme P. Telles,* and Daniela B. B. Trivella*

Cite This: *Anal. Chem.* 2024, 96, 7460−7469

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Natural products (or specialized metabolites) are historically the main source of new drugs. However, the current drug discovery pipelines require miniaturization and speeds that are incompatible with traditional natural product research methods, especially in the early stages of the research. This article introduces the NP^3 MS Workflow, a robust open-source software system for liquid chromatography−tandem mass spectrometry (LC−MS/MS) untargeted metabolomic data processing and analysis, designed to rank bioactive natural products directly from complex mixtures of compounds, such as bioactive biota samples. NP^3 MS Workflow allows minimal user intervention as well as customization of each step of LC−MS/MS data processing, with diagnostic statistics to allow interpretation and optimization of LC−MS/MS data processing by the user. NP^3 MS Workflow adds improved computing of the MS^2 spectra in an LC−MS/MS data set and provides tools for automatic $[M + H]^+$ ion deconvolution using fragmentation rules; chemical structural annotation against MS^2 databases; and relative quantification of the precursor ions for bioactivity correlation scoring. The software will be presented with case studies and comparisons with equivalent tools currently available. NP^3 MS Workflow shows a robust and useful approach to select bioactive natural products from complex mixtures, improving the set of tools available for untargeted metabolomics. It can be easily integrated into natural product-based drug-discovery pipelines and to other fields of research at the interface of chemistry and biology.

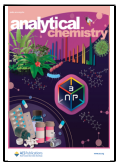
INTRODUCTION

Natural products (or specialized metabolites[1,2]) are well-known for their biological activity and historically are the main source of new drugs.[3] These natural molecules cover the chemical space of approved drugs[4,5] and further expand the biologically relevant chemical space. Natural products constitute attractive chemical probes to unveil new chemical (e.g., novel chemical scaffolds acting through a given biological target) and biological strategies (e.g., novel mechanisms of action) to fight diseases and understand biology.[6]

Although about 300 thousand natural products have been discovered by traditional methods in the last 100 years,[7] technical limitations inherent to natural product research had limited the use of natural products in modern drug discovery pipelines.[8−11] Such pipelines use large natural chemical libraries that are screened, and consequently, a large number of hits are found. Dereplication (the process of identifying known unknowns[11,12]) and prioritization[11,13] of the hits found after conducting a high-throughput screening campaign with unpurified natural products are nowadays possible and essential for natural product-based drug discovery pipelines.[7,11,13,14] Evolutions on dereplication approaches, aligned with recent advances in chromatographic and spectroscopic/spectrometric instrumentation and analysis,[7] have pushed the reemergence of natural products in modern drug discovery pipelines.

Untargeted metabolomics has played a central role in natural product dereplication as this approach allows rapid and small-scale investigation of natural products, even directly from complex mixtures, such as crude extracts and fractions derived from biota bioactive samples.[9,15−17] The advancement of mass spectrometry (MS) instrumentation and chemoinformatic tools is a determinant in the field of untargeted metabolomics. MS/MS fragmentation spectra (MS^2) represent the fingerprint of a chemical substance, and this information can be used to

Received: December 20, 2023
Revised: April 2, 2024
Accepted: April 5, 2024
Published: May 3, 2024



Downloaded via UNIV OF SAO PAULO on May 14, 2024 at 11:16:57 (UTC).
See https://pubs.acs.org/sharingguidelines for options on how to legitimately share published articles.

© 2024 The Authors. Published by American Chemical Society

https://doi.org/10.1021/acs.analchem.3c05829
Anal. Chem. 2024, 96, 7460−7469

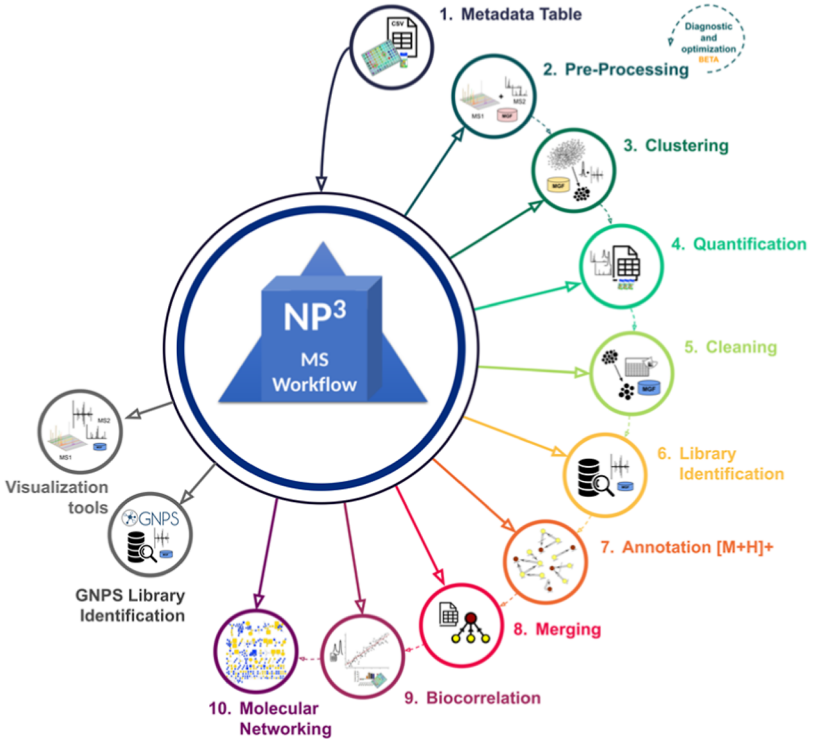


Figure 1. NP^3 MS Workflow.

scan the chemistry of biodiversity in a rapid and informative way. The MS^2 of virtually each natural product present in a sample from the biota can be readily collected in user-friendly mass spectrometers, which are nowadays available in most analytical chemistry laboratories[18,19] or in open facilities around the world.

MS^2 and liquid chromatography (LC)−MS/MS data processing have evolved over the last years, allowing more automated data treatment, noise reduction, and feature annotation.[20−22] In addition, MS^2 databases have been created, making available thousands of annotated MS^2 data for natural compounds[23] and hundreds of thousands MS^2 data predicted in silico[24] for most of the already reported natural products. Global Natural Products Social Molecular Networking (GNPS), a web-based and open-access MS knowledge base created for community-wide organization and sharing of MS^2 data, represents one of the most important efforts in the field of MS^2 for small molecules.[23] Algorithms for querying and comparing MS^2 databases have also been developed,[25−28] which allow readily structural annotation for virtually any MS^2 experimental data sets.

In addition to drug discovery, these advancements have also aided in the fields of chemical ecology,[29−31] chemotaxonomy,[32,33] human metabolomics,[34] microbiota characterization,[35] and synthetic biology,[36] among many other fields. However, the availability of integrated tools for the different scientific fields that can benefit from untargeted metabolomics is still being improved in the current days.

We present the NP^3 MS Workflow (Figure 1) developed to address this issue. NP^3 MS Workflow allows for automated LC−MS/MS data processing in a user-friendly environment, enabling its use by scientists from different disciplines. The NP^3 MS Workflow further brings customizations for natural product-based drug discovery pipelines, presenting robust shortcuts for bioactive natural product selection and structural annotation.

METHODS

NP^3 MS Workflow is a free software (https://github.com/danielatrivella/NP3_MS_Workflow), with 10 major steps to process and analyze LC−MS/MS data in positive mode. The steps can be executed through a command line interface, thoroughly or step-by-step, and their parameters may be redefined by the user. In step 1, the user provides a Metadata Table as input, describing the LC−MS/MS samples and sample groups. Preprocessing (step 2) detects LC−MS features (MS^1 peaks) for each sample, extracts their MS^2 spectra, determines correspondence among MS^1 and MS^2, and enriches MS^2 spectra with chromatographic information. Preprocessing results in an MS^2 file in Mascot Generic Format (MGF) for each sample and produces a diagnosis report with values for the most critical workflow parameters. It may be re-executed to optimize isomer resolution according to data quality and instrument accuracy.

