Isolation and Characterization of Compounds from Ochreinauclea maingayi (Hook. f.) Ridsd. (Rubiaceae) with the Aid of LCMS/MS Molecular Networking

Norfaizah Osman, Azeana Zahari, Hazrina Hazni, Wan Nurul Nazneem Wan Othman, Nurulfazlina Edayah Rasol, Nor Hadiani Ismail, Pierre Champy, Mehdi Beniddir, Marc Litaudon, Khalijah Awang

▶ **To cite this version:**

Norfaizah Osman, Azeana Zahari, Hazrina Hazni, Wan Nurul Nazneem Wan Othman, Nurulfazlina Edayah Rasol, et al.. Isolation and Characterization of Compounds from Ochreinauclea maingayi (Hook. f.) Ridsd. (Rubiaceae) with the Aid of LCMS/MS Molecular Networking. Separations, 2023, 10 (2), pp.74. 10.3390/separations10020074 . hal-04127296

HAL Id: hal-04127296
https://hal.science/hal-04127296v1

Submitted on 13 Jun 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

separations

MDPI

Article

Isolation and Characterization of Compounds from *Ochreinauclea maingayi* (Hook. f.) Ridsd. (Rubiaceae) with the aid of LCMS/MS Molecular Networking.

Norfaizah Osman[1], Azeana Zahari[1,2]*, Hazrina Hazni[1,2], Wan Nurul Nazneem Wan Othman[3], Nurulfazlina Edayah Rasol[3], Nor Hadiani Ismail[3], Pierre Champy[4], Mehdi A. Benniddir[4], Marc Litaudon[5] and Khalijah Awang[1,2]*.

[1] Department of Chemistry, Faculty of Science, Universiti Malaya, 50603, Kuala Lumpur, Malaysia
[2] Centre for Natural Products Research and Drug Discovery (CENAR), Universiti Malaya 50603 Kuala Lumpur, Malaysia
[3] Atta-ur-Rahman Institute for Natural Product Discovery, Universiti Teknologi MARA Puncak Alam Campus, Bandar Puncak Alam, Selangor, Malaysia
[4] Équipe "Chimie des Substances Naturelles" BioCIS, CNRS, Université Paris-Saclay, 5 Rue J.-B. Clément, 92290 Châtenay-Malabry, France
[5] Institut de Chemie des Substances Naturalles, CNRS, ICSN UPR 2301, Universite Paris-Sud, 91198, Gif-sur-Yvette, France.
* Correspondence: azeanazahari@um.edu.my (A.Z); khalijah@um.edu.my (K.A)
; Tel.: +603-79674064 ext 2522(A.Z)

Abstract: Phytochemical investigation of the dichloromethane crude extract of bark of *Ochreinauclea maingayi* with the aid of LCMS/MS based molecular networking guided isolation and accelerated elucidation of known and new indole alkaloids. Molecular networking analysis produces two main clusters, 41 non-prioritized clusters and self-loop nodes. Each cluster has several nodes which depict the fractions contained within those nodes. Implementation of a fraction mapping for each nodes represented the molecular weight and key fragment data of each compound. From the analysis of each cluster and nodes, we can deduce the indole alkaloids is scaffold of interest. Indole scaffold can be found between F5-F10 that contain several types of indole alkaloids. In total we have successfully purified nine indole alkaloids; 9H-β-carboline-4-carboxylate **2**, norharmane **3**, harmane **4**, naucledine **10**, neonaucline **15**, 1,2,3,4-tetranorharmane-1-one **16**, naulafine **19**, dihyrodeglycocadambine **7**, cadambine **9** and new monoterpene indole alkaloid **7** deglycocadambine from F5-F10 using chromatography technique. Their structures were confirmed by 1D-NMR, 2D-NMR, UV, IR, LCMS and MS2LDA. Several clusters and nodes contained ions could not be annotated thus suggesting that they may possess novel compounds that are yet to be discovered.

Keywords: indole alkaloid; molecular networking; dereplication; Rubiaceae; *Ochreinaculea maingayi*

Citation: To be added by editorial staff during production.

Academic Editor: Azeana Zahari and Khalijah Awang

Received: date
Revised: date
Accepted: date
Published: date

Copyright: © 2022 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Natural product studies usually deal with samples of complex chemical mixtures. Isolation and structural elucidation of the metabolites within these mixtures remain tedious and time-consuming despite the advanced performance of modern chromatographic and analytical technologies. Dereplication is one of the strategies to prioritize metabolites by having efficient isolation and elucidation using LC-MS/MS based molecular networking (MN). Using the literature and annotated molecular network as a guide, it significantly reduces the cost of drug discovery pipeline by increasing the speed of isolation of new and known compounds. Hence in this study, we describe the use of a targeted substructure-informed molecular networking screening approach to detect indole alkaloid (IA) scaffold and predict the molecular structure of possible compounds present. IA are

biosynthesized from tryptophan which undergo decarboxylation to form tryptamine. Follow by, Mannich reaction of latter and secologanin to produce strictosidine, which serves as the precursor in the production of indole alkaloids. [1, 2]. IA can be found in Apocynaceae, Icacinaceae, Nyssaceae, Gelsemiaceae, Loganiceae and Rubiaceae which have been known to exhibit various bioactivities [3-5]. In the framework of scientific collaboration between ICSN and Universiti Malaya, our laboratories have embarked on a project to find IA scaffold using MN from *Ochreinauclea maingayi* which belongs to the genus of *Ochreinauclea* (Rubiaceae). Throughout the world, there are only two species; *O. maingayi* and *O. missionis* which is understudied in phytochemical analysis [6, 7]. *O. maingayi* is a commercial timber that mostly distributed in Peninsular Malaysia, Borneo, Sumatra and Thailand [8, 9].

To execute this strategy, first, the LC-MS/MS data of crude and fractions were acquired then .RAW data LC-MS/MS were converted to .mzXML format using MSConvert (http://proteowizard.sourceforge.net). Every .mzXML file will be processed using MZmine 2 v53 to produce .mgf data [10]. The .mgf preclustered spectral data file and its corresponding.csv metadata file (for RT and integration) were exported and submit for GNPS (Global Natural Product Social Molecular Networking). Next, a molecular network will be constructed using a minimum cosine score value of 0.6 and network topK of 10 with six match fragment ions which indicate similarity in MS/MS spectra. The generated molecular network can be view and analyzed using Cytoscape (ver 3.8.2). To further prove this analysis, MS2LDA decomposes fragmentation spectra into blocks of co-occurring fragments and losses, referred to as "Mass2Motifs"[11].

2. Materials and Methods

2.1. General experimental procedures

Analytical and preparative TLC was carried out on Merck 60 F254 silica gel plates (absorbent thickness: 0.25 and 0.50 mm, respectively). Column chromatography (CC) was performed using silica gel (Merck 230–400 mesh, ASTM). IR spectra were recorded using a Perkin-Elmer Spectrum 400 FT-IR Spectrometer. NMR spectra were acquired in $CDCl_3$, C_5D_5N and CD_3OD (Merck, Germany) with tetramethyl silane as internal standard (Merck, Germany) using BRUKER Avance III 400 and BRUKER Avance III 600 spectrometers. The LC/MS profiling were identified using UHPLC-Orbitrap MS. Nitrogen was used as the sheath and auxiliary gas at flow rate of 0.25 mL/min. The optimum source (ESI) condition was spray voltage 3500V (4.5) and 2500V (3.7) kV for positive and negative respectively, S-Lens RF Level at 55.00, ESI capillary temperature 325 °C and probe heater temperature 40 °C. UV spectra were recorded using a Shimadzu 1650 PC UV–Vis Spectrophotometer. An AP-300 Automatic Polarimeter was used to measure optical rotations. All solvents were of analytical grade and were distilled prior to use.

