JOURNAL OF NATURAL PRODUCTS

Article

Cite This: J. Nat. Prod. 2018, 81, 758−767

pubs.acs.org/jnp

Bioactivity-Based Molecular Networking for the Discovery of Drug Leads in Natural Product Bioassay-Guided Fractionation

Louis-Félix Nothias,[†,‡,§,¶] Mélissa Nothias-Esposito,[§,∥,¶] Ricardo da Silva,[†,‡] Mingxun Wang,[†,‡] Ivan Protsyuk,[⊥] Zheng Zhang,[†,‡] Abi Sarvepalli,[†,‡] Pieter Leyssen,[▽] David Touboul,[§] Jean Costa,[∥] Julien Paolini,[∥] Theodore Alexandrov,[‡,⊥] Marc Litaudon,[§] and Pieter C. Dorrestein*[,†,‡]

[†]Collaborative Mass Spectrometry Innovation Center, University of California, San Diego, La Jolla, California 92093, United States

[‡]Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California, San Diego, La Jolla, California 92093, United States

[§]Institut de Chimie des Substances Naturelles, CNRS, ICSN UPR 2301, Université Paris-Sud, 91198, Gif-sur-Yvette, France

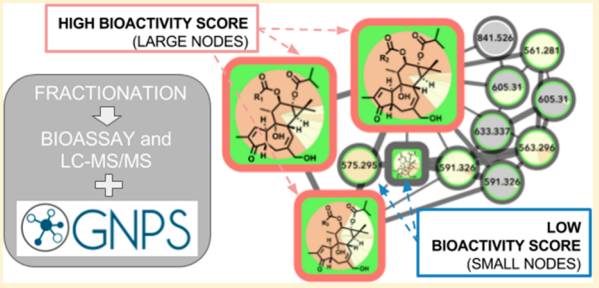
[∥]Laboratoire de Chimie des Produits Naturels, CNRS, UMR SPE 6134, University of Corsica, 20250, Corte, France

[⊥]European Molecular Biology Laboratory, EMBL, Heidelberg, Germany

[▽]Laboratory for Virology and Experimental Chemotherapy, Rega Institute for Medical Research, KU Leuven, 3000 Leuven, Belgium

Supporting Information

ABSTRACT: It is a common problem in natural product therapeutic lead discovery programs that despite good bioassay results in the initial extract, the active compound(s) may not be isolated during subsequent bioassay-guided purification. Herein, we present the concept of *bioactive molecular networking* to find candidate active molecules directly from fractionated bioactive extracts. By employing tandem mass spectrometry, it is possible to accelerate the dereplication of molecules using molecular networking prior to subsequent isolation of the compounds, and it is also possible to expose potentially bioactive molecules using bioactivity score prediction. Indeed, bioactivity score prediction can be calculated with the relative abundance of a molecule in fractions and the bioactivity level of each fraction. For that reason, we have developed a bioinformatic workflow able to map bioactivity score in molecular networks and applied it for discovery of antiviral compounds from a previously investigated extract of *Euphorbia dendroides* where the bioactive candidate molecules were not discovered following a classical bioassay-guided fractionation procedure. It can be expected that this approach will be implemented as a systematic strategy, not only in current and future bioactive lead discovery from natural extract collections but also for the reinvestigation of the untapped reservoir of bioactive analogues in previous bioassay-guided fractionation efforts.



Nature has inspired the discovery of numerous therapeutics from organisms such as microbes, plants, and insects.[1] The discovery of new bioactive natural products as leads for therapeutic development can be inspired by ethnopharmacological knowledge or achieved by screening a collection of extracts for bioactivity,[1−3] using in vitro, in cellulo, and even in vivo assays.[4] When a natural extract is found to be bioactive, a bioassay-guided fractionation is usually carried out and is generally comprised of the following steps: (i) extraction of metabolites from the biomass using solvents, (ii) fractionation of the resulting extract by chromatography, (iii) bioassay screening of each fraction, (iv) isolation of the molecule(s) from bioactive fractions, and (v) identification of the isolated molecules and evaluation of their bioactivity.[1,5,6] Although the bioassay-guided fractionation method seems to have been in use by chemists, pharmacologists, and toxicologists since the early 1900s,[7] it was formally conceptualized after the 1950s[7,8] and is still used today for the discovery of therapeutic leads by researchers in academia, government, and industrial laboratories worldwide.

Loss of activity or failure in isolating the bioactive compounds during the bioassay-guided fractionation workflow is a very common and costly pitfall in natural products isolation attempts. Generally, the unsuccessful outcome of these efforts are simply not published or are finally reported as investigations aiming at characterizing the chemical components. Some of the reasons for these pitfalls are (1) the bioactive compounds were degraded during the purification process, (2) the bioactive compounds were present in too low concentrations to be efficiently isolated, and (3) the bioactivity was due to synergistic effects between multiple compounds. It is therefore important to detect candidate bioactive molecules early in the

Received: August 28, 2017
Published: March 2, 2018

DOI: 10.1021/acs.jnatprod.7b00737
J. Nat. Prod. 2018, 81, 758−767

purification scheme in order to rationalize the isolation procedure applied toward these substances.

In addition, the application of bioassay-guided fractionation procedures in the last decades has resulted in the repeated isolation of molecules already described. To avoid the isolation of those known molecules, researchers have long used a preliminary structural assessment step called "dereplication". Dereplication aims at detecting known and bioactive compounds in natural product extracts and ideally prior to the start of in-depth investigation. In practice, dereplication is often performed when the molecule is at least partially, if not fully, purified.[9,10] Dereplication is generally achieved by first separating the molecular constituents using chromatography and then by looking at their spectral signatures obtained from mass spectrometry,[11−13] nuclear magnetic resonance,[14] or UV−vis spectroscopy, and then searching for compounds having identical properties (such as the molecular formula or core chromophores) in natural product databases[13] or in dedicated open source cheminformatics tools.[15] Some of the most commonly used natural product dereplication databases are "MarineLit",[16] "Antibase",[17] and the "Dictionary of Natural Products".[18] The use of untargeted tandem mass spectrometry to dereplicate known natural products has gained popularity, thanks to the introduction of a public spectral library and new bioinformatic tools.[13,19−22] The recent introduction of the Global Natural Product Social molecular networking (GNPS) Web-platform (http://gnps.ucsd.edu) enables the automatic spectral mining of large experiments with up to thousands of samples in a few hours.[22] In addition to the annotation of the molecules with reference MS/MS spectra publicly available on GNPS, it is also possible to propagate an annotation to an unknown molecule using the "analogues search mode" and MS/MS molecular networking, which expand and accelerate dereplication capabilities to similar structures.[22−25] Annotation can then be propagated using differences in molecular formulas of features. For example, a difference of 14 Da between two precursor ions would suggest a putative CH_2 increment, a difference of 16 Da on oxygenation, while a difference of 34 Da in the substitution of proton by a chlorine atom. Careful interpretation of the mass defects and the fragmentation patterns observed may even enable an investigator to decipher the part of the molecule that is modified. For this reason, GNPS and molecular networking have been used to prioritize biomass in in-depth investigations based on the observation of known and unknown analogues or of unique clusters of MS/MS spectra referred to as "molecular families".[22,24−27] However, because molecular networking on GNPS was designed for MS/MS spectral comparisons, this approach alone does not consider the relative abundance of the molecules between samples, which is crucial for downstream statistical analysis of data.

Bioactivity screening and LC-MS (i.e., without MS/MS acquisition) have been employed in order to (i) prioritize biomass selection[28] and (ii) speed up the isolation process for the discovery of bioactive molecules.[29−31] These studies used statistical approaches to estimate a bioactivity score, defined as the probability of a molecule as being bioactive. For each molecule, a bioactivity score can be calculated using its relative abundance between samples and the quantitative result of a bioassay for those same samples.[28] Multivariate statistical models can be used to identify bioactive compounds from LC-MS data.[29] Especially, partial least squares (PLS) regression has been shown to be an effective method to detect the association between the spectral data of a molecule and the result of a bioassay.[29] However, although the above-mentioned approaches can provide the molecular formula of the predicted bioactive moieties, they are unable to provide the high-throughput annotation that GNPS and molecular networking offer. Recently, molecular networking was introduced for the prioritization of bioactive natural extracts in large collections[26,32,33] and for the annotation of bioactive fractions.[34−36]

In the present work, this gap is bridged by providing a workflow procedure, named "*bioactive molecular networking*", which integrates MS/MS molecular networking and bioactivity scoring to assist in bioassay-guided fractionation. By highlighting molecular families and putative bioactive candidate molecules, prior to undertaking a time-consuming isolation step, the use of bioactivity scoring and molecular networking can accelerate natural-products-based drug discovery.[24] Herein, we demonstrate the utility of bioactive molecular networking by reinvestigating the data obtained in our previous study.[37] In this previous study, the chromatographic bioactive fractions from *Euphorbia dendroides* L. (Euphorbiaceae) latex, a Mediterranean tree spurge, were investigated to elucidate the components responsible for the potent and selective inhibition of the initial extract against chikungunya virus (CHIKV) replication. Despite a phytochemical effort that led to the isolation and the identification of nearly 20 molecules from the most bioactive fractions, no compound with selective and significant inhibition activity against CHIKV replication was found. By reinvestigating *E. dendroides* samples using bioactive molecular networking, it was possible to better capture the untapped nature of molecules associated with the bioactivity. The findings were validated by conducting a mass spectrometry targeted purification, and it was demonstrated that the new diterpene esters isolated in trace amounts (isolation yield from the extract ranging from 0.04% to 0.28%) were indeed active against the CHIKV replication in a cell-based bioassay. The workflow we have developed relies on a combination of open bioinformatic tools, including MZmine2[38] and Optimus workflow[39] (using OpenMS[40]), and the GNPS Web-platform. The approach can be readily adopted by researchers in the field of natural products drug discovery and also adapted to metabolomics and chemical ecology applications with other quantitative variables instead of the bioactivity level.