Clustering, Quantification, and Cleaning (steps 3−5) cluster preprocessed MS^2 data relying on spectral similarity, peak intensity, peak area, and MS^1 information, creating clean consensus spectral representatives. These steps minimize the redundancies of multiple spectra from the same ion[37] while keeping isomers separated. Similarity comparisons in Cleaning use the NP^3-shifted cosine function, a greedy version of the modified cosine[38] developed for the NP^3 MS Workflow.

Library Identification (step 6) uses the Tremolo tool[25] to search for clean consensus spectra in the In Silico MS^2 DataBase (ISDB)[24] to retrieve structure information about natural products present in the samples. NPClassifier,[39] ClassyFire,[40] and NPAtlas[41] are used to add origin and chemical class information to the UNPD compound database. Annotation and Merging (steps 7−8) reduce redundancies due

https://doi.org/10.1021/acs.analchem.3c05829
Anal. Chem. 2024, 96, 7460−7469

to different ionization variants, indicating that real metabolites ($[M + H]^+$ ions) are present in the samples. The NP^3 MS Workflow solution to resolve ionization variants is the pairwise chemical annotation of putative ionization variants. This is done by using an ionization variant annotation molecular network (IVAMN),[29] based on a "ionization rules table" and the use of the PageRank[42] algorithm to select the $[M + H]^+$ representative ions.

Biocorrelation (step 9) computes the bioactivity correlation coefficient (BCC) of each consensus spectra to the biocorrelation groups and bioactivity scores of each sample, as provided by the user. BCCs may be used to rank the consensus spectra according to their likelihood of representing a bioactive compound in a given bioactive mixture.

Molecular Networking (step 10) creates the spectral similarity molecular network (SSMN), which connects every pair of clean consensus spectra that have NP^3-shifted cosine similarity above a given cutoff. The SSMN is then filtered to remove links below the minimum number of matched peaks, keeping only the K-most similar neighbors of each node, thus limiting the size of each component using a smooth approach (damped pruning). The final filtered SSMN has components of similar spectra that represent chemical and structurally related compounds. Annotations from step 7 are included as attributes in SSMN links, when available. The SSMN can be opened by graph visualization tools, such as Cytoscape.[43]

NP^3 MS Workflow also has separated commands to open MS^1 and MS^2 viewers, to incorporate experimental chemical annotations from GNPS, to assist the software setup, and to execute the test suit. More information on NP^3 MS Workflow development and implementation is in Supporting Information Methods and Notes S1−S6.

Experimental Section. The MA9 strain (*Annulohypoxylon moriforme*) was cultivated in PDB medium (10 L, 28 °C) for 14 days and extracted with ethyl acetate (EtOAc). Dried extracts were dissolved in methanol and submitted to partition with *n*-hexane. Methanolic final extract (882 mg, MA9 crude extract) was submitted to fractionation using reverse-phase C18 bench column. A stepwise gradient of H_2O/MeOH was used, rendering fractions A−G. Fractions E−F inhibited the proteasome, were joined, and refractionated using reverse-phase, rendering fractions EF1-6. Bioactive fractions EF2-4 were joined and fractionated by semipreparative high-performance liquid chromatography (HPLC) in a C18 InterSustain 5 μm, 4.6 × 250 mm (GL Sciences) using a H_2O/MeCN/MeOH solvent gradient. A major UV peak was purified using C18 X-Terra column (5 μm, 4.6 × 250 mm), resulting in isolated TMCA/B (1.9 mg). The procedure was repeated to accumulate mass for NMR analysis, starting with 18 L of a fungal culture. Similar yields were obtained. All generated fractions were used in case study #4.

Streptomyces sp. BRA346 wild-type bacteria were grown in A1 media for 7 days (200 rpm, 30 °C) and extracted with EtOAc, as described in ref 17. Five milligrams were used forsemipreparative fractionation in a C18 column (Gemini, 5 μm, 110 Å, 150 × 10 mm, Phenomenex) in HPLC Autopurification (Waters) system, rendering fractions #1−#95. The fractions were dried, resuspended in 25 μL of dimethyl sulfoxide (DMSO) (Sigma, cell culture grade), and directly assayed against proteasome inhibition. Bioactive fractions #57−#58 were selected for case study #3.

For TMC-86A isolation, the heterologous organism *S. coelicolor* M1146-*epn*/*tmc*-BRA-346[36] was cultivated in A1 media for 72 h (8 × 100 mL), 200 rpm, 30 °C. The culture media was acidified with HCl 0.1 M (1:2) and extracted with EtOAc. $MgSO_4$ was added to the organic fraction to adsorb water. The organic fraction was filtered and dried under vacuum. The dried extract (100 mg) was subjected to reversed-phase HPLC (Autopurification Waters). The m/z 343 peak was dried and isolated by UPLC (Acquity HClass Waters), rendering 0.4 mg of the pure compound. The procedure was repeated from 1.6 L of culture, resulting in equivalent yields of m/z 343.

LC−MS/MS data was acquired as reported in ref 29 using a H-Class UPLC (Waters, column: BEH18 1.7 μm, 2.1 × 100 mm, Aquity, Waters) coupled to an Impac II qQToF (Bruker), operating in positive mode. LC was conducted in a 0.3 mL/min flow rate, column at 40 °C, sample at 20 °C. The electrospray source was set to the range of 30−2000 Da, a detection speed of 8 Hz 500 V end plate offset, a capillarity of 4500 V, nebulizer at 4.0 bar, and drying gas flow (nitrogen) at 10 L/min with a drying temperature of 200 °C. For MS/MS, the collision cell was 5.0 eV, with collision energy in the range of 20−70 V and an absolute fragmentation cutoff of 1000. Ions below 200 Da were excluded, and the "active exclusion" function was enabled. For internal calibration, 10 mM sodium formate solution was used.

Proteasome inhibition assays followed ref 17 using in-house purified yeast 20S proteasome and the suc-LLVY-Amc (R&D systems) fluorogenic substrate. Experiments were carried out in Greiner 384-well black plates. The fluorescence was acquired using a Clario Star plate reader (BMG Labtech). Positive (reaction carried out with buffer replacing the enzyme) and negative (DMSO vehicle) controls were used to normalize the inhibitory data. Experiments were conducted in three replicates.

For NMR analyses, the isolated compounds were resuspended in DMSO-d_6 (TMC-95) or MeOH-*d* (TMC-86) and analyzed in 600 MHz Varian/Agilent Inova (TMC-86A) or 600 MHz Bruker AVANCE III (TMC-95A and B). Detailed information is given in the Supporting Information (Extended Methods).

RESULTS AND DISCUSSION

Case Studies. NP^3 MS Workflow was initially evaluated by analyzing the results on three data sets made available by the scientific community: two from the GNPS Feature-Based Molecular Networking (FBMN) work[20] (case studies #1 and #5) and one from the GNPS Ion Identity Molecular Networking (IIMN) paper[44] (case study #2). Case studies #2 and #5 are presented in Supporting Information, Notes S10 and S11, respectively. FBMN data sets show NP^3 MS Workflow capabilities to separate isomers and to handle fragmented clusters. IIMN data sets help demonstrate NP^3 MS Workflow capabilities to annotate ionization variants and to select the putative $[M + H]^+$ ions. These results also provide a comparison of the NP^3 MS Workflow and GNPS-FBMN-IIMN.