2.2 Plant material

The stem bark of *O. maingayi* was collected from Ulu Sat Reserve Forest, Machang, Kelantan, Malaysia. The plant was identified by Prof. Colin E. Ridsdale (Leiden University, Netherland) and a voucher specimen (KL 5595) is deposited with the Universiti Malaya herbarium.

2.3 Extraction

Dried and powdered stem barks of *O. maingayi* (2.0 kg) were extracted by cold extraction procedure using three different solvents; hexane, dichloromethane (CH_2Cl_2) and methanol (CH_3OH) respectively to give three different crude extracts depending on the polarity of compounds. The dried grounded stem bark of *O. maingayi* were first defatted with hexane (10 L) for three days at room temperature. The resulting slurry was filtered, and the residual plant material was moistened with 25% ammonia solution (1 L) and left

for 2 hours to aggregate the nitrogen containing compounds in the plant. The basified residual plant material was then successively re-extracted with dichloromethane CH_2Cl_2 (15 L, 3×). CH_2Cl_2 extract was repeatedly extracted with a solution of 5% hydrochloric acid (0.5 L, 1×) until it gave a negative result for the Mayer's test. It was next basified with 25% ammonia solution to about pH 11 and re-extracted with CH_2Cl_2 (3 L, 1×) to yield 10 g of dark brown gummy CH_2Cl_2 crude. The CH_2Cl_2 extract (10 g) was subjected to silica gel column chromatography (CC), eluted with CH_2Cl_2/CH_3OH (100: 1 to 1: 100, *v*/*v*) yielded 10 fractions.

***2.3** Purification of compounds*

15 (3.2 mg, CH_2Cl_2: CH_3OH; 99:1) and **16** (2.2 mg, CH_2Cl_2: CH_3OH; 99:1) were obtained from F5 after purified using PTLC. Meanwhile F6 yielded **19** (2.8 mg, CH_2Cl_2: CH_3OH; 99:1) using PTLC. F7 was subjected to microcolumn eluted with CH_2Cl_2/CH_3OH (100: 1 to 1: 100, *v*/*v*) to give 5 fraction and then fraction 5(1) was purified with PTLC to obtained **3** (1.6 mg, CH_2Cl_2: CH_3OH; 99:1) and **4** (3.5 mg, CH_2Cl_2: CH_3OH; 99:1). F8 was purified to afford compound **2** (2.3 mg, CH_2Cl_2: CH_3OH; 98:2) by PTLC. **10** (3.6 mg, CH_2Cl_2: CH_3OH; 95:5) was obtained after purification by PTLC of F9. Purification of F10 by PTLC yielded **7** (6.0 mg, CH_2Cl_2: CH_3OH; 94:6) and **9** (110.0 mg, CH_2Cl_2: CH_3OH; 94:6) respectively.

2.4 Data dependent LCMS analysis, FBMN, and MS2LDA parameters

2.4.1 Data dependent LCMS Orbitrap MS/MS analysis

For LC-MS analysis, the crude extracts were dissolved in LC-MS grade MeOH at a concentration 0.1mg/mL. Analyses were performed on a Thermo Scientific Orbitrap Fusion Tribrid. Samples were injected and chromatographically separated on Accucore TM Vanquish C_{18}+ Dim. (2.1 x 100 mm) at a temperature 40^0C with an injection volume of 1.0 µL. Separation was achieved with a binary LC solvent system controlled by Mobile phase A consisted of 99.9% water/0.1% formic acid (LC/MS grade) and mobile phase B contained 100% MeOH (LC/MS grade), which were pumped at a rate 0.250mL/min.

2.4.2 Feature-Based Molecular Networking

The MS2 data files were converted from the.raw (Orbitrap) standard data format to .mzXML format using MSConvert Software, part of the ProteoWizard package [22]. All .mzXML were then processed using MZmine 2 v53 [10]. The mass detection was realized keeping the noise level at 5.0E5 at MS1 and at 5.0E4 at MS2. The ADAP chromatogram builder was used using a minimum group size of scan of 5, a group intensity threshold of 3.0E5, a minimum highest intensity 3.0E5, and m/z tolerance of 0.001 Da (or 5ppm). The ADAP wavelets deconvolution algorithm was used with the following standard setting: Chromatographic threshold=1%, search minimum in RT range min =0.30, minimum relative height =1%, minimum absolute height=5.0E5, minimum ratio peak top/edge=2 and peak duration range (min)= 0 to 2 min. Isotopologues were grouped using the isotopic peaks grouper algorithm with a m/z tolerance of 0.006 Da (20ppm), RT tolerance of 0.5 min. Peak alignment was performed using the join aligner module (m/z tolerance of 0.001 Da (or 5ppm), weight for RT= 2.0, absolute RT tolerance 0.3 min).The feature list row filler peak mz 100 to 1000, with keep the MS2 scan and reset the ID. The peak list was gap-filled with the same RT and m/z range gap filler module (m/z tolerance 0.001 or 5ppm). Eventually, the.mgf preclustered spectral data file and its corresponding.csv metadata file (for RT and integration) were exported using dedicated "Export/Submit for GNPS/FBMN" option.

2.4.3 Molecular Networking parameters

A molecular network was created using the online workflow (https://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=54bf2f40092c41e28f69b60e845f056d /)

on the GNPS [23] website (http://gnps.ucsd.edu). The data was filtered by removing all MS/MS fragment ions within +/- 17 Da of the precursor m/z. MS/MS spectra were window filtered by choosing only the top 6 fragment ions in the +/- 50Da window throughout the spectrum. The precursor ion mass tolerance was set to 0.02 Da and a MS/MS fragment ion tolerance of 0.03 Da. A network was then created where edges were filtered to have a cosine score above 0.6 and more than 6 matched peaks. Further, edges between two nodes were kept in the network if and only if each of the nodes appeared in each other's respective top 10 most similar nodes. Finally, the maximum size of a molecular family was set to 100, and the lowest scoring edges were removed from molecular families until the molecular family size was below this threshold. All matches kept between network spectra and library spectra were required to have a score above 0.7 and at least 6 matched peaks. The spectra in the network were then searched against GNPS' spectral libraries. The molecular networking data were analysed and visualized using Cytoscape (ver 3.8.2)[24]

2.4.4 MS2LDA Unsupervised Substructure Annotation

The fragmentation spectra were mined for co-occurring fragments and neutral losses using the latent Diriclet allocation algorithm on MS2LDA web app[25]. The following parameters were used for this task: isolation_window (0.5), min_ms2_int (500), n_its (1000), K (60), ms2_bin (0.005 Da), Zero MotiftDB. The result of the MS2LDA annotation is publicly available at the following link: https://ms2lda.org/basicviz/toggle_public/2353/ the inspection and the annotation of the Mass2Motifs were realized on the MS2LDA.org Web app.

3. Results and Discussions

The present work aims to explore and highlight the existing link between spectral molecular networks, predicted fractions, experimental fractions and how we deduce isolation and elucidation of known and unknown compounds.