RESULTS AND DISCUSSION

Development of the Workflow Procedure. In the context of a bioassay-guided fractionation study, the employment of the bioactive molecular networking workflow procedure (Figure 1) required first the performance of an untargeted LC-MS/MS analysis and the evaluation of the bioactivity level for each chromatographic fraction. The bioactivity molecular networking consisted of three steps (I−III). The first step (I) was LC-MS/MS spectral data processing with the Optimus workflow,[41] based on OpenMS algorithms,[40] or with the MZmine2 toolbox.[38] This step allowed the detection and relative quantification of LC-MS/MS spectral features (ions) across the chromatographic fractions. The second step (II) consisted in the calculation of a bioactivity score using the Pearson correlation between feature intensity across samples and the bioactivity level associated with each sample. The needed script was prepared as a Jupyter notebook using R language.[42] Jupyter notebook (http://jupyter.org/)[43] is an open-source Web application that facilitates reproducibility and does not require advanced computational skills to be

DOI: 10.1021/acs.jnatprod.7b00737
J. Nat. Prod. 2018, 81, 758−767

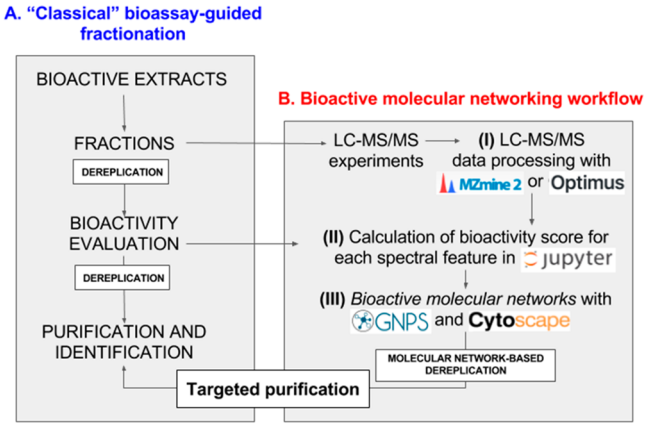


Figure 1. Bioactive natural products discovery pipelines. Workflow procedure (A) for the classical bioassay-guided fractionation approach and (B) by integrating bioactive molecular networking.

reemployed. The third and final step (III) was achieved by analyzing MS/MS data on the GNPS Web-platform[22] and visualizing the corresponding molecular networks in Cytoscape.[44] GNPS is a Web-platform that serves both as a data repository and a data analysis infrastructure with molecular networking and spectral library search capabilities. In recent years, this crowd-sourced platform has become an essential toolbox for both the natural products and metabolomics scientific communities. Indeed, public spectral libraries facilitate the dereplication of known molecules, and molecular networks allow annotation propagation to unknown related molecules. The objective of the last step of this workflow procedure was to annotate detected molecules using both spectral library annotations and the network topology, in particular those molecules with significant predicted bioactivity scores. For this purpose, the data curation relied on two complementary approaches. First a targeted approach, where prior knowledge can be used, such as chemotaxonomic information and the set of previously described bioactive molecules. Second, an untargeted approach can be undertaken by looking for consistent patterns of bioactive candidates in the networks, such as clusters with a high frequency of bioactive candidates, which could translate to the presence of a bioactive pharmacophore.

Evaluation of the Workflow Procedure. To evaluate the efficiency of bioactive molecular networking, an extract of *Euphorbia dendroides* latex was reinvestigated. This plant has documented ethnobotanical and ethnopharmacological uses to treat warts since the days of Ancient Rome.[37,45] Previous bioactivity screening of a Corsican specimen of *E. dendroides* showed that the latex extract displayed potent antiviral activity against CHIKV replication in a cell-based assay (Vero cells, green monkey kidney).[46,47] However, none of the compounds isolated during subsequent bioassay-guided fractionation of this *E. dendroides* extract showed selective antiviral activity against CHIKV.[37,48] Some of its fractions (F12, F13, and F17) exhibited a selective inhibition of CHIKV replication with an

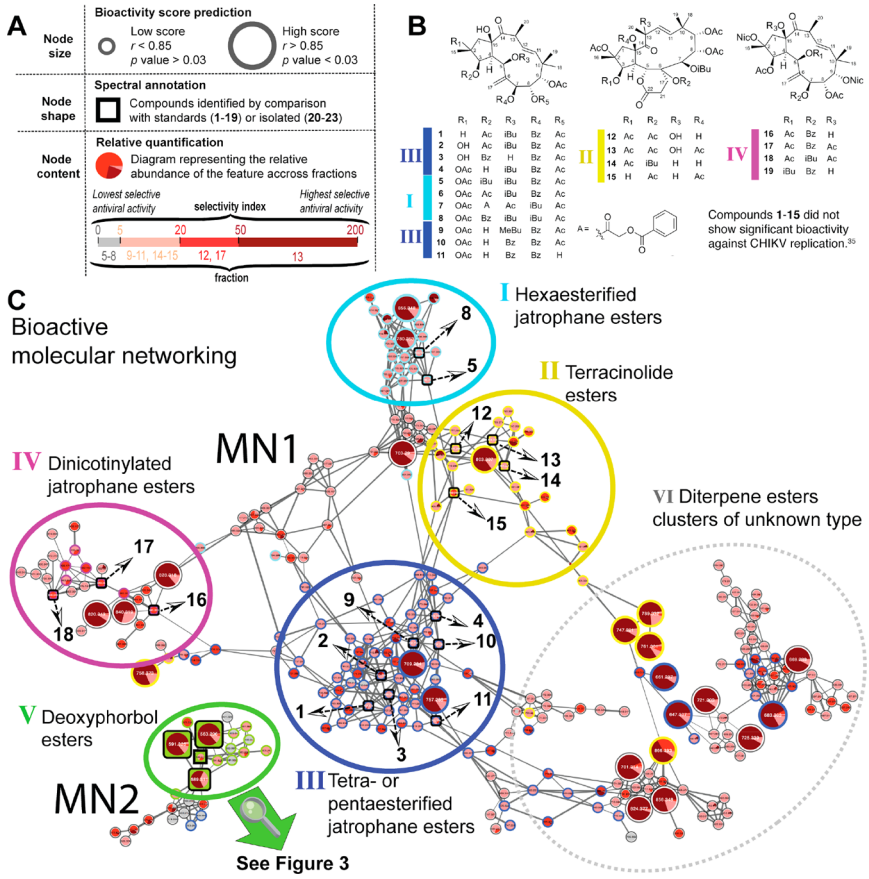


Figure 2. Bioactive molecular networking of bioassay-guided fractionation of *Euphorbia dendroides* latex extract. (A) Legend of bioactive molecular networks. (2) Previously isolated molecules (**1**–**19**). (C) Bioactive molecular network 1 (MN1) and 2 (MN2) of *E. dendroides* latex extract and cluster annotations (I–V). See Figure 3 for a detailed view for the cluster V of molecular network 2 (MN2).

DOI: 10.1021/acs.jnatprod.7b00737
J. Nat. Prod. 2018, 81, 758–767

effective concentration (EC_{50}) below 1.2 μg/mL (EC_{50} = 0.27, 0.17, and 1.13 μg/mL, respectively) and high selectivity indices of 41, 140, and 57, respectively. (EC_{50} is defined as the concentration that reduces the replication of the CHIKV by 50%, and the SI as the ratio EC_{50}/CC_{50} where CC_{50} is the cytotoxic concentration of the compound to reduce natural growth of host cells by 50%.) Nineteen diterpenoid jatrophane esters (**1**−**19**), including four terracinolides (**12**−**15**), were isolated from these bioactive fractions, but none of these molecules showed significant selective inhibition of CHIKV replication.[37] Thus, the nature of the compounds responsible for antiviral activity has not yet been elucidated. The assumption was made that the bioactive molecule(s) either occurred in trace amounts or were degraded, or that multiple molecules were required for activity (synergistic activity). To find the molecules with anti-CHIKV replication inhibition property, the extract and fractions were analyzed by LC-MS/MS and the presently described workflow was applied.