Additionally, two in-house data sets (case studies #3 and #4) are presented, reporting the application of NP^3 MS Workflow to deconvolute bioactive natural products from complex biota samples, representing typical natural product-based drug discovery projects, in which enzyme inhibitors from biota samples are searched for. These case studies report the identification of proteasome inhibitors in extracts and enriched chromatographic fractions of Brazilian bacteria and fungi, in

https://doi.org/10.1021/acs.analchem.3c05829
Anal. Chem. 2024, 96, 7460−7469

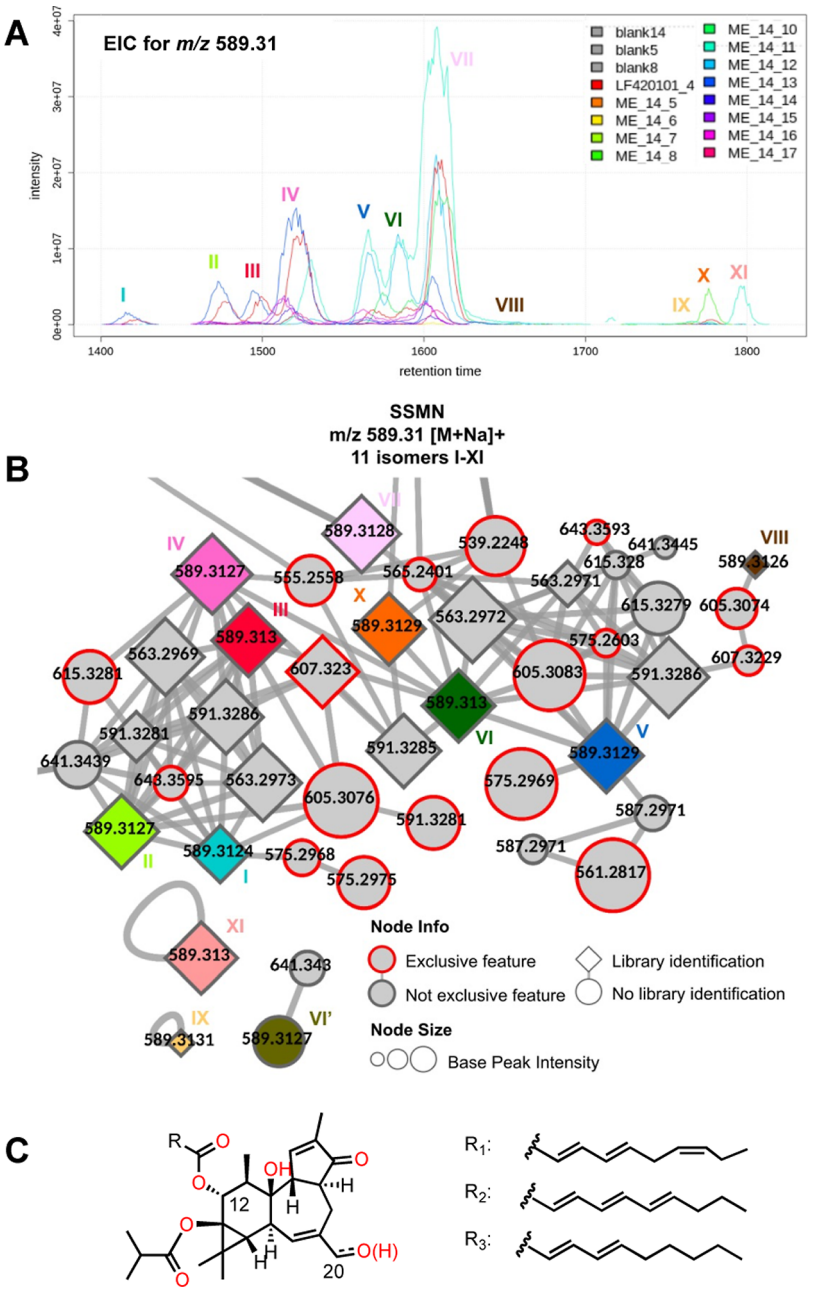


Figure 2. Accurate isomer separation by NP^{3}MS Workflow exemplified by the FBMN data set I. (A) EIC for m/z 589.31 highlights the presence of at least 11 isomers when all 16 samples from the *E. dendroides* data set are displayed (NP^3 MS Workflow MS^1 viewer command was used to plot the chromatogram). The number of isomers shown in the BPC matches the number of clean consensus spectra returned by the NP^3 MS Workflow, excluding the fragmented cluster of peak VI′. (B) Nine isomers, I−VIII and X, from m/z 589.31 had their clean consensus spectra clustered together in the same spectral cluster from the filtered SSMN, with 100 nodes (the figure focuses on 40 nodes). Two other isomers from peaks IX and XI were isolated, totalizing 11 isomers. A maximum component size of 150 would keep all of the isomers in the same component (data not shown). Nodes annotated with the GNPS spectral database are shown as diamonds; and nodes that were not annotated are represented by a circle. The node size is proportional to its MS^2 base peak intensity. (C) The molecules annotated with the GNPS MS^2 library for m/z 589.31 isomers are shown in three different arrangements of the 4-deoxyphorbol esters: isomer with R_1 corresponds to the GNPS spectrum CCMSLIB00000840552 and was matched with peaks V, VII, and VIII of spectra B and D; isomer with R_2 corresponds to CCMSLIB00006716184 and was matched with peaks I−VI and X of spectra A and C; and isomer with R_3 and 20-oxo corresponds to CCMSLIB00000840338 and was matched with peaks IX and XI of spectra E (Figure S7C).

which the use of the NP^3 MS workflow was essential for the early discovery of their bioactive components. Experimental proof-of-concept of the NP^3 results are also presented using traditional and orthogonal experimental methods.

Case Study #1: Isomer Separation. GNPS Data set I, made available by the GNPS-FBMN group,[20] consists of LC−MS/MS data (MSV000080502) from *Euphorbia dendroides* plant extract. The analysis focused on the target m/z 589.31, the 4-deoxyphorbol esters. Two isomers were isolated and characterized by NMR in the original paper[9] and deposited at GNPS, CCMSLIB00000840551:12b-*O*-[deca-2*Z*,4*E*,6*E*-trienoyl]-13-isobutyroyloxy-4b-deoxyphorbol and

https://doi.org/10.1021/acs.analchem.3c05829
Anal. Chem. 2024, 96, 7460−7469

CCMSLIB00000840552:12b-*O*-[deca-2*Z*,4*E*,7*Z*-trienoyl]-13-isobutyroyloxy-4b-deoxyphorbol). The EIC for ion *m*/*z* 589.31 in this data set shows the presence of at least 11 LC−MS peaks, with retention times between 1400 and 1800 s in all 16 samples (Figures 2A and S7A).

Two indicators were computed to assist the comparison of NP^3 MS Workflow to GNPS-FBMN: (1) the number of probable exclusive features and (2) the number of duplicated features or fragmented clusters. The first is the number of features (called clean consensus spectra in NP^3 MS Workflow), computed either by NP^3 MS Workflow or GNPS-FBMN within *m*/*z* and retention time tolerances. Isotopes are not counted. The second is the number of features that have the same *m*/*z* and retention time of another feature (within tolerance) in the results of each method. Neither indicator counts *m*/*z* from the blank samples.

Preprocessing was optimized focusing on the target *m*/*z*. Then, the NP^3 MS Workflow was run with similar parameters adopted by the GNPS[9] for data processing (Supporting Information, Note S7).