Our strategy begins with analysing LC-MS^2 profile (Figure 1) of dichloromethane CH_2Cl_2 extract from the bark of *O.maingayi*, under positive ion mode. More than 1000 MS/MS spectra over the range m/z 70 to 1000 were generated. The raw data of LCMS for extract and ten fractions (fractionation of crude using column chromatography) were converted to .mgf, then mapping was done using GNPS (https://gnps.ucsd.edu/ProteoSAFe/status.jsp?task=54bf2f40092c41e28f69b60e845f056d) to generate massive molecular networking (MN) (Figure 2). The MN contain two main clusters (A and B), 41 non-prioritized clusters and self-loop nodes. Each cluster has several nodes which depict the fractions contained within those nodes. Those nodes represented the molecular weight and key fragment data of each compound. From the analysis of each cluster and nodes, we can deduce the cluster that is of interest, in this case is indole alkaloids (IA). After analysis the massive MN data (Table 1), we deduce five fractions (F5-F10) of interest contain several types of indole alkaloids. In total we have successfully purified nine indole alkaloids from F5 (neonaucline **15**, 1,2,3,4-tetranorharmane-1-one **16** [12]), F6 (naulafine **19** [13]), F7 (norharmane **3** [14], harmane **4** [15]), F8 (methyl 9H-β-carboline-4-carboxylate **2** [16]), F9 (naucledine **10** [12]), F10 (dihyrodeglycocadambine **7**, cadambine **9** [17]). Compounds **2**, **3**, **4**, **10**, **15**, **16**, **19** are non-isoprenoid tryptophan indole alkaloid which is from simple *β*-carboline (**2**, **3**, **4, 16)**, indolopyridine (**10**), nauclefine (**15**) and naulafine (**19**) type, while **7**, **9** are IA belong to cadambine type. **7** is a new monoterpene indole alkaloid successfully elucidated with MS-guided purification.

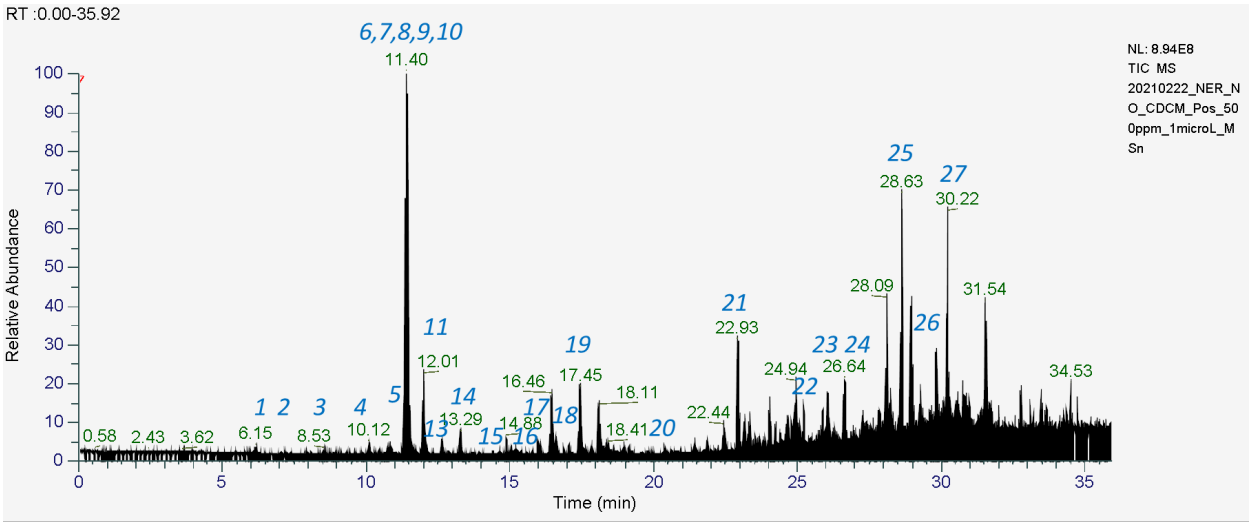

Figure 1. Dichloromethane crude extract of *Ochreinauclea maingayi* show the LCMS profile under positive mode $[M+H]^+$

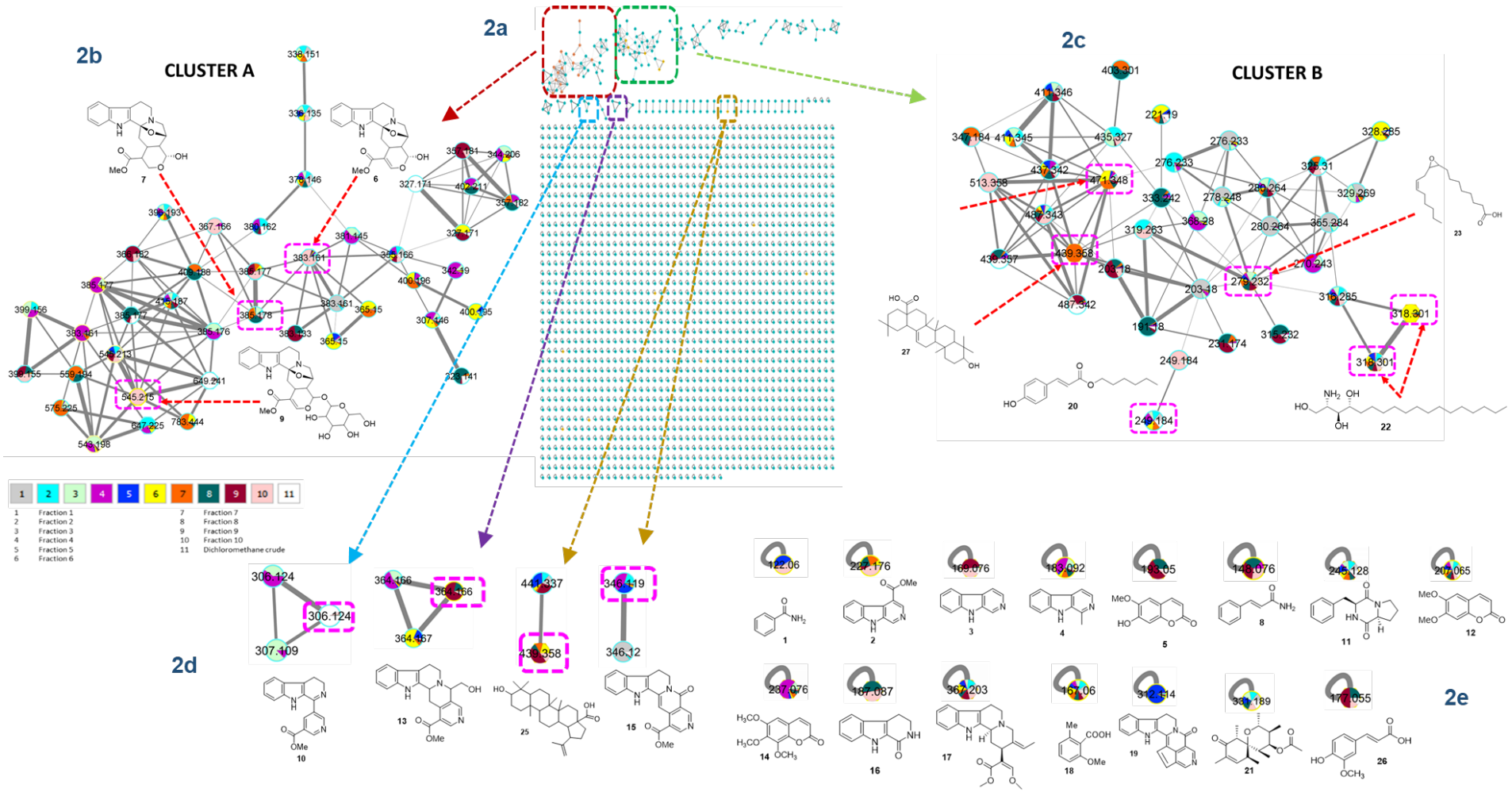

Figure 2. (a) Massive molecular networking of dichloromethane crude extract of the bark of *O.maingayi* (b) Implementation of a fraction mapping for molecular networking-guided purification for cluster A. (c) Implementation of a fraction mapping for molecular networking-guided purification for cluster B. (d) Implementation of a fraction mapping for molecular networking-guided purification for non-prioritized cluster. (e) Implementation of a fraction mapping for molecular networking-guided purification for self-loop nodes.