A crucial factor to predict a significant bioactivity score is to have multiple samples of known concentration that could have different, but related, molecular profiles with corresponding associated bioactivity levels. These could be chromatographic fractions of a single extract or multiple extracts from different but related sources. In this study, 18 fractions were obtained from a chromatographic separation of the extract of *E. dendroides* latex (F1−F18), and these fractions were subjected to LC-MS/MS and anti-CHIKV bioassays. Their mass spectrometric data, along with the results of their selective inhibition of CHIKV replication, were then processed with our procedure using the Optimus workflow, and bioactive molecular networks were generated in order to prioritize candidate molecules responsible for such bioactivity.

The bioactive molecular networks for the bioassay-guided fractionation of *E. dendroides* fractions are represented in Figures 2 and 3. As depicted in the legend (Figure 2A), the result of bioactivity scoring is visualized directly in the molecular network with the node size proportionally representing the molecule predicted bioactivity score. The bioactivity score was defined herein as the Pearson correlation coefficient (r) between the molecule relative abundance derived from the LC-MS peak (area under the curve) and the selectivity index (SI) value obtained from the evaluation of fraction antiviral activity. The largest nodes represent molecules with statistically significant bioactivity score [strong Pearson correlation value (r > 0.85) and a significance (FDR corrected, p value <0.03) after Bonferroni correction]. In addition, the relative abundance of detected molecule in each fraction can be visualized in each node with a pie chart diagram [colors correspond to the fraction bioactivity level]. In total, 8.4% (24/285) of molecules were predicted as having a statistically significant bioactivity score. These bioactive candidates represented 10.7% (3/28) of the nodes in MN1 and 8.2% (21/257) of the nodes in MN2.

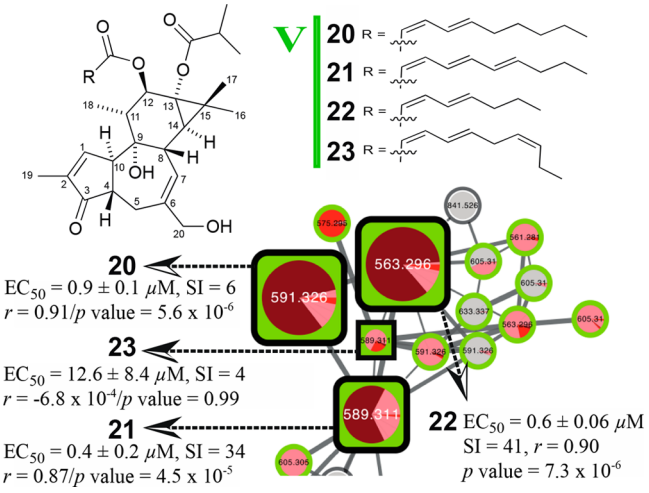


Figure 3. View of bioactive molecular networking results for the cluster V of molecular network 2 (MN2). The molecules **20**−**22** with significant bioactivity scores (large nodes, see legend) were annotated as deoxyphorbol ester derivatives and isolated in the present study. Evaluation of the biological activity indicated that **21** and **22** are selective inhibitors of chikungunya virus replication in the submicromolar range. The deoxyphorbol ester **23** was also isolated and found to have a lower selective antiviral activity in comparison to compounds **20**−**22**.

A detailed examination of the molecular networks allowed the annotation of two main molecular families (Figure 2C). The first molecular family (MN1) was composed of the known diterpene esters (**1**−**15**), including jatrophane esters (**1**−**11**) belonging to the 14-oxojatropha-6(17),11*E*-diene type and terracinolides (**12**−**15**). These molecules did not show strong anti-CHIKV activity in a previous study[49] and were readily annotated (square node in the molecular networks) because their spectral data were available in the GNPS public spectral library. We also observed that the aggregation of reference compounds in clusters I−IV was based on two main structural features: first, the number of esters such as hexa-esterified jatrophane derivatives (cluster I) versus penta- and tetra-esterified jatrophane derivatives (clusters III) or the presence of specific esters such as nicotinoyl esters (cluster IV) and the presence of a lactone for the terracinolides (cluster II).

In addition, it was possible to expand the annotation of MN1 to unknown related molecules using the "analogues search function" available on GNPS. Compounds that were annotated as analogues are represented with a circular colored node, with the color being representative of diterpene-type matching. Interestingly, it was observed that analogues were mostly found in the same clusters as the reference matching spectra. Reference compounds belonging to each cluster of MN1 (I−IV) were previously evaluated against CHIKV replication,[37,48] and none of them showed inhibition activity. For the above reasons, it was assumed that the analogues present in clusters I−V were not the most promising bioactive candidates, and subsequent efforts were focused on MN2, for which no candidate was previously isolated.

The molecular network 2 contained three spectral features with significant bioactivity score at m/z 591.326 (**20**), 589.311 (**21**), and 563.296 (**23**) and at the retention times 1626 s, 1520 s, and 1291 s, respectively. Still, no spectral annotation could be retrieved from public spectral libraries on GNPS, including using analogue searching. An available in silico annotation tool for molecular networking was used.[50] The results indicated heterogeneous candidate structures, including many unlikely candidates from a chemotaxonomic standpoint, which were not considered further. Examination of spectra showed that the ion at m/z 591.326 was a sodium adduct corresponding to a molecular formula of $C_{34}H_{48}O_7$ (m/z_{calc} 591.3298, $[M + Na]^+$, $\Delta m/z$ = 3.0 ppm). This observation explained the result of the in silico annotation tool, which considers all ions to be protonated adducts.[50] A search for molecules with the same

DOI: 10.1021/acs.jnatprod.7b00737
J. Nat. Prod. 2018, 81, 758−767

molecular formula was carried out in the Dictionary of Natural Products[18] and indicated a diterpene diester isolated from *Euphorbia prolifera*[51] named Euphorbia factor Pr4 [12-*O*-(2,4-decadienoyl)-13-*O*-(2-methylpropanoyl)-4-deoxyphorbol]. The examination of the MS/MS fragmentation pattern of this diterpenoid (Figure S3, Supporting Information) showed two neutral losses of 88 Da $[M - 88 + Na]^+$ and 168 Da $[M - 168 + Na]^+$, indicating that the detected molecules possessed a C_4H_7O acyl (potentially an isobutyryl or a butyryl) and a $C_{10}H_{17}O$ acyl moiety (potentially a decadienyl). Moreover, the observations of characteristic diterpene backbone fragment ions at *m/z* 313, 295, and 285 were supportive of this molecule being indeed of the phorboid diterpene ester, of either deoxyphorbol or ingenol type.[47]

The annotation of these diagnostic bioactive Euphorbiaceae diterpenoids[52−54] was of significance, as some phorboid derivatives are well-known inhibitors of human immunodeficiency virus replication,[46,55] and recently also against CHIKV,[56,57] due to their protein kinase C (PKC) modulatory ability. Thus, based on (i) the significant bioactivity score in MN2 for the nodes at *m/z* 591.326, 589.311, and 563.296 (compounds **20**−**22**, with respectively retention times at 1630, 1523, and 1289 s, an *r* value of 0.91, 0.87, and 0.90, and *p* values of 5×10^{-6}, 4.5×10^{-5}, and 7.3×10^{-6}) and (ii) the molecular network-based annotation indicating that they were likely phorboid diterpene esters, it was decided to perform a targeted isolation of these putative bioactive molecules. Inspection of the pie-chart diagram showed that the concentration of compounds **20**−**22** was higher in F13, and this fraction was hence selected for targeted isolation. In addition, it was proposed to isolate compound **23** (*m/z* 589.31, with a retention time 1609 s), because it was likely a positional isomer of compound **21** with low predicted bioactivity score ($r \simeq 0$ and *p* value 0.99), and it was thus valuable to test the approach.