The NP^3 MS Workflow results showed the presence of all 11 distinct isomers of *m*/*z* 589.31, annotated with the GNPS spectral database (job link). A fragmented cluster of peak VI, named peak VI′, was present in the NP^3 MS Workflow result, but its corresponding chemical structure was not annotated. In the filtered SSMN, nine isomers of *m*/*z* 589.31 were grouped together into the same spectral cluster (Figure 2B). Other 91 features were also observed in this spectral cluster, 39% of which were exclusively detected by NP^3 MS Workflow. Isomers IX and XI of *m*/*z* 589.31 were isolated in the filtered SSMN and represent a different MS^2 spectra type (Figure S7B). Peak VI′ is also in a distinct spectral cluster, with only one neighbor. The only exclusive feature identified by the NP^3 MS Workflow present in the *m*/*z* 589.31 spectral cluster that could be annotated was *m*/*z* 607.32 (Figure 2B). The signal at *m*/*z* 607.32 also appears only in NP^3 MS Workflow results. It was annotated as an analogue of 4-deoxyphorbol esters (Figures 2B and S10E, GNPS entry CCMSLIB00000840365), reflecting the validity of this clean consensus spectra and its annotation within the chemical family represented by this spectral cluster.

The MS^2 spectra for the precursor *m*/*z* 589.31, that were annotated in GNPS, present common fragment ions with different relative intensities that could establish four distinct spectral types (named spectra A, B, C and D—Figure S7B). Also, a fifth spectral type was observed with most intense fragment ions *m*/*z* having a 2 Da shift from the other spectral types (named Spectra E) that represent isomers IX and XI. This 2 Da difference is an indication of deoxyphorbol oxidation, which is a known behavior of this class of compounds.[45,46]

Regarding the filtered SSMN size and minor compound detection, the NP^3 MS Workflow resulted in a larger molecular network, with 1.9 times more nodes compared to FBMN (1413/765) and with 51.2% exclusive features. This and other NP^3 MS Workflow options that may influence the number of detected features are commented on in Supporting Information, Notes S7 and S8.

NP^3 MS Workflow achieved a good library annotation rate, equal to 5.1% (vs 5.4% in FBMN). This rate represents 1.8 times the number of unique annotations and 1.8 times the number of total annotations of FBMN results.[20] Together with a rate of duplicated consensus spectra equal to only 1.5% (vs 0.4% in FBMN) and a rate of single nodes equal to 35.32% (vs 44.71% in FBMN), it enables inferring that NP^3 MS Workflow kept the chemical diversity of the data set maintaining relevant features of good quality minor compounds. The distribution of the MS^2 base peak intensity and of the maximum peak area of NP^3 MS Workflow exclusive features also corroborate this statement (Supporting Information, Note S8). A full comparison of both methods plus an analysis of the noise cutoff impact to NP^3 MS Workflow results are listed in Table S1 and Figure S10. Analysis of the noise cutoff and chemical rules impact on the annotations of *m*/*z* 589.31 isomers in the IVAMN are in Supporting Information, Note S9. Case study #2 (Supporting Information Note S10) provides evaluation of IVAMN and chemical rule impact on $[M + H]^+$ ion selection by NP^3.

Case Study #3: Pointing Bioactive Natural Products in Complex Mixtures—*Streptomyces* sp. BRA-346 Data Set. A robust LC−MS/MS data processing is essential for metabolomic data mining. In natural product-based drug discovery projects, deconvoluting the bioactive compound from a mixture of natural products (i.e., extracts or enriched fractions) is one of the main challenges that can be assisted by untargeted metabolomics and was one of our main goals while developing the NP^3 MS Workflow. To evaluate the Biocorrelation step of NP^3 MS Workflow (step 9), we used two in-house data sets: *Streptomyces* sp. BRA-346 and *A. moriforme* MA9.

The BRA-346 data set consists of LC−MS/MS data from *Streptomyces* sp. BRA-346 extract and enriched fractions (MSV00091453). *Streptomyces* sp. BRA-346, isolated from a Brazilian endemic tunicate,[17] was reported to produce the epoxy-β-aminoketones TMC-86A, dihydroeponemycin, and eponemycin and analogues that exhibited high cytotoxicity against glioma cell lines through proteasome inhibition.[17,36]

Streptomyces sp. BRA-346 extract was fractionated, resulting in 96 fractions. The extract and derived chromatographic fractions were screened for the inhibition of the 20S proteasome core particle.[17,36] Fractions 57−58 and 83−93 showed the highest inhibition (Figure S16). Initially, only the first inhibition peak was used for computing the biocorrelation. The metadata table was constructed with the LC−MS/MS data and the inhibition scores for the extract and fractions 54−63 (Table S2).

NP^3 MS Workflow was executed for BRA-346 data set using the parameters presented in Supporting Information, Note S13. The BCCs were computed for the clean consensus spectra regarding proteasome inhibition, being filtered to only keep the *m*/*z* ions that did not appear in the blank or culture media samples. Then, the clean consensus spectra were ranked according to the BCCs (Table 1). One clean consensus spectrum was ranked with the highest value: *m*/*z* 343.187 (BCC 0.80).

The *m*/*z* ion 343.187 (rt 4.24 min) was detected only in the bioactive fractions of peak 1 (Figure 3), and its peak area follows the bioactivity value of the fractions (Figure S17). In the molecular networks, the node of the $[M + H]^+$ ion *m*/*z* 343.187 exhibited the colors of the bioactive fractions (Figure 3A,B); most of its area was painted in red (highest bioactive fractions 57 and 58) and the remaining area being painted orange (moderated bioactive fractions 56 and 59) or black (BRA-346 extract). The yellow color, that refers to the inactive fractions (fractions 54, 55, 60, 61, 62, and 63), is not visible. This reflects that the fractions that contained *m*/*z* 343.187

https://doi.org/10.1021/acs.analchem.3c05829
Anal. Chem. 2024, 96, 7460−7469

Table 1. Ranking of the Candidate $[M + H]^+$ Ions in *Streptomyces* sp. BRA-346 Samples, According to the NP^3 BCC for the Observed 20S Proteasome Inhibition[a]

m/z consensus	rt mean	BCC
343.1870	4.24	0.8072
218.1400	3.70	0.7439
366.1779	4.95	0.6536
433.2293	4.24	0.6307
408.9794	0.39	0.5986

[a]rt: retention time; BCC: COR_ALL_INHIBITION.

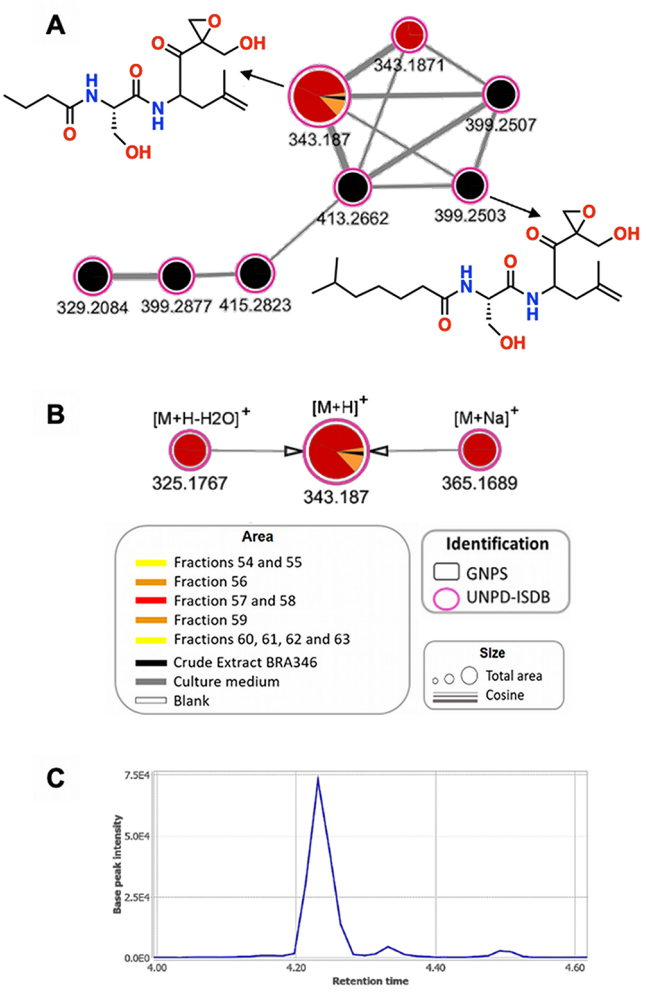


Figure 3. Selection of TMC-86A (*m/z* 343.187) as the bioactive compound in *Streptomyces* sp. BRA-346 data set using the NP^3 MS Workflow. (A) Filtered SSMN for TMC-86A and analogues present in *Streptomyces* sp. BRA-346 extract and derived chromatographic fractions. (B) IVAMN for TMC-86A. (C) Extracted ion chromatograms (EIC) of *m/z* 343.18 in fraction #58.

were bioactive (inhibited the 20S proteasome) and the fractions that did not contain this *m/z* were inactive.