Table 1. Chemical constituents identified in the dichloromethane crude extract of the bark of *O.maingayi* 212

Peak No	Compound Identifications	Precursor ion mass (m/z)	Molecular Formula	RT means (min)	Ion type	Predicted fractions (Figure 3)	Experimental Fraction	Key fragment (m/z)
	IA (Cluster A, non-prioritized cluster and selfloops)							
2	Methyl 9H-β-carboline-4-carboxylate[b]	227.176	$C_{13}H_{11}N_2O_2$	7.03	$[M+H]^+$	F10,**F8**,F9,F7,F5,F6	**F8**	209,100
3	Norharmane[ab]	169.076	$C_{11}H_9N_2$	8.88	$[M+H]^+$	F10,F9,F8,F5	F7	142,115
4	Harmane[bc]	183.091	$C_{12}H_{11}N_2$	10.41	$[M+H]^+$	F6,**F7**,F8,F9,F10,F1,F2,F3,F4	**F7**	142,115
6	Deglycocadambine[d]	383.161	$C_{21}H_{23}N_2O_5$	11.15	$[M+H]^+$	F8,F9,F10,F7,F6,F4		365,337,305 ,183
7	Dehydrodeglycocadambine[b]	385.177	$C_{21}H_{25}N_2O_5$	11.23	$[M+H]^+$	F8,F9,**F10**	**F10**	367,227,183,170,144
9	Cadambine[b]	545.215	$C_{27}H_{23}N_2O_{10}$	11.43	$[M+H]^+$	F9,**F10**	**F10**	383,365,227,170,139
10	Naucledine[bc]	306.124	$C_{18}H_{16}N_3O_2$	11.75	$[M+H]^+$	F10,**F9**,F8,F6	**F9**	189,163,144
13	Cadamine[c]	364.167	$C_{21}H_{22}N_3O$	13.2	$[M+H]^+$	F6,F7,F5		346,317,285,233,144
15	Neonaucline[bc]	346.119	$C_{20}H_{16}N_3O_2$	14.39	$[M+H]^+$	F4,**F5**,F2,F7,F6	**F5**	314,286,271,231
16	1,2,3,4-tetranorharmane-1-one [ab]	187.087	$C_{11}H_{11}N_2O$	15.21	$[M+H]^+$	F2,F8,F9,F10,F6,F4,F3	F5	158,142,130,115
17	Geissoschizine methyl ether[a]	367.203	$C_{22}H_{26}N_2O_3$	15.62	$[M+H]^+$	F8,F9,F10		170,144,130,108,75
19	Naulafine[b]	312.114	$C_{20}H_{14}N_3O$	17.15	$[M+H]^+$	F5,**F6,**F4,F8,F9,F10,F2,F4	**F6**	269,242,187,170,159,141, 116

	Triterpene (Cluster B and Selfloops)							
23	Glycyrrhetic acid[a]	471.348	$C_{30}H_{47}O_4$	26.57	$[M+H]^+$	F6,F7,F5,F4,F1		189,175,133,119,107,95
25	Betulinic acid[a]	439.358	$C_{30}H_{46}O_2$	28.72	$[M+H]^+$-H_2O	F9,F10, F7,F8,F6		393,259,243,213,179,137,95
27	Oleanolic acid[a]	439.358	$C_{30}H_{46}O_2$	30.47	$[M+H]^+$-H_2O	F7,F8,F9,F10		215,203,189,147,133,119,107,95,81, 69
	Coumarine (Selfloops)							
5	Scopoletine[ab]	193.050	$C_{10}H_9O_4$	10.75	$[M+H]^+$	F4,**F3**	**F3**	150,133,122,94,77,66
12	Scoparone[b]	207.065	$C_{11}H_{11}O_4$	12.73	$[M+H]^+$	F1,**F2**,F3,F4,F5,F6,F7,F8,F9	**F2**	163,151,146,135,118,107,91
14	Di-O-methylfraxetin[a]	237.076	$C_{10}H_9O_5$	13.98	$[M+H]^+$	F4,F3,F5,F7,F8		207,179,147,133,123,91
	Phenolic (Cluster B)							
20	Hexyl-*p* coumarate[b]	249.184	$C_{15}H_{20}O_3$	20.58	$[M+H]^+$	**F1**,F2,F3,F4,F5,F6,F7,F10	**F1**	175,163,153,147,125,109,95,69
26	Trans ferulic acid[a]	177.055	$C_{10}H_{11}O_4$	29.60	$[M+H]^+$	F8,F9,F10		117,89,78
	Fatty acid (Cluster B)							
24	9(10)-EpOME[a]	279.232	$C_{18}H_{33}O_3$	26.59	$[M+H]^+$-H_2O	F8,F9,F10,F7,F1,F4,F5,F6		135,123,109,95,81,67
	Lipid (Cluster B)							
22	Phytosphingosine[a]	318.301	$C_{18}H_{40}NO_3$	25.11	$[M+H]^+$	F6,F7,F2,F4,F5		300,282,270,252,95,83

	Organic acid (Selfloops)							
11	cyclo(L-Phe-D-Pro) [a]	245.128	$C_{14}H_{17}N_2O_2$	11.90	$[M+H]^+$	F3, F2,F1,F6,F5,F7,F8,F9,F4		154,120,98,70
18	2-Methoxy-6-methyl benzoic acid[b]	167.060	$C_9H_{11}O_3$	17.11	$[M+H]^+$	F1,**F2**,F3,F4,F5,F6,F7,F10	**F2**	140,113
	Amide (Selfloops)							
1	Benzamide[bc]	122.060	C_7H_8NO	6.15	$[M+H]^+$	F5,F1,F8,F10	F4	105,95,77
8	Cinnamide[bc]	148.076	$C_9H_{10}NO$	11.28	$[M+H]^+$	F4,F8,F10,F10	F5	131,103
	Polyketides							
21	Decarboxyportentol acetate[b]	331.188	$C_{18}H_{28}O_4$	22.96	$[M+Na]^+$	F1,**F2**,F3,F5,F8,F10	**F2**	173,147,137,109,95,83,69

Peak numbers correspond to the numbering of peaks that appear in total ion chromatogram (TICs). [a] Structural hit obtained from GNPS spectral library matching. [b] Compounds have been elucidated from fractions. [C] Compounds previously reported in leaves of *O. maingayi* [12].[d] MS/MS spectra matched with published literature [18].