The residual sample from the previous bioassay-guided fractionation of F13 (sub-F13) was fractionated by preparative HPLC using the UV trace as a proxy to the LC-MS/MS analysis. The targeted compounds **20**−**23** were purified using preparative HPLC and were then identified by 1D- and 2D-NMR data analysis (Tables 1 and 2). As suggested by molecular networking analysis (Figure 3), these four new compounds were indeed 4*β*-deoxyphorbol esters (**20**−**23**) (4-dPEs). The absolute configuration of each stereocenter was assumed based on prior studies on diterpenes of *Euphorbia* species (X-ray diffraction and total synthesis).[54,58] Following their identification, the reference MS/MS spectra of compound **20**−**23** were uploaded to the GNPS library and an analogue search was performed on molecular network MN2. Figure 3 showed that MN2 was essentially composed of deoxyphorbol ester analogues (nodes colored in green). The four isolated 4-dPEs (**20**−**23**) were evaluated for antiviral activity against CHIKV replication (Table S26, Supporting Information), and compounds **21** and **22** were found to be the most potent and selective CHIKV replication inhibitors with EC_{50} values of 0.40 ± 0.02 *μ*M and 0.60 ± 0.06 *μ*M, and SI = 34 and 41, respectively. The low SI values for compounds **20** and **23** indicated that they are associated with high host-cell cytotoxicity (SI = 6 and 4, respectively). These four 4*β*-deoxyphorbol esters possess the same isobutyrate group at C-13, but differ in the nature of the C-12 acyl moiety. Compound **20** has a 2*Z*,4*E*-decadienoyl group at position C-12, while compound **22** has a 2*Z*,4*E*-octadienyl group. The two compounds **21** and **23** are indeed isobaric compounds with a 12-acyl substituent of deca-2*Z*,4*E*,6*E*-trienoyl and deca-2*Z*,4*E*,7*Z*-trienoyl, respectively. While being closely related to compounds **20**−**22**, compound **23** had a lower selectivity index against chikungunya virus replication (EC_{50} = 12.6 ± 8.4 *μ*M, SI = 4), in agreement with the prediction of the bioactivity scoring. This result was consistent with previously described structure−activity relationships for phorbol esters,[59] showing that the nature of the acyl chain residue modulates the antiviral activity. It was remarkable to observe that the bioactive molecules **20** and **22** had also isobaric counterparts at *m/z* 591.33 and 563.296 with low bioactivity scores predicted. These isobaric counterparts had almost identical MS/MS spectra but showed different retention times (1707, 1607, and 1406 s). They were annotated as positional isomers, or epimers, of compounds **20**, **21**, and **22**, as commonly found in the Euphorbiaceae.[54,60,61] In the context of bioactivity score prediction, it was observed that despite the isobaric counter-

Table 1. ^{1}H NMR Spectroscopic Data for Compounds 20−22 (300 MHz) and 23 (500 MHz) in CDCl3 at 300 K (δ_H in ppm, *J* in Hz)

position	20	21	22	23
1	7.54 br t (1.7)	7.54 br t (1.7)	7.54 br t (1.7)	7.54 br t (1.7)
4	2.44 td (9.5, 4.5)	2.44 td (9.5, 4.5)	2.44 td (9.5, 4.5)	2.46 td (9.5, 4.5)
5*α*	2.82 dd (18.0, 10.0)	2.82 dd (17.7 9.6)	2.82 dd (18.0, 10.0)	2.82 dd (18.0, 10.0)
5*β*	2.12 dd (18.0, 10.0)	2.12 dd (17.7, 8.1)	2.12 dd (18.0, 10.0)	2.13 dd (18.0, 10.0)
7	5.52 d (5.4)	5.52 d (5.4)	5.52 d (5.4)	5.52 d (5.4)
8	2.35 t (5.4)	2.35 t (5.4)	2.35 t (5.4)	2.35 t (5.4)
10	3.23 m	3.23 m	3.23 m	3.24 m
11	1.56 m	1.56 m	1.56 m	1.58 m
12	5.43 d (9.8)	5.44 d (9.6)	5.43 d (9.8)	5.43 d (9.8)
14	1.01 d (5.4)	1.01 d (5.4)	1.01 d (5.4)	1.01 d (5.4)
16	1.18 s	1.18 s	1.18 s	1.21 s
17	1.18 s	1.18 s	1.18 s	1.20 s
18	0.91 d (6.4)	0.91 d (6.4)	0.91 d (6.4)	0.90 d (6.4)
19	1.70 dd (2.4, 1.2)	1.70 dd (2.4, 1.2)	1.70 dd (2.4, 1.2)	1.70 dd (2.4, 1.2)
20	4.01 b rs	4.01 dd (19.5, 13.3)	4.01 br s	4.00 br s
OH-9	5.81 br s	5.81 br s	5.81 br s	5.85 br s
OH-20	5.27 br s	3.46 br s	5.27 br s	5.28 br s
OR-14				
1′	5.52 d (11.3)	5.56 d (11.1)	5.52 d (11.3)	5.53 d (11.3)
2′	6.56 t (11.3)	6.59 t (11.1)	6.56 t (11.3)	6.57 t (11.3)
3′	7.29 dd (15.6, 11.5)	7.35 dd (15.1, 11.1)	7.29 dd (15.6, 11.5)	7.31 dd (15.6, 11.5)
4′	6.07 ddd (15.6, 7.1, 6.8)	6.46 dd (14.8, 10.8)	6.07 ddd (15.6, 7.1, 6.8)	6.04 ddd (15.6, 7.1, 6.8)
5′	2.15 dd (13.6, 7.1)	6.20 dd (14.8, 10.8)	2.16 dd (14.9, 7.1)	2.92 m
6′	1.41 m	5.92 ddd (14.8, 7.1, 6.8)	1.45 dd (14.9, 7.4)	5.33 dd (11.2, 5.7)
7′	1.27 m	2.11 dd (14.6, 6.8)	0.90 t (7.4)	5.46 dd (11.2, 8.0)
8′	1.28 m	1.42 dd (14.6, 7.5)		2.02 m
9′	0.87 t (6.9)	0.90 t (7.5)		0.94 t (6.9)
OiBu-13				
2″	2.57 q (7.0)	2.57 q (7.0)	2.57 q (7.0)	2.57 q (7.0)
Me-3″	1.18 d (7.0)	1.18 d (7.0)	1.18 d (7.0)	1.18 d (7.0)
Me-4″	1.15 d (7.0)	1.15 d (7.0)	1.15 d (7.0)	1.15 d (7.0)

DOI: 10.1021/acs.jnatprod.7b00737
J. Nat. Prod. 2018, 81, 758−767

Table 2. ^{13}C NMR Spectroscopic Data for Compounds 20–22 (75 MHz) and 23 (125 MHz) in $CDCl_3$ at 300 K (δ_C in ppm, J in Hz)

position	20	21	22	23
1	160.1	160.1	160.1	160.1
2	136.5	136.5	136.5	136.5
3	210.1	210.1	210.1	210.1
4	44.4	44.4	44.4	44.4
5	29.8	29.8	29.8	29.8
6	142.1	142.1	142.1	142.1
7	126.7	126.7	126.7	126.7
8	42.2	42.2	42.2	42.2
9	77.9	77.9	77.9	77.9
10	54.4	54.4	54.4	54.4
11	42.7	42.7	42.7	42.7
12	76.2	76.2	76.2	76.2
13	65.1	65.1	65.1	65.1
14	35.8	35.8	35.8	35.8
15	26.1	26.1	26.1	26.1
16	23.9	23.9	23.9	23.9
17	16.9	16.9	16.9	16.9
18	15.2	15.2	15.2	15.2
19	10.4	10.4	10.4	10.4
20	67.7	67.7	67.7	67.7
OR-12	166.1	166.1	166.1	166.2
1′	115	113.4	115	115.0
2′	146.3	143.5	146.3	146.3
3′	127	128.1	127	127.0
4′	146.6	140.4	146.6	144.0
5′	33.2	124.2	35.4	31.1
6′	28.6	138.6	22.1	124.7
7′	31.6	32.8	13.8	134.1
8′	22.7	19.7		20.6
9′	14.2	11.3		14.2
OiBu-13				
1″	179.7	179.7	179.7	179.7
2″	34.4	34.4	34.4	34.4
Me-3″	18.8	18.8	18.8	18.8
Me-4″	18.7	18.7	18.7	18.6

parts being abundant (9.2E7, 3.0E8, and 1.2E8 vs 6.5E7, 1.9E8, and 6.7E7 for compounds **20**–**22**, respectively, in the latex extract), they did not have a significant bioactivity score.

Bioactive molecular networks have permitted the detection of bioactive molecules from an antiviral fraction of *E. dendroides* latex, for which bioactive constituents were unknown despite the carrying out of a previous bioassay-guided fractionation procedure.[49] The present work showed that bioactive molecular networking can accelerate dereplication and a rationalized isolation procedure in bioassay-based drug discovery. It has been shown that, prior to any time-consuming isolation procedure, bioactive molecular networking is able to (i) narrow down the number of bioactive candidates by 10;, (ii) speed up spectral annotation with MS/MS molecular networking, and (iii) discriminate even isomers as a result of LC-MS feature detection tools. The workflow procedure can be readily implemented and adapted for any bioassay-guided fractionation with LC-MS/MS data associated. The approach can be easily reproduced since it relies on open tools (Optimus or MZmine2), the GNPS Web-platform, and a Jupyter notebook. Yet, like any other statistical approach, the accuracy and utility of a bioactive molecular network depends on (i) the reproducibility of the bioassay procedure and (ii) the number of samples included in the data set. Furthermore, the annotation capability of MS/MS spectra depends on public MS/MS spectral libraries, which are expanding rapidly as a result of contributions from the scientific community.