Regarding library annotation, Tremolo showed 10 suggestions of compounds from the UNPD database for *m/z* ion 343.187, based on its MS^2 fragmentation. The suggested compounds were ranked according to the highest Tremolo score and the lowest *m/z* error (Table S3). Among the suggestions, the compound with the lowest *m/z* error was TMC-86A (MQscore 0.27, *m/z* error 1.87 ppm), a bioactive epoxy-*β*-aminoketone known as a specialized metabolite produced by *Streptomyces*,[47] including BRA-346.[17,36]

The clean consensus spectrum of *m/z* ion 343.187 exhibited fragment ions of *m/z* 325.176, 251.103, 186.113, 168.102, 150.091, and 130.050 (Figures S18A and S19), which match the experimental spectra of isolated TMC-86A[48,49] (Figure S18C), supporting the initial structure annotated by Tremolo using the UNPD-ISDB database. Although TMC-86A was not ranked with the highest score, its lowest *m/z* error was a determining factor among Tremolo suggestions. This suggests that for high-resolution mass spectrometers, *m/z* error should be considered for the selection of annotated chemical structures, in conjunction with the Tremolo score (MS^2 spectra match). Importantly, the MS^2 spectral annotations, in particular with in silico databases such as UNPD-ISDB, should be used as a guide for structure annotations and must be followed by experimental validations.

For this case study, we have further compared the experimental MS^2 spectra of our putative TMC-86A with experimental literature data for TMC-86A[48,49] (Figure S18C). In addition, as a proof-of-concept of TMC-86A being produced by BRA-346 and acting as a proteasome inhibitor, we have also sequenced the genome of BRA-346 and found the biosynthetic gene cluster (BGC) producing TMC-86. This BGC was cloned and heterologously expressed, confirming the presence of TMC-86 in *Streptomyces* sp. BRA-346.[36] In addition to TMC-86 being a known proteasome inhibitor with IC_{50} value in the low micromolar range,[47] a duplicated copy of the bacterial proteasome gene was also found within the TMC-86A biosynthetic gene cluster in the *Streptomyces* sp. BRA-346 genome.[36] This further evidenced its proteasome inhibition capacity. Aiming to unequivocally define the chemical structure of *m/z* 343.187, we further purified the compound by traditional methods (Figures S29–S30) and confirmed its chemical structure using uni and bidimensional NMR analysis (Figures S33–38). The chemical shifts were compared to literature data, showing a perfect match with published TMC-86A[48,49] NMR data (Tables S7–S8). The isolated compound was further tested in concentration–response curves against proteasome enzymatic activity (Figure S39), showing IC_{50} values compatible with that previously reported.[47]

Furthermore, the top-ranked *m/z* 343.187 node was connected in the filtered SSMN to the nodes of $[M + H]^+$ ions *m/z* 343.187, 399.251, 399.250, and 413.266 (Figure 3A). The $[M + H]^+$ ion *m/z* 343.187 (rt = 4.50 min) was mostly detected in the bioactive fractions and showed the same mass fragmentation pattern as TMC-86A (Figures S18B and S19), an indication that it represents one of the possible diastereomers of this molecule. Other TMC-86A analogues were also detected in the BRA-346 extract (Figure 3A, in black), which also represent epoxy-*β*-aminoketone proteasome inhibitors produced by the same BGC.[36,49] These TMC-86A analogues, detected in the BRA-346 extract, are present in the second peak of activity (fractions 83–93), which were not used in this NP^3 MS Workflow analysis (more information in Supporting Information, Note S13).

Case Study #4: Pointing Bioactive Natural Products in Complex Mixtures—*A. moriforme* MA9 Data Set. The second in-house data set comprises the LC–MS/MS analysis of *A. moriforme* MA9 extract and derived chromatographic fractions (MSV000091455). The methanolic fraction of the *A. moriforme* MA9 extract was fractionated (series 1) and

https://doi.org/10.1021/acs.analchem.3c05829
Anal. Chem. 2024, 96, 7460–7469

screened for the 20S proteasome inhibition. Then, the most bioactive fractions (E−F) were fractionated (series 2), with EF2−3 showing the highest inhibition (Figure S40). The IC_{50} values for the *A. moriforme* MA9 samples are in Table S9. These data were used in the construction of the metadata file, which used two series with correlation applied for both series at the same time, assuming that series 2 was derived from series 1. The parameters for the NP^3 MS Workflow are presented in Supporting Information, Note S14.

The resulting BCC values for the clean consensus spectra, candidates to inhibit the proteasome, are shown in Table 2.

Table 2. Ranking the Candidate Ions in *A. moriforme* MA9 Samples, According to the NP^3 MS Workflow Correlation for the Observed 20S Proteasome Inhibition[a]

m/z consensus	rt mean (min)	BCC
679.2719	5.65	−0.8994
315.1349	3.97	−0.8875
257.1294	5.59	−0.8821
295.145	5.56	−0.8599
309.088	5.39	−0.8552

[a]rt: retention time; BCC: COR_FRACTIONS_IC_{50}.

Since IC_{50} values were used in the construction of the metadata, the biocorrelation values were inverse and the top candidates presented the lowest BCC (negative values in Table 2). Considering series 1 and 2, the *m/z* ion 679.272 was attributed with the lowest score (BCC = −0.90).

Extracted ion chromatograms for the ion at *m/z* 679.272 (rt 5.65 min) showed that it exhibited the highest intensities in the inhibitory fractions and was not detected in the inactive ones (Figure S41A). Regarding library annotation, the MS^2 fragmentation pattern of this ion had a match with the GNPS database, being annotated as TMC-95 (GNPS score 0.75 and *m/z* error 0.09 ppm). The clean consensus spectra of *m/z* 679.272 exhibited ion fragments of *m/z* 661.261, 622.214, 594.219, 549.198, and 295.107 (Figure S40), as expected for TMC-95 (Figure S41). TMC-95 are cyclic peptide metabolites produced by microorganisms that are known as proteasome inhibitors.[50]

The node of the clean consensus spectrum $[M + H]^+$ of *m/z* 679.2719 was marked with the color of the inhibitory fractions (red and orange) on the filtered SSMN of this data set (Figure 4B). The yellow color, which refers to the inactive fractions, was not displayed. Moreover, *m/z* 679.2719 was connected to the nodes of the clean consensus spectra $[M + H]^+$ of *m/z* 679.2732 (rt = 5.25 min) and 679. 2717 (rt = 5.50 min). These nodes were detected in the inhibitory fractions and showed the same fragmentation pattern of TMC-95 (Figure S42B,C). TMC-95 has four diastereomers (A to D)[50] what can explain the presence of three nodes in this cluster. Only one isomer, *m/z* 679.2719 (rt 5.65 min, represented by the largest node), was correlated to the inhibitory activity, and the others were attributed with worse biocorrelation values (−0.60 for *m/z* 679.2732 and −0.56 for *m/z* 679. 2717). For reference, TMC-95A and TMC-95B are low-nanomolar inhibitors of the proteasome, whereas TMC-95C and TMC-95D are 20−150 times weaker inhibitors.[50]