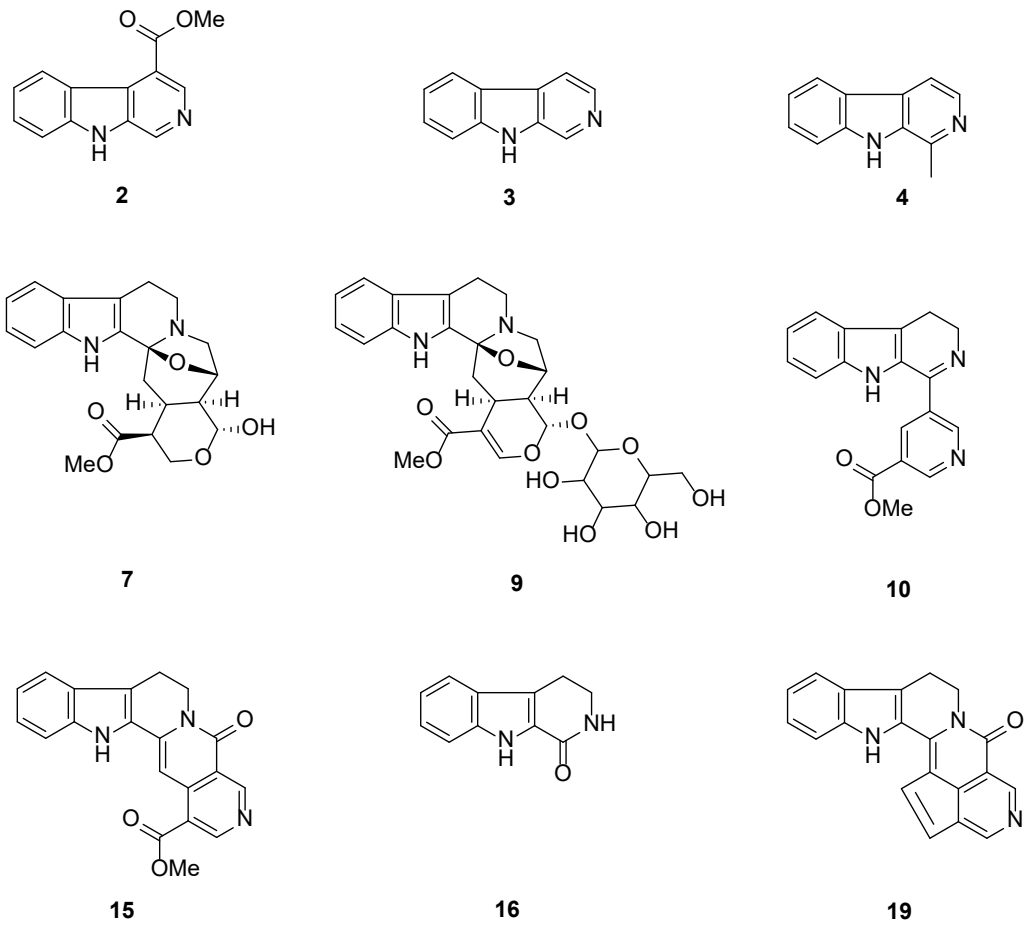

Figure 3. Compounds isolated from F5-F10; methyl 9H-β-carboline-4-carboxylate (2), norharmane (3), harmane (4), dihyrodeglycocadambine (7), cadambine (9), naucledine (10), neonaucline (15), 1,2,3,4-tetranorharmane-1-one (16), naulafine (19)

Next, to further prove our elucidation on **7**, we utilized the web-based MS2LDA [11, 19] to process our MS/MS data and explore the results through interactive visualizations. This process indicated that Cluster A exhibited a Mass2motif (motif_370) shared by 16 parent masses that possess IA scaffold. Out of these, two parent mass; Parent: 492 and Parent: 803 showed characteristic fragmentation patterns belonging to IA scaffold [20]. Details of Parent: 492 and Parent: 803 MS/MS spectra, the protonated molecular ions, and fragmentation pathway are shown in Figure 4 and Figure 5 respectively. Parent: 492 fragmentation pattern analysis led to the identification of cadambine **9** while the analysis of Parent: 830 led to discovery of new compound dihydrodeglycocadambine **7**.

Cadambine **9** (545.2153 Da) loss hexose sugar (162.075 Da) to give deglyocadambine **6** (383.1750 Da) [18]. With respect from this parent mass fragmentation that give **6** we can deduce new IA compound isolated is dihydrodeglycocadambine **7** (385.177 Da) from the loss of hexose sugar at C-21 and double bond at C16-C17 that have the similar pattern of fragmentation pathway as in cadambine **9** (Figure 4).