Herein is presented the concept of bioactive molecular networking, and it can be adopted as a systematic strategy for the annotation of bioactive molecules in a complex plant extract. It was applied to the discovery of antiviral diterpenoids in a *E. dendroides* latex extract, an extract on which a classical bioassay-guided effort was previously unsuccessful. From bioactive molecular networking, two bioactive compounds (**21** and **22**) were isolated. Compounds **21** and **22** were found to be active against CHIKV replication at submicromolar concentrations. Chikungunya virus is an emerging virus that is causing massive outbreaks (1 million estimated cases in South America and in the Caribbean region in 2014)[62] and associated with high morbidity and for which no antiviral chemotherapy is available. The present results demonstrate that bioactive molecular networking can speed up the process of drug-lead discovery by revealing bioactive candidates and facilitating their dereplication and/or annotation at the very first step of any bioassay-guided purification procedure. For this reason, we expect that bioactive molecular networking can be a cornerstone concept/methodology for the discovery of bioactive molecules from natural product extracts.

To facilitate its acceptance by various communities, the bioactive molecular networking workflow is freely available (https://github.com/DorresteinLaboratory/Bioactive_Molecular_Networks) and can be readily applied or adapted without the need for advanced computational skills, as it relies on popular LC-MS feature detection (MZmine2 or Optimus),[38,39] accessible bioinformatics tools (Jupyter notebook),[63] and the crowd-sourced GNPS Web-platform (http://gnps.ucsd.edu).[22] Moreover, the workflow will directly benefit from the development of new modules and bioinformatics tools, in particular those for improved LC-MS feature detection[38,40] or spectral[64] and in silico annotation.[65−67]

Bioactive molecular networking has the retroactive potential to reinvestigate successful bioassay-guided fractionation or those where investigators failed to find the active components, as long as corresponding MS/MS data were acquired or can still be generated. Currently, while 352 studies mentioned "bioassay-guided" and "MS/MS" for the year 2016,[68] none have deposited MS/MS data in public repositories such as MetaboLights (http://www.ebi.ac.uk/metabolights/)[69] or GNPS-MassIVE (http://massive.ucsd.edu). Therefore, as the scientific community pushes toward transparency[70] and as mandated by many funding agencies, the availability of all MS/MS data in public repositories will increase (similar to genomic efforts decades ago) and will enable both the detection of a presently untapped reservoir of bioactive analogues and the optimization of bioactivity scoring.

Finally, since bioactive molecular networking is agnostic to the nature of the quantitative variable employed (here the value of the bioassay results), any other quantitative outcome parameters can be employed instead. Thus, the workflow described herein can be adapted for applications outside of natural products research. Among the potential use of the current workflow, it could be employed in microbial ecology to annotate compounds impacted by an environmental variable (such as the pH), or perhaps it could be used in exposomics to

DOI: 10.1021/acs.jnatprod.7b00737
J. Nat. Prod. 2018, 81, 758−767

annotate biomarkers associated with a level of exposure to a chemical contaminant.

EXPERIMENTAL SECTION

General Experimental Procedures. Optical rotations ($[\alpha]_D$) were measured with the sodium D line monochromatic light source in EtOH at 24 °C. The instrument used was an MCP 300 Anton Paar polarimeter. A PerkinElmer Spectrum BX FT-IR system was used to record IR spectra. UV spectra were measured in a 1 cm quartz tank using a Varian Cary 100 scan spectrophotometer. The 1D (^{1}H and ^{13}C) and 2D (COSY, HSQC, HMBC, and ROESY) NMR spectra were obtained in $CDCl_3$ for compounds **20**–**22** on a Bruker Avance 300 MHz instrument, and for compound **23** on a 500 MHz instrument, using a 1.7 mm microprobe. HPLC separation was performed on analytical C_{18} columns (Kromasil, 250 × 4.6 mm i.d., 5 μm, Thermo Scientific) and on semipreparative C_{18} columns for fractionation. A Waters autopurification system equipped with a binary pump (Waters 2525), a UV–vis diode array detector (190–600 nm, Waters 2996), and a PL-ELS 1000 ELSD Polymer Laboratory detector was used. Preparative HPLC separation was performed on a semipreparative C_{18} column (Kromasil, 250 × 10 mm; i.d. 5 μm). A Dionex autopurification system equipped with a binary pump (P580), a UV–vis array detector (200–600 nm, Dionex UVD340U), and a PL-ELS 1000 ELSD Polymer Laboratory detector was used. A hybrid linear ion trap/Orbitrap mass spectrometer (LTQ-XL Orbitrap, ThermoFisher Scientific, Les Ulis, France) was used for LC-MS/MS analysis. An ESI source operating in positive ionization mode was used. Mass spectra were acquired between m/z 150 and m/z 1000. In the full-scan mode, full width at half-maximum mass resolution of the Orbitrap mass analyzer was fixed at 30 000 for mass spectra and at 15 000 for tandem mass spectra. The data-dependent MS^n mode was used to monitor the one to three most intense ions with an exclusion duration of 40 s after eight repetitions. Instrumental parameters were set as follows: source voltage 5 kV, lens 1 voltage −15 V, capillary temperature 275 °C, gate lens voltage −35 V, capillary voltage 25 V, tube lens voltage 65 V. The CID parameters were set as follow: CE at 30% of the maximum and an activation time of 30 ms. HPLC was performed with an HPLC Ultimate 3000 system (Dionex, Voisins-le-Bretonneux, France) consisting of a degasser, a quaternary pump, an autosampler, a column oven, and a photodiode array detector. Separation was achieved using an octadecyl silica gel column (Sunfire, 150 × 2.1 mm × 3.5 μm; Waters, Guyancourt, France), equipped with a guard column. The column oven temperature was set at 25 °C. Elution was conducted with a mobile phase consisting of water + 0.1% formic acid (A) and CH_3CN + 0.1% formic acid (B), following the gradient 5% to 95% B in 40 min, then maintaining 100% B for 10 min at a flow rate of 250 μL/min. To enable the semiquantification of the detected ions, the samples were prepared at a concentration of 2.5 mg/mL in MeOH, and the injection volume was fixed at 10 μL. The LC-MS^n analysis of the CH_3CN latex extract, the fractions of *E. dendroides*, and the reference MS/MS spectra were deposited on GNPS (http://gnps.ucsd.edu), under the reference number MSV000079385.

Plant Material. The latex of *Euphorbia dendroides* was collected by L.-F.N. in August 2012 on 20 specimens at Ficaghjola Beach (Piana) in the western region of Corsica. Botanical identification was performed by L.-F.N., and a voucher specimen (LF-026) was deposited at the Herbarium of the University of Corsica (Laboratoire de Chimie des Produits Naturels, Corte).

Extraction and Isolation. The latex extract of *E. dendroides* (4.5 g) was subjected to a bioassay-guided fractionation protocol as described in a previous publication,[37] suggesting that F13 (1600 mg) contains phorboid diterpenes with significant bioactivity score predictions. The previous bioassay-guided isolation procedure on fraction F13 (EC_{50} = 0.17 μg/mL, SI = 140) allowed the isolation of euphodendroidin E (804 mg, with an isolation yield of 17% from the extract) and of the subfraction sub-F13 (486 mg). This subfraction was analyzed by LC-MS/MS, and ions of interest were annotated and compounds present purified by preparative HPLC (Kromasil C_{18}, isocratic H_2O–CH_3CN/MeOH (5/5) + 0.1% formic acid, 8:92 in 40 min) using the UV–vis detector as a reference between chromatographic gradients. This purification procedure afforded four new compounds, compound **20** (12.8 mg), **21** (5.0 mg), **22** (5.1 mg), and compound **23** (1.8 mg), with isolation yields from the extract of 0.28%, 0.11%, 0.11%, and 0.04%, respectively.

*12β-O-[Deca-2Z,4E-dienoyl]-13α-isobutyryl-4β-deoxyphorbol (**20**):* amorphous powder; $[\alpha]^{25}_D$ −2 (*c* 1, EtOH); UV (EtOH) λ_{max} (log ε) 262 (4.32), 207 (4.08) nm; IR ν_{max} 2925, 1706, 1634, 1164, 1131, 1082, 780 cm^{-1}; ^{1}H (300 MHz) and ^{13}C (75 MHz) NMR data, see Tables 1 and 2; HRESIMS m/z 591.3272 $[M + Na]^+$ (calcd for $C_{34}H_{48}O_7Na$, 591.3298) deposited in the GNPS spectral library, https://gnps.ucsd.edu/ProteoSAFe/gnpslibraryspectrum.jsp?SpectrumID=CCMSLIB00000840337#%7B%7D.