As a proof-of-concept of the NP^3 MS Workflow data mining results, the active compounds and annotated structures were confirmed by traditional bioguided isolation methods and IC_{50} curves. TMC-95A and TMC-95B were purified from MA9

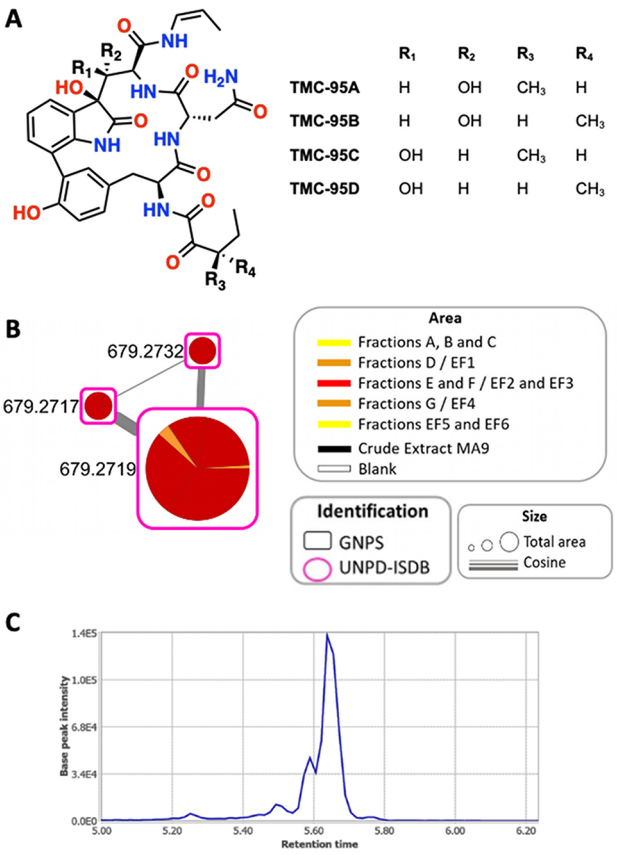


Figure 4. Selection of TMC95 as the bioactive natural product in *A. moriforme* MA9 fractions data set using the NP^3 MS Workflow. (A) TMC-95 structure and diastereomers A−D. (B) Filtered SSMN for the MA9 data set, highlighting the TMC-95 (*m/z* 679.27) spectral cluster. (C) EIC for *m/z* 679.27.

active fractions (fraction EF) and ^1H and ^{13}C NMR experiments were carried out (Figures S46−S47 and Tables S10−S12). The chemical shifts were compared to literature data, showing a perfect match with published[50] TMC-95A and TMC-95B NMR data. The isolated compounds were further tested in concentration−response curves against proteasome enzymatic activity, showing that both compounds isolated from *A. moriforme* MA9 active fraction EF (TMC-95A and TMC-95B) are low-nanomolar proteasome inhibitors (Figure S45), as expected.[50]

According to Figure S44C,D, isolated TMC-95A and TMC-95B elute as a double peak in the chromatogram (rt 5.6 min). Therefore, this peak in MA9 fraction EF2 (Figures S44B and 4C) represents a mixture of these two diastereomers, which are low-nanomolar proteasome inhibitors. The minor peaks (rt 5.25 and 5.5 min) would represent the less-potent TMC-95 diastereomers TMC-95C and TMC-95D. This is an example showing that the NP^3 MS Workflow is capable of detecting and attributing different BCCs even for diastereomers of a given bioactive compound, correctly selecting the most bioactive compounds from a complex mixture.

Figure S15 shows ionization variant profile distribution of the in-house data sets and percent of clean consensus spectra annotated in these data sets. The execution times for the NP^3 MS Workflow are reported in Supporting Information, Note S15.

https://doi.org/10.1021/acs.analchem.3c05829
Anal. Chem. 2024, 96, 7460−7469

CONCLUSIONS

We are now in an era in which very large prefractionated natural product libraries are screened (currently on the 10−100 thousand fractions and soon on the 1 million fraction scale[14]); therefore, it is common to find hundreds of hit samples after a screening campaign. It is impossible to isolate and perform the NMR structure elucidation of all these hits. The drug discovery team must prioritize which hits to follow. Rediscovery of known compounds, for example, is something that can be easily addressed by the NP^3 MS Workflow, directing the team to the discovery of the desired new and chemically relevant compounds for a given drug discovery project.

The case studies presented highlight that the NP^3 MS Workflow is a suitable and viable method for LC−MS/MS data processing and analysis. Optimization of the preprocessing step resolved adjacent isomers and prevented splitting large elution peaks into fragmented clusters. Blank separation and MS^1/MS^2 match allowed for accurate reduction of the MS^2 ions through the clustering methods. The use of the MS^2 base peak intensity for a quality check resulted in noise removal while keeping good quality minor compounds, what can be important for novel, very potent bioactive compound discovery. Structure annotation with UNPD-ISDB[24] showed that it covers a wide range of natural products for which no experimental data is available, serving as a guide for structural annotation of natural products in bioactive samples. The BCC score also showed a robust and readily useful approach to select bioactive natural products from complex mixtures, assisting decision making in natural product-based drug discovery projects. Finally, the molecular networks (SSMN and IVAMN) aid in LC−MS/MS data visualization and mining. The NP^3 MS Workflow may be easily installed and executed on personal computers, with the results automatically obtained within minutes for small data sets. The modularity of this pipeline further allows future upgrades to follow the developments on LC−MS/MS data processing and mining.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.analchem.3c05829.

Details and comparisons on methods and studied cases (PDF)

"dendroides_metadata": metadata table for GNPS-FBMN Data set I (CSV)

"MSV000084118_metadata": metadata table for GNPS-IIMN data set II (CSV)

"EDTA_metadata ": metadata table for GNPS-FBMN data set III (CSV)

AUTHOR INFORMATION

Corresponding Authors

Guilherme P. Telles − *Institute of Computing, University of Campinas (UNICAMP), Campinas 13083-852 State of São Paulo, Brazil;* orcid.org/0000-0003-2608-4807; Email: gpt@ic.unicamp.br

Daniela B. B. Trivella − *Brazilian Biosciences National Laboratory (LNBio), Brazilian Center for Research in Energy and Materials (CNPEM), Campinas 13083-970 State of São Paulo, Brazil;* orcid.org/0000-0002-7505-2345; Email: daniela.trivella@lnbio.cnpem.br

Authors

Cristina F. Bazzano − *Brazilian Biosciences National Laboratory (LNBio), Brazilian Center for Research in Energy and Materials (CNPEM), Campinas 13083-970 State of São Paulo, Brazil; Institute of Computing, University of Campinas (UNICAMP), Campinas 13083-852 State of São Paulo, Brazil*

Rafael de Felicio − *Brazilian Biosciences National Laboratory (LNBio), Brazilian Center for Research in Energy and Materials (CNPEM), Campinas 13083-970 State of São Paulo, Brazil*

Luiz Fernando Giolo Alves − *Brazilian Biosciences National Laboratory (LNBio), Brazilian Center for Research in Energy and Materials (CNPEM), Campinas 13083-970 State of São Paulo, Brazil;* Present Address: DiveGen, São Paulo, SP01451902, Brazil

Jonas Henrique Costa − *Brazilian Biosciences National Laboratory (LNBio), Brazilian Center for Research in Energy and Materials (CNPEM), Campinas 13083-970 State of São Paulo, Brazil;* Present Address: Baylor College of Medicine, Houston, TX77030, USA.; orcid.org/0000-0002-6415-3984

Raquel Ortega − *Brazilian Biosciences National Laboratory (LNBio), Brazilian Center for Research in Energy and Materials (CNPEM), Campinas 13083-970 State of São Paulo, Brazil; Institute of Biology, University of Campinas (UNICAMP), Campinas 13083-852 State of São Paulo, Brazil;* Present Address: University of Auckland, Auckland, 92019, New Zealand.