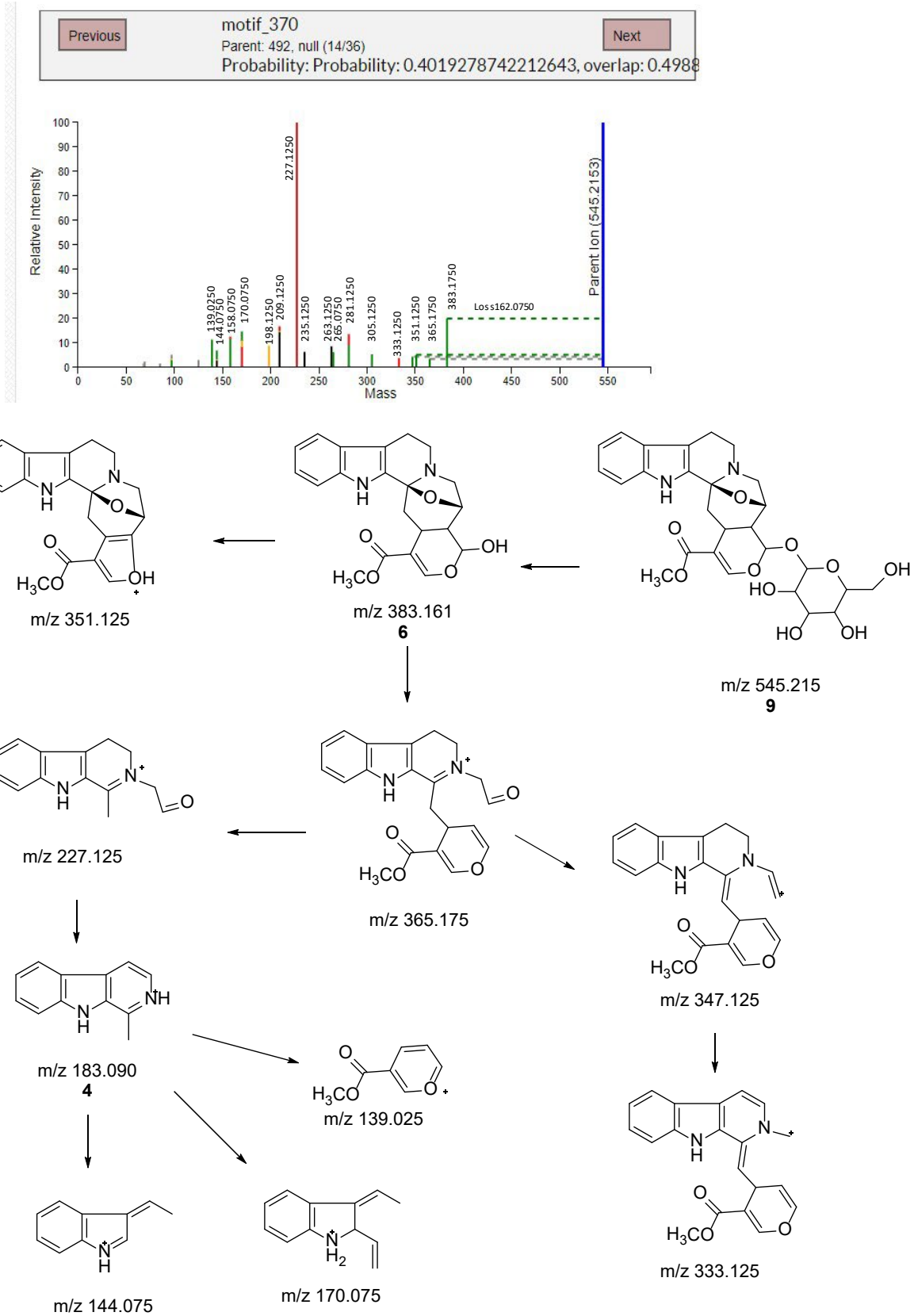

Figure 4. Mass spectra and fragmentation pathway of mass motif for parent 492

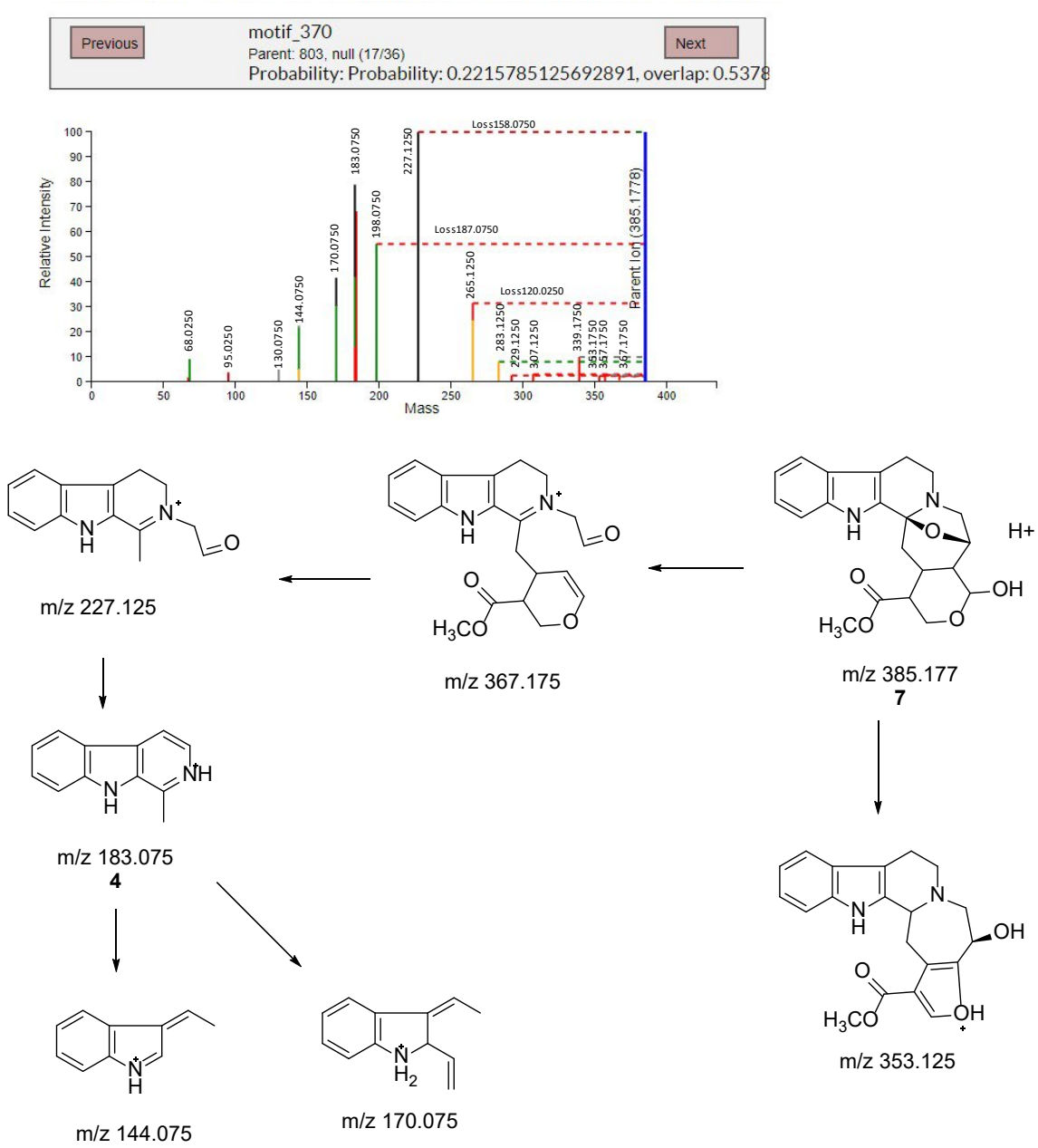

Figure 5. Mass spectra and fragmentation pathway of mass motif for parent 830

Dihydrodeglycocadambine **7** was obtained as an optically active light yellow amorphous solid $[\alpha]^{25}{}_{D}$= +57.3 (*c*=0.0004, MeOH), UV(MeOH) λmax (log ε) 222 (4.18), 274 (3.64) nm;. It has the molecular formula $C_{21}H_{23}N_2O_5$ as deduced from LC-ESI-HR-MS/MS analysis $[M+H]^+$, *m/z* 385.1754 (calcd. for $C_{21}H_{24}N_2O_5$, 385.1685). The UV spectrum revealed maximum absorption at λ_{max} 222 and 274 nm which revealed an indole chromophore [21] while IR gave broad band at 3339.6cm^{-1} due to O-H stretching and a sharp strong band at 1731.0 cm^{-1} due to C=O group. Analysis of the ^{1}H NMR and HSQC indicated four aromatics signal characteristic of unsubstituted indole moiety appeared at δ_H 7.74, (d, *J*=7.7) H-9, δ_H 7.28, (t, *J*=7.5), H-10, δ_H 7.34, (t, *J*=7.4), H-11 and δ_H 7.67, (d, *J*=8.1),H-12.

The ^{13}C NMR and DEPT indicated a total of twenty-one carbons in the structure which belong to five quaternary carbons (δ_c 135.1,C-2; 92.3,C-3; 111.1 C-7; 127.2, C-8; 138.4,

C-13), nine methines (δ_c 119.9,C9; 119.7,C10; 122.7,C11; 111.4,C12; 29.8,C15; 48.9,C16; 74.8,C19; 50.3,C20), five methylenes (δ_c 53.2,C-5; 22.8,C-6; 47.9,C-14; 67.8,C-17; 57.3,C-18), one hemiacetal carbon (δ_c 97.1,C21), one methoxy group (δ_c 52.0,C-23) and one carbonyl groups (δ_c 173.0,C-22).

Further analysis of ^{1}H and ^{13} C NMR spectroscopic data (Supplementary 1, Table 2) indicated that **7** was structurally related to **9**, except for the following three differences; first the absence of glucosyl moiety at C-21, which suggested that **7** was the aglycone to **9**, second the absent of double bond at C16-C17, and thirdly the presence of an ester group at C-16 (δ_c 48.9). The presence of ester group at C-16 is further proven by the existence upfield methine δ_H 2.65, H-16 and appearance oxygenated methylene at δ_H 3.97 (m) H-17α, δ_H 4.33(dd, *J*=11.5,4.3) H-17β.