*12β-O-[Deca-2Z,4E,6E-trienoyl]-13α-isobutyryl-4β-deoxyphorbol (**21**):* amorphous powder; $[\alpha]^{25}_D$ +5 (*c* 1, EtOH); UV (EtOH) λ_{max} (log ε) 301 (4.04), 262 (4.10), 208 (4.10) nm; for ^{1}H (300 MHz) and ^{13}C NMR data, see Tables 1 and 2; HRESIMS m/z 589.3127 $[M + Na]^+$ (calcd for $C_{34}H_{46}O_7Na$, 589.3141); deposited in the GNPS spectral library, https://gnps.ucsd.edu/ProteoSAFe/gnpslibraryspectrum.jsp?SpectrumID=CCMSLIB00000840338#%7B%7D.

*12β-O-[Octa-2Z,4E-dienoyl]-13α-isobutyryl-4β-deoxyphorbol (**22**):* amorphous powder; $[\alpha]^{25}_D$ −1 (*c* 1, EtOH); UV (EtOH) λ_{max} (log ε) 262 (4.10), 207 (3.90) nm; for ^{1}H (300 MHz) and ^{13}C NMR data, see Tables 1 and 2; HRESIMS m/z 563.3010 $[M + Na]^+$ (calcd for $C_{32}H_{44}O_7Na$, 563.2985); deposited in the GNPS spectral library, https://gnps.ucsd.edu/ProteoSAFe/gnpslibraryspectrum.jsp?SpectrumID=CCMSLIB00000840336#%7B%7D.

*12β-O-[Deca-2Z,4E,7Z-trienoyl]-13α-isobutyryl-4β-deoxyphorbol (**23**):* amorphous powder; $[\alpha]^{25}_D$ +3 (*c* 1, EtOH); UV (EtOH) λ_{max} (log ε) 261 (3.79), 206 (3.76) nm; for ^{1}H and ^{13}C NMR data, see Tables 1 and 2; HRESIMS m/z 589.3167 $[M + Na]^+$ (calcd for $C_{34}H_{46}O_7Na$, 589.3141); deposited in the GNPS spectral library, https://gnps.ucsd.edu/ProteoSAFe/gnpslibraryspectrum.jsp?SpectrumID=CCMSLIB00000840339#%7B%7D.

Spectral Feature Detection. This step was performed with the Optimus workflow for feature detection, integration, filtering, and visualization (https://github.com/MolecularCartography),[39] based on using OpenMS algorithms.[40] The Optimus workflow was executed with the KNIME Analytics Platform version 3.2.1 on macOS 10.12. The workflow must be run with the option "Presence of MS/MS". The .MGF file generated by the Optimus workflow (version 1.2.0) that contains MS/MS spectra (*spectral_data_MS2.mgf*) for each feature was uploaded on GNPS, and a Data Analysis was performed as described below. The CSV table generated by Optimus (*feature_quantification_matrix.csv*) was used for both the calculation of the bioactivity score and the visualization of the relative abundance into the molecular networks. More details on parameters used for Optimus and OpenMS are provided in the Supporting Information. Note that alternatively the popular MZmine2 toolbox (release 2.30) can be used for the LC-MS feature detection step. The Dorrestein laboratory implemented the needed features in MZmine2.[38,71] See the release of MZmine 2.28 in July 2017 and the corresponding online documentation on GitHub (https://github.com/DorresteinLaboratory/Bioactive_Molecular_Networks). MZmine2 can now output similar files to those outputted by the Optimus workflow thanks to the option named "filter MS/MS for GNPS" and "Reset the peak number ID" in the peak list rows filter module and the MGF export module "Export for GNPS". The CSV table with relative quantitative information is exported using the CSV export module. Note that an offline multisoftware workflow procedure was recently developed for the integration of MZmine2 and molecular networking.[72] While the use of this workflow is theoretically compatible with the present bioactivity molecular networking workflow procedure, it has not been tested.

Molecular Networking. Tandem mass spectrometry molecular networks were created using the GNPS platform (http://gnps.ucsd.edu).[22] Data were first converted to the.mzML format with MSConvert.[73] The spectral information (MGF file) was generated by Optimus and uploaded on GNPS. This file was used to generate an

DOI: 10.1021/acs.jnatprod.7b00737
J. Nat. Prod. 2018, 81, 758–767

MS/MS molecular network using the GNPS Data Analysis workflow, with the MS-Cluster deactivated. The precursor ion mass tolerance was set to 0.01 Da and to a product ion tolerance of 0.0075 Da (allowing a maximum error of 25 ppm at m/z 300 and 12 ppm at m/z 600). The fragment ions below 10 counts were removed from the MS/MS spectra. MNs networks were generated using 12 minimum matched peaks and a cosine score of 0.7. Data were open and visualized using Cytoscape 3.4.0 software.[44] A force-directed layout modulated by cosine score factor was used for data visualizations. Data of LC-MS/MS analysis were deposited in the MassIVE Public GNPS data set (http://massive.ucsd.edu, MSV000079856). The molecular networking job on GNPS can be found at http://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=169c80d3192d4c29a6399729cdc8c9a4 and with analogue search http://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=ce2a564dbd704c0595494e04798b0233. The corresponding Cytoscape file is available as Supporting Information. The annotated MS/MS spectra were deposited in the GNPS spectral library under references CCMSLIB00000840337, CCMSLIB00000840338, CCMSLIB00000840336, and CCMSLIB00000840339 for compounds **20**–**23**, respectively.

Prediction of Bioactivity Score Significance and Mapping onto the Molecular Networks. An R-based Jupyter notebook,[43] called Bioactive Molecular Networks, was created and is available on GitHub, https://github.com/DorresteinLaboratory/Bioactive_Molecular_Networks, along with a step-by-step tutorial and the representative input/output files needed. The table of spectral features' intensity (CSV format) obtained with Optimus workflow was used in the analysis. Note that to enable LC-MS-based semiquantitation, and thus a reliable prediction of the bioactivity score, each sample has to be analyzed at a known concentration. If the concentration differs between samples, a normalization has to be applied prior to using the workflow. The bioactive molecular networks Jupyter notebook is composed of three steps: scaling samples by normalizing the intensity to the TIC,[74] calculating the Pearson correlation score (r) and its significance (r_pval) between each feature and the bioactivity value, as well as the multiple hypothesis testing correction (the Bonferroni correction), and outputting the node attribute table. This table is conveniently formatted for import into the Cytoscape software to map back the bioactivity score onto the corresponding MS/MS molecular network. For the visualization of bioactive molecular networks in Cytoscape, the *Select* function in Cytoscape was used to filter nodes based on their r and the FDR corrected p value. The style of these nodes is then bypassed with a larger node size. In the current study, only molecules having a bioactivity score above $r > 0.85$ and p value of <0.03 were represented as having a large node in the bioactive molecular networks.

Virus-Cell-Based Antialphavirus Assay. The protocol used was described in a previous publication.[49]

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jnatprod.7b00737.

Results of bioassays against CHIKV replication and 1D and 2D NMR spectra (PDF)

■ AUTHOR INFORMATION

Corresponding Author

*Tel: 858-534-6607. Fax: 858-822-0041. E-mail: pdorrestein@ucsd.edu.

ORCID

Louis-Félix Nothias: 0000-0001-6711-6719
David Touboul: 0000-0003-2751-774X
Julien Paolini: 0000-0002-3109-1430
Marc Litaudon: 0000-0002-0877-8234
Pieter C. Dorrestein: 0000-0002-3003-1030

Author Contributions

[¶]L.-F. Nothias and M. Nothias-Esposito contributed equally to this article.

Notes

The authors declare the following competing financial interest(s): P. C. Dorrenstein is an advisor for Sirenas, a company that employs molecular networking for the discovery of bioactive natural products from marine resources. The work in that company does not overlap with the work presented in this paper.

Reference MS/MS spectra of isolated compounds were deposited in the GNPS public spectral library (CCMSLIB00000840316 to CCMSLIB00000840340). LC-MS data have been deposited in the MassIVE repository http://gnps.ucsd.edu under accession number MSV000079856. The MS/MS molecular networks can be consulted at http://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=ce2a564dbd704c0595494e04798b0233. The corresponding Cytoscape file of the molecular networks and the code for bioactive molecular network Jupyter notebook are available at https://github.com/DorresteinLaboratory/Bioactive_Molecular_Networks.