Bruna Domingues Vieira − *Brazilian Biosciences National Laboratory (LNBio), Brazilian Center for Research in Energy and Materials (CNPEM), Campinas 13083-970 State of São Paulo, Brazil*

Raquel Peres Morais-Urano − *Instituto de Química de São Carlos, Universidade de São Paulo, São Carlos CEP 13560-970 State of São Paulo, Brazil;* Present Address: CNPEM Campinas, SP13083970, SP, Brazil.

Luciana Costa Furtado − *Department of Pharmacology, Institute of Biomedical Sciences, University of São Paulo, São Paulo 05508-000 State of São Paulo, Brazil;* orcid.org/0000-0002-2306-7959

Everton L. F. Ferreira − *Instituto de Química de São Carlos, Universidade de São Paulo, São Carlos CEP 13560-970 State of São Paulo, Brazil;* Present Address: Univ.Federal do Vale do São Francisco, São Raimundo Nonato, PI6477000, Brazil.

Juliana R. Gubiani − *Instituto de Química de São Carlos, Universidade de São Paulo, São Carlos CEP 13560-970 State of São Paulo, Brazil*

Roberto G. S. Berlinck − *Instituto de Química de São Carlos, Universidade de São Paulo, São Carlos CEP 13560-970 State of São Paulo, Brazil;* orcid.org/0000-0003-0118-2523

Leticia V. Costa-Lotufo − *Department of Pharmacology, Institute of Biomedical Sciences, University of São Paulo, São Paulo 05508-000 State of São Paulo, Brazil;* orcid.org/0000-0003-1861-5153

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.analchem.3c05829

Author Contributions

CFB, GPT, and DBBT wrote the manuscript with contributions of all authors. All authors have given approval to the final

https://doi.org/10.1021/acs.analchem.3c05829
Anal. Chem. 2024, 96, 7460−7469

version of the manuscript. CFB, LFGA, and GPT developed the NP^3 MS Workflow codes with inputs and design from DBBT and RF. JHC and RF processed the biological data. LCF, JG, BDV, RPMU, EL, RGSB, and LCVL provided chemical samples for the in-house case studies. RO, BDV, and DBBT performed the proteasome inhibition analyses. DBBT and GPT designed and coordinated the project.

Notes

The authors declare no competing financial interest.

The *Streptomyces* sp. BRA-346 and the *A. moriforme* MA9 data sets are derived from Brazilian biodiversity. Sample collection and bioprospection authorizations were granted by the Biodiversity Authorization and Information System (SISBIO, number 48522−2 for the collection of tunicates, from which *Streptomyces* sp. BRA-346 was isolated) and the National System for the Management of Genetic Heritage and Associated Traditional Knowledge (SISGen), numbers AC0781C, AD99AF7 (BRA-346), and ABDEAEA (MA9).

■ ACKNOWLEDGMENTS

This research was funded by the Serrapilheira Institute, Serra-1709-19681 (DBBT); the São Paulo Research Foundation, 2014/10753-9 (DBBT), 2015/17177-6 (LVC-L), 2013/50228-8, 2015/01017-0, and 2019/17721-9 (RGSB), and CNPEM - through the Brazilian Ministry of Science, Technology and Innovation (MCTI) and the Brazilian National Fund for Scientific and Technological Development(FNDCT). Fellowships from the Coordination for the Improvement of Higher Education Personnel (RO), the National Council for Scientific and Technological Development (LCF-140146/2020-2, LVC-L −306913/2017-8, and GPT-317249/2021-5), the Serrapilheira Institute (CFB, LFGA, JHC), and FAPESP (LCF, 2017/18235-5 and 2020/08987-2) are acknowledged. CNPEM is also acknowledged for the use of the facilities, especially the Chemistry Laboratory (LQPN), Nuclear Magnetic Resonance (NMR) Laboratory, and the HPC-Tepui. Dr. M. L. Sforca, J. B. Guedes, and Dr. J. K. Rustiguel are also acknowledged for assistance with NMR analyses, TMC-86A purification, and biochemical assays. P. Ballone and H. Niero are also acknowledged for their feedback as the first users of NP^3 MS Workflow.