The location of substituents and rings arrangement were further confirmed by cross peaks in HMBC spectra (Supplementary 1, Figure 6). The correlation of H-16/C-22 (δ_c 173.0), OMe-/C-22 (δ_c 173.0) and H-16/C-20 (δ_c 50.3) proving the ester group situated at δ_c 48.9, C-16. The cross peaks between H-17/ C-21 (δ_c 97.1), H-20/C-21 indicated location of hydroxyl group at C-21. Finally, the linkages between ring C, D, E were assigned with the aid of the HMBC spectrum. The cross peak between H-18/ C-5 and H-14/ C-3 further confirmed that indole ring (A-C) and D were fused via C-3-N-4 junction. The cross peak between H-18/C-20, H-16/C-20 inferred that ring D was connected through tetrahydropyran ring E through C-20.

The relative stereochemistry of **7** was confirmed from the NOESY spectrum and correlation (Figure 7) by comparison of its correlations with those of similar skeleton [17]. The stereochemistry at C-20 could be ascertained by the NOESY relationship between H19 and H-20. The presence of chiral carbon at C-19 indicated by *O*-bridged between C-3 and C-19 that exist as β-configuration. This β-configuration led to the assignment of proton at C-19 exist as α-configuration. NOESY spectrum H-19 correlated with H-20 as α-configuration. In addition, H-20 correlated with H-21 in COSY spectrum, however there is no correlation between H-20/H-21 in NOESY spectrum, thus it can be deduced that H-21 have different configuration with H-20 and assigned as β-configuration. There are two geminal proton for C-17 (H17α and H17β). H-17β showed correlation with H-21β. It also proved that H16 exist as α -configuration due to the direct correlation of H17α and H16 α. There is direct correlation between H-15/H-16 and H-15/H-20, therefore we can presume that H-15 also exist as α-oriented.

The ^{1}H NMR and ^{13}C NMR (600 MHz) spectral assignments performed by extensive 2D-NMR experiments (COSY, HSQC and HMBC) and summarized in Figure 6 and Table 2.

Table 2. ^{1}H-NMR (600 MHz) and ^{13}C-NMR (150 MHz) Spectral Data of dihydrodeglycocadambine 7 in C_5D_5N

	Dihydrodeglycocadambine 7		**Cadambine 9**		**Cadambine [17]**	
position	δ_H (*m*, *J* in Hz) in Pyridine, C_5D_5N	δ_C in Pyridine, C_5D_5N	δ_H (*m*, *J* in Hz) in Pyridine, C_5D_5N	δ_C in Pyridine, C_5D_5N	δ_H (*m*, *J* in Hz) in CD_3OD	δ_C in CD_3OD
NH	12.48 (*br s*)					
2		135.1		134.6		133.2
3		92.3		92.1		93.0
5	2.83 (*m*) overlap 3.18 (*d*, *J*= 8.1)	53.2	3.01 (*m*) 2.74 (*m*)	53.0	3.16 (*m*) 2.79 (*m*)	53.9
6	2.83 (*m*) overlap	22.8	2.80 (*m*) 4.63 (*dd*, *J*=3.0, 12.2)	22.8	2.80 brs	22.8
7		111.1		110.8		111.6
8		127.2		127.0		127.0
9	7.74 (*d*, *J*=7.6)	119.9	7.69 (*d*, *J*=7.7)	119.9	7.47 (*d*, *J*=7.8)	120.2
10	7.28 (*t*, *J*=7.6)	119.7	7.23 (*t*, *J*=7.7)	119.9	7.00 (*t*, *J*=7.8)	120.1
11	7.34 (*t*, *J*=7.6)	122.7	7.27 (*t*, *J*=7.7)	122.8	7.11 (*t*, *J*=7.8)	123.4
12	7.67 (*d*, *J*=7.6)	111.4	7.51 (*d*, *J*=7.7)	112.6	7.33 (*d*, *J*=7.8)	112.6
13		138.4		138.4		136.3
14a	2.28 (*dd*, *J*=12.0,12.0) 2.09 (*dd*, *J*=12.0, 4.7)	47.9	2.38 (*dd*, *J*=5.9, 12.7)	44.1	2.06 (*m*)	43.1
14b			2.33 (*t*, *J*=12.7)			
15	2.83 (*m*) overlap	29.8	3.48 (*m*)	26.41	3.26 *(m)*	26.9
16	2.65 (*m*)	48.9		110.8		111.3
17a	4.33 (*dd*, J=11.2,4.3) 3.97 (*t*, *J*=11.2)	67.8	7.65 (*s*)	153.4	7.57 s	154.4
17b						
18a	2.83 (*m*) overlap	57.3	3.14 (*d J*=10.2)	59.4	3.51 (*brd J*=10.8)	59.4
18b	2.62 (*d*, *J*=10.1)		2.86 (*m*)		3.02 (*dd J*=10.8,7.3)	
19	5.22 (*dd*, *J*=6.2, 2.0)	74.8	5.10 (*d J*=7.0)	73.8	4.94 (*brd J*=7.3)	74.6
20	2.20 (*m*)	50.3	1.78 (*t J*=7.0*)*	40.9	1.76 (*m*)	41.1
21	5.04 (*d*, *J*=8.6)	97.1	5.82 (*d J*=9.2)	98.2	5.84 (*d J*=9.3)	97.6
22-C=O		173.0		167.6		168.9
23-OCH3	3.49 (*s*)	52.0	3.60 (*bs*)	51.5	3.65 s	51.9
1′			5.19 (*d J*=7.8)	103.0	4.80 *d J*=7.9	101.7
2′			4.16 (*m*)	75.4	3.30 (*m*)	74.9
3′			4.30 (*t J*=8.9)	78.0	3.42 *t J*=9.0	78.1
4′			4.20 (*m*)	72.4	3.28 (*m*)	71.7
5′			3.97 (*dddd J*=2.4, 7.0)	79.0	3.34 (*m*)	78.5
6′			4.48 (*dd J*=11.8, 2.4) 4.17 (*m*) overlapp	63.6	3.87 (*dd J*=12.1, 2.3) 3.61 (*dd J*=12.1, 6.4)	62.9

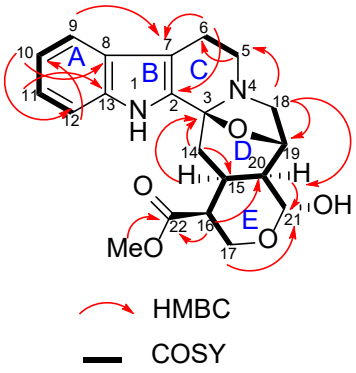

Figure 6. Selected key correlation of dihydrodeglycocadambine **7**

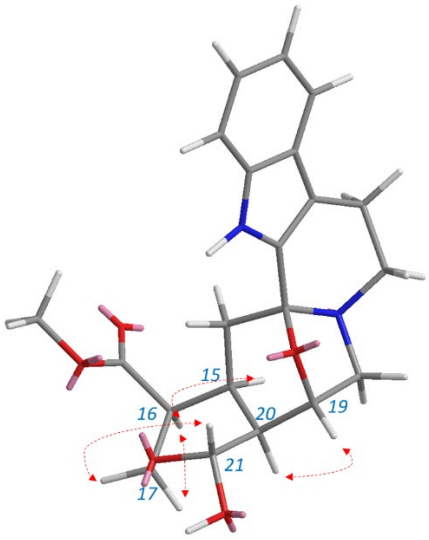

Figure 7. Selected NOESY correlation of dihydrodeglycocadambine **7**

5. Conclusions

The combination of MN, analysis of the fragmentation patterns and manual dereplication proved to be a useful strategy for facilitated dereplication of IA scaffold from *O. maingayi*. The MN not only aided the detection of the presence of the new compound; dihydrodeglycocadambine **7**, but also enables the selection of fraction containing targeted compounds, detection and identification of known IA scaffold. A total of nine indole alkaloids, two coumarins and one fatty acid were positively identified in bark of *O. maingayi*. Moreover, several clusters contained ions that could not be annotated suggesting that they may also possess novel compounds that are yet to be discovered.