■ ACKNOWLEDGMENTS

The authors would like to thank J.-F. Gallard (CNRS) for help with the NMR acquisition and I. Schmitz-Afonso (CNRS) for the LC-MS/MS analysis. We thank the NIH for supporting this work under NIH-UCSD Center for Computational Mass Spectrometry P41 GM103484 and the NIH grant on reuse of metabolomics data R03 CA211211. This work was also supported by an "Investissement d'Avenir" grant managed by Agence Nationale de la Recherche (CEBA, ANR-10-LABX-25-01). D.T. was supported by the Agence Nationale de la Recherche (Grant ANR-16-CE29-0002-01 CAP-SFC-MS). This project benefited from European Commission funding (Horizon 2020 Research and Innovation Program) from the Marie Skłodowska-Curie Action MSCA-IF-2016, 3D-Plant2-Cells, No. 704786 (to L.F.N), and from the grant agreement No. 634402 (to T.A. and I.P).

■ REFERENCES

(1) Newman, D. J.; Cragg, G. M. *J. Nat. Prod.* **2012**, *75*, 311−335.
(2) Kingston, D. G. I. *J. Nat. Prod.* **2010**, *74*, 496−511.
(3) Fabricant, D. S.; Farnsworth, N. R. *Environ. Health Perspect.* **2001**, *109*, 69−75.
(4) Potterat, O.; Hamburger, M. *Nat. Prod. Rep.* **2013**, *30*, 546−564.
(5) Colegate, S. M.; Molyneux, R. J. *Bioactive Natural Products Detection, Isolation, and Structural Determination*; CRC Press: Boca Raton, FL, 1993.
(6) Bucar, F.; Wube, A.; Schmid, M. *Nat. Prod. Rep.* **2013**, *30*, 525−545.
(7) Weller, M. G. *Sensors* **2012**, *12*, 9181−9209.
(8) Schneider, H. G.; Tener, G. M.; Strong, F. M. *Arch. Biochem. Biophys.* **1952**, *37*, 147−157.
(9) Wolfender, J.-L.; Marti, G.; Thomas, A.; Bertrand, S. *J. Chromatogr. A* **2015**, *1382*, 136−164.
(10) Gaudêncio, S. P.; Pereira, F. *Nat. Prod. Rep.* **2015**, *32*, 779−810.
(11) Henke, M. T.; Kelleher, N. L. *Nat. Prod. Rep.* **2016**, *33*, 942−950.
(12) Covington, B. C.; McLean, J. A.; Bachmann, B. O. *Nat. Prod. Rep.* **2017**, *34*, 6−24.
(13) Kind, T.; Fiehn, O. *Phytochem. Lett.* **2017**, *21*, 313−319.

DOI: 10.1021/acs.jnatprod.7b00737
J. Nat. Prod. 2018, 81, 758−767

(14) Williams, R. B.; O'Neil-Johnson, M.; Williams, A. J.; Wheeler, P.; Pol, R.; Moser, A. *Org. Biomol. Chem.* **2015**, *13*, 9957−9962.
(15) Zani, C. L.; Carroll, A. R. *J. Nat. Prod.* **2017**, *80*, 1758−1766.
(16) Munro, M.; Blunt, J. W. MarineLit http://pubs.rsc.org/marinlit/ (accessed 2016).
(17) Laatsch, H. *AntiBase, a Database for Rapid Dereplication and Structure Determination of Microbial Natural Products*; Wiley-VCH, 2010.
(18) Buckingham, J. *Dictionary of Natural Products, Supplement 4*; CRC Press: Boca Raton, FL, 1997.
(19) Horai, H.; Arita, M.; Kanaya, S.; Nihei, Y.; Ikeda, T.; Suwa, K.; Ojima, Y.; Tanaka, K.; Tanaka, S.; Aoshima, K.; Oda, Y.; Kakazu, Y.; Kusano, M.; Tohge, T.; Matsuda, F.; Sawada, Y.; Hirai, M. Y.; Nakanishi, H.; Ikeda, K.; Akimoto, N.; Maoka, T.; Takahashi, H.; Ara, T.; Sakurai, N.; Suzuki, H.; Shibata, D.; Neumann, S.; Iida, T.; Tanaka, K.; Funatsu, K.; Matsuura, F.; Soga, T.; Taguchi, R.; Saito, K.; Nishioka, T. *J. Mass Spectrom.* **2010**, *45*, 703−714.
(20) Smith, C. A.; O'Maille, G.; Want, E. J.; Qin, C.; Trauger, S. A.; Brandon, T. R.; Custodio, D. E.; Abagyan, R.; Siuzdak, G. *Ther. Drug Monit.* **2005**, *27*, 747−751.
(21) Gowda, H.; Ivanisevic, J.; Johnson, C. H.; Kurczy, M. E.; Benton, H. P.; Rinehart, D.; Nguyen, T.; Ray, J.; Kuehl, J.; Arevalo, B.; Westenskow, P. D.; Wang, J.; Arkin, A. P.; Deutschbauer, A. M.; Patti, G. J.; Siuzdak, G. *Anal. Chem.* **2014**, *86*, 6931−6939.
(22) 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*, 828−837.
(23) Watrous, J.; Roach, P.; Alexandrov, T.; Heath, B. S.; Yang, J. Y.; Kersten, R. D.; van der Voort, M.; Pogliano, K.; Gross, H.; Raaijmakers, J. M.; Moore, B. S.; Laskin, J.; Bandeira, N.; Dorrestein, P. C. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, E1743−E1752.
(24) Quinn, R. A.; Nothias, L.-F.; Vining, O.; Meehan, M.; Esquenazi, E.; Dorrestein, P. C. *Trends Pharmacol. Sci.* **2017**, *38*, 143−154.
(25) 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.; Wong, W. R.; Linington, R. G.; Zhang, L.; Debonsi, H. M.; Gerwick, W. H.; Dorrestein, P. C. *J. Nat. Prod.* **2013**, *76*, 1686−1699.
(26) Winnikoff, J. R.; Glukhov, E.; Watrous, J.; Dorrestein, P. C.; Gerwick, W. H. *J. Antibiot.* **2014**, *67*, 105−112.
(27) Nguyen, D. D.; Wu, C.-H.; Moree, W. J.; Lamsa, A.; Medema, M. H.; Zhao, X.; Gavilan, R. G.; Aparicio, M.; Atencio, L.; Jackson, C.; Ballesteros, J.; Sanchez, J.; Watrous, J. D.; Phelan, V. V.; van de Wiel, C.; Kersten, R. D.; Mehnaz, S.; De Mot, R.; Shank, E. A.; Charusanti, P.; Nagarajan, H.; Duggan, B. M.; Moore, B. S.; Bandeira, N.; Palsson, B. Ø.; Pogliano, K.; Gutiérrez, M.; Dorrestein, P. C. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, E2611−E2620.
(28) Kurita, K. L.; Glassey, E.; Linington, R. G. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112*, 11999−12004.
(29) Kellogg, J. J.; Todd, D. A.; Egan, J. M.; Raja, H. A.; Oberlies, N. H.; Kvalheim, O. M.; Cech, N. B. *J. Nat. Prod.* **2016**, *79*, 376−386.
(30) Bertrand, S.; Azzollini, A.; Nievergelt, A.; Boccard, J.; Rudaz, S.; Cuendet, M.; Wolfender, J.-L. *Molecules* **2016**, *21*, 259.
(31) Aligiannis, N.; Halabalaki, M.; Chaita, E.; Kouloura, E.; Argyropoulou, A.; Benaki, D.; Kalpoutzakis, E.; Angelis, A.; Stathopoulou, K.; Antoniou, S.; Sani, M.; Krauth, V.; Werz, O.; Schütz, B.; Schäfer, H.; Spraul, M.; Mikros, E.; Skaltsounis, L. A. *ChemistrySelect* **2016**, *1*, 2531−2535.
(32) Olivon, F.; Allard, P.-M.; Koval, A.; Righi, D.; Genta-Jouve, G.; Neyts, J.; Apel, C.; Pannecouque, C.; Nothias, L.-F.; Cachet, X.; Marcourt, L.; Roussi, F.; Katanaev, V. L.; Touboul, D.; Wolfender, J.-L.; Litaudon, M. *ACS Chem. Biol.* **2017**, *12*, 2644−2651.
(33) Brito, Â.; Gaifem, J.; Ramos, V.; Glukhov, E.; Dorrestein, P. C.; Gerwick, W. H.; Vasconcelos, V. M.; Mendes, M. V.; Tamagnini, P. *Algal Res.* **2015**, *9*, 218−226.
(34) Naman, C. B.; Rattan, R.; Nikoulina, S. E.; Lee, J.; Miller, B. W.; Moss, N. A.; Armstrong, L.; Boudreau, P. D.; Debonsi, H. M.; Valeriote, F. A.; Dorrestein, P. C.; Gerwick, W. H. *J. Nat. Prod.* **2017**, *80*, 625−633.
(35) Esposito, M.; Nothias, L.-F.; Retailleau, P.; Costa, J.; Roussi, F.; Neyts, J.; Leyssen, P.; Touboul, D.; Litaudon, M.; Paolini, J. *J. Nat. Prod.* **2017**, *80*, 2051−2059.
(36) Nothias, L.-F.; Boutet-Mercey, S.; Cachet, X.; De La Torre, E.; Laboureur, L.; Gallard, J.-F.; Retailleau, P.; Brunelle, A.; Dorrestein, P. C.; Costa, J.; Bedoya, L. M.; Roussi, F.; Leyssen, P.; Alcami, J.; Paolini, J.; Litaudon, M.; Touboul, D. *J. Nat. Prod.* **2017**, *80*, 2620−2629.
(37) Esposito, M.; Nothias, L.-F.; Nedev, H.; Gallard, J.-F.; Leyssen, P.; Retailleau, P.; Costa, J.; Roussi, F.; Iorga, B. I.; Paolini, J.; Litaudon, M. *J. Nat. Prod.* **2016**, *79*, 2873−2882.
(38) Pluskal, T.; Castillo, S.; Villar-Briones, A.; Orešič, M. *BMC Bioinf.* **2010**, *11*, 395.
(39) Protsyuk, I.; Melnik, A. M.; Nothias, L.-F.; Rappez, L.; Phapale, P.; Aksenov, A. A.; Bouslimani, A.; Ryazanov, S.; Dorrestein, P. C.; Alexandrov, T. *Nat. Protoc.* **2018**, *13*, 134−154.
(40) 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*, 741−748.
(41) Optimus https://github.com/MolecularCartography/Optimus (accessed May 24, 2017).
(42) Ihaka, R.; Gentleman, R. *J. Comput. Graph. Stat.* **1996**, *5*, 299−314.
(43) Perez, F. Project Jupyter, 2015 http://jupyter.org/about.html.
(44) Shannon, P.; Markiel, A.; Ozier, O.; Baliga, N. S.; Wang, J. T.; Ramage, D.; Amin, N.; Schwikowski, B.; Ideker, T. *Genome Res.* **2003**, *13*, 2498−2504.
(45) Appendino, G.; Szallasi, A. *Life Sci.* **1997**, *60*, 681−696.
(46) Nothias-Scaglia, L.-F.; Dumontet, V.; Neyts, J.; Roussi, F.; Costa, J.; Leyssen, P.; Litaudon, M.; Paolini, J. *Fitoterapia* **2015**, *105*, 202−209.
(47) Nothias-Scaglia, L.-F.; Schmitz-Afonso, I.; Renucci, F.; Roussi, F.; Touboul, D.; Costa, J.; Litaudon, M.; Paolini, J. *J. Chromatogr. A* **2015**, *1422*, 128−139.
(48) Esposito, M.; Nim, S.; Nothias, L.-F.; Gallard, J.-F.; Rawal, M. K.; Costa, J.; Roussi, F.; Prasad, R.; Di Pietro, A.; Paolini, J.; Litaudon, M. *J. Nat. Prod.* **2017**, *80*, 479−487.
(49) Esposito, M.; Nothias, L.-F.; Nedev, H.; Gallard, J.-F.; Leyssen, P.; Retailleau, P.; Costa, J.; Roussi, F.; Iorga, B. I.; Paolini, J.; Litaudon, M. *J. Nat. Prod.*. **2016**.79287310.1021/acs.jnatprod.6b00644
(50) 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*, 3317−3323.