■ REFERENCES

(1) Price-Whelan, A.; Dietrich, L. E. P.; Newman, D. K. *Nat. Chem. Biol.* **2006**, *2* (2), 71−78.
(2) Davies, J. *J. Antibiot.* **2013**, *66* (7), 361−364.
(3) Newman, D. J.; Cragg, G. M. *J. Nat. Prod.* **2020**, *83* (3), 770−803.
(4) Pye, C. R.; Bertin, M. J.; Lokey, R. S.; Gerwick, W. H.; Linington, R. G. *Proc. Natl. Acad. Sci. U.S.A.* **2017**, *114* (22), 5601−5606.
(5) Dobson, C. M. *Nature* **2004**, *432*, 824−828.
(6) Bruder, M.; Polo, G.; Trivella, D. B. B. *Nat. Prod. Rep.* **2020**, *37* (4), 488−514.
(7) Berlinck, R. G. S.; Monteiro, A. F.; Bertonha, A. F.; Bernardi, D. I.; Gubiani, J. R.; Slivinski, J.; Michaliski, L. F.; Tonon, L. A. C.; Venancio, V. A.; Freire, V. F. *Nat. Prod. Rep.* **2019**, *36* (7), 981−1004.
(8) Beutler, J. A. *Curr. Protoc. Pharmacol.* **2009**, *46*, 9.11.1.
(9) Nothias, L.-F.; Nothias-Esposito, M.; da Silva, R.; Wang, M.; Protsyuk, I.; Zhang, Z.; Sarvepalli, A.; Leyssen, P.; Touboul, D.; Costa, J.; et al. *J. Nat. Prod.* **2018**, *81* (4), 758−767.
(10) Olivon, F.; Allard, P.-M.; Koval, A.; Righi, D.; Genta-Jouve, G.; Neyts, J.; Apel, C.; Pannecouque, C.; Nothias, L. F.; Cachet, X.; et al. *ACS Chem. Biol.* **2017**, *12* (10), 2644−2651.
(11) Yang, J. Y.; Sanchez, L. M.; Rath, C. M.; Liu, X.; Boudreau, P. D.; Bruns, N.; Glukhov, E.; Wodtke, A.; de Felicio, R.; Fenner, A.; et al. *J. Nat. Prod.* **2013**, *76* (9), 1686−1699.
(12) Gerwick, W. H.; Moore, B. S. *Chem. Biol.* **2012**, *19* (1), 85−98.
(13) Trivella, D. B. B.; de Felicio, R. *mSystems* **2018**, *3* (2), No. e00160.
(14) Grkovic, T.; Akee, R. K.; Thornburg, C. C.; Trinh, S. K.; Britt, J. R.; Harris, M. J.; Evans, J. R.; Kang, U.; Ensel, S.; Henrich, C. J.; et al. *ACS Chem. Biol.* **2020**, *15* (4), 1104−1114.
(15) Kurita, K. L.; Glassey, E.; Linington, R. G. *Proc. Natl. Acad. Sci. U.S.A.* **2015**, *112* (39), 11999−12004.
(16) Aminov, R. *Expert Opin. Drug Discovery* **2022**, *17* (9), 1047−1059.
(17) Furtado, L. C.; Bauermeister, A.; de Felicio, R.; Ortega, R.; Pinto, F. d. C. L.; Machado-Neto, J. A.; Trivella, D. B. B.; Pessoa, O. D. L.; Wilke, D. V.; Lopes, N. P.; et al. *Front. Mar. Sci.* **2021**, *8*, 644730.
(18) Géhin, C.; Holman, S. W. *Anal. Sci. Adv.* **2021**, *2* (3−4), 142−156.
(19) Bouslimani, A.; Sanchez, L. M.; Garg, N.; Dorrestein, P. C. *Nat. Prod. Rep.* **2014**, *31* (6), 718−729.
(20) Nothias, L.-F.; Petras, D.; Schmid, R.; Dührkop, K.; Rainer, J.; Sarvepalli, A.; Protsyuk, I.; Ernst, M.; Tsugawa, H.; Fleischauer, M.; et al. *Nat. Methods* **2020**, *17* (9), 905−908.
(21) Domingo-Almenara, X.; Montenegro-Burke, J. R.; Guijas, C.; Majumder, E. L. W.; Benton, H. P.; Siuzdak, G. *Anal. Chem.* **2019**, *91* (5), 3246−3253.
(22) Graça, G.; Cai, Y.; Lau, C.-H. E.; Vorkas, P. A.; Lewis, M. R.; Want, E. J.; Herrington, D.; Ebbels, T. M. D. *Anal. Chem.* **2022**, *94* (8), 3446−3455.
(23) 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.; et al. *Nat. Biotechnol.* **2016**, *34* (8), 828−837.
(24) Allard, P. M.; Péresse, T.; Bisson, J.; Gindro, K.; Marcourt, L.; Pham, V. C.; Roussi, F.; Litaudon, M.; Wolfender, J. L. *Anal. Chem.* **2016**, *88* (6), 3317−3323.
(25) Wang, M.; Bandeira, N. *J. Proteome Res.* **2013**, *12* (9), 3944−3951.
(26) Huber, F.; Ridder, L.; Verhoeven, S.; Spaaks, J. H.; Diblen, F.; Rogers, S.; van der Hooft, J. J. J. *PLoS Comput. Biol.* **2021**, *17* (2), No. e1008724.
(27) Allen, F.; Pon, A.; Wilson, M.; Greiner, R.; Wishart, D. *Nucleic Acids Res.* **2014**, *42* (W1), 94−99.
(28) Ruttkies, C.; Schymanski, E. L.; Wolf, S.; Hollender, J.; Neumann, S. *J. Cheminf.* **2016**, *8* (1), 3.
(29) de Felício, R.; Ballone, P.; Bazzano, C. F.; Alves, L. F. G.; Sigrist, R.; Infante, G. P.; Niero, H.; Rodrigues-Costa, F.; Fernandes, A. Z. N.; Tonon, L. A. C.; et al. *Metabolites* **2021**, *11* (2), 107.
(30) Männle, D.; McKinnie, S. M. K.; Mantri, S. S.; Steinke, K.; Lu, Z.; Moore, B. S.; Ziemert, N.; Kaysser, L. *mSystems* **2020**, *5* (3), No. e00125.
(31) Creamer, K. E.; Kudo, Y.; Moore, B. S.; Jensen, P. R. *Microb. Genomics* **2021**, *7* (5), 000568.
(32) Negrin, A.; Long, C.; Motley, T. J.; Kennelly, E. J. *J. Agric. Food Chem.* **2019**, *67* (19), 5687−5699.
(33) Maciá-Vicente, J. G.; Shi, Y.-N.; Cheikh-Ali, Z.; Grün, P.; Glynou, K.; Kia, S. H.; Piepenbring, M.; Bode, H. B. *Environ. Microbiol.* **2018**, *20* (3), 1253−1270.
(34) Schrimpe-Rutledge, A. C.; Codreanu, S. G.; Sherrod, S. D.; McLean, J. A. *J. Am. Soc. Mass Spectrom.* **2016**, *27* (12), 1897−1905.
(35) Bauermeister, A.; Mannochio-Russo, H.; Costa-Lotufo, L. V.; Jarmusch, A. K.; Dorrestein, P. C. *Nat. Rev. Microbiol.* **2022**, *20* (3), 143−160.
(36) Domingues Vieira, B.; Niero, H.; de Felício, R.; Giolo Alves, L. F.; Freitas Bazzano, C.; Sigrist, R.; Costa Furtado, L.; Felix Persinoti, G.; Veras Costa-Lotufo, L.; Barretto Barbosa Trivella, D. *Front. Microbiol.* **2022**, *13*, 786008.
(37) Frank, A. M.; Bandeira, N.; Shen, Z.; Tanner, S.; Briggs, S. P.; Smith, R. D.; Pevzner, P. A. *J. Proteome Res.* **2008**, *7* (1), 113−122.

https://doi.org/10.1021/acs.analchem.3c05829
Anal. Chem. 2024, 96, 7460−7469

(38) 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.; et al. *Proc. Natl. Acad. Sci. U.S.A.* **2012**, *109* (26), No. E1743−52.
(39) 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. *J. Nat. Prod.* **2021**, *84* (11), 2795−2807.
(40) Djoumbou Feunang, Y.; Eisner, R.; Knox, C.; Chepelev, L.; Hastings, J.; Owen, G.; Fahy, E.; Steinbeck, C.; Subramanian, S.; Bolton, E.; et al. *J. Cheminf.* **2016**, *8* (1), 61.
(41) van Santen, J. A.; Poynton, E. F.; Iskakova, D.; McMann, E.; Alsup, T.; Clark, T. N.; Fergusson, C. H.; Fewer, D. P.; Hughes, A. H.; McCadden, C. A.; et al. *Nucleic Acids Res.* **2022**, *50* (D1), D1317−D1323.
(42) Page, L.; Brin, S.; Motwani, R.; et al. *PageRank Citation Ranking: Bringing Order to the Web*; University of Pennsylvania, 1999.
(43) Shannon, P.; Markiel, A.; Ozier, O.; Baliga, N. S.; Wang, J. T.; Ramage, D.; Amin, N.; Schwikowski, B.; Ideker, T. *Genome Res.* **2003**, *13* (11), 2498−2504.
(44) Schmid, R.; Petras, D.; Nothias, L.-F.; Wang, M.; Aron, A. T.; Jagels, A.; Tsugawa, H.; Rainer, J.; Garcia-Aloy, M.; Dührkop, K.; et al. *Nat. Commun.* **2021**, *12* (1), 3832.
(45) Goel, G.; Makkar, H. P. S.; Francis, G.; Becker, K. *Int. J. Toxicol.* **2007**, *26* (4), 279−288.
(46) Haas, W.; Sterk, H.; Mittelbach, M. *J. Nat. Prod.* **2002**, *65* (10), 1434−1440.
(47) Koguchi, Y.; Kohno, J.; Suzuki, S.-I.; Nishio, M.; Takahashi, K.; Ohnuki, T.; Komatsubara, S. *J. Antibiot.* **2000**, *53* (1), 63−65.
(48) Zabala, D.; Cartwright, J. W.; Roberts, D. M.; Law, B. J. C.; Song, L.; Samborskyy, M.; Leadlay, P. F.; Micklefield, J.; Challis, G. L. *J. Am. Chem. Soc.* **2016**, *138* (13), 4342−4345.
(49) Huang, C.; Zabala, D.; de los Santos, E. L. C.; Song, L.; Corre, C.; Alkhalaf, L.; Challis, G. *Nucleic Acids Res.* **2023**, *51* (3), 1488−1499.
(50) Koguchi, Y.; Kohno, J.; Nishio, M.; Takahashi, K.; Okuda, T.; Ohnuki, T.; Komatsubara, S. *J. Antibiot.* **2000**, *53* (2), 105−109.

https://doi.org/10.1021/acs.analchem.3c05829
Anal. Chem. 2024, 96, 7460−7469