Supplementary Materials: The following supporting information can be downloaded at: www.mdpi.com/xxx/s1, Figure S1: LCMS of Dihyrodeglycocadambine (**7**); Figure S2: ^{1}H NMR spectrum (600 MHz, C_5D_5N) of Dihyrodeglycocadambine (**7**); Figure S3: ^{13}C NMR spectrum (600 MHz, C_5D_5N) of Dihyrodeglycocadambine (**7**); Figure S4: COSY spectrum (600 MHz, C_5D_5N) of Dihyrodeglycocadambine (**7**); Figure S5: DEPT 90 spectrum (600 MHz, C_5D_5N) of Dihyrodeglycocadambine (**7**); Figure S6: HSQC spectrum (600 MHz, C_5D_5N) of Dihyrodeglycocadambine (**7**); Figure S7:

HMBC spectrum (600 MHz, C_5D_5N) of Dihyrodeglycocadambine (**7**); Figure S8: NOESY spectrum (600 MHz, C_5D_5N) of Dihyrodeglycocadambine (**7**).

Author Contributions: N.O. and A.Z. contributed to the experiments and analyses, participated in the discussion of results and wrote the manuscript; K.A. participated in material preparation and manuscript writing; H.H., N.E.R, and N.I. participated in experiments; W.N.W.O., P.C., M.D., and M.L. provided intellectual input in the analysis of the results. All authors have read and agreed to the published version of the manuscript.

Funding: This work was financial supported by the Ministry of Higher Education (MOHE) under the Fundamental Research Grant Scheme (FRGS) FRGS/1/2020/SKK0/UM/01/5), FP054-2020 and University Malaya Research Fund Assistance (BKP) BK040-2017. International Funding, French-Malaysian Collaborative grant on Phytochemical Analysis on Malayan Flora, project number 57-02-03-1007 were used for travel grant on molecular networking studies.

Data Availability Statement: Not applicable

Acknowledgments: The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article.

Conflicts of Interest: Declare conflicts of interest or state "The authors declare no conflict of interest."

References

1. O'Connor, S.E. and J.J. Maresh, *Chemistry and biology of monoterpene indole alkaloid biosynthesis.* Natural Product Reports, 2006. **23**(4): p. 532-547.
2. Szabó, L.F., *Rigorous biogenetic network for a group of indole alkaloids derived from strictosidine.* Molecules, 2008. **13**(8): p. 1875-96.
3. Škubník, J., et al., *Vincristine in Combination Therapy of Cancer: Emerging Trends in Clinics.* Biology (Basel), 2021. **10**(9).
4. Shamon, S.D. and M.I. Perez, *Blood pressure-lowering efficacy of reserpine for primary hypertension.* Cochrane Database Syst Rev, 2016. **12**(12): p. Cd007655.
5. Rtibi, K., et al., *Vinblastine, an anticancer drug, causes constipation and oxidative stress as well as others disruptions in intestinal tract in rat.* Toxicol Rep, 2017. **4**: p. 221-225.
6. Dalal, N. and R. Ravishankar, *in vitro propagation of Ochreinauclea missionis (Wall. EX G. Don), an ethnomedicinal endemic and threatened tree.* In Vitro Cellular & Developmental Biology - Plant, 2001. **37**: p. 820-823.
7. Ravishankar, R., *Genetic fidelity in micropropagated plantlets of Ochreinauclea missionis an endemic, threatened and medicinal tree using ISSR markers.* Afr J Biotechnol, 2009. **8**.
8. Ridsdale, C.E., *A revision of the tribe Naucleeae s.s. (Rubiaceae).* Blumea: Biodiversity, Evolution and Biogeography of Plants, 1978. **24**(2): p. 307-366.
9. Lim, S., K. Gan, and C. Kheng Ten, *Identification and utilisation of lesser-known commercial timbers in Peninsular Malaysia 1: Ara, Bangkal, Bebusok and Bekoi.* 2022.
10. Pluskal, T., et al., *MZmine 2: Modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data.* BMC Bioinformatics, 2010. **11**(1): p. 395.
11. Van der Hooft, J.J.J., et al., *Topic modeling for untargeted substructure exploration in metabolomics.* Proceedings of the National Academy of Sciences, 2016. **113**(48): p. 13738-13743.
12. Mukhtar, M.R., et al., *Neonaucline, a new indole alkaloid from the leaves of Ochreinauclea maingayii (Hook. f.) Ridsd. (Rubiaceae).* Molecules, 2011. **17**(1): p. 267-74.
13. Repke, D.B., et al., *Synthesis of naucléfine, angustidine, angustine, (±)-13b,14-dihydro-angustine and naulafine.* Tetrahedron, 1989. **45**(9): p. 2541-2550.
14. Parameswaran, P.S., C.G. Naik, and V.R. Hegde, *Secondary Metabolites from the Sponge Tedania anhelans: Isolation and Characterization of Two Novel Pyrazole Acids and Other Metabolites.* Journal of Natural Products, 1997. **60**(8): p. 802-803.
15. Seki, H., A. Hashimoto, and T. Hino, *The ^{1}H and ^{13}C-Nuclear Magnetic Resonance Spectra of Harman. Reinvestigation of the Assignments by One- and Two-Dimensional Methods.* Chemical & Pharmaceutical Bulletin, 1993. **41**(6): p. 1169-1172.
16. Lai, Z.-Q., et al., *Seven Alkaloids from Picrasma quassioides and Their Cytotoxic Activities.* Chemistry of Natural Compounds, 2014. **50**(5): p. 884-888.
17. Yuan, H.L., et al., *Anti-inflammatory and analgesic activities of Neolamarckia cadamba and its bioactive monoterpenoid indole alkaloids.* J Ethnopharmacol, 2020. **260**: p. 113103.
18. Wu, X.-D., et al., *Two New Indole Alkaloids from Emmenopterys henryi.* Helvetica Chimica Acta, 2013. **96**(12): p. 2207-2213.

19. Wandy, J., et al., *Ms2lda.org: web-based topic modelling for substructure discovery in mass spectrometry.* Bioinformatics, 2018. **34**(2): p. 317-318.
20. Gai, Y., et al., *Analysis of the traditional medicine YiGan San by the fragmentation patterns of cadambine indole alkaloids using HPLC coupled with high-resolution MS.* J Sep Sci, 2013. **36**(23): p. 3723-32.
21. Albinsson, B. and B. Norden, *Excited-state properties of the indole chromophore: electronic transition moment directions from linear dichroism measurements: effect of methyl and methoxy substituents.* The Journal of Physical Chemistry, 1992. **96**(15): p. 6204-6212.
22. Chambers, M.C., et al., *A cross-platform toolkit for mass spectrometry and proteomics.* Nature Biotechnology, 2012. **30**(10): p. 918-920.
23. Wang, M., et al., *Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking.* Nature Biotechnology, 2016. **34**(8): p. 828-837.
24. Otasek, D., et al., *Cytoscape Automation: empowering workflow-based network analysis.* Genome Biol, 2019. **20**(1): p. 185.
25. Wandy, J., et al., *Ms2lda.org: web-based topic modelling for substructure discovery in mass spectrometry.* Bioinformatics, 2017. **34**(2): p. 317-318.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.