DOI: 10.1021/acs.jnatprod.7b00737
J. Nat. Prod. 2018, 81, 758−767

(51) Wu, D.; Sorg, B.; Hecker, E. *Phytother. Res.* **1994**, *8*, 95−99.
(52) Evans, F. J.; Kinghorn, A. D. *J. Chromatogr. A* **1973**, *87*, 443−448.
(53) Evans, F. J.; Kinghorn, A. D. *Bot. J. Linn. Soc.* **1977**, *74*, 23−35.
(54) Appendino, G. *Prog. Chem. Org. Nat. Prod.* **2016**, *102*, 1−90.
(55) Gulakowski, R. J.; McMahon, J. B.; Buckheit, R. W., Jr.; Gustafson, K. R.; Boyd, M. R. *Antiviral Res.* **1997**, *33*, 87−97.
(56) Abdelnabi, R.; Staveness, D.; Near, K. E.; Wender, P. A.; Delang, L.; Neyts, J.; Leyssen, P. *Biochem. Pharmacol.* **2016**, *120*, 15−21.
(57) Abdelnabi, R.; Amrun, S. N.; Ng, L. F. P.; Leyssen, P.; Neyts, J.; Delang, L. *Antiviral Res.* **2017**, *139*, 79−87.
(58) Vasas, A.; Hohmann, J. *Chem. Rev.* **2014**, *114*, 8579−8612.
(59) Nothias-Scaglia, L.-F.; Pannecouque, C.; Renucci, F.; Delang, L.; Neyts, J.; Roussi, F.; Costa, J.; Leyssen, P.; Litaudon, M.; Paolini, J. *J. Nat. Prod.* **2015**, *78*, 1277−1283.
(60) Shi, Q.-W.; Su, X.-H.; Kiyota, H. *Chem. Rev.* **2008**, *108*, 4295−4327.
(61) Vasas, A.; Rédei, D.; Csupor, D.; Molnár, J.; Hohmann, J. *Eur. J. Org. Chem.* **2012**, *2012*, 5115−5130.
(62) Powers, A. *Res. Rep. Trop. Med.* **2015**, *6*, 11−19.
(63) Project Jupyter website, 2018, http://jupyter.org/install.
(64) Scheubert, K.; Hufsky, F.; Petras, D.; Wang, M.; Nothias, L.-F.; Dührkop, K.; Bandeira, N.; Dorrestein, P. C.; Böcker, S. *Nat. Commun.* **2017**, *8*, 1494.
(65) Gerlich, M.; Neumann, S. *J. Mass Spectrom.* **2013**, *48*, 291−298.
(66) Hufsky, F.; Scheubert, K.; Böcker, S. *Nat. Prod. Rep.* **2014**, *31*, 807.
(67) Dührkop, K.; Shen, H.; Meusel, M.; Rousu, J.; Böcker, S. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112*, 12580−12585.
(68) "bioassay-guided" MS/MS - Google Scholar https://scholar.google.com/scholar?q=%22bioassay-guided%22+MS%2FMS&hl=en&as_sdt=0%2C5&as_ylo=2016&as_yhi=2016 (accessed May 31, 2017).
(69) Haug, K.; Salek, R. M.; Conesa, P.; Hastings, J.; de Matos, P.; Rijnbeek, M.; Mahendraker, T.; Williams, M.; Neumann, S.; Rocca-Serra, P.; Maguire, E.; González-Beltrán, A.; Sansone, S.-A.; Griffin, J. L.; Steinbeck, C. *Nucleic Acids Res.* **2013**, *41*, D781−D786.
(70) Spicer, R. A.; Steinbeck, C. *Metabolomics* **2018**, *14*, 16.
(71) Katajamaa, M.; Miettinen, J.; Oresic, M. *Bioinformatics* **2006**, *22*, 634−636.
(72) Olivon, F.; Grelier, G.; Roussi, F.; Litaudon, M.; Touboul, D. *Anal. Chem.* **2017**, *89*, 7836−7840.
(73) Chambers, M. C.; Maclean, B.; Burke, R.; Amodei, D.; Ruderman, D. L.; Neumann, S.; Gatto, L.; Fischer, B.; Pratt, B.; Egertson, J.; Hoff, K.; Kessner, D.; Tasman, N.; Shulman, N.; Frewen, B.; Baker, T. A.; Brusniak, M.-Y.; Paulse, C.; Creasy, D.; Flashner, L.; Kani, K.; Moulding, C.; Seymour, S. L.; Nuwaysir, L. M.; Lefebvre, B.; Kuhlmann, F.; Roark, J.; Rainer, P.; Detlev, S.; Hemenway, T.; Huhmer, A.; Langridge, J.; Connolly, B.; Chadick, T.; Holly, K.; Eckels, J.; Deutsch, E. W.; Moritz, R. L.; Katz, J. E.; Agus, D. B.; MacCoss, M.; Tabb, D. L.; Mallick, P. *Nat. Biotechnol.* **2012**, *30*, 918−920.
(74) van den Berg, R. A.; Hoefsloot, H. C. J.; Westerhuis, J. A.; Smilde, A. K.; van der Werf, M. J. *BMC Genomics* **2006**, *7*, 142.

DOI: 10.1021/acs.jnatprod.7b00737
J. Nat. Prod. 2018, 81, 758−767